Towards Automated Human-Robot Mutual Gaze

Frank Broz*, Hatice Kose-Bagci†, Chrystopher L. Nehaniv* and Kerstin Dautenhahn*
* *University of Hertfordshire*
Department of Computer Science
Hatfield, UK
{*f.broz, c.l.nehaniv, k.dautenhahn*}*@herts.ac.uk*
† *Istanbul Technical University*
Computer Engineering Department
Istanbul, Turkey
hatice.kose@itu.edu.tr

***Abstract*—The role of gaze in interaction has been an area of increasing interest to the field of human-robot interaction. Mutual gaze, the pattern of behavior that arises when humans look directly at each other's faces, sends important social cues communicating attention and personality traits and helping to regulate conversational turn-taking. In preparation for learning a computational model of mutual gaze that can be used as a controller for a robot, data from human-human pairs in a conversational task was collected using a gaze-tracking system and face-tracking algorithm. The overall amount of mutual gaze observed between pairs agreed with predictions from the psychology literature. But the duration of mutual gaze was shorter than predicted, and the amount of direct eye contact detected was, surprisingly, almost nonexistent. The results presented show the potential of this automated method to capture detailed information about human gaze behavior, and future applications for interaction-based robot language learning are discussed. The analysis of human-human mutual gaze using automated tracking allows further testing and extension of past results that relied on hand-coding and can provide both a method of data collection and input for control of interactive robots.**

***Keywords*-mutual gaze; human-robot interaction; psychology; Markov model**

I. INTRODUCTION

Mutual gaze is an ongoing process between two interactors jointly regulating their eye contact, rather than an atomic action by either person [1]. This social behavior is important from an early developmental stage; even young infants are responsive to being the object of a caretaker's gaze [2]. Mutual gaze behavior is the basis of and precursor to more complex task-oriented gaze behaviors such as visual joint attention [3].

Mutual gaze is also an important part of face-to-face communication. It is a component of turn-taking "proto-conversations" between infants and caretakers that set the stage for language learning [4] and is known to play a role in regulating conversational turn-taking in adults [5]. There is evidence that children learn this coordination of gaze with conversational turns during early language acquisition, shifting towards an adult-like pattern as they gain more language skills [6].

Rather than measuring actual eye contact, "mutual gaze" is typically defined as subjects looking at one another's faces [1]. This is primarily due to the measurement limitations of having gaze direction coded by a human observer. There is limited evidence that people themselves may have trouble distinguishing gazes to different features within the face [7]. But the gaze patterns used in these experiments were highly unnatural, and people may access direction more accurately from natural conversational gaze.

Recently, the field of human-robot interaction has become increasingly interested in the role of gaze in a variety of conversational tasks, and robots have been programmed to produce natural-appearing mutual gaze behavior. But these robots base their behavior on models that typically focus on one important aspect of mutual gaze, such as reactivity or timing, while ignoring others. In work by Yoshikawa and colleagues, the robot responds to human gaze but do not take any action to regulate the duration of mutual gaze itself [8]. In the story-telling robot study by Mutlu, Forlizzi and Hodgins, a robot produces human-directed gaze behavior based on a model with realistic timings that is not responsive to real-time gaze information [9].

The characteristics of human gaze when interacting with robots is also an active area of research. Yu and colleagues performed a temporal analysis of human gaze and speech behavior from a human-robot interaction word teaching task with a robot that autonomously performed a simple form of joint attention [10]. While this study provides insight into patterns of human gaze at a robot, the simplicity of the robot's controller makes it unlikely that humans found the gaze interaction to be natural or its dynamics to be similar to gaze between two humans. Also, gaze was analyzed by looking at the entire robot rather than examining whether people fixated on the robot's face or particular facial features as they would with a human partner. In another human-robot study, Vollmer and colleagues found a significant decrease in gaze directed at the learner in a tutoring task when the

 ISBN: 978-1-61208-117-5

learner was a childlike virtual robot with a simple salience-based attention model rather than a human child [11]. They cite differences in the robot's visual feedback as a likely reason for the differences in tutor behavior. Gaze behavior in dyads is an interaction, and the robot's gaze policy will have an impact on the human. In order to support natural and effective gaze interaction, it is worthwhile to first look at gaze behavior in human-human dyads. By examining human gaze interactions, we can gain insight into how to build better gaze policies for robots that interact with people.

For a robot to successfully negotiate humanlike mutual gaze, it must both be responsive to the human's immediate gaze behavior and possess an internal model of mutual gaze based on time and other significant factors. Robotic systems designed to learn language through interaction by exploiting the structure of child-directed speech (e.g., [12]) could especially benefit from a gaze model that supports social engagement. Building models by using data collected from human-human pairs is likely to improve the quality of interaction with these systems.

There has been some previous research into using human-human gaze data to produce agent gaze. Raidt and colleagues conducted a study into face-to-face real time communication and gaze direction [13]. However, people interacted through a pair of video displays. While this is appropriate to their computer-agent model, it unnaturally constrains people's options for movement (as opposed to co-located face-to-face conversation). Also the speech task involved was one of repetition and memorization rather than natural conversation. Given these constraints on user behavior, it is unclear whether the data collected is representative of human conversational gaze behavior.

The rest of this paper describes a conversational gaze interaction experiment and its results. In Section II, the implementation details of the system used for data collection is described. Section III presents the design and setup of the experiment itself. Experimental results are discussed in Section IV, including overall amounts of mutual gaze and gaze duration, differences in behavior of individual pairings and gaze at specific areas of the face. In Section V a Markov model of the gaze data is discussed, both in terms of what it reveals about the data set and in terms of how such a model could be used for robot control.

II. System Overview

The automated detection of mutual gaze requires a number of signal-processing tasks to be carried out in real time and their separate data output streams to be combined for further processing. Note that if the goal of this work were solely to study mutual gaze in humans rather than to provide input for a robot control system, there would be no requirement for real-time operation. The video could be collected and then analyzed later offline. The system is a mixture of off-the-shelf programs and custom-written software combining

Figure 1. A pair of subjects engaging in conversation during an experiment.

and processing their output. The interprocess communication was implemented using YARP [14].

ASL MobileEye gaze tracking systems were used to collect the gaze direction data [15]. The output of the scene camera of each system was input into face-tracking software based on the faceAPI library [16]. Each participant also wore a microphone which was used to record a simple sound level (though this data was collected, it is not included in the experimental analysis in this paper). Timestamped data of gaze direction (in x,y image pixel coordinates) and the location of the partner's facial features (in pixel coordinates) were recorded at a rate of 30 hertz. In order to synchronize time across machines to maintain timestamp accuracy, a Network Time Protocol (NTP) server/client setup was used. This setup is typically able to maintain clock accuracy among machines within a millisecond or less over a local area network [17].

III. Experiment

Experiment participants were recruited in pairs from the university campus. A requirement for participation was that the members of each pair know one another. This requirement was chosen to limit one possible source of variability in the data, as strangers have been shown to exhibit less mutual gaze than people who are familiar with one another.

The pairs were seated approximately six feet apart with a desk between them. This distance was chosen to allow comparisons to be made to earlier studies. They were informed that they would engage in an unconstrained conversation for ten minutes while data was recorded. The participants were asked to avoid discussing potentially upsetting topics (so that extreme emotional reactions would not effect their behavior) and given a list of suggestions should they need one, which included: hobbies, a recent vacation, restaurants, television shows, or movies. After filling out a consent form and writing down their demographic information, each participant was led through the procedure to calibrate the

 ISBN: 978-1-61208-117-5

gaze tracking system by the experimenter before the trial began.

After the trial, participants were asked to complete a short questionnaire. This questionnaire's purpose was to collect additional data about individual traits that may have an influence on gaze behavior. Participants were asked which country they'd lived in for most of their life and how and how well they knew their partner for the experiment. They also filled out a ten item personality inventory (TIPI), a short personality test which assesses people on big five personality traits [18].

IV. RESULTS

Ten pairs of people participated in the study. Of these pairs, five experienced errors during data collection that resulted in their data being discarded from the study. The nature of these errors were: loss of gaze tracker calibration due to the glasses with the camera mount slipping or being moved by the participant, failure of the face tracker to acquire and track the face of a participant, and failure of the firewire connection that was used to transmit the video data to the computers for analysis. These failures reflect the difficulty of deploying a real-time system for mutual gaze tracking due to the complexity of the necessary hardware and software components. This experiment was the first of its kind conducted by this group, and quickly and reliably calibrating the system for each different individual was a process that required practice. Participants were promised that the experiment would last no longer than thirty minutes. In the case of difficulties with calibration or hardware failure, we continued the experiment and collected incomplete data from the working portions of the system so as not to inconvenience the participants. The five remaining pairs of participants for whom complete face and gaze tracking data were available were used for data analysis. They ranged in age from 23 to 69. Of the pairs, two were male-male, two were male-female, and one was female-female.

For each pair, the contiguous two minute period of their data with the lowest number of tracking errors was selected for analysis. The gaze behavior was divided into a set of high-level gaze states. In all pairs observed, one participant looked at their partner noticeably more than the other. The participant with the high face-directed gaze level will be referred to as the "high" participant and the partner with the lower level of face-directed gaze will be referred to as "low". The gaze states and their descriptions are given below:

- Mutual - mutual gaze, as defined as both participants' looking at one another's face area
- At Low - the high gaze level partner looks at the face of the low level partner while they look elsewhere
- At High - the low gaze level partner looks at the face of the high level partner while they look elsewhere
- Away - both partners look somewhere other than their partner's face

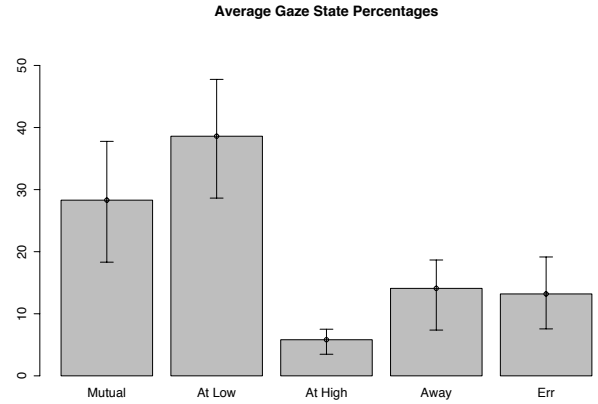

Figure 2. The average percentage of time spent in each gaze state by all pairs, with 95% bootstrap confidence intervals.

- Err - gaze state could not be classified due to missing gaze direction or face location readings

The average percentage of time spent by all pairs in each gaze state is shown in Figure 2. The mutual gaze results are in agreement with results from earlier studies which predict conversational mutual gaze at this distance to be around 30% [19]. It can also be seen that there is a marked asymmetry in the partners' gaze behavior, with one partner looking at the other's face far more often.

While the percentage of timesteps that cannot be classified due to tracking error is relatively high, we do not believe that it has a dramatic impact on the mutual gaze estimate. The reason for this is the source of the errors in the tracking system. Transient errors in the system occur for one of two reasons, lost gaze tracking readings or lost face tracking readings. Face tracking is usually lost when either of the participants move their head very quickly or when the partner's face is outside of the head-mounted camera's field of view (which happens when a person's head is directed away from their partner). Gaze tracking is lost when the system cannot find the participant's pupil, which typically happens when they are looking at a point at the periphery of their vision. Therefore, tracking errors occur most frequently when one or both participants have their head and/or gaze directed away from the other's face rather than towards it. This hypothesis will be supported by an examination of the pattern of transitions between gaze states (to be discussed in Section V).

A. Results for individual pairings

Figure 3 shows the gaze state percentages for each individual pair of participants. It can be seen that there are noticeable differences in gaze behavior between pairs. In particular, pairs three and five exhibit far more mutual gaze

 ISBN: 978-1-61208-117-5

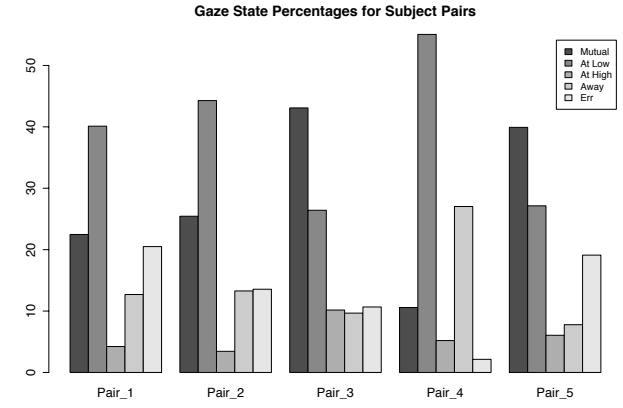

Figure 3. The average percentage of time spent in each gaze state by each pair.

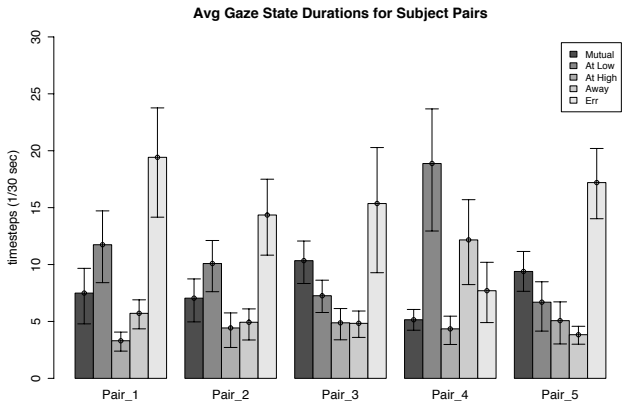

Figure 4. The average duration of time spent in each gaze state by each pair, with 95% bootstrap confidence intervals.

than the other pairs (these were a male-female and a male-male pair, respectively). We examined characteristics that might explain this difference, such as personality traits of extroversion and social dominance, how well the partners knew one another, and gender. The personality traits were measured through either single responses or a combination of responses (in the case of social dominance) from the short personality test administered. However, we failed to find a characteristic that could explain the difference in this small sample. In future studies with a larger number of participants we will focus on identifying characteristics that lead to different gaze dynamics and how knowledge about such characteristics might be incorporated into a gaze model for a robot to support more natural-seeming behavior.

One surprising result of this study is that the duration of each mutual gaze ranges from around one third to one sixth of a second (see Figure 4). These measurements are far shorter than a previously reported average figure of 1.18 seconds [19]. This may be due to the fact that automated techniques are more capable of measuring gaze at a fine temporal resolution than a human experimenter making judgements during observation. There is also a possibility that transient tracking error may cause the system to underestimate gaze durations. This will be investigated in future experiments using offline video analysis.

B. Gaze at significant facial features

Both the mouth and the eyes are features of great visual interest during communication because of the information they transmit through gaze and speech. The face tracking software used in this system allows the tracking of these features. In order to better understand what aspects of the face people attend to during conversation, the timesteps during which the pairs were in mutual gaze (looking at each other's faces) were classified according to what features were attended to. The regions of interest examined were:

- Mouth - this area is defined by the outside of the lip contour
- Eye - this area contains two separate regions for both the left and right eye
- Face - this area is defined as all of the face other than the mouth or eyes

The resulting percentages of mutual gaze time are shown in Figure 5. The vast majority of mutual gaze was made up of the pairs looking at somewhere other than the significant features on the face (the Face-Face state). The next most common state was "Mouth-Face", where one person looked at the other's mouth. The only other state that occurred with regularity for all of the pairs was "Eye-Face", though this was far less common than looking at the mouth.

Perhaps the most surprising result from this data is the lack of eye contact. Of all the participants, only Pair 1 exhibited any eye contact at all (and it was momentary). Because the measurement limitations we discussed earlier have prevented the study of direct eye contact in psychology, we have no way of knowing whether this result is out of line with normal human behavior, though it violates our naive expectations about gaze. There are a few possible explanations for this result. One is that it is possible, given that the eyes are a small target in the scene camera's image, that minor calibration errors in the gaze tracker may prevent at-eye gaze from being registered as such. The fact that eye-directed gaze was registered (in both the Mouth-Eye and Eye-Face states) proves that it can be detected by the system, though it could still be underestimated. Another possible explanation is that people actually don't frequently look into the center of other's eyes, preferring instead to glance at the area surrounding them. Given previous research [7] suggesting that people may have problems correctly assessing

 ISBN: 978-1-61208-117-5

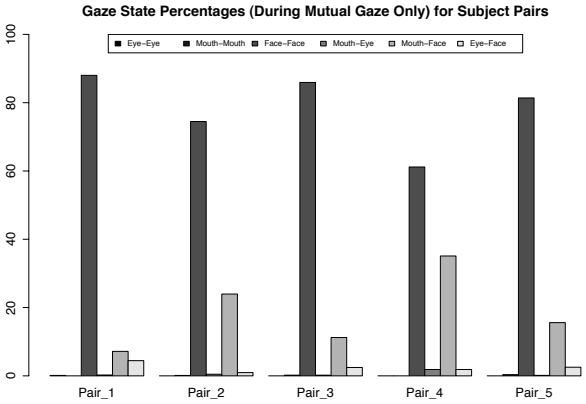

Figure 5. The average percentage of time spent by each pair attending to different facial features while in mutual gaze.

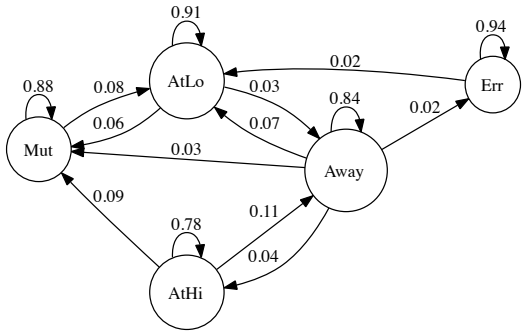

Figure 6. Markov model of the gaze state transitions for all pairs. For clarity, transitions of less than two percent probability are not displayed.

gaze direction even when it is directed at themselves, they may experience this as direct eye gaze. It is also possible that people typically exhibit more eye-directed gaze, and that the glasses used in the gaze tracking system cause people to fixate on the eyes of their partner less often than they would if their partner were not wearing the glasses. Artificially doubling the size of the eyes during analysis did not lead to a significant increase in the amount of eye contact detected (though it at least quadrupled the percentage of gaze classified as Eye-Face for each pair). So if near-eye gaze occurs frequently in the data either because of minor gaze tracker miscalibration or natural human behavior, it still does not explain the lack of eye contact registered. Future experiments will seek to further explore this unexpected result and verify that it is a real phenomenon and not a system limitation. Raw video will be recorded and analyzed offline to access the real-time system's accuracy.

Given that this was a conversational task, it is not surprising that the relative amount of mouth directed gaze was high. But it is unclear what is happening in the broadly defined Face-Face state. For controller design, the exact properties of gaze at seemingly non-significant parts of the face may make a difference between realistic and non-realistic gaze behavior. Further analysis will explore what areas of the face are being attended to and what the typical length and patterns of gaze fixations on them are.

V. Towards a model of gaze behavior

As a method of further analysis and as a first step towards using this data to implement a gaze controller for a robot, we created a Markov model of the interaction using data from all five pairs. A Markov model (or Markov chain) is a graphical probabilistic model that describes the state transitions of a system or process [20]. Data from the contiguous two minute period with the lowest error rate for each pair was combined to construct a model of their average behavior. This model is shown in Figure 6. Each gaze state of the interaction is a node in the model. The chance of reaching any other state from a given state at the next timestep is given by the probabilities on the outgoing edges from that state. The probability of staying in the same state at the next timestep is the probability of the state's edge that points back to itself. These self-transitions cause the time spent in each state to follow a geometric distribution, which agrees well with the form of the data observed. In order to improve the readability of the model and emphasize its major dynamics, transitions of less that 0.02 probability are not shown. From this model, it can be seen that gaze rarely alternates between being mutual and having both partners look away. Typically, one partner holds the face-directed gaze longer than the other. Also, note that the only state with a transition into the Err state with a large enough probability to be displayed is the Away state, which supports our hypothesis that tracking errors are non-uniformly distributed and most frequently occur when gaze is directed away from a partner's face.

VI. Conclusion

In this paper, a system for the real-time automated detection of mutual gaze was described, and results were presented from natural conversational interactions between human pairs. The overall level of mutual gaze observed was in line with predictions from the psychology literature on mutual gaze. But the durations of the mutual gaze episodes was far shorter. Additionally, an analysis of gaze at specific facial features found virtually no evidence of simultaneous direct eye contact between the conversational partners. These results highlight the potential for obtaining different, possibly more accurate measures of behavior using automated methods.

This real-time system is designed not purely for analysis,

 ISBN: 978-1-61208-117-5

but to provide gaze information as input to a controller for a humanoid robot in the future. As a demonstration of how we intend to use this human-human gaze data to produce a robotic gaze controller, we created a Markov model from the data collected and discussed how it captures the gaze behavior dynamics of the humans.

There are numerous opportunities for future work that could improve the sophistication and realism of conversational gaze controllers. This data has not yet been analyzed based on participants' conversational role (speaker or listener). We plan to do so in the near future, comparing the results we obtain with what is predicted by the psychology literature. Our resulting robotic gaze controller will use conversational role as an input in order to better support natural conversational interaction.

Another opportunity for analysis is to further examine people's patterns of gaze at facial features such as the eyes and mouth, as well as which non-significant areas of the face people gaze at during conversation. This could improve the life-likeness of the gaze actions taken by the robot. The human-human results for such an analysis is interesting in and of itself. There is a lack of results of this sort because of the difficulty of accurately measuring gaze in this manner without supporting technologies. Automated analysis could provide new insights into how humans interact through gaze as well as how they could interact with robots.

ACKNOWLEDGMENT

This research was conducted within the EU Integrated Project ITALK (Integration and Transfer of Action and Language in Robots) funded by the European Commission under contract number FP7-214668.

REFERENCES

[1] M. Argyle, *Bodily communication*, 2nd ed. Routledge, 1988.

[2] S. M. Hains and D. W. Muir, "Infant sensitivity to adult eye direction." *Child development*, vol. 67, no. 5, pp. 1940–1951, October 1996.

[3] T. Farroni, "Infants perceiving and acting on the eyes: Tests of an evolutionary hypothesis," *Journal of Experimental Child Psychology*, vol. 85, no. 3, pp. 199–212, July 2003.

[4] C. Trevarthen and K. J. Aitken, "Infant intersubjectivity: Research, theory, and clinical applications," *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, vol. 42, no. 01, pp. 3–48, 2001.

[5] C. Kleinke, "Gaze and eye contact: A research review." *Psychological Bulletin*, vol. 100, no. 1, pp. 78–100, 1986.

[6] L. D'Odorico, R. Cassibba, and N. Salerni, "Temporal relationships between gaze and vocal behavior in prelinguistic and linguistic communication," *Journal of Psycholinguistic Research*, vol. 26, no. 5, pp. 539–556, September 1997.

[7] M. v. Cranach and J. H. Ellgring, "Problems in the recognition of gaze direction," in *Social communication and movement : studies of interaction and expression in man and chimpanzee*, M. von Cranach, Ed. London: Academic Press, 1973, pp. 419–443.

[8] Y. Yoshikawa, K. Shinozawa, H. Ishiguro, N. Hagita, and T. Miyamoto, "The effects of responsive eye movement and blinking behavior in a communication robot," in *IROS*, 2006, pp. 4564–4569.

[9] B. Mutlu, J. Forlizzi, and J. Hodgins, "A storytelling robot: Modeling and evaluation of human-like gaze behavior," in *Humanoids*, 2006, pp. 518–523.

[10] C. Yu, M. Scheutz, and P. Schermerhorn, "Investigating multimodal real-time patterns of joint attention in an hri word learning task," in *HRI '10: 5th ACM/IEEE international conference on Human-robot interaction*. New York, NY, USA: ACM, 2010, pp. 309–316.

[11] A.-L. Vollmer, K. S. Lohan, K. Fischer, Y. Nagai, K. Pitsch, J. Fritsch, K. J. Rohlfing, and B. Wredek, "People modify their tutoring behavior in robot-directed interaction for action learning," in *DEVLRN '09: Proceedings of the 2009 IEEE 8th International Conference on Development and Learning*. Washington, DC, USA: IEEE Computer Society, 2009, pp. 1–6.

[12] J. Saunders, C. L. Nehaniv, and C. Lyon, "Robot learning of lexical semantics from sensorimotor interaction and the unrestricted speech of human tutors," in *2nd Intl Symp. on New Frontiers in HRI, AISB*, 2010.

[13] S. Raidt, G. Bailly, and F. Elisei, "Analyzing and modeling gaze during face-to-face interaction," in *7th International Conference on Intelligent Virtual Agents, IVA'2007 7th International Conference on Intelligent Virtual Agents, IVA'2007*, ser. 17-19 September 2007, Paris, France, Paris France, 09 2007, pp. 100–101.

[14] G. Metta, P. Fitzpatrick, and L. Natale, "Yarp: Yet another robot platform," *International Journal of Advanced Robotics Systems, special issue on Software Development and Integration in Robotics*, vol. 3, no. 1, 2006.

[15] Applied Science Laboratories, "Mobile Eye gaze tracking system." [Online]. Available: http://asleyetracking.com/

[16] Seeing Machines, Inc., "faceAPI." [Online]. Available: http://seeingmachines.com/

[17] D. L. Mills, "Improved algorithms for synchronizing computer network clocks," *SIGCOMM Computer Communication Review*, vol. 24, pp. 317–327, October 1994. [Online]. Available: http://doi.acm.org/10.1145/190809.190343

[18] S. D. Gosling, P. J. Rentfrow, and J. Swann, W. B., "A very brief measure of the big five personality domains," *Journal of Research in Personality*, vol. 37, pp. 504–528, 2003.

[19] M. Argyle and R. Ingham, "Gaze, mutual gaze, and proximity," *Semiotica*, vol. 6, no. 1, pp. 32–49, 1972.

[20] S. P. Meyn and R. L. Tweedie, *Markov Chains and Stochastic Stability*. Springer-Verlag, London, 1993.

 ISBN: 978-1-61208-117-5