

Tracking Aggregate vs. Individual Gaze Behaviors during a Robot-led Tour Simplifies overall Engagement Estimates

Heather Knight Reid Simmons

Carnegie Mellon's Robotics Institute
5000 Forbes Avenue
Pittsburgh, PA

heatherbot@cmu.edu, reids@cs.cmu.edu

ABSTRACT

As an early behavioral study of what non-verbal features a robot tourguide could use to analyze a crowd, personalize an interaction and maintain high levels of engagement, we analyze participant gaze statistics in response to a robot tour guide's deictic gestures. There were thirty-seven participants overall, split into nine groups of three to five people each. In groups with the lowest engagement levels aggregate gaze response to the robot's pointing gesture involved the fewest total glance shifts, least time spent looking at indicated object and no intra-participant gaze. Our diverse participants had overlapping engagement ratings within their group, and we found that a robot that tracks group rather than individual analytics could capture less noisy and often stronger trends relating gaze features to self-reported engagement scores. Thus we have found indications that aggregate group analysis captures more salient and accurate assessments of overall *humans-robot interactions*, even with lower resolution features.

Categories and Subject Descriptors

[HRI Communication]: Conveying Intention, Gaze and Gestures,
[Robot Perception & Prediction]: Modeling social situations.

General Terms: Measurement, Experimentation, Human Factors.

Keywords: Human tracking, Humans-robot interaction, Social dynamics modeling, Low resolution sensing.

1. INTRODUCTION

Our experiment used a Nao robot, which made tourguide-style presentations in three distinct locations of the unusual Gates building at Carnegie Mellon University. In this paper, we analyze a sub-set of our larger video dataset in order to populate hypothesizes for further analysis and future human-tracking software. We examine participant gaze behaviors leading up to and following the robot's indication of a fire safety ceiling vent. The vertical location made gaze directions very distinguishable, ideal for tracking gaze statistics (see Fig. 1). The frequent convergence of within-group engagement levels meant that, given the high levels of variation in gaze data (not due to sensing limitations, but rather the complexity of human behavior), it is more fruitful to use aggregate features to estimate engagement than to look at each individual's gaze behaviors separately.

2. RELATED WORK

In [5], they evaluate robot behaviors that might attract a larger audience, however, to our knowledge, no other Robot Tour guide has investigated aggregate group engagement and attention tracking, though [1][3][5][6] discuss tour guides bots that can play games, button-in topic selection, have simple dialog and/or dance.

3. EXPERIMENTAL SETUP

In this study, we assess human gaze in response to a robotic tourguide to evaluate statistical trends relating to their engagement level. To generate a wide range of group reactions, we vary both the location and presentation style (*friendly, neutral, sarcastic*) over three Latin square combinations. Overall, both friendly and sarcastic presentations of the same factual material resulted in higher self-reported engagement scores as compared to neutral, based on survey results. In this initial analysis, we label participant gaze, using RGB video data to code viewer gaze statistics (the subject of this report), while reserving the Kinect depth data for training our human tracking software (future work).



Figure 1. a) Study capture of group gaze toward Vent, b) Pre-study testing of Kinect sensor mounted next to robot

The groups each had three to five participants and were unfamiliar with most the material the robot presented, as outlined in Table 1.

Table 1. Participant Demographic Information

Total #	Gender	Age	Tech Savvy	Comfortable with robots
34	13 Female / 24 Male	14-61	92%	5 No / 3 Sort-of / 26 Yes

4. RESULTS

Videos were hand-labeled by the author to indicate the gaze direction of each participant, e.g. Fig. 2, wherein horizontal height represents distinct gaze directions; *vent (topmost), window, each-other, robot, floor*. Group 1 had high engagement ratings; note the frequency of glance shifts and the small bumps toward the end at 12.5 and 13.5 seconds, which indicate intra-participant gaze. In Group 3 (low engagement) there is a fast drop off of gaze activity. The default gaze direction in all groups is toward the robot, as it is the one talking. Thus, participants at all engagement levels spend large portions of time gazing at the robot.

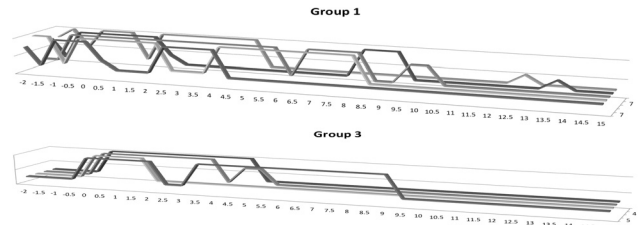


Figure 2. Aggregate Group 1 and 3 data, high/low engagement

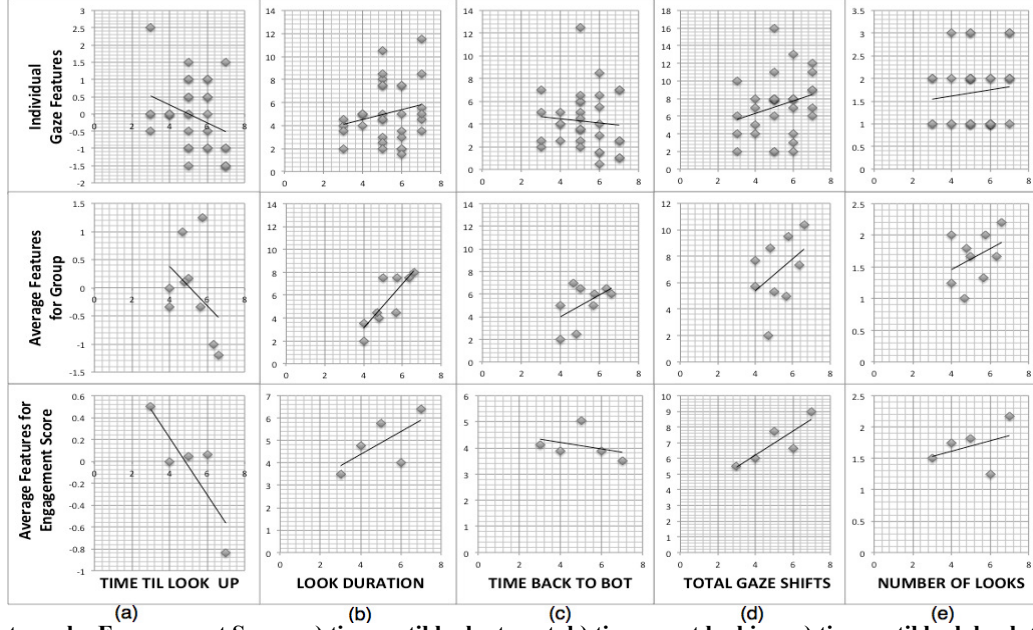


Figure 3. Features by Engagement Score: a) time until look at vent, b) time spent looking, c) time until look back to bot, d) total number of gaze shifts, e) number of glance-backs. The top row presents all individual data, middle averages the group data, and bottom averages all data from the same engagement score self-reports. (One outlier removed, as not enough samples at score=2)

4.1 Gaze Directions

To parse our data, we first isolate participant gaze directions (using data from 3 seconds before the robot's pointing gesture to 15 seconds following). For example, Fig. 4a displays all 34 participants' window-gaze data, organized from least-engaged (front-most) to most-engaged (back-most). Horizontal height indicates the binary presence of a gaze-direction. Note the trend for participants with lower engagement scores to gaze out the (non-topical) window sooner. Fig. 4b displays intra-participant gaze, notably occurring with higher engagement levels only.

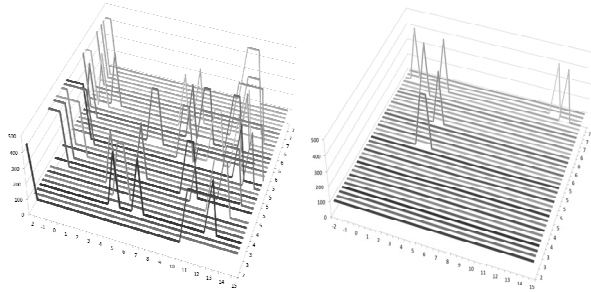


Figure 4. Participant Gaze toward (a) Window, (b) Each-other

4.2 Gaze Features

We next combine statistics from all gaze directions to identify higher-level gaze-based features we could use to classify audience engagement. Fig. 3 displays features such as 'time to look up,' 'total time spent looking at the vent,' 'total number of gaze shifts,' etc. The data shows weak trends correlating individual self-engagement reports with these individual features (top row), though the data spread is high.

When we instead aggregate features across the group (middle row) or all people with similar engagement scores (bottom row), we find clearer trends, e.g., in feature columns 9a (time-to-look-up) and 9d (number-of-gaze-shifts) aggregate data reduces overall

noise. In 9b (total-time-spent-looking), and 9e (total-number-of-vent-glances), aggregate results also find stronger positive trends.

5. CONCLUSION

In this experiment, we begin to uncover several possible advantages to tracking the aggregate gaze behaviors of a group over tracking individuals one-by-one:

1. Aggregate analysis means the robot can use lower resolution features that may be easier to detect.
2. Group tracking is more likely to capture unique and salient events that individual analysis might miss, e.g., shared glances or synchrony/divergence of behavior.

In separate work, we are developing algorithms to estimate gaze direction. The behavioral findings presented here will inform the development of our human tracking software. In the case of a bifurcated group, we also wish to assess if storing the engagement data spread better models the dynamics of how its topic and tour selection are received.

6. REFERENCES

- [1] Gunhee K, et al. 2004. The Autonomous Tour-Guide Robot Jinny. (IROS '04).
- [2] Knight, H. A Savvy Robot Standup Comic: Online Learning through Audience Tracking. Workshop paper (TEI '10)
- [3] Nourbakhsh, I.R., Kunz, C., Willeke, T. 2003. The mobot museum robot installations: five year experiment. (IROS '03)
- [4] Shiomi, M. Kanda, T. Ishiguro, H. Hagita. 2010. A larger audience, please! — Encouraging people to listen to a guide robot. (HRI '10).
- [5] Thrun, S., et al. 1999. MINERVA: a second-generation museum tour-guide robot. In Proc. Int. Conf. on Robotics and Automation (ICRA '99)