Human Factors, Business Management and Society, Vol. 215, 2026, 110–120
https://doi.org/10.54941/ahfe1007592

Cognitive Mechanisms in Drone Delivery Interaction Interfaces: An Eye-Tracking Study

Yutong Zhang, Yuxin Wu, Chaoyi Zeng, Chunyun Tan, Jiale Zhuang, and Jialin Cai

Department of Industrial Design, Huaqiao University, Xiamen, China

ABSTRACT

With the growing adoption of drone delivery, ground control station interfaces have become critical to operator performance and safety. This study examined how two interface pop-up types, immediate-triggering and continuous-display, affect operators' cognitive processing, and further explored the role of training. A 2 (training: trained vs. untrained) × 2 (pop-up type) × 4 (task difficulty) mixed-design experiment was conducted. Forty participants were recruited, and 32 valid samples were retained. Behavioral performance, subjective cognitive load, and eye-tracking measures were analyzed. The results showed that task difficulty was the primary factor affecting operational performance, and the practice effect weakened under high-difficulty tasks. Although overall behavioral performance was similar across interface conditions, eye-tracking measures revealed clear differences in cognitive processing. Compared with continuous-display, immediate-triggering supported faster orientation to key information. These findings provide empirical evidence for optimizing drone delivery interfaces from the perspective of situation awareness and cognitive load regulation.

Keywords: Drone human-machine interaction, Cognitive load, Eye tracking, Information presentation, Training effect

INTRODUCTION

With the increasing use of UAVs in delivery and urban operations, human–machine interfaces at ground control stations are critical to safety and efficiency. Despite high autonomy in flight control, key decisions such as task management and anomaly handling still depend on human operators. In complex environments, operators must process multi-source information simultaneously, leading to cognitive overload and attentional demands, which constrain large-scale UAV deployment (Tezza and Andujar, 2019).

Existing studies have mainly focused on performance outcomes, with limited attention to underlying cognitive processes. In particular, there is a lack of systematic evidence on how interface pop-up type, training, and task difficulty jointly affect attention allocation and visual search.

This study addresses this gap using a mixed experimental design with behavioral, subjective, and eye-tracking measures. It examines the effects of pop-up type, training, and task difficulty on cognitive processing, and explores the potential moderating role of gender.

Received March 30, 2026; Revised April 26, 2026; Accepted May 11, 2026; Available online July 20, 2026

THEORETICAL FOUNDATIONS AND LITERATURE REVIEW

Modern UAV operation has evolved from traditional flight control into complex monitoring and decision-making tasks. In dynamic delivery scenarios, operators must integrate multi-source information under time pressure, which increases cognitive load and may impair situation awareness and error management (Endsley, 1995). This challenge is further intensified in one-to-many control contexts, where frequent attention shifts across multiple UAVs place additional demands on cognitive resources.

To assess such demands, both subjective and objective measures have been widely adopted. NASA-TLX is commonly used because of its practicality (Hart, 2006), whereas physiological indicators such as pupil diameter and eye tracking can provide more continuous evidence of cognitive resource allocation (Di Stasi et al., 2016a; Orlandi and Brooks, 2018). Prior studies have also shown that interface design can significantly influence cognitive workload and visual processing (Zhang, Liu and Kaber, 2024). At the same time, training may help operators build more complete mental models and reshape visual attention patterns, while interface strategies such as layout and pop-up design may reduce extraneous cognitive load and improve task performance (Politowicz, 2024; Chen Heyuan et al., 2025).

Despite these advances, existing research still has several limitations. First, most studies focus on single interface factors or overall performance outcomes, with limited attention to how pop-up type, training, and task difficulty jointly influence cognitive processing. Second, although training has been shown to alter visual attention, its effectiveness under high task difficulty remains unclear (Politowicz, 2024; Wiggins, 2021). Third, individual differences, especially gender-related differences in visual attention, remain underexplored in UAV human-machine interaction, even though prior studies suggest that users may respond differently to the same interface design (Djamasbi et al., 2010; Liu Tingwei et al., 2025).

To address these gaps, the present study uses eye-tracking measures within a mixed experimental design to examine the combined effects of pop-up type, training, and task difficulty on cognitive processing, while also exploring the potential moderating role of gender.

EXPERIMENTAL DESIGN AND IMPLEMENTATION PROCEDURE

This study used a mixed design to examine how training, pop-up type, and task difficulty affect behavioral performance and cognitive processing. Based on cognitive load theory, we expected training effects to weaken under high task difficulty, pop-up types to shape visual search and attention differently, and gender to moderate these effects.

The DJI flight simulator was used as the experimental interface, with Python scripts implementing pop-up triggering logic. Training consisted of a 3-minute instructional video and a 5-minute tutorial. The apparatus included a laptop with an NVIDIA GeForce RTX 4060 GPU, a DJI UAV remote controller, a Tobii Pro X3-120 eye tracker, and a timer.

A total of 40 undergraduate participants were recruited, all with less than 2 hours of UAV experience and normal vision. Based on a G*Power analysis, at least 32 participants were required (Faul et al., 2007). After data screening, 32 valid samples were retained.

The experiment consisted of a training phase and a testing phase. During testing, participants completed four task units twice. Eye-tracker calibration and basic preparation were conducted prior to the experiment to ensure data quality. All procedures followed standardized protocols to ensure experimental consistency and reproducibility.

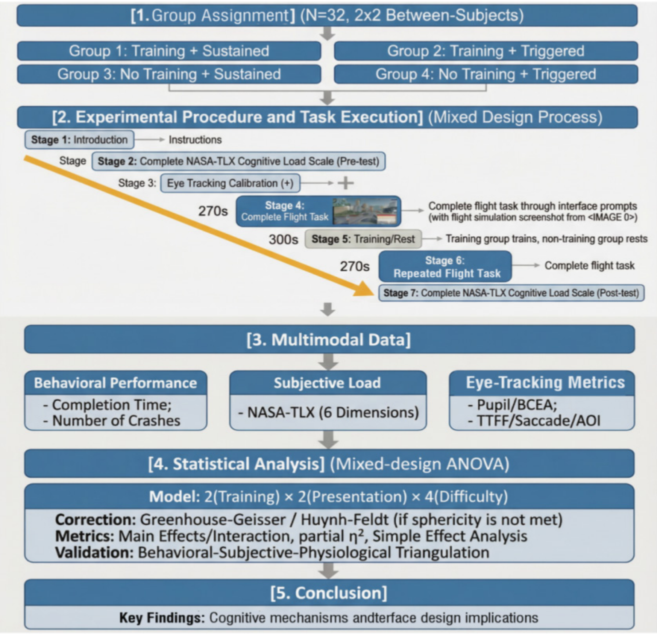


Figure 1: Framework of the experimental procedure

ANALYSIS OF BEHAVIORAL PERFORMANCE AND SUBJECTIVE COGNITIVE LOAD RESULTS

A 2 (training: trained vs. untrained) × 2 (pop-up type: immediate-triggering vs. continuous-display) × 4 (task difficulty) mixed design was used, with training and pop-up type as between-subject factors and task difficulty as a within-subject factor. Task completion times were averaged across two rounds for analysis.

For subjective cognitive load (TLX difference), no significant main effects or interaction were found (all $p > .05$). Similarly, no significant effects were observed for overall task efficiency based on mean completion time.

Task completion time increased significantly across task units (Unit 1: 86.44 ± 22.90 s; Unit 2: 134.06 ± 33.04 s; Unit 3: 189.59 ± 36.90 s). A multivariate test confirmed a significant main effect of task unit ($F = 1122.113$, $p < .001$, partial $\eta^2 = 0.987$), while the interaction between pop-up type and task unit was not significant ($F = 0.954$, $p = .397$, partial $\eta^2 = 0.062$).

Between-subject analyses for each task unit showed no significant effects of training, pop-up type, or their interaction (all $p > .05$). At the highest difficulty level (Unit 4), the model explained slightly more variance ($R^2 = 0.145$), but no effects reached significance.

Mauchly's test indicated that the assumption of sphericity was met ($p = .120$). Repeated-measures ANOVA showed a significant main effect of task unit ($F(2, 56) = 546.693, p < 0.001$), with no significant interactions involving training or pop-up type (all $p > .05$).

Overall, task difficulty significantly affected performance, while training and pop-up type showed no significant effects on behavioral outcomes, suggesting that behavioral performance was primarily determined by task difficulty, while the effects of interface design and training are more evident in cognitive processing indicators such as eye tracking.

Table 1: Descriptive statistics of task completion time across task units by training condition and pop-up type (M ± SD).

Descriptive Statistics

Task Unit	Untrained		Trained		Total
	Immediate	Continuous	Immediate	Continuous	Overall
Unit 1	82.25±18.11	82.50±19.07	86.88±21.63	94.13±32.50	86.44±22.90
Unit 2	131.25±19.98	134.75±39.39	133.38±32.78	136.88±42.31	134.06±33.04
Unit 3	183.50±21.85	183.00±40.55	193.75±40.26	198.13±45.76	189.59±36.90

Table 2: Tests of within-subjects effects in the three-factor mixed design.

Tests of Within-Subjects Effects

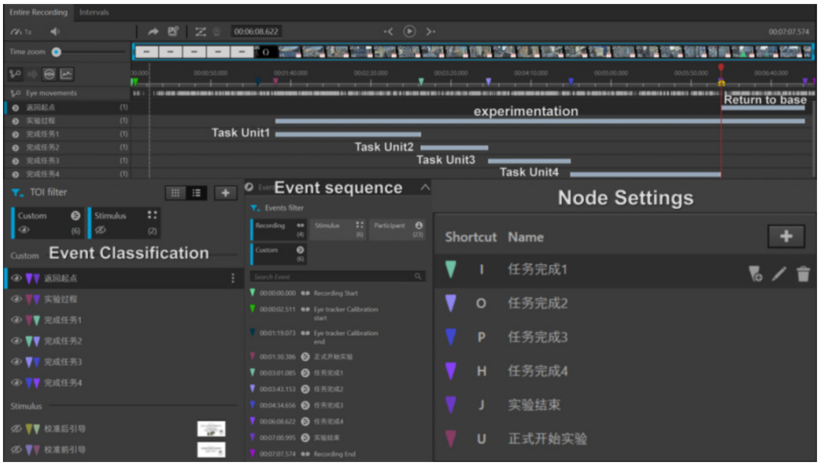
Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	p
Task Unit	Sphericity Assumed	170592.771	2	85296.385	546.693	0.000
Task Unit x Training Types	Sphericity Assumed	449.021	2	224.510	1.439	0.246
Task Unit x Pop-up Types	Sphericity Assumed	15.438	2	7.719	0.049	0.952
Task Unit x Training Types x Pop-up Types	Sphericity Assumed	51.521	2	25.760	0.165	0.848

ANALYSIS OF EYE-TRACKING RESULTS

To ensure the reliability of the analysis results, eye-tracking data in this experiment were collected using a Tobii Pro X3-120 eye tracker and cleaned in Tobii Pro Lab by defining Areas of Interest (AOIs) and Times of Interest (TOIs) to exclude invalid eye-movement data. In the AOI setting, four areas of interest were defined based on key task-related information regions. These included the main view area, which contained core flight parameters such as altitude, speed, and attitude; the primary map, which displayed the global route and obstacle information; the secondary map, which provided a locally enlarged view; and the order pop-up window, which dynamically presented operational instructions. In the TOI setting, five time segments of interest were defined based on the experimental procedure, covering the complete process from takeoff to landing, and these were further divided into four task stages.

Figure 2: Example of AOI configuration.



Shortcut	Name
I	任务完成1
O	任务完成2
P	任务完成3
H	任务完成4
J	实验结束
U	正式开始实验

Figure 3: Example of TOI configuration.

The core eye-tracking indicators selected in this study strictly corresponded to the predefined research questions. To address the issue of attention allocation, total fixation duration was adopted as the primary indicator. To assess visual search efficiency, two complementary indicators were selected: time to first fixation (TTFF) and total saccade amplitude. To reflect cognitive load, average pupil diameter (APD) was used as an objective physiological indicator. Under different task difficulty levels, the fixation proportion in the main view area was used to represent participants' attentional investment in the core information region. The results showed that the fixation proportion in the main view area generally increased with task difficulty, indicating that under high-difficulty tasks, participants allocated more attentional resources to the core flight information area. Descriptive results further showed that, at the high-difficulty stage, the fixation proportion was relatively higher in the trained group and the continuous-display group, suggesting that the effects of training and interface strategy became more prominent under high-load conditions. However, these differences still require further statistical verification.

Table 3: Means and standard deviations (M ± SD) of fixation time proportion in the main view area across task units for each group

Group	Task unit 1	Task unit 2	Task unit 3	Task unit 4
Trained	56.7±8.3	62.1±7.9	68.4±6.5	76.8±5.2
Untrained	59.8±9.1	63.5±8.2	67.9±7.0	70.2±6.1
Immediate-triggering	57.3±8.6	61.7±7.5	66.2±6.8	71.6±5.8
Continuous-display	59.2±8.0	64.0±7.7	69.5±6.3	75.3±4.9
Trained+Continuous-display	58.1±8.2	63.2±7.6	68.9±6.4	79.1±4.7

In the analysis of visual search efficiency, TTFF generally decreased from Task Unit 1 to Task Unit 4, whereas total saccade amplitude showed an increasing trend in the later stages of the task. This indicates that as task complexity increased, operators were required not only to orient more rapidly toward key instructions, but also to expand the scope of visual search accordingly. A further comparison of different interface pop-up types showed that the continuous-display group exhibited overall higher TTFF and relatively smaller total saccade amplitude than the immediate-triggering group, suggesting that the two pop-up types corresponded to different visual search patterns. In contrast, under the immediate-triggering condition, operators oriented to key information areas earlier (Fitz et al., 2019).

To further verify the above visual search trends, repeated-measures analyses of variance were conducted for TTFF and total saccade amplitude. The results showed that, after Greenhouse-Geisser correction, the main effect of task stage on total saccade amplitude was significant, $F(1.968, 57.084) = 7.835$, $p = .001$, partial $\eta^2 = 0.213$, indicating that participants' visual search range changed systematically as the task progressed. By contrast, the main effect of task stage on TTFF was not significant, $F(1.662, 49.852) = 1.830$, $p = .176$, partial $\eta^2 = 0.057$. These findings indicate that the increase in task difficulty was mainly reflected in an expansion of the overall visual search range, whereas its effect on the speed of initial attentional orientation was relatively limited.

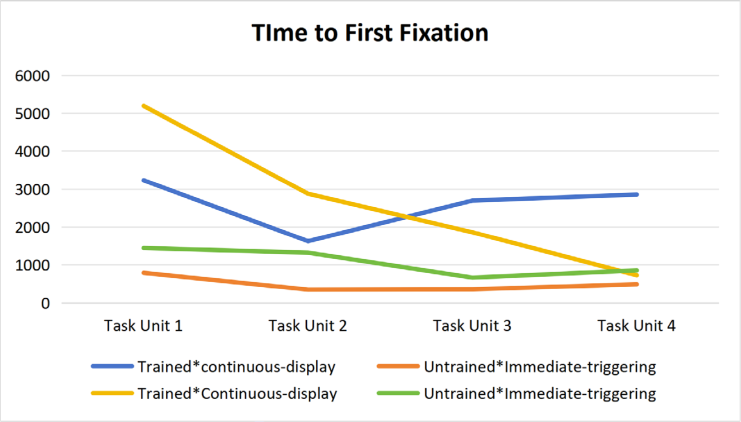


Figure 4: Line chart of time to first fixation (TTFF) (unit: ms).

At the eye-movement physiological level of cognitive load, descriptive results showed that APD exhibited a certain upward trend as the task progressed, and dynamic differences were also observed across different training and pop-up types. This pattern is consistent with the typical tendency for pupil diameter to increase with task difficulty (Di Stasi et al., 2016b), suggesting that interface strategy and training condition may correspond to different patterns of cognitive resource allocation. Further combined with indicators of fixation stability, the results showed that as task difficulty increased, overall fixation dispersion tended to increase, indicating that under high-complexity tasks, fixation stability generally declined. At the highest difficulty level, namely Task Unit 4, the trained group and the trained + continuous-display combination showed relatively lower dispersion, suggesting that training intervention may be associated with more stable attentional allocation under high-load conditions.

It should be noted that the repeated-measures analysis of variance showed that the main effect of task unit on APD did not reach significance, F(2, 56) = 1.008, p = .372, partial $\eta^2 = 0.035$. The interaction between task unit and experimental group was also not significant. Therefore, the above differences are better interpreted as trend-like changes and still require further verification with a larger sample.

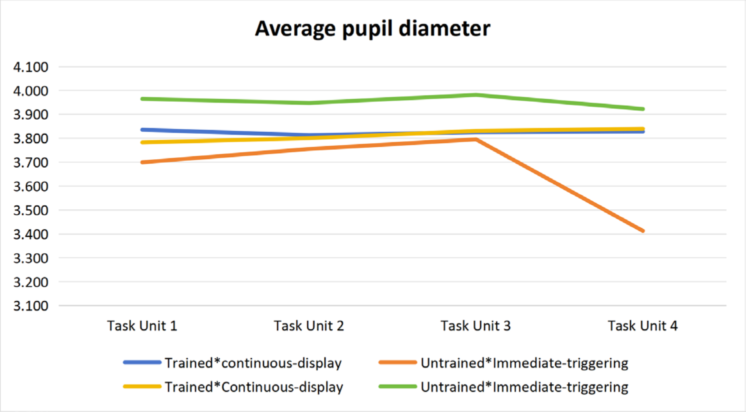


Figure 5: Line chart of average pupil diameter (APD) (unit: mm).

INTEGRATED DISCUSSION OF BEHAVIORAL PERFORMANCE AND EYE-TRACKING INDICATORS

In the present study, the practice effect was mainly reflected in improved efficiency during repeated task execution, but this effect was moderated by task difficulty. Task completion time decreased significantly in the second round from Task Unit 1 to Task Unit 3, whereas no significant difference was observed in Task Unit 4. Eye-tracking results further showed an improving trend in orientation to key information after task repetition, although the magnitude of change varied across pop-up types.

The 5-minute training did not produce a significant behavioral advantage, and only limited differences were observed in eye-tracking indicators. Although the trained group showed slightly longer fixation duration in the main view area, this difference was not significant. In contrast, the two pop-up types were associated with different patterns of visual search and attention allocation (Castelhano, Mack and Henderson, 2009). Compared with the continuous-display condition, the immediate-triggering condition showed shorter TTFF and supported faster orientation to key information, whereas the continuous-display condition was associated with greater fixation duration and total saccade amplitude, indicating broader interface scanning (Wang et al., 2025). These findings suggest that, although the two designs did not significantly affect final behavioral outcomes, they involved different cognitive processing pathways.

Figure 6: Scatter plot of time to first fixation.

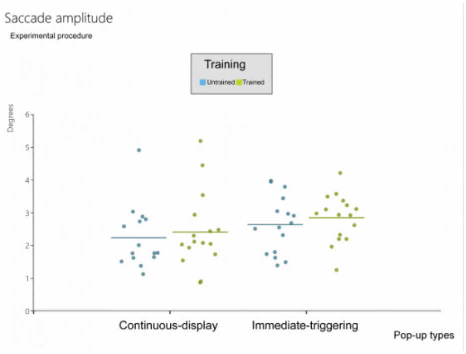


Figure 7: Scatter plot of saccade amplitude.

Gender-related findings should be interpreted cautiously, as they were exploratory and constrained by the small sample size. Although male and female operators showed similar behavioral performance, differences may exist in eye-movement patterns and responses to interface design. Female operators generally showed longer fixation duration and more stable attentional allocation, whereas male operators showed larger saccade amplitudes under the continuous-display condition (Halpern, 2012). These patterns suggest a possible moderating role of gender, but further validation with larger samples is required.

Overall, task difficulty primarily determined behavioral performance, whereas pop-up type and training had greater effects on cognitive processing. In particular, task difficulty significantly influenced some process-based eye-tracking measures, especially total saccade amplitude, while differences in TTFF and pupil diameter were relatively limited. Even when behavioral indicators showed no significant differences, eye-tracking measures still revealed distinct patterns of visual search and attention allocation. Therefore, the evaluation of UAV human-machine interfaces should combine behavioral and process-based indicators in order to more comprehensively assess the effects of different design strategies.

CONCLUSION

Through a systematic experimental design and multidimensional data analysis, this study explored the effects of training condition, pop-up type, and task difficulty on operators' cognitive processing and behavioral performance in drone delivery interaction interfaces. The results showed that task difficulty was the primary factor affecting operational performance. Under high-difficulty tasks, short-term training did not produce a stable performance advantage. At the same time, although macroscopic behavioral performance was generally similar across different interface conditions, eye-tracking indicators still revealed clear differences in cognitive processing.

With regard to different pop-up types, the immediate-triggering condition was more strongly associated with rapid orientation to key information, whereas the continuous-display condition was related to longer fixation duration and broader interface scanning characteristics. This indicates that although different pop-up types did not significantly change final task completion outcomes, they gave rise to different visual search and attention allocation strategies. From the perspective of cognitive processing, the value of optimizing drone delivery interfaces lies not only in improving task efficiency, but also in supporting cognitive load regulation and attentional stability. This finding provides empirical evidence for the design of situation-aware interfaces for complex low-altitude delivery scenarios.

In addition, although the training effect did not directly translate into a behavioral advantage, it still reflected differences in attentional allocation stability under high-difficulty tasks, suggesting that its boundary may be constrained by task complexity. Exploratory analyses further indicated that gender may influence operators' response patterns to interface pop-up types, although this result still requires further verification with a larger sample. Future research may further deepen this line of inquiry by improving ecological validity, expanding sample types, extending the training period, and introducing real-time cognitive state monitoring.

REFERENCES

Castelhano, M.S., Mack, M.L. and Henderson, J.M. (2009) Viewing task influences eye movement control during active scene perception. Journal of Vision, 9(3), p.6. https://doi.org/10.1167/9.3.6

Chen, H., Zhong, W., Yang, H., Gao, S., Yang, W. and Li, B. (2025) Comprehensive evaluation method of human-machine interface based on cognitive load: Taking forest firefighting command vehicle display console as an example. Journal of Beijing Forestry University, 47(6), pp.141–151. https://doi.org/10.12171/j.1000-1522.20250070

Di Stasi, L.L., Diaz-Piedra, C., Rieiro, H., et al. (2016b) Effects of long and short simulated flights on the saccadic eye movement. Physiology & Behavior, 153, pp.91–96. https://doi.org/10.1016/j.physbeh.2015.10.022

Di Stasi, L.L., Diaz-Piedra, C., Rieiro, H., Sánchez Carrión, J.M., Martin Berrido, M., Olivares, G. and Catena, A. (2016a) Gaze entropy reflects surgical task load. Surgical Endoscopy, 30(11), pp.5034–5043. https://doi.org/10.1007/s00464-016-4851-8

Djamasbi, S., Siegel, M. and Tullis, T. (2010) Generation Y, web design, and eye tracking. International Journal of Human-Computer Studies, 68(5), pp.307–323. https://doi.org/10.1016/j.ijhcs.2009.12.006

Endsley, M.R. (1995) Toward a theory of situation awareness in dynamic systems. Human Factors, 37(1), pp.32–64. https://doi.org/10.1518/001872095779049543

Faul, F., Erdfelder, E., Lang, A.G. and Buchner, A. (2007) G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behavior Research Methods, 39(2), pp.175–191. https://doi.org/10.3758/BF03193146

Fitz, N., Kushlev, K., Jagannathan, R., Lewis, T., Paliwal, D. and Ariely, D. (2019) Batching smartphone notifications can improve well-being. Computers in Human Behavior, 101, pp.84–94. https://doi.org/10.1016/j.chb.2019.07.016

Halpern, D.F. (2012) Sex differences in cognitive abilities. 4th edn. New York: Psychology Press.

Hart, S.G. (2006) NASA-Task Load Index (NASA-TLX); 20 years later. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 50(9), pp.904–908. https://doi.org/10.1177/154193120605000909

Kim, S., Nussbaum, M.A. and Gabbard, J.L. (2019) Influences of augmented reality head-worn display type and user interface design on performance and usability in simulated warehouse order picking. Applied Ergonomics, 74, pp.186–193. https://doi.org/10.1016/j.apergo.2018.08.026

Liu, B., Pan, H., Qiu, Y. and Ding, J. (2025) The impact of eye-tracking-based full-touchscreen car dashboards on driving safety. (Automotive and Safety), (4), pp.65–70.

Orlandi, L. and Brooks, B. (2018) Measuring mental workload and physiological reactions in marine pilots: Building bridges towards redlines of performance. Applied Ergonomics, 69, pp.74–92. https://doi.org/10.1016/j.apergo.2018.01.005

Politowicz, M.S. (2024) Visual attention in remote vehicle supervision: Examining the effects of mental models and information bandwidth. Master's thesis, Old Dominion University. https://doi.org/10.25777/xske-8625

Tezza, D. and Andujar, M. (2019) The state-of-the-art of human-drone interaction: A survey. IEEE Access, 7, pp.167438–167453. https://doi.org/10.1109/ACCESS.2019.2953900

Wang, J., Lu, G., Gao, Y., Hu, B. and Yang, M. (2025) Enhancing UAV human-machine interaction with multimodal behavioral data: A gaze-posture synergistic approach. IEEE Transactions on Consumer Electronics, 71(3), pp.8033–8044. https://doi.org/10.1109/TCE.2025.3593889

Wiggins, M.W. (2021) A behaviour-based approach to the assessment of cue utilisation: implications for situation assessment and performance. Theoretical Issues in Ergonomics Science, 22(1), pp.46–62. https://doi.org/10.1080/1463922X.2020.1758828

Zhang, W., Liu, Y. and Kaber, D.B. (2024) Effect of interface design on cognitive workload in unmanned aerial vehicle control. International Journal of Human-Computer Studies, 189, p.103287. https://doi.org/10.1016/j.ijhcs.2024.103287