Naturalistic Conversational Gaze Control for Humanoid Robots - A First Step

Hagen Lehmann[1], Ingo Keller[1], Reza Ahmadzadeh[2], and Frank Broz[1]

[1] Heriot-Watt University, Edinburgh, EH14 4AS, Scotland
hagen.lehmann@gmail.com, ijk1@hw.ac.uk, f.broz@hw.ac.uk
[2] Georgia Institute of Technology, Atlanta, USA
reza.ahmadzadeh@gatech.edu

Abstract. In this paper we present a novel approach to conversational gaze control for robots with physical eyes. We will introduce our gaze model, locate it in current relevant research in social robots, and present the results of a set of exploratory experiments. The work illustrated here is a first step of a larger research program dedicated to the development of a reactive mutual gaze controller (MGC) for human-robot interaction (HRI). The study presented in this paper has pilot character and was designed to inform more complex setups in the future. Despite the limitations of the current version of our MGC, our results are in line with the findings of previous studies and illustrate the importance of interpretable eye movements for HRI.

Keywords: Eye Gaze; Conversational Gaze Control; Humanoid robots; Social robotics; Human-Robot Interaction

1 Introduction

The ability to interpret and to generate understandable head- and eye-gaze cues is vital for an individual to be successfully integrated in human society, and is a key factor in primate social evolution. It stands to reason that the special role gaze had in shaping our sociality makes it as well an important features for human-robot co-evolution. One of the defining characteristics of humans is our predisposition to live in differentiated social groups and form complex societies. Gaze as an easily accessible and partly subconscious form of communication plays a central role in human-human nonverbal communication. Because of its distinct role in social evolution, humans are extremely sensitive to unnatural gaze patterns. They quickly induce a feeling of uneasiness and untrustworthiness, and individuals exhibiting such patterns are at best treated with suspicion and precaution, at worst they are avoided.

Contemporary social robots have a wide variety of physical embodiments. One common factor however is, that they usually have a face with different features. This is to give their potential users a point of reference for how and where to address the robot during an interaction. Even though endowing robots with facial features certainly helps to increase the intuitiveness of the potential social interaction for the human user, it also generates very specific expectations of how these features should be used by the robot. This means if a robot is constructed with physical eyes, it also needs to be equipped with the according behaviors for these eyes. These behaviors don't have to be exactly human-like, but they have to be a least inspired by human eye-gaze behavior in order to be interpretable and ideally to make the robot appear reliable and trustworthy.

The aim of this paper is to give an insight into the ongoing development process of a conversational gaze controller for humanoid robots with physical eyes. In order to develop this gaze controller we collected a large set of gaze tracker data during dyadic human conversations [8] and analyzed it with respect to the probabilities of gaze shift location and timing. The data from this analysis was used to build the gaze model we introduce in this paper. We also describe a pilot study in which we tested the first iteration of our gaze controller. At the end of the paper we will discuss the issues we found and give an outlook of our next steps.

2 Background

In their very comprehensive overview article Admoni & Scassellati [1] define and illustrate three different gaze research categories in HRI. According to their definition our research approach is located between design-focused and technology-focused. We are specifically interested in face-to-face conversational gaze, which largely consists of instances of mutual gaze.

Mutual gaze can be defined as a continuously ongoing process between two interlocutors jointly regulating their eye contact [3]. It is at the basis of complex task-oriented gaze behaviors such as visual joint attention [10], is a component of turn-taking between infants and caregivers, is vital for language learning [19] and is known to play a role in regulating conversational turn-taking in adults [12]. This illustrates the importance of mutual gaze for human social development and during human face-to-face interactions. It explains also, to a certain degree, why humans are very sensitive to inappropriate gaze behaviors, specifically in the context of dyadic conversations.

In HRI the development and implementation of naturalistic gaze behaviors has been a central topic for almost two decades, but until now, due to technical limitations, humanoid robots with moving physical human-like eyes have been limited to a few examples (e.g. [13, 7, 16]). Additional to the increase in technological complexity moving physical eyes imply, and related reliability issues, the conversational reactiveness of robots has been an issue in the past. This was largely due to limitations in speech recognition and speech generation. But lately developments in this area have brought the generation of true reactive conversational settings with robots into the reach of, at least, laboratory research.

3 Method

3.1 Ethics statement

The research was approved by the Heriot-Watt University's MACS department's ethics board for studies involving human participants. The participants provided their informed consent before being included in the experiments.

3.2 Experimental Setup

We conducted a study with 20 participants and three different conditions. Each participant saw all three conditions. The order of the conditions was randomized between the participants. The research was done at the Robotics Lab at Heriot-Watt University. After each condition the participants were asked to complete a questionnaire including three subscales of the Godspeed Questionnaire [6] and the Inclusion of Other in Self scale [4]. For each participant there were on average two days between the experimental trials. Our participants were recruited in the School of Mathematics and Computer Science at Heriot-Watt University and included students, and academic and administrative staff.

Experimental Conditions During each of the conditions the participants were seated in front of the robot with a table between them, that separated the face of the participant and the face of the robot by approximately 60 cm. The participants were wearing *Tobii Pro2* eyetracking glasses. The data from the gaze tracker were collected and will be used for the further analysis. In this paper we present only the subjective measures evaluated with the above mentioned questionnaires.

During each condition the iCub told a short (1 minute) story to the participant. The story was the same for all conditions. The only instruction given to the participants was to listen to the robot's story. Besides its gaze movements, the robot behaved exactly the same during all conditions. The participants were not given any information about the purpose of the study or the reasons behind the robot's behavior.

- **Condition 1 (Model)**: In the first experimental condition the robots gaze behavior was controlled by our gaze model.
- **Condition 2 (Artificial)**: In the second experimental condition the robot exhibited eye gaze that was controlled by a model with state change probabilities dissimilar from the human data. With this condition we wanted to test whether the fact that the robot moved its eyes, even though it appeared to gaze randomly, would have a positive effect compared to the control condition. Previous research has shown that animated artificial agents are in general anthropomorphized and ascribed social roles and behaviors [11, 14].
- **Condition 3 (Control)**: In the control condition the robot was not moving its eyes and was staring at the participant during the entire experiment. This condition controlled for the overall effect of robot eye movement.

Model The model is based on a dataset of gaze tracking data collected from 32 pairs of people engaged in face-to-face conversation (for details, see [8]). A coarse discretization of the image from the head-mounted gaze tracker was used to classify the gaze location of the participants at each timestep. The size and location of the conversation partner's eyes and mouth in the image were used to compute bounding boxes that divided the face into eye and mouth regions. When the gaze point falls outside the face, it is classified into one of four possible away directions depending on which quadrant of the image it falls in (see Figure 1). A Markov model was created with state transition probabilities from counts of the transitions present in the dataset. For the random condition, a model with the same set of states was used. Each state transition weight was set to $w_{max} + w_{min} - w_i$ ($p_i = \frac{w_i}{\sum w_i}$ is the probability in the original model), reversing the relative likelihoods of each transition. This was done to produce a gaze pattern that was dissimilar to the observed patterns of human gaze behavior.

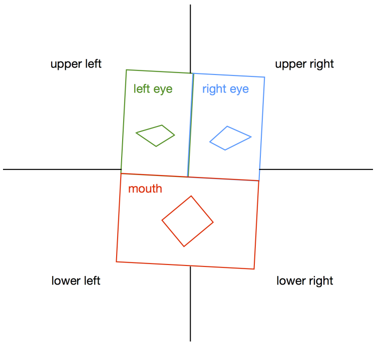

Fig. 1. The image regions corresponding to the model states.

Software Architecture The gaze controller is implemented as a collection of YARP modules. The image from the robot's eye is processed using the CLM library to locate a person's eyes and mouth locations [5]. The robot sends either the facial feature's center location or a face-relative away gaze point to the iKinGazeCtrl module [18], which computes the correct eye movement for the robot to look at the given image coordinate. Additional modules process text-to-speech for the robot and accompanying lip animations. See Figure 2 for a diagram.

The parameters for iKinGazeCtrl were set so that only the eyes moved (there was no neck movement). For the away gaze states, the robot looked 0.1 m to the side and 0.08 m up or down from the mouth. These distances were chosen to prevent the robot from losing the human's face in the camera image while looking away. The system runs at approximately 30 frames per second, the frame rate of the eye camera. However, the random and model based controllers were both slowed down by a factor of 4 (a new gaze state is generated every 4th incoming video frame) to allow large gaze shifts time to complete.

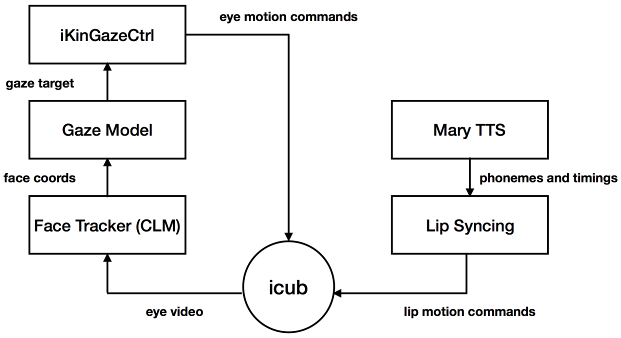

Fig. 2. The modules that make up the control software for the experiment.

Robotic Platform For the testing of our conversational gaze controller we used an iCub robot with an iCub Talking Head. The iCub has eyes that include human features like a black pupil, white sclera, and moveable upper and lower eyelids. These features let the robot's gaze behavior appear very clear and pronounced.

The iCub Talking Head has 4 DOF for the movements of the lips (up, down, right, left) (see Figure 3). The position of the motors, resembling a so called viseme, had to be mapped to given phonemes in order to achieve a realistic lip synchronisation [9]. As there is no common standard for these mappings established yet, the phoneme-to-visemes mapping was adapted from Annosoft's Lipsync tool [2], a software widely used in 3D character animation. Visemes were adjusted to the limited resolution and variability of the lip movements. MaryTTS [15] was chosen as speech synthesizer due its capability of generating phoneme duration information.

3.3 Research Questions and Hypotheses

We evaluated the impression the robot induced in our participants in order to answer two research questions:

- Research Question 1 (RQ1): Are eye movements, between different facial features of the interlocutor, by an anthropomorphic robot with physical eyes sufficient to positively influence participants perception of the robot?
- Research Question 2 (RQ2): If RQ1 can be positively answered, what role does the naturalness, concerning frequency and direction, of the gaze shifts play?

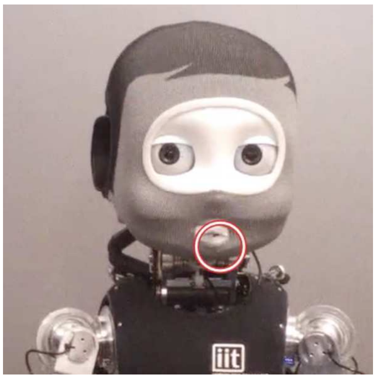

Fig. 3. icub robot with gaze point overlay from gaze tracker.

For RQ1, we expected that human-like movements such as eye gaze shifts between different facial features would induce an emotional reaction in the user. This hypothesis is based on the hypothesis that human-like movements are important for robotic agent with a human-like appearance [17].

For RQ2, we expected that the robot would be perceived more anthropomorphic, intelligent, and likeable, when exhibiting movements that are based on our model compared to random movements.

3.4 Measures

The evaluation of the impression the three different conditions induced in our participants is based on two standardized, validated measures. We used three subscales of the Godspeed Questionnaire, an instrument constructed and validated by Bartneck et al. [6]. The second evaluation tool we used is the Inclusion of Other in Self (IOS) scale, which can be found in Aron et al. [4].

The Godspeed Questionnaire The Godspeed questionnaire is a HRI-specific measure of participant perceptions across several dimensions, where each dimension is addressed using a set of semantic differential scale [6]. We chose 3 dimensions from this questionnaire that are most relevant to the presented experiments - anthropomorphism, likeability and perceived intelligence.

Inclusion of Other in Self Scale The IOS Scale in this study was used, based on Aron et al. [4], as a pictorial scale of closeness in which participants can describe their relationship with an 'other' by selecting a picture from a set of Venn-like diagrams which depict two circles that overlap to differing degrees. The overlapping area of the different circles changes linearly from one picture to the next, and can be compared visually by the participant, in terms of absolute degrees of overlap rather than merely relative to the adjacent images. Because of this we treated participant responses to this scale as a seven-point interval scale.

In addition to the experimental items in the questionnaires we collected demographic data from the participants including age, gender, and prior experience with robots.

4 Results

4.1 Characteristics of the Sample

Our sample consisted of 20 participants. The mean age was 33.7, ranging from 25 to 56, and with a median age of 30.5. The sample included 5 females and 15 males. Experience with robots was evaluated with a 5 point Likert-Scale. The average experience reported in the sample was 3.2, ranging from 1 to 5, with a median of 3.

4.2 Results for Godspeed Questionnaire Subscales

To address our research questions, we evaluated the differences in participant ratings of the robot's behavior along the three above mentioned subscales of the Godspeed Questionnaire between our different experimental conditions. We calculated the subscale scores for each dimension for each condition and treated all subsequent analyses of the Godspeed Questionnaire subscales on the dimension subscales rather than the individual items.

The descriptive statistics for anthropomorphism are presented in Table 1. These results suggest that for the Anthropomorphism dimension participants overall rated the robot with a higher than neutral score of 3 in the model condition (1), and the artificial condition (2). The participants rated the robot lower than 3 only in the still condition (3). The effect of the different conditions on participant ratings along the anthropology dimension was assessed using a repeated measures ANOVA. The repeated measures ANOVA found a significant effect for condition (F (2, 18) = 14.32, p = 0, partial 2 = 0.614). Post-hoc tests suggest that there were significant differences between the model (1) and the control condition (3), and the artificial (2) and control condition (3), with the control condition (3) always receiving the lowest scores.

The results presented in Table 1 suggest that for the likeability dimension participants scored the robot higher than a neutral score of 3 in all conditions, with the highest score in the artificial condition . The effect of condition on participant ratings along this dimension was likewise assessed using a repeated measures ANOVA. The repeated measures ANOVA found a significant effect for Condition (F (2, 18) = 5.2, p = .017, partial 2 = 0.366). Post-hoc tests found significant differences between the artificial (2) and the control condition (3), with the control condition (3) receiving the lower score.

The descriptive statistics for the perceived intelligence subscale are presented in Table 1. The results suggest that for the perceived intelligence dimension participants overall scored the robot higher than a neutral score of 3 in all three conditions. The effect of the different conditions on participant ratings along the Perceived Intelligence dimension was assessed using a repeated measures ANOVA. The repeated measures ANOVA found no significant effect for condition (F (2, 16) = 2.666, p = 0.1, partial 2 = 0.25).

4.3 Results for Inclusion of Other in Self Scale

The results in Table 1 show that all the mean scores for the IOS scale are below the middle rating of 4. The repeated measures ANOVA found a significant effect for Condition (F (2, 18) = 9.656, p = 0.001, partial 2 = 0.52). Post-hoc tests suggest that there were significant differences between the model (1) and the control condition (3), and between the artificial (2) and the control condition (3), with the control condition (3) always receiving the lowest scores. The low average rating in the different conditions indicate, in combination with the large standard deviations, that the Inclusion of Other in Self scale might not be an appropriate evaluation instrument in this kind of setup.

Table 1. Descriptive Statistics

	Anthropo.		Likability		Perc. Intell.		IOS Scale	
Condition	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Model	3.16	0.78	3.92	0.56	3.4	0.65	3.25	1.41
Artificial	3.07	0.84	4.1	0.65	3.41	0.72	3.2	1.47
Control	2.37	0.73	3.52	0.81	3.1	0.62	2.25	1.12

5 Discussion

Our results show that the participants preferred the interactions in which the robot moved its eyes. For the anthropomorphism subscale, both the model condition (1) and the artificial condition (2) were rated higher than the control condition (3) in which the robot looked straight at the person. Interestingly for the likeability subscale only the artificial condition (2) was rated significantly higher than the control condition (3), and for the perceived intelligence subscale we didn't find any significant differences between the different conditions. The results for the Inclusion of Other in Self scale show the same pattern as the anthropomorphism subscale, with both the model and the artificial condition being rated higher than the control condition, but with no significant difference between them.

At first glance these results are in line with the large body of previous research on gaze in HRI, i.e. if a robot has eyes, humans prefer them to be animated. Synchronized lip movements were in our case not enough to make the robot appear as anthropomorphic and likeable as in the condition with random eye movements. It might be the case that eye movements are more important than lip movements for comfortable and intuitive interactions between robots and humans.

However these findings also illustrate some of the limitations of this study. At the current state of our implementation the participants were not able to tell the difference between eye gaze movements based on human data and the model with incorrect state transitions. This might have a few reasons.

The robot's gaze targets for the each state are fixed relative to the face location. The robot always looks to the center of a facial feature or (more noticably) a fixed distance and direction away from the face. This simplified away gaze behavior is not very naturalistic. How to generate natural away gaze is also unclear. There is a great deal of variability between individuals in away gaze locations and directions in our human data. Anecdotally, some participants reported afterwards that they couldn't understand why the robot would look at certain positions at certain times. For them this made the robot's behavior unpredictable. This might be one of the reasons why we didn't find any significant differences between the model and the control condition in the results for the likeability subscale and perceived intelligence subscales.

While based on human data, the model is not a model of mutual gaze. It is not reactive to the human partner's gaze direction and does not model the relationship between partners' gaze states. People may be sensitive enough to the contingency of gaze that differences in noncontingent gaze patterns are not noticed or don't elicit a preference.

Another important reason why people were not able to distinguish between the model (1) and the artificial (2) condition might be that the participants only listened to the robot. There was no conversational turn taking between the robot and the participants. Our model however was based on gaze data of dyadic human conversations. We were aware of this limitation from the beginning, but wanted to test the gaze controller without considering speaker role as a preliminary step before modeling speech's relationship to gaze.

6 Conclusions and Future Work

Despite the limitations of this study, the results are promising for future work on naturalistic gaze control for humanoid robots. We could verify that gaze behavior is important for robots with physical eyes and encountered a number of issues that help us to understand how to improve and expand our gaze controller in the next experimental iteration.

The work presented here is a first step in larger experimental cycle. The next steps will be extending the model to be reactive to human gaze and integrating speech detection on the robot in order to enable conversational turn taking. We assume that being engaged in an dyadic interaction with the robot, instead of passively listening to it, will change the perception of the robot's gaze behavior. Additionally we are re-analysing our original data set in order to more accurately model gaze shifts away from a partner's face.

7 Acknowledgements

This research was supported by the Scottish Informatics and Computer Science Alliance SICSA.

References

1. Admoni, H., Scassellati, B.: Social Eye Gaze in Human-Robot Interaction : A Review. J. Human-Robot Interact. (2017)
2. Annosoft: Phoneme mapping. http://www.annosoft.com/docs/Visemes17.html (2015)
3. Argyle, M.: Bodily communication, 2nd ed. Routledge (1988)
4. Aron, A., Aron, E.N., Smollan, D.: Inclusion of other in the self scale and the structure of interpersonal closeness. J. Pers. Soc. Psychol. (1992)
5. Baltrušaitis, T., Robinson, P., Morency, L.P.: 3d constrained local model for rigid and non-rigid facial tracking. In: Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. pp. 2610–2617. IEEE (2012)
6. Bartneck, C., Kulić, D., Croft, E., Zoghbi, S.: Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots (2009)
7. Breazeal, C., Scassellati, B.: How to build robots that make friends and influence people. Proceedings of the IEEE International conference on intelligent robots and systems (IROS 1999) 2, 858–863 (1999)
8. Broz, F., Lehmann, H., Nehaniv, C., Dautenhahn, K.: Mutual gaze, personality, and familiarity: Dual eye-tracking during conversation. IEEE International Symposium on Robots and Human Interactive Communications (RO-MAN) (2012)
9. Cappelletta, L., Harte, N.: Phoneme-to-viseme mapping for visual speech recognition. Proceedings of ICPRAM 2, 322–329 (2012)
10. Farroni, T.: Infants perceiving and acting on the eyes: Tests of an evolutionary hypothesis. Journal of Experimental Child Psychology 85(3), 199–212 (2003)
11. Heider, F., Simmel, M.: An experimental study of apparent behaviour. Am. J. Psychol. (1944)
12. Kleinke, C.: Gaze and eye contact: A research review. Psychological Bulletin 10(1), 78–100 (2003)
13. Kozima, H., Ito, A.: Towards language acquisition by an attention-sharing robot. Proceedings of the international conference on new methods in language processing and computational natural language learning (CoNLL) pp. 245–246 (1998)
14. Lester, J.C., Converse, S.a., Kahler, S.E., Barlow, S.T., Stone, B.a., Bhogal, R.S.: The Persona Effect: Affective Impact of Animated Pedagogical Agents. Proc. SIGCHI Conf. Hum. Factors Comput. Syst. - CHI '97 (1997)
15. MaryTTS: http://mary.dfki.de/ (2015)
16. Metta, G., Natale, L., Nori, F., Sandini, G., Vernon, D., Fadiga, L., Von Hofsten, C., Rosander, K., Lopes, M., Santos-Victor, J., Bernardino, A.: The icub humanoid robot: An open-systems platform for research in cognitive development. Neural Networks 23(8), 1125–1134 (2010)
17. Mori, M., MacDorman, K.: The uncanny valley. Energy (1970)
18. Roncone, A., Pattacini, U., Metta, G., Natale, L.: A cartesian 6-dof gaze controller for humanoid robots. Proceedings of Robotics: Science and Systems (2016)
19. Trevarthen, C., Aitken, K.: Infant intersubjectivity: Research, theory, and clinical applications. The Journal of Child Psychology and Psychiatry and Allied Disciplines 42(1), 3–48 (2001)