Do you see what I see? Gaze understanding in people, 3D-rendered robot heads, and virtual reality

Akash Singh, Abrar Anwar, Justin Hart

The University of Texas at Austin
Department of Computer Science
College of Natural Sciences

LARG
Learning Agents Research Group
The University of Texas at Austin

AUTONOMOUS ROBOTS

Motivation

- Implicit communication refers to communication which does not require an explicit communicative act.
- Gaze direction is a powerful cue for implicit communication.
  - Gaze reveals what the viewer is paying attention to. The viewer does not need to look at an object with the intention of communicating, but an observer can look to their face and identify where they are looking [1].
  - For example, our lab has previously validated gaze's importance in coordinating passing behaviors in hallways [3].
- It is often difficult to interpret the gaze of a 3D-rendered virtual agent on a computer monitor [2]. This is often referred to as the Mona Lisa effect, named for the fact that it always looks like the Mona Lisa painting is looking at the observer. This effect has also been used in film to create ambiguous gaze.
- This experiment investigates this by conducting a study on the gaze interpretation across various agents -- 3D, 2D, and VR.

Contributions

- An accurately modeled gaze algorithm which exhibits ocular vergence - where if, for each eye, a ray was taken from the center of the pupil of the robot head, the rays would intersect at the target the eyes are focusing on.
- This study uses a methodology to characterize the loss of gaze accuracy at various granularities and distances in people, 3D-rendered robot heads, and virtual reality
- Contrary to the Mona Lisa effect, we show that there is little to no loss in gaze accuracy between the rendered head and other agents.

Method

- Participants are asked to view an agent in person, on a monitor and in virtual reality and asked to identify what target the agent is focusing its gaze on
- The following target design is used to measure the distance between the true gaze of an agent and a user's perception of the gaze, seen in Figure 1
  - Designed a square target with differing levels of granularity.
  - Each level of granularity is subdivided into four quadrants; it is first distinguished by color, then letter, then number.
- 3 conditions:
  - **Monitor**: A 3D-rendered robot head in Unity produces a gaze into the physical world, where the virtual head is optimally placed such that it is the same size on the monitor as a physical head would appear from the perspective of the viewer.
  - **VR**: Using the same setup as above, an identical condition is reproduced in virtual reality, where the participant is able to perceive depth.
  - **Person**: A human produces a gaze into the physical world to provide a baseline for an interaction between two physical agents.
- For each condition, gazes are produced at 3 varying distances (tabletop, medium, and far) for targets on the left- and right-hand side of the participant, seen in Figure 4.
- We recruited 7 participants from the BWI Lab,, as we were unable to recruit participants due to COVID-19 restrictions.
  - Each participant completed 30 trials per condition.

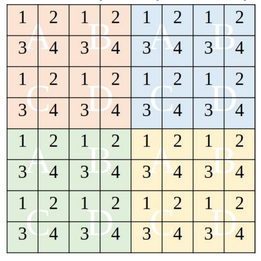

Figure 1. The design of the target consists of three levels of granularity. Each level of granularity is subdivided into four quadrants; it is first distinguished by color, then letter, then number. For example, the (Red, C, 3) refers to the number granularity, while (Red, C) is in the letter granularity.

Figures and Results

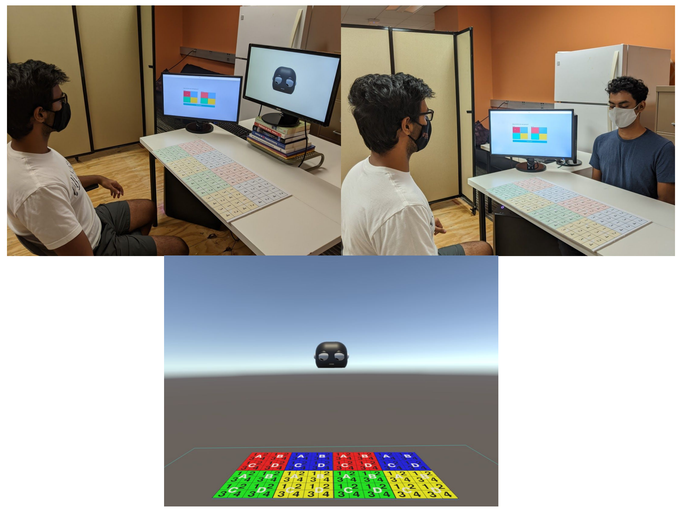

Figure 2. (top left, top right, bottom) the 3D-rendered robot head, human, and head in virtual reality, respectively, produces gazes onto the targets on the tabletop. The participant identifies where each agent type is focusing their gaze and records it onto the appropriate input interface.

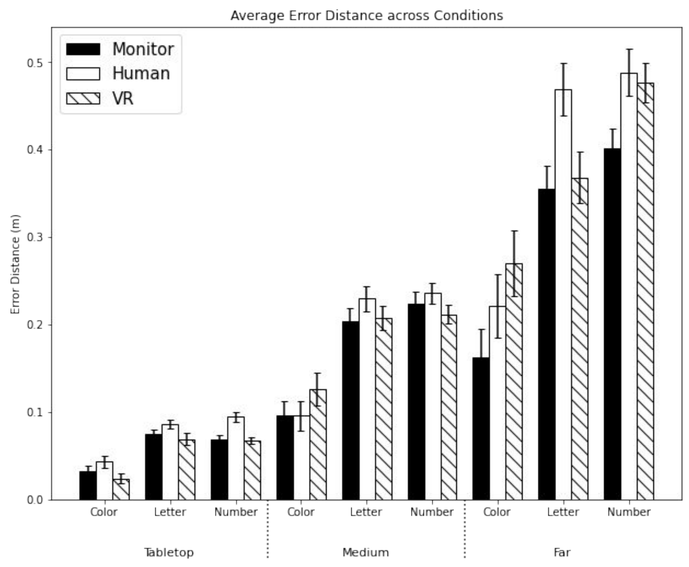

Figure 3. Average error distances across targets of varying distance and granularity in different conditions. Error distance is defined as the distance between the cell selected by the user and the cell which the agent was looking at. These results indicate that as distance increases and granularity gets finer, the error increases. Moreover, across the different conditions, we achieved a similar gaze accuracy. This is contrary to the Mona Lisa effect which expects a worse performance on virtual agents

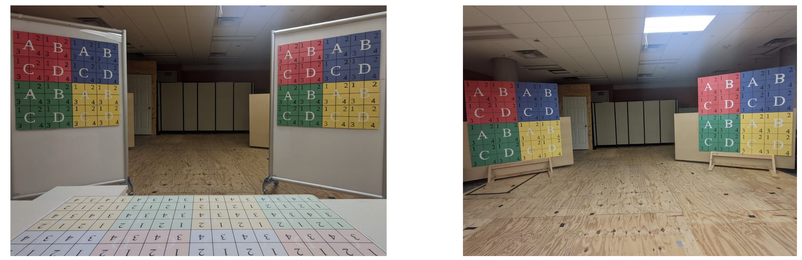

Figure 4. (left) The tabletop target is shown on the bottom, and the two targets above are the left/right medium distance targets located 2.5 meters away. (right) The far target located 5 meters away from the participant is shown. The agent makes a gaze at one of these targets at the various granularities (color, letter, and number), and the participant identifies the location which the agent is looking at.

Conclusion

- **Future Work**:
  - Recently, we have built a mathematical model of when we expect for the Mona Lisa Effect to arise.
  - We can model two renderedrobot heads in which the eyes appear to be looking at the same point, but are looking at different points due to a combination of the magnification of the camera's lens and the range at which the head is placed. This introduces ambiguity in where a virtual agent is looking.
  - We plan to redo this experiment by intentionally introducing this ambiguity to see if there is a difference in gaze interpretation.

Acknowledgments

This work has taken place in the Learning Agents Research Group (LARG) at UT Austin. LARG research is supported in part by NSF (CPS-1739964, IIS-1724157, NRI-1925082), ONR (N00014-18-2243), FLI (RFP2-000), ARO (W911NF-19-2-0333), DARPA, Lockheed Martin, GM, and Bosch. Peter Stone serves as the Executive Director of Sony AI America and receives financial compensation for this work. The terms of this arrangement have been reviewed and approved by the University of Texas at Austin in accordance with its policy on objectivity in research.

References

[1] Samer Al Moubayed, Jonas Beskow, Jens Edlund, Björn Granström, and DavidHouse. 2011. Animated Faces for Robotic Heads: Gaze and Beyond. In Analysis of Verbal and Nonverbal Communication and Enactment. The Processing Issues, Anna Esposito, Alessandro Vinciarelli, Klára Vicsi, Catherine Pelachaud, and Anton Nijholt (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 19–35.
[2] Samer Al Moubayed, Jonas Beskow, Gabriel Skantze, and Björn Granström. 2012.Furhat: a back-projected human-like robot head for multiparty human-machinei nteraction. In Cognitive behavioural systems. Springer, 114–130.
[3] Justin Hart, Reuth Mirsky, Xuesu Xiao, Stone Tejeda, Bonny Mahajan, JaminGoo, Kathryn Baldauf, Sydney Owen, and Peter Stone. 2020. Using Human-Inspired Signals to Disambiguate Navigational Intentions. InProceedings of the12th International Conference on Social Robotics (ICSR'20). Springer-Verlag
[4] Samer Al Moubayed, Jens Edlund, and Jonas Beskow. 2012. Taming Mona Lisa:communicating gaze faithfully in 2D and 3D facial projections.ACM Transactionson Interactive Intelligent Systems (TiiS)1, 2 (2012), 1–25

Contact

Akash Singh akashsingh@utexas.edu
Abrar Anwar abraranwar@utexas.edu