MULTIMODAL CUEING IN PRECISION HOVERING TASKS: PREDICTING PILOT COGNITIVE WORKLOAD VIA PHYSIOLOGICAL MEASUREMENTS

Gabriele Luzzani
Postdoctoral Fellow
Maryland Robotics Center
University of Maryland
College Park, MD, USA

Michael T. Morcos
Graduate Research Assistant
Department of Aerospace Engineering
University of Maryland
College Park, MD, USA

Umberto Saetti
Assistant Professor
Department of Aerospace Engineering
University of Maryland
College Park, MD, USA

ABSTRACT

This study investigates the relationship between mental workload (MWL) and full-body haptic haptic feedback during an ADS-33 MTE-like precision hovering task involving fourteen participants. The task was conducted under four conditions combining good and degraded visual environments (GVE/DVE) with or without haptic cues. Perceived MWL was measured using the Bedford Workload Questionnaire (BWQ), while physiological signals, including cardio-respiratory activity, brain activity (fNIRS), skin temperature, and electrodermal activity, were used as objective indicators of cognitive load. Data were analyzed using univariate statistical tests and multivariate Generalized Linear Mixed Models (GLMM). Results showed that haptic feedback significantly reduced perceived MWL, particularly in degraded visual conditions. Physiological signals demonstrated sensitivity to workload variations, and the GLMM explained up to 88% of the variance. Prediction performance peaked under the most demanding condition, suggesting a stronger physiological response to increased cognitive load. These findings underscore the potential of haptic feedback to mitigate MWL in safety-critical operations and highlight physiological monitoring as a reliable approach for future adaptive flight systems.

NOTATION

*

Acronyms

BWQ Bedford Workload Questionnaire

CLS China Lake Situational Awareness Scale

DVE Degraded Visual Condition

DVEH Degraded Visual Condition + Haptic

ECG Electrocardiography

EDA Electrodermal Activity

EMS Electrical Muscular Stimulation

FAA Federal Aviation Administration

fNIRS functional Near-Infrared Spectroscopy

GLMM Generic Linearized Mixed Model

GVE Good Visual Condition

GVEH Good Visual Condition + Haptic

HF Human Factors

HMI Human-Machine Interaction

K-W Kruskal Wallis Test

LASSO Least Absolute Shrinkage and Selection Operator

MTE Mission Task Element

M-W Mann-Whitney U Test

MWL Mental Workload

NDI Nonlinear Dynamic Inversion

RCAH Rate Command/Attitude Hold

SA Situational Awareness

SPO Single Pilot Operation

INTRODUCTION

Recent advances in neuroengineering have increased interest in human-robot interaction, pushing the application of Human Factors (HF) principles to Human-Machine Interaction (HMI) systems. Users are more frequently required to interact with autonomous platforms, demanding high standards of safety and reliability. These interactions occur across both everyday activities, such as driving, and safety-critical domains, including air traffic control, aviation, space operations, and healthcare.

The aeronautical sector plays a key role in HF research due to the forecasted transition toward Single Pilot Operations (SPOs) and the continuous evolution of helicopter pilot training and flight operations. According to the Federal Aviation Administration (FAA) (Ref. 1), helicopter missions rise substantial HF challenges due to their operational complexity and demanding conditions. These environments contribute to elevated levels of physical and mental fatigue, increased stress, and a higher risk of human error, all of which degrade communication and decision-making performance. The U.S. Government Accountability Office (GAO) reports that approximately 90% of rotorcraft accidents are attributable to human error (Ref. 2). Cognitive demands are particularly significant during pilot training, where individuals often encounter unfamiliar scenarios without adequate experience. These challenges are further intensified by the high financial and temporal costs associated with training. In this context, reducing excessive mental workload (MWL) while enhancing situational awareness (SA) is a central concern.

MWL is a complex, multidimensional construct that has been extensively studied in the HF literature (Refs. 3–5). It relates to attentional capacity, cognitive effort, and task performance, varying according to task demands and individual capabilities (Refs. 6–8). These cognitive processes are reflected in physiological responses, with MWL being closely associated with activity in the autonomic nervous system (Ref. 9). SA, on the other hand, refers to the perception of environmental elements, their comprehension, and the projection of future states relative to task objectives (Ref. 10). It is closely linked to MWL, as increased cognitive demand can impair SA (Ref. 11). Similarly, vigilance, the ability to sustain attention over time, decreases with increasing workload, particularly in high-demand operational contexts (Ref. 12). Given the sensitivity of both SA and vigilance to MWL, these constructs offer a unified framework for evaluating how multimodal cueing strategies can support cognitive performance in complex environments.

Related Works

In aviation, HMI systems, where control and perception are shared between a pilot and an intelligent agent, offer a promising solution for managing complex vehicles such as airplanes, helicopters, and drones (Refs. 13, 14). However, current HMI models rely primarily on visual and vestibular signals, often neglecting auditory and somatosensory inputs, which can result in perceptual dissociation, particularly in teleoperation scenarios, leading to reduced SA and increased MWL (Refs. 15, 16).

Somatosensory feedback, including tactile and haptic cues, has been shown to mitigate this dissociation by reinforcing pilot awareness and engagement (Ref. 17). Multimodal cueing, extensively studied in neuroscience and aviation, has demonstrated benefits in SA, pilot training, and flight envelope protection (Refs. 18–21). Haptic interfaces, such as force-feel sticks, support motor learning and task performance (Refs. 22, 23), while full-body tactile systems and Electrical Muscular Stimulation (EMS) improve perception in Degraded Visual Environments (DVE) by encoding motion through continuous cueing strategies (Refs. 24–26). Similarly, spatial audio cueing has shown potential in maintaining Pilot-Vehicle System performance under sensory degradation (Ref. 27). These findings highlight the promise of multimodal feedback in enhancing pilot performance, though further validation in real-world aviation operations is needed.

In this context, varying cueing modalities could significantly affect MWL (Ref. 28), making its accurate estimation essential for optimizing sensory cue combinations. MWL has been commonly assessed through subjective questionnaires, behavioral observations, and psychophysiological measurements. However, the first two approaches lack real-time feasibility and scalability in dynamic contexts, respectively (Refs. 29, 30). Current biomedical sensing technologies allow real-time MWL estimation through physiological signals such as cardio-respiratory activity, electrodermal activity (EDA), skin temperature, eye movements, and brain activity (Refs. 31–35). Due to the complex and context-dependent nature of MWL, no single physiological signal has proven universally reliable. This complexity necessitates a multimodal approach, wherein multiple physiological signals with demonstrated relevance to MWL are acquired, and a comprehensive analysis for identifying the most informative subset for a specific application.

Motivation of work

Understanding the relationship between perceived MWL and physiological responses is critical for designing reliable human-machine interfaces in aviation. This study investigated that relationship during a hover task based on the Mission Task Element (MTE) described in the ADS-33E-PRF standard (Ref. 36), conducted using a motion-based VR simulator, as introduced in (Ref. 37). Participants performed the task under four cueing conditions, combining good and degraded visual environments with or without haptic feedback. MWL was assessed using both subjective and physiological metrics. Physiological signals included ECG, respiration, functional near-infrared spectroscopy (fNIRS), electrodermal activity (EDA), and skin temperature. These were selected based on (i) prior evidence linking them to MWL, (ii) suitability for integration into wearable systems for aviation use, and (iii) compatibility with a unified, real-time acquisition platform (Refs. 38, 39). Subjective workload was measured using the Bedford Workload Questionnaire (BWQ) (Ref. 40), a validated 10-level self-report scale commonly employed in aviation research. The investigation was structured around three hypotheses:

Hypothesis 1 (H1): Perceived workload (BWQ ratings) differs significantly between cueing modalities.

Hypothesis 2 (H2): Physiological responses vary meaningfully across cueing modalities when grouped by self-reported MWL levels.

Hypothesis 3 (H3): A combined set of physiological features can predict perceived MWL, with key predictors identifiable through multivariate modeling.

To evaluate *H2*, univariate analyses were used to examine the sensitivity of each physiological metric to perceived workload levels. For *H3*, we implemented a multivariate framework combining LASSO regularization with Generalized Linear Mixed Models (GLMM), enabling robust feature selection. This approach identified the most predictive physiological markers and quantified their contributions to a multimodal model of MWL.

This paper is structured as follows. The *Materials and Methods* section describes the test population, equipment, hover test procedure, and data processing. The *Results* section outlines the analysis strategy to evaluate the relationship between MWL and physiological responses. The *Discussion* interprets the findings, and the *Conclusions* summarize the results and suggest future directions.

MATERIALS AND METHODS

Overview

This study investigated the impact of multimodal cueing strategies on pilot physiological signals and MWL during a precision hover task, as illustrated in Fig. 1. The task setup followed the MTE framework defined in the ADS-33E-PRF standard (Ref. 36). Four experimental conditions were tested:

1. **GVE** (Good Visual Environment): Clear visibility with no wind or external disturbances affecting the aircraft.
2. **DVE** (Degraded Visual Environment): Simulates dawn or low-light daytime conditions, reducing visibility and visual reference reliability.
3. **GVE + H** (Good Visual Environment with Haptic Feedback): Same as GVE, with the addition of haptic cues.
4. **DVE + H** (Degraded Visual Environment with Haptic Feedback): Same as DVE, augmented by haptic feedback.

Precision Hover Task Participants were instructed to maintain a stable hover within specified tolerance limits for position (longitudinal, lateral, and vertical) and heading. Desirable performance required staying within ± 3 ft laterally and longitudinally, ± 2 ft in altitude, and $\pm 5^\circ$ in heading. Adequate performance thresholds were defined as ± 6 ft, ± 4 ft, and $\pm 10^\circ$, respectively. Each trial involved sustaining the hover for 60 s, with participants completing 5 trials in total. During the experiment, primary outcome measures included positional deviation in the x, y, and z axes of the North-East-Down (NED) coordinate frame, and heading error in degrees. Additionally, pilot control input time histories were recorded for further analysis.

Equipment

A full-body haptic system composed of the `bHaptics TactSuit X40` and `Tactosy for Arms` was used to deliver vibrotactile feedback. The TactSuit included 40 actuators across the chest and back, while each arm sleeve contained 6 actuators. Both components were wireless and commercial devices. Participants used a motion-based VR simulator (`Brunner NovaSim VR/MR`), equipped with programmable rotorcraft inceptors, a 6-DOF motion platform (±30°, 2 rad/s), vibration seat cueing, and Varjo Aero VR goggles for immersive visual feedback.

Physiological signals, including ECG, respiration, EDA, skin temperature, and fNIRS, were recorded using *biosignalsplux Professional Kit* and *OpenSignals* software at 1000 Hz. Sensor placement was as follows:

- **ECG Sensor:** Measured heart activity (ECG); placed on the thorax.
- **Inductive RIP Sensor:** Measured respiration via thoracic expansion; placed on the thorax.
- **fNIRS Sensor:** Captured hemodynamic response; positioned on the forehead (prefrontal cortex, PFC).
- **EDA Sensor:** Recorded electrodermal activity (EDA); placed on the index and middle fingers.
- **Temperature Sensor:** Measured skin temperature; placed on the ring fingertip.

Population

Fourteen pilots (13 males and 1 female) participated in the experiments, with a mean age of 35 years ($M = 35$, $SD = 15$). Among them, three held rotorcraft licenses, and three were licensed fixed-wing civilian pilots. The remaining participants were unlicensed but had varying degrees of fixed-wing or rotorcraft flight experience. Only one participant had no prior flight experience but was adequately trained to meet the performance requirements of the tasks.

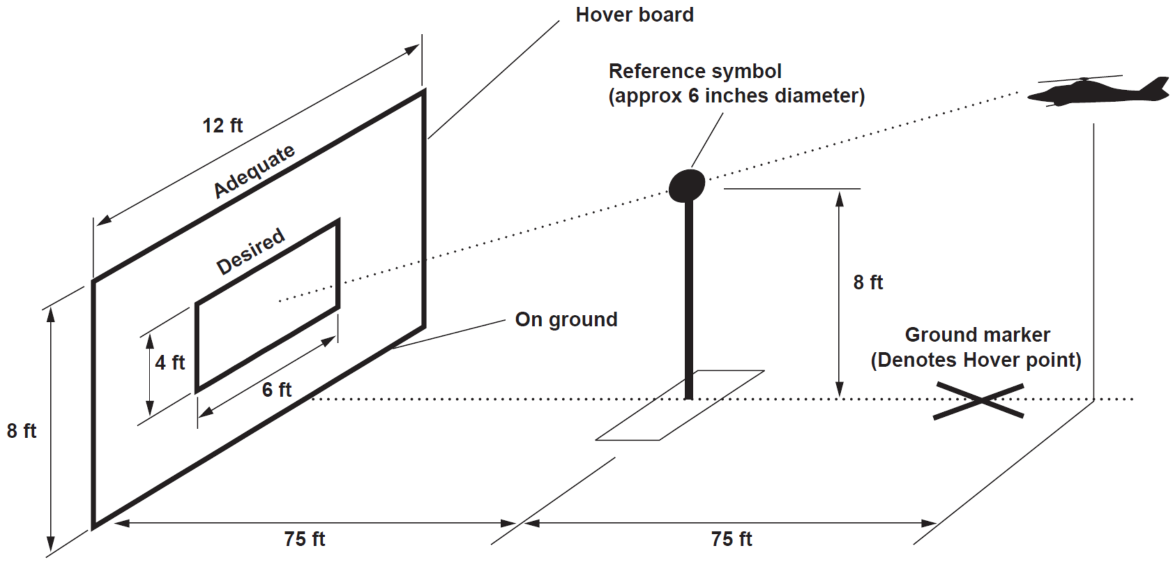


Figure 1: Hover MTE scheme (Ref. 36).

Experimental Procedure

The experimental procedure, illustrated in Fig. 2, began with a briefing session in which participants received detailed information regarding the task objectives, procedures, and safety considerations. Each session was conducted in a climate-controlled room, maintained at a temperature of 71 °F and relative humidity of 50 %. After providing informed consent, participants were equipped with the *bHaptics* suit and *PLUX* physiological sensors.

A setup and calibration phase followed, lasting approximately 15 min, during which the haptic system was individually calibrated to determine personalized vibration intensity coefficients based on each participant's tactile sensitivity. Once calibration was complete, participants underwent a 30 min instruction and practice session. This phase provided structured guidance on the hover task and allowed participants to complete practice trials under all cueing conditions to ensure familiarity and competency. Unlicensed participants received prior training sessions on separate days to achieve the required proficiency before the experimental session.

After training, participants were instructed to rest while baseline physiological data were collected across the five monitored signals (ECG, respiration, EDA, skin temperature, and fNIRS).

Each participant completed the hover task under four cueing conditions: GVE, GVE + H, DVE, and DVE + H. Each condition was repeated 5 times in a randomized order, with the condition order counterbalanced across participants. Each trial lasted 60 s, followed by a 30 s rest period. A longer rest of approximately 5 min was provided between different cueing conditions to minimize fatigue.

The full hover task session lasted approximately 1.5 h. Including consent, calibration, testing, rest breaks, and post-test procedures, the total experimental duration was up to 2.5 h per participant.

Following each cueing condition, participants completed two subjective assessments: the BWQ for mental workload and the China Lake Situational Awareness Scale (CLS) for SA (Ref. 41). At the conclusion of the session, a 15 min debrief and cleanup period was conducted, during which participants removed the equipment, the haptic suits were sanitized, and qualitative feedback was collected.

Simulation and Control Architecture

Flight Simulation Model The simulation and control framework used in this study was based on the high-fidelity infrastructure described in Morcos et al. (Ref. 37). The rotorcraft flight dynamics were modeled using an in-house MATLAB®/Simulink framework developed at the University of Maryland. The software supported flexible configurations of multi-rotor and fixed-wing systems, with aerodynamic interactions between rotors and wings modeled through dynamic inflow, lifting-line theory, and Biot-Savart circulation-based formulations. The fuselage was treated as a rigid body, and rotor-on-rotor and rotor-wing aerodynamic effects were explicitly accounted for. The flight simulation model had been validated against flight test data and supports integration with real-time human-in-the-loop simulation.

Flight Control Design The control architecture was based on nonlinear dynamic inversion (NDI) (Refs. 42–55), providing a model-based approach to feedback linearization and tracking of a desired reference model. The inner-loop design supported Rate Command/Attitude Hold (RCAH) laws for roll, pitch, and yaw, and altitude hold for the heave axis. An outer-loop control law was implemented to track position and velocity commands in the inertial frame using a reduced-order

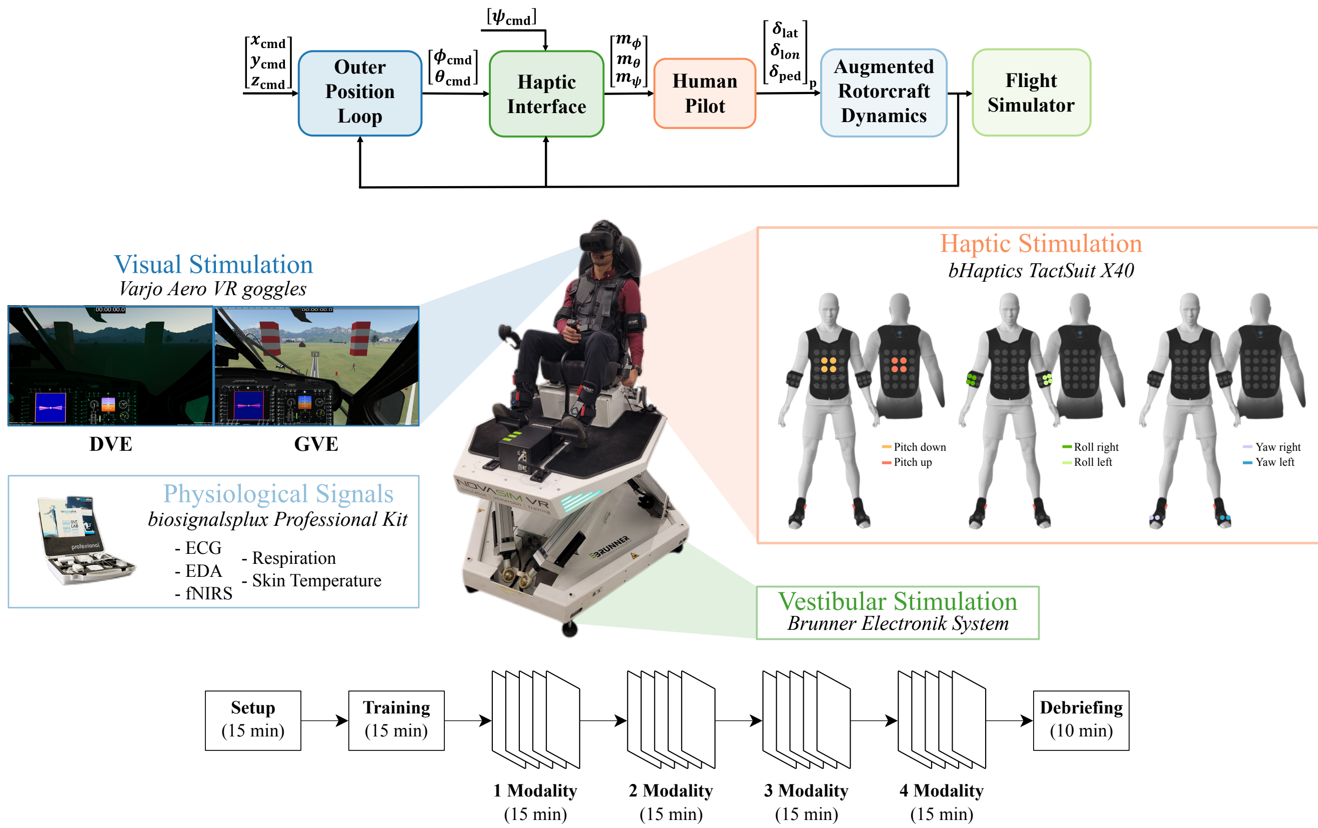


Figure 2: Overview of the experimental setup, including the procedure, test equipment, and configuration. The four modalities were randomly assigned to each subject among: GVE, GVE+H, DVE, and DVE+H. The control system's block diagram is also reported.

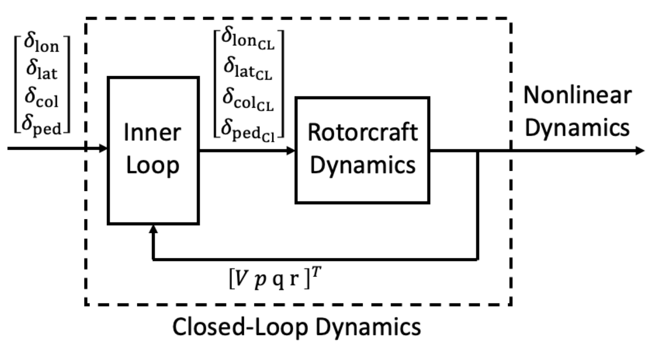


Figure 3: Schematic of the closed-loop rotorcraft dynamics (Ref. 37).

dynamic model. The outer-loop controller used PID compensation with feedforward acceleration commands to ensure trajectory tracking while maintaining sufficient frequency separation from the inner loop to avoid pilot-induced oscillations. Gains were selected to satisfy Level 1 handling qualities and performance requirements as defined in ADS-33. The closed-loop rotorcraft dynamics is reported in Fig. 3.

Haptic Feedback Cueing Strategy Haptic cues were assigned to the control deviations in roll, pitch, and yaw, and were generated by a cueing algorithm that translated attitude commands into actuator activation patterns. The haptic system enabled intuitive, multi-axis guidance during position and velocity tracking tasks. In hover tasks, the haptic feedback directed the pilot toward a fixed point in space. The output provided to the pilot through the *bHaptics* and the activation of the tactors for pitch, roll, and yaw controls is reported in Figure 2.

Software Architecture The full system was compiled into C++ code using Simulink Coder and executed in a real-time simulation loop. The simulation architecture comprised modules for vehicle dynamics, flight control, haptic cueing, and visual rendering. A wrapper code synchronized data logging and module communication. The pilot interacted with the system through a VR-based motion simulator using Lockheed Martin's Prepar3D® with a VR plugin, ensuring full immersion and accurate representation of cockpit dynamics. The result was a robust, modular platform for investigating pilot behavior, human-machine interaction, and haptic feedback efficacy in complex rotorcraft tasks.

Physiological Signal Processing

The signal processing of the five selected physiological signals (ECG, EDA, respiration, skin temperature, and fNIRS) followed the pipeline and the features extracted reported in (Refs. 38, 39). A total of 81 features were extracted per trial: 11 from heart rate (HR) and heart rate variability (HRV), 8 from EDA including skin conductance level (SCL) and skin conductance response (SCR), 10 from skin temperature, 14 from respiration (breath rate-related), and 38 from fNIRS (temporal, statistical, and spectral domains).

Data Analysis

BWQ Scores Analysis To evaluate Hypothesis 1 (*H1*), perceived MWL was assessed using the BWQ (Ref. 40), a 10-level scale where 1 indicates minimal workload and 10 indicates maximal workload. The BWQ was administered after each cueing condition. To determine whether MWL scores differed significantly across the four experimental modalities, a Kruskal–Wallis (K–W) test was conducted. This non-parametric test was chosen following preliminary checks for normality (Shapiro–Wilk test) and homogeneity of variances (Levene's test). The significance threshold for all analyses was set at $p = .05$.

Univariate Analysis of Physiological Features To test Hypothesis 2 (*H2*), we performed univariate analyses to investigate whether individual physiological features varied significantly across perceived MWL levels. The analysis included three phases: data preprocessing, workload classification, and statistical testing.

Preprocessing. To account for inter-individual variability, each physiological feature was baseline-corrected using the percentage change from the subject's resting baseline, as defined in Equation 1:

$$X_{\text{corr}} = \frac{X_{\text{raw}} - X_{\text{baseline}}}{X_{\text{baseline}}} \tag{1}$$

After baseline correction, feature-level outliers were identified and removed on a per-participant basis using the interquartile range (IQR) method, thereby minimizing noise within each condition. Finally, all features were standardized using z-score normalization for each subject individually, ensuring scale consistency while preserving within-subject variability.

Workload Classification. Due to the limited sample size ($N = 14$) and the fine granularity of the BWQ, we grouped the scores into three broader workload levels: *Relax*, *Low MWL*, and *High MWL*. Grouping was guided by distributional metrics (e.g., normality, skewness, kurtosis). The baseline condition was retained as a reference class.

Statistical Testing. A one-way Kruskal–Wallis test was used to assess whether each physiological feature differed significantly across workload classes. For features showing significant main effects ($p < .05$), we conducted pairwise Mann–Whitney U tests for post hoc comparisons, applying Benjamini–Hochberg correction to control the false discovery rate ($\alpha = .05$).

Multivariate Analysis To test Hypothesis 3 (*H3*), we applied a multivariate regression approach to evaluate whether combinations of physiological features could predict BWQ scores. The model design accounted for repeated measures and individual variability using mixed-effects modeling. Unlike the univariate analysis, this approach used raw (non-averaged) trial data and preserved the original 10-point BWQ scale.

Preprocessing. As in the univariate case, features were baseline-corrected using Equation 1, but individual trial-level data were retained. Z-score standardization was again applied per subject. Because the BWQ was scored at the condition level, the same BWQ value was assigned to all five trials of each modality per participant.

Modeling Pipeline. The multivariate framework consisted of two main stages:

1. **Feature Selection (LASSO):** To reduce dimensionality and avoid multicollinearity, we applied L1-regularized linear regression (LASSO), selecting only features with non-zero coefficients:

$$\hat{\beta} = \arg\min_{\beta} \left\{ \frac{1}{2n} \|\mathbf{y} - \mathbf{X}\beta\|_2^2 + \lambda \|\beta\|_1 \right\} \tag{2}$$

   The optimal regularization parameter λ was determined via 10-fold cross-validation.

2. **Generalized Linear Mixed Model (GLMM):** Selected features were then used in a GLMM where subject identity was modeled as a random intercept:

$$\text{BWQ}_i = \mathbf{X}_i\beta + b_{s(i)} + \varepsilon_i \tag{3}$$

   - $\mathbf{X}_i$ is the vector of physiological features for trial i,
   - β are fixed-effect coefficients,
   - $b_{s(i)} \sim \mathcal{N}(0, \sigma_b^2)$ is the random effect for subject s,
   - $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$ is the residual error.

   The dataset per modality included five trials and one baseline per subject, leading to:

$$n_{\text{modality}} = (5+1) \times 14 \times 81 = 6804$$

Two evaluation strategies were used:

- **Global Fit:** The overall model performance was quantified using marginal R^2 (fixed effects) and conditional R^2 (fixed + random effects), as described by Nakagawa and Schielzeth (Ref. 56). We also reported AIC, BIC, and log-likelihood metrics.
- **Leave-One-Subject-Out (LOSO):** To assess generalizability, a LOSO cross-validation was conducted. For each modality, we reported the average R^2_{LOPO}, RMSE, and visual comparisons between true and predicted BWQ scores.

RESULTS

This section reports the *Data Analysis* results in three parts: (i) subjective MWL ratings from the BWQ, (ii) univariate analysis of physiological features clustered by MWL levels, and (iii) multivariate analysis using feature selection and mixed-effects modeling.

BWQ Perceived MWL

The subjective MWL perceived by participants, as measured by the BWQ, is illustrated in Figure 4. Specifically, Figure 4a presents boxplot distributions for each cueing modality, while Figure 4b shows the overall distribution across all modalities.

Table 1 summarizes the results of the normality and homogeneity of variance tests. As the GVEH and DVE conditions violated the normality assumption ($p < .05$), the non-parametric K-W test was used to evaluate *H1*. The analysis revealed a significant main effect of cueing modality, $H(3) = 9.69$, $p = .02$, indicating that perceived MWL differed significantly across conditions.

Furthermore, the overall distribution shown in Figure 4b exhibited a significant deviation from normality ($p < .001$, $W = 0.90$), though the distribution remained approximately symmetrical, as indicated by a low skewness value of 0.275. Despite the departure from normality, the distribution was considered roughly centered around the median (median = 4), justifying the use of non-parametric analysis as a first approximation.

CLS results related to SA subjective perception for each modality are reported in the *Appendix*.

Table 1: Shapiro–Wilk test for normality and Levene's test for homogeneity of variance for each BWQ distribution across cueing modalities.

Modality	Normality Test		Homogeneity of Variance	
	p	**W**	**F(3,52)**	p
GVE	.106	.898	0.239	.8685
GVEH	.025	.854		
DVE	.021	.849		
DVEH	.351	.934		

Table 2: Frequency counts of encoded BWQ ratings (0 = low, 1 = medium, 2 = high) per cueing modality in the Bedford dataset.

Modality	Relax	Low MWL	High MWL
DVE	0	10	57
DVEH	0	32	35
GVE	0	45	25
GVEH	0	55	14
Baseline	14	0	0

One-way analysis

To assess *H2*, BWQ scores were discretized into three cognitive states based on their symmetrical distribution: *Relax* (baseline), *Low MWL* (BWQ ≤ 4), and *High MWL* (BWQ > 4). This classification led to an unbalanced number of observations across modalities and workload levels, as shown in Table 2. The variation in counts per condition reflects the removal of outliers discussed in the previous section.

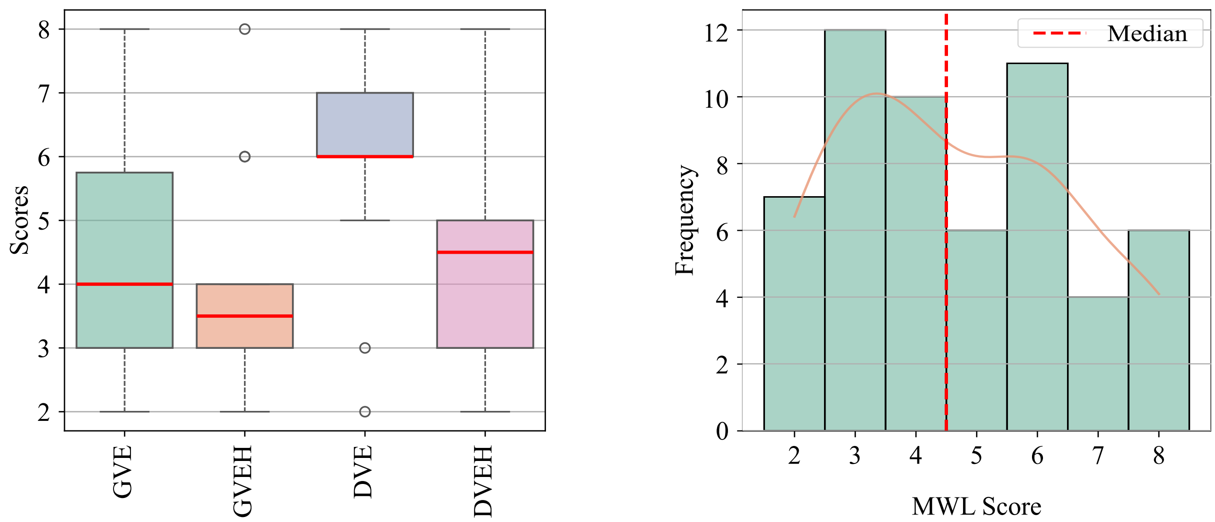
Given the large number of statistical comparisons (81 features across 4 modalities), the results are summarized visually. Figure 5b shows, for each modality, the number of features as percentage respect to the initial number that reached significance at successive stages of the analysis: (i) Kruskal–Wallis (K-W) test, (ii) one, (iii) two, or (iv) all three Mann–Whitney (M-W) post hoc comparisons (with Benjamini–Hochberg correction).

Figure 5c to 5f further presents these results by physiological signal, ECG, EDA, temperature (T), respiration (RESP), and fNIRS, highlighting the contribution of each signal to statistical significance in all modalities.

Over 30% of features reached significance in the K-W test, and over 20% in at least two post hoc comparisons. However, all modalities showed less than four features consistently significant across all MWL levels. Notably, **DVE** showed no fully significant features. Cardio-respiratory, temperature, and fNIRS signals demonstrated the highest sensitivity to MWL variation, while EDA remained the least informative, highlighting its limited sensitivity in workload classification within this experimental context, in accordance with previous studies reported in (Ref. 26).

Multivariate Analysis

The results of the multivariate analysis using the LASSO + GLMM pipeline are summarized in Table 3. Overall, the total variance explained by each model, as indicated by the conditional R^2 (R^2_{cond}), reached 85-88% across all modalities. The variance explained by the fixed effects alone (R^2_{marg}), driven by the physiological features, was substantial in each case, and particularly high in the most cognitively demanding condition, **DVE**. This is further supported by the ΔR^2 (the difference between conditional and marginal R^2), which reaches its minimum in the **DVE** condition, indicating a lower relative contribution from subject-specific random effects.



(a) BWQ rating per modality.

(b) Overall BWQ rating.

Figure 4: BWQ ratings output. Figure 4b shows the questionnaire results grouped by modality, while Figure 4b reports its overall distribution, highlighting the median value.

Table 3: Model performance metrics for each modality using LASSO feature selection followed by GLMM with LOSO cross-validation.

Modality	Features	λ	R^2_{LOSO}	RMSE	R^2_{marg}	R^2_{cond}	ΔR^2	AIC	BIC	logLik
GVE	17	0.249	0.13	2.24	0.47	0.88	0.42	292.0	340.0	-126.0
GVEH	25	0.079	0.01	2.06	0.50	0.85	0.35	298.0	366.0	-121.0
DVE	15	0.238	0.40	2.12	0.62	0.87	0.25	304.0	347.0	-134.0
DVEH	16	0.189	0.30	1.85	0.49	0.86	0.36	284.0	329.0	-123.0

Model fit indices (AIC, BIC, and log-likelihood) reported comparable values across modalities, suggesting similar model adequacy. Variations in the LASSO regularization parameter λ highlight differences in model sparsity. Specifically, a lower λ value in the **GVE** condition resulted in a denser model with a greater number of selected features.

Regarding generalization performance, the LOSO cross-validated R^2 (R^2_{LOSO}) remained below 40% across all modalities. This outcome reflects the limited number of participants relative to the complexity of the cognitive model under investigation, and the necessity of incorporating a substantial number of physiological features. Despite this, the results are informative: the **DVE** condition, associated with the highest workload, led the best predictive performance with $R^2_{\text{LOSO}} = 0.40$, suggesting strong potential for future development. In contrast, the lowest-performing model was observed in the **GVEH** condition, which corresponded to the least cognitively demanding task and exhibited near-zero cross-validated predictive ability.

Figure 6 shows the mean standardized regression coefficients for each physiological signal across the four cueing modalities. These values reflect the relative contribution of each signal to the GLMM prediction of perceived MWL, based on the features selected through LASSO regularization. Consistent with findings reported in (Ref. 26), cardio-respiratory (ECG, RESP) and neural (fNIRS) signals exhibit the highest importance across all conditions. Temperature (T) shows intermediate influence, while electrodermal activity (EDA) consistently contributes the least to the model, indicating a lower predictive value relative to the other physiological variables. The regression prediction results are reported in the *Appendix*, confirming a trend of higher predictive performance in the more cognitively demanding conditions (**DVE** and **DVEH**), and lower performance in less demanding scenarios such as **GVE** and **GVEH**.

DISCUSSION

The purpose of this research was to study the relationship between perceived MWL and the response of physiological signals during a hover task in a motion-based VR immersive simulation environment. The hypotheses *H1* to *H3* were proposed to address each aspect of this relationship (subjective perception, univariate and multivariate physiological characteristics responses). Since each hypothesis addressed a different aspect

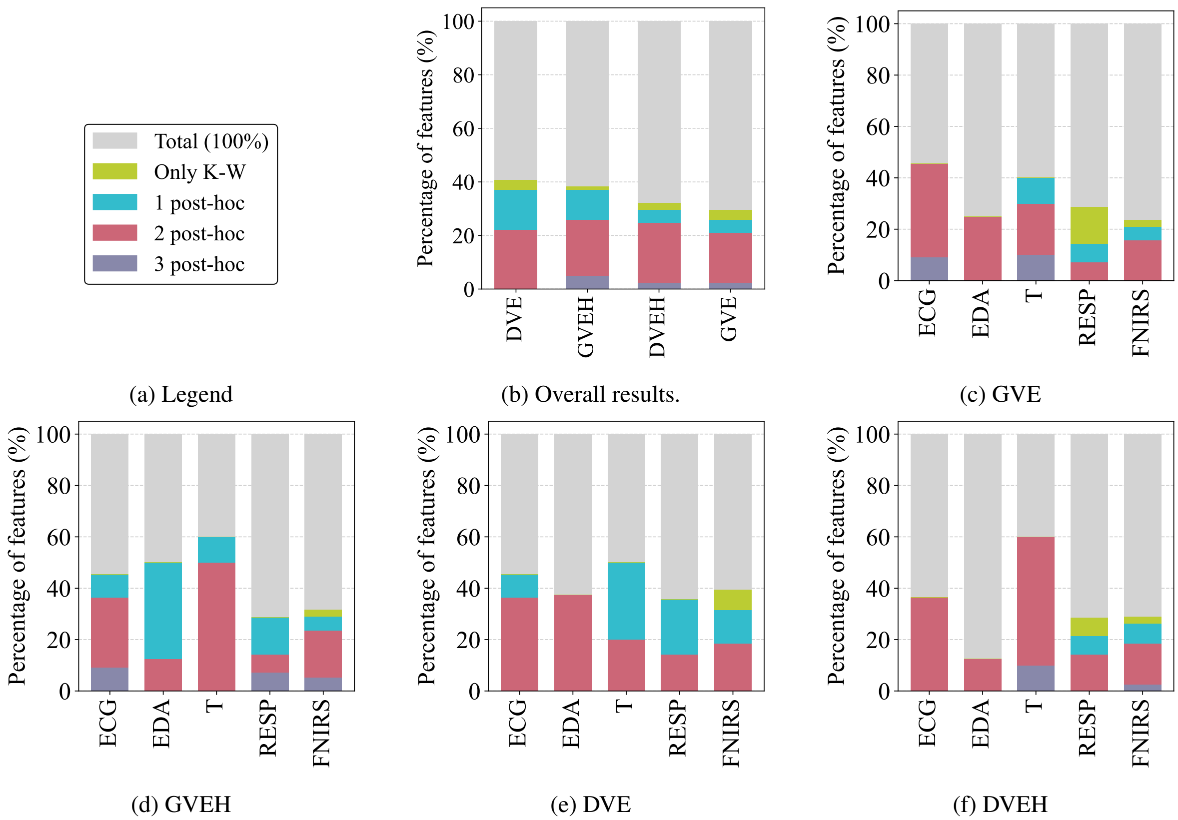


Figure 5: This Figure reports for each modality and physiological signal the percentage of significant features for each statistical analysis step.

of this relationship, the following section is structured around them to highlight and discuss key findings and their impact on HF research.

BWQ perceived mental workload

Hypothesis *H1* proposed that perceived workload, as measured by the BWQ, would significantly differ across cueing modalities. This hypothesis was supported by the results shown in Figure 4a and confirmed by the non-parametric K-W statistical analysis. In particular, modalities involving haptic feedback, both in good visual environments (**GVEH**) and degraded visual environments (**DVEH**), showed a clear and significant reduction in perceived MWL compared to conditions without haptic cues. This difference was especially evident under degraded visual conditions.

These findings are consistent with previous studies (Refs. 26, 37), reinforcing the idea that haptic feedback can help mitigate cognitive load even in high-fidelity simulation environments. This effect is particularly relevant in demanding operational scenarios, such as hovering tasks, where reducing MWL is critical for maintaining safety.

The results support the hypothesis that when a primary sensory modality, such as vision, is impaired due to external factors (e.g., degraded visual conditions), the introduction or enhancement of alternative sensory channels (e.g., haptics), which are not typically relied upon, can compensate for this deficit. This multisensory integration may restore the perceived cognitive effort to a level comparable to less demanding situations, such as those with good visual conditions (**GVE**). In particular, the similarity in perceived workload distributions between the **GVE** and **DVEH** conditions reports empirical support for this interpretation.

One-Way Analysis of Physiological Features

Hypothesis *H2* aimed to investigate the effect of varying workload levels on individual features extracted from five physiological signals. Given the limited sample size, the original ten-point BWQ ratings were grouped into a three-level scale to ensure more robust and reliable statistical comparisons.

As illustrated in Figure 5b, each cueing modality reported over 30% of the initially extracted 81 features as statistically significant. Significance was determined based on the presence of at least one statistically meaningful difference in the three post-hoc Mann–Whitney comparisons. As discussed in prior studies (Refs. 38, 39), this result suggests that many physiological features are effective at distinguishing between a relaxed (baseline) state and a cognitively engaged state. However, fewer features are sensitive enough to differentiate reliably across all three workload levels.

Figure 4a further contextualizes this by showing that, despite variability in the perceived MWL distributions across modal-

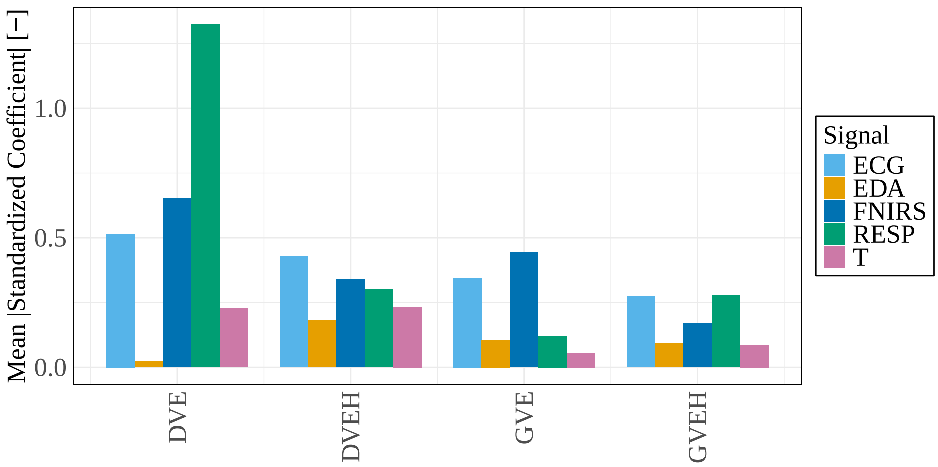


Figure 6: Physiological signals importance in multivariate model regression per modality.

ities, the absolute median BWQ ratings typically fall within a narrow range of approximately 3.5. This indicates that while differences in perceived workload exist between conditions, they are relatively subtle. As a result, clustered analysis may not capture fine-grained distinctions in feature responsiveness, even though it remains effective in identifying general trends.

Nonetheless, consistent patterns were observed. Features derived from cardiorespiratory activity, skin temperature, and brain signals showed the highest number of significant effects across conditions. Conversely, EDA showed the fewest significant features, likely because EDA is more responsive to acute stress responses (e.g., "fight-or-flight" response) than to the more sustained cognitive demands emphasized in this experimental task.

Multivariate Analysis

Since Hypothesis *H2* aimed to explore general relationships between physiological signals and MWL, a multivariate approach was necessary to capture the complexity of this multifactorial construct.

To this end, a LASSO + GLMM model was applied with two main goals: (i) to evaluate the model's ability to explain variance in the dataset, and (ii) to identify which physiological signals contributed most significantly to the regression.

Table 3 presents the model results. Despite relatively subtle variations in MWL, the model explained up to 60% of the variance (R^2_{marg}), particularly in the most demanding condition (**DVE**). The gap between marginal and conditional R^2 values (ΔR^2) highlights the importance of inter-subject variability, reinforcing the idea that MWL is highly dependent on both individual and contextual factors. By modeling subject-specific effects as random factors, the conditional R^2 (R^2_{cond}) increased to values between 0.85 and 0.88, indicating that a physiologically-based model accounting for subjectivity can explain the majority of the variance. As expected, the limited sample size resulted in moderate predictive performance. RMSE remained around 2, and LOSO cross-validation R^2 (R^2_{LOSO}) peaked at 0.4 in the **DVE** condition. Nevertheless, these values are promising considering the dataset size and suggest that larger datasets could lead to substantially improved predictive accuracy. Notably, the best prediction performance occurred in the highest workload condition, in line with previous findings (Ref. 26) showing that more demanding tasks produced more reliable physiological responses for MWL estimation.

Figure 6 shows the mean standardized coefficients of the features selected through LASSO regularization, quantifying the contribution of each signal type to the regression. Consistent with the one-way analysis, EDA was the least influential across all modalities. In contrast, signals more strongly linked to autonomic activation in response to external workload demands, such as cardiorespiratory features and fNIRS, provided the highest contributions. These effects were most pronounced in the **DVE** condition, followed by **DVEH** and **GVE**, with the lowest contributions observed in **GVEH**. Interestingly, this trend contrasts with the subjective MWL levels reported via the BWQ in Figure 4a, suggesting a possible divergence between perceived workload and physiological indicators in specific cueing conditions.

Limitations and Future Directions

This study has several limitations, primarily the small sample size, which reduces the generalizability of the findings, as shown by the modest LOSO cross-validation performance in the multivariate regression. However, the main goal was to provide foundations for future experimental campaigns that can build on these results and re-evaluate the proposed hypotheses with a larger and more diverse participant pool. As noted earlier, the hovering task stimulated varying MWL levels, but the granularity of these differences was limited. Future studies should incorporate more distinct workload conditions, for example, through external disturbances or secondary tasks, to produce stronger physiological responses and better validate the observed trends. To address these limitations, future work should increase the number of participants and improve task difficulty differentiation. Additionally, integrating behavioral and performance metrics could strengthen model predictions and support the development of a hybrid behavioral-physiological framework for MWL estimation.

CONCLUSION

This study explored the relationship between physiological signals and MWL in a motion-based VR helicopter hover task, using a high-fidelity flight model (Ref. 37). The experiment involved four cueing modalities based on the ADS-33 MTE hover task (Ref. 36), combining visual conditions (good and degraded) with or without haptic feedback. ECG, EDA, respiration, fNIRS, and skin temperature were recorded and analyzed in relation to perceived MWL. Key findings include:

1. Full-body haptic feedback significantly reduced perceived MWL in both visual conditions, with the **DVEH** condition lowering workload levels to those comparable with **GVE**. This confirms the benefit of multimodal cueing in challenging flight scenarios.
2. Physiological signals, particularly those reflecting autonomic activity (cardio-respiratory and brain-related), were sensitive to changes in MWL, supporting their relevance for cognitive state monitoring in realistic operational contexts (Ref. 26).
3. The multivariate LASSO + GLMM model explained up to 88% of the variance when accounting for individual differences, though limited generalizability was observed due to the small sample size.
4. Prediction performance was highest under the most demanding condition (**DVE**), indicating stronger physiological responses under higher cognitive load.

Overall, the results highlight the potential of haptic feedback to reduce MWL in safety-critical scenarios like helicopter hovering. Physiological monitoring proved to be a reliable approach for assessing MWL, and future work should explore hybrid behavioral-physiological models to enhance prediction accuracy and support integration into next-generation flight systems.

REFERENCES

1. Baumgartner, H. M., DiDomenica, R., Durham, J., and Hu, P. T., "Human Factors in Helicopter Air Ambulance Operations Annotated Bibliography (2014 – 2022)," Technical report, February 2023.
2. GAO, "National Guard Helicopters: Additional Actions Needed to Prevent Accidents and Improve Safety," Technical report, 2023.
3. Longo, L., Wickens, C. D., Hancock, G., and Hancock, P. A., "Human mental workload: A survey and a novel inclusive definition," *Frontiers in Psychology*, Vol. 13, 2022, pp. 883321. doi: 10.3389/fpsyg.2022.883321
4. Wickens, C. D., "Multiple Resources and Mental Workload," *Human Factors*, Vol. 50, (3), 2008, pp. 449–455. doi: 10.1518/001872008X288394
5. Sweller, J., "Cognitive load during problem solving: Effects on learning," *Cognitive Science*, Vol. 12, (2), 1988, pp. 257–285. doi: 10.1016/0364-0213(88)90023-7
6. Alsuraykh, N. H., Wilson, M. L., Tennent, P., and Sharples, S., "How Stress and Mental Workload Are Connected," Proceedings of the 13th EAI International Conference on Pervasive Computing Technologies for Healthcare, 2019.
7. Hoang, I., Ranchet, M., Derollepot, R., Moreau, F., and Paire-Ficout, L., "Measuring the Cognitive Workload During Dual-Task Walking in Young Adults: A Combination of Neurophysiological and Subjective Measures," *Frontiers in Human Neuroscience*, Vol. 14, 2020. doi: 10.3389/fnhum.2020.592532
8. Debie, E. *et al.*, "Multimodal Fusion for Objective Assessment of Cognitive Workload: A Review," *IEEE Transactions on Cybernetics*, Vol. PP, September 2019, pp. 1–14.
9. Rojas, R. F., Debie, E., Fidock, J., Barlow, M., Kasmarik, K., Anavatti, S., Garratt, M., and Abbass, H., "Electroencephalographic Workload Indicators During Teleoperation of an Unmanned Aerial Vehicle Shepherding a Swarm of Unmanned Ground Vehicles in Contested Environments," *Frontiers in Neuroscience*, Vol. 14, 2020. doi: 10.3389/fnins.2020.00040
10. Endsley, M. R., "Toward a Theory of Situation Awareness in Dynamic Systems," *Human Factors*, Vol. 37, (1), 1995, pp. 32–64. doi: 10.1518/001872095779049543
11. Vidulich, M. A. and Tsang, P. S., "Mental Workload and Situation Awareness," *Handbook of Human Factors and Ergonomics*, edited by G. Salvendy, Wiley, fourth edition, 2012, pp. 243–273. doi: 10.1002/9781118131350.ch8
12. Neigel, A. R., Dever, D. A., Claypoole, V. L., and Szalma, J. L., "Task Engagement and the Vigilance Decrement Revisited: Expanding Upon the Work of Joel S. Warm Using a Semantic Vigilance Paradigm," *Human Factors*, Vol. 61, (3), 2019, pp. 462–473. doi: 10.1177/0018720819835086
13. Hoc, J.-M., "From human – machine interaction to human – machine cooperation," *Ergonomics*, Vol. 43, (7), 2000, pp. 833–843. doi: 10.1080/001401300409044
14. Kolling, A., Walker, P., Chakraborty, N., Sycara, K., and Lewis, M., "Human Interaction With Robot Swarms: A Survey," *IEEE Transactions on Human-Machine Systems*, Vol. 46, (1), 2016, pp. 9–26. doi: 10.1109/THMS.2015.2480801

15. Baldwin, C. L., Spence, C., Bliss, J. P., Brill, J. C., Wogalter, M. S., Mayhorn, C. B., and Ferris, T. K., "Multimodal Cueing: The Relative Benefits of the Auditory, Visual, and Tactile Channels in Complex Environments," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 56, (1), 2012, pp. 1431–1435. doi: 10.1177/1071181312561404

16. Schmerwitz, S., Müllhäuser, M., Wartmann, J., and Ernst, J. M., "Multimodal cueing and short-term pilot interaction during helicopter low-level operation under automatic trajectory following," Virtual, Augmented, and Mixed Reality (XR) Technology for Multi-Domain Operations III, Vol. 12125, 2022. doi: 10.1117/12.2618877

17. Duxbury, D. J., Paton, N. W., and Keane, J. A., "Dynamic decision making for situational awareness using drones: Requirements, identification and comparison of decision support methods," *Expert Systems with Applications*, Vol. 252, 2024, pp. 124057. doi: 10.1016/j.eswa.2024.124057

18. White, T. L. and Hancock, P., "Specifying advantages of multi-modal cueing: Quantifying improvements with augmented tactile information," *Applied Ergonomics*, Vol. 88, 2020, pp. 103146. doi: 10.1016/j.apergo.2020.103146

19. Reynal, M., Imbert, J. P., Arico, P., Toupillier, J., Borghini, G., and Hurter, C., "Audio Focus: Interactive spatial sound coupled with haptics to improve sound source Address in poor visibility," *International Journal of Human-Computer Studies*, Vol. 129, 2019, pp. 116–128. doi: 10.1016/j.ijhcs.2019.04.001

20. Deldycke, P. J., Van Baelen, D., Pool, D. M., van Paassen, M. M., and Mulder, M., "Design and Evaluation of a Haptic Aid for Training of the Manual Flare Manoeuvre," AIAA 2018-0113, Proceedings of the AIAA SciTech Forum, Jan 8–12, 2018. doi: 10.2514/6.2018-0113

21. Van Baelen, D., Ellerbroek, J., van Passen, M. M., and Mulder, M., "Design of a Haptic Feedback System for Flight Envelope Protection," *Journal of Guidance, Control, and Dynamics*, Vol. 43, (4), 2020, pp. 700–714. doi: 10.2514/1.G004596

22. Malik, H. A., Rasool, S., Maqsood, A., and Riaz, R., "Effect of Haptic Feedback on Pilot/Operator Performance During Flight Simulation," *Applied Sciences*, Vol. 10, (11), 2020. doi: 10.3390/app10113877

23. Olivari, M., Nieuwenhuizen, F. M., Bülthoff, H. H., and Pollini, L., "Pilot Adaptation to Different Classes of Haptic Aids in Tracking Tasks," *Journal of Guidance, Control, and Dynamics*, Vol. 37, (6), 2014, pp. 1741–1753. doi: 10.2514/1.G000534

24. McGrath, B. J., Tactile Instrument for Aviation, Technical report, NAMRL Monograph 49, Naval Aerospace Medical Research Laboratory, Pensacola, FL, 2000.

25. Morcos, M. T., Fishman, S. M., Cocco, A., Saetti, U., Berger, T., Godfroy-Cooper, M., and Bachelder, E. N., "Full-Body Haptic Cueing Algorithms for Augmented Pilot Perception in Degraded/Denied Visual Environments," Proceedings of the 79th Annual Forum of the Vertical Flight Society, May 15-18, 2023. doi: 10.4050/F-0079-2023-18072

26. Luzzani, G., Morcos, M. T., Fischer, M. R., U., S., Guglieri, G., Demarchi, D., and Buraioli, "Multimodal Cueing in Attitude Tracking Tasks: Toward the Prediction of Pilot Cognitive Workload via Physiological Measurements," Proceedings of the 50th European Rotorcraft Forum, Sep 10–12, 2023.

27. Morcos, M. T., Fishman, S. M., Saetti, U., Berger, T., Godfroy-Cooper, M., and Bachelder, E. N., "Spatial Audio Cueing Algorithms for Augmented Pilot Perception in Degraded/Denied Visual Environments," Proceedings of the 49th European Rotorcraft Forum, Sep 5–7, 2023.

28. Oskarsson, P. A., Eriksson, L., and Carlander, O., "Enhanced Perception and Performance by Multimodal Threat Cueing in Simulated Combat Vehicle," *Human Factors*, Vol. 54, (1), 2011, pp. 122–137. doi: 10.1177/0018720811424895

29. Longo, L., "Subjective Usability, Mental Workload Assessments and Their Impact on Objective Human Performance," Human-Computer Interaction – INTERACT 2017, edited by M. Kurosu, Vol. 10514, Lecture Notes in Computer Science, 2017.doi: 10.1007/978-3-319-67684-5_13

30. Butmee, T., Lansdown, T. C., and Walker, G. H., "Mental Workload and Performance Measurements in Driving Task: A Review Literature," Proceedings of the 20th Congress of the International Ergonomics Association (IEA 2018), edited by D. Schmorrow and C. Fidopiastis, Vol. 823, Advances in Intelligent Systems and Computing, 2019. doi: 10.1007/978-3-319-96074-6_31

31. Klyde, D. H., Lampton, A., Mitchell, D. G., Berka, C., and Rhinehart, M., "A New Approach to Aircraft Handling Qualities Prediction," Proceedings of the AIAA Scitech 2021 Forum (Virtual), 2021. doi: 10.2514/6.2021-0178

32. Lampton, A. K., Klyde, D. H., Musso, D., Mitchell, D., and Rhinehart, M., "Further Development of an Approach to Aircraft Handling Qualities Prediction," Proceedings of the AIAA Scitech 2024 Forum (Orlando, FL), 2024. doi: 10.2514/6.2024-2479

33. Dell'Agnola, F., Pascucci, F., Priore, P. C., Jacovitti, G., Mennini, A., and Placidi, G., "Machine-Learning-Based Monitoring of Cognitive Workload in Rescue Missions With Drones," *IEEE Journal of Biomedical and*

Health Informatics, Vol. 26, (9), 2022, pp. 4751–4762. doi: 10.1109/JBHI.2022.3186625

34. Wei, W., Fu, X., Zhu, Y., Lu, N., and Ma, S. E., "Classification and Prediction of Driver's Mental Workload Based on Long Time Sequences and Multiple Physiological Factors," *IEEE Access*, Vol. 11, 2023, pp. 81725–81736. doi: 10.1109/ACCESS.2023.3300721

35. Luzzani, G., Buraioli, I., Demarchi, D., and Guglieri, G., "A review of physiological measures for mental workload assessment in aviation: A state-of-the-art review of mental workload physiological assessment methods in human-machine interaction analysis," *The Aeronautical Journal*, 2023, pp. 1–22. doi: 10.1017/aer.2023.101

36. Anon, "Aeronautical design standard performance specification handling qualities requirements for military rotorcraft," Technical report, U.S. Army Aviation and Missile Command Aviation Engineering Directorate, Redstone Arsenal, AL, 2000.

37. Morcos, M. T., Saetti, U., Geiger, D. H., Kubik, S. T., Breed, A. R., Crane, C. J., Luzzani, G., Fischer, M. R., Jun, D., and Gary, E., "Full-Body Haptic Cueing for Augmented Pilot Perception in Position/Velocity Tracking Tasks," Proceedings of the 81st Annual Forum of the Vertical Flight Society, 2025. doi: 10.4050/F-0081-2025-0136

38. Luzzani, G., Buraioli, I., Guglieri, G., and Demarchi, D., "EDA, PPG and Skin Temperature as Predictive Signals for Mental Failure by a Statistical Analysis on Stress and Mental Workload," *IEEE Open Journal of Engineering in Medicine and Biology*, Vol. 6, 2025, pp. 248–255. doi: 10.1109/OJEMB.2024.3515473

39. Luzzani, G., Pogliano, M., Buraioli, I., Colavincenzo, M., Martorana, S., Guglieri, G., and Demarchi, D., "ECG, Respiration, fNIRS, and Eye Tracking for Stress and Mental Workload Monitoring in Human-Machine Interaction," *IEEE Access*, 2025, pp. 1–1. doi: 10.1109/ACCESS.2025.3588384

40. Roscoe, A. H. and Ellis, G. A., "A Subjective Rating Scale for Assessing Pilot Workload in Flight: A Decade of Practical Use," RAE Technical Report 90019, Royal Aerospace Establishment, 1990.

41. Hicks, J. S., Durbin, D. B., Morris, A. W., and Davis, B. M., "A Summary of Crew Workload and Situational Awareness Ratings for U.S. Army Aviation Aircraft," Technical Report ARL-TR-6955, U.S. Army Research Laboratory, Aberdeen Proving Ground, MD, 2014.

42. Saetti, E. J., U., , and Horn, J. F., "Flight Dynamics and Control of an eVTOL Concept Aircraft with a Propeller-Driven Rotor," *Journal of the American Helicopter Society*, **67**, 032012 (2022). doi: https://doi.org/10.4050/JAHS.67.032012

43. Berger, T., Tischler, M. B., and Horn, J. F., "Outer-Loop Control Design and Simulation Handling Qualities Assessment for a Coaxial-Compound Helicopter and Tiltrotor," Proceedings of the 77th Annual Forum of the Vertical Flight Society, Virtual, Oct 5-8, 2020.

44. Berger, T., Tischler, M. B., and Horn, J. F., "High-Speed Rotorcraft Pitch Axis Response Type Investigation," Proceedings of the 77th Annual Forum of the Vertical Flight Society, Virtual, May 10-14, 2021.

45. Saetti, U., Horn, J. F., and Berger, T., "On the Effects of Rotor Induced Vibrational Stability on Helicopter Flight Dynamics," *CEAS Aeronautical Journal,* Article in Advance, 2024. doi: https://doi.org/10.1007/s13272-024-00718-w

46. Saetti, U., Rogers, J. D., Alam, M., and Jump, M., "Tau Theory-Based Flare Control in Autonomous Helicopter Autorotation," *Aerospace*, Vol. 33, (1), 2024. doi: https://doi.org/10.3390/aerospace11010033

47. Saetti, U. and Bugday, B., "Tiltrotor Simulations with Coupled Flight Dynamics, State-Space Aeromechanics, and Aeroacoustics," *Journal of the American Helicopter Society*, Vol. 69, (1), 2024. doi: https://doi.org/10.4050/JAHS.69.012003

48. Saetti, U. and Guner, F., "Interactional Aerodynamics Modeling and Flight Control Design of Multi-Rotor Aircraft," Proceedings of the 6th Decennial Aeromechanics Specialits' Conference, Santa Clara, CA, Feb 6–8, 2024.

49. Sahani, N. A. and Horn, J. F., "Adaptive Model Inversion Control of a Helicopter with Structural Load Limiting," *Journal of Guidance, Control, and Dynamics*, Vol. 29, (2), 2006, pp. 411–420. doi: https://doi.org/10.2514/1.13391

50. Theron, J. P., Horn, J. F., Wachspress, D. A., and Enciu, J., "Nonlinear dynamic inversion control for urban air mobility aircraft with distributed electric propulsion," Proceedings of the 76th Annual Forum of the Vertical Flight Society, October 5–8, 2020. doi: https://doi.org/10.4050/F-0076-2020-16396

51. Walter, A., Mckay, M., Niemiec, R., Gandhi, F., and Berger, T., "Hover Dynamics and Flight Control of a UAM-Scale Quadcopter With Hybrid RPM and Collective Pitch Control," *Journal of the American Helicopter Society*, **68**, 022012 (2023). doi: https://doi.org/10.4050/JAHS.68.022012

52. Enciu, J., Horn, J. F., and Langelaan, J. W., "Formation Control of a Rotorcraft Multilift System," *Journal of the American Helicopter Society*, **62**, 042011 (2022). doi: https://doi.org/10.4050/JAHS.62.042011

53. Spires, J. M. and Horn, J. F., "Multi-Input Multi-Output Model-Following Control Design

Methods for Rotorcraft," *Journal of the American Helicopter Society*, **66**, 032002 (2021). doi: https://doi.org/10.4050/JAHS.66.032002

54. Scaramal, M. and Horn, J. F., "Load Alleviation on Compound Rotorcraft Using Load Feedback and Extremum Seeking Control," *Journal of Guidance, Control, and Dynamics*, Vol. 46, (12), 2023. doi: https://doi.org/10.2514/1.G007192

55. Saetti, U., "Dynamic Inversion Flight Control Laws for Autonomous Transition of Tilt-Rotor/Wing Aircraft," *Journal of the American Helicopter Society*, Article in Advance, 2025.

56. Nakagawa, S. and Schielzeth, H., "A general and simple method for obtaining R^2 from generalized linear mixed-effects models," *Methods in Ecology and Evolution*, Vol. 4, (2), 2013, pp. 133–142. doi: 10.1111/j.2041-210X.2012.00261.x

CLASSIFICATION & COPYRIGHT

APPENDIX

CLS Perceived SA

As discussed in the *Introduction*, situational awareness (SA) and mental workload (MWL) are closely related cognitive constructs. While the primary focus of this study was on cognitive load, subjective SA was also evaluated to align with literature suggesting that increasing the number of sensory channels, particularly under high-demand conditions, can enhance SA. To assess this, the China Lake Situational Awareness scale (CLS) (Ref. 41) was employed. This 5-point scale ranges from 1 (very good SA) to 5 (extremely poor SA perception).

Figure 7 presents the SA scores reported by the 14 participants across the different conditions. Consistent with trends observed in MWL assessments and prior studies, haptic feedback led to improved SA, particularly under degraded visual conditions. Specifically, the median difference between the **GVE** and **GVEH** conditions was small (0.5), whereas a more pronounced difference (1.5) and distinct distribution patterns were observed between the **DVE** and **DVEH** conditions, reinforcing the positive impact of haptic feedback on SA in challenging environments.

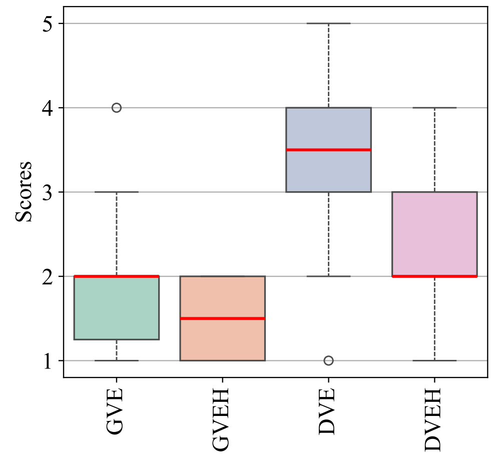


Figure 7: CLS ratings output.

Multivariate Regression Prediction Results

The results of the multivariate regression analysis are presented in the following figures. These results were obtained using Leave-One-Subject-Out (LOSO) cross-validation. Figures 9 to 12 show the predicted versus actual MWL (BWQ) scores for each participant across the four experimental conditions. Each data point represents a single trial, with participants distinguished by color, as indicated in the legend (Figure 8). This visualization highlights the model's ability to generalize across individuals and conditions.

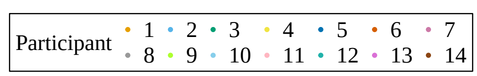


Figure 8: Legend for true vs. predicted MWL (BWQ) plots.

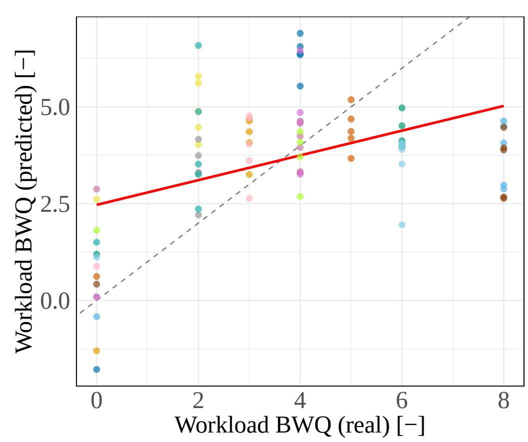


Figure 9: True vs. predicted MWL (BWQ) in **GVE**.

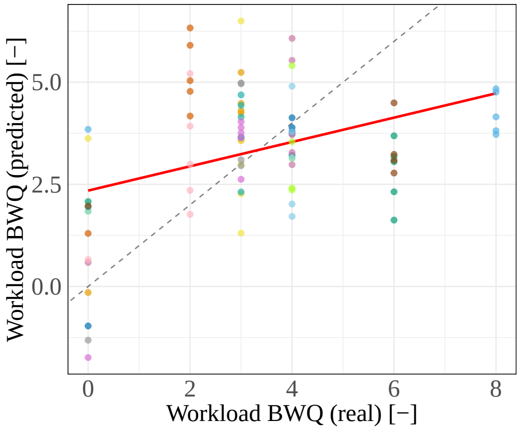


Figure 10: True vs. predicted MWL (BWQ) in **GVEH**.

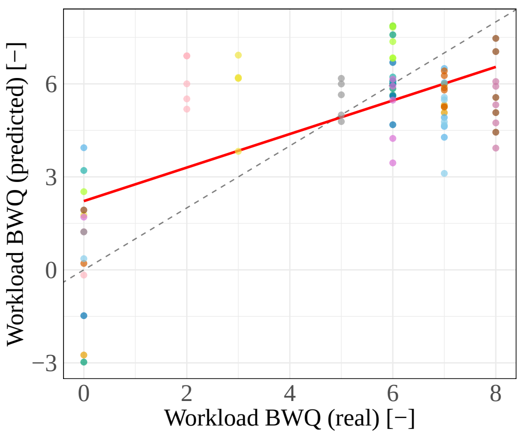


Figure 11: True vs. predicted MWL (BWQ) in **DVE**.

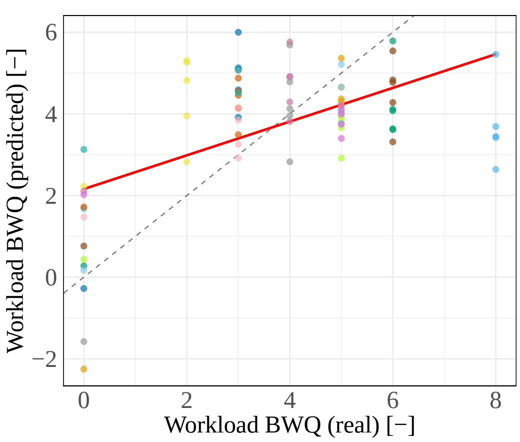


Figure 12: True vs. predicted MWL (BWQ) in **DVEH**.