Effectiveness of verbal communication with conjunction of gestures and facial expression in human-robot interaction

Rahul Ramachandran
Faculty of Engineering and Technology
University of the West of England
Bristol, United Kingdom
rahul2.ramachandran@live.uwe.ac.uk, Group 14

***Abstract*—We conducted a user study to evaluate the optimal level of verbal and non-verbal communication and the use of facial emotion recognition which the robot should use during a human-robot interaction for a given specific task. This research question was chosen based on previous studies done by psychologists on multi-modal communication in human-human interaction and to see if it applies the same in human-robot interaction. It was also our aim to explore if participants rate the robot interacting with the human only with verbal communication different from a robot interacting with both verbal and non-verbal communication and which reactions people show during the interaction in both cases. Our results show that efficiency of the task improved when the robot uses more relevant social signals during the interaction. We also found that use of facial emotion recognition by the robot was not helpful during the task but the participants found it to be more enjoyable and friendly.**

***Index Terms*—emotion recognition, human-robot interaction, human-human interaction, social signals.**

I. Introduction

It is quite natural for us humans to make physical movements even when doing nothing, but it is not the same when it comes to robots. Robots are motionless when doing nothing and are usually stock-still while having a verbal communication with a human. Previous studies have presented related knowledge to generate speech-based motions of virtual agents on computer screens. However, social robots during a task collaboration share space and time with humans, and thus, numerous speechless or inactive situations occur were the robot cannot be hidden from users. Gestures or some idle motions are used a lot in movie animations and gaming (3; 4). However, only few studies investigated the role of gestures and postures in relation to making robots more social entities. For example, (5) mimicked some basic gestures and postures made by a clerk on a robot, but the effect on social interaction was not studied.

Since robots in human-robot interaction (HRI) are social actors, they elicit mental models and expectations known from human-human interaction (HHI) (1). One aspect that we know from HHI is that during a HHI, there is a common understanding that the use of verbal and non-verbal signals influence perceivers to produce responses in others that are beneficial to signalers and also guides the perceivers to achieve the specific task effectively and comfortably in a task collaboration. Marc Mehu in (2) argued that use of verbal and non-verbal social signals in an interaction play different roles in the processes of information transfer and social influence.

In our effort to make robots more social entities and in order to minimize the gap between the human and a robot (that is "machine like feel"), we propose to specifically explore what level of each components of social signals of a robot to be used during a task specific HRI by studying the effects on social interaction with the humans. The term social signal is used to describe verbal and non-verbal signals that the robot use in a conversation to communicate their intentions. We report on a user study where we studied the effect of use of different levels of social signal by the robot during a task specific social interaction.

II. Related Work

In an earlier work, (6) performed a user study in which a robot help a human to unpack a cardboard moving box containing sixteen objects. There aim was to study the social effects of idle and meaningful motions of the robot during a task specific interaction. They found that when the robot makes idle motions, humans perceive the robot to be more human-like and alive but not intelligent. Only when the robot used meaningful motions, it was perceived to be socially competent and intelligent. In their experiment, they studied only the effects of gestures and postures whereas our study extends the boundaries to effects of verbal, gestures, and use of facial emotion recognition by the robot during task specific interaction with humans.

A user study by (8) evaluates the different modalities for attracting attention of the robots human partner using social signals. The results from (8) shows that actions which involve sound generate the fastest reaction times and are better perceived by participants and non-verbal meaningful gestures have been demonstrated to improve user experience and communication efficiency. They also found that attempting to attract attention by establishing eye-contact resulted in worse participants perception. Our study and that of (8) cover different aspects: (a) all the participants were elderly people between the age of 62 to 70 years, but in our case the age ranged from 25 to 58 years; (b) There user study was not task specific, ours is task specific; (c) They used eye gaze to establish in a communication with the participant, whereas in our work, it was not necessary since the users attention was on the objects and task completion and not on the robot all the time.

A recent illustrative experimental study by (7) shows how non-verbal cues can both interfere and facilitate communication when passing a message to a user in HRI. They conducted two experiments, in experiment one they studied the effect of non-verbal cues based on behavioural styles (that is permissive and authoritative) being perceived as verbal by participants. And in experiment two they studied the facilitation of non-verbal understanding by congruent verbal signals in their users. They found that verbal and non-verbal communication can facilitate the understanding of information conveyed by the robot when combined appropriately. Our work and that of (7) cover different aspects: (a) their interaction is not task related, ours is task specific; (b) they did not consider the facial emotion recognition of the user, while we looked into the understanding of users emotions by the robot and its effects during the social interaction.

III. Methods

We setup a Wizard of Oz(WOz) user study to specifically evaluate the effectiveness of verbal communication with conjunction of gestures and facial emotion recognition by the robot. A human and a robot interacted with each other in four verbal sessions. All the four sessions had a similar sequence were the robot helps the human by giving instructions to complete a Lego task. In the Lego task the human interacts with the robot and follows the instructions given by the robot to build a simple object. We chose this task in order to keep the task simpler and less time consuming.

The user study was performed within subjects, with each participant taking part in all of the following conditions:

- No robot gestures (experimental condition - only verbal communication between human and the robot).
- Task specific gestures (experimental condition in which the robot makes some gestures related to the task).
- excessive use of gestures (experimental condition in which the robot uses many gestures while giving instructions).
- Face and emotion recognition (experimental condition in which the robot uses face and emotion recognition alongside gestures and verbal communication).

A. Hypotheses

This user study drives hypotheses in three key areas, which are verbal communication, optimal use of gestures and use of facial expression recognition by the robot. The following hypotheses were postulated for this user study.

H1: Use of task specific gestures by robot alongside verbal communication during a task specific interaction is perceived as better understandable than a robot which does not use any gestures and only verbal communication.

H2: A robot that uses task specific gestures is perceived as more efficient than a robot that uses too many gestures during a task specific interaction.

H3: A robot that uses facial emotion recognition is perceived as more enjoyable than a robot that doesn't.

B. User Study Design

The participants were asked to interact with a NAO robot [1] developed by Aldebaran Robotics in our WOz user study. We set the interaction in four LEGO sessions. As mentioned in III each LEGO session had a experimental condition and all the sessions were run within subjects.

All the participants were asked to seat opposite to the robot, ensuring they are at eye level, with three different sets of coloured blocks placed equidistant between the participant and robot (see figure 2 for the setup). The participants were then instructed by the robot to select certain blocks out of a total of 8 blocks of three different colours (red, green and blue) and place in a designated area with a specific block pose.

In all the four LEGO sessions, the participants had to assemble LEGO blocks according to the robot's instructions. The tasks were assigned in the same order and all the four LEGO sessions were studied in the same order with all the participants. The first LEGO session lasted for an average of 1 min and 85 s (SD = 36 s), for second LEGO session the average time taken was 1 min and 33 s (SD = 11 s), for third LEGO session the time taken on an average of 1 min and 27 s (SD = 3 s) and for the fourth LEGO session the average time taken was 1 min 61 s (SD = 26 s).

The user study was performed in robotics teaching room at the Bristol Robotics Lab. The LEGO blocks and the robot was placed on a table in front of a wall as you can see from figure 1. The participant was sitting in front of the robot at a distance of approximately 1.5 m and the researcher was behind the participant at a distance of approximately 4 m operating the system and two other researchers were standing to the right side of the participant approximately at a distance of 4 m, one researcher controlled the video recording and the other researcher was observing the session and was readily available to handle any technical difficulties during the session and also to rearrange the LEGO blocks and the NAO robot at the end of each session for the next session.

The within-subjects design required each participant to take part in all four sessions of the user study. In the first session, the experimental condition was that the robot uses only verbal communication (speech) to give the instructions. In second session, the robot uses meaningful or task specific gestures along with speech to interact with the participant, whereas in third session the experimental condition was that the robot uses many gestures and some body postures along with meaningful

[1]https://www.softbankrobotics.com/emea/en/nao

gestures and speech. It also does some gestures slowly to get the attention of the participant. In fourth session of the user study, the robot uses face and emotion recognition alongside gestures and verbal communication. Except for the fourth session, in all other sessions the researcher did not interfere during the study. The researchers only intervened in the rare cases of the fourth session when ever the robot fails to recognize the participants emotion.During such situations the researcher simply requested the participant to gaze at the robot with a normal face. Table I gives an overview on the LEGO tasks and NAO robot's action types.

Fig. 1: Study setup with the participant interacting with the robot.

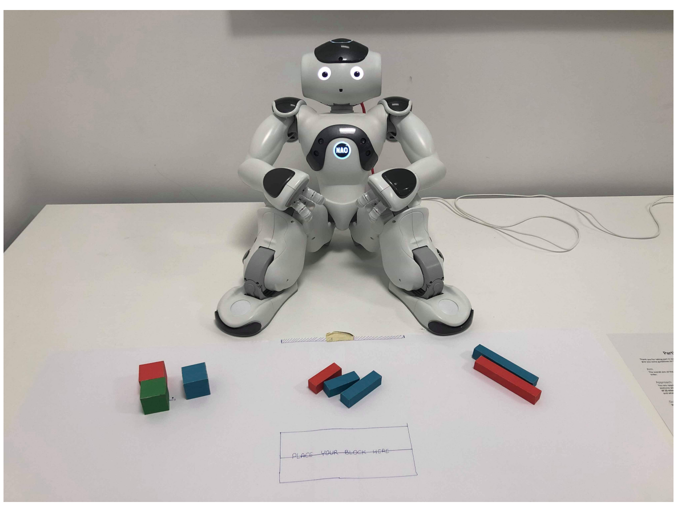

Fig. 2: Lego blocks that were provided to the participants.

C. User Study Procedure

The participants were welcomed to the Bristol robotics lab's robotics teaching room were our user study took place. Participant information sheet was provided first to give the participants with the necessary information about the user study and what will happen to the data collected during the study. After that a short briefing note explaining the aim, approach about the user study and also explaining the necessary guidelines to be followed once the experiment starts. Soon after the short briefing, the participants were requested to sign an informed consent. Next, the participants were asked to complete questionnaires form to asses their demographics. The participant was introduced to the robot and they were given an overview on the process of the first session. As soon as the participants took their position opposite to the robot, the user study began. First the robot offered greetings to the participant and later offered to assist in completing a Lego block building task. The robot would instruct the participant to pick up a Lego block based on colour (e.g. Pick Up Red Piece) and await any further questions from the participant to aid them in selecting the correct block (e.g. Which red piece?) to which the robot would respond, if necessary, with a description of the block shape (e.g. The long one). The second part of the verbal instructions was the robot stating where to place the block (e.g. Place next to blue piece) followed by, if necessary, a question from the participant to further clarify (e.g. how?). At the end of each session, the participants were again asked to complete the questionnaire assessing about their interaction with the robot. The study was finished with an ending interview taken by the researcher to the participants. The researcher asked the participants three open-ended questions, which were followed by a short debriefing in which the purpose of the study was explained. The study procedure is depicted in figure 3.

D. Dependent Measures

Before the start of user study, we asked the participants to fill a questionnaire. We used to this questionnaire to analyze the participants technology affinity and pre-experience with the robots and their age. The questionnaire consisted of six items out of which for two items we used 5-point Likert scale range.

For the dependent measures from the four sessions with each participant we looked into two factors, efficiency and enjoyability of task completion since they are important factors in a task specific human robot interaction. In both factors we used quantitative and qualitative measurements. We used 5-point Likert scale range (1 - strongly disagree to 5 - strongly agree) for quantitative measurements.

E. Participants

A total of 7 participants took part in our user study among them 5 were male and 2 were female. The participants were recruited over Bristol robotics laboratory mailing list. They were primarily staff from Bristol robotics laboratory and all of them had prior experience with robots. Their age ranged from 25 to 58 years, with a mean age of 38.28 years (SD = 11.12). With regards to the conditions, all the conditions were tested within-subjects. The participants technology affinity was rated on average with a mean of 4.14 (SD = 1.35; 5-point Likert-scaled ranging from 1 - "not technical to 5 - "very technical") and their pre-experience with the robots was above average

TABLE I: LEGO Task Session

Step	***Who?***	***Action Type***	***Action***	***Robot gesture***
1	User	Voice Recognition	Hello	
2	Robot	Animated Say	Do you need help with Lego, Again?	
3	User	Voice Recognition	Yes	
4	Robot	Animated Say	Pick Up Red Piece, OK ?	Pointing with right hand
5	User	Voice Recognition	Yes	Back to rest position
6	Robot	Animated Say	place that in the middle	
7	Robot	Animated Say	then, Pick up blue piece, OK ?	Pointing with right hand
8	Robot	Voice Recognition	Yes	Back to rest position
9	Robot	Animated Say	Place next to red piece	
10	User	Voice Recognition	where?	
11	Robot	Animated Say	Parallel to red piece	
12	Robot	Animated Say	OK?	
13	User	Voice Recognition	Yes	
14	Robot	Animated Say	Pick Up Red Piece	Pointing with left hand
15	Robot	Animated Say	Attach at the end of both pieces	
16	Robot	Animated Say	OK?	Back to rest position
17	User	Voice Recognition	Yes	
18	Robot	Animated Say	Pick up blue piece	Pointing with left hand
19	Robot	Animated Say	Attach other end	
20	Robot	Animated Say	OK?	Back to rest position
21	User	Voice Recognition	Yes	
22	Robot	Animated Say	Pick up green piece	Pointing with left hand
23	Robot	Animated Say	Attach to middle	
24	Robot	Animated Say	OK?	Back to rest position
25	User	Voice Recognition	Yes	
26	Robot	Animated Say	great, now it's done	

This LEGO Task session is for the second condition

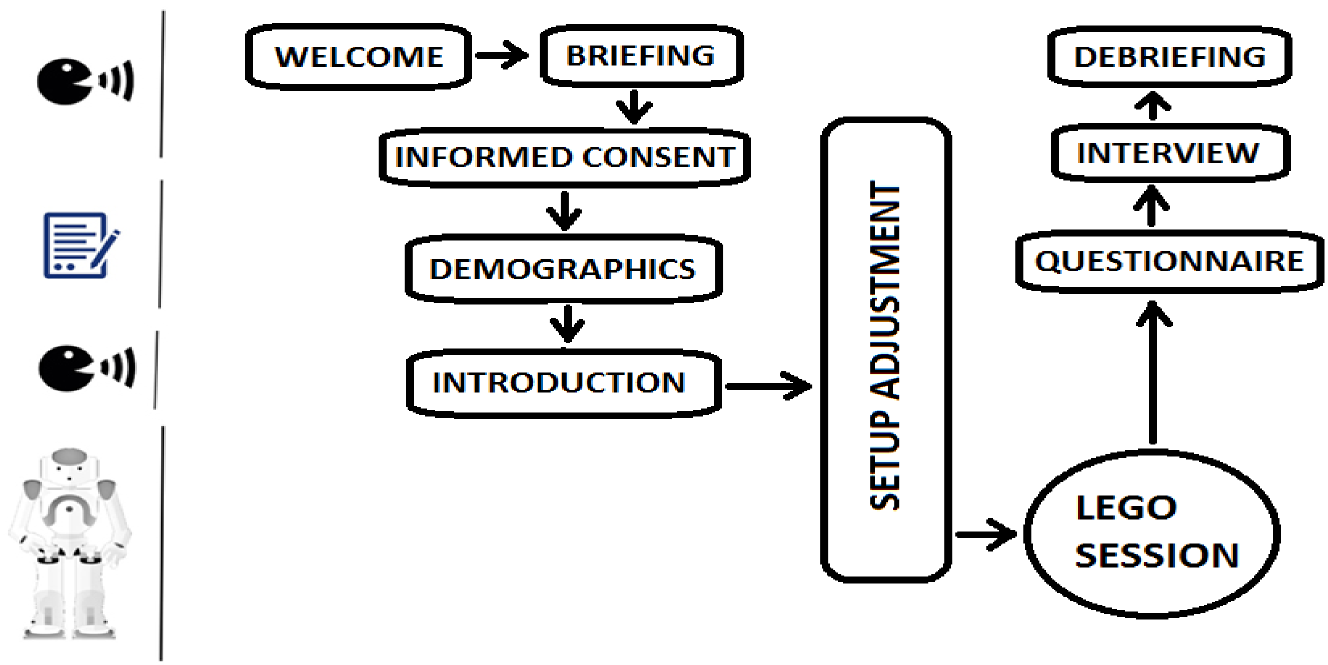

Fig. 3: Study Procedure

with a mean of 3.71 (SD = 1.03; 5-point Likert-scaled ranging from 1 - "not at all familiar" to 5 - "extremely familiar").

IV. Results

For the data analysis, we used non-parametric statistical test procedures since our data were mostly not normally distributed. Mann-Whitney-U tests were used to compare the two experimental conditions related to each hypotheses. For the hypotheses one, the two conditions were condition one (LEGO session 1) and condition two (LEGO session 2) and for the second hypotheses, it was between condition two and three (i.e LEGO session 2 and 3) and for the third hypotheses the conditions were one and four (i.e LEGO session 1 and 4).

We measured participants rating of the robot at the end of each session. To do so, we asked the participants to complete a questionnaire. We used 5-point Likert scale range (1 -

strongly disagree to 5 - strongly agree) for the questionnaire. In order to explore if the participants who experienced the robot interacting with some task specific gestures different from the robot which interacted with only speech, we conducted Mann-Whitney-U tests (see Table II). Participants have rated the robot using task specific gestures and speech as more understandable than the robot which uses only speech during the interaction. This confirms that our first hypotheses is true.

It was also our aim to study how the participants rate the robot when it uses irrelevant gestures along with task specific gestures and speech during the interaction. Our test results from the users rating shows that the task efficiency decreases as the gestures used by the robot became more irrelevant. For the mean values(SD) and effect size see table IV. This difference yielded in a medium effect size.

We also conducted the same test between experimental condition one and four to see how the participant rate when the robot uses facial emotion recognition during the interaction. Interesting the participants rated the use emotion recognition to be not helpful for the task but they found it to be enjoyable during the interaction. See table III for the mean values(SD) and effect size. From this test, we can confirm that our third hypotheses is true.

TABLE II: Mean values(SD) compared between the conditions for the first hypotheses

Measure	Condition 1	Condition 2
Mean	2.43	3.57
SD	1.17	1.39
d	0.89	0.89
Rank Sum	41.5	63.5
Mann-Whitney-U statistic	13.5	13.5

Q:I found the instructions given by the robot clear and easy to understand

TABLE III: Mean values(SD) compared between the conditions for the second hypotheses

Measure	Condition 2	Condition 3
Mean	4.28	3.28
SD	0.45	1.27
d	1.16	1.16
Rank Sum	63.5	41.5
Mann-Whitney-U statistic	13.5	13.5

Q:I found the amount of gesturing adequate

TABLE IV: Mean values(SD) compared between the conditions for the third hypotheses

Measure	Condition 1	Condition 4
Mean	3.0	3.28
SD	1.06	1.15
d	0.25	0.25
Rank Sum	47.5	55.5
Mann-Whitney-U statistic	19.5	19.5

Q:I found the engagement with robot more enjoyable

V. Discussion

Our results show that the efficiency of the task improved as the robot uses more relevant social signals in a human robot interaction. The results also indicate that use of facial expression recognition by the robot is not helpful during the task but was enjoyable during the interaction by the participants. Our result confirm the findings of (6), who also studied the effect of idle and meaningful motions (gestures) in a task specific HRI.

We conducted the user study within subjects since it requires fewer participants but in future we will do the user study between subjects and with as many participants as possible so that we can have a more reliable data to conduct the use many tests to evaluate the data. One of the limitations in our user study was that we had many people inside the robotics teaching room were we conducted the user study, which once influenced the study with one participant and also the researchers of this study were not completely wizarded from the participant and the experimental setup. Although we video recorded all the LEGO sessions, we were unable to do the qualitative data analysis due to technical difficulties.

We asked all our participants for the opinions about the user study after the completion of the LEGO session, it was interesting to note that many participants mentioned that the gestures made by the robot in third LEGO session was slow and two participants mentioned that the robot in fourth LEGO session was found to be more friendly but at the same time they were expecting the robot to interact more and not just being task specific.

The author of this report and his group member Chathura Semasinghe programmed the NAO Robot for the conditions two, three and four. During the user study the group members responsibilities were as follows, Ross Harrison was responsible for greeting the participant, briefing about the study and Debriefing. Huiwen Tan was responsible for video recording all the sessions, Zihao Xiong was responsible for filling the qualitative measurement sheet and getting the signature on the information consent form. The author of this report (Rahul Ramachandran)and Chathura Semashighe were the Robot technicians. This report is written by the author and other group members didn't make any contribution towards this report writing.

VI. Conclusion

With our user study we explored how people rated a robot when it used gestures and facial emotion recognition during a task specific HRI. We measured the robots effectiveness when it uses different social signals and user experience during the same. We found out that the robot was more enjoyable but not helpful during the LEGO task session and also found that the efficiency of the task improves when task specific gestures were used by the robot during the interaction with the users. Our results confirm existing HRI research on social signals of robots such as (6) and (8), showing a pattern that the effectiveness of interaction with the humans can be improved by using appropriate social signals by the robot. As part of next steps and future work, it needs to be studied the influence of non-verbal social signals over verbal communication between the human and robot and set of guidelines can be proposed based on an extensive user study with more number of participants.

Acknowledgment

The author of this paper would like to thank the module leader for Human-Robot Interaction module Professor Manuel Giuliani and the module team Professor Praminda caleb-Solly and Dr. Severin Lemaignan for their support and technical advice. The author would like to thank the Human-Robot Interaction module practical session instructor Anouk van Maris for her advice and support throughout the user case study. The author would also like to thank Chathura Semasinghe, Ross Harrison, Zihao Xiong and Huiwen Tan for their contribution during the user case study.

References

[1] Manja Lohse, “The role of expectations and situations in human-robot interaction,”New Frontiers in Human-Robot Interaction, pp. 35–56, 2011.

[2] Mehu and Marc, “The integration of emotional and symbolic components in multimodal communication” Frontiers in Psychology, vol. 6, pp. 961, 2015.

[3] Ribeiro T. and Paiva A., “The illusion of robotic life: principles and practices of animation for robots,” in Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction, pp. 383–390, 2012.

[4] A. Egges, T. Molet, and N. Magnenat-Thalmann, “ Personalised real-time idle motion synthesis,” in Proceedings of the 12th Pacific Conference on Computer Graphics and Applications, pp. 121–30, 2004.

[5] Hyunsoo Song, Min Joong Kim, Sang-Hoon Jeong, Hyen-Jeong Suk and Dong-Soo Kwon, “Design of idle motions for service robot via video ethnography,” in Proceedings of the 18th IEEE International Symposium on Robot and Human Interactive Communication, pp. 195–99, 2009.

[6] Raymond H. Cuijpers and Marco A. M. H. Knops, “Motions of Robots Matter! The Social Effects of Idle and Meaningful Motions,” in Social Robotics, Springer International Publishing, pp. 174–183, 2015.

[7] Wafa Johal and Gaëlle Calvary, and Sylvie Pesty, “Nonverbal Signals in HRI: Interference in Human Perception,” in Social Robotics, Springer International Publishing, pp. 275–284, 2015.

[8] Elena Torta, Jim Van Heumen, Raymond H. Cuijpers, James F. Juola, “How can a robot attract the attention of its human partner? a comparative study over different modalities for attracting attention,” in Social Robotics, Springer Berlin Heidelberg, pp. 288–297, 2012.