*The 33rd Annual Conference of the Japanese Society for Artificial Intelligence, 2019*

# Eye-gaze in Social Robot Interactions – Grounding of Information and Eye-gaze Patterns

Koki Ijuin[*1] Kristiina Jokinen[*2] Tsuneo Kato[*1] Seiichi Yamamoto[*1]

[*1] Doshisha University [*2] AIRC, AIST Tokyo Waterfront

This paper examines human-robot interactions and focuses on the use of eye-gaze patterns in evaluating the partner's understanding process. The goal of the research is to understand better how humans focus their attention when interacting with a robot and to build a model for natural gaze patters to improve the robot's engagement and interaction capabilities. The work is based on the AIST Multimodal Corpus which contains human-human and human-robot interactions on two different activities: instruction dialogues and story-telling dialogues. The preliminary experiments show that there are differences in the eye-gaze patterns given expected and non-expected responses, which affects their understanding and grounding of the presented information. The paper corroborates with the hypothesis that eye-gaze patterns can be used to predict grounding process and provide information to the speaker about how to proceed with the presentation, so as to support the partner's understanding and building of the mutual knowledge. Some consideration is given to future improvements in methodology.

## 1. Introduction

The goal of the research is to understand better how humans focus their attention when interacting with a robot and to build a model for natural gaze patters to improve the robot's engagement and interaction capabilities. The work follows from the pilot study (Jokinen 2018) in which human gaze patterns were studied when they interacted with a humanoid Nao robot using the WikiTalk application (Jokinen and Wilcock 2014) which allowed the user to navigate among Wikipedia topics, and is also related to eye-gaze in second-language learning with robots (Fujio et al. 2018).

In this paper, we focus on eye-tracking technology and its use in instruction giving and story-telling activities. The hypothesis examined in the paper is that there is a difference between the interlocutor's eye-gaze patterns depending on how their understanding proceeds, i.e. if the partner's utterance is understood, misunderstood or non-understood. By measuring eye-gaze activity in the communicative context we build a model that enables estimation of the partner's level of understanding, and consequently, modification of the presentation if the partner eye-gaze signals problems in the grounding of information. Such a model can help the humanoid robot to better tailor its presentations to the human user, i.e. to enable the use of the partner's eye-gaze signals to establish an appropriate way to continue. In particular, it will enable us to study how eye-gaze is used in grounding information and creating mutual understanding of the discussion topic.

Smooth interaction requires that the partners can easily understand each other and are able to build their conversation on mutual knowledge of what has been discussed. The process of creating such mutual knowledge is called grounding, i.e. the partners ground the semantics of their utterances in the context of their interaction and the context of their world knowledge, Clark & Wilkes-Gibbs (1986).

Earlier work shows that in social situations, humans are sensitive to another person's gaze: it constructs new shared knowledge, communicates experiences, and creates social linkages (Argyle & Cook, 1976; Kendon, 1967). Visual attention is important in cognitive studies (Skarrat et al. 2012), and e.g. turn-taking is commonly coordinated by gaze (Jokinen et al., 2013). Broz et al. (2015) provides an overview of the work on eye gaze and human robot interaction.

Contact: Kristiina Jokinen, AIRC AIST Tokyo Waterfront, Koto, Aomi, Tokyo, *kristiina.jokinen@aist.go.jp*

## 2. Data collection setup

Experiments were set-up using the lab's SMI eye-tracker and the Nao robot, to collect eye-tracking data. Fig 1. Shows the setup. Each participant had two conversations, one with a human partner (HHI) and one with a humanoid robot (HRI). The conversations were about 10 minutes long. One of the experimenters played the role of the human partner but was different from the one who gave instructions to the participant.

The experiment was conducted in Japanese or English depending on the participant's preferred language. The instructions were the same for both HRI and HHI conditions. Before the experiments, the participants signed a consent form and filled in a pre-experiment questionnaire of their background and expectations. After each interaction (HRI and HHI), they filled in another questionnaire focussing on their experience in the interaction.


Data consists of 30 participants (20 Japanese, 10 English), each having both HHI and HRI conversation. The participants (10 female) were students and researchers, age 20-60, with experience on IT, but no experience on robots. Of the participants, 14 had instruction and 16 chat dialogues.

Data analysis has started using the standard gaze frequency and duration measurements, annotations and statistical analysis,

to analyse the user's gaze patterns during interaction with the robot and with a human partner. Special attention is paid to gesturing and head nodding, and conversational instances such as turn-taking, feedback, and problem cases.

## 3. Annotation and Analysis

### 3.1 Annotation Method

Annotation for duration of utterances were done with automatic silence segmentation of ELAN. The audio files which were recorded with eye tracker were used to annotate utterances. Automatic silence segmentation was conducted two times with different thresholds of loudness for determining the silence. The one with low threshold annotates both participant's and partner's utterances, and the other one with high threshold annotates only participant's utterances. The values of those thresholds were manually set by each conversation. The segmentation of partner's utterances was calculated by subtracting those automatic segmentations.

After the segmentation, the participant's utterances in human-robot conversations were manually classified into four types from the perspective of robot's feedback to that utterances: Correct-Understood, Miss-Understood, None-Understood, and Other. Correct-Understood (CU) were tagged to the utterance which robot recognized and gave the correct feedback, Miss-Understood (MU) were tagged to the utterance which robot recognized but gave the unexpected or wrong feedback, and None-Understood (NU) was tagged to the utterance which robot did not recognized and did not give any feedback.

To annotate the eye gaze activities of participants automatically, we created the robot detection system with OpenCV3. We used cascade classifier to detect the position of robot's face in video recorded with eye sight camera of eye tracker. The robot's face was detected as rectangle, and the rectangle of robot's body was estimated with the position of robot's face. After detecting the robot's face and body, the eye gaze activities were automatically annotated into two groups; Gaze Face and Gaze Body, by judging whether the coordinates of gaze point captured by the eye tracker were in those detected rectangles or not. Fig 2 shows the result of robot detection system and gaze point of the participant.

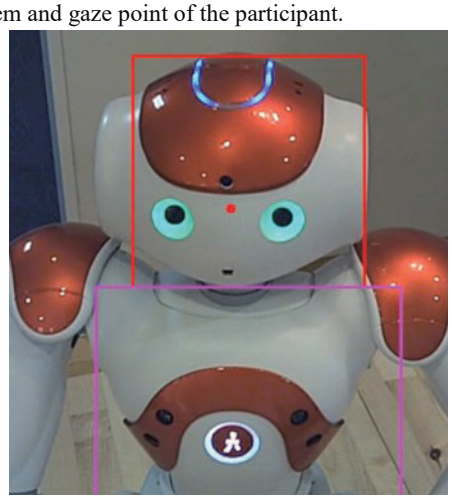

**Figure 2 Snapshot of the result of robot detection system. Red rectangle represents the robot's face, purple rectangle represent, the estimated robot's body, and red dot represents the point of participant's gaze**

### 3.2 Methodology of analyses of eye gaze activities

In order to verify how the participant uses his/her eye gaze activities, we conducted quantitative analyses of eye gaze activities during utterances, pauses just before the beginning of utterances, and pauses just after the end of utterances.

We used Gazing Ratios in order to understand how the participant uses eye gaze activities during the human-robot conversations.

Gazing Ratio is defined as:

$$\frac{1}{N}\sum^{N}\left(\frac{DG_{(i)}}{\text{duration of } i\text{-th } w}\right.$$

where $DG_{(i)}$ represents that the duration of participant's gaze toward the robot during *i*-th window. Three types of windows were used: before utterance window, during utterance window, and after utterance window. *N* represents the total number of windows. Gazing Ratio was calculated for each utterance type (CU, MU, and NU).

## 4. Preliminary Results

The Gazing Ratios were calculated with the data of 19 participants. Figure (ppt p. 26) shows that the temporal change of Gazing ratio for each utterance type in human-robot conversations. The results of eye gaze activities show that the participants tend to gaze away from the robot after they finishes speaking regardless of correctness of robot's feedback. After the robot gives feedback, the participants shift their gaze to the robot again. However, when the robot does not give any feedback to the participants, the participants keep gaze away to the robot for a while, and then they gaze at the robot again.

These results suggest that after the participants answer the

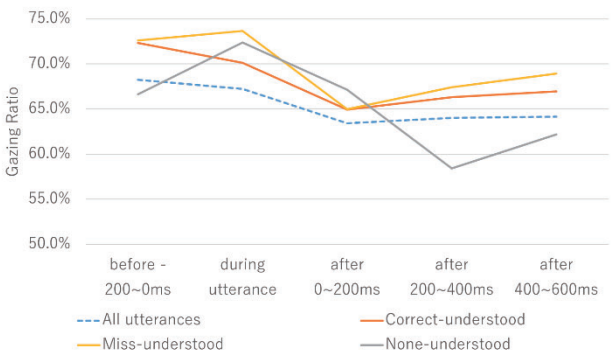


**Figure 3 Gazing Ratios of participant during utterances, before the beginning of utterances, and after the end of utterances**

question from the robot, they gaze away from the robot during the interval that the robot starts speaking, and if the robot does not start speaking, the participants soon realize that there is something wrong with the conversation so that they gaze at the robot in order to monitor what is going on to the robot.

This quantitative difference of eye gaze activities according to the robot's reaction might be useful to predict the participant's state whether he/she is waiting for the robot's feedback or not.

## 5. Conclusions

This paper has presented preliminary studies concerning eye-gaze in human-robot interaction and focused especially on the understanding of the presented information. Such grounding is important for the human-robot interaction to progress smoothly and for the robot to exhibit context-aware capability, i.e. be able to take the user's multimodal signals in the given conversation al context into account. The work is based on the AIST Multimodal Corpus which includes eye-tracking data on natural interactions between two humans and between a human and a robot.

The work continues on further analysis and annotation of the corpus and building computational models of the use of eye-gaze in signaling the interlocutors' understanding. As the corpus contains both human-human and human-robot interactions in similar conversational situations, further research is focused on studying the differences in the human gaze behaviour in the two types of conversational settings. This will deepen our knowledge of the function of gazing in interaction in general and the role of being able to detect and analyse gaze-patterns also in human-robot interactions.

## Acknowledgement

This paper is based on results obtained from *Future AI and Robot Technology Research and Development Project* commissioned by the New Energy and Industrial Technology Development Organization (NEDO).

## References

Argyle, M. and Cook, M. (1976). Gaze and Mutual Gaze. Cambridge University Press, Cambridge.

Broz, F., Lehmann, H., Mutlu, B., and Nakano, Y. (Eds.) (2015). Gaze in Human-Robot Communication. John Benjamins Publishing Company.

Clark, H. & Wilkes-Gibbs D. Referring as a collaborative process. *Cognition 22:*1-39, 1986

Fujio, S., Ijuin, K., Kato, T., Yamamoto, S. (2018). Measurement of Gaze Activities of Learners with Joining-in-type RALL System, Proceedings of the 2018 IEICE General Conference, March 2018 (In Japanese)

Jokinen, K., Furukawa, H., Nishida, M., Yamamoto, S. (2013). Gaze and Turn-taking behaviour in Casual Conversational Interactions. ACM Transactions on Interactive Intelligent Systems (TiiS) Journal, Special Section on Eye-gaze and Conversational Engagement, Vol 3, Issue 2.

Jokinen, K. and Majaranta, P. (2013). Eye-Gaze and Facial Expressions as Feedback Signals in Educational Interactions. In D. Griol Barres, Z. Callejas Carrión, R. López-Cózar Delgado (Eds.) Technologies for Inclusive Education: Beyond Traditional Integration Approaches. Chapter 3, pp.38-58. Hershey, PA: Information Science Reference, IGI Global.

Jokinen, K. and Wilcock, G. "Multimodal Open-domain Conversations with the Nao Robot." In Natural Interaction with Robots, Knowbots and Smartphones - Putting Spoken Dialog Systems into Practice. Springer Science+Business Media, 2014. pp. 213-224

Kendon, A. (1967). Some functions of gaze–direction in social interaction. *Acta Psychologica*, 26 (1): 22–63.

Skarratt, P.A. et al. 2012. Visual cognition during real social interaction. *Frontiers in human neuroscience*. 6, (Jan. 2012), 196