Human Factors in Design, Engineering, and Computing, Vol. 238, 2026, 1273–1284
https://doi.org/10.54941/ahfe1008356

Is Situation Awareness Related to Pupillary Changes During a Flight Navigation Simulation?

Shannon P. D. McGarry[1], Allison M. Bayro[2,3], Michael Kukenberger[4,5], Heejin Jeong[2], and Joseph T. Coyne[1]

[1]U.S. Naval Research Laboratory, Washington, DC 20375, USA
[2]Arizona State University, Tempe, Arizona 85281, USA
[3]Strategic Analysis, Inc., Arlington, Virginia 22203, USA
[4]Naval Aerospace Medical Institute, Pensacola, Florida 32508, USA
[5]University of New Hampshire, Durham, New Hampshire 03824, USA

ABSTRACT

A lack of situation awareness (SA) is often cited as contributing to Naval aviation mishaps. Yet, SA is not measured in real-time as doing so is too intrusive. Previous work suggests eye tracking metrics may be able to serve as a non-intrusive measure of SA. If so, it is critical to study how these metrics trend over time and their relation to SA. The present work begins to investigate this with nonlinear, longitudinal models of pupil size of Naval aviation trainees who completed a flight simulation in virtual reality. Specifically, pupil size during the four seconds before and after approaching the assigned waypoint was modeled via a generalized additive mixed model (GAMM). It found that pupil size differed per the waypoint and the person's overall SA. This preliminary investigation suggests pupil size changes over time may inform a real-time measure of SA in Naval aviation environments and the like.

Keywords: Situation awareness (SA), Pupillometry, Virtual reality (VR), Generalized additive mixed modeling (GAMM)

INTRODUCTION

In a review of Naval aviation mishaps occurring from 2011-2016, Miranda (2018) found the leading preconditions to human errors were related to situation awareness (SA). And yet, even after nearly three decades of research, Naval aviation is not measuring SA in situ. A large reason as to why is because the best SA measures intrude upon the human's current task load (Moens et al., 2026). Several works have explored psychophysiology to address this (see review in Biswas et al., 2024), including pupillometry (Zhang et al., 2023). Recent cognitive neuroscience research suggests pupillary dynamics may be informative of attentional reorienting, multitasking, and even team navigation performance in virtual reality (VR) (Lapborisuth et al., 2023; Qin et al., 2024). In this work, we explore how pupil size fluctuations during a VR flight simulation differ over time, per the individual, and/or across

Received October 5, 2026; Revised November 4, 2026; Accepted November 10, 2026; Available online December 1, 2026

© 2026 The Authors. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License.
For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

SA performance. These results, although preliminary, inform how real-time measures of SA should be theorized and developed.

BACKGROUND

SA is formally defined as the perception of objects in time and space, the comprehension of their meaning, and the accurate projection of future states (Endsley, 1995b). The most robust measure of SA is the Situation Awareness Global Assessment Technique (SAGAT), although it is not without flaws. For example, it requires generating queries that are not generalizable or validated beyond the task at hand. Then, when answering the queries, the participant is interrupted from completing tasks and not allowed to seek information from the environment. Not only is this not realistic, it means that SA is only measured in discrete intervals, i.e., after each round of queries (Moens et al., 2026). In total, SAGAT is not compatible with high stake, dynamic environments such as Naval aviation. de Winter et al. (2019) suggests SA measures should match the temporal and embedded features of SA development, which psychophysiological measures naturally offer given their inherent time resolution, cognitive underpinnings, and relative unobtrusiveness (Zhang et al., 2023). Also, the way in which people develop SA may not be identical, so capturing intraindividual changes is likely more informative than comparison of averages between people or conditions. These two ideals could be addressed via statistical models of in-task, psychophysiological data.

Pupillometry can inform real-time cognitive processes, such as the amount of mental workload or fatigue, but it has not been particularly successful in indexing SA (Zhang et al., 2023). This could be because previous work tends to study it in aggregate, such as by trial or person (e.g., Wu et al., 2019). When considering the "high degree of covariation between changes in pupil size and [locus coeruleus] LC activation [i.e., the primary control center of norepinephrine in the brain]" (Strauch et al., 2022), tracking its real-time change may be informative of SA. Approaches such as growth curve modeling, detrended fluctuation analysis, and linear discriminant analysis have informed aspects of human behavior, performance, and cognitive state in a way average-based approaches could not (see reviews in Fink et al., 2024 and van Rij et al., 2019; Lapborisuth et al., 2023). Among those methods is generalized additive mixed modeling (GAMM), which uses thin plate spline regression to model the nonlinear effect while accounting for random effects (Sóskuthy, 2017). Previous work has implemented GAMMs to more accurately characterize the temporal fluctuations of pupil size, revealing differences in fatigued versus alert states, for example (Yin et al., 2026). They also have indicated temporal differences in pupil size across conditions (see Korkmaz & Aydın, 2026).

To the authors' knowledge, investigating SA measurement via GAMMs of pupil size is novel. Some similar previous work includes analyzing different psychophysiological measures [e.g., electroencephalogram (EEG) and functional near-infrared spectroscopy (fNIRS)] over event-based time intervals to explain differences in a behavior-derived measure of SA

(Festa et al., 2024). When research has studied SA with pupillometry in a VR environment, it finds pupil size and SA positively correlated (Luo et al., 2024). Previous research that has studied differences in EEG measures between those with low and high SA have found the former has lower alertness and cognitive arousal in sensory and motor systems (Feng et al., 2022; Kästle et al., 2021). Several works have predicted low or high SA from multimodal psychophysiology (e.g., Chen et al., 2024), but *when* these differences emerge and *how* they trend over time is less understood.

The Present Work

Of interest was to investigate how pupil size trends over time could indicate differences in SA. We explored this by modeling the pupil size of Naval aviation trainees who completed a flight navigation simulation in VR with GAMMs. Our work is some of the first to analyze this psychophysiological measure with this method and purpose. Nonetheless, we predict that those with higher SA will have larger pupils overall and as they navigate to waypoints, in line with what was observed in previous work (Feng et al., 2022; Kästle et al., 2021; Luo et al., 2024).

METHOD

All of the experimental details and some of the participants in Bayro et al. (2026) are in the present work. The aforementioned assessed the utility, usability, workload, and valence of the VR scenario used herein. The current work differs in that it analyzes the real-time fluctuations in pupil size when participants are navigating in the scenario.

Participants

This research was approved by Institutional Review Board at the U.S. Naval Research Laboratory. Participants were Naval aviation trainees and had to be at least 18 years old and have normal (20/20) or corrected-to-normal vision (e.g., via glasses or contact lenses). In total, there were 151 participants in the study (age: M = 23.8 years, SD = 4.2 years; 90.4% male; 61.6% officer versus 38.4% enlisted). Twenty-one participants were removed as the psychophysiological data was not correctly synchronized and another four were removed due to simulator sickness. Another eight were removed for insufficient eye tracking data quality (i.e., more than 50% of their eye tracking data was missing). Finally, six participants did not answer all SAGAT queries, making the final sample size N = 112.

Equipment

Participants completed tasks in both standard desktop (24" ViewSonic monitor, standard keyboard and mouse) and VR-enabled workstations. The tasks in VR were completed with a Varjo XR-3 headset, which provides "human-like eye resolution" and built-in, high-fidelity head, gaze, and pupil tracking (f_s = 200 Hz, reported sub-degree accuracy; Varjo, n.d.), and a

Logitech Saitek X52 H.O.T.A.S (Logitech, n.d.). Concurrently, a Shimmer 3 GSR+ recorded galvanic skin response and photoplethysmography signals (Shimmer, n.d.), but its findings are omitted as they are not pertinent to the present research goals.

Experimental Design

SA was assessed in our custom VR flight navigation task called *ASANA*, which is thoroughly detailed in Bayro et al. (2026). The main task in *ASANA* is to pilot an aircraft to assigned waypoints that are presented on a map as a red dot. When a waypoint was successfully flown through, the map rotated to one of eight orientations (e.g., 0°, 45°, 90°, 135°, etc.) and updated to show the next assigned waypoint (as seen in Figure 1). Participants had to rely on their spatial ability to navigate to waypoints, specifically by using terrain association and internal spatial representations.



Figure 1: Progression of flying towards and then through a waypoint with the map rotating thereafter.

Additionally, when it was raining and thundering in the environment, participants were instructed to complete the "storm procedure", meaning they deviated from navigating to the assigned waypoint and navigated to a specific, predetermined location and then pressed a designated button on the H.O.T.A.S. Participants were incentivized to visit all waypoints and complete the storm procedure as quickly and accurately as possible to maximize their final score. As described by Endsley (2000), the SAGAT queries were developed per the results of a goal-directed task analysis and were iteratively refined by the first and second authors. Each query assessed the person's perception (e.g., "What is directly under you, currently?"), comprehension (e.g., "What action series leads to the most points?"), or projection of the environment (e.g., "Which of these is most likely to occur next?").

Procedure

Written informed consent was obtained before participants began one of two workflows (order was counterbalanced across participants): either a series of cognitive assessments and surveys during the "Desktop Workflow" or the "VR Workflow", which included collecting baseline psychophysiological data, completing training and full *ASANA* sessions in VR, and answering debriefing surveys. Participants read and were quizzed on *ASANA* instructions before completing a practice session where they navigated to four waypoints, completed one storm procedure, and answered three SAGAT queries. They could ask the research administrator questions during this time. After

the VR headset was removed, participants rested for at least three minutes or until they indicated they were ready to resume. Participants then completed the full *ASANA* session, during which eye tracking data were recorded (where eye tracking calibration was completed prior). After flying to the 12 waypoints, completing one storm procedure, and answering 15 SAGAT queries, participants completed three surveys and were debriefed from the experiment. Bayro et al. (2026) has more details about this procedure and its accompanying results.

RESULTS

Data Processing and Analysis

All data processing and analyses were conducted in RStudio (version 2026.01.1; Posit Team, 2026) with R version 4.5.2 (R Core Team, 2026). First, the eye tracking data was down sampled from 150 Hz to 60 Hz then linearly interpolated using pupillometryR package (e.g., downsample(), interpolate(), etc.). Pupil size was standardized and winsorized at $\pm$ 3.3 standard deviations within each individual in order to more accurately compare pupil size fluctuations across and within participants. To make the data more uniformly representative of what participants were experiencing in the environment and because it took approximately four seconds to fly to a waypoint once it was visible, the data was then filtered to be the four seconds before and after crossing a waypoint.

In the GAMMs, the four seconds before and after crossing a waypoint (i.e., Time), was modeled as a smooth term and an interaction effect with waypoint. In other words, pupil size fluctuations over time were not constrained to be linear and could significantly vary per the waypoint. Participant was estimated as a random effect to account for individual differences in overall pupil size. GAMMs were built via the mgcv package, specifically with the function bam(). The function gam.check() was used to assess the assigned k value, (i.e., the number of effective degrees of freedom (edf) that could be estimated for a given smooth term), and models were adjusted accordingly (e.g., when |edf-k'| < 4, where k'=k-1). When models estimated autocorrelation of the residuals, it was incorporated into the next formulation. Models were tuned ad-hoc and compared against each other via analysis of deviance. For a furhter introduction and explanation of GAMMs, see Sóskuthy (2017) and van Rij et al. (2019).

Baseline Model: Pupil Fluctuations When Flying to Assigned Waypoints

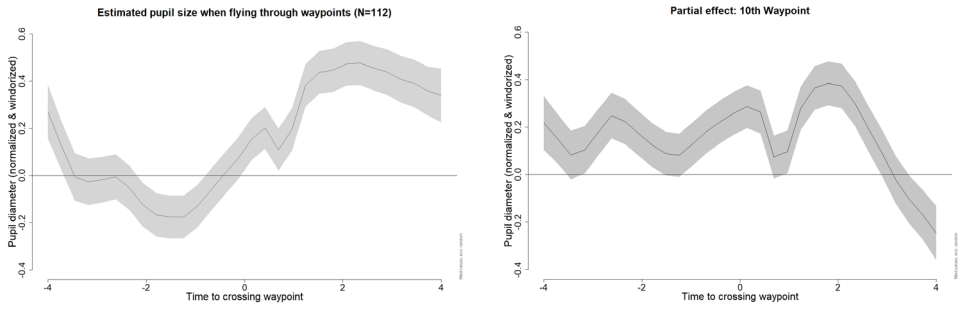
Pupil size was first modeled to assess its general fluctuation and variability and the pertinent results are in Table 1. Firstly, it shows pupil size was estimated to be larger on average when approaching and leaving waypoints than during other times in *ASANA* (e.g., time between waypoints, during the storm procedure, etc.). It also estimated pupil size to fluctuate significantly during this time (see s(TimeToWaypoint) in Table 1). Except for the 4th and

6th waypoints, the model estimated pupil size fluctuations during all the waypoints to be significantly different from what was estimated on average (see the rows under s(TimeToWaypoint) in Table 1). As seen in the final row of Table 1, participant was a significant random effect meaning pupil size was significantly different across participants when navigating to and from assigned waypoints. The model estimated 27.7% of the total variance (Adjusted $R^2 = 0.277$). Figure 2 shows the estimated pupil size given all waypoints versus the 10th waypoint, where it was estimated to significantly deviate.

Table 1: Output from baseline GAMM that was significant.

	b	*SE*	*t*	*p*
Intercept	0.158	0.019	8.51	<.0001****
	edf	**Ref. df**	***F***	***p***
Smooth Term				
s(TimeToWaypoint)	41.170	46.715	20.089	<.0001****
s(TimeToWaypoint):2ndWaypoint	11.508	15.963	2.264	0.0028**
s(TimeToWaypoint):3rdWaypoint	1.003	1.005	5.966	0.0144*
s(TimeToWaypoint):5thWaypoint	4.872	6.729	2.650	0.0108*
s(TimeToWaypoint):7thWaypoint	5.330	7.387	4.553	<.0001****
s(TimeToWaypoint):8thWaypoint	4.254	5.840	3.194	0.0045**
s(TimeToWaypoint):9thWaypoint	6.261	8.717	2.446	0.0118*
s(TimeToWaypoint):10thWaypoint	12.991	17.916	3.116	<.0001****
s(TimeToWaypoint):11thWaypoint	14.129	19.380	4.511	<.0001****
s(TimeToWaypoint):12thWaypoint	21.623	28.171	6.136	<.0001****
s(Participant)	84.219	111.000	3.145	<.0001****

Significance codes: <.0001 '****' <0.001 '***' <0.01 '**' <0.05 '*'



(a) **(b)**

Figure 2: Estimated pupil size given (a) all waypoints versus (b) the 10th waypoint, which was estimated to significantly differ.

SA Model: Pupil Fluctuations of Different SA Groups

Next, SA was added as an interaction effect to the baseline model to determine whether pupil size fluctuations differed by SA group. SA group

was determined by Kelley's 27% rule (Kelley, 1939), meaning participants were split into three groups depending on their SAGAT accuracy score. The Low (n = 23, M = 43.4%, SD = 13.4%) and High SA group (n = 14, M = 89.5%, SD = 3.3%) included the bottom and top 27% of scores, respectively, whereas the Middle SA group included the remaining middle 46% (n = 75, M = 69.8%, SD = 7.5%). Incorporating SA into the model as an interaction effect led to a significantly better model fit ($F(414.59, 99{,}099) = 45.309$, $p < .0001$) as it estimated about 10% more variance (Adjusted $\mathbf{R^2} = 0.383$). Table 2 shows pertinent results of this model.

Table 2: Output from GAMM that incorporates SA that was significant.

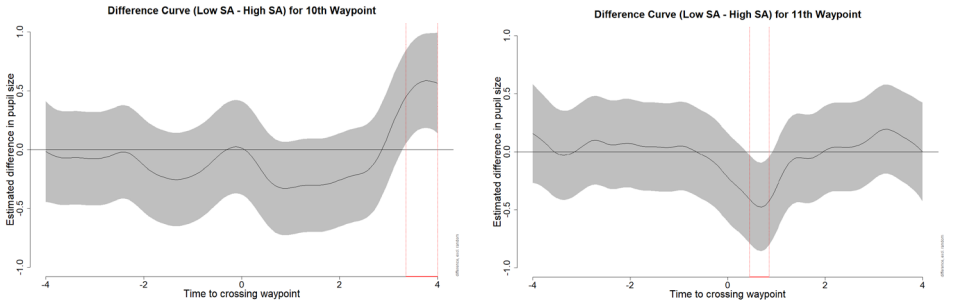
	b	*SE*	*t*	*p*
Intercept	0.361	0.077	4.684	<.0001****
LowSA.8thWaypoint	−0.635	0.097	−6.528	<.0001****
HighSA.8thWaypoint	−0.490	0.126	−3.903	<.0001****
LowSA.9thWaypoint	−0.232	0.097	−2.384	0.0171*
HighSA.9thWaypoint	−0.321	0.126	−2.563	0.0104*
LowSA.10thWaypoint	−0.255	0.097	−2.623	0.0087**
LowSA.11thWaypoint	−0.810	0.097	−8.321	<.0001****
HighSA.11thWaypoint	−0.789	0.126	−6.284	<.0001****
Smooth Term	**edf**	**Ref. df**	***F***	***p***
s(TimeToWaypoint):LowSA.3rdWaypoint	6.827	9.457	6.162	<.0001****
s(TimeToWaypoint):HighSA.3rdWaypoint	8.850	12.212	3.919	<.0001****
s(TimeToWaypoint):LowSA.4thWaypoint	13.162	17.683	2.936	<.0001****
s(TimeToWaypoint):HighSA.4thWaypoint	4.865	6.698	2.263	0.0272*
s(TimeToWaypoint):LowSA.5thWaypoint	13.500	14.890	3.896	<.0001****
s(TimeToWaypoint):HighSA.5thWaypoint	6.467	8.956	3.144	0.0008***
s(TimeToWaypoint):LowSA.6thWaypoint	7.659	10.602	4.711	<.0001****
s(TimeToWaypoint):HighSA.6thWaypoint	11.769	15.985	2.874	<.0001****
s(TimeToWaypoint):LowSA.7thWaypoint	4.288	5.877	6.660	<.0001****
s(TimeToWaypoint):HighSA.7thWaypoint	4.399	6.035	4.769	<.0001****
s(TimeToWaypoint):LowSA.8thWaypoint	15.835	20.724	3.105	<.0001****
s(TimeToWaypoint):HighSA.8thWaypoint	7.010	9.710	2.482	0.0081**
s(TimeToWaypoint):LowSA.9thWaypoint	17.646	22.596	3.213	<.0001****
s(TimeToWaypoint):HighSA.9thWaypoint	5.178	7.143	2.725	0.0077**
s(TimeToWaypoint):LowSA.10thWaypoint	13.015	17.507	1.640	0.0479*
s(TimeToWaypoint):HighSA.10thWaypoint	11.576	15.743	2.996	<.0001****
s(TimeToWaypoint):LowSA.11thWaypoint	16.381	21.306	4.878	<.0001****
s(TimeToWaypoint):HighSA.11thWaypoint	5.816	8.045	6.067	<.0001****
s(TimeToWaypoint):LowSA.12thWaypoint	10.681	14.609	14.762	<.0001****
s(TimeToWaypoint):HighSA.12thWaypoint	6.724	9.313	15.040	<.0001****
s(Participant)	82.798	109.000	3.160	<.0001****

Significance codes: <.0001 '****' <0.001 '***' <0.01 '**' <0.05 '*'

To contextualize, Table 3 details the timing and magnitude of significant pupil size differences between the Low and High SA groups. In all but one instance, the significant differences were found after the waypoint was flown through (i.e., time to crossing waypoint was greater than zero) and in four out of seven instances, pupil size was estimated to be significantly smaller for the Low versus the High SA group. Figure 3 depicts an example of these differences graphically, as it shows the magnitude and timing of the estimated difference between the Low and High SA group for the 10th and 11th waypoint.

Table 3: Significant pupil size differences between Low and High SA groups.

Waypoint	Time of Significant Differences Between Low SA and High SA (s)	Magnitude of Difference (Low SA – High SA)
2nd	[–1.899 - –0.848]	[–0.656 - –0.002]
4th	[3.838 - 4.000]	[0.056 - 0.943]
6th	[3.919 - 4.000]	[0.135 - 0.989]
8th	[0.444 - 1.091]	[–0.880 - –0.019]
10th	[3.353 - 4.000]	[0.052 - 0.994]
11th	[0.444 - 0.848]	[–0.858 - –0.027]
12th	[0.283 - 0.768]	[–0.810 - –0.011]



(a) **(b)**

Figure 3: Plot of the estimated pupil size differences between low and high SA while flying to the (a) 10th and (b) 11th waypoint (red lines indicate when difference was significant).

DISCUSSION

The goal of this work was to begin to investigate how pupil size changes over time were informative of differences in SA. Even though there were some informative differences, the present results indicate that pupil size fluctuations were more dependent on what waypoint was being flown to, with greater differences being observed later in the task. These findings suggest differentiating between low and high SA with pupil data requires understanding the content and timing of a person's current task, both of which are important to understanding SA in real-life contexts.

The baseline model indicated that pupil size fluctuated rather constantly when participants navigated to waypoints. However, the way in which it fluctuated was not universal, given this model's significant interaction effect between time and waypoint. Specifically, almost all of the waypoints had more fluctuations than what was estimated on average and the largest deviations were from the last three waypoints in the session. Relatedly, the significant random effect of participant suggested pupil size differed for each participant overall. In other words, a large pupil size for one person (e.g., greater than 1 standard deviation from their mean) could be another's average pupil size. Collectively, it seems time on task and individual differences impact pupil size fluctuations during an immersive, multi-featured spatial navigation task, meaning a universal, homogenous model to determine SA would most likely not be accurate.

Relatedly, adding each participant's SA accounted for approximately 10% more variance, suggesting pupil size and its changes over time are related to SA. Other work has not always observed this (Zhang et al., 2023, p. 744), but as noted before, pupil size as it relates to SA has not been analyzed with nonlinear nor longitudinal models. Future work should continue to explore pupil size with statistical models to better understand why such a discrepancy exists, as it could better detail how informative pupil size is of SA when in operational environments. Except for the 2nd waypoint, significant differences in pupil size between SA groups were found after the waypoint was flown through (i.e., time to crossing waypoint was greater than zero, see 2nd column in Table 3). In other words, those with different SA recruited and deployed cognitive resource differently after flying through the assigned waypoint. We speculate that this corresponds to when people were figuring out the location of the newly assigned waypoint. Overall, it suggests real-time, physiological measures of SA will likely be more informative during periods of active cognitive tasking.

Pupil size was inconsistently different between those with low and high SA. It was estimated to be significantly larger for those with higher SA in four of seven instances, but significantly smaller for the remaining three (review 3rd column in Table 3). Therefore, our a priori prediction, (i.e., those with more SA will have larger pupils), was only partially supported. One possible explanation for these varied findings is that those of higher SA recruited and deployed more resources more immediately. Notably, those of higher SA had larger pupil size right after crossing the waypoint (i.e., all < 1.1 s), whereas instances of significant smaller pupil size for this group always happened later on (i.e., all > 3 s). Therefore, successfully completing *ASANA* as a whole, i.e., flying to assigned waypoints and answering SA queries as quickly and accurately as possible, requires recruiting and deploying more cognitive resources as assigned waypoints are reached. This pattern also suggests that SA differences are not going to physiologically manifest the same way over the course of a navigation task, further emphasizing the need to account for this and the individual completing the task.

CONCLUSION

GAMMs of in-task pupil size were informative of differences in SA, however nuances of what that means when trying to detect it online still persist. These models also suggest that patterns of pupil size fluctuations significantly change over time and are dependent on the individual. Consistent with de Winter et al. (2019), this work is an important first step toward modeling and measuring SA as a dynamic cognitive process. However, future work needs to incorporate the specific context from the environment if SA is to be measured in situ by Naval aviation.

LIMITATIONS AND FUTURE WORK

The error bands for partial effects/difference curve were large, suggesting unaccounted variability in pupil size differences between those of low and high SA exists. These models did not account for where a person was looking, which may impact pupil size measurement as well as interpretation, as that information may account for more variance above and beyond what was observed with these models. Also, this work did not further explore how map rotation itself could impact pupil size, which may be legitimate as different map rotations have found to be more challenging than others (Wickens et al., 2005), and task difficulty impacts pupillary responses. Our future analyses could add in the eye's location and/or map rotation as a predictor and/or covariance to future GAMM formulations to understand if they impact the pupillary trends that occur when a waypoint is assigned. Additionally, future work should explore how pupil and other psychophysiological measures could be modeled to predict SA-based errors, such as attentional tunneling, errors of omission, or over-reliance on default values (Endsley, 1995a) and not SAGAT accuracy. This would improve the predictive and external validity of the current findings.

ACKNOWLEDGMENT

This work received funding support from the Office of Naval Research through the U.S. Naval Research Laboratory (NRL) Base Program. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of NRL or the U.S. Government. This work was also supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. 026257-001 and the Link Foundation Fellowship.

REFERENCES

Biswas, A., Gupta, P., Khurana, S., Held, D., & Admoni, H. (2024, October 22). Modeling drivers' situational awareness from eye gaze for driving assistance. *8th Conference on Robot Learning (CoRL 2025)*. https://HARPLab.github.io//DriverSA

Bayro, A., McGarry, S. P. D., NeSmith, R., Coyne, J. T., & Jeong, H. (2026). From design to user experience: The creation and assessment of ASANA, an immersive VR task for situation awareness and spatial ability. *Computers and Graphics*, *135*. https://doi.org/10.1016/j.cag.2026.104541

Chen, J., Chen, A., Jiang, B., & Zhang, X. (2024). Physiological records-based situation awareness evaluation under aviation context: A comparative analysis. *Heliyon*, *10*(5). https://doi.org/10.1016/j.heliyon.2024.e26409

de Winter, J. C. F., Eisma, Y. B., Cabrall, C. D. D., Hancock, P. A., & Stanton, N. A. (2019). Situation awareness based on eye movements in relation to the task environment. *Cognition, Technology and Work*, *21*(1), 99–111. https://doi.org/10.1007/s10111-018-0527-6

Endsley, M. R. (2000). Direct measurement of situation awareness: Validity and use of SAGAT. In M. R. Endsley & D. J. Garland (Eds.), *Situation awareness analysis and measurement* (pp. 1–21). Lawrence Erlbaum Associates.

Endsley, M. R. (1995a). A taxonomy of situation awareness errors. In R. Fuller, N. Johnston, & N. McDonald (Eds.), *Human Factors in Aviation Operations* (pp. 287–292). Ashgate Publishing Ltd. https://www.researchgate.net/publication/285731357

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

Feng, C., Liu, S., Wanyan, X., Chen, H., Min, Y., & Ma, Y. (2022). EEG feature analysis related to situation awareness assessment and discrimination. *Aerospace*, *9*(10). https://doi.org/10.3390/aerospace9100546

Festa, E. K., Bracken, B. K., Desrochers, P. C., Winder, A. T., Strong, P. K., & Endsley, M. R. (2024). EEG and fNIRS are associated with situation awareness (hazard) prediction during a driving task. *Ergonomics*, *67*(12), 1993–2008. https://doi.org/10.1080/00140139.2024.2367163

Fink, L., Simola, J., Tavano, A., Lange, E., Wallot, S., & Laeng, B. (2024). From pre-processing to advanced dynamic modeling of pupil data. *Behavior Research Methods*, *56*(3), 1376–1412. https://doi.org/10.3758/s13428-023-02098-1

Kästle, J. L., Anvari, B., Krol, J., & Wurdemann, H. A. (2021). Correlation between situational awareness and EEG signals. *Neurocomputing*, *432*, 70–79. https://doi.org/10.1016/j.neucom.2020.12.026

Kelley, T. L. (1939). The selection of upper and lower groups for the validation of test items. *Journal of Educational Psychology*, *30*(1), 17–24. https://doi.org/10.1037/h0057123

Korkmaz, H. A., & Aydın, Ö. (2026). The effect of orthographic neighbourhood and semantics on lexical processing in a transparent orthographic language: A pupilometry study. *Journal of Psycholinguistic Research*, *55*(4). https://doi.org/10.1007/s10936-026-10267-4

Lapborisuth, P., Koorathota, S., & Sajda, P. (2023). Pupil-linked arousal modulates network-level EEG signatures of attention reorienting during immersive multitasking. *Journal of Neural Engineering*, *20*(4). https://doi.org/10.1088/1741-2552/acf1cb

Logitech G. (n.d.). *X52 H.O.T.A.S.* https://www.logitechg.com/en-us/shop/p/x52-space-flight-simulator-controller

Luo, Y., Yang, Q., Seo, J., & Ahn, S. (2024). Theoretical framework for utilizing eye-tracking data to understand the cognitive mechanism of situational awareness in construction hazard recognition. *Journal of Management in Engineering*, *40*(4). https://doi.org/10.1061/jmenea.meeng-5905

Miranda, A. T. (2018). Understanding human error in naval aviation mishaps. *Human Factors*, *60*(6), 763–777. https://doi.org/10.1177/0018720818771904

Moens, L. L., Lydon, S., Cucurachi, S., O'Connor, P., Sauter, T. C., Töndury, G. A., & Manser, T. (2026). Measuring situation awareness: A meta-review across domains. *Human Factors*, *68*(5), 632–672. https://doi.org/10.1177/00187208251412110

Posit Team. (2026). *RStudio: Integrated development environment for R* (Version 2026.01.1) [Computer software]. Posit Software, PBC. https://posit.co/

Qin, Y., Zhang, W., Lee, R., Sun, X., & Sajda, P. (2024). Pupil-linked arousal is predictive of team performance in a virtual reality (VR) sensory-motor task. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*. https://doi.org/10.1109/EMBC53108.2024.10782124

R Core Team. (2026). *R: A language and environment for statistical computing* (Version 4.5.2) [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Strauch, C., Wang, C. A., Einhäuser, W., van der Stigchel, S., & Naber, M. (2022). Pupillometry as an integrated readout of distinct attentional networks. *Trends in Neurosciences*, *45*(8), 635–647. https://doi.org/10.1016/j.tins.2022.05.003

Shimmer. (n.d.). Shimmer3 GSR+ Unit. https://www.shimmersensing.com/product-tag/shimmer3-gsr-unit/

Sóskuthy, M. (2017). Generalised additive mixed models for dynamic analysis in linguistics: A practical introduction. In *arXiv preprint* (Issue arXiv:1703.05339). http://arxiv.org/abs/1703.05339

van Rij, J., Hendriks, P., van Rijn, H., Baayen, R. H., & Wood, S. N. (2019). Analyzing the time course of pupillometric data. *Trends in Hearing*, *23*. https://doi.org/10.1177/2331216519832483

Varjo. (n.d.). Varjo XR-3. https://varjo.com/products/varjo-xr-3

Wickens, C. D., Vincow, M., & Yeh, M. (2005). Design applications of visual spatial thinking: The importance of frame of reference. In P. Shah & A. Miyake (Eds.), *The Cambridge handbook of visuospatial thinking* (pp. 383–425). Cambridge University Press.

Wu, S., Guo, J., Zeng, S., & Lu, Z. (2019). Direct measurement of situation awareness in abnormal situation during aircraft landing. *2019 International Conference on Quality, Reliability, Risk, Maintenance, and Safety Engineering (QR2MSE 2019)*.

Yin, Z., Pan, W., Wang, A., Fu, Q., Yang, Z., & Ni, S. (2026). The evolution of fatigue in remote tower controllers: Evidence from eye-tracking analysis. *Scientific Reports*, *16*(1). https://doi.org/10.1038/s41598-026-42161-3

Zhang, T., Yang, J., Pitts, B. J., Prakah-Asante, K. O., Curry, R., Duerstock, B. S., Wachs, J. P., & Yu, D. (2023). Physiological measurements of situation awareness: A systematic review. *Human Factors*, *65*(5), 737–758. https://doi.org/10.1177/0018720820969071