Computational Linguistics and Intellectual Technologies:
Proceedings of the International Conference “Dialogue 2019”

Moscow, May 29—June 1, 2019

THE ROLE OF ORIENTED GESTURES DURING ROBOT'S COMMUNICATION TO A HUMAN[1]

Zinina A. (zinina_aa@nrcki.ru),
Arinkin N. (arinkin_na@nrcki.ru),
Zaydelman L. (zaydelman_ly@nrcki.ru),
Kotov A. (kotov_aa@nrcki.ru)

Kurhcatov Institute, Moscow, Russia;
Russian State University for the Humanities, Moscow, Russia

The role of oriented gestures is crucial while solving spatial problems. We analyze the influence of a robot, using oriented gestures, on a human. In an experimental situation robot F-2 was helping a human to solve a “tangram” puzzle. Robot was indicating in speech, which game element to take and where to place it. In a half of the tasks the robot was using oriented communicative actions (hand gestures, head movements and gaze) to indicate the required game element, and then—the game position to place it in. In the other half of tasks, the robot was using non-oriented gestures. We show, that the use of oriented gestures increases the attractiveness of a robot to human and rises the general satisfaction of the interaction with the robot.

Keywords: Multimodal communication, oriented gestures, robot-to-human interaction

Several sciences—linguistics, psychology and social robotics—cooperate in order to explore the capabilities of a robot to maintain natural communication with humans. This area has a wide research potential—in particular, the search for “natural” and easy-to-use user interfaces is of fundamental importance [Breazeal, Scassellati, 2002]; [Beuter, Spexard et al., 2008]; [Klamer et al., 2011]. At the same time, this behavior of a robot should be as close as possible to the communicative behavior of a human, therefore—should be complex and diverse.

[1] The research is supported by the Russian Science Foundation (project No 19-18-00547).

The interaction between a robot and a human can be modelled and studied in three modules: multimodal, cognitive and emotional [Lee et al., 2005]. Cognitive interaction is the ability of a robot to understand user's commands, emotional interaction is necessary to maintain positive relations between a robot and a user, and the multimodal module represents the means of interaction that are most convenient and familiar to humans. Many studies have shown that non-verbal communication plays an important role in coordinating actions when a robot and a person work together. In [Breazeal, 2003] it is experimentally proved that on the one hand, people send a robot to perform a physical task using speech and gesture. On the other hand, the nonverbal behavior of the robot has a positive effect on the success of solving the problem during human-computer interaction. In [Cabibihan, et al, 2009] it is shown that the pointing gestures accompanying the speech facilitate the understanding of spatial information in a videoconference. The authors in an experimental study proved that the use of pointing gestures increases the speed and accuracy of the task. Researchers [Häring, et al, 2012] studied the influence of gaze and pointing gestures of a humanoid robot on human performance in solving abstract puzzles of varying complexity. The authors confirmed that the directional gaze of the robot usually improves the interaction. At the same time, the authors have shown that additional pointing gestures are often necessary to make the interaction between the robot and a user more effective.

In [Salem et al., 2012] it is shown that participants evaluate the robot more positively when its nonverbal behavior (hand gestures) is reproduced along with speech, even if speech and gestures are semantically incongruent. The interaction between the robot and a user was evaluated in three experimental conditions: (1) unimodal—speech statements only, (2) congruent multimodal—semantically matching speech and gestures and (3) incongruent multimodal—semantically non-matching speech utterances and gestures. The authors have revealed an interesting effect: in the third condition, the robot was evaluated as more lively, active, friendly, sociable and cheerful compared to the robot in the second condition. That is, the robot was perceived more positively when the gesture did not correspond to the statement. The researchers suggested that the communicative behavior of the robot is positively evaluated by the user when it is potentially less predictable, and the robot is "imperfect".

In our work, we decided to test the effect of robot's oriented gestures on participants in an experimental study. We assumed that participants would prefer the robot that helps them and also uses pointing communicative actions. The study was conducted using F-2 robot, an experimental platform for studying the interaction between humans and robots. On the one hand, F-2 robot can construct a semantic text representation using the syntax parser [Kotov et al, 2015], [2017], [2018]. On the other hand, the robot selects communicative responses to the constructed meaning and reproduces the gestures and expressional patterns using the behavior management system [Kotov et al, 2019]; [Zinina et al., 2018].

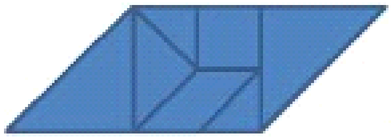
1. Research procedure

In this experiment the robot helped participants to complete a Tangram puzzle. This puzzle is a well-known experimental media for studies of natural human communication [Clark, Wilkes-Gibbs, 1986], development of linguistic resources [Shore et al, 2018], [Gnjatović, Rösner, 2019] as well as for the design of robot communicative strategies [Kirschner et al., 2016]. The puzzle consists of 7 elements of different color, shape and size (two big triangles: red and blue, two small triangles: yellow and dark-blue, a middle green triangle, an orange parallelogram and a purple square—see **Fig. 1**). The task of a participant was to arrange the elements within a given contour on a white sheet. It was allowed to turn the tangram elements upside down.

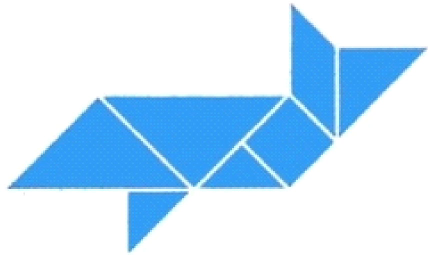
Fig. 1. Tangram puzzle

During the experiment a participant was to complete: *Parallelogram*, *Whale*, *Triangle* and *Ship* figures (**Fig. 2**). The order of tasks presentation was random. The robot was located on the table in front of the participant and used speech instructions, gestures and gaze, instructing the participant to put a certain element on a certain place. Before each task the game elements were placed in front of the participant on the left and right sides of the playing field. Two paired elements (large triangles; small triangles) were always placed on different sides of the playing field. In its speech instructions the robot has always been referring to an element by its shape and size (not by color). Thus, an ambiguous reference in speech had been appearing when the robot mentioned one of the paired elements, such as *Take a big triangle!*

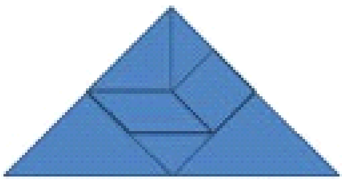
The behavioral scripts of the robot were organized as sequences of BML (Behavior Markup Language) packets: one sequence per task. The experiment was organized in the paradigm of the Wizard of Oz [Kelley, 1984] in which the moves by the player were evaluated as successful or not by a remotely located human operator. The robot was controlled through a Python script that has been sending BML packets to the robot. Special groups of BMLs were developed for a successful move by the player and for an unsuccessful move: wrong position or wrong game element. If the user's actions were correct, the operator gave the robot a command to praise the user and move on. If the participant was mistaken, the robot, according to the operator's command, informed the participant of the error and repeated the previous instruction. If the user has solved the entire figure before the end of BML protocol, the operator gave a command to the robot to praise the participant for successful work. The whole experiment was recorded from two viewpoints: the experimental situation from the side with the view on a player (Fig. 3) and the top view of the playing field (**Fig. 4**).



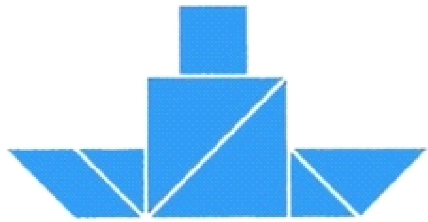
(a) Parallelogram



(b) Whale



(c) Triangle



(d) Ship

Fig. 2. Tangram figures for experimental tasks

Fig. 3. Experimental setup: side view

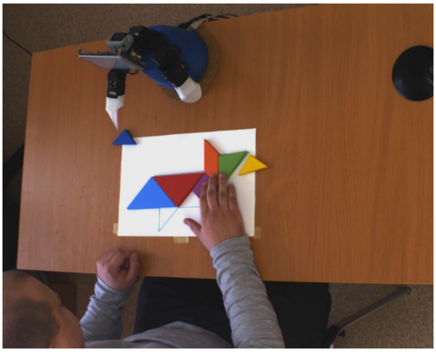


Fig. 4. Experimental setup: top view

The experiment involved 31 participants (12 female, 19 male), at a mean age 27.4. Among respondents 29 (90,7%) held tertiary education qualifications, 2 (9.3%) were students.

2. Experimental conditions

Two experimental conditions had to examine the role of pointing gestures/gaze within the game: in these conditions the robot helped a participant in different ways.

Condition 1: The robot accompanied its instructions by oriented communicative actions: it used pointing hand gestures, head and eye movements (Fig. 5). The robot performed the oriented gestures while instructing the participant, which element to take and where to place it. For example, the robot pronounced: *Take this* (pointed and looked at the element) *little triangle*. In this experimental condition each participant completed *Parallelogram* and *Whale* figures.

Condition 2: The robot did not use pointing gestures, its speech instructions were accompanied by non-oriented movements of hands, head and eyes (Fig. 6). The movements of the robot were selected in such a way as to exclude any directions to the game elements or sides of the playing field. For example, the head tilts were performed along a straight vertical path, eyes could only move up and down, and hand gestures were strictly symmetrical, performed by both hands. In this experimental condition the participant completed *Triangle* and *Ship* figures.

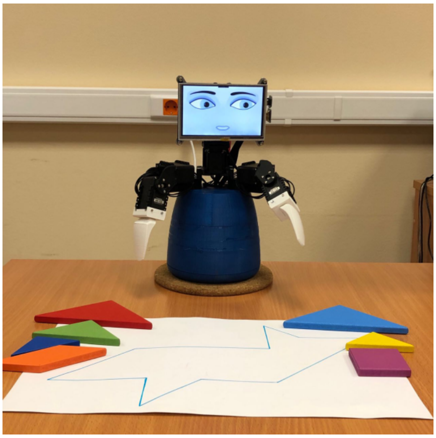


Fig. 5. Condition 1: Pointing gestures

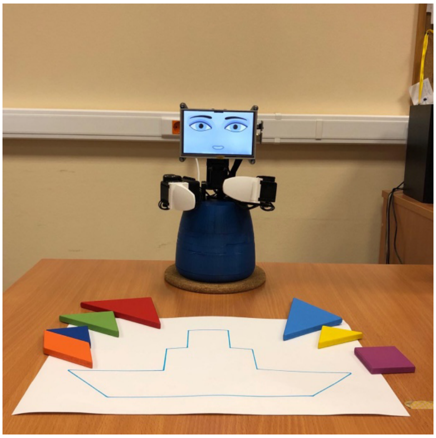


Fig. 6. Condition 2: Symmetric gestures

Tasks with different experimental conditions were presented in random order. After the experiment participants had to fill out a questionnaire: describe the difference in robot's behavior, choose the preferred condition and rate the robot in two experimental conditions on five-point semantic differential scales. In addition, the experiment recorded objective indicators—the speed of solving each task and the number of participants' errors.

3. Results

The experiment took 8.5 minutes on average. *Whale* took the longest time to complete (1 minute 44 seconds) and *Ship* took the shortest time (1 minute 14 seconds) (Fig. 7). *Whale* was also the most complex figure—the participants made the most mistakes while completing this figure. The simplest figure for the participants was *Triangle* (Fig. 8).

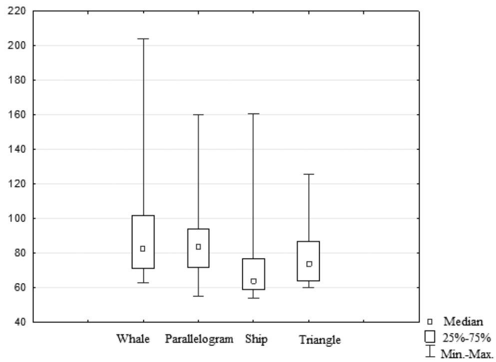


Fig. 7. Tasks solving time (s)

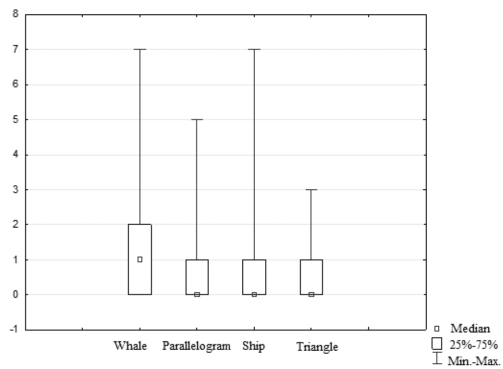


Fig. 8. The number of mistakes

64.5% of the participants preferred the robot pointing to the tangram elements and their locations; 32.3% of respondents equally liked the robot in two experimental conditions. The participants noted that they were waiting for clues from the robot when the conditions were changed. As soon as the robot stopped using pointing gestures, the participants chose the elements longer and more often doubted their positiosn inside the contour. Only one participant (3.2% of the entire sample) preferred the robot that did not use oriented gestures in its instructions. This could be explained, as the participant was mistaken four times while completing *Whale* figure (a task with pointing gestures). Because of this, the robot repeated the instruction four times and the participant, as he has indicated in the self-report, "felt foolish."

The difference between the experimental conditions was not obvious for participants: only half of the players (15 people; 48.4% of the total group) noticed the difference between robot's performance with and without pointing gestures. According to the results of the experiment, most participants (20 people, 64.5% of the total group) significantly preferred the robot with oriented gestures (**Condition 1**) (chi-squared 18.1, $p < 0.01$). Therefore, the assessment can be implicit because the participants didn't always distinguish two experimental conditions, but much more often preferred the robot that indicates the necessary element and its location in the contour with the help of head, eye and hand movements.

At the same time, a nonverbal instruction was significant even for the participants, who did not notice the difference. We have evaluated the participants' choice for a Tangram paired element to analyze the implicit perception of robot's pointing gestures. When two paired elements are not yet used and are placed on the two sides of the playing field, the robot may refer to such an element ambiguously as *big triangle* or *small triangle*. When the robot used a pointing gesture with this reference (on left or right side), the participants followed this nonverbal indication in 91.1% of cases and took the element the robot was pointing at. This can indicate the substantial influence of oriented gestures on the user's behavior, even if the user did not reflect this influence in the self-report.

In the study, we did not identify a link between the preferred condition and whether the participants noticed the difference between the conditions (chi-squared 2.7, $p > 0.05$). According to the collected data, 16 respondents (51.6% of the whole group) stated that they did not notice the difference between the robot that used pointing gestures and the robot that accompanied its instructions with non-oriented movements. These participants in 43.7% of the cases equally evaluated the attractiveness of the robot. However, even these participants followed the robot's gestural instructions in 78.5% of cases. Therefore, one could speak of the implicit influence of the robot oriented gestures on a user behavior, even if the user did not clearly realize this influence.

Moreover, there are cases when participants for several seconds wait for the robot's pointing gesture to resolve the ambiguity (for example, when the participant selects one of the small or large triangles). In addition, we can observe "reverse" in participants' gestures. Reverse is a reciprocating motion when choosing a certain figure. For example, the robot instructs the user: *Take the little triangle*. After that the user brings one hand to one triangle, and another—to the second triangle. The participant can look like "frozen" while he waits for the pointing gesture of the robot.

11 people (73.3%) out of those who noticed the difference in the robot's behavior (15 people, 48.4% of the whole sample) prefer robots using pointing gestures (**Condition 1**). This distribution does not correspond to the normal (chi-squared 9.9, $p < 0.01$), therefore, these parameters are correlated. There are also those participants who equally evaluate robots in different conditions. It can be assumed that these participants preferred verbal instructions, when evaluating robots. The differences between the conditions obtained with the standard scales of the semantic differential is not revealed.

4. Conclusion

We conducted a study that provides further insights into the question of robot-to-human interaction. As we have shown, the robot's oriented gestures in solving a spatial problem are important to give a positive impression on a user. Results showed that participants significantly preferred when the robot used oriented gestures rather than it did not use pointing gestures. This effect is observed even if the participants did not explicitly notice the difference between pointing and non-pointing behavior.

The obtained results open up perspective for a further research on the interaction between robot and humans. Within future studies we plan to evaluate the contribution of expressive means of the robot to its attractiveness for a user, the influence of the shape designation method on the efficiency of solving the problem, etc.

Moreover, the developed system becomes helps to test communicative strategies (for example, using positive or negative feedback) and styles in a dialogue between robot a user as well as is to evaluate the effectiveness of such strategies in different communication situations and in different socio-demographic groups.

References

1. *Beuter N., Spexard T., Lutkebohle I., Peltason J., Kummert F.* (2008) Where is this?—gesture based multimodal interaction with an anthropomorphic robot, 8th IEEE-RAS International Conference on Humanoid Robots (Humanoids 2008), pp. 585–591.
2. *Breazeal C.* (2003) Emotive qualities in lip-synchronized robot speech, Advanced Robotics, V. 17(2), pp. 97–113.
3. *Breazeal C., Scassellati B.* (2002) Robots that imitate humans, Trends in Cognitive Science, V. 6(11), pp. 481–487.
4. *Brugman H., Russel A.* (2004) Annotating Multimedia Multi-modal resources with ELAN, Proceedings of the 4th International Conference on Language Resources and Language Evaluation (LREC 2004), pp. 2065–2068.
5. *Cabibihan J. J., So W. C., Nazar M., Ge S. S.* (2009) Pointing Gestures for a Robot Mediated Communication Interface. In: Xie M., Xiong Y., Xiong C., Liu H., Hu Z. (eds) Intelligent Robotics and Applications. ICIRA 2009. Lecture Notes in Computer Science, V. 5928, pp. 67–77.
6. *Clark H. H., Wilkes-Gibbs D.* (1986) Referring as a collaborative process, Cognition, pp. 1–39.
7. *Gnjatovic, M., Rosner, D.* (2010). Inducing genuine emotions in simulated speech-based human-machine interaction: The nimitek corpus. IEEE Transactions on Affective Computing, 1(2), 132–144.
8. *Häring M., Eichberg J., André E.* (2012) Studies on Grounding with Gaze and Pointing Gestures in Human-Robot-Interaction, Social Robotics. ICSR 2012. Lecture Notes in Computer Science, V. 7621, pp. 378–387.
9. *Kelley J. F.* (1984) An iterative design methodology for user-friendly natural language office information applications, ACM Transactions on Office Information Systems, March 1984, V. 2(1), pp. 26–41.
10. *Kirschner, D., Velik, R., Yahyanejad, S., Brandstötter, M., Hofbaur, M.* (2016). YuMi, come and play with Me! A collaborative robot for piecing together a tangram puzzle. In International Conference on Interactive Collaborative Robotics, pp. 243–251.
11. *Klamer T., Allouch S. Ben, Heylen D.* (2011) Adventures of Harvey: Use, acceptance of and relationship building with a social robot in a domestic environment, Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, V. 59 LNICST, pp. 74–82.

12. *Kotov A., Arinkin N., Filatov A., Zaidelman L., Zinina A.* (2017) Semantic Comprehension System for F-2 Emotional Robot, BICA 2017: Biologically Inspired Cognitive Architectures, pp. 126–132.
13. *Kotov A., Zinina A., Filatov A.* (2015) Semantic Parser for Sentiment Analysis and the Emotional Computer Agents, Proceeding of the AINL-ISMW FRUCT Conference, pp. 167–170.
14. *Kotov A. A., Arinkin N. A., Zaydelman L. Y., Zinina A. A.* (2019) Linguistic Approaches to Robotics: From Text Analysis to the Synthesis of Behavior, Language, Music and Computing. LMAC 2017. Communications in Computer and Information Science, V. 943, pp. 207–214
15. *Lee K. W., Kim H.-R., Yoon W. C., Yoon Y.-S., Kwon D.-S.* (2005) Designing a human-robot interaction framework for home service robot, ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication, DOI: 10.1109/ROMAN.2005.1513793
16. *Salem, M., Kopp, S., Wachsmuth,* (2012) Generation and Evaluation of Communicative Robot Gesture, International Journal of Social Robotics, V. 4(2), pp. 201–2017
17. *Shore, T., Androulakaki, T., Skantze, G.* (2018). KTH Tangrams: A Dataset for Research on Alignment and Conceptual Pacts in Task-Oriented Dialogue. In Eleventh International Conference on Language Resources and Evaluation (LREC 2018), pp. 768–775.
18. *Zinina A. A., Arinkin N. A., Zaydelman L. Y., Kotov A. A.* (2018) Development of communicative behavior model for f-2 robot basing on «REC» multimodal corpora [Razrabotka modeli kommunikativnogo povedeniya robota f-2 na osnove mul'timodal'nogo korpusa «REC»], Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference "Dialog 2018" [Komp'yuternaya Lingvistika i Intellektual'nye Tekhnologii: Trudy Mezhdunarodnoy Konferentsii "Dialog 2018"], V. 17 (24), pp. 831–844.