

A Robotic Framework for Controlling Human's Attention

Mohammed Moshui Hoque^{1*}

¹ Graduate School of Science and Engineering, Saitama University, 255 Shimo-Okubo, Sakura-Ku, 338-8570 Saitama, Japan

Received 2012/08/13

Accepted 2012/11/22

Abstract

It is a major challenge in HRI to design a social robot that is able to selectively direct a target human's attention towards an intended direction. For this purpose, the robot may first turn its gaze toward him/her in order to establish eye contact. However, such a turning action of the robot may not in itself be sufficient to make eye contact with the target person in all situations, especially when the robot and the person are not facing each other or the human is intensely engaged in a task. In this paper, we propose a conceptual model of attention control with five phases: attention attraction, eye contact, attention avoidance, gaze back, and attention shift. Evaluation experiment by using a robotic head reveals the effectiveness of the proposed model in different viewing situations.

Keywords

human-robot interaction · attention control · eye contact

1. Introduction

We would like to facilitate the increased introduction of robots into the human world. For robots to be accepted in the real world, they must be capable of behaving in ways analogous to those used by humans in their interactions with each other. It is certain that a number of challenges remain unsolved [1] in the realm of natural HRI. However, a robot that is able to control a particular human's attention away from his/her current attentional focus in a desired manner is also an important research issue. Controlling someone's attention is a fundamental skill in human social interaction and cognition. This ability plays a critical role in a wide range of social behaviors; it sets the stage for learning, develops the capacity for mutual understanding, and facilitates communication [2]. Establishing eye contact plays an important role in initiating an interaction process between two partners [3]. However, it is difficult for the initiating agent to establish such a communication channel through non-verbal means when the responding agent is not looking at his/her face, or when the responding agent is intensely focused on his/her task.

For instance, consider the scenarios illustrated in Fig. 1 where the initiating agent (R) would like to control (i.e., shift) the target participant's (TP's) attention to the *goal direction* (GD). Under these constraints, we consider how a robot can attract a human's attention depending on the relative position between the human and the robot. The human field of view is wide and is divided into central and peripheral visions. In these scenarios, we illustrate the positional relationship between the human and the robot in terms of the robot's position within the human's field of view, that is to say, where the robot is seen from the perspective of the human. We consider four positional relationships in which the robot is seen: (i) in the central field of view (CFOV) (Fig. 1 (a)), (ii) in the near peripheral field of view (NPFOV) (Fig. 1 (b)), (iii) in the far peripheral field of view (FPFOV) (Fig. 1 (c)), and (iv) in the out of field of

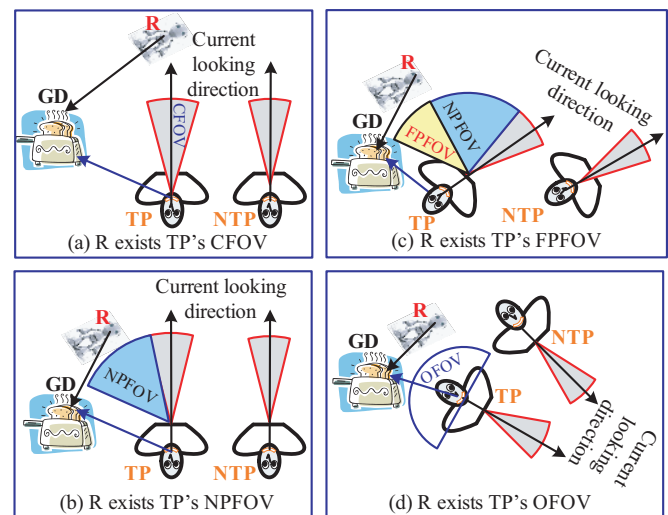


Figure 1. Scenarios where the robot (R), the target participant (TP), and the non-target participant (NTP) are not facing each other. R would like to shift the TP's attention toward the GD. Red cones indicate the central viewing zone. Black and blue arrows indicate the current attentional focus of the TP and R, while red arrows indicate the intended goal direction.

view (OFOV) (Fig. 1 (d)) of the target participant. In all these situations except for CFOV, it is difficult to establish mutual gaze. Thus, the initiating agent should display some form of pro-active behaviors (verbal or non-verbal) to attract the target human's attention and prompt him/her to look towards it, in order to easily make eye contact.

It is readily apparent that a robot can initiate an interaction process with humans through utilizing voice or sound cues, but vocal utterances are also certain to attract the attention of others who are not intended tar-

*E-mail: moshuiul@cv.ics.saitama-u.ac.jp

gets. However, in this work our main concern is to design an attention control process for the social robot that can attract and control a particular person's attention. In other words, we would like the robot to be able to selectively attract only the target person while avoiding attracting other people's attention as much as possible. Thus, non-verbal behaviors are ideal candidates for our purpose. Besides voice, non-verbal behavior has been emphasized as another effective channel of communication [4]. However, it is apparent that visual stimuli offered by the robot's nonverbal behaviors cannot affect a person if he/she is in a position from which he/she cannot see the robot. In this situation (such as the scenario represented in Fig. 1 (d)), the use of touch or sound cues should be considered a last resort.

For initiating an interaction process, the robot should pro-actively approach the intended human when the intended human and the robot are not initially facing each other. This pro-active approach enables the robot to help people who have potential needs and convey some information about an object or a particular direction in which the human should focus. In order to ensure that the robot has acquired attention control of the target agent, it must inform him/her in a socially acceptable manner that it (the robot) intends to direct his/her attention to an object requiring attention. Simultaneously, the robot must inform the remainder of the group that it has no intention of controlling their attention. Thus, the question becomes: how can the robot effectively and reliably convey this intention to the target member of a group through its attention control process? In order to enable the robot to control the target person's attention, we propose an attention control approach that consists of five phases: attention attraction, eye contact, attention avoidance, gaze back, and attention shift. We hypothesize that if the robot performs these phases consecutively, it will be able to successfully direct the target participant's attention in the necessary direction.

2. Related Work

The capabilities of social robots that can control human attention are still at a rudimentary stage. Moreover, there is yet to be significant work conducted about how humans attract the attention of others to initiate an interaction, beyond the primary facts that they stop at a certain distance [5], start the interaction with a greeting [6, 7], and arrange themselves in a spatial formation [8]. Several previous HRI studies have addressed the use of greeting behaviors to initiate human-robot conversation [9, 10]. Some robots were equipped with the capability to encourage people to initiate interaction by offering cues such as approach direction [11], approach path [12], and standing position [13]. Robotic systems have also been developed that are able to establish eye-contact by gaze crossing [14–16] and to shift a human's attention by gaze turning [17, 18], reference terms, and pointing gestures [19, 20]. 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. Although such a passive attitude can work in some situations, many situations require a robot to employ a more active approach [21–23]. In this paper, we consider some situations in which a target human and a robot are not arranged face-to-face. In such situations, it is difficult to establish mutual gaze through basic non-verbal action alone (e.g., the robot turning its head toward the target person [24]).

The use of attention avoidance behaviors in the realm of human-robot interaction is limited. Yoshikawa et al. [25] used the head away action of the robot to produce its gaze avoidance behavior. However, this work dealt with the single participant and considered a face-to-face situation. In addition to that, they did not consider the case of a collaborative situation between the robot and the participant. A recent work

[26] has used the *no look cue* for the robot as the gaze avoidance behavior in a multiple person environment. Their robot does not actively orientate its head towards the target participant and it lowers its head from horizontal. The *no look cue* has been designed to convey a coarse message, such as 'the robot is not looking at anyone'. However, the issue of how a robot behaves if the non intended participant looks at the robot in place of the intended participant it remains to be unexplored. Shiomi et al. [27] implemented a gaze back function in Robovie in which it gazed in certain directions (looked away) for a few seconds and then looked back to the human face. In the Ksera project [28] a robot, Nao, uses a motion tracking algorithm for its gaze back function. This system dealt with the single participant and it did not describe how the robot should revert its gaze to the target participant after averting its gaze from the non-target participant.

In our previous work, we demonstrated that a robot's head shaking actions can attract human attention [29], and moreover that a robot can shift a human target's attention in a desired direction by using head directed cues [30]. However, we dealt only with cases involving a single participant in general viewing conditions, and the success rate was limited due to the limitations of the robot's attention attraction and attention shift behaviors. In this paper, we focus on the following issues that go beyond our previous research: (i) we present an approach to attention control that can selectively attract and control a particular participant's attention from multiple participants; (ii) We develop a robotic platform that can avoid the gaze of non-target participants in order to not attract their attention; and finally (iii) the robot can gaze back if the target participant responds later than a non-target, and can establish eye contact with him/her.

3. Attention Control in HRI

In this section, we operationalize the attention control process. We also describes how this process can apply in an HRI platform to control the target human's attention.

3.1. An Attention Control Process

Although eyes are the primary source of information about the direction of attention, there are many situations where determining the viewer's eye direction is impossible or infeasible. Especially when people are further apart or are not facing each other, head orientation becomes a stronger cue than information from the eyes in determining direction of attention [31, 32]. This explains 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 the person's eye gaze, because the head is a much larger visual stimulus than the eyes, and so the social signals associated with the facing direction of the head are still quite accessible in peripheral vision despite the lower visual acuity [33]. In our attention control process, we employ head direction to monitor the focus of the participants' attention. Thus, we may define attention control as a means of gaze (i.e., head) control, in which one can shift someone's gaze from an initial direction to another direction as intended by the initiator. A canonical attention control episode, between the robot and the target participant, can be described in five major steps: attention attraction, eye contact, attention avoidance, gaze back, and attention shift. Fig. 2 shows a simple depiction of the attention control process.

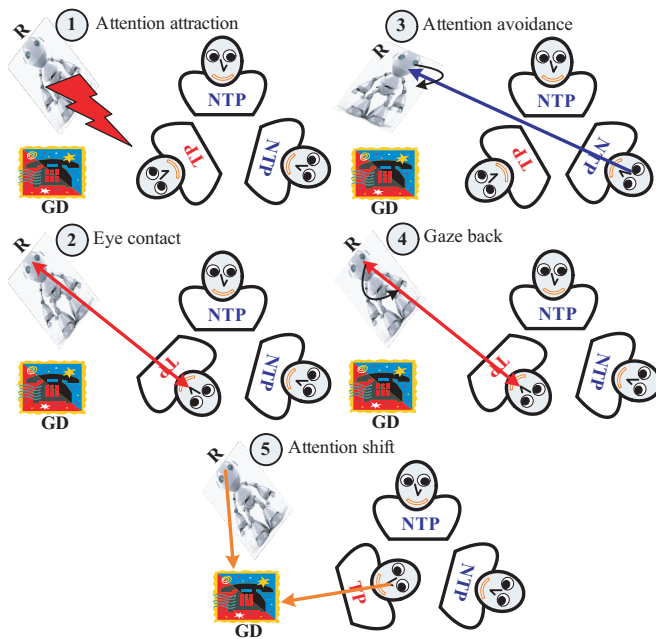


Figure 2. Attention control process in interaction. Direction of arrows indicates the agent's attentional focus.

Attention attraction

This refers to one's ability to capture another's attention, shifting it from its existing direction towards him/her for establishing a communication channel through eye contact. Humans usually first turn their head toward the person with whom they would like to communicate. If the target human does not respond, he/she will attempt to gain their attention with stronger signals/cues (e.g., waving a hand, shaking the head, moving the body, utilizing sound cues or even using their voice). In a natural human-robot interaction scenario, robots should utilize the same conventions as humans. In other words, the robot should start with a weak action to avoid attracting the attention of individuals other than the target person, and then use stronger actions if its initial attempt fails to attract his/her attention. Based on surveying previous literature [14, 34], we decided to have our robot turn its head (to look towards the person) as the weakest action. If the robot does not attract the attention of the target agent through this head turn, we decided that it would next employ head shaking as a stronger action because object motions effectively draw people's attention [35]. Peripheral vision is good at detecting motion [36], and abundant psychological evidence shows that abrupt motions are effective to capture human attention toward the object in peripheral view [37, 38].

In this work, we propose these two kinds of motions (i.e. head turning, and head shaking) of the robot for the attention attraction process. Any kind of motion may attract both participant's (target and non-target) attention. Thus, the robot should observe the current viewing situation of the target participant (TP) before applying its attention attraction action. This is because previous studies [39] have shown that the success of attracting someone's attention by an action depends on his/her viewing situation. This study also revealed that the robot can attract the target participant's attention alone from multiple participants. For example, the head turning action of the robot can attract the TP's attention alone when the TP observes the robot in his/her central or near peripheral

field of view situation, and the non-target participant (NTP) observes it in his/her far peripheral or out of field of view situation. In that case, the robot does not need to activate or use its attention avoidance component.

Eye contact

Attracting the attention of the target participant can produce observable behavioral responses such as eye or head movements, or shifts in body orientation. Therefore, if the target participant is attracted by the robot's behavior, he/she will turn toward it, which will make establishing eye contact simple. It appears that humans can make eye contact if they establish gaze crossing, that is, looking at each other's eyes or face [32]. Psychological studies show, however, that this gaze crossing action alone may not be sufficient to establish eye contact. Gaze-awareness is also important because it enables humans to feel that they have made eye contact with each other [40]. Thus, to be most effective the robot should also display gaze awareness explicitly through some actions (e.g., facial expressions or nodding) after gaze crossing.

Attention avoidance

A gaze avoidance phenomenon occurs when a human (H1) avoids looking at another human (H2), especially if H1 is aware of being looked at, and/or moves their head away from that person (H2) [32, 33]. Examples of behaviors employed by humans to show gaze aversion include closing the eyes, or turning the eyes or head away [41]. Previous studies [39] demonstrate that a simple head-turning action of the robot can attract a non-target participant's attention in some cases. For examples, when the robot exists in the central or near peripheral field of view situation of the NTP and the far peripheral or out of field of view situation of the TP. In that case, the NTP experiences a stronger pull on his/her attention from the robot than does the TP. Moreover, in this case, the robot would need to use a strong action (i.e., head-shaking in our current design) to capture the TP's attention, making attracting the NTP's attention even more likely.

It is thus certain that in some situations the attention of the non-target participant can be gained, prompting them to look at the robot before the intended target of the action does so, due to his/her larger response delay. In these situations, it is necessary for the robot to display an explicit behavior to the non-target participant to convey that it has no intention of initiating communication with him/her. The robot should therefore be capable of behavior by which non-target agents can easily and readily infer that it is not willing to start an interaction with them. This gaze avoidance behavior of the robot will therefore convey the feeling of non-communicative intention.

Gaze back

This phenomenon occurs when a target agent looks at the initiating agent later than a non-target agent. Naturally, there will be times when a non-target agent may respond earlier than the intended target to a robot's actions. The target agent may respond within a brief interval or he/she may respond during the gaze avoidance phase. In that case, the robot should return its gaze to the target agent and continue its attention control process.

Attention shift

This involves an agent gazing at or turning toward the object referred to by another agent. The robot carries out an attention shift request by switching the target agent's attention (via eye gaze, head turn, shift in body orientation, pointing gesture, or reference terms [42]) to the object which needs to be addressed. The target agent then responds to the

request by attending to the referential focus. However, after addressing the focus, the initiating agent checks whether or not the attention shift attempt is successful within the expected time-frame. If the target agent is not yet attending to the referential object, the robot normally tries to seek their attention with different strategies or employs the same behavior again. Finally, an attention shift will only be successful if both interacting agents direct their attention to the same target.

3.2. Our Approach

An attention control event is executed by a finite-state-machine model as shown in Fig. 3. To initiate the attention control, the robot begins to observe the current direction of the agent's attention by tracking his/her head. After recognizing the viewing situation of the target agent (TA), the robot usually turns its head first toward the TA, and commences shaking its head up to three times (if necessary) to attract his/her attention. For the head turning action, we adjusted the pan speed of the pan-tilt unit at $120^\circ/\text{second}$. For the head shaking action, the robot shook its head back and forth ($\pm 30^\circ$) from its initial position. This meant that the robot turned its head 30° left and 30° right. The head-shaking speed was adjusted at $240^\circ/\text{second}$. However, the robot waits about 4 seconds after each attempt for the TA to respond by looking in its direction. Psychological studies have shown that silences of more than 4 seconds become embarrassing because they imply a break in the thread of communication [43].

If the robot is successful in attracting the TA's attention, the two agents will experience gaze crossing. Thus, the robot considers the TA to have responded to its actions if he/she looks at the robot within the expected time-frame. It is able to recognize whether this is so by detecting the front of his/her face in the camera image. The robot performs a blinking action to display gaze awareness. A previous study has shown that blinking is one of the most important cues for forming a person's impressions [44]. The robot's eye blinks are produced by the rapid closing and opening of the eyelids depicted in the CG images, and displayed through the LED projector on the robot's eyes. The TA feels that he/she has made eye contact with the robot, and comes to expect the initiation of an attention shift phase. The robot meets the TA's gaze, and blinks its eyes in response to only them in the event that both target and non-target agents are looking at it simultaneously. The robot avoids the attention of non-target agents (NTA) by turning its head away by -120° at a pan speed of $180^\circ/\text{second}$ in the event that an NTA looks at the robot before the TA does. If the TA responds later (within 4 seconds) or during the gaze avoidance period, the robot turns its head back to its previous position, and continues to seek to control his/her attention.

The robot can be judged to have successfully acquired control of the TA's attention when the latter is gazing at or turning toward the object referred to by the robot. In our design, we have implemented the attention shift component of the robot in the form of it first directing its eyes (up to 3 times) and then turning its head (if necessary) toward the relevant object. After addressing the focus based on each cue, the robot waits for up to 2 seconds and checks whether or not the attention shift attempt has been successful. In a successful outcome, both the TA and the robot are looking at the same object, prompting the robot to generate a beep to indicate attention control success (this is for experimental purposes). Otherwise, the robot will try again by shifting its eyes or head back toward the TA, and once again focus on the object. The robot considers the case to be a failure if it cannot shift the TA's attention within two attempts of employing head turning.

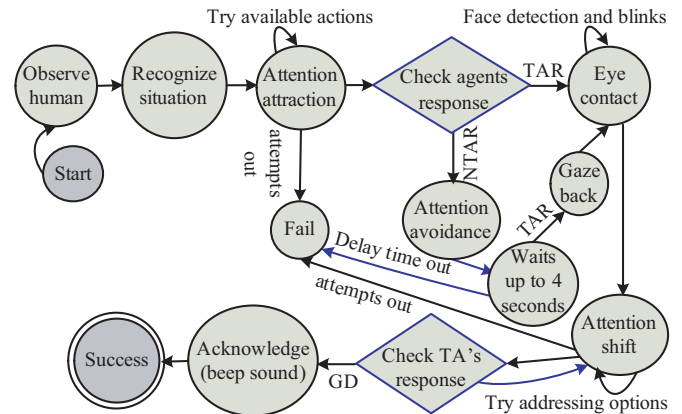


Figure 3. Behavioral protocol of the attention control process. The acronyms TAR, NTAR, and GD represents the target agent's response, the non-target agent's response, and goal direction respectively. Arrows indicate the information flow from initial state to goal state.

4. System Overview

For the HRI experiments we developed a robot head, an overview of which is given in Fig. 4. This figure shows a prototype robotic head and the corresponding outputs of its several software modules. The head consists of a spherical 3D mask, an LED projector (3M pocket projector, MPro150), a laser range sensor (URG-04LX by Hokuyo Electric Machinery), three USB cameras (Logitech Inc., Qcam), and a pan-tilt unit (Directed Perception Inc., PTU-D46). The LED projector projects CG-generated eyes on the mask as in [45]. In the current implementation, one USB camera is attached to the robot's head (as shown in Fig. 4). The other two cameras and the laser sensor are affixed to the tripods, placed at an appropriate position for observing participants' heads as well as their bodies.

The system utilizes two general purpose computers (Windows XP), each connected to a USB camera. One PC is installed with the head detection and tracking modules (HDTM) for tracking the head of the non-target participant. To detect, track, and compute the direction of the participant's head in real time (30 frame/sec), we use FaceAPI [46] by Seeing Machines Inc. It can measure 3D head direction within a 3° error at 30 frame/second. The other PC is installed with three software modules, namely HDTM, the body tracking module (BTM), and the robot control module (RCM). This PC detects and tracks the bodies of multiple people in the range sensor data using the BTM. The human body model is consequently represented with the center coordinates of an ellipse $[x, y]$ and its principle axis direction (θ). These parameters are estimated in each frame by a particle filter framework. BTM can locate a human body within errors of 6 cm for position, and 6° for orientation. A detailed description of the BTM is provided in [47]. Both PC's are connected through a wired network. The RCM consists of four sub-modules: a situation recognition module (SRM), face-detection module (FDM), eye-blinking module (EBM), and pan-tilt unit control module (PTUCM). The RCM integrates all sensor processing results of both computers.

To assess the current situation (i.e., where the human is currently looking), one observes the head direction estimated by the HDTM. From the results of the HDTM, the SRM recognizes the existing viewing situation (with 99.4% accuracy) and the direction of the relevant object in terms of yaw (α), and pitch (β) movements of the head by using a

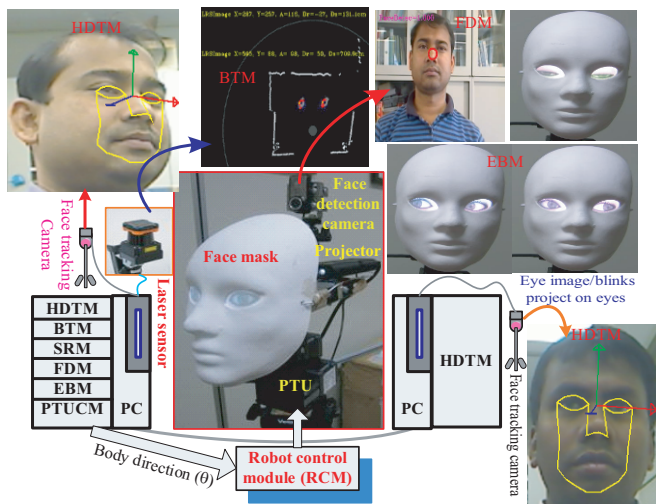


Figure 4. A prototype robotic head that consists of six modules: HDTM (head detection and tracking module), SRM (situation recognition module), BTM (body tracking module), FDM (face detection module), EBM (eye blinks module), and PTUCM (pan-tilt unit control module).

set of predefined rules. For each rule, we set the values of (α) , and (β) by observing several experimental trials. For example, *if the current head direction (of the human with respect to the robot) is within $-10^0 \leq \alpha \leq +10^0$ and $-10^0 \leq \beta \leq +10^0$ and remains in the same direction for 30 frames, system recognizes the situation as the central field of view*. The results of the SRM are sent to the PTUCM to initiate the attention control process, and the robot turns toward the human based on the results provided by the BTM. The robot considers that the participant has responded successfully to its actions if the TA looks at the robot within the expected time-frame. If this step is successful, the FDM uses the image from the forehead camera to detect the front of the target participant's face [48]. After detecting the target agent's face, the FDM sends the results to the EBM for exhibiting eye blinks to let the human know that the robot is aware of his/her gaze. The EBM generates the eye blinks at a rate of 1 blink/second up to three times because previous studies [49] revealed that three times blinking of the robot are appropriate to create the feeling of making eye contact. All the robot's head actions are performed by the PTU, with the actual control signal coming from several modules.

5. Experiment

We conducted an experiment to verify the effectiveness of the five-step attention control process for directing the target human's attention in the direction intended by the robot.

5.1. Participants

A total of thirty six graduate students (30 males, 6 females) from Saitama University participated in the experiment. Their ages ranged from 22 to 35 with an average of 26.8 ($SD = 3.72$). Twenty males and two females participants had engineering majors such as electrical engineering and civil engineering, while ten males and four females came from a science background including biological, computer, and

information sciences. There was no remuneration paid to the participants. Two participants interacted at a time, one of them interacting with each robot as a target and the another as a non-target.

5.2. Design and Procedure

As a low attention-absorption task, we considered a scenario: ‘watching the paintings’. Fig 5 (a) shows the schematic setting of the experiment.

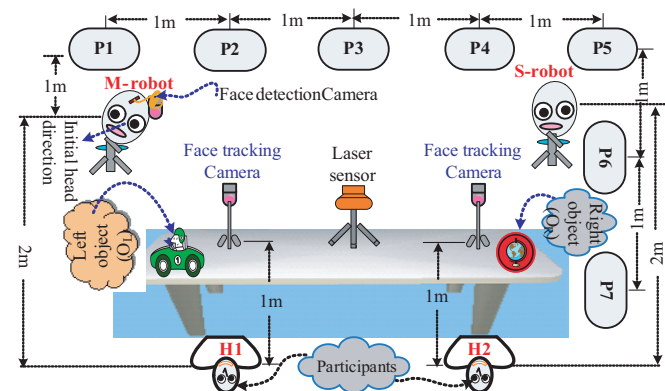


Figure 5. Schematic setting of the experiment where P1-P7 indicates the paintings, M-robot means the moving robot, and S-robot means the static robot. H1 and H2 represents the human 1 and human 2.

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. Only the eyes of the robotic head are retro-projected and the rest of the face is opaque. 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 [50]. Thus, two robots were placed in the participant's left and right monocular fields of view. One robot is called the static robot (S-robot) and it was stationary at all times. The other is called moving robot (M-robot) and it performed the attention control process. The placement of robots was exchanged randomly. Two objects were placed on the table as attention shift targets.

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 pair of participants were asked to sit down on chairs and to look around at the paintings. Since the positions of the robots were fixed, the participants detected them in their various fields of view as they moved their heads and bodies around. The participants took part in six trials in each condition. The moving robot randomly chose one of the participants as the target and the other as the non-target for the first 3 trials (each trial corresponds to each viewing direction). In the last 3 trials, the roles of the participants were exchanged.

Before running the experiment, participants were shown a demo behavior by the experimenter to watch the paintings by moving the head. 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 three types of viewing situations: the central field of view (e.g., looking at

pictures P1 or P2), the near peripheral field of view (e.g., looking at pictures P3, P4 or P5), and the far peripheral field of view (e.g., looking at pictures P6 or P7). The moving robots initiated the attention control process after recognizing the viewing situation. In this experimental scenario, if the target participants did not gaze at the robot or the object within the expected time-frame following the robot's actions, then the robot considered the case to be a failure and completed the experiment. Fig 6 depicts some images from the experiment. Two video cameras were placed in appropriate positions to capture all interactions during the experiment.

5.3. Experimental Conditions

The experiment had a within-participant design, and the order of all experimental trials was counterbalanced. It is important to compare the proposed robot with another one, that lacks eye blinks and gaze back functions in order to assess how these functions affect the performance of an interactive task. Moreover, it is also necessary to evaluate the effectiveness of the attention avoidance function by comparing the proposed robot to a robot that lacks this function. Thus, the proposed model was compared with three alternative methods. That means we programmed the moving robot in four different ways including the proposed robot.

- **Method 1 (proposed):** This utilized our proposed method as outlined in section 3.2.
- **Method 2:** The robot did not blink its eyes and would stay at its current position for about 3 seconds when the target participant (TP) looked at it. The robot did not move its head from its current position when the non-target participant (NTP) looked at it before the TP. In other words, the robot did not seek to avoid the NTP's attention.
- **Method 3:** The robot performed eye blinks and gaze avoidance behaviors as in method 1. However, it did not divert its gaze from a gaze avoidance position if the TP looked at it after the NTP did.
- **Method 4:** In order to attract attention, the robot always applied the head turning action regardless of the viewing situation. If the TP was attracted by its action, the robot made eye contact with him/her by blinking its eyes. If only the NTP responds before the TP, it avoided his/her gaze by turning its head away. The robot returned its head toward the TP and made eye contact with him/her, if s/he responded within 4 seconds. The robot would meet the gaze, and blink its eyes only at the TP, if both were looking at it at the same time.

For all conditions, the robot demonstrated a basic behavior: after gaze crossing with the target participant, it attempted to shift his/her attention toward a particular object (to the left or the right).

5.4. Evaluation Measures

Subjective measures

We asked participants to fill out a questionnaire after interactions were complete. The measurement was a simple rating on a Likert scale of 1 to 7. The questionnaire contained two items. The first question was evaluated based on the target participant's scores and the second question was evaluated based on the non-target participant's scores.

- **Impression of attention control:** Did you think that the robot was successful in controlling your attention toward the direction it intended?

- **Impression of communicative intention:** Did you think that the robot had no intention to communicate with you during the interaction?

Quantitative measures

For overall evaluation, we measured the following item by observing the experimental videos.

- **Success ratio (SR):** This refers to the object (that the robot indicated through its first eye shifting cue) being looked at by the target participant. This was measured by the ratio of the number of target participants looking at the object and the number of times that the robot attempted to shift their attention to that object.

5.5. Hypotheses and Predictions

We proposed a robot that seeks to control a target participant's attention with attention attraction, eye contact, and attention shift functions. Moreover, it is also necessary to perform the attention avoidance function and gaze back functions while multiple people are present in the interaction scenarios. While three of the methods try to control people's attention, they lack some functions, which are implemented in the proposed method. Thus, our hypotheses state that if the robot is successful in performing the five-step process, the proposed method produce the most appropriate interaction for controlling human attention. Based on this consideration, we expected that the following predictions (P1-P3) would be verified by the experiment.

- **P1:** Target participants perceive that the proposed robot is better at controlling their attention.
- **P2:** Non-target participants feel that the proposed robot more effectively conveys its non-communicative intention in all four conditions.
- **P3:** The proposed method outperforms the other three methods for the overall evaluation.

6. Results

Fig 7 shows the results of the questionnaire assessment. We conducted a repeated-measure of analysis of variance (ANOVA) on the target participant scores. The results show that the differences among conditions were statistically significant for impression of attention control ($F(3, 105) = 27.23, p < 0.01, \eta^2 = 0.52$). Multiple comparisons with the Bonferroni method show significant differences between pairs (method 1 vs. method 2: $p < 0.001$, method 1 vs. method 3: $p < 0.001$, method 1 vs. method 4: $p = 0.0001$). However, no significant differences were found between other pairs (method 2 vs. method 3: $p = 0.13$, method 2 vs. method 4: $p = 0.86$, method 3 vs. method 4: $p < 0.09$). Fig. 7 (a) also illustrates these results which verify prediction 1.

Concerning the impression of communicative intention, a significant main effect was found ($F(3, 105) = 76.43, p < 0.01, \eta^2 = 0.77$) (Fig 7 (b)). Here too the Bonferroni method comparisons revealed significant differences between pairs (method 1 vs. method 2: $p < 0.001$, method 1 vs. method 4: $p < 0.01$). But there are no significant difference was found between the methods 1 and 3 ($p = 0.75$). This happen because of the same gaze avoidance behaviors displayed by the robots in methods 1 and 3. These results partially verified prediction 2.

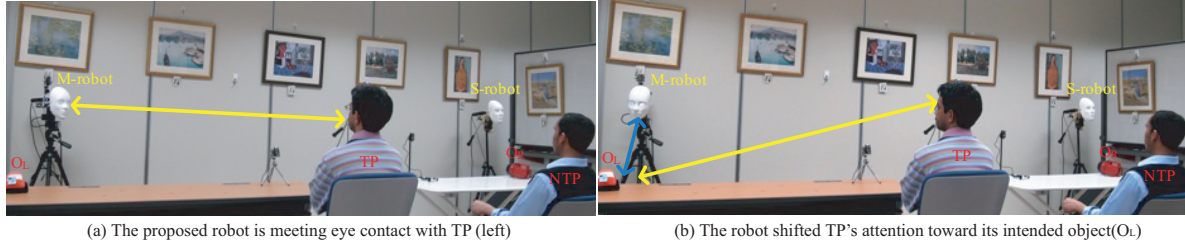


Figure 6. Experimental scenes. TP, NTP, O_L , O_R represents the target participant, the non-target participant, left object, and right object respectively.

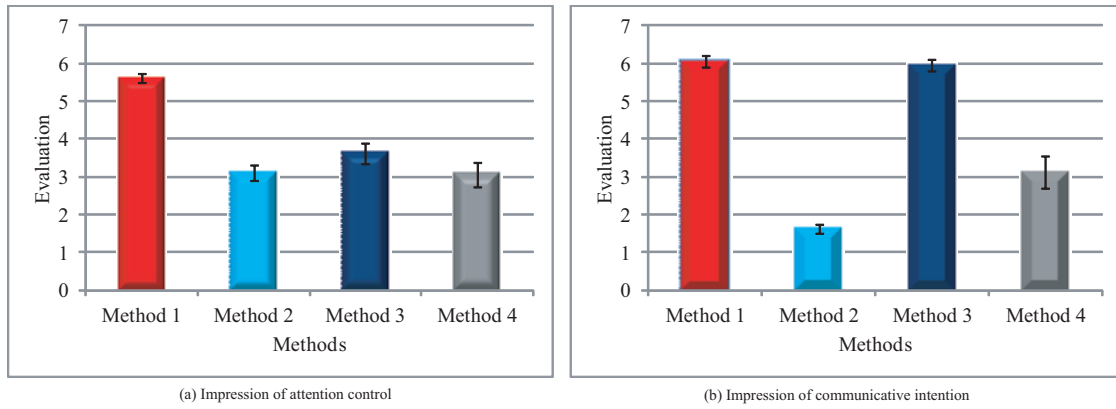


Figure 7. Results of subjective evaluation measures. The error bars indicate the standard deviation (SD).

All 36 participants interacted in 3 trials as the target with 4 different methods. Thus, we observed a total of 432 ($36 \times 3 \times 4$) interactions of the target participants for all conditions. Table 1 shows the results of the quantitative measurements.

Table 1. Quantitative measures. M and SD indicates the mean and the standard deviation of the success ratio (SR) in each condition

Measures	Conditions			
	Method 1 (proposed)	Method 2	Method 3	Method 4
SR				
M	2.72	1.22	1.56	1.78
SD	0.46	0.73	0.51	0.43
%	90.7	40.7	51.8	59.3

We conducted a repeated measures of analysis of variance (ANOVA) on the success ratio that showed significant differences between the conditions ($F(3, 105) = 23.22, p < 0.01, \eta^2 = 0.52$). We conducted multiple comparisons with the Bonferroni method that showed significant differences between methods 1 and 2 ($p < 0.001$), between methods 1 and 3 ($p < 0.001$), between methods 1 and 4 ($p < 0.01$), and between methods 2 and 4 ($p < 0.05$) while the difference between other pairs (methods 2 and 3: $p = 0.08$, and methods 3 and 4: $p = 0.24$) were not significant.

Results also revealed that a substantial 90% (100% for CFOV, 97% for NPFOV, and 75% for FPFOV) of target participants' attention was controlled by the proposed method, while only 40% (47% for CFOV, 41% for NPFOV, and 33% for FPFOV), 51% (61% for CFOV, 50% for NPFOV, and 44% for FPFOV), and 59% (92% for CFOV, 83% for NPFOV, and 2% for FPFOV) of their attention was controlled by methods 2, 3, and 4 respectively. These results mean that the attention control performance of the proposed robot is clearly more effective, in terms of producing a higher success ratio, when it employs the head shaking, blinking and gaze back actions. Therefore prediction 3 has been verified by the experiment.

7. Discussion

The experimental results reveal that the proposed method can effectively control a target participant's attention in terms of initially attracting his/her attention, making eye contact, and finally shifting his/her attention in the intended direction. We found that the robot was able to capture its intended target's attention by turning its gaze to him/her and shaking its head (if necessary).

Results also reveal that the proposed method can control participant's attention more effectively than the other methods. The head turning action of method 4 was not effective compared to the head shaking action of the proposed robot to attract the participants' attention in all cases, especially when they were looking at paintings P6 and P7 (i.e., the robot existed in his/her far peripheral field of view). That means,

the action of turning the head toward the participants is effective up to the NPFOV situation. More stronger actions (such as head shaking in this study) may be needed when the human is looking at it in the far peripheral field of view. This may happen because abrupt motions are effective at drawing human attention toward an object in peripheral vision.

Gaze back and eye blinking actions proved helpful to relay 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 explained why the human participant could quickly identify the robot's shift in attention focus. 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. Moreover, the robot was successful at conveying its non-communicative intention to the non-target participants using the gaze avoidance function. Although we need further study using more participants, the preliminary results of this experiment confirm that the attention avoidance cue is useful to convey to non-targets the fact that the robot has no intention of interrupting their current focus and directing their attention elsewhere.

Although our purpose is to develop a robot to control a target person's attention in a friendly manner, any kind of motion may annoy people to some extent. Especially since our robot produces mechanical sounds in making actions, unlike humans, people may be surprised by this sound and feel disturbed. In this paper, we used a single pattern of head shaking action for a stronger attention-attraction cue. We need to consider the size and speed of the action to design a less annoying and more acceptable behavior for robots. In the experiments, the robot turned its head right and left by 30° at a speed of $240^\circ/\text{second}$, which we do not think was too fast or large a motion. In fact, we did not observe any surprising behaviors of the participants during the head shaking actions. Moreover, we measured the sound level in the head shaking action to check whether or not the sound was too obtrusive. We measured the maximum sound pressure level over 10 seconds while the robot was iterating the head shaking action, from the participant's position at a height of 1.1 meters. The mean value of the sound level was 45.1 db (std=0.60) with the background noise (mean=41.3, std=0.35). The result suggested that the noise with the action may not be too obtrusive. We chose the head shaking action as an example of strong action with noticeable motions. We will investigate other actions such as raising or waving a hand in the future.

Although there may be many possible applications for attention control capabilities of robots, this paper demonstrates how robots can control a target human's attention alone among multiple humans, from his/her current attentional direction toward the robot's goal direction when s/he is not looking toward the robot initially or s/he is engaging such a task that does not require much attention. The robot may be applied this capability when serving as a guard, information provider, or personal assistance in the office environment where people are usually engaged in their work.

7.1. Limitations

There are still several issues that have not been addressed in the current model. Firstly, we limited the robot's behavior to head and eye movements only. However, robots may need to use other bodily actions depending on the situation. Robots may also need to use voice or sound cues in some situations, although this may attract the attention of others as well as the target person. Secondly, we tested our model for a particular scenario where two people were engaged in a task that demanded a relatively low level of attention. The strategies required to attract and control only a target human's attention when the target 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. Thirdly, we must consider timing issues. Robots do not need to consider the question of timing if the need to gain a person's attention is urgent. However, robots should usually consider timing and decide upon an appropriate time at which to interrupt a human. A robot should be able to observe a target person to find an occasion where his/her attention toward the current work decreases, and then take such an opportunity to attract his/her attention. The robot should wait for an opportunity when it is not in a hurry or emergency and the participant is engaged in important work. The robot should continuously monitor the mental states or the level of attention of the target participant during his/her involvement in a task and try to attract his/her attention when the level of attention is lowest. Fourthly, 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 two participants at a time. Therefore, the robot should interpret more peoples' responses and respond appropriately to them to cope with a real world situation. Moreover, building real-time interactivity into social robots will require combining speech and nonverbal behavior recognition, generation and cognitive representations of the world that adapt to new input from users and the environment. Lastly, the behavior of the robot was hard-wired and the robot worked on a simple rule-base. Human behaviors are different in different contexts. Thus, it is very difficult to choose proper parameters for generating natural non-verbal behaviors of the robot without extensive human-human interaction studies. Learning approaches may be used as an alternative to generate better non-verbal behaviors of robots with less effort. Some recent works [51, 52] used unsupervised learning techniques to design the robot's non-verbal behaviors and these systems can dynamically update their behavioral protocol directly from human-human interaction. Lee et al. [53] used supervised learning techniques for the robot to perform contingency detection. Although the adaptability and naturalness of the systems are much better, these systems require extensive amount of training data in different contexts to train the system.

8. Conclusion

The primary focus of our work is to develop a robot that can control a particular person's attention among multiple persons by non-verbal means. For this purpose, we have proposed an integrated attention control approach that consists of five phases. Although there may be various non-verbal behaviors, we incorporated head turning and shaking, eye blinking, turning the head away, turning the head back, and shifting the eyes and/or the head in respective phases. We have shown that our method can function to control a target human's attention in a situation where he/she is not initially looking toward the robot and is involved in a task that does not demand much attention to perform. Although several improvements are needed, we believe that our proposed approach has paved the way for controlling human attention by robots.

References

- [1] A. Weiss, T. Scherndl, M. Tscheligi, A. Billard, Evaluating the ICRA 2008 HRI Challenge. In: ACM/IEEE International Conference on Human-Robot Interaction, San Diego, USA, (ACM Press, New York, 2009), 261

- [2] G. O. Deak, I. Fasel, J. Movellan, The Emergence of Shared Attention: Using Robots to Test Developmental Theories. In: International Workshop on Epigenetic Robotics, Lund, Sweden, (Lund University Cognitive Studies, 2001), 95
- [3] T. Striano, V. M. Reid, S. Hoehl, Neural Mechanisms of Joint Attention in Infancy. *The Euro. J. Neurosci.* 23, 2819 (2006)
- [4] M. L. Patterson, *Nonverbal Behavior: A Functional Perspective*, (Springer-Verlag, New York, 1983)
- [5] E. T. Hall, *The Hidden Dimension: Man's Use of Space in Public and Private*, (The Bodley Head Ltd., London, 1966)
- [6] E. Goffman, *Behavior in Public Place*, (The Free Press, New York, 1963)
- [7] A. Kendon, Features of the Structural Analysis of Human Communicational Behavior, In: W. R. Engel (Ed.), *Aspects of Nonverbal Communication* (Swets and Zeitlinger, Lisse, 1980)
- [8] A. Kendon, Spatial Organization in Social Encounters: The Formation System, In: A. Kendon (Ed.), *Conducting Interaction: Patterns of Behavior in Focused Encounters* (Cambridge University Press, Massachusetts, 1990) 209
- [9] K. Hayashi, D. Sakamoto, T. Kanda, M. Shiomi, S. Koizumi, H. Ishiguro, T. Ogasawara, N. Hagita, Humanoid Robots as a Passive-Social Medium-A Field Experiment at a Train Station. In: *ACM/IEEE International Conference on Human Robot Interaction*, Arlington, Virginia, USA, (ACM Press, New York, 2007), 137.
- [10] M. P. Michalowski, S. Sabanovic, R. Simmons, A Spatial Model of Engagement for a Social Robot. In: *IEEE International Workshop on Advanced Motion Control*, Istanbul, Turkey, (IEEE, 2006), 762.
- [11] K. Dautenhahn, M. Walters, S. Woods, K. L. Koay, C. L. Nehaniv, How May I Serve You?: A Robot Companion Approaching A Seated Person in A Helping Context. In: *ACM/IEEE International Conference on Human Robot Interaction*, Salt Lake City, Utah, USA, (ACM Press, New York, 2006), 172
- [12] C. Shi, M. Shimada, T. Kanda, H. Ishiguro, N. Hagita, Spatial Formation Model for Initiating Conversation, In: *International Conference on Robotics: Science and Systems*, Los Angeles, USA, (USC, Los Angeles, 2011)
- [13] F. Yamaoka, T. Kanda, H. Ishiguro, N. Hagita, A Model of Proximity Control for Information Presenting Robot. *IEEE Tran. on Robot.*, 26, 187 (2010)
- [14] B. Mutlu, T. Shiwa, T. Kanda, H. Ishiguro, N. Hagita, Footing in Human Robot Conversations: How Robots Might Shape Participant Roles using Gaze Cues, In: *ACM/IEEE International Conference on Human Robot Interaction*, San Diego, CA USA, (ACM Press, New York, 2009), 61
- [15] C. Breazeal, Toward Sociable Robots. *J. of Robo. & Autono. Sys.*, 42, 162 (2003).
- [16] D. Miyauchi, A. Nakamura, Y. Kuno, Bidirectional Eye Contact for Human-Robot Communication. *IEICE Trans. of Info. & Syst.*, 88-D, 2509 (2005).
- [17] Z. Yucel, A. A. Salah, C. Mericli, T. Mericli, Joint Visual Attention Modeling for Naturally Interacting Robotic Agents, In: *International Symposium on Information and Computer Sciences*, METU, North Cyprus, (IEEE, 2009), 242
- [18] N. Kubota, T. Shimizu, M. Abe, Joint Attention Between a Human Being and a Partner Robot Based on Computational Intelligence. *J. of Adv. Comput. Intelli. & Inte. Info.*, 11, 1274 (2007).
- [19] O. Sugiyama, T. Kanda, M. Imai, H. Ishiguro, N. Hagita, Y. Anzai, Human-Like Conversation with Gestures and Verbal Cues Based on Three-Layer Attention Drawing Model. *J. Conn. Sci.*, 18, 379 (2006).
- [20] G. Heidemann, R. Rae, H. Bekel, I. Bax, H. Ritter, Integrating Context-Free and Context-Dependant Attentional Mechanisms for Gestural Object Reference. *J. of Mac. Vis. & App.*, 16, 64 (2004).
- [21] S. Satake, T. Kanda, D. F. Glas, M. Imai, H. Ishiguro, N. Hagita, How to Approach Humans? Strategies for Social Robots to Initiate Interaction. In: *ACM/IEEE International Conference on Human Robot Interaction*, San Diego, CA USA, (ACM Press, New York, 2009), 109
- [22] M. Buss, D. Carton, B. Gonsior, K. Kuehnlenz, C. Landsiedel, N. Mitsou, Towards Proactive Human-Robot Interaction in Human Environments. In: *International Conference on Cognitive Infocommunications*, Budapest, Hungary, (IEEE, 2011), 1
- [23] A. Cesta, G. Cortellessa, V. Giuliani, F. Pecora, R. Rasconi, M. Scopelliti, L. Tiberio, Psychological Implications of Domestic Assistive Technology for the Elderly. *Psychology J.*, 5, 229 (2007).
- [24] M. Finke, K. L. Koay, K. Dautenhahn, C. L. Nehaniv, M. L. Walters, J. Sauters, Hey, I'm Over Here-How can a Robot Attract People's Attention? In: *IEEE International Workshop on Robot and Human Interactive Communication*, Nashville, USA, (IEEE Press, New York, 2005), 7
- [25] Y. Yoshikawa, K. Shinozawa, H. Ishiguro, N. Hagita, T. Miyamoto, Impression Conveyance with Responsive Robot Gaze in a Conversational Situation. In: *IEEE International Symposium on Robot and Human Interactive Communication*, Hatfield, UK, (IEEE Press, New York, 2006), 457
- [26] N. Kirchner, A. Alempijevic, G. Dissanayake, Nonverbal Robot-Group Interaction Using an Imitated Gaze Cue. In: *ACM/IEEE International Conference on Human-Robot Interaction*, Lausanne, Switzerland, (ACM Press, New York, 2011) 497
- [27] M. Shiomi, T. Kanda, N. Mirallers, T. Miyashita, Face-to-Face Interactive Humanoid Robot. In: *IEEE/RSJ International Conference on Intelligent Robots and Systems*, Sendai, Japan, (IEEE, 2004), 1340
- [28] D. V. Pol, J. Juola, L. Meesters, C. Weber, A. Yan, S. Wermter, Deliverable D3.1: Human Robot Interaction. <http://www.ksera-project.eu>
- [29] M. M. Hoque, T. Onuki, Y. Kobayashi, Y. Kuno, Y. Sato, S. Kodama, In: R. Koch (Ed.), *An Empirical Framework to Control Human Attention by Robot*, (Springer, Heidelberg, 2010) 430
- [30] M. M. Hoque, T. Onuki, Y. Kobayashi, Y. Kuno, Controlling Human Attention through Robot's Gaze Behaviors, In: *International Conference on Human System Interaction*, Yokohama, Japan, (IEEE, 2011), 95
- [31] M. Argyle, M. Cook, *Gaze and Mutual Gaze*, (Cambridge University Press, Oxford, 1976)
- [32] M. V. Cranach, J. H. Ellgring, *The Perception of Looking Behaviour*, (Academic Press, London, 1973)
- [33] N. J. Emery, *The Eyes Have It: The Neuroethology, Function and Evolution of Social Gaze*, *Neurosci. & Biobeha. Rev.*, 24, 581 (2000)
- [34] M. Imai, T. Kanda, T. Ono, H. Ishiguro, K. Mase, Robot Mediated Round Table: Analysis of The Effect of Robot's Gaze, In: *IEEE International Workshop on Robot and Human Interactive Communication*, Berlin, Germany, (IEEE Press, 2002), 411
- [35] W. James, *The Principle of Psychology*, (Dover, New York, 1950)
- [36] Y. Okano, S. Fukushima, M. Furukawa, H. Kajimoto, Embedded Motion-Generating the Perception of Motion in Peripheral Vision. In: *International Conference and Exhibition on Computer Graphics and Interactive Techniques in Asia*, Seoul, Korea, (ACM Press, New York, 2010)
- [37] J. Jonides, A. Arbor, Voluntary vs. Automatic Control Over the Mind's Eye's Movement. (Lawrence Erlbaum Associates Inc., Hillsdale, NJ, 1981)
- [38] S. Yantis, J. Jonides, Abrupt Visual Onsets and Selective Atten-

- tion: Voluntary Versus Automatic Allocation. *J. of Expt. Psy.: Human Percep. & Per.*, 16, 121 (1990)
- [39] M. M. Hoque, D. Das, T. Onuki, Y. Kobayashi, Y. Kuno, Model for Controlling a Target Human's Attention in Multi-Party Settings, In: *IEEE International Symposium on Robot and Human Interactive Communication*, Paris, France, (IEEE, 2012), 476
 - [40] V. M. Cranach, In: A. H. Esser (Ed.), *The Role of Orienting Behavior in Human Interaction, Behavior and Environment*, (Plenum Press, New York, 1971), 217
 - [41] M. Cook, M. C. Smith, The Role of Gaze in Impression Formation, *Brit. J. Soc. Clin. Psycho.*, 14, 19 (1975).
 - [42] H. H. Clark, *Using Language*, (Cambridge University Press, Cambridge, 1996)
 - [43] R. B. Miller, Response Time in Man-Computer Conversational Transactions, In: *AFIPS Fall Joint Computer Conference*, San Francisco, USA, (ACM Press, New York, 1968), 267
 - [44] K. Takashima, Y. Omori, Y. Yoshimoto, Y. Itoh, Y. Kitamura, F. Kishino, Effects of Avater's Blinking Animation on Person Impressions. In: *International Conference on Graphics Interface*, Windsor, Canada, (Canadian Information Processing Society, 2008), 169
 - [45] F. Delaunay, T. Belpaeme, Towards Retro-Projected Robot Faces: An Alternative to Mechatronic and Android Faces, In: *IEEE International Workshop on Robot and Human Interactive Communication*, Toyama, Japan, (IEEE, 2009), 306
 - [46] FaceAPI: Face Tracking for OEM Product Development, Seeing Machines Int., (2010), <http://www.faceapi.com>
 - [47] Y. Kobayashi, Y. Kuno, People Tracking using Integrated Sensors for Human Robot Interaction, In: *IEEE International Conference on Industrial Technology*, Vina del Mar, Chile, (IEEE Press, 2010), 1597
 - [48] G. Bradsky, A. Kaehler, V. Pisarevsky, Learning Based Computer Vision with Intel's Open Source Computer Vision Library, *J. Intel Tech.*, 9, 119 (2005)
 - [49] M. M. Hoque, T. Onuki, Y. Kobayashi, Y. Kuno, Effect of Robot's Gaze Behaviors for Attracting and Controlling Human Attention, *International Journal of Advanced Robotics* (Accepted)
 - [50] P. E. Downing, C. M. Dodds, D. Bray, Why Does the Gaze of Others Direct Visual Attention. *Vis. Cog.*, 11, 71 (2004)
 - [51] Y. Mohammad, T. Nishida, Unsupervised Learning of Interactive Behavior for HRI, In: *RSS 2010 Workshop on Learning for Human Robot Interaction Modeling*, Zaragoza, Spain, (June 2010)
 - [52] Y. Mohammad, S. Okada, T. Nishida, Autonomous Development of Gaze Control for Natural Human-Robot Interaction, In: *International IUI 2010 Workshop on Eye Gaze in Intelligent Human Machine Interaction* (ACM, 2010)
 - [53] J. Lee, J. F. Kiser, A. F. Bobick, A. L. Thomaz, Vision-based Contingency Detection, In: *ACM/IEEE International Conference on Human Robot Interaction*, Lausanne, Switzerland, (ACM, 2011), 297