Investigating Mutual Gaze

Frank Broz, Hatice Kose-Bagci, Chrystopher L. Nehaniv, Kerstin Dautenhahn
School of Computer Science
University of Hertfordshire, UK
{f.broz, h.kose-bagci, c.l.nehaniv, k.dautenhahn}@herts.ac.uk

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 and is a developmental precursor in humans to joint attention and language learning. In preparation for learning a computational model of mutual gaze, data from human-human pairs in a conversational task was collected using a gaze-tracking system and face detection algorithm. The results presented show the potential of this automated method. Future applications for interaction-based robot language learning are discussed.

1. Introduction

Mutual gaze is an ongoing process between two interactors jointly regulating their amount of eye contact rather than an atomic act by a single person (Argyle, 1988). This important social phenomenon is also one that becomes significant at an early developmental stage; even young infants are responsive to being the object of a caretaker's gaze (Hains and Muir, 1996). Mutual gaze behavior is the basis of and developmental precursor to more complex gaze behaviors such as visual joint attention (Farroni, 2003). It is also a component of turn-taking "proto-conversations" between infants and caretakers that set the stage for language learning (Trevarthen and Aitken, 2001). Mutual gaze is known to play a role in regulating conversational turn-taking in adults (Kleinke, 1986). There is evidence that children in the earliest stages of language acquisition also coordinate their gaze patterns with conversational turns, shifting towards an adult-like pattern as they gain more language skills (D'Odorico et al., 1997).

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 in these existing systems, robots either respond to human gaze but do not take any action to regulate the duration of mutual gaze itself (e.g., (Yoshikawa et al., 2006)), or they produce behavior based on a model with realistic timings that is not responsive to real-time gaze information (e.g., (Mutlu et al., 2006)). 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 such as (Saunders et al., 2010) could especially benefit from a gaze model that supports social engagement. A promising way of building such models is by using data collected from human-human pairs.

2. Experiment

The purpose of this research is two-fold. One goal is to collect data for the design of a gaze controller for a robot which will be capable of producing socially appropriate mutual gaze behavior. The other is to verify and further investigate human-human mutual gaze behavior. Studies of mutual gaze in the psychology literature have been conducted using human observation to encode the gaze data. If we hope to build on these findings to produce gaze behavior for robots, it would be useful to first confirm that we can duplicate them in data that was collected using the automated gaze detection methods necessary for robot control.

In light of these goals, an exploratory study was conducted involving a face-to-face conversational task between pairs of participants. In order to record their gaze direction and the location of their faces, each participant wore a head-mounted gaze tracking system and face detection software (based on the openCV library) was applied to each tracker's scene camera to locate the other participant's face in their visual field .

Seven pairs of people participated in the experiment, recruited from the staff and student population of the University of Hertfordshire. Because the level of familiarity between interactors has been show to have an impact on the level of eye contact, the pairs were all workplace acquaintances. During the experiment, two participants were seated face to face, approximately six feet apart, with a desk between them (this distance was selected so that a direct comparison could be made with existing results). They engaged in an unconstrained conversation for ten minutes. The participants were instructed to speak about emotionally neutral topics (nothing very personal or potentially upsetting). During the conversation, data was recorded for three trials of a duration of eighty seconds each.

3. Results

Two pairs had their data excluded from analysis because of obvious face-tracking errors (where the face

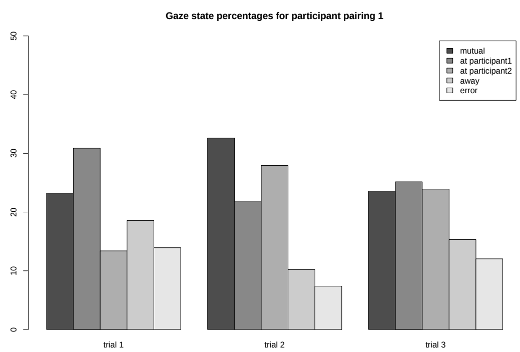

Figure 1: Results for participant pair 1

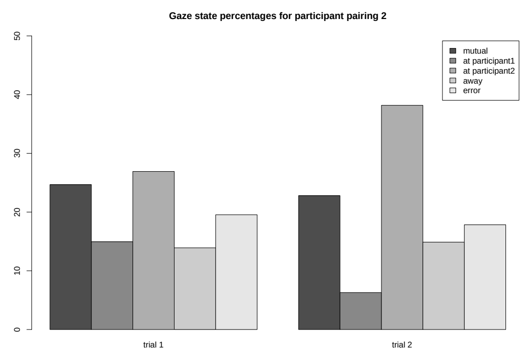

Figure 2: Results for participant pair 2

tracker became fixed on a non-face object during the trials). Of the remaining five pairs, three had an average error rate (the percentage of timesteps where the face detection or gaze tracking did not obtain valid coordinates) of forty percent or greater, making it impossible to draw conclusions from the analysis of their results. The two remaining pairs had an average error rate of below twenty percent, though one trial for the second pair also had to be discarded due to face tracking errors (see Figs 1 and 2).

While this error rate is relatively high, it allows a rough comparison to be made with results from earlier studies (Argyle, 1988). The predicted percentage of mutual gaze is 30%, and both pairs are near this figure for all trials. As this figure is an average over many individuals and conversations, individual pairs are not guaranteed or even expected to be close to it. But the fact that the data collected does not diverge wildly from the prediction is promising. It can be seen from the differences between the pairs (and their individual trials) that similar levels of mutual gaze can arise from varying levels of other-directed gaze by each individual, which is dependent on personality traits and other external factors. The average duration of mutual glances (continuous periods of mutual gaze) was predicted to be 1.5 seconds on average. The average duration observed for these trials was considerably shorter, ranging from roughly 0.33 to 0.66 second. The reason for this discrepancy will be explored further if it is still present in data from future studies.

The major source of error in this experiment was due to failures in face-tracking. The skin-tone-based tracking algorithm employed could become confused by wooden objects in the background (which are close to skin in the color space used) or transfer tracking to a hand when a participant touched their face. Future experiments will use a feature-based face tracking algorithm. This also has the potential to allow analysis of gaze at individual facial features, which may improve the realism of a data-derived controller. A secondary source of error was loss of calibration of the gaze tracker due to the glasses slipping or the participants touching or readjusting them.

4. Future Work

Further experiments will use improved face tracking algorithms which will also allow data to be collected about gaze at individual facial features. Periods of speech will also be automatically recorded, so that differences between speaker and listener conversational roles may be analyzed and compared with results from the literature. Data collected will be used to design a controller for a humanoid robot that can respond to a person's gaze behavior in real-time. This controller's ability to support naturalistic language teaching interactions between a childlike robot and a human participant will be explored.

Acknowledgements

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

Argyle, M. (1988). *Bodily communication*. Routledge, 2nd edition.

D'Odorico, L., Cassibba, R., and Salerni, N. (1997). Temporal relationships between gaze and vocal behavior in prelinguistic and linguistic communication. *J. of Psycholinguistic Research*, 26(5):539–556.

Farroni, T. (2003). Infants perceiving and acting on the eyes: Tests of an evolutionary hypothesis. *J. of Experimental Child Psychology*, 85(3):199–212.

Hains, S. M. and Muir, D. W. (1996). Infant sensitivity to adult eye direction. *Child Development*, 67(5):1940–1951.

Kleinke, C. (1986). Gaze and eye contact: A research review. *Psychological Bulletin*, 100(1):78–100.

Mutlu, B., Forlizzi, J., and Hodgins, J. (2006). A storytelling robot: Modeling and evaluation of human-like gaze behavior. In *Humanoids*, pages 518–523.

Saunders, J., Nehaniv, C. L., and Lyon, C. (2010). 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*.

Trevarthen, C. and Aitken, K. J. (2001). Infant intersubjectivity: Research, theory, and clinical applications. *J. of Child Psychology and Psychiatry and Allied Disciplines*, 42(01):3–48.

Yoshikawa, Y., Shinozawa, K., Ishiguro, H., Hagita, N., and Miyamoto, T. (2006). The effects of responsive eye movement and blinking behavior in a communication robot. In *IROS*, pages 4564–4569.