![DiVA logo](http://www.diva-portal.org)
Postprint

This is the accepted version of a paper presented at *16th International Conference on Manufacturing Research, incorporating the 33rd National Conference on Manufacturing Research, University of Skövde, Sweden, September 11–13, 2018.*

N.B. When citing this work, cite the original published paper.

Permanent link to this version:
http://urn.kb.se/resolve?urn=urn:nbn:se:oru:diva-70706

Accessing your navigation plans! Human-Robot Intention Transfer using Eye-Tracking Glasses

Ravi Teja Chadalavada[1a], Henrik Andreasson[a], Maike Schindler[b], Rainer Palm[a],
Achim J. Lilienthal[a]

[a]*AASS MRO Lab, Örebro University, Sweden.*
[b]*Faculty of Human Sciences, University of Cologne, Germany.*

Abstract. Robots in human co-habited environments need human-aware task and motion planning, ideally responding to people's motion intentions as soon as they can be inferred from human cues. Eye gaze can convey information about intentions beyond trajectory and head pose of a person. Hence, we propose eye-tracking glasses as safety equipment in industrial environments shared by humans and robots. This paper investigates the possibility of human-to-robot implicit intention transference solely from eye gaze data. We present experiments in which humans wearing eye-tracking glasses encountered a small forklift truck under various conditions. We evaluate how the observed eye gaze patterns of the participants related to their navigation decisions. Our analysis shows that people primarily gazed on that side of the robot they ultimately decided to pass by. We discuss implications of these results and relate to a control approach that uses human eye gaze for early obstacle avoidance.

1. Introduction

During human-human interaction, humans rely on implicit and explicit, verbal and non-verbal cues to communicate [1], [2]: for mutual understanding, predicting future actions, and decision making. Similarly, robots operating in human co-habited environments should be capable of human-aware task and motion planning, ideally responding to expressions of human motion intentions as soon as they can be inferred from human cues. In public spaces, for instance an airport, a robot can typically deduce human intentions only with its onboard sensors, e.g., by using RGB-D cameras for trajectory and human head pose estimation [3]. In industrial environments, however, it is possible to issue regulations that require human workers to wear special safety equipment, e.g., safety vests or, as we suggest in this paper, safety eye-tracking glasses. Eye gaze can convey information about intentions beyond trajectory and head pose of a person [4]. Thus, we propose eye-tracking glasses as safety equipment in environments shared by humans and robots, eye-tracking glasses that are accessed wirelessly by a robot for implicit human intention transference. Especially in safety-critical, co-habited work places, e.g., warehouses or distribution centers with autonomous forklift trucks, this could reduce the number of accidents and enable more efficient operation.

This work presents an investigation of the possibility to recognize human navigation intent *implicitly* expressed through gaze patterns. We set out with the hypothesis that navigational intent can be identified at least to some extent and in some situations from gaze patterns. In this paper, we use data collected for a previous study on *'Communicating Motion Intentions in Human Robot Interaction (HRI)'* [5], [6]. During the experiments, participants encountered a robot while wearing eye-tracking glasses, which recorded eye gaze patterns and a video of the participants' view on the scene. We analyzed gaze patterns in relation to navigational decisions during human-robot encounters and found that people primarily gazed on that side of the robot they ultimately decided to pass by.

It was demonstrated that eye gaze is linked, in time and location, to momentary task requirements, e.g., [7]–[10]. Patla and Vikers [8], [10] report that people fixate points on which they will step approximately one second before reaching them. In addition to this "footprint fixation", the dominant pattern found was gaze being stable and travelling at the speed of the body of the person ("travel fixation") [8]. These findings support our hypothesis that navigational intent can be inferred from gaze patterns.

Gaze points and fixations are measures routinely used in eye-tracking studies. Gaze points are identified by an eye-tracker through back-projection of light hitting the fovea from the surrounding scenery, thus identifying areas a person looks at with acute vision. Fixations are detected as spatiotemporal clusters of gaze points and typically represented by the cluster center. Essentially, during fixations the eyes stop scanning and focus on one part of the scene so that the visual system can collect detailed information about what is being looked at [11]. Since fixations are typically detected in image space, however, standard fixation detectors do not work well in dynamic settings where the eye-tracker is generally in motion. We thus investigate the distribution of gaze points instead of fixations. This is different to the works of Huang et al. [12], Admoni et al. [13], Li and Zhang [14], [15] and Castellanos et al. [16], who also address the problem of recognizing implicit human intent, however in static settings.
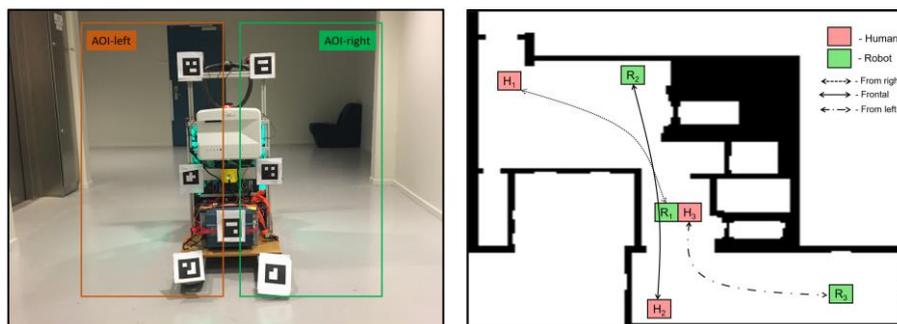
2. Setup and proposed approach

Figure 1. (Left) Robot platform used for the experiments and the pre-defined Areas of Interest (AOI): AOI-left (Orange) and AOI-right (Blue). (Right) Experimental layout design: During the experiment, the humans (denoted by H_i) moved from $H_i \rightarrow R_i$ and the small forklift robot (denoted by R_i) moved from $R_i \rightarrow H_i$.

The experiments took place in the basement of a university building using the robotic platform shown in Fig.1 (left). The experimental layout displayed in Fig.1 (right) was designed to reproduce different real-life encounters – from right, frontal, and from left

– in narrow situations. During the experiments, the participants wore eye-tracking glasses, which recorded gaze-overlaid videos of the participants' perspective while encountering the robot. Each participant had 12 encounters with the robot, corresponding to 12 decisions whether to pass the robot to the left or the right. During the encounters, the robot projected its navigational intent on the shared floor space in four patterns: *Line, Arrow, Blinking Arrow and No Projection*. We investigated the suitability of these different robot-to-human intention projection patterns in our previous work [5], [6]. For this paper we analyze whether the projected robot intention had a significant influence on the observed gaze patterns but do not investigate further reciprocations between human and robot intentions. Every participant encountered each pattern an equal number of times and the order in which the participants encountered the different patterns was randomized to compensate for learning effects.

2.1. Robotic Platform

The mobile base, see Fig.1 (left), was built upon a model of a manually operated forklift equipped with motorized forks. The forklift has been retrofitted with a steering mechanism and a commercial AGV control system. The latter is used to interface the drive mechanism, as well as the steering servo. To assure safe operation, the vehicle is equipped with two SICK S300 safety laser scanners facing in forward and backward directions. During the experiments, the robot followed predefined paths at 0.6m/s, which were generated by using a two-step approach which combines a lattice based motion planner and a smoothing operation [17].

2.2. Eye Tracker

For acquiring gaze data, Pupil eye-tracking glasses [18] were used. The world camera resolution was 1920x1080 pixels (at 30 fps) and the resolution of the two infrared spectrum eye cameras was 640x480 (at 120 fps). Scene capturing was done using the open source software Pupil Capture. The Pupil Player software was used for categorization and analysis. As can be seen in Fig.1 (left), fiducial markers were attached to the robot to define the areas of interest, thus, enabling an automatic categorization of the detected gaze points. We conducted manual marker calibration [18], where participants focus on a printed marker at different positions in the their field of view. With this calibration method, it was possible to cover a greater field of view and to calibrate on greater distances (2 meters) than with a screen-based calibration.

2.2.1. Areas of Interest (AOI)

In this work, we are primarily interested in finding how the gaze pattern of the participants was distributed on the left and right side of the robot in relation to their navigation decisions. Hence, two areas of interest (AOI): AOI-left and AOI-right were defined as shown in Fig.1 (left). The dimensions of AOI-left and AOI-right were chosen after manually checking the gaze-overlaid videos so that the defined AOIs captured most of the relevant gaze points around the robot at the encountered distances.

2.3. Experimental Procedure

The participants were greeted, introduced to the experiments, asked to fill a general questionnaire and to sign a consent form. The robot was then shown in a stationary position to familiarize with the platform. The path the participants had to take during

the experiment (see Fig.1, right) was then shown and the participants were informed about safety considerations. Finally, the eye-tracking glasses were calibrated, see Sec. 2.2, and the trials were started. The experiments took about 20-30 min.

3. Results

During the encounters, the number of gaze points on AOI-left and AOI-right were extracted and used to compute $GazePt_{left}$ and $GazePt_{right}$, denoting the relative percentage of gazes on AOI-left and AOI-right, respectively. For each encounter those gaze points were considered that fall into the time span where the fiducial markers could be tracked. For 22 participants (1:1 gender ratio with an age of 28.5±6.5 years and various backgrounds) and 12 encounters per participant, there were 264 decisions, which we all use in our analysis. The decisions of the participants to go left or right were identified manually from the recorded video and assigned to categorical numbers: decision = 0 for left, decision = 1 for right. *Gaze support*, denoting the percentage of gaze points on the AOI corresponding to the decision taken, was then computed as follows:

$$Gaze\ support = GazePt_{left} \cdot (1 - decision) + GazePt_{right} \cdot decision \qquad (1)$$

The distribution of *Gaze support* over all decisions is shown in Fig. 3 by two histograms, one with 20 bins (dark red/green bars) and one with 2 bins (light red/green bars). Overall, gaze support was in agreement with the navigation decision in 72.3% of the encounters, i.e. people moved in 191 out of the 264 encounters to the side they were looking at more often before passing the robot (they evaded the robot 106 times to the left and 85 times to the right). In 27.7% of the encounters gaze support was not in agreement with the navigation decision, i.e. people decided to move to the side they gazed at less frequently. Out of 73 such instances people passed by 43 times to the left and 30 times to the right.
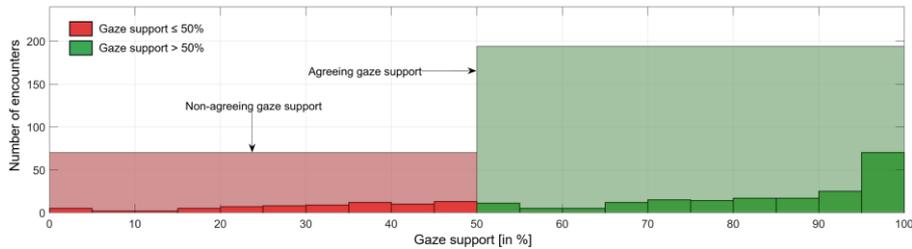

Figure 3. Histograms showing the distribution of *Gaze support* over all participants, all patterns, and all types of encounters. Red bars represent cases where people predominantly looked on one side and still moved to the other side. Green bars represent encounters where gaze was predominantly on the side chosen. Dark bars show a histogram with 20 bins. Light bars show another histogram of the same distribution with two bins.

Next, we evaluated the influence of the projected robot intention on the observed gaze patterns. As mentioned above, the robot used four different projection patterns: *Line, Arrow, Blinking Arrow and No Projection*. The leftmost four bar pairs in Fig. 4, show the distribution of *Gaze support* over all participants and all types of encounters, separately for each of those patterns. The red and green bars in Fig. 4 correspond to the light red and green bins in Fig. 3. The 2-bin histogram in Fig. 3 is reproduced for comparison in Fig. 4 (labelled "All Cases"). We observed more encounters with agreeing gaze support when a line was projected to communicate the navigation intent

of the robot. It is subject to further investigations to validate whether these differences are statistically significant and, if so, to identify possible reasons for this.

Finally, we analyze whether the type of encounter had a significant influence on the observed gaze patterns and counted cases of agreeing and non-agreeing gaze support over all participants, all projections but for the different types of encounters separately. The corresponding results are shown in the three rightmost bar pairs in Fig. 4. We observed that most cases with agreeing gaze support occurred during frontal encounters. Again, further work is required to validate whether the observed differences are statistically significant and, if so, to identify possible reasons for this.
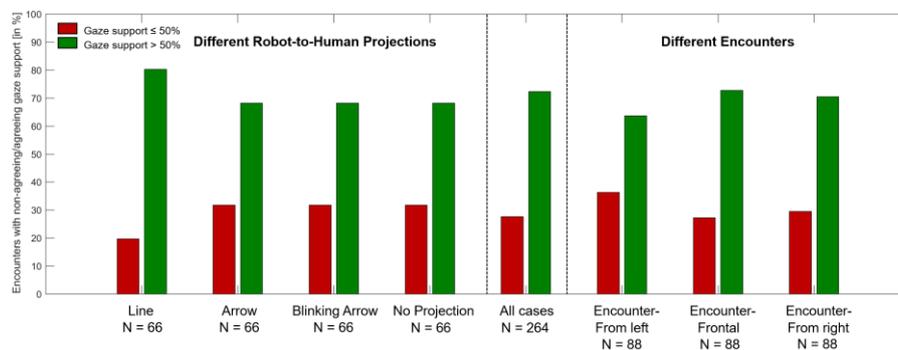

Figure 4. Percentage of encounters in which non-agreeing (red) and agreeing (green) gaze support was observed. Leftmost four bar pairs: all participants and all types of encounters, separated by robot-to-human projection pattern (Line, Arrow, Blinking Arrow, and No Projection). Fifth bar pair: all participants, all tasks, and all projections (corresponding to the 2-bin histogram in Fig. 3). Rightmost three bar pairs: all participants, all projections, separated by types of encounters.

4. Discussion

This work investigates the possibility of human-to-robot implicit intention transference through recognition of implicitly expressed navigation intentions solely from eye gaze data. We presented experiments in which humans wearing eye-tracking glasses encountered a robot in different situations and evaluated how the observed gaze patterns of the participants related to their navigation decisions. Our results show that, in the given scenario, a navigation intent predictor based on the simple rule, "If people look more often to one side of the robot they intend to go to that side", would have predicted the correct navigation intention in 72.3% of the encounters.

This result is encouraging and a springboard for further research. More experiments in different scenarios are needed to establish stronger statistical evidence of the findings in this paper. Further studies in varied settings are needed to reach general conclusions (independent of our particular setting) about differences with respect to the projection pattern of the robot or the type of encounter and to find out about possible reasons for these differences. Finally, future research should address the question at which distance to a robot navigation intent shows most strongly in the gaze of humans.

Gaze is not the only modality from which navigation intent can be inferred. Head and body pose and a person's recent trajectory also allow inference of navigation intent. Future research should aim to find confidence models for predictors based on different modalities with the goal to derive more reliable *joint* predictors. Finally, navigation

intent prediction needs to be integrated into human-aware motion planning. A suitable control approach based on implicit intention transference using eye-tracking, considering transmission to the robot, computation of states of human and robot and planning of the next motion step based on the velocity obstacles method is described in [19].

References

[1] B. Mutlu, F. Yamaoka, T. Kanda, H. Ishiguro, and N. Hagita, "Nonverbal leakage in robots: Communication of intentions through seemingly unintentional behavior," in *2009 4th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2009, pp. 69–76.

[2] C. Breazeal, C. D. Kidd, A. L. Thomaz, G. Hoffman, and M. Berlin, "Effects of nonverbal communication on efficiency and robustness in human-robot teamwork," in *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2005, pp. 708–713.

[3] R. Triebel *et al.*, "SPENCER: A Socially Aware Service Robot for Passenger Guidance and Help in Busy Airports," in *Field and Service Robotics*, Springer, Cham, 2016, pp. 607–622.

[4] O. Palinko, F. Rea, G. Sandini, and A. Sciutti, "Robot reading human gaze: Why eye tracking is better than head tracking for human-robot collaboration," in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2016, pp. 5048–5054.

[5] R. T. Chadalavada, H. Andreasson, R. Krug, and A. J. Lilienthal, "That's on my mind! robot to human intention communication through on-board projection on shared floor space," in *2015 European Conference on Mobile Robots (ECMR)*, 2015, pp. 1–6.

[6] R. T. Chadalavada, H. Andreasson, R. Krug, and A. Lilienthal, "Empirical evaluation of human trust in an expressive mobile robot," in *DIVA*, 2016.

[7] B. W. Tatler, M. M. Hayhoe, M. F. Land, and D. H. Ballard, "Eye guidance in natural vision: Reinterpreting salience," *J. Vis.*, vol. 11, no. 5, pp. 5–5, May 2011.

[8] A. E. Patla and J. N. Vickers, "How far ahead do we look when required to step on specific locations in the travel path during locomotion?," *Exp. Brain Res.*, vol. 148, no. 1, pp. 133–138, Jan. 2003.

[9] J. Jovancevic-Misic and M. Hayhoe, "Adaptive Gaze Control in Natural Environments," *J. Neurosci.*, vol. 29, no. 19, pp. 6234–6238, May 2009.

[10] A. E. Patla and J. N. Vickers, "Where and when do we look as we approach and step over an obstacle in the travel path?," *Neuroreport*, vol. 8, no. 17, pp. 3661–3665, Dec. 1997.

[11] K. Rayner, "The 35th Sir Frederick Bartlett Lecture: Eye movements and attention in reading, scene perception, and visual search," *Q. J. Exp. Psychol.*, vol. 62, no. 8, pp. 1457–1506, Aug. 2009.

[12] C.-M. Huang and B. Mutlu, "Anticipatory robot control for efficient human-robot collaboration," 2016, pp. 83–90.

[13] H. Admon and S. Srinivasa, "Predicting User Intent Through Eye Gaze for Shared Autonom," p. 6.

[14] S. Li, X. Zhang, F. J. Kim, R. D. da Silva, D. Gustafson, and W. R. Molina, "Attention-aware robotic laparoscope based on fuzzy interpretation of eye-gaze patterns," *J. Med. Devices*, vol. 9, no. 4, p. 041007, 2015.

[15] S. Li and X. Zhang, "Implicit Intention Communication in Human #x2013;Robot Interaction Through Visual Behavior Studies," *IEEE Trans. Hum.-Mach. Syst.*, vol. PP, no. 99, pp. 1–12, 2017.

[16] J. L. Castellanos, M. F. Gomez, and K. D. Adams, "Using machine learning based on eye gaze to predict targets: An exploratory study," in *2017 IEEE Symposium Series on Computational Intelligence (SSCI)*, 2017, pp. 1–7.

[17] H. Andreasson, J. Saarinen, M. Cirillo, T. Stoyanov, and A. J. Lilienthal, "Fast, continuous state path smoothing to improve navigation accuracy," in *2015 IEEE International Conference on Robotics and Automation (ICRA)*, 2015, pp. 662–669.

[18] M. Kassner, W. Patera, and A. Bulling, "Pupil: An Open Source Platform for Pervasive Eye Tracking and Mobile Gaze-based Interaction," *ArXiv14050006 Cs*, Apr. 2014.

[19] R. Palm and A. J. Lilienthal, "Long distance prediction and short distance control in human-robot systems," in *2017 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, 2017, pp. 1–6.