An Intelligent Human-Robot Interaction Framework to Control the Human Attention

Mohammed Moshiul Hoque, Kaushik Deb
Chittagong University of Engineering & Technology, Bangladesh
{moshiul_240, debkushik}@cuet.ac.bd

Dipankar Das, Yoshinori Kobayashi, Yoshinori Kuno
Saitama University, Japan
{dipankar, kuno}@cv.ics.saitama-u.ac.jp

***Abstract*— Attention control can be defined as shifting someone's attention from his/her existing attentional focus to another. However, it is not an easy task for the robot to control a human's attention toward its intended direction, especially when the robot and the human are not facing each other, or the human is intensely attending his/her task. The robot should convey some communicative intention through appropriate actions according to the human's situation. In this paper, we propose a robotic framework to control the human attention in terms of three phases: attracting attention, making eye contact, and shifting attention. Results show that the robot can attract a person's attention by three actions: head turning, head shaking, and uttering reference terms corresponding to three viewing situations in which the human vision senses the robot (near peripheral filed of view, far peripheral field of view, and out of field of view). After gaining attention, the robot makes eye contact through showing gaze awareness by blinking its eyes, and directs the human attention by the combination of eye and head turning behavior to share an object. Experiments using sixteen participants confirm the effectiveness of the propose framework to control human attention.**

Keywords: human-robot interaction; attention control; eye contact; shifting attention

I. Introduction

Controlling someone's attention from the current target to another is a fundamental skill in human social interaction and cognition. People direct their gaze at the object that they are paying attention to [1]. However, there are plenty of situations where determining the looker's eye direction is impossible or infeasible. Specially, 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 [2]. Therefore, we may define 'attention control as a means of gaze (i.e., head) control in which one can shift someone's gaze from one direction to that of his/her interest'. Controlling one's attention plays a critical role in a wide range of social behaviors; it sets the stage for learning, develops shared capability to the other's interest, and facilitates communication both in overt and covert.

In order to control the target human's attention to its intended direction, robots should be able to perform three tasks consecutively: attracting attention (AA), making eye contact (MEC), and shifting attention (SA). It is certain that a single action alone cannot always be successful to attract human attention. Although there may be various situations, we concentrate to consider a general important situation in this paper where a human and a robot are not facing each other initially and the human is engaging in some task (i.e., watching paintings). Under these constraints, we consider how the robot can attract the human attention depending on the relative position between the human and the robot. In this paper, we define the three positional relations between the human and the robot by the robot's position in the human's field of view, such as, near peripheral field of view (NPFOV), far peripheral field of view (FPFOV), and out of field of view (OFOV). In such situations, it is difficult to establish eye contact by using non-verbal behaviors, which is the prime condition to initiate any conversation [3].

It is apparent that the robot can attract the attention of the person by voice, but using voice is certain to attract other people's attention. Thus, voice should be used as a final resort. The robot should start with a weak action to avoid attracting other people than the target person and use stronger actions if it cannot attract his/her attention. This is the basic design concept of our robot. Turning the head toward the target person is the most fundamental cue (action) of the robot [4]. Simple head turning or eye movements may be enough when the robot exists in the human's central view but not effective in all cases [5]. Therefore, the robot may need to use stronger signal in some situations than others. We determined to use head shaking, and uttering reference terms in this order if the robot cannot attract the target person's attention.

Attracting people attention can produce observable behavioral responses such as eye movements, head movements, or body orientation. If the target person is attracted by the robot behaviors, the target person will turn toward the robot, which will make eye contact easy. However, only face-to-face orientation is not enough to establish eye contact. The robot needs to make the person notice clearly that it is looking at none other than him/her. That means, the robot should be able to display its awareness explicitly by some actions (e.g., facial expressions, eye blinking or nodding).

Attention shift involves an agent gazing at or turning to the object referred by another agent [6]. In our current design, the robot turns its eyes first and a bit later the head follows toward the target object. Humans are exceptionally good at inferring where others are looking and quite accurate at detecting others' face-directed gaze at normal conversation distances. Thus we can predict that the human will gaze at the object if the robot shifts its gaze from the person to it and that the robot gaze leads people to make assumptions

978-1-4799-0400-6/13/$31.00 ©2013 IEEE

about what the intended referents are.

The success of a robot to control a target human's attention depends on the existing situation and actions played by the robot. However, it is difficult to consider all actions and social situations together in early stages of research. Thus, as a first step, we limit our scope of research as follows. (i) How a robot can use subtle cues to attract human attention if s/he is not facing to the robot? (ii) How a robot can ensure that the human is responding and how it can tell when it has captured attention (i.e., eye contact)? (iii) How a robot can convey its intention to the human for shifting his/her attention to a particular object? In this paper, we propose three actions corresponding to three situations (i.e., head turning for NPFOV, head shaking for FPFOV, and uttering reference terms for OFOV) for attracting human attention. We show that attracting attention, making eye contact, and shifting attention behaviors of the robot can control the human attention.

II. Related Work

The capability of robots that can control human attention is in a rudimentary stage still now. One approach used a wizard-of-Oz to control human attention by several behaviors (playing, moving, and touching) [7]. Moreover, there are also not sufficient work about how humans attract others' attention to initiate an interaction beyond the primary facts that they stop a certain distance [8], and start the interaction with a greeting [9]. Some robots were equipped with the capability to encourage people to initiate interaction by some means such as the approach direction [10], the standing position [11], following behaviors [12], and uttering greeting terms [13]. 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. Although such a passive attitude can work in some situations, many situations require robots to use an active approach [14]. In this paper, we consider a general situations where a human and a robot are not in face-to-face. In such situations, it is difficult to establish mutual gaze.

It seems that we can make eye contact if we establish gaze crossing (i.e., looking at each other). Several robotic systems are able to establish eye-contact by gaze crossing [15], [16]. Psychological studies show, however, this gaze crossing action alone may not be enough to establish eye contact. Gaze-awareness is also necessary for humans to feel that they have made eye contact with others [17]. Some robotic agents shift human attention in several ways including eye gaze [18], head orientation [19], [20], reference terms and pointing gestures [21]. Most of these also assumed that a human faces to the robot when their interaction begins.

III. Attention Control Phases

In order to control the human's attention, we hypothesize that robots should perform three consecutive tasks: (i) attracting attention (AA), (ii) making eye contact (MEC), and (iii) shifting attention (SA). Fig. 1 illustrates the conceptual process of attention control (AC) in terms of sub tasks. To perform a successful AC process, both robot (R) and human (H) need to show some explicit behaviors and to respond appropriately to them by communicative behaviors in each phase. That means, R and H perform a set of behaviors, $R = \{\phi, \psi, \omega\}$ and $H = \{\lambda, \delta, \mu\}$. We show that signals for AA, MEC and SA of the robot are able to control the human attention from one direction to that indicated by the robot through experiments. In this work, we apply a set of behaviors of robot such as ϕ={head turn, head shaking, reference terms} in the attracting attention phase, ψ={frontal face detection, eye blink} in the making eye contact phase, and ω={eye turn, head turn} in the shifting attention phase respectively. We expect that humans also perform some responsive behaviors, such as λ={head/gaze turn toward the robot, body turn toward the robot} in AA phase, δ={keep looking toward the robot while blinking} in MEC, and μ ={turning head/gaze toward the robot intended object} in SA phase respectively.

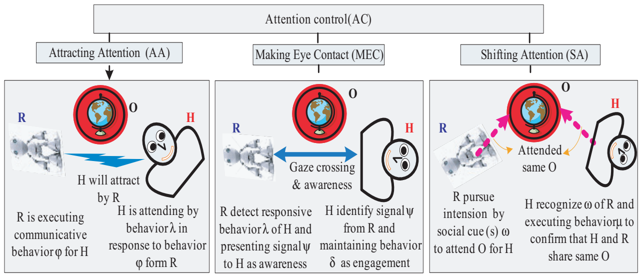


Fig. 1. Prerequisites for an attention control episode. Symbols (H, R, and O) represent the human, the robot and the object, respectively.

IV. System Overview

We have developed a robot head for HRI experiments. Fig. 2 shows an overview of the robotic head. The head consists of a 3D mask, an LED projector (3M pocket projector, MPro150), Laser range sensor (URG-04LX by Hokuyo Electric Machinery), two USB cameras (Logicool Inc., Qcam and PTZ Logicool Inc., Orbit AF) and a pan-tilt unit (Directed Perception Inc., PTU-D46). The LED projector projects CG generated eyes on the mask.

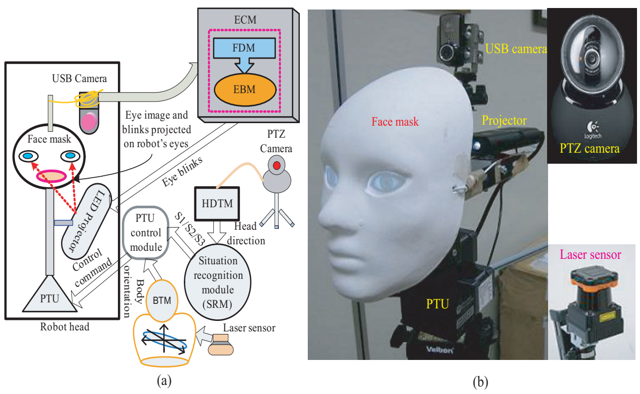


Fig. 2. System configuration: (a) system consists of five modules: *HDTM, SRM, BTM, ECM, and PTUCM* (b) prototype of robot head.

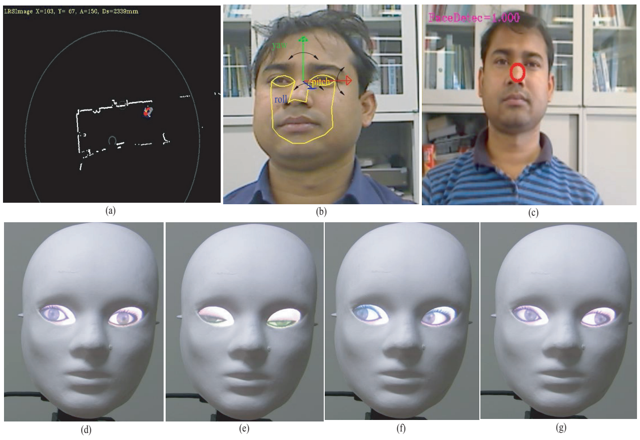


Fig. 3. (a) Tracking results of BTM (b) result of HDTM (c) face detection. EBM results:(d) fully opened eyes (d) partially open eyes (e) (right) eye turn (g) (left) eye turn.

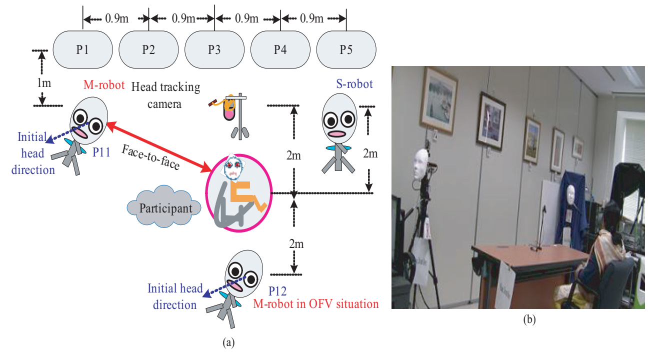


Fig. 4. (a) Schematic representation of attention attraction experiment (b) Human-robot interaction scenes.

The proposed system has five main software modules: head detection and tracking module (HDTM), body tracking module (BTM), situation recognition module (SRM), eye contact module (ECM), and pan-tilt control unit module (PTUCM). The last module controls the head turn toward the human based on BTM and provides AA signals based on the SRM results. The cross section of a human body can be modeled as an ellipse. The human body model is consequently represented with the center coordinates of ellipse $[x, y]$ and rotation of ellipse (θ). To detect, track and computes the direction of head (30 frame/sec), we use FaceAPI by Seeing Machines Inc. By extrapolating information from the HDTM and BTM, the SRM determines which situation exists. From the results of tracking modules, the system recognizes the situation and the direction of attended object in terms of yaw (α), pitch (β) movements of head and/or body orientation (θ) respectively using a set of rules. In any situation, if the target person looks toward the robot in response to AA behaviors then it considers that his/her face is directed toward the robot. In that case, FDM uses the image of the forehead camera to detect the frontal face. After detection, FDM sends the results to EBM for exhibiting eye blinks. The robot performs head turning by the PTUCM with proper control signals coming from several modules. The details description of the robotic head has described in [22]. Fig. 4 illustrates the results of several modules.

V. Preliminary Study 1: Cues for Attracting Attention

Although it might be apparent that the motion of the robot may attract human attention, we conducted the first study to confirm this hypothesis. A total of 36 graduate students of Saitama University participated in the experiment. Each participant experienced two trials and each session lasted approximately 120 seconds.

A. Experimental Conditions

M-robot produces actions (during the last 60 seconds) depending on the viewing relation between the participant and the robot as shown bellow.

(i) **Near peripheral field of view (NPFOV)**: If the participant is looking at the picture (P2 or P3), and M-robot at position P11, then the situation is considered as NPFOV.
(ii) **Far peripheral field of view (FPFOV)**: While the human is looking at the picture (P5) and M-robot at P11, the situation is considered as FPFOV.
(iii) **Out of field of view (OFOV)**: The participant is looking at pictures and M-robot is placed at position P12 (i.e., M-Robot cannot capture the face of the participant). The situation is considered as OFOV.

B. Experimental Design and Procedure

We prepared two robotic heads. Both were the same in appearance. One was (i) Moving robot (M-robot). Initially it is static and is looking in a direction not toward the human face. After recognizing the situation, it will produce three actions: (i) head turning toward the human, (ii) head shaking (moves ± 20 degrees from current position, lasting about 3 seconds), and (iii) uttering reference terms ('excuse me'). The second was (ii) Static robot (S-robot). It is stationary all times and is looking toward the human. We placed two robot heads to the left of the leftmost picture and to the right of the rightmost picture (for NPFOV and FPFOV situations). In OFOV situation, M-robot was placed back of the human. Fig. 4 (a) shows the experimental settings.

In order to simulate the three viewing situations, we hung five paintings (P1-P5) on the wall at the same hight (a bit above the eye level of participants sitting on the chair). Participants looking direction will be varied according to the viewing situations when they watch the different paintings. The USB camera was positioned in front of the participant to track his/her face. Two video cameras were placed in appropriate positions to capture all interactions. We let participants to watch the paintings, M-robot did not perform any action during the first 60 seconds.

Participants were asked to look around the paintings by moving their head. In each conditions, M-robot shows all actions (i.e., head turning, head shaking, and reference terms) to the participant one after another. However, M-robot waits about 4 seconds for human response after giving each signal. If the participant looks at M-robot within 4 seconds, the

TABLE I

PARTICIPANT'S RESPONSES ON DIFFERENT ACTIONS OF ROBOT IN THREE CONDITIONS

	Actions			Probability (p) value		
Situations	HT	HS	RT	HT/HS	HS/RT	RT/HT
NPFOV	1.91	2.0	2.0	0.38	0.32	1.0
FPFV	0.16	1.83	1.92	0.0001*	0.55	0.0001*
OFOV	.083	0.083	1.83	1.0	0.0001*	0.0001*

*-means significant differences.

robot considers that the human has been attracted. Fig. 4 (b) shows a scene of the experiment. We videotaped all sessions to analyze human behaviors.

C. Hypotheses

We performed this study to check the effect of robot's actions in various viewing situations of human. We expected that the following hypotheses would be verified by the experiment.

(i) **H1**: For NPFOV, head turning toward the participant is enough to attract his/her attention. However, this simple turning action is not enough to capture human attention when the robot exists in his/her FPFOV. More strong signals (i.e., head shaking) are needed in the FPFOV.

(ii) **H2:** For OFOV, any kind of motions cannot attract human attention. The robot needs to use voice/sound signals to attract their attention toward the robot.

D. Results

The study conducted was a within participants design. To observe the effect of each action, we first examined whether the participants responded to M-robot after its actions. We counted the numbers of times that the participants responded to robot's actions. Table I represents the mean value of participant's response with respect to the robot's behaviors.

We conducted the measures of analysis of variance (ANOVA) in all conditions. We observed a total of $216(12 \times 3 \times 2 \times 3)$ interactions for all conditions. Concerning NPFOV, no significant differences were found (F(2,22)=0.47, p=0.62). We conducted multiple comparisons with Bonferroni method that showed no significant differences between HT and HS ($p = 0.38$), HS and RT ($p = 0.32$), and RT and HT ($p = 1.0$) in attracting participant's attention. Concerning FPFOV, there are significant differences were found (F (2,22)=98.5, $p < 0.01$). Multiple comparisons also shows a significant difference between both pairs(HT and HS: p=0.0001, RT and HT: p=0.0001). However, there is no significant difference was found between HS and RT (p=0.55). These results verified our hypothesis 1

In the case of OFOV, significant differences were found among actions (F=(2,22)=124.38, $p < 0.01$). Multiple comparisons showed that the participants respond more while the robot uttering reference terms than HT (p=0.0001) and HS (p=0.0001). But no significant difference was found between HT and HS (p=1.0). This result confirmed our hypothesis 2.

Therefore, the results indicate that the robot needs stronger actions to attract human attention when the situation becomes tougher. In OFOV cases, the robot needs to use voice (except in one case in where the participant responded to HT and HS. This may be due to the sound/noise produced by the PTU while shaking).

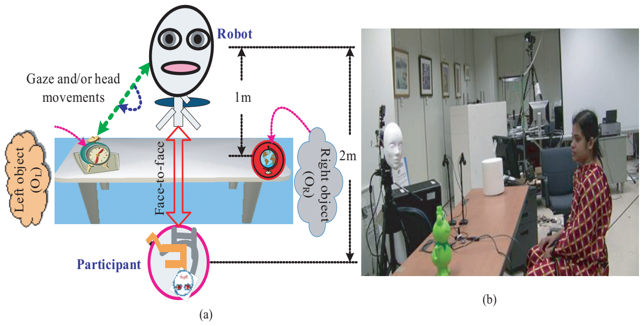


Fig. 5. (a) Schematic set up, and (b) an experiment scene.

VI. PRELIMINARY STUDY 2: CUES FOR SHIFTING ATTENTION

In this study, we examined the effect of eye and head movements to direct the human attention to a particular object.

A. Experimental Design and Procedure

Two objects (left and right) were placed on the table as attention targets. Fig. 5(a) shows the experimental setting. A participant is asked to sit down on the chair and to look at the robot. The robot head is facing the participant. After a few seconds, the robot presents one of the following motions three times of each: (i) turning eyes (TE): it turns only its eyes toward a target object, (ii) turning head (TH): it turns its head only toward an object, and (iii) turning both eye and head (TEH): it first turns its eyes toward an object then the head follows to turn toward the object. The robot waits about 4 seconds after each attempt. We videotaped the participants' behaviors.

Fig. 5 (b) shows a scene of the experiment. After the session, we evaluate participant's responses (i.e., looking at the object that the robot indicated by its motion) after the first attempt only and we would like to find out the answer to the question, 'which robot motion made the participants shift their attention most'?

B. Results

We used a total of 30 university students (10 students in each motion) to perform this study. The result indicates that the third one (turning both eyes and head) gained more votes (50%) than those of only head turn (40%) and only eye turn (20%). Although, we need to further study the combination of head and eye movements by using more participants, the preliminary result of this experiment confirms that the combination of eye and head movements is better than the others (i.e., only head or eye turn) to convey robot's attended direction and more useful to shift participants attention to the target object.

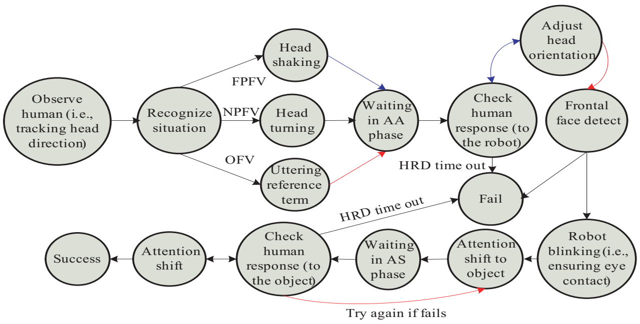


Fig. 6. Robot behaviors to control the human attention.

VII. Behavioral Protocol of Robot

In order to initiate an attention control process, the robot first observes the current attentional situation of the human, applies an appropriate action to capture his/her attention, and establishes a communication channel through eye contact. Fig. 6 illustrates how our robot system works to control a human's attention when the human and the robot are not facing each other initially.

Based on the preliminary studies, we hypothesized that the head turning is enough when the robot is present in the human's NPFV. However, head shaking is necessary for FPFV. Moreover, the robot should use reference terms (i.e., voice) if the robot is present in the out of field of view (OFV) of the human. If the human responds within the expected times (1 second in NPFV, 2 seconds in FPFV, and 3 seconds in OFV), the robot adjusts its head orientation toward the human based on the result of body tracking module, detects his/her frontal face and blinks its eyes (about 2 seconds) to create awareness. If the human does not gaze at the robot within the delay period after its action, the robot considers the case as failure and gives up attention control. After establishing eye contact, the robot makes an attention shift request by switching its attention through turning both eyes and head (a bit later) in our current implementation to the object which it intends to address. After addressing the focus, the robot waits (2 seconds) and checks whether or not the attention shift attempt is successful. If the responding agent is not attending to the referential object, the robot will try again by shifting its attention back to the face-to-face position, and focus again to the object.

VIII. Attention Control Experiments

To evaluate the system, we performed experiments using 16 participants, total 64 trials. All of them were graduate student of Saitama University. We evaluated the robot capabilities in terms of objective and subjective ways.

A. Experimental Design and Procedures

The experimental setting was almost similar as that described in section V with some modifications. Two objects were placed on the table as attention targets in the same way as in experiment 2. We programmed M-robot as the attention control robot ('AC-robot') and placed at left of the leftmost picture. The placement of S-robot was unchanged. AC-robot behaved as described in section VII. To create the out of field of view situation, we added one more picture (P6) with the existing 5 pictures (i.e., the participant cannot see AC-robot while watching P6). Fig. 7 shows some snapshots. We asked the participants to watch the hanging paintings. We captured the participants' behaviors using video cameras and logged the head and body direction data.

(a) (b)

Fig. 7. Experimental scenes: (a) Human and AC-robot in face-to-face and make eye contact, and (b) AC-robot and human shift attention to O_L.

TABLE II

The effectiveness (η_{AC}) of the robot to control human attention.

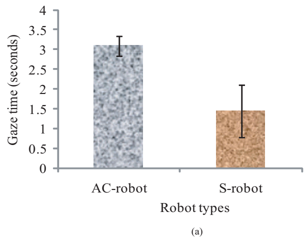
Situations	Number of responses/attempts			Effectivness (η_{AC})
	N_R/N_{AA}	N_B/N_{EC}	N_R/N_{AS}	
NPFOV	18/20	17/18	14/17	70%
FPFOV	23/26	19/23	16/19	61.5%
OFOV	14/18	12/14	10/12	55.5%

IX. Experimental Results

We counted the number of times that the participants responded to the robot (this number provides a measure of success) in each phase and measured the total time needed to complete the attention control process. We also evaluated the robot behaviors in terms of subjective measures.

Behavioral Analysis: Table II summarizes the numbers of attempts made by the robot in the attention attraction, eye contact, and attention shift phases. Participants responded 18 times among 20 in NPFOV, 23 times out of 26 AA attempts in FPFOV, and 14 times out of 18 AA attempts in OFOV respectively. The robot proceeded to the next eye contact step in these successful cases. Results show that the attention shifting capabilities are somewhat common in all situations. However, the eye contact effort was higher in NPFOV (94%) than that of FPFOV (82%) and OFOV (85%) situations. These differences may be due to orientation mismatch (between the participants face and the robot) and hence failure to detect the human face. A substantial 70% participants' attention was controlled by the robot in NPFOV while only 61% and 55% were controlled in FPFOV and OFOV. Participants were somewhat more frequently responded in the frontal case than in the others because they may be able to notice robot behaviors easily due to its physical appearance in his/her field of view.

Comparing the time that the participants spent on gazing at the two robots by using paired t-test shows that participants gaze 2.2 times more at AC-robot (3.09 seconds), than at the



Situations	τ_{AA} (secs)	τ_{EC} (secs)	τ_{AS} (secs)	$\tau_{AC}=\tau_{AA}+\tau_{EC}+\tau_{AS}$ (secs)
NPFV	2.50	2.50	4.0	9.0
FPFV	7.50	2.50	4.0	14.0
OFV	5.50	2.50	4.0	12.0

(b)

Fig. 8. (a) Total times spent on gazing two robots after AA signals (b) times to perform an attention control episode.

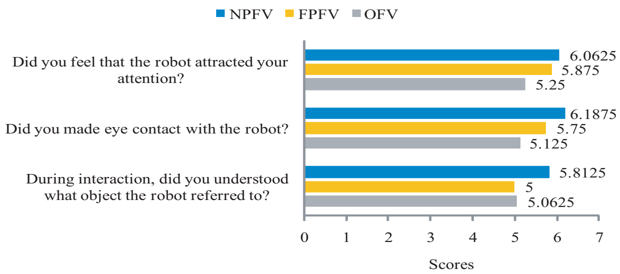


Fig. 9. Result of the post-questionnaire for the attention control experiment.

S-robot (1.43 seconds). That means participants gazed at AC-robot significantly longer ($t(47) = 17.18, p < 0.0001$) due to its eye contact behaviors (Fig. 8(a). Fig. 8(b) shows the times to control attention in various situations. The value of τ_{AA} is greatly affected by the situation, hence the τ_{AC} too. This is because of the different kinds of actions as well as human response delay involved in each situation.

Subjective Evaluation: After all sessions, we asked the participants questionnaire to obtain their evaluations on a 7-point Likert scale (Fig. 9). The attention control robot can attract their attention, making eye contact with them, and convey its intended direction to them effectively in all situations. A 2-way ANOVA analysis shows that participants' rating depends significantly both on the question types ($F(2,143) = 5.15$, $p = 0.0075$, η^2= 0.03) as well as situations ($F(2,143) = 13.7$, $p < 0.05$, η^2= 0.14).

Finally, we asked to describe how effective they found our robot to control their attention from the current object to the one that it chooses. Participants were asked to subjectively rate the robot's effectiveness on a seven-point Likert scale. The average score was $5.06(SD = 0.99)$. The experimental results confirm that the proposed robot can effectively attract a target person's attention, set up eye contact, shift attention to an intended object and hence control his/her attention.

X. Conclusions

The main objective of our work is to develop a robot system that can shift a particular person's attention from one direction to another by considering the social situations. It is difficult to design effective non-verbal behaviors for robots in various situations. The success rate of a particular action of robot depends on several factors, such as distance between the robot and the human, spatial angle between them, duration of the action, direction of attention, level of attention of human to the task involved and so on. In this study, we consider only situation dependent actions with respect to the direction of attention (such as head turn for NPFV, head shaking for FPFV, and reference terms for OFV). We have shown that robot's actions are effective to gain human attention, to make a person feel that the robot looks at him/her not others and shift his/her attention to an intended object. If a human is paying attention to a particular object or intensely engaged in a task, the robot needs to use more strong actions such as waving hand or the combination of verbal and nonverbal actions. This is left for future work.

References

[1] S. B. Cohen, "The eye direction detector (edd) and the shared attention mechanism (sam): Two cases for evolutionary psychology", in Joint Attention: Its Origins and Role in Development, Lawrence Erlbaum Associates Inc., 1995, pp. 41-58.

[2] M. Argyle, and M. Cook, Gaze and Mutual Gaze, Cambridge University Press, Oxford, 1976.

[3] E. Goffman, Behavior in Public Place, New York: The Free Press, USA, 1963.

[4] B. Mutlu, "The design of gaze behaviors for embodied social interface", CHI EA'08, pp. 2661-2664, 2008.

[5] M. Finke, et al., "Hey, I'm over here-How can a robot attractpPeople's attention?, IEEE ROMAN, pp. 7-12, 2005.

[6] G. Butterworth, The ontogeny and phylogeny of joint visual attention, in Natural Theories of Mind: Evolution, Development, and Simulation of Everyday Mindreading, Basil Blackwell, 1991, pp. 223-232.

[7] M. Shiomi, T. Kanda, S. Koizumi, H. Ishihuro, and N. Hagita, "Group attention control for communication robots with wizard of oz approach", ACM International Conference on HRI, pp. 121-128, 2007.

[8] E. T. Hall, The Hidden Dimension: Man's Use of Space in Public and Private, London: The Bodley Head Ltd., 1966.

[9] A. Kendon, Features of the Structural Analysis of Human Communicational Behavior, Holland: Swets and Zeitlinger, 1980.

[10] K. Dautenhahn et al., "How may I serve you?: A robot companion approaching a seated person in a helping context", ACM/IEEE International Conference on HRI, pp. 172179, 2006.

[11] F. Yamaoka, T. Kanda, H. Ishiguro, and N. Hagita, "A model of proximity control for information presenting robot, IEEE Tran. Robo., vol. 26, pp. 187195, 2010.

[12] R. Gockley, J. Forlizzi, and R. G. Simmons, "Natural person following behavior for social robots", ACM/IEEE International Conference on HRI'07, pp. 1724, 2007.

[13] C. M. Huang, and A. L. Thomaz, "Joint attention in human-robot interaction, Dialog with Robots: AAAI, pp. 3237, 2010.

[14] M. M. Hoque et al., "An empirical framework to control human attention by robot, in ACCV 2010 Workshops, Part I, LNCS, Springer, Heidelberg, vol. 6468, pp. 430-439, 2011.

[15] B. Mutlu, T. Shiwa, T. Kanda, H. Ishiguro, and N. Hagita, "Footing in human robot conversations: How robots might shape participant roles using gaze cues", International Conference on HRI, pp. 61-68, 2009.

[16] T. Kanda, H. Ishiguro, T. Ono, M. Imai, and R. Nakatsu, "Development and evaluation of an interactive humanoid robot 'robovie'", IEEE ICRA, pp. 1848-1855, 2002.

[17] D. Miyauchi, A. Nakamura, and Y. Kuno, "Bidirectional eye contact for human-robot communication", IEICE Trans. of Info. and Syst., vol. 88D, pp. 2509-2516, November 2005.

[18] Z. Yucel, et al., "Joint visual attention modeling for naturally interacting robotic agents, ISCIS, pp. 242-247, 2009.

[19] M. Staudte and M. W. Crocker, "Investigating joint attention mechanisms through spoken human-robot interaction", J. Cognit., vol. 120, pp. 268-291, January 2011.

[20] Y. Nagai et al., "A constructive model for the development of joint attention", J. Conn. Sci., vol. 15, pp. 211-229, 2003.

[21] O. Sugiyama et al., "Human-like conversation with gestures and verbal cues based on three-layer attention drawing model, J. Conn. Sci., vol. 18, pp. 379-402, December 2006.

[22] M. M. Hoque, K. Deb, D. Das, Y. Kobayashi, Y. Kuno, "Design an intelligent robotic head to interacting with humans", ICCIT'12, pp. 539-545, 2012.