Eye Gaze Reveals Intentions in Shared Autonomy

Short Paper

Henny Admoni
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15213
henny@cmu.edu

Siddhartha Srinivasa
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15213
siddh@cs.cmu.edu

KEYWORDS

human-robot interaction, intentions, nonverbal behavior, eye gaze

ACM Reference format:
Henny Admoni and Siddhartha Srinivasa. 2017. Eye Gaze Reveals Intentions in Shared Autonomy. In *Proceedings of Intentions in HRI Workshop at HRI 2017, Vienna, Austria, March 2017 (Intentions in HRI '17),* 4 pages.
DOI: 10.475/123_4

1 INTRODUCTION

Physical human-robot collaborations involve people and robots coordinating their actions to manipulate objects in their shared environment. The success of these collaborations depend partly on the robot's ability to predict and respond to what its human partner intends to do. For example, assistive robots for people with motor disabilities that can predict user intentions—such as which object a user wants to grasp [10]—can proactively assist users in achieving those goals.

However, user intentions are unobservable. Instead, robots must attempt to predict intentions and other mental states through observable behaviors. One observable behavior that conveys intentions—often unconsciously—is eye gaze. When people talk about objects in their environment, they look at them before verbally referencing them [4, 5, 21]. When performing a task, people gaze almost exclusively to objects involved in that task, rarely looking elsewhere in the environment [6, 14]. When the task involves manipulating objects, gaze is almost always directed to task-based locations, like obstacles and grasp locations on target objects [11].

Automatically reading and interpreting *human* eye gaze remains a problem for autonomous robots, in large part because gaze is difficult to automatically detect. Eyes are difficult to track from video recordings or the cameras typically used on robots. Sometimes, a person's head pose may be substituted for a direct gaze measurement [2, 22], but studies suggest this may be insufficient for precisely detecting where people are looking [12].

However, recent technological advances in eye trackers have made it possible to obtain fine-grained, real-time data about eye gaze expressed during human-robot interactions. Head-mounted eye trackers have become more popular in HRI research, because of their increasing affordability and data quality. These devices are worn like a pair of eye glasses (Figure 1). A world-facing camera records a participant's view of the environment, while an inward-facing eye camera tracks pupil direction. The trackers can then report the direction of eye gaze relative to the participant's view by overlaying gaze direction on the streaming world video. Trackers like the Tobii Pro Glasses [18] and Pupil Labs Pupil Headset [17] provide a direct channel for accessing participants' eye gaze.

Figure 1: Eye gaze detected in real time through a head-mounted eye tracker can reveal intentions and guide robot behavior in human-robot collaborations.

Some recent HRI research has used head-mounted eye trackers to detect human gaze behavior and adapt robot actions based on this gaze data in real time [9, 15, 19, 20]. This research shows that eye trackers can successfully be used to detect and respond to human gaze.

Because eye tracker technology is relatively new, however, the field does not yet have a deep understanding of eye gaze behavior during human-robot interactions, the way psychologists have achieved for human-only interactions. Our research aims to develop such a model for eye gaze during physical human-robot collaboration. We focus on a *shared autonomy* interaction, where the user controls a robot while the robot simultaneously attempts to assist the user by producing some autonomous behavior.

2 SHARED AUTONOMY

Shared autonomy systems are especially important for assistive robots, such as wheelchair-mounted robot arms. These robots provide a flexible, mobile, highly dexterous tool for performing activities of daily living without a caregiver's assistance. Assistive robots, such as the Kinova JACO and MICO [13] and the Exact Dynamics iARM [3], are typically teleoperated through a joystick or other

Intentions in HRI '17, Vienna, Austria

DOI: 10.475/123_4

input device (Figure 1). These robots enable users to grasp and move objects, like pouring water from a pitcher into a glass.

However, people with motor impairments that necessitate the physical assistance of a robot often do not have the physical capability to easily control the robot. To address this, shared autonomy systems try to predict the user's goal from their control input and then seamlessly assist the user in completing that goal while maintaining user control over the robot. Shared autonomy algorithms can reduce the amount of time required to grasp an object [10] and can automatically put the robot in the correct control mode for the user's current goal [7].

State of the art shared autonomy systems still depend on direct user input, such as through a joystick, to predict intentions. Our research aims to leverage the unconscious, natural, revealing signals of intent contained in human eye gaze to create better intention predictors for shared autonomy systems. Using indirect signals like eye gaze in addition to direct robot control can yield more robust systems. Furthermore, understanding mental states like cognitive load and stress—which can be detected through pupil size, for example—will allow shared autonomy systems to tune their assistance to the individual's immediate needs.

In short, this research has the eventual goal of amplifying the functionality of assistive robots by having them recognize and use natural human behavior like eye gaze to predict intent and other mental states, and so provide timely, automatic assistance through shared autonomy.

3 DATA COLLECTION

To build a model of eye gaze behaviors during physically assistive human-robot collaboration, we collected data about eye gaze during robot teleoperation with and without shared autonomy. To replicate a realistic activity that might be performed with an assistive robot arm, we tasked participants with spearing one of three pieces of food on a plate using a fork held in the hand of a Kinova MICO robot. To do so, participants had to maneuver the robot using a 2-axis joystick into a spearing position (vertically) above their desired piece of food. They then pressed a button that prompted the robot to autonomously move the fork down to table height to spear the food and return it to a serving position.

Participants completed the task under three levels of robot assistance:

(1) Teleoperation: the robot provided no assistance, so participants had to fully control the robot to position it above their desired piece of food
(2) Autonomy: the robot ignored all joystick input and autonomously identified the pieces of food, then selected one at random and speared it without participant intervention
(3) Shared autonomy: the robot attempted to predict the participant's target piece and assist toward retrieving that piece using a state-of-the-art shared autonomy framework [10]

Participants completed five trials under each level of robot assistance, for a total of fifteen trials. Each trial lasted between 30 seconds and 6 minutes, depending on user success at positioning the fork. Participants wore a Pupil Labs head-mounted eye tracker [17] while they completed the task. This eye tracker was individually calibrated at the beginning of the experiment, and recorded participant eye gaze during each trial.

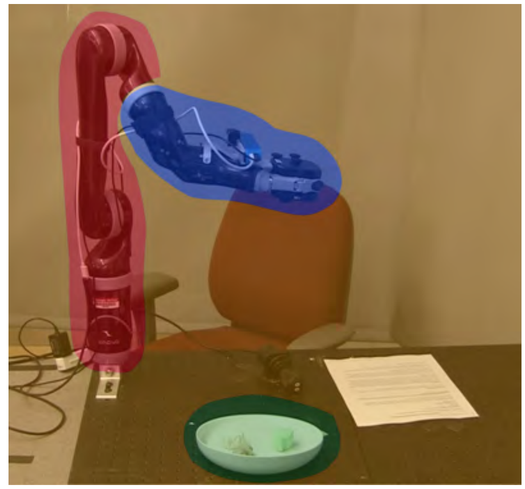

Figure 2: We coded the eye gaze videos based on the target of eye gaze: blue shows the area for "end effector," red shows "robot other," green shows "plate," and orange shows "other."

3.1 Preliminary Quantitative Analysis

We analyzed eye gaze data from nine participants, who represent a subset of a larger study (N=24) investigating task performance under shared autonomy. Participants were compensated $10 for their participation in this larger study. These nine participants were selected for gaze analysis because a significant portion of their gaze data was high quality, meaning that the eye tracker was able to identify a gaze position with at least 60% confidence. Of the nine participants, we excluded all teleoperation trials from one user because of a technical problem with the eye tracker. In total, we analyzed 85.5 minutes of high-quality gaze data drawn from approximately 3 hours of total trial time.

When people are completing tasks in human-only interactions, their eye gaze is focused almost exclusively on task-based targets, and rarely to locations or objects that do not have to do with the task at hand [6]. To investigate whether this pattern is the same during interactions with robots, we can measure gaze durations to various targets and analyze how gaze is distributed to task-based and non-task-based objects or locations.

In our case study, which involves spearing a piece of food with a fork held in the robot's hand, task-based locations are the robot's end effector (which includes the fork) and the plate of food. Non-task-based locations are other parts of the robot and other locations in the environment.

To extract these gaze durations, we coded the videos of eye gaze for the target of gaze locations in each frame (Figure 2). Gazes to the robot's hand, wrist, and fork were labeled as "end effector." Gazes to other parts of the robot, including the joints commonly referred to as elbow and shoulder, were labeled as "robot other." Gazes to the plate or pieces of food (while on the plate) were identified as "plate."

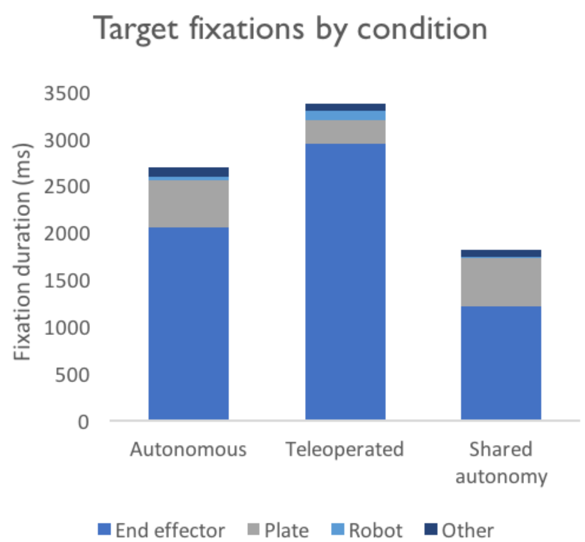

Figure 3: Duration of fixations to each world target (end effector, plate, robot, or other) by operation condition (autonomous, teleoperated, or shared autonomy). A significant proportion of user time was spent fixating on the end effector in all conditions.

Any gazes that did not fall in one of these zones were labeled as "other."

To establish whether there was a significant influence on operation condition or gaze target, we conducted a repeated measures multivariate analysis of variance (MANOVA). The repeated measures MANOVA tests for the effect of a single independent variable (condition) with three or more levels (autonomy, teleoperation, and shared autonomy) on two or more dependent variables (gaze timings to different targets) in a within-subjects design.

The repeated measures MANOVA revealed that there was a significant difference between gaze timings to different targets ($F(3,6) = 24.908, p = 0.001$, partial $\eta^2 = 0.926$). A post-hoc test using a Bonferroni correction for multiple comparisons showed that significant differences in gaze durations exist between end effector and robot other (mean difference 124.25 seconds, $p = 0.002$) as well as between end effector and other (mean difference 123.03 seconds, $p = 0.003$). No other significant difference between targets was found. There was also no significant effect of condition ($F(2,7) = 3.825, p = 0.075$), and no significant interaction between condition and target ($F(6,3) = 0.973, p = 0.555$).

Statistical analysis reveals that gaze toward one task-based target was significantly greater than gaze toward either non-task-based targets across all conditions. Qualitatively, this manifested as consistent gaze toward the object being manipulated (the end effector), with short monitoring glances toward the target of manipulation (the pieces of food on the plate). Thus, as in previous literature on human-only interactions [6], gaze in human-robot interactions is dominated by task-based targets. However, the large amount of gaze toward the robot's end effector emphasizes that the robot's hand is an object being manipulated, and not an extension of the user like the human's own hand. In human-only tasks, people rarely look at their own hands [11], whereas in this task, the large majority of gaze was directed at the robot's end effector.

3.2 Preliminary Qualitative Analysis

Task performance is not always smooth, however, especially when teleoperating the robot. As can be observed from the gaze timings in Figure 3, teleoperation had significantly more gaze to all targets than conditions that involved robot assistance. This was often the case because of the complexities of robot kinematics. The six DOF MICO arm make it simultaneously robust and difficult to manage. Therefore, the robot would sometimes enter tricky kinematic poses that were difficult to exit.

For example, one of the robot's joints would occasionally reach a limit, restricting further movement in that direction. For conditions with robot autonomy, the motion planners took joint limits into account and could plan trajectories that did not require exceeding them. However, because people were not aware of the robot's joint limits until encountering them, they sometimes controlled the robot into a position where the limit was encountered, and had difficulty returning to a viable pose to complete the task. These difficulties led to longer task completion times or even task failures when users could not get the robot out of its tricky kinematic position.

Eye gaze patterns changed when people encountered these kinds of difficulties in the task. Figure 4 shows one example of this: the user begins the trial by focusing on task-based objects like the plate and the robot's end effector. However, they encounter difficulty when the robot's elbow reaches a joint limit, restricting their ability to maneuver the end effector. At this point, gaze shifts from task-based objects to obstacles, specifically the robot's elbow joint. This shift is visible in the timeline of gaze targets.

Though not strictly an intention, task difficulty represents another mental state that can be identified through gaze. By recognizing when users are experiencing task difficulty, assistive robots can take action to correct that difficulty. In this case, the robot could plan a motion that brings it out of the problematic configuration.

4 FUTURE WORK

Once we understand the dynamics of eye gaze, we can make predictions about people's intentions based on observations of their gaze behavior. These predictions can be used to modify robot behavior to make robots more assistive, more responsive, and more pleasant to use. For example, in our assistive scenario, the robot can use predictions of a user's intent—the target object they want to manipulate—to provide assistance through shared autonomy. Researchers have already had success designing shared autonomy systems that predict user intent through direct inputs, such as joystick control [10] or even brain-computer interfaces [16]. A method for integrating eye gaze into shared autonomy is proposed in [1].

In addition to recognizing a user's intended target of manipulation, eye gaze can help reveal when people are intending to manipulate the robot into or out of a certain configuration. One example of this, as described above, is to use gaze to recognize when people are stuck in joint limits and other problematic kinematic configurations that make task completion difficult. Another

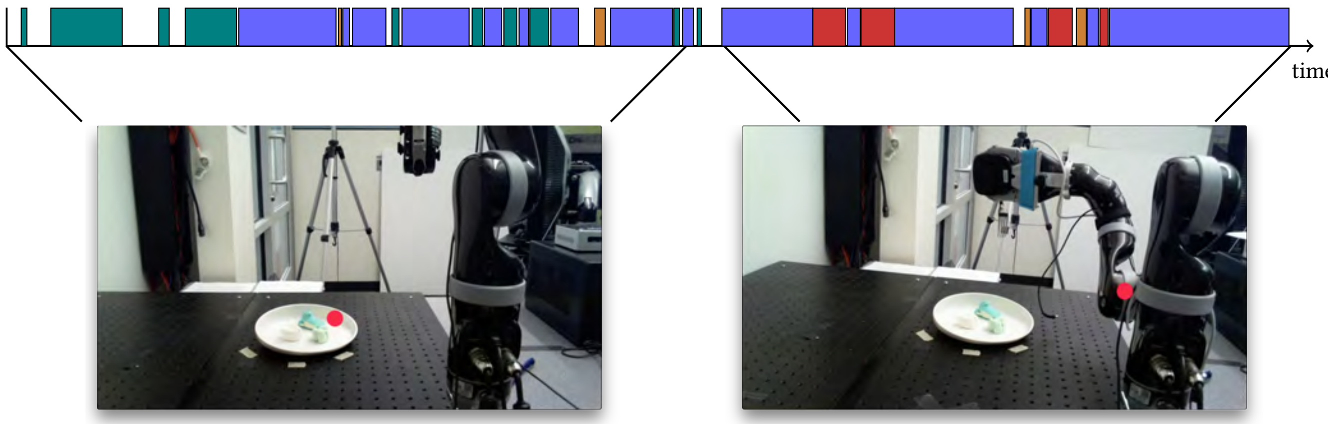

Figure 4: Eye gaze can reveal difficult parts of a task. The timeline shows which target the participant is fixating (purple: end effector; green: plate; red: robot other; orange: other). Still images show the eye tracker view with eye gaze overlayed as a red dot. The participant begins by focusing on task-oriented objects like the plate and end effector. However, when the robot enters a tricky kinematic configuration, the participant's eye gaze shifts to glances toward the robot's elbow joint instead.

example is to use eye gaze to maintain the visibility of the target during manipulation [8].

As we continue to analyze the data from this collection, we expect to draw additional insights into how peoples' eye gaze reveals their intentions during a shared autonomy manipulation task. These insights will help develop better human-robot collaborations by taking advantage of people's natural gaze behaviors to reveal their intentions.

REFERENCES

[1] Henny Admoni and Siddhartha S. Srinivasa. 2016. Predicting User Intent Through Eye Gaze for Shared Autonomy. In *Proceedings of the AAAI Fall Symposium: Shared Autonomy in Research and Practice*. 298–303.

[2] E.T. Bekele, Uttama Lahiri, AR. Swanson, J.A Crittendon, Z.E. Warren, and Nilanjan Sarkar. 2013. A Step Towards Developing Adaptive Robot-Mediated Intervention Architecture (ARIA) for Children With Autism. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 21, 2 (2013), 289–299. DOI: http://dx.doi.org/10.1109/TNSRE.2012.2230188

[3] Exact Dynamics. Exact Dynamics. http://www.exactdynamics.nl. (????).

[4] Zenzi M. Griffin and Kathryn Bock. 2000. What the Eyes Say About Speaking. *Psychological Science* 11, 4 (July 2000), 274–279. DOI: http://dx.doi.org/10.1111/1467-9280.00255

[5] Joy E. Hanna and Susan E. Brennan. 2007. Speakers' eye gaze disambiguates referring expressions early during face-to-face conversation. *Journal of Memory and Language* 57 (2007), 596–615.

[6] Mary Hayhoe and Dana Ballard. 2005. Eye movements in natural behavior. *Trends in Cognitive Sciences* 9, 4 (2005), 188–194. DOI: http://dx.doi.org/10.1016/j.tics.2005.02.009

[7] Laura V Herlant, Rachel M Holladay, and Siddhartha S Srinivasa. 2016. Assistive Teleoperation of Robot Arms via Automatic Time-Optimal Mode Switching. In *ACM/IEEE HRI*. 35–42.

[8] Rachel Holladay, Laura Herlant, Henny Admoni, and Siddhartha S. Srinivasa. 2016. Visibility Optimization in Manipulation Tasks for a Wheelchair-Mounted Robot Arm. In *RO-MAN Workshop on Human-Oriented Approaches for Assistive and Rehabilitation Robotics (HUMORARR)*.

[9] Chien-Ming Huang and Bilge Mutlu. 2016. Anticipatory robot control for efficient human-robot collaboration. In *ACM/IEEE HRI*. 83–90.

[10] Shervin Javdani, Siddhartha S. Srinivasa, and J. Andrew Bagnell. 2015. Shared autonomy via hindsight optimization. In *RSS*.

[11] R S Johansson, G Westling, A Bäckström, and J R Flanagan. 2001. Eye-hand coordination in object manipulation. *The Journal of Neuroscience* 21, 17 (Sept. 2001), 6917–6932.

[12] James Kennedy, Paul Baxter, and Tony Belpaeme. 2015. Head Pose Estimation is an Inadequate Replacement for Eye Gaze in Child-Robot Interaction. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts (HRI'15 Extended Abstracts)*. ACM, 35–36.

[13] Kinova Robotics, Inc. Kinova Robotics, Inc. http://kinovarobotics.com. (????).

[14] M F Land and M Hayhoe. 2001. In what ways do eye movements contribute to everyday activities? *Vision Research* 41, 25-26 (Jan. 2001), 3559–65.

[15] Gregor Mehlmann, Markus Häring, Kathrin Janowski, Tobias Baur, Patrick Gebhard, and Elisabeth André. 2014. Exploring a Model of Gaze for Grounding in Multimodal HRI. In *Proceedings of the 16th International Conference on Multimodal Interaction (ICMI)*. ACM, New York, NY, USA, 247–254.

[16] Katharina Muelling, Arun Venkatraman, Jean-Sebastien Valois, John Downey, Jeffrey Weiss, Shervin Javdani, Martial Hebert, Andrew B. Schwartz, Jennifer L. Collinger, and J. Andrew (Drew) Bagnell. 2015. Autonomy Infused Teleoperation with Application to BCI Manipulation. In *RSS*.

[17] Inc. Pupil Labs. 2016. Pupil Headset Website. (2016).

[18] Inc. Tobii. 2016. Tobii Pro Website. (2016).

[19] Tian Xu, Hui Zhang, and Chen Yu. 2013. Cooperative gazing behaviors in human multi-robot interaction. *Interaction Studies* 14, 3 (2013), 390–418.

[20] Yuichiro Yoshikawa, Kazuhiko Shinozawa, Hiroshi Ishiguro, Norihiro Hagita, and Takanori Miyamoto. 2006. The effects of responsive eye movement and blinking behavior in a communication robot. In *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS '06)*. 4564–4569. DOI: http://dx.doi.org/10.1109/IROS.2006.282160

[21] Chen Yu, Paul Schermerhorn, and Matthias Scheutz. 2012. Adaptive Eye Gaze Patterns In Interactions with Human and Artificial Agents. *ACM Transactions on Interactive Intelligent Systems* 1, 2 (January 2012).

[22] Abolfazl Zaraki, Daniele Mazzei, Manuel Giuliani, and Danilo De Rossi. 2014. Designing and Evaluating a Social Gaze-Control System for a Humanoid Robot. *IEEE Transactions on Human-Machine Systems* 44, 2 (April 2014), 157–168. DOI: http://dx.doi.org/10.1109/THMS.2014.2303083