

<https://helda.helsinki.fi>

Conversational Gaze modelling in First Encounter Robot Dialogues

Jokinen, Päivi Kristiina

European Language Resources Association (ELRA)
2018

Jokinen , P K 2018 , Conversational Gaze modelling in First Encounter Robot Dialogues . in
H Koiso & P Paggio (eds) , Proceedings of the Eleventh International Conference on
Language Resources and Evaluation (LREC 2018) . European Language Resources
Association (ELRA) , Paris , LREC workshop on Language and Body in Real Life and
Multimodal Corpora , Miyazaki , Japan , 07/05/2018 . <
http://lrec-conf.org/workshops/lrec2018/W20/summaries/16_W20.html >

<http://hdl.handle.net/10138/299798>

cc_by
publishedVersion

Downloaded from Helda, University of Helsinki institutional repository.

This is an electronic reprint of the original article.

This reprint may differ from the original in pagination and typographic detail.

Please cite the original version.

Conversational Gaze Modelling in First Encounter Robot Dialogues

Kristiina Jokinen

AIRC AIST Waterfront

Aomi, Koto-ku, Tokyo JAPAN

{ [kristiina.jokinen](mailto:kristiina.jokinen@aist.go.jp) }@aist.go.jp

Abstract

Given the popularity of humanoid social robots which can talk with humans and maintain human-like communication patterns, an interesting question is whether the users engage themselves with such systems in a manner similar to human-human communication. If the humanoid robot is perceived as a communicative agent, it can be hypothesized that the user's engagement with the robot resembles social interaction rather than tool manipulation. This paper reports on a pilot study that explores if the hypothesis is supported in the context of a humanoid robot application which reads a digital newspaper interactively for the user. Human eye-gaze patterns are used as an objective measure of the engagement with the robot. The study found support for the hypothesis, but concludes that the interaction is socially less binding than with humans.

Keywords: conversation management, gaze modelling, eye-tracking, first encounter human-robot interaction

1. Introduction

The popularity of humanoid social robots which can talk with humans and appear having human-like communication patterns, has brought in interesting questions about whether the users engage themselves with such robots in a similar manner as they do when communicating with human partners. Although natural language interactions between humans and intelligent agents are often problematic due to limited communicative capabilities of the system, they nevertheless give rise to expectations that the system functions more like a communicating agent than a voice-controlled tool. This may be due to the people's tendency to anthropomorphize computers and other media, i.e. treat them as if they were real people (Reeves and Nass 1996). However, when interactions are conducted with a humanoid robot which does not only speak, but also acts in a human like manner (i.e. moves and gestures), such expectations are reinforced and easily lead the robot to be perceived as an intelligent agent with near-human communicative competence. Consequently, if the users perceive a humanoid robot as a communicating agent, it can be hypothesised that their behaviour towards the robot resembles social interaction with other humans, rather than tool manipulation.

One of the fundamental characteristics of human-human conversations is that the interlocutors look at their partners' face (not necessarily straight into the eyes but in the facial area), i.e. eye-gaze is an important means for joint control and coordination of the interaction. According to Gullberg & Holmqvist (1999), gaze is fixated on the partner's face about 90% of the interaction time, and a similar fixation pattern is carried over to conversations conducted through videoconference technology, although gaze tends to wander around the screen and the overall environment especially if the partner's video is not life-size (Gullberg and Holmquist, 2006). The gaze of another person is a strong cue of where to focus one's visual attention (Friesen and Kingstone, 1998), and in developmental psychology, gaze-following and visual joint attention are regarded as social phenomena learnt through interaction with the others, and children learn them early, at the age of about 1 year (Meltzoff and Brooks 2007).

As the fundamental function of eye-gaze is related to monitoring the partner's gaze direction so as to establish joint attention, it can be assumed that also in human-robot

interactions, visual attention plays an important role. Even if the robot cannot reciprocate the gaze, the users may apply visual attention to their robot partners in a similar way as they do with their human partners, i.e. their gaze patterns follow social expectations found in human interactions.

In this paper we set out to study if the hypothesis that the robot is perceived as a communicative agent rather than an interface tool can be supported by the user's eye-gaze behaviour. We report of a small experimental study that explored if the hypothesis is supported in the context of a humanoid robot application which reads a digital newspaper interactively for the user. Using eye-tracker technology, human eye-gaze patterns are detected and used as an objective measure for the user's engagement with the robot. The study found support for the hypothesis, but also concludes that the interaction is socially less binding with the robot agent than with humans. Due to the case study nature of the experiment, the results will be investigated later with a large set of participants.

The paper is structured as follows. Section 2 provides a brief overview of the background and related research in interaction studies and eye gaze as a social signal. Section 3 presents the experimental setup, and Section 4 describes the results. Section 5 concludes with future views.

2. Gaze as social signal

The role of eye-gaze as a means of social signalling has long been established, see Kendon (1967), Argyle & Cook (1976), Goodwin (1980). Its fundamental function is related to visual attention, and in communicative situations this means monitoring the partner's gaze so as to establish joint attention and enable construction of shared context and mutual understanding. Land (2006) points out that gaze is also proactive in nature since it anticipates actions: we often gather visual information from our surroundings before performing motor actions.

Conversational feedback can be effectively mediated by gaze behaviour. The partner's willingness to continue interaction can be inferred from their looking at or looking away from the partner, and in general, direct and averted gaze can signal the speaker's interest to approach or to avoid the object of attention (Mutlu et al. 2012).

Also turn-taking is coordinated by gaze: a quick shared gazing at each other, mutual gaze, is used to agree on the change of the speaker (Kendon, 1967; Brennan et al. 2008;

Jokinen et al., 2010, 2012, Mutlu et al. 2006). Levitski et al. (2012) observed different gaze patterns within a one second window at the beginning and at end of the utterance and noticed that in the beginning of the utterance, mutual gaze is quickly broken by the speaker, whereas at the end of the utterance, the speaker's gaze fixates onto the partner quite a long time before their speaking ends. As the speaker needs to focus their attention to the next speaker to facilitate smooth turn-taking, the interlocutors also fixate their eyes more often and longer in the beginning than at end of one's utterance, whereas in the middle of their speaking, the gaze wanders off since the speaker focusses on producing their own utterance.

See Jokinen (2014) for a longer description of eye-tracker and gaze research, and Ruhland et al. (2015) and Broz et al. (2015) for overviews of the work on eye gaze and human-robot interaction.

3. Experimental setup

The main hypotheses that the experiment focus on, are:

1. Majority of human eye gaze focus on the robot's head.
2. Gaze focus in the beginning of the interaction differs from the gaze focus at the end of the interaction.
3. There is more focus on the robot's face in the beginning than at the end of the interaction (the user becomes more familiar with the robot).
4. There is not much focus on the robot's gesturing.

The study used two female participants who were between 20-40 years of age and worked as researchers at the university. Neither of them had prior contact with robots and they also had neutral expectations of the interaction with the robot. Both participants had normal vision.

Eye gaze was measured using SMI Mobile eye-tracking glasses (SMI ETG 2 Wireless 60 Hz), and the data created using a Lenovo X230 laptop with Intel® Core TM i7-3520M CPU 2.90 GHz. Statistics were calculated using IBM SPSS Statistics 22.0.



Figure 1. The NAO robot in the centre of the study setting.

The robot was the humanoid NAO robot developed by Softbank/Aldebaran Robotics (Figure 1). The robot was installed with MoroTalk, a newspaper reading application for the national newspaper (Jokinen & Wilcock, 2013). This had been developed in collaboration with the Koti ('Home') project aiming to improve well-being and health care in future smart homes. The application is based on the WikiTalk technology (Wilcock, 2012) which allows interactive access to digital repositories. The robot supports open-domain conversations, and the user can shift to related topics or switch to a totally new topic by spelling the first few letters.

The setting was a brightly lit classroom and the robot stood on a table facing the participants so that the robot and the human were at similar eye level. The participants were first asked to fill in a short demographic form and survey on their previous experience with robots. Then they were fitted with the eye tracking glasses which were calibrated during a three-point calibrating session until accurate. The participants were instructed how to interact with the robot and a short description of the robot's abilities was given. The participants were told that the robot will read them news from today's newspaper and that they could select interesting news for the robot to read. The session started when the robot began its introduction speech, after which the participant began commanding the robot.

The experiment leader controlled the beginning and the end of the session, and, intervened if necessary, e.g. if the robot shut down. All interactions were videotaped using the eye tracker and two extra video cameras. The interactions took 10-15 minutes and afterwards the participants filled in a short feedback form about their experience.

The data from the eye gaze videotapes were annotated using the Elan Linguistic Annotator version 4.1.0 (Wittenburg et al. 2006). The eye movement data were coded for the duration and the location of the fixations. Five different categories for the targets of eye gaze were used:

1. gaze focused on the robot's head;
2. gaze focused on the robot's hand;
3. gaze focused on another part of the robot;
4. gaze focused on the study conductor;
5. gaze focused on background.

Annotations were done on the first and the last three minutes of each of the two robot human interaction sequences (altogether 12 minutes), so as to be able to compare the participants' gaze behaviour at the beginning and at the end of the interaction.

The data sets were annotated by two annotators, who were blind to each other's annotations. A two-minute section in the beginning of one of the videotapes was annotated by both annotators to determine consistency among the annotators. The interrater reliability was found to be Kappa = 0.42 ($p < .001$), 95% CI (0.249, 0.597). According to the scale proposed by Rietveld and van Hout (1993), these values indicate fair agreement between the two scorers.

For the analyses, three measures were of interest:

1. the amount of changes in gaze focus;
2. the length of individual fixations on the five eye gaze targets coded for;
3. the accumulated fixations time on the five gaze targets.

The frequencies and time durations were then compared between the first two minutes and the last two minutes of the annotated interaction sequences. The difference in the changes of gaze fixations between the end and the beginning of the interaction sequences was assessed using the Chi-Square test. Due to uneven sizes of cases categorized into the five different coding categories, the assumption of homoscedasticity for analyses of variance was not met. However, to assess whether the mean duration of the human participants' gaze focus differed between the beginning and the end of the interactions one paired t-test was conducted.

4. Results and discussion

Changes in the human participants' gaze focus between the different parts of the robot and the background is summarized in Table 1 and illustrated in Figure 2. The data on the two naïve users' eye tracking patterns suggests that there were differences between the beginning and the end of the interaction period, and overall, the three original hypotheses are corroborated by the data. First, most of the fixations in the beginning and in the end of the interactions were on the robot's face. Moreover, there were more changes in gaze focus in the beginning than at the end of the human robot interactions for both of the naïve participants. As illustrated in Figure 2, in the beginning of the interactions there were more fixations on the robots head, and other parts of the robot. In contrast, at the end of the interactions there were fewer fixations on the robot's head and body, and more fixations on the background for one of the participants. There were no differences between the beginning and the end of the interactions regarding fixation counts on the hands of the robot. Chi-Square tests indicated a statistically significant association between the target of gaze focus changes and the time during the interaction ($\chi^2(4, N=205) = 15.378, p = .004$). That is to say, the five different gaze targets were focused on differently between the beginning and end of the interactions. Cramer's V test of the strength of association indicated a medium effect size ($\phi_{\text{Cramer}} = .274$).

Gaze focus	Beginning of interaction		End of interaction	
	Count	Percentage	Count	Percentage
Participant 1				
Robot's head	29	42.6	14	41.2
Robot's hands	3	4.4	3	8.8
Other part of robot	23	33.8	10	29.4
Background	11	16.2	7	20.6
Study conductor	2	2.9	0	0
Participant 2				
Robot's head	29	43.3	18	50
Robot's hands	10	14.9	10	27.8
Other part of robot	25	37.3	0	0
Background	3	4.5	8	22.2
Study conductor	0	0	0	0
Total	135	100	70	100

Table 1. Counts of changes in gaze focus in the beginning and end of the interactions; visualisation in Figure 1.

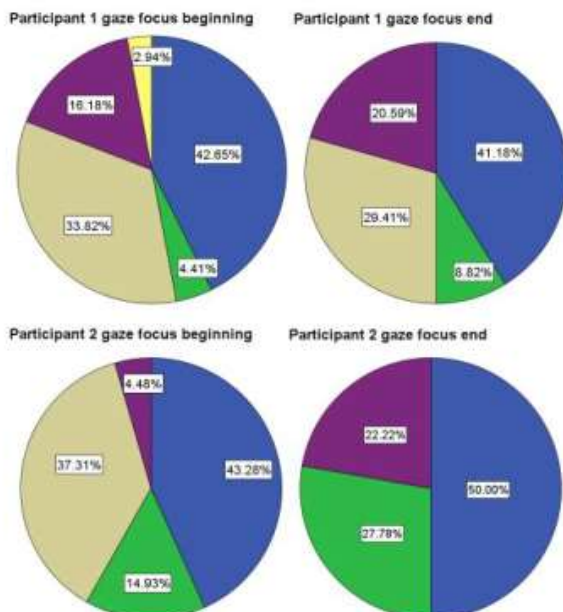


Figure 2 Gaze focus changes for the two human participants. Blue= head, green = hand, grey = other part, violet = background, yellow = study conductor.

Gaze focus	Beginning of interaction			End of interaction		
	Mean	Std. Deviation	Sum	Mean	Std. Deviation	Sum
Participant 1						
Robot's head	2.88	4.31	83.60	2.60	2.86	36.45
Robot's hands	0.63	0.58	1.90	0.26	0.13	0.77
Other part of robot	0.74	0.72	16.95	0.51	0.44	5.05
Background	1.49	2.07	16.38	1.15	1.15	8.04
Study conductor	0.89	0.27	1.78			0
Participant 2						
Robot's head	3.42	3.73	99.15	6.18	7.81	111.21
Robot's hands	0.73	0.67	7.26	0.20	0.14	2.00
Other part of robot	0.89	1.07	22.17			0
Background	0.26	0.15	0.79	2.25	5.43	18.02
Study conductor			0			0

Table 2 Average length of gaze fixations in the beginning and end of the interaction sequences for the two human participants.

Descriptive statistics of the eye gaze durations are shown in Table 2. The mean durations of the eye gazes on the five different targets are roughly similar between the end and the beginning of the interactions. An exception is that for participant 2 the fixations on the robots head are on average longer in the end compared with the beginning of the interaction (means 3.42 and 6.18 respectively) like the average fixations on the background (means .26 and 2.25 respectively). For most categories the mean gaze durations are inconsistent between the participants (one participant has longer durations in the beginning when the other has longer durations in the end) or there are too few cases for comparison (e.g. only participant 1 has fixations on the study director) However, the mean duration of fixations on the robot's hands is longer in the beginning of the interactions. A paired t-test found that this difference in mean gaze duration on the robot's hands was significant ($t(12)=2.811, p=.016$).

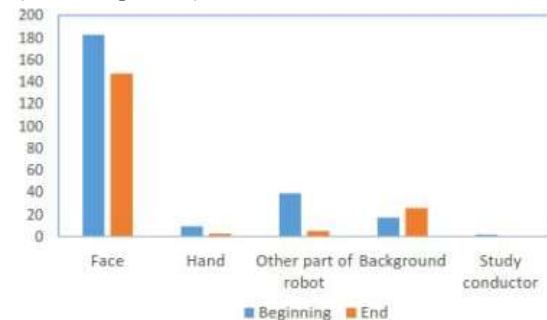


Figure 3 Totals of gaze durations between the end and beginning of interactions.

Figure 3 shows the accumulated sums of the length of fixations on the five different gaze targets between the beginning and end of the interactions. As can be seen from Figure 3, overall the longest duration of time was spent on focusing on the robot's face. When comparing the beginning and end of the interactions, the participants spent less time on focusing on the robot and more time focusing on the background in the end of the interaction.

5. Conclusions and future work

This experimental study provided support for the hypothesis that human-robot interaction is social interaction as opposed to interaction with a tool, and the results were in line with previous study results that have found human robot interaction to resemble that of human to human interactions (Jokinen et al, 2012; Yonezawa et al., 2007; Yu et al., 2012). Overall, the counts of eye gaze fixations and the duration on fixations was largest for the robot's head at all times during the interaction and for both the study participants, which supported our hypothesis (1)

in Section 3. Furthermore, the pilot study provided support for the hypothesis (2) that adapting to the robot changes human's gaze fixations. There were more changes in gaze focus in the beginning than at the end of the human robot interactions. Chi-square analysis indicated that the targets of the gaze fixations differed significantly between the end and the beginning of the interactions. Finally, the pilot study found support for the hypothesis (3) that when the robot becomes more familiar to the human, there is less focus on the robots head and gaze starts to wonder elsewhere. There were more fixations on the robots head, and other parts of the robot in the beginning than during the end of the interactions. The total length of the fixations on the robot's head and other parts of the robot's body was also longer in the beginning of the interactions. In contrast, there were more fixations on the background and the total length of fixations on the background was longer in the end as compared to the beginning of the interactions.

An interesting finding is that while the counts of fixations on the hands of the robot were the same in the beginning and the end of the interactions, the duration of these hand fixations were longer in the beginning than in the end. This can mean that in the beginning of the interaction the user found the robot's gesture behaviour novel and focused attention longer on the gestures to gather more information about them, whereas at the end of the interaction the user had already got familiar with the robot's gesturing and did not need to spend so much time on them.

The study is an experimental study with small sample size of naïve participants, which makes the results less generalizable. More data with more participants and more interactions with the robot are needed to fully investigate how people learn to interact with humanoid robots in the future. However, the strength of the study was the novel topic of study and its explorative nature. The results give a good indication of how people new to humanoid robots may react to them. Based on averaged and subjective estimations, it seems that our initial hypothesis regarding the fixation points and durations was partially correct, although not sufficiently accurate.

Regarding the practical setup, care should be paid to the eye-tracking glasses. They were not always held correctly during the experiments, but slipped down the bridge of the participants' noses, in particular when the user laughed. In this experiment it did not seem to present a major issue but should be taken into attention in future studies.

The purpose of the study was to investigate how humans without any previous experience of humanoid robots begin to interact and adapt their gaze behaviour when they first meet and interact with a humanoid robot. The results suggest that humanoid robot interaction is social, but it is not as encaptivating and smooth as interaction between humans. Naïve participants instantly focused mainly on the robot's head and perhaps learned to ignore hand gestures as the interaction progressed (although hand gestures were designed to support presentation and rhythm of the robot's utterance). After the novelty of the beginning was worn out there were less changes in gaze fixations.

While the Nao robot is a cute humanoid robot, its facial expressions are limited to flash lights. According to Media Equation Hypothesis this does not prevent the user to bond

with the robot and interact in a natural manner since people's interactions with computers and new media are "fundamentally social like interactions in real life" (Reeves and Nash 1996). On the other hand, the robot's human-like appearance is known to have impact on the interaction and the participants' social behaviour (e.g. industrial robots are not designed to arouse affection or social effects, so social gaze in industrial robots does not create affective and emotional effects (yet supports floor management and makes the users feel more responsible for the task, Fisher et al. (2013)). It seems obvious that human-like appearance as such does not guarantee agenthood, since this is a complex phenomenon and requires the robot to exhibit human-like behaviour as well, i.e. the robot's appearance needs to conform to the robot's level of social competence. The view of an automated system as an intelligent agent can be related to *affordance*, the concept originally discussed by Gibson (1979), applied to HCI by Norman (1988), to robotic control by Chemero and Turvey (2007) and suggested by Jokinen (2010) to account for the flexible use of natural language dialogue systems: the system's communicative competence affords natural language interaction and lends itself to the intuitive use of the system where the system is communicative agent, not just a tool.

However, in the case of Nao, participants commonly perceive it likable, intelligent and safe, and the gaze fixations onto its face may thus indicate the human partners' initial attraction and benevolence towards the robot and its face in general, rather than "agenthood". To replicate the experiment using a robot with a more human-like, expressive head and compare the results along the robot's perceived agenthood and appearance will be an interesting future study: we may be able to infer how the user's engagement in interaction, as measured by eye-gaze behaviour, is related to the humanoid's appearance and communication skills. This task also has implications to the famous Uncanny Valley hypothesis (Mori 1977), according to which the artefact's increasing human-likeness will, at some point close to the real resemblance, cause the user's acceptance of the artefact suddenly drop. Moore (2015) explains the Uncanny Valley effect on the basis of category boundaries and the uncomfortable feeling that humans experience when typical or normal boundaries are crossed. A humanoid robot may cause uncomfortableness as it is not a typical member of either the classes "human" or "robot", and its accommodation into the existing world requires that a new category is created. The uncomfortable feeling can be overcome by more regular encounters with the untypical object, and thus autonomous and communicating robots can become more acceptable as the audience have more interactions with them, and as their social communication capability increases.

6. Acknowledgements

The main research was carried out when the author was affiliated with University of Helsinki. Thanks go to the students and colleagues for useful discussions and taking part in the experiments, and to Trung Ngo Trong and to Graham Wilcock for their contributions to the work. The work was supported by the Academy of Finland project Digital Citizens. The author would also like to thank the Japanese Nedo project for support in the final editing phase of the paper.

7. Bibliographical References

- Argyle, M. and Cook, M. (1976). *Gaze and Mutual Gaze*. Cambridge University Press, Cambridge
- Brennan S., Chen X., Dickinson C., Neider M., and Zelinsky G. (2008). Coordinating cognition: The costs and benefits of shared gaze during collaborative search. *Cognition*, 1465-1477.
- Broz, F., Lehmann, H., Mutlu, B., and Nakano, Y. (Eds.) (2015). *Gaze in Human-Robot Communication*. John Benjamins Publishing Company.
- Chemero, A., Turvey, M.T. (2007). Gibsonian Affordances for Roboticians. *Adaptive Behaviour* 15(4): 473-480.
- Fischer K., Lohan K.S., Nehaniv C., & Lehmann H. (2013). Effects of Different Kinds of Robot Feedback. In: Herrmann G., Pearson M.J., Lenz A., Bremner P., Spiers A., Leonards U. (eds) *Social Robotics. ICSR 2013*. Lecture Notes in Computer Science, vol 8239. Springer, Cham.
- Friesen, C.K. and Kingstone, A. (1998). The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. *Psychonomic Bulletin & Review*. 5(3): 490–495.
- Gibson, J.J. (1979). *The Ecological Approach to Visual Perception*. Houghton Mifflin.
- Goodwin, C. (1981). *Conversational Organization. Interaction between Speakers and Hearers*. New York, London
- Gullberg, M. and Holmqvist, K. (2006). What speakers do and what addressees look at. Visual attention to gestures in human interaction live and on video. *Pragmatics and Cognition*, 14(1), 53-82.
- Gullberg, M. and Holmqvist, K. (1999). Keeping an eye on gestures: Visual perception of gestures in face-to-face communication. *Pragmatics and Cognition*, 7(1), 35-63.
- Jokinen, K. (2014). Eye-trackers and Multimodal Communication. In: Paggio, P. and Wessel-Tolvig, B. (eds.) *Proceedings from the 1st European Symposium on Multimodal Communication*, University of Malta, Valletta, October 17–18, 2013. Linköping Electronic Conference Proceedings 101. Linköpings universitet.
- Jokinen, K. 2010. Rational communication and affordable natural language interaction for ambient environments. In: *Spoken Dialogue Systems for Ambient Environments: Second International Workshop on Spoken Dialogue Systems Technology (IWSDS 2010)*, pp.163-168, Springer (2010)
- Jokinen, K. (2009). *Constructive dialogue management – Speech interaction and rational agents*. John Wiley & Sons.
- Jokinen, K., Furukawa, H, Nishida, M., Yamamoto, S. (2012). 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., Harada, K., Nishida, M., Yamamoto, S. (2010). Turn alignment using eye-gaze and speech in spoken interaction. *Proceedings of Interspeech 2010*. Makuhari, Japan.
- Jokinen, K. and Wilcock, G. (2013). Multimodal Open-domain Conversations with the Nao Robot. In: Mariani, J., Devillers, L., Garnier-Rizet, M., Rosset, S. (eds.) *Natural Interaction with Robots, Knowbots, and Smartphones – Putting Spoken Dialog Systems into Practice*. Springer Science+Business Media
- Kendon, A. (1967). Some functions of gaze direction in social interaction. *Acta Psychologica*, 26, 22–63.
- Land, M.F. (2006). Eye movements and the control of action in everyday life. *Progress in Retinal and Eye Research* 25, pp. 296-324.
- Meltzoff, A.N. and Brooks, R. (2007). Eyes wide shut: The importance of eyes in infant gaze following and understanding other minds. In R. Flom, K. Lee, & D. Muir (Eds.), *Gaze following: Its development and significance* (pp. 217-241). Mahwah, NJ: Erlbaum.
- Moore, R. (2014). A Bayesian explanation of the ‘Uncanny Valley’ effect and related psychological phenomena. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3499759/>
- Mori M. (1970). The Uncanny Valley. *Energy*, 7(4), 33–35. English translation by Karl F. MacDorman and Takashi Minato.
- Mutlu B., Hodgins J. and Forlizzi J. (2006). A Storytelling Robot: Modeling and Evaluation of Human-like Gaze Behavior. *Proceedings of the 6th IEEE-RAS International Conference on Humanoid Robots HUMANOIDS’06*.
- Mutlu, B., Kanda, T., Forlizzi, J., Hodgins, J., Ishiguro, H. (2012). Conversational gaze mechanisms for humanlike robots. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, Vol.1, Issue 2.
- Norman, D.A. (1988). *The Psychology of Everyday Things*. Basic Books: New York.
- Reeves, B. and Nass, C. (1996). *The Media Equation: How People Treat Computers, Television, and New Media like Real People and Places*, Cambridge University Press.
- Ruhland, K., Peters, C., Andrist, S., Badler, J., Badler, N., Gleicher, M., Mutlu, B., McDonnell, R. (2015). A review of eye gaze in virtual agents, social robotics and HCI: Behaviour generation, user interaction and perception. *Comp Graphics Forum* 34(6): 299-326.
- Rietveld, T. and Van Hout, R. (1993). *Statistical Techniques for the Study of Language and Language Behaviour*. Berlin: Walter de Gruyter.
- Stratou, G. and Morency, LP. (2017). MultiSense—Context-Aware Nonverbal Behavior Analysis Framework: A Psychological Distress Use Case. In: *IEEE Transaction on Affective Computing*, 8(2).
- Wilcock, G. (2012). WikiTalk: A spoken Wikipedia-based open-domain knowledge access system. In *Proceedings of the COLING-2012 Workshop on Question Answering in Complex Domains*. Mumbai, India, 57-69. 2012.
- Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., Sloetjes, H. (2006). ELAN: a Professional Framework for Multimodality Research. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC 2006)*, 1556-1559.
- Yonezawa, T., Yamazoe, H., Utsumi, A., and Abe, S. (2007). Gaze-communicative behavior of stuffed-toy robot with joint attention and eye contact based on ambient gaze-tracking. *Proceedings of the 9th ICMI*, pp. 140-145.
- Yu, C., Schermerhorn, P., Scheutz, M. (2012). Adaptive eye gaze patterns in interactions with human and artificial agents. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 1(2), 13.