Improving Human-Robot Team Transparency with Eye-tracking based Situation Awareness Assessment

Favour Aderinto*
Oregon State University
Corvallis, OR
aderintf@oregonstate.edu

Josh Bhagat Smith*
Oregon State University
Corvallis, OR
bhagatsj@oregonstate.edu

Mark-Robin Giolando
Oregon State University
Corvallis, OR
giolandm@oregonstate.edu

Prakash Baskaran
Oregon State University
Corvallis, OR
baskarap@oregonstate.edu

Julie A. Adams
Oregon State University
Corvallis, OR
adamsjuli@oregonstate.edu

ABSTRACT

Human-robot interactions rely on transparency to foster effective collaboration. Transparency can be assessed through metrics associated with factors such as situation awareness. This manuscript presents an ocular metric to assess situation awareness for human-machine teams. Participants used a decision support system to select a grasp for underwater manipulation. The participants' gaze behavior and visual awareness was analyzed using a wearable eye tracker. An initial analysis that measures saccadic distance provides insight into the requirements of future techniques for objectively assessing situation awareness.

CCS CONCEPTS

• **Empirical studies in HCI**; • **HCI theory, concepts and models**;

KEYWORDS

situation awareness, transparency, eye-tracking, human-robot teams

1 INTRODUCTION

Transparency is the principle of providing easily exchangeable information to enhance humans' comprehension [13]. Human-robot interfaces must provide transparency into the robot's behavior in order for humans and robots to collaborate effectively. Transparency fundamentally impacts many aspects of a human's interactions with a robot (e.g., cooperation, communication, team performance, and team efficiency). Providing objective transparency metrics is critical to improving the capabilities of existing human-robot teams.

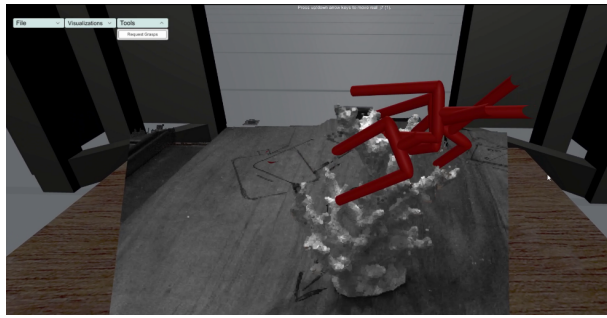
*These authors contributed equally.

Transparency is a complicated construct that is impacted by a broad range of factors (e.g., performance, usability). Situation Awareness (SA) is an indirect transparency factor that has broad implications. Two major aspects of transparency (i.e., performance and trust) are directly impacted by humans' SA [13]. Prior work demonstrated that in situ SA Probe Questions can be used to assess a human's awareness of the system but suggested that further insight can be gained via more objective metrics [12].
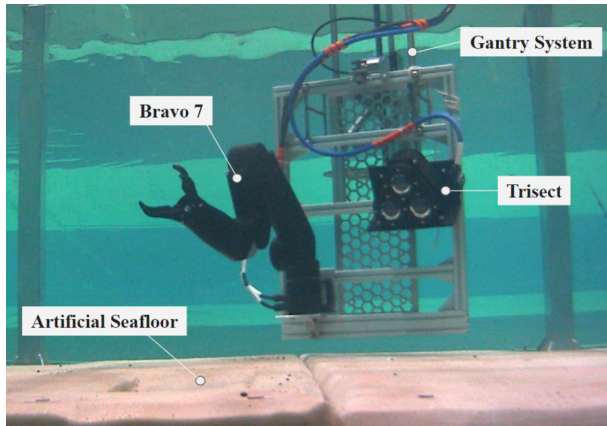
Eye-tracking is becoming an objective means of assessing SA [9, 16]; however, these metrics have never been directly applied to assessing transparency. Prior work leveraged metrics based on where the operator was interacting with a user interface as a proxy metric for operator gaze location to gauge SA, but these proxy metrics cannot be validated as representing aspects of the operator's SA [12]. This work also leveraged SA Probe Questions representative of the three levels of SA (i.e., perception, comprehension, and prediction). Combining eye tracking metrics (e.g., saccadic distance, gaze locations) related to the SA Probe Question timing (e.g., before, during, and after prompt) can provide a more objective assessment of SA and deeper insight into the system's transparency [12].

This manuscript presents an initial analysis of gaze-based metrics for objectively assessing SA and discusses requirements for future assessment techniques such that transparency in a human-robot team can be measured more reliably. A supervisory, shared autonomy human subjects evaluation was conducted for an underwater autonomous manipulation system [15], shown in Figure 1b. The decision support system provided potential grasp locations [10], from which the human selected what they believed was the best option to allow the robot to successfully grasp an object autonomously. Example grasp options and the decision support system interface are shown in Figure 1a. Participants executed a series of marine robotic manipulation tasks where ocular metrics were collected with a wearable eye-tracker.

Generally, the initial overall results demonstrate that saccadic distance is more variable in periods of lower SA. These initial results do specifically demonstrate that participants are evaluating how to best grasp an object at a time when they are seeking to gather information (Level 1 SA: perception) and think through which presented grasp option (Level 2 SA: comprehension) is most likely to result in a prediction of successful grasp (Level 3 SA: projection) [2]. However, the results on a per-object basis were not clear and

Favour Aderinto*, Josh Bhagat Smith*, Mark-Robin Giolando, Prakash Baskaran, and Julie A. Adams

(a) User Interface

(b) Underwater System

Figure 1: Candidate grasps as presented to the operator (a) were executed autonomously by the Bravo 7 manipulator (b).

Figure 2: The objects grasped by set in order from left to right.

require further analysis of how gaze-based metrics can be used to assess SA objectively.

2 RELATED WORK

Marine robotic environments present numerous challenges (e.g., noisy perception, constrained communication). The noisy, delayed, and potentially incomplete environmental perceptions can negatively impact an operator's SA and transparency into the true situation underwater. Interface design and effective decision support systems can intelligently process and convey information that enhances transparency and SA [10].

Transparency is critical to developing more effective human-robot interfaces. Performance, trust, explainability, and usability are the highest-degree direct factors of transparency [13]. SA directly impacts performance and trust and is indirectly related to both explainability and usability. More objective SA assessment techniques will result in more reliable measurements of these high-degree factors; thus, leading to deeper insights into transparency.

SA assessment techniques have primarily relied on subjective questionnaires administered either while an individual is operating or after an individual has finished operating a given system [2, 4, 14]. Some techniques (e.g., SA Rating Technique [2]) have individuals self-report va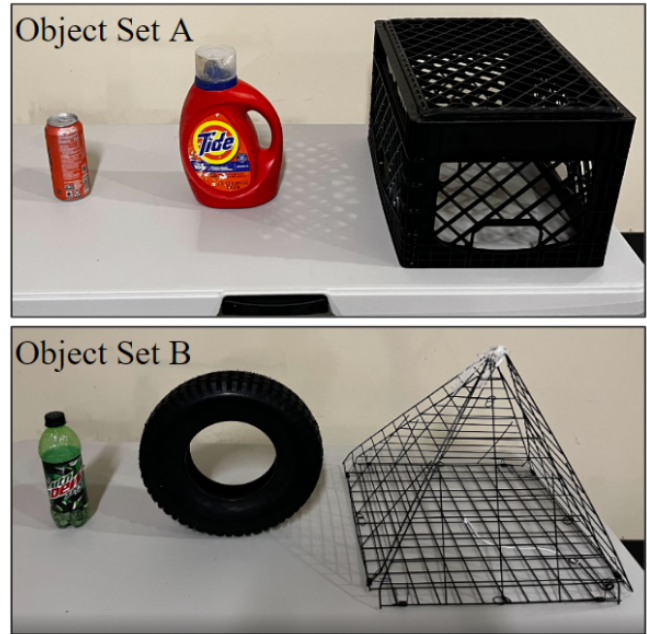rious internal aspects of their mental state to generate an aggregate SA metric. Other techniques (e.g., SA Global Assessment Technique [2], SA Probe Question [4]) administer queries concerning a system's current state to provide a more objective assessment. Prior work related to transparency used less intrusive but domain-appropriate in situ SA Probe Questions for each of Endsley's three SA Levels [1]: perception (Level 1), comprehension (Level 2), and prediction (Level 3). These questionnaires often distract from the primary task, which hinders performance and may confound other key aspects of transparency; thus, unobtrusive, real-time SA assessment techniques are preferred.

Real-time SA assessment techniques rely on an individual's physiological and behavioral patterns to estimate mental constructs related to SA [16]. For example, Level 1 SA (i.e., perception) is directly connected to where individuals dedicate their visual attention [3]. The distribution of visual attention correlates with SA, as spending more time looking for information indicates an individual exhibits lower comprehension of the system (i.e., Level 2 SA).

Eye trackers can provide insights into SA, as they can collect the necessary ocular metrics to capture visual attention [5, 9, 11]. Many ocular metrics have been shown to correlate with SA, but some are more useful than others. Unconscious ocular metrics (e.g., pupil dilation) are primarily involuntary responses to environmental stimuli and not directly connected to SA [16]. Conscious ocular metrics (e.g., fixations, saccades) are a direct indication of how visual attention resources are allocated, indicating where operators are looking for information required to understand the system's state. Both fixation-based (e.g., area of interest, fixation duration) and saccade-based metrics (e.g., saccadic amplitude, saccadic frequency) have been used in physiological SA assessments [6–8, 17]. Incorporating these metrics into objective SA assessment techniques can result in

a more informed transparency analysis and lead to better design recommendations [12]. Further, combining subjective and objective SA metrics can improve the analysis of transparency.

3   EXPERIMENTAL DESIGN

Participants were tasked with remotely grasping marine debris with an underwater robot arm. The participants completed a consent form and a demographic questionnaire before donning several wearable sensors, including the Pupil Labs Core eye tracker. Participants were presented with a brief instructional video before a training session where they grasped coral branches.

The trial session required participants to use a computer-based decision support system to complete six trials, each of which required grasping a different object. Two object sets containing objects of similar complexity were used (see Figure 2). The object set presentation order was randomized, but the objects within a set were always presented in the same order.

Fifteen participants (9 females and 6 males) completed the evaluation. Six participants were between the ages of 18 and 30, three were 31–40, three were 41–50, and three were over the age of 51. Four held a Bachelor's degree, six held a Master's degree, and the remaining five were high school graduates. All but one participant was right-handed. Participants rated their proficiency with computers (mean = 5.93, standard deviation = 1.22) on a scale from 1-low to 7-high. Five participants reported playing video games regularly, and ten participants reported never playing video games.

3.1   Methodology

The Pupil Lab's Core eye tracker, employed to monitor participants' gaze behavior during the grasping trials, records gaze location in the camera's reference frame. The Pupil Lab's Core eye tracker necessitates a transformation to align the gaze location with the screen's coordinates. Fiducial markers positioned at the monitor's four corners were used to calculate the disparity between the eye tracker's camera-based coordinates and the screen's reference frame in order to correct the participant's gaze locations. This transformation aligned the participant's gaze coordinates with the screen's reference frame using OpenCV to detect the fiducial marker positions. Gaze locations outside the screen's boundaries were excluded.

Gaze locations were recorded at three distinct time points: 15s before, during, and 15s after the selection of a potential grasp. Gaze locations over a 5-second window were collected; specifically, 2.5s before and after each of these three time points. For example, all gaze locations between 12.5s and 17.5s were considered for the *15s Before* selection time point. The primary metric utilized in this work is saccadic distance, which is defined as the length of a saccade between consecutive gaze locations. The mean and standard deviation of saccadic distance were computed for each object at all three time points. Additionally, the overall mean and standard deviation for each object set, represented by Object Set A and Object Set B, were calculated to provide an overall analysis. The presented results are for 13 of 15 participants' results due to data recording errors. It is important to note that although this evaluation did not incorporate the in situ SA Probe Questions, the evaluation does support developing objective eye tracking-based SA metrics.

4   RESULTS

Participants' gaze variability was assessed by calculating the mean and standard deviation of saccadic distance for each object and object set (see Table 1). Generally, when comparing the results across the timestamps, the highest mean saccadic distance occurred for the less complex objects 15s before selecting a grasp, while the more complex objects (e.g., crate and cage) had higher variability at the time of selecting a grasp. The magnitude of these increases varies between objects, with the objects in Set A demonstrating more variability. Additionally, the mean saccadic distance *15s Before* selecting the grasp is correlated with an object's grasp complexity for objects within an object set, where the easy objects (i.e., can, bottle) have the highest mean distance and the difficult grasp objects (i.e., crate, cage) have the lowest.

Table 1: Saccadic Distances Values: Mean (Standard Deviation) in mm

Object	15s Before	At Selection	15s After
Can	0.412 (4.332)	0.174 (1.263)	0.236 (2.763)
Jug	0.303 (3.463)	0.129 (0.330)	0.204 (2.376)
Crate	0.172 (1.920)	0.305 (3.626)	0.156 (2.124)
Object Set A	0.295 (3.389)	0.203 (2.225)	0.199 (2.432)
Bottle	0.240 (2.020)	0.160 (0.493)	0.119 (0.224)
Tire	0.158 (1.661)	0.218 (1.903)	0.242 (2.034)
Cage	0.113 (0.587)	0.143 (1.364)	0.157 (0.452)
Object Set B	0.169 (1.531)	0.174 (1.394)	0.173 (1.212)

The mean saccadic distance is similar for all objects during the *At Selection* and *15s After* periods, with the exception of the crate. The crate exhibited the highest mean and standard deviation across objects during the *At Selection* period. Mean saccadic distances were higher with larger standard deviations during the *15s After* period for four of the six objects when compared to the *At Selection* period. Objects of similar grasp complexity also had a similar mean saccadic distances across object sets during the same period, but easy and difficult objects had lower standard deviations. For example, the crate and the cage results 15s after selection had a mean of approximately 0.157, but the crate's standard deviation was five times that of the cage.

5   DISCUSSION

The increased variability of saccadic distance *15s Before* grasp selection suggests that participants were seeking information (Level 1 SA), processing (Level 2 SA), and projecting potential outcomes (Level 3 SA) to improve their SA. The mean saccadic distance roughly corresponds to scenarios where participants are either looking for information or assessing the system's state. Higher standard deviations are indicative of participants searching for information frequently. These behaviors are characteristic of lower SA [16], and were both present *15s Before* selecting a grasp.

These results are in conflict with the fact that mean saccadic distance decreased as object grasp complexity increased, which may suggest that participants required less information *15s Before* selecting a grasp for a more complex object. It is intuitive to suggest that increases in object complexity correspond with decreases in

SA due to the object's larger size and more difficult grasp execution. Complex objects were always presented after simpler objects, which may have been confounding, but more complex objects also provide more unique grasping options, which may have been easier for participants to interpret. Further investigation into the system interactions is required to understand these aspects of the results.

The results during the *At Selection* and *15s After* periods tell an equally complex story. Consistent mean saccadic distance between these two time periods is likely due to the fact that the participant's decision had been made and the system is communicating the selected grasp, planning the path from the manipulator's resting position to a grasping position, or executing the autonomous navigation and grasping behaviors. The higher values exhibited during the *At Selection* period for the crate object are possibly the result of a more difficult decision requiring further verification during the selection. Unlike other domains, where the human will have other tasks to attend to while the grasping actions are autonomously executed, this system does not require any other tasks. Thus, the generally higher standard deviations during the *15s After* period are the likely result of disengagement (i.e., lower SA). The decision support system takes from 10 to 20 seconds to begin executing a grasp; participants may reduce their vigilance during this period.

Overall, these results suggest the potential of saccadic distance to be used for objective SA assessment, but further analysis is required to draw a definitive conclusion. Future work must investigate the exceptions in the data patterns to better understand their cause. Gaze behavior may only partially capture SA, but incorporating more diverse ocular metrics (e.g., saccadic amplitude, fixation duration, areas of interest) may result in a more reliable SA assessment.

Additional metrics also allow for a more in-depth analysis of the implications on transparency. Distinguishing between SA Levels is difficult when relying on a singular ocular metric. Understanding the distribution of SA across the three levels allows for a more in-depth transparency analysis that can lead to more targeted design recommendations. Simple gaze-based metrics appear to easily capture SA Levels 1 and 2. Fixating on key interface elements is indicative of the participants' perception, and the frequency with which participants look around the screen may correspond to their overall comprehension [16]. However, the direct connection between Level 3 SA and gaze metrics remains unclear.

Future work must leverage either system logs or SA Probe questions to validate potential gaze metrics that correspond to specific interface elements at specific times. The presented underwater robotic system collects information about selected grasp positioning, grasp quality, and grasp execution time that will be analyzed to calculate more complex gaze-based metrics to develop an objective SA assessment method. Delineating the relationship between specific ocular metrics and their corresponding SA Levels will allow for a more detailed assessment. This level-based decomposition helps provide a more indepth and objective SA assessment for evaluating human-robot teaming transparency.

6 CONCLUSION

Transparency is fundamental for effective human-robot collaboration. Leveraging eye-tracking metrics such as saccadic distance and gaze locations can offer an objective assessment of SA. The preliminary analysis of these metrics from a marine robotic object manipulation system is presented. These metrics serve as crucial indicators of transparency, impacting the performance and trust within human-robot teams. Assessing SA objectively through ocular metrics advances the understanding of transparency, paving the way for improved collaboration between human-robot teams.

7 ACKNOWLEDGMENTS

REFERENCES

[1] M. R. Endsley. Toward a theory of situation awareness in dynamic systems. *Human Factors*, 37(1):32–64, 1995.

[2] M. R. Endsley, S. J. Selcon, T. D. Hardiman, and D. G. Croft. A comparative analysis of sagat and sart for evaluations of situation awareness. In *Human Factors and Ergonomics Society Annual Meeting*, volume 42, pages 82–86. SAGE Publications Sage CA: Los Angeles, CA, 1998.

[3] G. Hauland. Measuring individual and team situation awareness during planning tasks in training of en route air traffic control. *The International Journal of Aviation Psychology*, 18(3):290–304, 2008.

[4] C. M. Humphrey. *Information abstraction visualization for human-robot interaction*. Vanderbilt University, 2009.

[5] L. H. Ikuma, C. Harvey, C. F. Taylor, and C. Handal. A guide for assessing control room operator performance using speed and accuracy, perceived workload, situation awareness, and eye tracking. *Journal of loss prevention in the process industries*, 32:454–465, 2014.

[6] N. Liang, J. Yang, D. Yu, K. O. Prakah-Asante, R. Curry, M. Blommer, R. Swaminathan, and B. J. Pitts. Using eye-tracking to investigate the effects of pre-takeover visual engagement on situation awareness during automated driving. *Accident Analysis & Prevention*, 157:106143, 2021.

[7] L. Paletta, A. Dini, C. Murko, S. Yahyanejad, and U. Augsdörfer. Estimation of situation awareness score and performance using eye and head gaze for human-robot collaboration. In *ACM Symposium on Eye Tracking Research & Applications*, pages 1–3, 2019.

[8] L. Paletta, A. Dini, C. Murko, S. Yahyanejad, M. Schwarz, G. Lodron, S. Ladstätter, G. Paar, and B. J. Velik. Towards real-time probabilistic evaluation of situation awareness from human gaze in human-robot interaction. In *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, pages 247–248, 2017.

[9] S. Peißl, C. D. Wickens, and R. Baruah. Eye-tracking measures in aviation: A selective literature review. *The International Journal of Aerospace Psychology*, 28(3-4):98–112, 2018.

[10] N. G. Pusalkar, M.-R. Giolando, and J. A. Adams. Decision support system for autonomous underwater robot grasping. In *Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*, pages 198–202, 2023.

[11] R. M. Ratwani, J. M. McCurry, and J. G. Trafton. Single operator, multiple robots: an eye movement based theoretic model of operator situation awareness. In *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 235–242. IEEE, 2010.

[12] K. A. Roundtree, J. R. Cody, J. Leaf, H. O. Demirel, and J. A. Adams. Transparency's influence on human-collective interactions. *ACM Transactions on Human-Robot Interaction (THRI)*, 11(2):1–48, 2022.

[13] K. A. Roundtree, M. A. Goodrich, and J. A. Adams. Transparency: Transitioning from human–machine systems to human-swarm systems. *Journal of Cognitive Engineering and Decision Making*, 13(3):171–195, 2019.

[14] S. J. Selcon, R. M. Taylor, and E. Koritsas. Workload or situational awareness?: Tlx vs. sart for aerospace systems design evaluation. In *Human Factors Society Annual Meeting*, volume 35, pages 62–66. SAGE Publications Sage CA: Los Angeles, CA, 1991.

[15] A. Shaevitz, M. L. Johnston, and J. R. Davidson. Design, characterization, and modeling of barometric tactile sensors for underwater applications. In *2023 IEEE*

International Conference on Soft Robotics (RoboSoft), pages 1–6. IEEE, April 2023.

[16] T. Zhang, J. Yang, N. Liang, B. J. Pitts, K. Prakah-Asante, R. Curry, B. Duerstock, J. P. Wachs, and D. Yu. Physiological measurements of situation awareness: a systematic review. *Human factors*, 65(5):737–758, 2023.

[17] F. Zhou, X. J. Yang, and J. C. De Winter. Using eye-tracking data to predict situation awareness in real time during takeover transitions in conditionally automated driving. *IEEE Transactions on Intelligent Transportation Systems*, 23(3):2284–2295, 2021.