Author: Megha Joshi

Title: Investigation of Robots Influencing Social Context Based on Gaze Behavior

Adviser: Marynel Vazquez

Fall 2021

Abstract

Group interactions are an everyday occurrence for most of us, and often something we don't think much about. Such interactions occur in our lives all the time, however they can often be stunted by an unproductive social dynamic [1]. Such dynamics require an understanding of specific body language, such as our gaze. This prompts the question: how might a robot be able to influence a social context through gaze behavior? Shutter is a robot photographer designed to catch the attention of a passerby and ask if they would like their picture taken. It has the ability to follow a subject with its “eyes” and make some remarks about an individual's photo after it was taken. This technology could be utilized to study the possibility of gaze mechanisms in a social context, having Shutter be a part of a group dynamic by acting as the photographer for them. To determine the ethical implications of such work IRB training was conducted as well as ROS tutorials to understand how Shutter operates. In addition to this, past research was analyzed for the way in which they determined how a robot in a group context changed the gaze behaviors and attentiveness of individuals in a group. It was found that when the robot addressed all participants in a group, the attentiveness in a group context increased significantly. Shutter was then deployed in various instances at a spot where many students and faculty walked past it. We found that many times individuals were afraid to interact with Shutter, since they were unsure if they were allowed to. Finally Shutter’s body tracking capabilities were tracked empirically. We

concluded that Shutter struggled with individuals walking at different depths and speeds at different times, which most closely modeled how individuals walk past Shutter during the day.

Background

Perceiving gaze in a social context is one of the most important aspects of interaction. It is a task that seems fairly trivial to humans but can be nuanced in human-robot interaction settings. For example, for screen-based robotic heads, it is known that people might experience a sort of “Mona Lisa effect” when interacting with them [2]. They may perceive gaze from the robot as if it was establishing mutual gaze with them more often than is actually intended or geometrically correct, based on the true gaze direction of the robot. While this effect may seem like a detriment to interactions, perhaps it could be leveraged from an interaction design perspective to balance participation in group conversational settings. This project will build foundational technologies to study this possibility in the future.

To achieve effortless conversation, a robot's screen face should be able to address an individual in a group interaction, prompt conversation, and suggest participation through indirect methods such as gaze. This work is motivated by previous efforts in Human Computer-Interaction, specifically in speech recognition and multi-modal mechanisms for robots to convey gaze direction (figure 1).

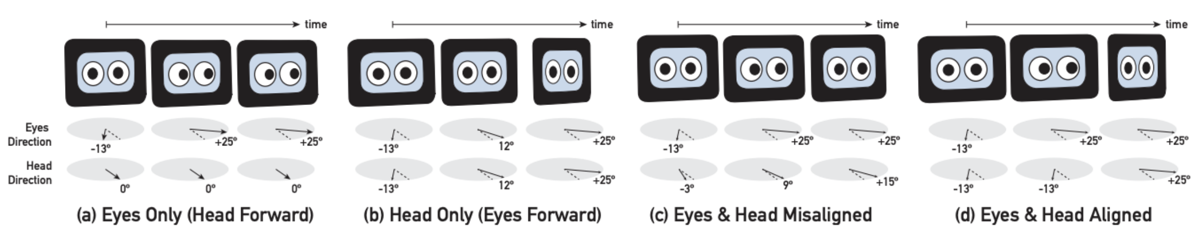

Figure 1: Prior research on development of different gaze behaviors based on eyes direction and head direction. Perceptions of the robot’s gaze direction, mutual gaze, width of the robot’s cone of direct gaze, and naturalness ratings were studied [1].

There are two primary objectives for this research: the first is to identify who is speaking in a social interaction and for how long they have been speaking, the second is to utilize prior research and code on gaze behaviors and request specific behaviors to be shown to an individual depending upon how long they have been speaking.

Project

Exploration

To begin research with Shutter, I needed to complete IRB training to understand the ethical implications of the project and human interactions with Shutter. I completed numerous modules to obtain certification which taught me about how to maintain privacy and confidentiality of test subjects, research with individuals of different age groups and avoid conflicts of interest. After taking these training sessions, I was able to understand why overt facial recognition in Shutter may not pass ethical principles, given that this takes away a subject's personal privacy they are entitled to. I also understood the importance of signage informing all subjects that they are being recorded whenever they approach Shutter.

To understand Shutter's operating system, I took a series of ROS (Robot Operating System) tutorials in which I installed and configured my ROS environment on a lab computer. I also learned how to build ROS packages with the latest build system for ROS called catkin. These tutorials helped me read the preexisting Shutter code better and helped me familiarize myself with how I could implement my own functionality for Shutter. Finally I set up my own workspace within Shutter, and worked through permission errors with people in the lab to ensure that it was set up correctly.

Finally I had to learn how Shutter is actually structured and how it processes commands [4]. Shutter is organized via a behavior tree, which enables modularity, hierarchical structure and an ability to efficiently identify where a problem may lie [5]. I discussed this structure with members of the lab as well as how trees use backwards chaining [6]. Through this process they are able to create a sequence of actions that lead to a desired outcome. So when Shutter gets inputs from a user about taking a photo, it processes these inputs down a behavior tree and then generates a sequence of actions that usually involve taking multiple photos of the user and generating comments based on these photos.

Research

My research included an analysis of papers that had shown success with the use of gaze mechanisms in multimodal group settings. Below I have shown my analysis of one of the notable papers from this analysis. Which will be particularly useful when evaluating the results of Shutter in a group setting.

<u>Conversational gaze mechanisms for humanlike robots [3]</u>

This paper focused on nonverbal mechanisms that establish who participates in a conversation and the context for which they participate [3]. It establishes the importance of gaze mechanisms in normal conversation and evaluates them in 36 trials of three-party human-robot conversations . These three party formations model those that exist in normal human interactions. It does so through three phases: computational modeling of key conversational gaze mechanisms, interaction design of these mechanisms for humanlike robots, and experimental evaluation of the social outcomes of manipulations in these mechanisms.

In order to analyze where the Speaker looked in a pattern, a frame-by-frame analysis of gaze behaviors was conducted. The target and time of each gaze change was recorded as were the speakers gaze targets. It was concluded that in all scenarios, the speaker looked at the Addressee for the majority of the time; 74%, 76%, and 71%. However it was also noted that most of the time the speaker looked at the individual's body rather than their eyes, which would have an influence when a robot is the Addressee.

The same process was then used with a robot, having all three members of the experiment take on the three different roles. Those that had the role of Speaker or Addressee had a raised attentiveness and felt more valued by their group. When the robot chose to acknowledge the presence of the side participant, their attentiveness also increased compared to when it was not acknowledged.

Deployment

I learned how to properly deploy Shutter and aided and deployed Shutter in the last few weeks of the semester. Deployments occurred in Becton hall at different times in the day. At some points Shutter would be active right after a lecture, having easily more than 30 individuals see Shutter and either choose to interact with it or walk past it. Some initial observations that were made after deployment were that the field of view was very shallow if the cart was pushed far back against the wall, and it seemed that subjects were apprehensive about interacting with Shutter because they were not sure if they were allowed to.

I then helped test Shutter's Kinect body tracking empirically. We observed the following situations:

1) Three individuals standing in different distances from shutter and each other
2) Three individuals walking at the same time in different directions
3) Three individuals walking at different times and different speeds in different directions

From these observations it seemed that Shutter struggled most with individuals walking at different speeds, since it kept switching its gaze when it saw a new individual and was unsure what to do if these individuals overlapped. Another observation was that the further back an individual was from Shutter in Becton the harder it was to detect, however this may have also been because the further back an individual was the closer it was to a large window emitting natural light. This may have been harder for an infrared sensor to detect.

Future Work

I will continue to help with the deployment of Shutter and identify a strong experimental design that allows us to track an individual's interaction with Shutter without interfering in the way a subject behaves around the robot. In addition to this I am interested in determining a social group influence to try to motivate someone in a group to wear a mask. Lastly, I am interested in identifying proper functionality for a mic array on Shutter. For example, having the users' inputs be vocal or through buttons to increase accessibility.

References

[1] Shamekhi, Ameneh, and Timothy Bickmore. "A Multimodal Robot-Driven Meeting Facilitation System for Group Decision-Making Sessions." 2019 International Conference on Multimodal Interaction. 2019.

[2] Vázquez, Marynel, et al. "Gaze by Semi-Virtual Robotic Heads: Effects of Eye and Head Motion." 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2020.

[3] Mutlu, Bilge, et al. "Conversational gaze mechanisms for humanlike robots." ACM Transactions on Interactive Intelligent Systems (TiiS) 1.2 (2012): 1-33.

[4] Rovida, Francesco, Bjarne Grossmann, and Volker Krüger. "Extended behavior trees for quick definition of flexible robotic tasks." 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2017.

[5] Banerjee, Bikramjit. "Autonomous acquisition of behavior trees for robot control." 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2018.

[6] Ögren, Petter. "Convergence analysis of hybrid control systems in the form of backward chained behavior trees." IEEE Robotics and Automation Letters 5.4 (2020): 6073-6080.