7th Annual International IEEE EMBS Conference on Neural Engineering
Montpellier, France, 22 - 24 April, 2015

Gaze based robot control: the communicative approach

A. A. Fedorova, S. L. Shishkin, Y. O. Nuzhdin and B. M. Velichkovsky

Abstract— **We propose a novel way of robotic device control with communicative eye movements that could possibly help to solve the problem of false activations during the gaze control, known as the Midas touch problem. The proposed approach can be considered as explicitly based on communication between a human operator and a robot. Specifically, we employed gaze patterns that are characteristic for "joint attention" type of communication between two persons. "Joint attention" gaze patterns are automatized and able to convey information about object location even under a high cognitive load. Therefore, we assumed that they may make robot control with gaze more stable. In a study with 28 healthy participants who were naive to this approach most of them easily acquired robot control with "joint attention" gaze patterns. The study did not reveal higher preference for communicative type of control, possibly because the participants did not practice before the tests. We discuss potential benefits of the new approach that can be tested in future studies.**

I. INTRODUCTION

Gaze follows attention effortlessly, in particular attention to objects on which we want to act. One may expect that this ability of gaze would be widely employed to control robotic devices, especially for assisting people with motor disabilities. Developing gaze based control of certain computer applications, especially text typing, is indeed an active area of assistive engineering [1]. However, gaze based solutions for robotic control are still rare and seem to be effective only when they are given a limited role within an advanced system (e.g., [2]). This can be related to the Midas touch problem [3]: if gaze was given a function of controlling devices, multiple false commands will be generated due to inability of humans to prevent unintended eye movements. Indeed, the gaze primarily serves for the visual system and is typically under unconscious control [4]. While in gaze typing distractors can be almost totally excluded from the visual field, much more dynamic visual environment is inevitable during robot control, and various objects or unexpected events can often attract the gaze. The existing approaches to solving Midas touch problem (e.g., [5]) often make the control unnatural and tedious.

The natural function of gaze that is employed in gaze based control of devices is likely related not to vision but to communication. Communicative function of gaze was the focus of many psychological studies (e.g., [6], [7]). However, to our knowledge, the results of these studies have never been explicitly used in developing methods for gaze based control of devices.

In this paper, we propose an approach that makes explicit use of the knowledge obtained in the studies of gaze based communication in humans. Specifically, we propose using gaze patterns described by developmental psychologists for situations known as "joint attention". This term denotes the "simultaneous engagement of two or more individuals in mental focus on one and the same external thing" [8]. "Joint attention" as an ability to actively and passively share an external attentional locus is usually formed at around 1 year of age and plays an important role in social development [9]. "Joint attention" is a very fast and automatized process that works even in tasks that require a high cognitive load [10]. We assumed that effortlessness and high speed of mechanisms involved in "joint attention" can lead to improvement in usability. Moreover, spontaneous eye movements related to vision function seem to be, to a certain extent, suppressed in "joint attention", otherwise they could significantly interfere with communication; activation of "joint attention" gaze behaviour, therefore, appears a prospective remedy against the Midas touch problem.

As the first step in exploring the usefulness of the proposed approach, we designed a testing procedure aimed on collecting experience related to gaze based control enriched with "joint attention" patterns and on estimating the easiness of adopting such type of control. In this study, participants could control a robotic device with "communicative" and "instrumental" control strategies. A control strategy is defined here as a sequence of fixations within predefined regions that result in a proper response of a robotic device. By the terms "communicative control strategies" and "instrumental control strategies" we understand using the eye movement patterns specific and not specific for communication, respectively. In a communicative strategy, a controlled robotic device can be considered as an autonomous partner who can show that it "attends" to the user in anthropomorphic way, and the act of sending a command can be considered as a communicative act. In the case of commonly used instrumental strategies, however, the controlled device is not a partner but just an instrument for its user; it may provide a feedback but not in an anthropomorphic way. In this study, only such instrumental strategies that required the use of an additional "switch" to prevent unintended activations were considered.

Each participant was first asked to find a way to control the robot with her/his gaze but was not told what kind of gaze patterns should be used, so that we could record which strategy from the existing set was found first. Then, we asked the participant to find and try other pre-programmed strategies; finally, the participant performed a simple task using a control strategy he or she chose.

This work was supported by Russian Science Foundation Grant 14-28-00234 to NRC "Kurchatov Institute".

NBICS Centre, National Research Centre "Kurchatov Institute", Moscow, Russia `anastasya.teo@gmail.com`

We hypothesized that communicative strategies will be perceived as more natural and more usable by the participants comparing to commonly used instrumental strategies. The following two specific hypotheses were set up: (1) participants will find communicative strategies more frequently than instrumental strategies in a situation of an uninformed search; (2) participants will use communicative strategies more frequently than instrumental strategies when provided an opportunity to choose a strategy freely.

II. METHOD

A. Participants

28 healthy right-handed volunteers (mean age $M \pm SD$ 24 ± 4.5 , 19 women) participated in the testing procedure after signing an informed consent. All of them were naive to gaze based robot control. The study was conducted in accordance with the Declaration of Helsinki (2000 revision).

B. Materials

Eyelink 1000 Plus (SR Research, Canada) eye tracker was used for gaze direction detection in a remote mode without a chinrest. R12-six robotic arm (ST Robotics, UK) with an attached plane paper mask and a soft-tip pen under its "eyes" was used as a device to be controlled by eye movements. A green round plastic disc ("button") used as a "switch" in instrumental control strategies was placed right to the robot on a white board (Figure 1).

The fixation detection algorithm was based on top of the Eyelink 1000 plus built-in parser [11]. Fixations at predefined "sensitive" regions (robot's eyes, "button", target positions) were detected, and if their durations and order matched a pattern specific to any instrumental or communicative strategy (see below), a command appropriate for the recognized strategy was sent to the robot.

C. Procedure

A participant was sat in a chair in 150 cm from a white board where targets were placed. In the initial position, the end of the robotic arm with the mask was to the left of the participant close to the white board. The plane of the mask was perpendicular to the board and its "eyes" were "looking" to the right. During command execution, the arm moved to certain positions on the board (see below) and the mask either turned to the board or "looked" at the participant. The eye tracker was at 65 cm distance in front of a participant (Figure 1). The participants were given instructions for each upcoming stage just before it began.

Two communicative strategies could be used. The "joint attention triangle" [9] strategy consisted of a fixation at robot's "eyes" for 500 ms, robot activation ("a communicative look": robot turned to a participant and "looks" at his/her eyes), fixation at robot's eyes for 300 ms or longer and a fixation at the target position on the board for 300 ms or longer (Figure 2, A). The "recursive mind reading approach" [9] strategy included series of short (128 ms or longer) gaze switching between robot's eyes (without waiting for robot activation) and a target position, at least 4 fixations.

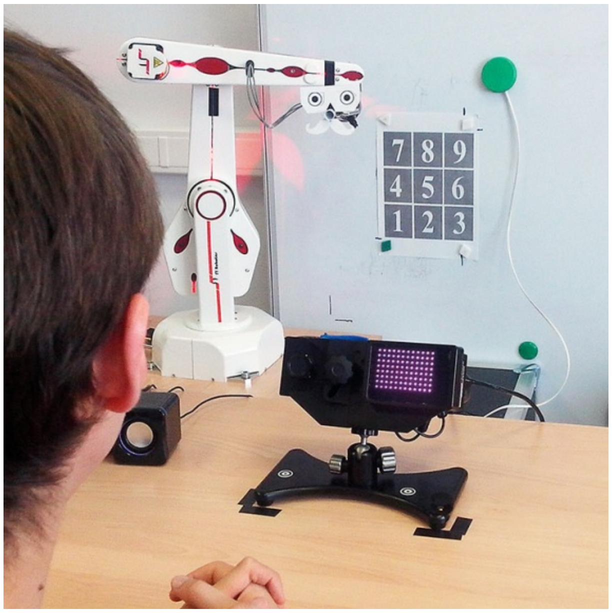

Fig. 1. A view of the experimental setup from the back of a participant. Here, the third stage of the testing procedure is shown at a time when the robot is "looking" at the participant following a communicative strategy pattern. A green "button" with a wire that connected a LED in it to an Arduino controller is attached to the board above the table with target digits. The eye tracker is on the table in front of the participant.

Two instrumental strategies used similar time patterns but the participants had to look at the "button" to activate the robot rather than to look directly at the robot, and the robot was also not "looking" at the participant. The "instrumental triangle" strategy consisted of a fixation at the green "button" for 500 ms or longer, receiving the "instrumental feedback" (the LED switched on inside the green "button" and the robot's "head" turned to the board), a fixation at the green "button" for 300 ms or longer, and a fixation at the target position on the board for 300 ms or longer (Figure 2, B). The "instrumental loop of fixations" included series of short (128 ms or longer) gaze switching between the green "button" and the target position (without waiting for the instrumental feedback).

After detecting a command sent using any of the strategies the robot made the following actions: (1) took an initial position "looking" to the right, (2) moved to a target position, (3) turned to the board and touched it, (4) moved back to the initial position (Figure 2, A and Figure 2, B, 1st frame). The duration of the action chain depended on the target position but did not depend on the strategy. All strategies were available for usage at all stages of the procedure.

The testing procedure included the following three stages:

1. "Uninformed search": after the eye tracker calibration the participant was instructed that her/his eye movements were tracked and she/he had to find a way, using their gaze only, to make the robot point the target on the white board. No additional information about the possible strategy or about the quantity of strategies was provided at this stage. As soon as the participant was sure that she/he found a method,

the first stage was over.

2. “Find all”: right after the first stage (without a recalibration of the eye tracker) a participant was told that there were more than one strategy to control the robotic arm and that she/he should try to find all of them (without telling the exact quantity of strategies). This stage lasted for 5 minutes, then the participant answered part one of the questionnaire (see subsection “Questionnaire”).

3. “Apply strategies found”: after a calibration procedure the single target was replaced by a paper keyboard with nine digits. A participant was asked to make the robot sequentially point nine digits (once for each digit from 1 to 9) in a predefined order. This stage ended when the whole sequence was pointed correctly. After that the participant was asked to answer the second part of the questionnaire.

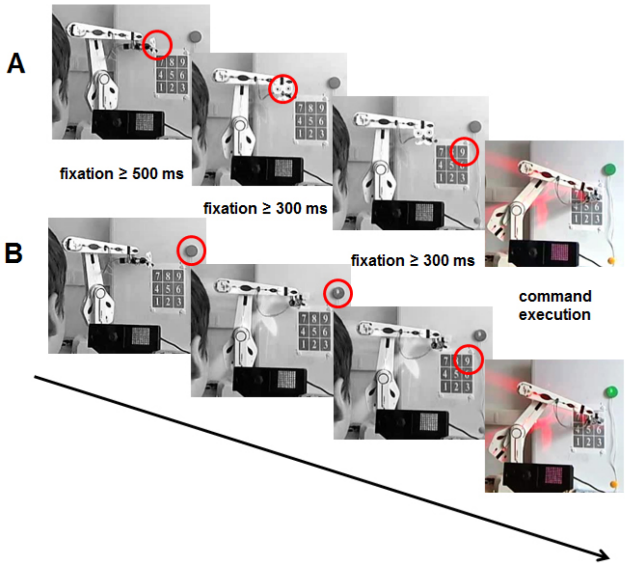

Fig. 2. The main elements of communicative (A) and instrumental (B) strategies of gaze based control. Red circles show gaze fixation positions.
A. “Joint Attention Triangle” strategy consisted of (1) fixation at robot’s “eyes” for 500 ms that led to robot “activation”, (2) (after robot turned its “face” to the participant) fixation at the robot eyes for 300 ms or longer, (3) fixation at the target position on the board for 300 ms or longer.
B. “Instrumental Triangle” strategy consisted of (1) fixation at the green “button” for 500 ms or longer, (2) (after the green “button” switched on and robot turned its “face” away from the participant) fixation on the green “button” for 300 ms or longer, (3) fixation at the target position on the board for 300 ms or longer.

At the first and second stages of the testing procedure a single target was used, while a table with nine target digits (Figure 1) was used at the third stage.

D. Questionnaire

The questionnaire was designed specifically for this study to access the participants’ subjective experience. It was divided into two parts. The first part was presented to a participant after the second stage of the testing procedure and contained questions that presumably could make following strategies more deliberate in the third stage: a participant was asked how many strategies she/he found and then to devise a name for each strategy and describe it. After that she/he was asked to evaluate usability of strategies they named.

In the second part of the questionnaire presented after the third stage the participant was asked to evaluate a sense of agency (sense of control) that she/he experienced while controlling the robot.

In the analysis of the responses, strategies were identified based on analysis of descriptions provided by the participants.

III. RESULTS

In their answers to the questionnaire, none of the participants explicitly described any of looping strategies (“recursive mind reading approach” and “instrumental loop of fixations”), possibly because they turned out to be quite difficult to use: it was hard to fixate at the same region repeatedly in a very short interval of time. However, some participants actually used those strategies, probably implicitly. We assumed that even when participants used such strategies they did not recognize their actions as a separate strategy different from “joint attention triangle” strategy or “instrumental triangle strategy” respectively. Therefore, the data were collapsed over each pair of strategies, resulting in two data pools: instrumental and communicative strategies.

There was no significant difference between strategy preference in the first stage of the testing procedure: 13 participants found an instrumental strategy and 15 participants found a communicative strategy first. More participants made first fixation which duration was 500 ms or longer at the robotic “face” (19 participants) than at a green “button” (9 participants) (chi-square goodness of fit test, $p = 0.06$).

At the third stage of the experiment, 8 participants used only a communicative strategy(-ies), 7 participants used only an instrumental strategy(-ies), and 13 participants used strategies of both types. For the number of control attempts (an attempt to “dial” one digit) collapsed over participants, 51% were made with a communicative strategy(-ies) and 49% with an instrumental strategy(-ies). The participants who used both types of strategies tended to make errors slightly more frequently with communicative comparing to instrumental strategies ($M \pm SD$ 0.40 ± 0.39 and 0.33 ± 0.37, respectively; the difference was not significant, according to paired t-test, $p = 0.6$). In the participants who used only one type of strategy the error rate did not depend on strategy type (0.10 ± 0.11 and 0.10 ± 0.08, respectively; the difference was not significant, according to t-test, $p = 0.6$). Six participants did not make any mistake in the third stage. When their data were collapsed, 55% of all control attempts were found to be with a communicative strategy(-ies) and 45% were with an instrumental strategy(-ies).

In the responses to the questionnaire, 17 participants described more than one strategy and found strategies both of communicative and instrumental type. The data from this subgroup was analyzed in detail. Nine of these participants said that a communicative strategy was the most usable, while eight of them said that an instrumental strategy was the most usable. Some of the subjects who chose a communicative strategy as the most usable mentioned that they preferred this strategy because it imitated communicative situation, was natural and/or emotional (e.g., amusing). Some of those who chose an instrumental strategy said that the robot's turn to them distracted them from the task, or that it was slower to use a communicative strategy. This could be related to the fact that the averaged duration of fixation in the beginning of the act of control was slightly longer (1402 ± 375.5 ms) at robot's eyes (the first two slides in Figure 2 A) comparing to fixation at the green "button" (the first two slides in Figure 2 B, 1229 ± 392.1 ms), although this difference was not significant (paired Student's t-test, $p = 0.22$). Average feeling of agency was 1.8 for those who stated that communicative strategy was more usable and 1.5 for those who preferred the instrumental strategy (at the scale, where an answer "1" was "The robot was fully under my control" and an answer "5" was "The robot's actions were not related to my commands at all"). The difference was not significant (t-test, $p = 0.4$).

IV. DISCUSSION

To the best of our knowledge, this study was the first one where the "joint attention" metaphor was applied for gaze based control. This metaphor was already employed to improve the quality of interaction between a human and a robot ([12], [13]), but in these studies the robot was not given commands by gaze.

Although the participants did not know about the gaze control strategies available in our system, all of them found one or few strategies without any hint from the experimenter. However, no preference for the "joint attention" based communicative strategies was found on the group level comparing to the common instrumental strategies in the sense of the first found strategy or the strategy used for control in the final test. The evaluation of usability, feeling of agency and accuracy at the third stage by the participants who used both types of strategies did not differ between the strategy types. The robotic "face" was near significantly more often the first position of a fixation that evoked robot's action comparing to the "button" (19 vs. 9 participants). However, what caused this preference could be the fixed position of the robot in the left while the "button" was always in the right; exploring space from the left to the right could be natural for the participants (e.g., because they read left to right in their native language).

It is important to note that "communication with the robot" was the most unusual part of the testing procedure, and that the study design did not include practice that could help to adopt it. The preference for a non-communicative (instrumental) strategy found in quite many participants could be, therefore, a result of the unusualness of the communicative strategies. It was possible that there were large variations across the participants in communicative style, experience, learning capabilities, strategy chosen during the study etc. that could smear possible positive effects of joint attention. Nevertheless, quite many participants liked the communicative strategies. It seems likely that other participants could also adopt the communicative strategies if they could practice enough. It is also likely that the communicative strategies of robot control can be useful not for all people. In further studies, we plan to assess the efficiency of different strategies in more objective way and after significant practice, so that the both negative and positive effects of novelty were eliminated or reduced.

It is important to note that the ways of control we studied were relatively slow. However, they were intended to provide a mean for reliable control in situations where strong distractors can be common and false alarms can have a high cost. The current initial study did not test the hypothesis of the protective value of the "joint attention" based mode of control, and this is also one of the possible direction for future studies. In addition, the proposed strategies possibly can be further elaborated to adjust to specific needs, e.g., faster control can be obtained by showing, after robot activation, several target positions in a sequence.

REFERENCES

[1] P. Majaranta, *Text entry by eye gaze*. Tampereen yliopisto, 2009.

[2] A. Frisoli, C. Loconsole, D. D. Leonardis, F. Banno, M. Barsotti, C. Chisari, and M. Bergamasco, "A new gaze-bci-driven control of an upper limb exoskeleton for rehabilitation in real-world tasks." *IEEE Transactions on Systems, Man, and Cybernetics, Part C*, vol. 42, no. 6, pp. 1169–1179, 2012.

[3] R. J. K. Jacob, "The use of eye movements in human-computer interaction techniques: What you look at is what you get." *ACM Trans. Inf. Syst.*, vol. 9, no. 2, pp. 152–169, 1991.

[4] J. M. Findlay and I. D. Gilchrist, *Active Vision: The Psychology of Looking and Seeing*. Oxford University Press, 2003.

[5] B. Velichkovsky, A. Sprenger, and P. Unema, "Towards gaze-mediated interaction: Collecting solutions of the "midas touch problem"." in *INTERACT*, ser. IFIP Conference Proceedings, S. Howard, J. Hammond, and G. Lindgaard, Eds., vol. 96. Chapman & Hall, 1997, pp. 509–516.

[6] M. Carpenter, K. Nagell, M. Tomasello, G. Butterworth, and C. Moore, "Social cognition, joint attention, and communicative competence from 9 to 15 months of age," *Monographs of the society for research in child development*, pp. i–174, 1998.

[7] B. M. Velichkovsky, "Communicating attention: Gaze position transfer in cooperative problem solving," *Pragmatics & Cognition*, vol. 3, no. 2, pp. 199–223, 1995.

[8] D. A. Baldwin, "Understanding the link between joint attention and language," *Joint attention: Its origins and role in development*, pp. 131–158, 1995.

[9] M. Carpenter and K. Liebal, "Joint attention, communication, and knowing together in infancy," *Joint Attention: New Developments in Psychology, Philosphy of Mind, and Social Neuroscience*, pp. 159–182, 2011.

[10] S. Xu, S. Zhang, and H. Geng, "Gaze-induced joint attention persists under high perceptual load and does not depend on awareness," *Vision Research*, vol. 51, no. 18, pp. 2048–2056, 2011.

[11] D. M. Stampe, "Heuristic filtering and reliable calibration methods for video-based pupil-tracking systems," *Behavior Research Methods, Instruments, & Computers*, vol. 25, no. 2, pp. 137–142, 1993.

[12] C. M. Huang and A. L. Thomaz, "Effects of responding to, initiating and ensuring joint attention in human-robot interaction," in *RO-MAN, 2011 IEEE*. IEEE, 2011, pp. 65–71.

[13] M. Imai, T. Ono, and H. Ishiguro, "Physical relation and expression: Joint attention for human-robot interaction," *Industrial Electronics, IEEE Transactions on*, vol. 50, no. 4, pp. 636–643, 2003.