Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020), pages 559–564
Marseille, 11–16 May 2020

The AICO Multimodal Corpus – Data Collection and Preliminary Analyses

Kristiina Jokinen
AI Research Center AIST Tokyo Waterfront
2-4-7 Aomi Koto-ku
Tokyo 135-0064 JAPAN
Kristiina.Jokinen@aist.go.jp

Abstract

This paper describes data collection and the first explorative research on the AICO Multimodal Corpus. The corpus contains eye-gaze, Kinect, and video recordings of human-robot and human-human interactions, and was collected to study cooperation, engagement and attention of human participants in task-based as well as in chatty type interactive situations. In particular, the goal was to enable comparison between human-human and human-robot interactions, besides studying multimodal behaviour and attention in the different dialogue activities. The robot partner was a humanoid Nao robot, and it was expected that its agent-like behaviour would render human-robot interactions similar to human-human interaction but also high-light important differences due to the robot's limited conversational capabilities. The paper reports on the preliminary studies on the corpus, concerning the participants' eye-gaze and gesturing behaviours, which were chosen as objective measures to study differences in their multimodal behaviour patterns with a human and a robot partner.

Keywords: eye-tracking, gesturing, multimodal corpus collection, human-human and human-robot dialogues

1. Introduction

Current development of interactive robot agents is backed by extensive research on methods and tools concerning neural models and big data, as well as symbolic and rule-based systems incorporating models for knowledge, reasoning and cooperation (Siciliano and Khatib, 2016). Research has been conducted on verbal and non-verbal communication and building of multimodal systems (see an overview in Almeida et al. 2018), but investigations comparing human multimodal behaviour in interactions with a human or a robot partner are few.

In human-human interactions (HHI), multimodal signals play a fundamental role in turn management, feedback, and meaning creation: they are related to coordination of conversation and building of a shared context in which to achieve task goals, seek for information, and form social bonds. By extension, such behaviour is important also in human-robot interaction (HRI), since the manner of interaction by which humans effectively respond to signals that indicate the partner's (mis)understanding, agreement and emotional state is intuitively used also when interacting with social robots (Jokinen, 2019).

Humans perceive verbal and non-verbal communication in an effortless manner, however, modelling of social signals in HRI is still less common, less smooth, and less effective for serving communicative goals. In experimental settings users often evaluate the robot's communicative patterns as inflexible and monotonous, and comment that the robot talks too much: the robot agent does not provide similar feedback or non-verbal engagement as human partners.

This paper discusses our data collection as a starting point to compare human behaviour in HHI and HRI. The main goal is to study understanding, engagement, and attention of human participants in various interaction activities, and to enable comparison between similar human-human and human-robot interactions. It is expected that interactions with an agent-like robot show similarities with human-human interactions, but also differ due to the robot's limited conversational capabilities (turn-taking, feedback, understanding). We explore the differences through the participants' multimodal behaviour and concentrate especially on visual attention (eye-tracker data). The corpus also contains video data which has been used for gesture studies and personality experiments, and Kinect data which is available for further investigations on the participants' movement in HHI and HRI. The corpus provides a useful starting point for systematic comparisons and modelling of the human partner's engagement and understanding depending on the conversational partner.

The paper is structured as follows. Section 2 discusses the setup of the data collection and gives basic presentation of the data. Section 3 provides preliminary analyses based on the data so far, with the focus on human gaze-patterns and gesturing. Finally, Section 4 presents conclusions and future research directions.

2. Data Collection

2.1 General Overview

The main goal of the research was to study human-human and human-robot interactions, and consequently this influenced the design and general setup of the data collection. The exercise focused on three general aspects of interaction, known to affect human dialogue behaviour:

1. Dialogue partner
2. Gender of the interlocutor,
3. Dialogue activity, and
4. Language and culture.

The main focus of the study is the comparison related to the type of the dialogue partner: the differences between dialogue behaviours when the partner is a fellow human or a speaking humanoid robot. Since present-day robots lack the capability for fully flexible dialogue interactions, we expect to find important and interesting differences in the subjects' gaze and gesture behaviour depending on the type of the partner, studying failures, mismatching expectations, and misunderstandings concerning the flow of dialogue.

Since Japanese dialogues are known to have different characteristics depending on the gender of the interlocutor, gender was also included as an important feature in the data collection (cf. Maynard, 1997).

The dialogue activity in which the participants are involved has impact on the content of the interaction as well as on the strategies and ways of presenting information to the partner. For instance, dialogues conducted in connection with a particular task (the participants either collaboratively work on a joint effort or exchange information on the conditions and ways to do such a task) have a clear goal (to have the task done), and the dialogue flow closely follows the task structure avoiding subtopics and subconversations, to reach the goal quickly, whereas more chatty type interactions usually aim at maintaining contact and have an associative topic structure which includes many phatic conversational means to provide for a relaxed and entertaining context. We focus on the two opposite types of activities: the participants engage either in a task-based instruction-giving activity where they are expected to produce more structured and professional dialogues, or in a chat-type conversation on music and movies where they are expected to show more spontaneous and relaxed conversational behaviour.

Finally, conversations are constrained by contextual issues ranging from specific dialogue contexts to the language and cultural environment in which the interaction is conducted. We do not intend to conduct intercultural communication studies but will take advantage of the possibility to explore if any differences can be observed related to the different languages available (English and Japanese) and the larger cultural background of the participants in general.

Much consideration was directed to ethical issues in data collection and complying with the regulations and rules related privacy issues, safety, and appropriate conduct. These aspects are discussed more in Jokinen et al. (2019).

2.2 Setup

The data collection follows methodological triangulation, i.e. it involves more than one method to gather data (eye-tracker, motion capture, video, questionnaires). The within-subject design includes two interactive tasks for each subject under different conditions, and the task rotations were assigned in a random order, bearing in mind the goal for a balanced corpus with respect to the above-mentioned dialogue aspects.

Each participant conducted two conversations, one with a human partner (HHI) and one with a humanoid robot (HRI). The conversations dealt with task-based instructions on the best practices for particular care-giving tasks, or a chat-type conversation on music and movies. The subjects participated only in one activity type (instruction or chat), but the topics in the selected activity were different to avoid memory effect from the previous interaction (e.g. if the first dialogue was about transferring a person from one place to another, the second dialogue was about changing clothes, and the other way round). The subject was always the one who received instructions or to whom the robot told a story, whereas the robot (and the other human partner) was always the instructor. One of the experimenters played the role of the human partner and was different from the person who instructed the subject on the course of the experiment.

Prior to the experiments, the participants signed a consent form and they also filled in a pre-experiment questionnaire of their background and expectations of the conversation. After each interaction (HRI and HHI), they filled in a questionnaire on the content of the presented information, as well as a matching 7-point Likert-scale questionnaire of their socio-emotional stance and experience of the interaction. The instructions, questionnaires, and dialogues with the robot/human partner were conducted in Japanese or English depending on the subject's preferred language.

The instructions were the same for both HRI and HHI conditions, modulo human/robot partner. The subjects were told about the goals of the data collection and that they will interact with a human and a robot in a domain dealing with care-giving. To encourage interaction and understanding, the participants were also told that at the end of the session they will be asked questions about the content of the conversation.

Before their interaction with the robot, the participants had a short training dialogue, and the experiment leader told about the robot and its behaviour, e.g. of its movements and that its motors make noise, which was not to be worried about. The experiment leader also emphasised that the participant must speak with a clear and loud voice so that the robot can "hear".

The experiment leader helped to mount the eye-tracker, made sure that it sat securely and safely on the head, and calibrated the eye-tracker. After starting all the recording devices, the experiment leader officially started the recording with a clap, which also had the extra function of enabling synchronisation of the different media afterwards. The participants then conducted their dialogues alone with the robot/human partner, although the experiment leader monitored the session in the next room and could intervene if needed.

In the instruction dialogues, the participant's role was to act as a novice care-giver with the goal to learn some basic care-giving tasks such as how to transfer a patient, while the robot and the human partner acted as an instructor. They could ask questions from the human partner and ask the robot to repeat the given instructions. In the chatty, story-telling task, the participant's role was again a novice care-giver, but the dialogue was to take place during a break so the activity was related to story-telling and chatting with the partner on some light familiar entertaining topic.

In human-human interactions, the interlocutors could conduct dialogues freely, with the experimenter partner usually taking the lead and after about 9-10 mins also winding up the conversation naturally. For the human-robot dialogues, the humanoid Nao robot was installed with a software which allowed the user to conduct spoken English and Japanese natural language dialogues with the robot (no wizard-of-oz dialogues). The topics for the instruction dialogues dealt with certain care-giving tasks (implementation described in Jokinen et al. 2018), while another software was designed especially for the experiment on chat conversations, which enabled the robot to tell stories of some favourite music and films as well as suggest the partner plays a short quiz game *Two truths and a lie.* The original versions of both dialogue systems were designed by the author. Due to privacy considerations the collected dialogues cannot be put in public websites, but demo dialogues similar to the care-giving instruction dialogues are available in English:

https://drive.google.com/open?id=1yq_YtjCwP42xTllCvsc7l46vtWmuys2E

and in Japanese:
https://drive.google.com/file/d/1x8lD9Bba-2WjQee_8MgcADqSNB6NtjKE/view?usp=sharing

The HRI session started when the robot noticed a human face and began to talk. The dialogues were robot-initiated to avoid problems with speech recognition and to allow the participants to have as natural spoken dialogue as possible. The participants could listen to the robot's instructions or play the game as many times as they wanted, and they could finish the interaction any time by saying *thank you – bye*.

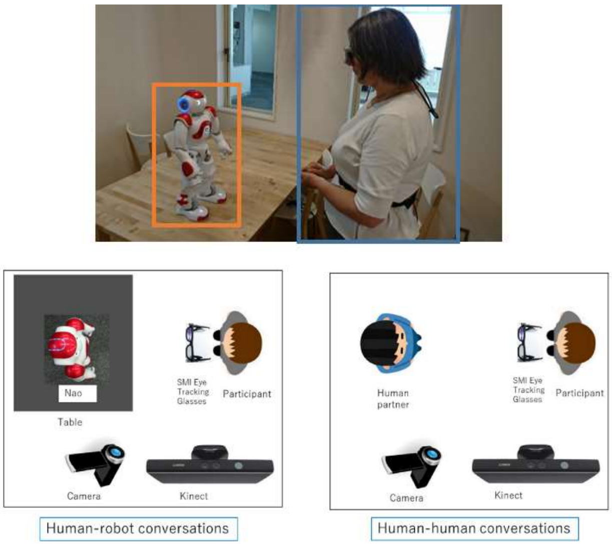

Figure 1: Experimental setup with a participant interacting with the Nao robot. Adapted from Ijuin and Jokinen (2019).

The recording used the SMI Mobile eye-tracking glasses (SMI ETG 2 Wireless 60 Hz) and the humanoid Nao robot (Softbank, formerly Aldebaran Robotics). The dialogue software used Nao Choregraphe and Python programming language, as well as the speech components installed in the Nao robot. A Kinect motion tracker was used to record the partner's movement, and a video camera to record the interactive situation sideways. The setup is shown in Fig 1.

2.3 Data and participants

The data collection included 30 participants (20 Japanese, 10 English), each having both HHI and HRI conversation. The participants were students and researchers aged 20-60, with experience of IT, but no experience of robots. 20 of the participants were male and 10 were female. 14 of the participants had instruction dialogues and 16 had chatty story-telling dialogues. In each group, half had human-human interaction first and the other half human-robot interaction first, to eliminate the effect of the dialogue order. Altogether there are about 13 hours of data, of which about 6 hours are instruction dialogues and 7 hours are chatting. About 9.5 hours are in Japanese and 3.5 hours in English. The data is summarized in Table 1.

Data property	Value
Number of conversations	30 human-human and 30 human-robot
Number of participants	20 male, 10 female
Language	20 Japanese and 10 English
Conversational activity	14 instruction and 16 story-telling
Conversation duration	Approx. 9 minutes
Recording devices	Eye-tracker, Kinect sensor, video camera

Table 1 Summary of the AICO Multimodal Corpus data.

The video corpus consists of videos of the two participants standing sideways, eye-tracker videos of the subject's gaze fixations on the partner (human or robot, see Figure 2), and Kinect recordings for each dialogue. The corpus is automatically analysed with respect to eye-gaze events, and also manually annotated concerning gestures, head movements, and body posture. The Japanese part is also transcribed and annotated with dialogue acts, while the English dialogues are analysed for personality tags.

3. Preliminary Analyses

Recent analyses of the AICO corpus have used user questionnaires, eye-gaze experiments, and gesture studies. The eye-gaze studies concern the user's eye-gaze patterns in dialogue breakdowns, i.e. when the robot's answer is not as expected, and comparison of gaze patterns in HHI and HRI. The gesture studies concern the analysis of co-speech gesturing and on neural models to classify communicative gestures and their role in personality studies. Below we briefly summarize the research activities so far.

3.1 User Questionnaire

After each interaction with a robot and with a human, the participants filled in a questionnaire about their impression of the interactions with the human and with the robot. The 7-point Likert questionnaire focussed on the participants' socio-emotional stance: the participants self-evaluated the interactions by stating their (dis)agreement concerning affective impressions in terms of six adjectives (*enjoyable, impressive, relaxed, natural, interesting, friendly*) and the corresponding negative ones (*awkward, ordinary, tense, unnatural, boring, anxious*), cf. Jokinen (2012). The score for the negative adjectives was expected to be opposite to the positive ones, and thus they also functioned as control statements for the consistency of the subjects' self-evaluation. The order of the statements was random but was kept the same for each participant.

The questionnaire is analysed in Jokinen (2019). In general, there was no statistically significant difference in the users' socio-emotional stance between HHI and HRI conditions. However, the analysis seems to support the view that the subjects are certain and unanimous about their impressions when it comes to human interaction, but less sure and less unanimous about their interaction with a robot partner. Interesting conclusion can also be drawn from the result that the subjects' socio-emotional stance shows a similar tendency in HHI and HRI but differs in the strength of the subjects' confidence ratings.

3.2 Eye-gaze studies

The goal in the eye-gaze studies is to enable humanoid robots to understand human visual attention and thus better tailor their interactions with users. The work follows the pilot study (Jokinen 2018) in which human gaze patterns were studied in interactions with the Nao robot using the WikiTalk application (Jokinen and Wilcock 2014).

3.2.1 Eye-gaze and understanding

The hypothesis examined in Ijuin et al. (2019) is that there is a difference in the speaker's eye-gaze activity depending on their expectations of the communicative situation and monitoring of the partner's understanding, i.e. if the partner is perceived as having understood, misunderstood or not understood the speaker's utterance. A model to estimate the perceived understanding was constructed by measuring the speaker's eye-gaze activity in different dialogue contexts. The study focussed on human-robot interactions only.

Automatic tools were employed to annotate eye-gaze in the eye-tracker videos. We used image processing libraries from OpenCV to detect face and body of the partner and the robot in the videos, with bounding boxes to mark the estimated positions (Figure 2). The Areas of Interest (gaze locations) were automatically calculated from the gaze data within the bounding boxes, and interpreted as follows:

- Face: gaze point is in the face rectangle,
- Body: gaze point is in the body rectangle,
- Other: gaze point is in neither rectangle.

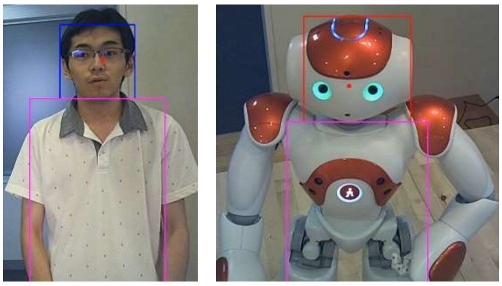

Figure 2 Automatic face recognition for eye-gaze analysis on a human and a robot.

Detection of the participants' utterances was based on pause detection and conducted with the Silence Recognizer provided by Elan (Wittenburg et al. 2006). It separated the subject's and the partner's speech from backchannelling, and the dialogue context was annotated in Elan with utterances coded as Correctly-Understood (CU), Mis-Understood (MU), Not-Understood (NU), and Other (O).

Gazing Ratio was used to measure how the participant uses eye gaze in the human-robot conversations. Gazing Ratio measures the duration of participant's gaze towards the target in a time window, averaged over all gaze durations in the window and the total number of windows. Three types of windows were used: before, during, and after the utterance window, and Gazing Ratio was calculated for each utterance type (CU, MU, and NU).

The results of eye gaze activity, as measured by Gazing Ratio for each utterance type and window type, show that the participants tend to gaze away from the robot after they finish speaking (cf. mutual gaze and breaking of the gaze in human-human conversations (Kendon 1967)) and shift their gaze back to the robot when the robot gives feedback (about 200ms after the turn change). However, if the robot does not start speaking or give feedback to the participants, they keep looking away from the robot longer as if waiting for the robot to take the turn, and only when realising that there is something wrong with the conversation, they shift gaze to the robot again (about 400ms after the turn change). This kind of quantitative difference in the user's eye-gaze behaviour can be useful to predict whether the user is waiting for the robot's feedback or not.

3.2.2 Gaze patterns in HHI and HRI

The analysis in Laohakangvalvit and Jokinen (2019) concerned the participants' focus of attention in interactive situations and used only the English-speaking dialogues. Areas of Interests (AoI) were defined as above (face, body, and other areas), but the analysis used two speech metrics (number of utterances and utterance duration) and two eye-tracking metrics (number of fixations and fixation duration) for each body part and each speaker, averaging from the results of 10 participants.

The participants mostly focussed their visual attention on the Face AoI of both the human and robot partners, but there are interesting differences in the gaze patterns. The number of fixations and the fixation duration were largest when participants were looking at Face AoI in HHI and Body AoI in HRI. Moreover, the participants looked at Other AoI rather than Body AoI in HHI, unlike in HRI. These results indicate the different focussing areas during HHI and HRI conversations. Similar tendencies were also found for the number of first and last fixations, which represent visual focus areas during turn-taking. The reason for the differences may be that in HHI looking at the partner's face is socially pertinent, as the partner's verbal reaction is accompanied by gaze and facial expressions which convey information about one's understanding, interests, timing for turn-taking, etc. In HRI, however, the participants did not look at the robot's face perhaps because they recognized that not much information is conveyed thorough the robot's face or gaze: the NAO robot conducted interaction mostly by means of speech and body gesturing due to limitations in its facial expressions.

The results show that the human tends to have much less interaction with a robot than with a human partner. The data also strongly supports the fact that the dialogue activity has a big impact on the interaction structure and style. In HHI, the participants tended to talk more when engaged in a story-telling activity than in an instruction activity, for the obvious reason that the former encourages topic shifts with no goal or role-related restrictions, while the latter typically has a clear goal which requires one partner to listen while the other partner provides information (cf. Section 2.1).

In HRI, no big differences in the subjects' conversational behaviour were found between the two activity types. Since the robot usually initiated the dialogues, the subjects were constrained to utter short phrases or commands regardless of the activity, and although topic contents of the dialogues were different, the robot's interaction strategy was largely the same. However, it is interesting that the participants demonstrated different gaze behaviours during the two activity types. When listening to the robot in the chatty story-telling activity, they tended to fixate their gaze at both the Face and the Body AoIs, whereas in the instruction-giving activity, gazing at various parts of the robot was not observed. This may be because the body movements and the blinking red chest button attracted the subject's attention besides the face, but also because in story telling situations, the subjects tried to observe whether the robot was actually responding to them or not. In the instruction-giving tasks, the participants needed to focus on listening to the given information rather than intuitively observing whether the robot is responding or not. Different dialogue activities thus induce different conversational behaviours, even if the overt interaction strategy stays the same, and the differences are then realised via non-verbal signals.

Applied to human-robot interactions, the results suggest that robots should integrate gaze models that enable observations of the partner's gaze behaviour and increase the robot's own expressiveness and interaction capability.

3.3 Gesture studies

Gesturing is another important social signal in human-human interactions, and has been widely studied (Kendon 2004). In robotics, gesturing can be divided into movements that the robot performs as part of the task (grasping and moving objects), and movements that a humanoid robot should detect and produce as part of its interaction with a human partner (hand, head, and body gesturing). Research has mainly focussed on smooth motor control and action-related movements in task contexts of industrial robotics, while communicative gestures have been studied in social robotics, in order to gain better understanding of the type of gestures that users would perceive as natural and helpful when interacting with robots. For instance, the WikiTalk application models presentation and beat gestures to provide livelier presentations to the user (Jokinen and Wilcock 2014, Meena et al. 2012). Gesture modelling with deep learning techniques deal with gesture and action recognition in the computer vision field (e.g. Asadi-Aghbolaghi et al. 2018 for an overview), but the studies usually concern single gestures rather than continuous co-speech gesturing with social robots.

Gesture studies on the AICO corpus have dealt with the participants' communicative gesturing in order to explore the function, timing, and detection of gestures and their correlation with speech, eye-gaze and personality. The goal is to build experimental models in order to enhance social robots' natural presentation capability in dialogue contexts.

In Mori et al. (2020) we provide an analysis of the functions and forms of co-speech gestures and continue with detailed models concerning the gesture correlations with gaze and the content of the utterances. The annotations are based on the MUMIN Coding Scheme (Allwood et al., 2004), while the gaze and dialogue content go to the earlier gaze studies.

Personality studies draw on the previous work summarized in Vinciarelli and Mohammadi (2014). In Ijuin and Jokinen (2020) we explore if some traits of people's personalities can be inferred by studying multimodal signals (gesturing, body posture, utterances) from human-human interactions in order to make the robot's behaviour more suited for the person it is interacting with. Personality is linked to emotion and empathy, and the AICO corpus is thus used to explore research questions such as “How does gesturing influence our perception of the other's personality and emotional state in human-human interaction? Is it possible to use this information to adapt the robot's behaviour to the perceived personality of human partners while interacting with them?” We also explore machine learning models to enable the robot to recognize the human interlocutor's affective state and personality during real-time interactions.

4. Conclusions and Future Work

The paper has presented the AICO corpus which is a multimodal corpus of corresponding human-human and human-robot interactions. It is a systematic collection of eye-tracking and video data which takes into consideration different interaction activities and languages, with the aim to compare engagement and attention in human-human and human-robot interaction. The main purpose of the corpus is to be used as training and testing data to bootstrap studies on engagement, awareness, and attention in naturally occurring interactions (i.e. data in the wild), using both qualitative and quantitative research methods as well as neural modelling (e.g. transfer learning and attention networks). The corpus is available for research by contacting the author.

Several preliminary analyses of the AICO corpus have already been conducted and reported in other publications. Currently the corpus is being further analysed with speech and dialogue acts, and by building models for the fusion of gaze and gesture behaviour with spoken utterance analysis, and to coordinate dialogue interactions.

Future research will aim at more detailed analyses on gaze and gesturing to deepen our understanding of the use and correlation between visual attention, action, collaboration and engagement in interactive situations. Moreover, the data can be used for computational modelling of natural and engaging interactions and for explorations concerning neural techniques to design and develop interactive systems. Such models and systems can be applied to a variety of contexts, including everyday tasks and context-aware applications for care-giving and educational domains. Finally, the corpus provides a starting point for discussions concerning ethical, legal, and privacy issues with robot agents. Some important aspects are also discussed in Jokinen et al (2019).

Acknowledgements

A big data collection project would not have been possible without many people who have assisted in the collection and analysis of the data. First of all, thanks go to all the participants who volunteered to take part in the data collection. Besides the raw data, they also provided useful comments and view-points on human-robot interaction in lengthy conversations with the experiment leaders. Special thanks go to Koki Ijuin for his help in the preparation, collection, post-processing, and analysis of the data during his internship and work at AIRC, to Yuki Yoshida, Kazuo Barada, and Naoki Koyama for taking care of the devices, recruiting participants, and preparing for and assisting in the actual data collection, and to Kentaro Watanabe and Takuichi Nishimura for their help in preparing the ethical document. Concerning the analyses of the corpus, thanks are due to Tipporn Laohakangvalvit for discussions and analysis of the gaze patterns, to Marta Romeo for the work on CNN, gestures, and personality studies, and to Taiga Mori for gesture annotation and analysis. The work is based on results obtained from a project commissioned by the New Energy and Industrial Technology Development Organization (NEDO).

Bibliographical References

Allwood et al. (2004). The MUMIN Multimodal Coding Scheme.https://www.researchgate.net/publication/228626291_The_MUMIN_multimodal_coding_scheme

Alameda-Pineda, X., Ricci, E., Sebe, N. (Eds.) (2019). Multimodal Behaviour Analysis in the Wild - Advances and Challenges. Academic Press.

Ijuin, K., Jokinen, K., Kato, T., Yamamoto, S. (2019). Eye-gaze in Social Robot Interactions – Grounding of Information and Eye-gaze Patterns. JSAI 2019.

Ijuin, K., Jokinen, K. (2019). Utterances in Social Robot Interactions - Correlation Analyses between Robot's Fluency and Participant's Impression. Procs of ACM Conference on Human-Agent Interaction (HAI '19).

Ijuin, K., Jokinen, K. (2020). Exploring Gaze Behaviour and Perceived Personality Traits. Proceedings of the International Conference on Human-Computer Interaction (HCII).

Jokinen, K. (2012). Explorations in the Speakers' Interaction Experience and Self-assessments. Procs of the COLING-2012, Mumbai, India.

Jokinen, K. (2018). Conversational Gaze Modelling in First Encounter Robot Dialogues. Proceedings of the International LREC Workshop on Language and Body in Real Life (REAL-MM), Miyazaki, Japan.

Jokinen, K. (2019). Social Emotions for Social Robots - Studies on Affective Impressions in Human-Human and Human-Robot Interactions. In: Social Robotics (ICSR 2019). Lecture Notes in Computer Science, vol 11876. Springer, Cham, pp. 267-277.

Jokinen, K., Furukawa, H, Nishida, M., Yamamoto, S. (2013). Gaze and Turn-taking behaviour in Casual Conversational Interactions. Special Issue on Eye Gaze in Intelligent Human-Machine Interaction, ACM Trans. Interactive Intelligent Systems.

Jokinen, K., Nishimura, S., Watanabe, K., Nishimura, T. (2018). Human-Robot Dialogues for Explaining Activities. The 9th International Workshop on Spoken Dialogue System Technology. Lecture Notes in Electrical Engineering, vol 579. Springer, Singapore.

Jokinen, K. et al. (2019). Privacy and Sensor Information in the Interactive Service Applications for Elder People. Proceedings of the 7th National Conference of Serviceology, Tokyo.

Jokinen, K., Wilcock, G. (2017). Expectations and First Experience with a Social Robot. Procs of ACM Conference on Human-Agent Interaction (HAI '17).

Jokinen, K., Wilcock, G. (2014). Multimodal Open-domain Conversations with the Nao Robot. In: Mariani, J. et al. (eds.) Natural Interaction with Robots, Knowbots and Smartphones - Putting Spoken Dialog Systems into Practice, pages 213–224. Springer, New York.

Kendon, A. (2004) Gesture: Visible Action as Utterance. Cambridge: Cambridge University Press.

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

Laohakangvalvit, T. and Jokinen, K. (2019). Eye-gaze Behaviors between Human-Human and Human-Robot Interactions in Natural Scene. Proceedings of the ECEM conference.

Maynard, S. (1997). Japanese communication. Honolulu: University of Hawaii Press.

Meena, R., Jokinen, K., Wilcock, G. (2012). Integration of gestures and speech in human-robot interaction. Proceedings of 3rd IEEE International Conference on Cognitive Infocommunications (CogInfoCom 2012), Kosice, Slovakia, pp. 673-678.

Mori, T., Jokinen, K., Den, Y. (2020). Analysis of Body Behaviours in Human-Human and Human-Robot Interactions. Procs of the International LREC Workshop *peOple in laNguage, vIsiOn and the mind*.

Siciliano, B, Khatib, O. (2016). Springer Handbook of Robotics. Springer-Verlag Berlin Heidelberg.

Vinciarelli, A., Mohammadi, G. (2014). A survey of personality computing. IEEE Transactions on Affective Computing, vol. 5, no. 03, pp. 273–291.

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., Sloetjes, H. (2006). ELAN: a Professional Framework for Multimodality Research. Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC).