Proceedings of the 35th European Safety and Reliability & the 33rd Society for Risk Analysis Europe Conference
Edited by Eirik Bjorheim Abrahamsen, Terje Aven, Frederic Bouder, Roger Flage, Marja Ylönen

doi: 10.3850/978-981-94-3281-3_ESREL-SRA-E2025-P4944-cd

Comparison Between Baseline and Recovery ECG Data in an Experiment with a Unmanned Aerial Vehicle Human-Machine Interface

Andrew G. P. Sarmento

Competence Center in Manufacturing (CCM), Mechanical Engineering Division of the Aeronautics Institute of Technology (ITA), Brazil. E-mail: andrewgp@ita.br

Ivan S. Rehder

Competence Center in Manufacturing (CCM), Mechanical Engineering Division of the Aeronautics Institute of Technology (ITA), Brazil. E-mail: ivani@ita.br

Moacyr M. Cardoso Júnior

Competence Center in Manufacturing (CCM), Mechanical Engineering Division, Aeronautics Institute of Technology (ITA), Brazil. E-mail: moacyr@ita.br

Petter Krus

Division of Fluid and Mechatronic Systems (FLUMES), Linköping University (LiU), Sweden. E-mail: petter.krus@liu.se

Emilia Villani

Department of Mechatronics, Mechanical Engineering Division of the Aeronautics Institute of Technology (ITA), Brazil. E-mail: evillani@ita.br

This study investigates the cardiovascular impact of cognitive load on Unmanned Aerial Vehicle (UAV) operators through a comparative analysis of baseline and recovery Electrocardiogram (ECG) data. Twenty-four participants from diverse aviation backgrounds underwent simulated UAV missions with varying Human-Machine Interface (HMI) configurations, including voice command and multi-operator settings. ECG data was collected pre- and post-simulation to assess physiological responses. Statistical analysis, including Z-transform normalization and Student's t-test, was employed to examine heart rate variability (HRV) differences. The results reveal significant changes in HRV previously and after the simulations, highlighting the substantial impact of cognitive load on cardiovascular function. These findings underscore the importance of monitoring physiological responses to optimize human-machine interface design and mitigate operator workload in demanding UAV operations.

Keywords: Cognitive Load, ECG Analysis, UAV Operator, HMI, T-test, Z-transform.

1. Introduction

Occupational health and human factors research increasingly rely on physiological measures to quantify mental workload Abdul Rahman et al. (2018); Mansikka et al. (2019); Charles and Nixon (2019); Sriranga et al. (2023); Luzzani et al. (2024). Among these measures, the Electrocardiogram (ECG) stands out for its ability to illuminate the intricate relationship between cognitive demands and cardiovascular responses Ziegler et al. (2016); Mühlhausen et al. (2019); Wang et al. (2024). By analyzing ECG baseline and recovery phases, researchers gain insights into how individuals navigate and recuperate from mental stressors, offering valuable implications for occupational health and performance enhancement Nixon and Charles (2017); Radüntz et al. (2020); Wang et al. (2024).

Early investigations into the application of ECG in assessing cognitive tasks laid the foundation for understanding its utility in capturing physiological changes pre- and post-exposure to cognitive stimuli Willemain (1978). While not initially focused on mental workload, pioneering studies like those by Willemain (1978) paved the way for subsequent research exploring the dynamic interplay between cognitive stressors and cardiovascular responses Willemain (1978). This led to the integration of ECG monitoring in contemporary studies on mental workload assessment, spanning diverse contexts from aviation to healthcare Nixon and Charles (2017); Abdul Rahman et al. (2018); Sarmento et al. (2023); Wang et al. (2024).

Comparing baseline and recovery phases of ECG signals offers valuable insights into the cardiovascular system's adaptability to cognitive demands Ziegler et al. (2016); Mansikka et al. (2019); Wang et al. (2024). Variations in heart rate (HR), heart rate variability (HRV), and other ECG-derived parameters during these phases serve as key indicators of the impact of mental workload Abdul Rahman et al. (2018); Radüntz et al. (2020); Wang et al. (2024). Such research not only enhances our understanding of the physiological underpinnings of stress and cognitive performance but also holds implications for designing interventions to mitigate the adverse effects of mental workload across various professional domains Nixon and Charles (2017); Radüntz et al. (2020).

This study builds upon ongoing experimental efforts previously outlined in related works, aiming to evaluate cognitive load in the context of enhancing the human-machine interface (HMI) for Unmanned Aerial Vehicles (UAVs) Silva et al. (2022); Sarmento et al. (2023); Rehder et al. (2023a,b). Within the experimental framework, a diverse set of data points was collected, including electrocardiogram (ECG) readings. Expanding on findings from a prior study involving 24 participants Sarmento et al. (2023), this paper centers on the comparative analysis of ECG data during the baseline and recovery phases to identify which phase more accurately represents the normal relaxed ECG pattern.

2. Experiment Procedure

The experimental process, conducted in a fully simulated environment, was divided into subsets of scenario design, experiment execution, and processing of collected data.

2.1. *Scenario Design*

As part of this research, the work Silva et al. (2022) describes a critical operational scenario for UAV missions, emphasizing the need for effective human-machine interaction during high-stress combat situations. It defines the context of the mission as seen in Sarmento et al. (2023), involving the UAV being deployed in a simulated enemy combat scenario. This configuration is designed to evaluate pilot situational awareness, decision-making, and overall system performance under pressure, highlighting the importance of a well-designed HMI to increase operational effectiveness and safety in such demanding environments. The UAV ground station simulator can be shown in Fig. 1, and a complete description of the controls and functions used in this experiment is presented in Sarmento et al. (2023).

Fig. 1. Ground station simulator

2.2. *Experiment Execution*

The procedure for collecting the first (before the simulation) and second (after the simulation) baseline signals during the experiment was carefully structured to ensure accurate measurement of the operator's mental workload. After the initial training and once the operator felt prepared for the experiment, informed consent was obtained. Following this, electrocardiogram (ECG) sensors were attached to the operator to monitor physiological responses. A baseline signal, capturing 5 minutes of data, was collected to establish the operator's resting state before the commencement of the simulated flights Sarmento et al. (2023).

The experiment entailed two simulated flights, each under different operational settings—either with a single operator, two operators, or a single operator utilizing voice commands Sarmento et al. (2023). These flights were divided into several segments, including navigating between villages and identifying targets, with specific tasks assigned to each segment Sarmento et al. (2023).

After completing the second flight, a second baseline or recovery signal was collected to gauge the post-experiment mental workload, again capturing 5 minutes of data Sarmento et al. (2023). This approach allowed for a comparison between the operator's physiological state before and after the experiment, providing insights into the mental workload imposed by the UAV operation under different HMI configurations Sarmento et al. (2023). The experiment involved 24 volunteers, both men and women, from various backgrounds, including pilots and students from the Brazilian Air Force, as well as individuals with experience in piloting or aeronautical engineering, to ensure a diverse set of data for analysis Sarmento et al. (2023).

2.3. *Data Processing*

During the data processing and classification process, related work resulted in the assessment of mental demand during the experiment, which is extremely important to verify how the volunteer reacts between baseline and recovery during the experiment.

Rehder et al. (2023a) investigates the influence of the human-machine interface (HMI) on decision-making and command and control in unmanned aerial systems (UAS), using an HMI prototype in combat scenarios together with performance data, subjective mental workload data (NASA-TLX and ISA), and physiological measurements (Eye Tracker, GSR, and ECG).

In another study, Rehder et al. (2023b), evaluated the usability of an interactive system for UAS using the Systems Usability Scale (SUS). An average SUS score of 67.4 was found, indicating a reasonably positive perception of usability. The study emphasizes the importance of considering users' history when controlling the UAS for usability improvements.

Data processing from the experiment described in Sarmento et al. (2023) involved assessing the operator's mental workload through the use of electrocardiogram (ECG) and instantaneous self-assessment (ISA) sensors as sources of physiological and subjective data, respectively. A scientific software was then used to analyze the ECG signal, focusing on RR intervals (space between two consecutive R waves = one heartbeat) for pilot data standardization and ECG feature collection. The variance in both data sets was analyzed using Analysis of Variance (ANOVA), which considered factors such as flight number, flight configuration, and mission segment.

3. Z-Transform

The Z-transformation, commonly known as Z-score standardization, is a statistical technique used to standardize the values of a dataset. It transforms the data so that the resulting distribution has a mean of zero and a standard deviation of one. This process makes different datasets comparable and is widely used in various fields, including psychology, education, and other social sciences, as well as in machine learning data preprocessing Snedecor and Cochran (1989).

The origin of Z-score standardization can be traced back to the work of Karl Pearson, a pioneering figure in the early development of statistics, around the late 19th and early 20th centuries. While the specific term "Z-score" might not have been used by Pearson himself, the concept of standardizing variables to compare them on a common scale is deeply rooted in his and his contemporaries' work on standard deviation and normalization of distributions Snedecor and Cochran (1989).

A specialized form of Z-transformation for standardizing physiological data is shown in Nählinder (2002). This approach adjusts individual participant data to match the group's average values and standard deviation, facilitating homogeneous analysis across participants. Unlike a normal Z-transformation, which standardizes data to a mean of zero and a standard deviation of one for all participants, this method ensures each participant's data aligns with the group's overall mean and variance, thereby maintaining individual variation while enabling direct interpretability of the standardized values Nählinder (2002).

As described in Nählinder (2002) and shown in Eq. (1), The original value, x_i is replaced with x_j. The equation terms are: $\overline{x}$ is the individual average, s is the individuals' standard deviation, μ is the entire group average, and σ is the group standard deviation.

$$x_j = \frac{x_i - \overline{x}}{s}\sigma + \mu \tag{1}$$

4. STUDENT'S T-TEST

The Student's t-test is a statistical method used to determine if the difference in outcomes between two groups is significant. This test is characterized by its reliance on the Student's t-distribution for the test statistic under the assumption of the null hypothesis. Typically applied when a test statistic would adhere to a normal distribution if a certain scale factor was known (which often isn't, making it a nuisance parameter), the t-test adjusts for this by estimating the scale factor from the data Kalpić et al. (2011). Under specific conditions, this results in the test statistic following a Student's t-distribution. Its primary use is in comparing the means of two populations to see if they differ significantly Kalpić et al. (2011). This test has been used in some recent works to analyze electrocardiogram (ECG) data:

- In Lahmiri (2023), this analysis is used for distinguishing congestive heart failure (CHF), arrhythmia (ARR), and normal sinus rhythm (NSR). Specifically examined whether short or long-term fluctuations in ECG records could statistically differentiate these heart conditions;
- In Falsaperla et al. (2018) the cardiac effects of spinal muscular atrophy (SMA), comparing electrocardiogram (ECG) parameters between SMA patients and controls. Utilizing the Student t-test for statistical analysis, significant differences were found in PR intervals, P-wave and QRS amplitudes, and heart rates, indicating altered cardiac conduction and function in SMA patients;
- And in Kopeć et al. (2015) assessed ECG interpretation skills among Polish medical students, revealing that competency in ECG interpretation significantly correlates with self-directed learning rather than attendance at regular classes. Through the application of the Student t-test and other statistical analyses, it was found that students in their clinical years showed better competency in interpreting primary ECG parameters compared to their junior counterparts.

5. ECG Baseline and Recovery

5.1. *Data Record*

Each baseline and recovery lasted 5 minutes, and each volunteer performed the two measurements in the order shown in Fig. (2). The ECG data in the experiment were collected using the TEA CAPTIV T-SENS ECG wireless sensor with an acquisition frequency of 256 Hz (Fig. 3).

5.2. *Separation of Data Groups*

The HR data is processed through a sliding window of 30 seconds width, applied at 5-second intervals. This approach effectively filters out the HR fluctuations, prioritizing the observation of slower, significant variations over rapid, short-term changes. For the purposes of this study, fluctuations within a brief span are deemed less relevant, with a focus instead on identifying trends that unfold over a longer duration, specifically between 10 to 30 seconds, Nählinder (2002).

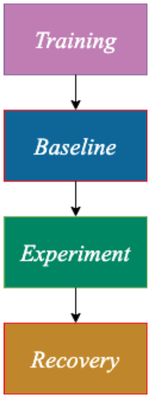

Fig. 2. Order of data collection in the experiment

Fig. 3. TEA CAPTIV T-SENS ECG available in Team (2024)

After observing the quality of data from each operator and evaluating the results of the time measurements between RR peaks, the presence of disturbances in the measurements was noted. These disturbances in the RR averages were caused by noise during collection and/or problems in signal transmission to the recording equipment. Due to these significant discontinuities in the signals, samples from 3 volunteer operators were excluded from the analysis of this work, resulting in the analysis of 21 volunteers. After validating volunteers' signals, 260 seconds of valid signals were established for each baseline and recovery.

To use the Z-transformation, the data were organized into groups, as shown in Fig. 4. All valid data collected are considered to calculate the group average and standard deviation. On the other hand, each pilot's Baseline and Recovery data are considered to calculate individual averages and standard deviations.

After separating the data into groups and calculating the group and individual average and standard deviations, the Z-transform was applied to each baseline and recovery data sequence, resulting in a pair of normalized signals for each pilot. Thus, all data considered for analysis in this work are normalized by the group and can, therefore, be compared with each other.

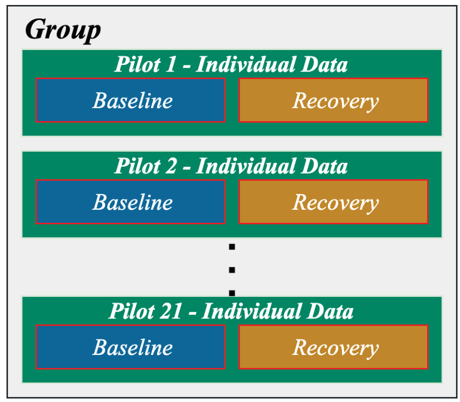

Fig. 4. Total data group

A student's t-test was performed comparing the resulting baseline and recovery values for each pilot to check whether the differences found were consistent to enable the sequence of analyses between these two measurement moments for each pilot in the experiment concerning the general group.

6. Results

To obtain the results of this study, we analyzed heart rate data from 24 volunteers to examine variations between the initial and recovery phases, and as previously explained, only 21 pilots had enough rich data to continue the analysis. After applying the Z-transform, the remaining data was verified through statistical comparison of baseline and recovery data for each pilot, employing T-Tests. These tests revealed no significant difference in the data for Pilot 17, leading to its exclusion. The complete results of the T-Test analysis are detailed in Table 1.

Figures 5 and 6 depict both the individual and average heart rates during the baseline and recovery phases, respectively. In these figures, individual pilot data are shown in lighter shades, whereas

Table 1. Baseline x Recovery T-Test Results for each pilot and for the average.

Column	T-Statistic	P-Value	Sig. Diff.
Pilot 2	-26.59581	$<$1e-5	Yes
Pilot 4	-2.58815	0.01106	Yes
Pilot 5	24.02538	$<$1e-5	Yes
Pilot 6	-5.22428	$<$1e-5	Yes
Pilot 7	4.44674	0.00002	Yes
Pilot 8	3.77078	0.00027	Yes
Pilot 9	17.12348	$<$1e-5	Yes
Pilot 11	-12.20834	$<$1e-5	Yes
Pilot 12	-3.50029	0.00069	Yes
Pilot 13	3.72457	0.00032	Yes
Pilot 14	-4.24375	0.00005	Yes
Pilot 15	13.35279	$<$1e-5	Yes
Pilot 16	5.69791	$<$1e-5	Yes
Pilot 17	0.90796	0.36604	No
Pilot 18	3.47718	0.00075	Yes
Pilot 19	-16.48392	$<$1e-5	Yes
Pilot 20	-44.22309	$<$1e-5	Yes
Pilot 21	13.16447	$<$1e-5	Yes
Pilot 22	6.49732	$<$1e-5	Yes
Pilot 23	20.53134	$<$1e-5	Yes
Pilot 24	-5.48600	$<$1e-5	Yes
Average	2.85011	0.00529	Yes

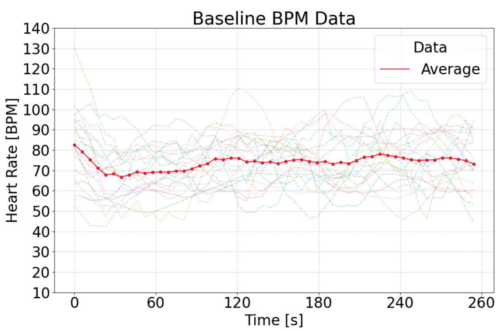

Fig. 5. Average baseline heart rate and individual pilot heart rate plot

average data are in red. A comparative analysis of these figures indicates a broader distribution of heart rates during the baseline phase compared to the recovery phase.

Figure 7 is a box-plot comparison of baseline and recovery heart rate data segmented into 30-second intervals. This visualization highlights the relative stability of recovery data in contrast to

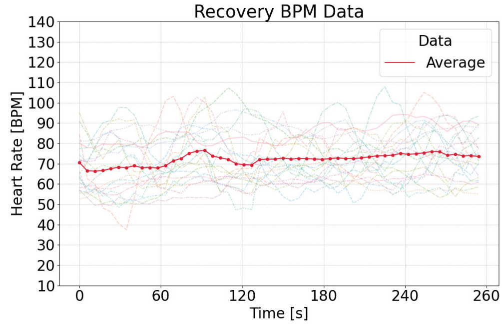

Fig. 6. Average recovery heart rate and individual pilots heart rate plot

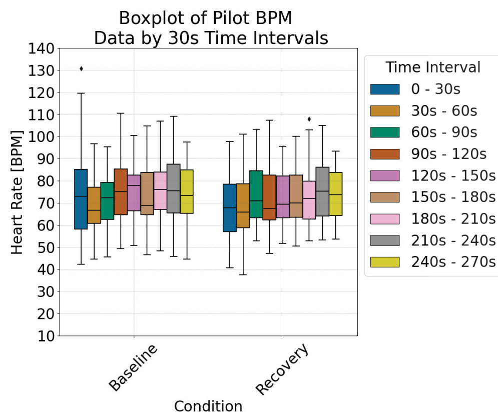

Fig. 7. Boxplot of the baseline and recovery heart rate grouped in time intervals of 30s

baseline data. Additionally, Fig. 8 focuses exclusively on the distribution of average heart rates between the baseline and recovery phases, revealing a lower and more precise distribution during recovery.

Based on the analysis of the results, the primary contribution of this study is the identification of recovery phase data as the most appropriate nominal reference for volunteers. Employing this nominal reference facilitates a more accurate assessment of heart rate variability (HRV) and mental workload fluctuations throughout the entirety of the experimental data.

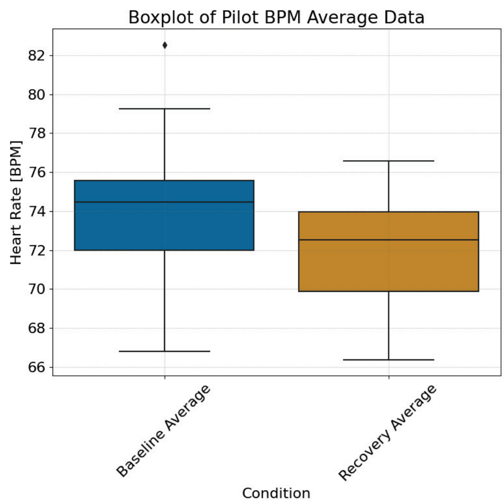

Fig. 8. Boxplot of average the baseline and recovery heart rate

7. Conclusion

This study demonstrates the utility of recovery phase ECG data as a more reliable nominal reference for assessing HRV and mental workload variations in UAV operators. By applying rigorous statistical analyses, including Z-transformation and Student's t-tests, we established that recovery data provide a more stable and precise representation of cardiovascular responses compared to baseline measurements. These findings underscore the importance of focusing on recovery phase data when designing and evaluating HMI systems for UAV operations.

The observed differences in heart rate variability between the baseline and recovery phases may be due to the pilots' emotional responses to the experimental conditions. Factors such as familiarity with experimental procedures, the feeling of participating in a study, or the excitement associated with the experimental equipment could influence heart rate variability, potentially elevating both the frequency and variance of heart rate measurements.

Building on these findings, future research will incorporate electrodermal activity (EDA) responses and eye-tracking data to provide a more comprehensive analysis of physiological and cognitive reactions during flight simulations. Additionally, the study will reassess the initial experiment's results by comparing recovery measurements against baseline data. A new approach in this research will involve dividing the data between men and women, considering potential differences in their natural physiological characteristics. Furthermore, training artificial intelligence models to interpret these multifaceted conditions may offer deeper insights into performance variability under different cognitive loads. This multidisciplinary approach seeks to advance our understanding of human physiological responses in diverse scenarios while leveraging AI to analyze complex data sets with greater precision.

Acknowledgement

Our thanks go to the Aeronautics Institute of Technology - (ITA) within the Brazilian Air Force (FAB) and the Competence Center of Manufacturing - (CCM) for their essential support and resources. We also gratefully acknowledge funding from the Brazilian Funding Authority for Studies and Projects - (FINEP), National Council for Scientific and Technological Development – (CNPq), Swedish-Brazilian Research and Innovation Centre - (CISB), and Svenska Aeroplan Aktiebolaget - (SAAB AB) which is crucial for the success of our research.

References

Abdul Rahman, S. M. B., M. F. Sidik, M. H. I. Mohd Shukri, and M. H. Nazarudin (2018, aug). Controller response behaviour during procedural control with surveillance information. *IOP Conference Series: Materials Science and Engineering 405*(1), 012004.

Charles, R. L. and J. Nixon (2019). Measuring mental workload using physiological measures: A systematic review. *Applied Ergonomics 74*, 221–232.

Falsaperla, R., G. Vitaliti, A. D. Collotta, C. Fiorillo, A. Pulvirenti, S. Alaimo, C. Romano, and M. Ruggieri (2018). Electrocardiographic evaluation in patients with spinal muscular atrophy: A case-control study. *Journal of Child Neurology 33*(7), 487–492. PMID: 29687752.

Kalpić, D., N. Hlupić, and M. Lovrić (2011). *Student's t-Tests*, pp. 1559–1563. Berlin, Heidelberg: Springer Berlin Heidelberg.

Kopeć, G., W. Magoń, M. Hołda, and P. Podolec (2015). Competency in ECG interpretation among medical students. *Medical Science Monitor 21*, 3386–3394.

Lahmiri, S. (2023). A nonlinear analysis of cardiovascular diseases using multi-scale analysis and generalized hurst exponent. *Healthcare Analytics 3*, 100142.

Luzzani, G., I. Buraioli, D. Demarchi, and G. Guglieri (2024). 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 128*(1323), 928–949.

Mansikka, H., K. Virtanen, and D. Harris (2019). Comparison of nasa-tlx scale, modified cooper-harper scale and mean inter-beat interval as measures of pilot mental workload during simulated flight tasks. *Ergonomics 62*(2), 246–254.

Mühlhausen, T., T. Radüntz, A. Tews, H. Gürlük, and N. Fürstenau (2019). Research design to access the mental workload of air traffic controllers. In T. Ahram, W. Karwowski, and R. Taiar (Eds.), *Human Systems Engineering and Design*, Cham, pp. 415–421. Springer International Publishing.

Nixon, J. and R. Charles (2017). Understanding the human performance envelope using electrophysiological measures from wearable technology. *Cognition, Technology & Work 19*, 655–666.

Nählinder, S. (2002, 01). Similarities and differences in psychophysiological reactions between simulated and real air-to-ground missions. *International Journal of Aviation Psychology - INT J AVIAT PSYCHOL 12*, 49–61.

Radüntz, T., C. Meffert, M. Mucke, and G. Grote (2020). Cardiovascular biomarkers for mental workload and stress assessment in laboratory and field settings. *Applied Psychophysiology and Biofeedback 45*(4), 249–257.

Rehder, I., A. Sarmento, A. Russo, M. Cardoso Junior, and E. Villani (2023a). Enhancing the human-machine interface: An analysis of unmanned aerial systems (uas) using the system usability scale (sus). In *Anais Estendidos do XXII Simposio Brasileiro de Fatores Humanos em Sistemas Computacionais*, Porto Alegre, RS, Brasil, pp. 84–88. SBC.

Rehder, I., A. G. P. Sarmento, A. Russo, M. Cardoso Junior, and E. Villani (2023b). A human-machine interface analysis: Under critical operation scenario. In *Anais Estendidos do XXII Simposio Brasileiro de Fatores Humanos em Sistemas Computacionais*, Porto Alegre, RS, Brasil, pp. 27–30. SBC.

Sarmento, A. G. P., I. S. Rehder, T. R. Paula, J. O. B. Possamai, M. M. C. Junior, E. T. Silva, H. C. Marques, and E. Villani (2023). Evaluation of Human Performance in the Operation of a UAV in a Joint Operation Scenario with Troops on the Ground. In *Proceedings of the 33rd European Safety and Reliability Conference*, Number September, pp. 748–755.

Silva, E. T., A. G. P. Sarmento, T. R. Paula, A. S. Oliveira, J. Possamai, H. C. Marques, M. M. Cardoso, and E. Villani (2022). Unmanned Aerial Vehicle – Human Performance Evalution Under Critical Operational Scenario. In *33rd Congress of the International Council of the Aeronautical Sciences, ICAS 2022*, Volume 2, pp. 1357–1366.

Snedecor, G. W. and W. G. Cochran (1989). *Statistical Methods* (8 ed.). Ames, Iowa: Iowa State University Press.

Sriranga, A. K., Q. Lu, and S. Birrell (2023). A systematic review of in-vehicle physiological indices and sensor technology for driver mental workload monitoring. *Sensors 23*(4).

Team, W. S. (2024). T-sens: Electrocardiogram (ecg). Accessed: 2025-01-03.

Wang, P., R. Houghton, and A. Majumdar (2024). Detecting and predicting pilot mental workload using heart rate variability: A systematic review. *Sensors 24*(12).

Willemain, T. R. (1978). Before-and-after experiment with a baseline period of random duration - analysis of a contingent experimental design. *Massachusetts Institute of Technology, Operations Research Center 4*(2), 123–136.

Ziegler, M. D., A. Kraft, M. Krein, L.-C. Lo, B. Hatfield, W. Casebeer, and B. Russell (2016). Sensing and assessing cognitive workload across multiple tasks. In D. D. Schmorrow and C. M. Fidopiastis (Eds.), *Foundations of Augmented Cognition: Neuroergonomics and Operational Neuroscience*, Cham, pp. 440–450. Springer International Publishing.