

Preliminary study of pilot stress and mental workload monitoring through physiological signals

Original

Preliminary study of pilot stress and mental workload monitoring through physiological signals / Luzzani, G; Buraioli, I; Demarchi, D; Guglieri, G. - In: JOURNAL OF PHYSICS. CONFERENCE SERIES. - ISSN 1742-6588. - ELETTRONICO. - 2716:(2024). (Intervento presentato al convegno 13th EASN International Conference on: Innovation in Aviation & Space for opening New Horizons tenutosi a Salerno, Italy nel 05/09/2023 - 08/09/2023) [10.1088/1742-6596/2716/1/012075].

Availability:

This version is available at: 11583/2987012 since: 2024-03-14T21:56:41Z

Publisher:

IOP Publishing Ltd

Published

DOI:10.1088/1742-6596/2716/1/012075

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

PAPER • OPEN ACCESS

Preliminary study of pilot stress and mental workload monitoring through physiological signals

To cite this article: G Luzzani *et al* 2024 *J. Phys.: Conf. Ser.* **2716** 012075

View the [article online](#) for updates and enhancements.

You may also like

- [Continuous decoding of cognitive load from electroencephalography reveals task-general and task-specific correlates](#)
Matthew J Boring, Karl Ridgeway, Michael Shvartsman et al.
- [Using Eye-Tracking for Adaptive Human-Machine Interfaces for Pilots: A Literature Review and Sample Cases](#)
Michalis Xenos, Andreas Mallas and Dimosthenis Minas
- [Heart rate variability predicts decline in sensorimotor rhythm control](#)
Marius Nann, David Haslacher, Annalisa Colucci et al.



PRIME
PACIFIC RIM MEETING
ON ELECTROCHEMICAL
AND SOLID STATE SCIENCE

HONOLULU, HI
Oct 6–11, 2024

Abstract submission deadline:
April 12, 2024

Learn more and submit!



Joint Meeting of

The Electrochemical Society
•
The Electrochemical Society of Japan
•
Korea Electrochemical Society



Preliminary study of pilot stress and mental workload monitoring through physiological signals

G Luzzani¹, I Buraioli², D Demarchi² and G Guglieri¹

¹ Politecnico di Torino, Department of Mechanical and Aerospace Engineering, Corso Duca degli Abruzzi 24, 10129 - Turin, Italy

² Politecnico di Torino, Department of Electronics and Telecommunication, Corso Duca degli Abruzzi 24, 10129 - Turin, Italy

E-mail: gabriele.luzzani@polito.it

Abstract. The aviation industry is moving towards single-pilot operations due to increased operating expenses and a shortage of pilots. The necessity of developing a digital cockpit assistant leads to discovering methods to assess the stress and mental workload of pilots. This study used twenty-eight healthy volunteers to conduct preliminary computerised cognitive tasks while recording their physiological data for PPG, EDA, and temperature under four different stress and workload situations. The results highlight how they are sensible to a binary classification between a relaxed and more cognitively demanding condition.

1. Introduction

Human-machine interaction (HMI) systems have consistently prioritized achieving optimal safety levels. This focus on safety is driven by today's rapid expansion of the technology market, resulting in more frequent scenarios where operators interact with machines equipped with varying degrees of automation. It is crucial to maintain an adequate level of safety during such interactions [1]. One of the most influential sectors in this context is aviation, according to the EASA AI Roadmap [2]. In modern commercial aviation, two pilots share the cockpit and are assigned either Captain (CA) or First Officer (FO) roles. The CA has the primary legal responsibility for the aircraft's safe operation and has the final say in decision-making. In situations where the CA is unable to perform its duties, the FO takes control of the plane and is responsible for ensuring its safe operation. However, with the aviation industry's continuous evolution, there has been a growing emphasis on SPOs. [3]. The term SPO stands for Single Pilot Operation, which denotes a scenario in which one pilot is responsible for flying an aircraft while receiving piloting support services from either equipment or ground operators. [4]. According to the National Aeronautics and Space Administration (NASA) [5], SPOs would make it possible to address two major issues affecting the aviation industry today, namely the declining trend in pilot availability and the rising operational expenses, particularly on short and medium-haul flights [6]. Hence, it's crucial to create a cockpit assistant capable of understanding the cognitive workload of pilots and their capacity to control the aircraft, potentially supplanting the current First Officer role. [7]. However, estimating an operator's mental workload (MWL) is complex. This context is demonstrated by a unique way to assess this cognitive condition, which still needs to be added to the literature. In particular, three strategies are mainly adopted to assess MWL: subjective



questionnaires, behavioural analysis (by exploiting the so-called primary and secondary tasks evaluation methods), and physiological measures [8]. The study of the physiological response to alterations in cognitive load is an increasingly crucial approach, particularly with today's expansion of the biomedical industry. However, a thorough method for evaluating this mental status has yet to be discovered because of the complexity and diverse nature of the cognitive workload, necessitating the development of new approaches to address this problem. Indeed, physiological signals such as cardiorespiratory measures, brain and electrodermal activity, body temperature, and monitoring of eye parameters have been investigated in recent years to study their relationship with MWL. Nevertheless, a robust solution to this issue has yet to be found [9] [11] [13]. Hence, this paper's primary purpose is to examine the correlation between stress variation and mental workload by analyzing three physiological signals: heart rate, skin activity, and body temperature. These particular signals have been chosen due to the non-invasive nature of the sensors, which is vital for potential future applications within SPO's cockpit. Due to the lack of reliable and consistent solutions for this kind of application, we decided to perform this preliminary study by exploiting two specific cognitive tests, Stroop and N-Back, on a population of 32 volunteer subjects. These two tests were adopted to induce different levels of external stress and MWL, respectively. This paper is organized into the following sections: Section 2 introduces a brief state-of-the-art analysis related to stress and MWL assessment in aviation; Section 3 presents our methods adopted in this preliminary study; Section 4 reports the obtained results. Section 5 discusses the conclusions.

2. Background

Due to the disruptive growth of Human-AI interaction systems in the aviation sector and, thus, the necessity of pilots' stress and MWL monitoring [2], it is necessary to clarify what these terms refer to. For many years, the Human Factors and Ergonomics (HFE) field has explored the idea of *mental workload*. This condition has no one single definition because of its complexity and multidimensional character. In fact, over the years, a number of researchers have attempted to establish MWL as a common ground by offering various definitions pertinent to their respective application domains. Nevertheless, a comprehensive study developed by Van Acker et al. [15] tried to find a general framework by defining MWL as a physiological processing state that is experienced subjectively and shows how one's finite and diverse cognitive resources interact with the cognitive task demands they are being exposed to. In particular, as shown in Figure 1, it is clear how MWL is part of the decision-making process, where attention and working memory are significant factors that limit operators in obtaining and interpreting information from the environment. Mental models and goal-directed behavior represent key methods for overcoming these limitations [16]. Furthermore, Figure 1 shows how, in this case, the workload of the pilot and stress conditions can affect the correct perception of the future status of events, inducing errors in the decision-making process. Also, there is no unique definition of *stress*. In fact, this condition is different for different people in different situations and depends on the application field [18]. In the aviation field, a definition can be found by Martins et al., defining stress as "the response of the body to stimuli that affect the normal physiological balance of a person, causing physical, mental or emotional strain" [19]. Assessing this condition can be challenging as what may cause stress for one individual might not be the same for another.

2.1. Stress and mental workload assessment methods

As reported in [10], aviation safety requires a collective effort across the industry due to the complex and subjective nature of mental health. MWL and stress could not be seen as two distinct things. The Debie et al. model [8] views stress as a depleting element that influences an operator's MWL, suggesting a clear cause-and-effect relationship between the two. The literature shows primarily three methods for evaluating pilots' stress and MWL:

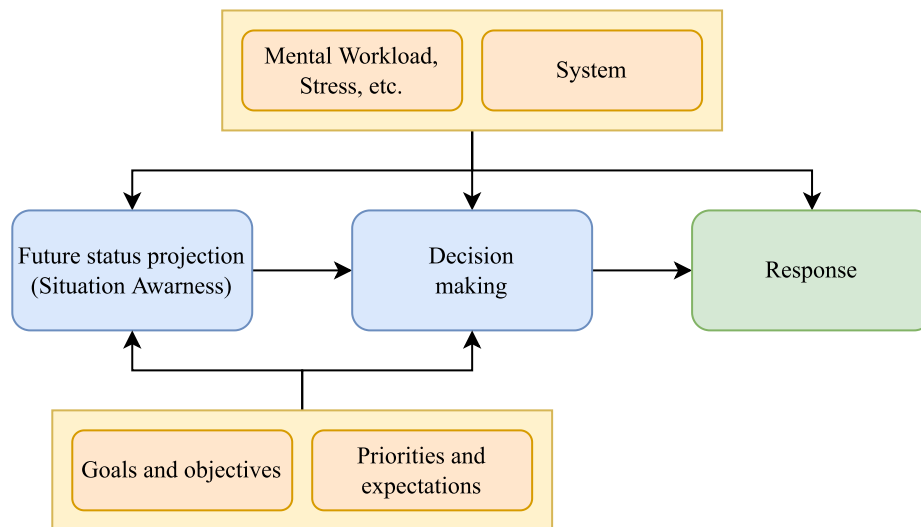


Figure 1: This figure shows the decision-making process. Situation Awareness, Mental Workload, and Stress influence an operator's response to external stimuli [17].

subjective evaluations, behavioral analysis, and physiological measurements. *Post-performance questionnaires*, consisting of ad-hoc questions that gauge a pilot's subjective perception of mental workload and stress, are commonly utilized as an evaluation tool. This approach is favored because it's relatively easy to implement and cost-effective [20]. *Behavioral measures* involve monitoring a pilot's conduct during a flight and comparing them with a predetermined mission plan. By checking the amount of wrong or right actions, it is possible to infer the MWL and stress levels of the operators [21]. We can obtain information about these two cognitive conditions by processing *physiological signals*. It is a method that is gaining more and more interest in recent years thanks to the rapid expansion of biomedical technologies [22]. The literature reveals several indicators that are responsive, albeit distinctively, to fluctuations in both mental workload (MWL) and stress. These signals include heart rate, skin conductance, eye movement, brain activity, respiration, body temperature, muscle activation, and voice patterns, as demonstrated in [12]. However, a complete, reliable physiologically based solution is still lacking [13].

3. Materials and Methods

This section presents the materials and methods adopted in this analysis. In particular, this paper aims to expand the research already shown in the previous work described in [14].

3.1. Test characteristics

In this study, two different computerized cognitive tests widely adopted in this research area were performed to stimulate the participants' different levels of stress and mental workload. The *Stroop test* [23] is a method utilized to replicate external stress situations. It involves displaying a sequence of color-related words written in the language of the country where the test is being conducted but colored with different hues and requesting the user to click on the button that represents the color of the word. The *N-Back test* [24] is used to induce MWL. The test involves a square grid projection comprising nine boxes on the screen. A grey square will move within the grid and change its position every 2.25 seconds. Participants are instructed to press a button when the square returns to the same position as it was 1, 2, or 3 steps before. The test is

Table 1: This table reports the g.tec sensors adopted and the high- and low-pass frequencies considered in the g.Recorder software during the real-time acquisition sessions.

Signal	g.tec Sensor	Low-pass filter cut-off frequency (Hz)	High-pass filter cut-off frequency (Hz)
EDA	g.GSRsensor2	30	-
PPG	g.SENSOR Oxygen Saturation	30	0.1
Temperature	g.SENSOR temperature	30	-

conducted audibly and visually, concluding with both auditory and visual N-Back combined.

3.2. Population involved and equipment

The study was conducted with the approval of the Politecnico di Torino ethics committee (P.N. 1606). A total of 32 participants (64% men and 36% women) were recruited, with a mean age of 26. The *g.tech* hardware and software equipment was used to acquire synchronized multi-channel signals. The *g.HIAMP 144 Biosignal Amplifier* was used as the signal acquisition device with sensors listed in Table 1. The sampling frequency was fixed at 1200 Hz, a notch filter at 50 Hz was applied to all, and digital filters were applied as per Table 1. Four participants were excluded due to issues during acquisition.

3.3. Procedure and signal processing

We followed the procedure described in [14]. Initially, the volunteers' physiological baseline was established through a relaxation phase. The first test administered was the Stroop test, which consisted of three sub-phases with increasing external stressors. Following this, a five-minute rest phase was implemented to restore the physiological signal to the baseline level. Subsequently, the N-Back test was conducted, which comprised a visual, an auditory, and a dual phase (where both visual and auditory elements were combined). Each phase of the N-Back test was further divided into three sub-phases, corresponding to the 1, 2, and 3 back concepts.

3.4. Signal Processing

Also, in this case, we followed the procedure described in detail in [14]. We extracted key characteristics pertaining to heart activity, skin activity, and body temperature. Heart monitoring characteristics were analyzed using PPG signal processing, yielding 19 features associated with PPG shape, Heart Rate (HR), and Heart Rate Variability (HRV). Parameters such as mean, median, and standard deviation were computed for PPG amplitude, duration, and rise time. Subsequently, the beat per minute (BPM) trend was analyzed to extract HR and HRV features in the time domain. Further assessment involved evaluating HRV features in the frequency domain through spectrogram analysis of the BPM. We opted to analyze Electrodermal Activity (EDA) to monitor skin activity, which comprises slow and fast components: Skin Conductance Level (SCL) and Skin Conductance Response (SCR). For SCL analysis, mean, standard deviation, and slope were determined. Additionally, the mean and standard deviation of SCR peak amplitude and rise time, along with the average number of peaks per phase, were evaluated. We also examined certain aspects of body temperature raw signals and their first derivative, including initial, final, delta (final minus initial), mean, standard deviation, variation over time (change in value over a specified time interval), and variation over time slope (the first coefficient of linear regression of temperature or its first derivative within the relative phase).

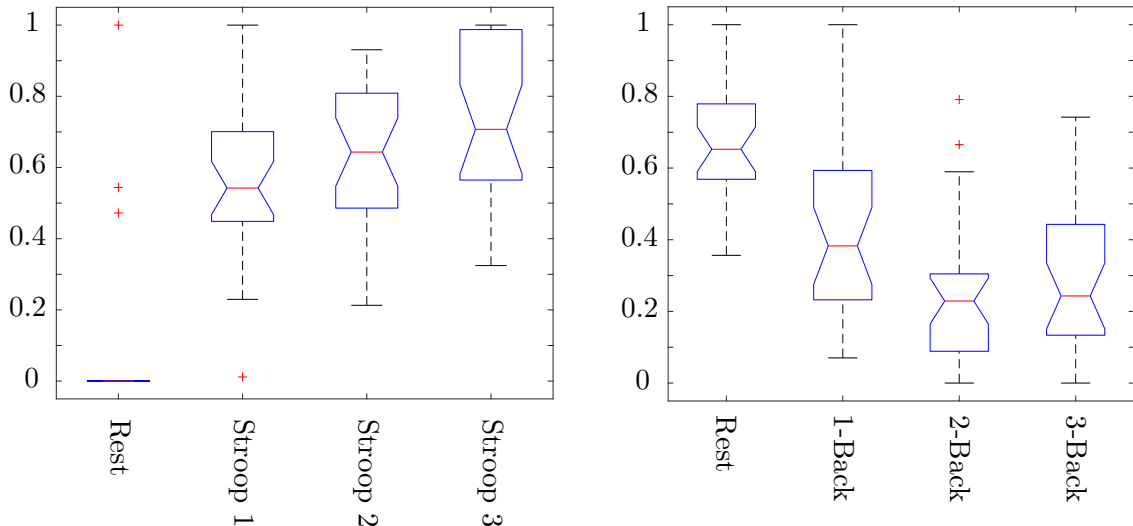
4. Results

The previous section detailed the signal-processing approach that resulted in the extraction of 43 features for every participant across each testing phase. To analyze the outcomes, we chose to treat each test individually, necessitating the consideration of four distinct datasets: Stroop, N-Back Visual, N-Back Auditory, and N-Back Dual. Each of these tests was further segmented into three smaller components. Therefore, it becomes evident that each dataset comprised an equal volume of data partitioned across the four test segments, inclusive of an additional rest phase derived from the physiological data collected during the initial relaxation phase, as elucidated in [14].

4.1. Features statistical analysis

Once the four datasets mentioned above were composed, we decided to implement a statistical analysis to understand if the features obtained for each phase (corresponding to different external stress and mental workload conditions) show some differences among their distributions. This procedure allows us to understand if there is a link between the variation of external induced stress or MWL and the considered physiological signals. The same process is repeated four times to analyse the features' significance in the Stroop and the Visual, Auditory, and Dual N-Back tests. The procedure is structured as follows:

- (i) The first step was to implement a *Kruskal-Wallis (K-W)* test [25] for each feature for each dataset between the four distributions of data corresponding to the four phases of the test. The choice of the K-W test is justified by the impossibility of assuming a normal distribution of the data. More specifically, if the *p-value* output was $p < 0.05$, we considered the test significant and could deeply analyse our tests' effect. The positive outcome of the K-W means that among the four distributions, at least one of them is different from the others; therefore, our tests generated a physiological response to the induced stress (related to the Stroop) or MWL (related to N-Backs). Two examples are reported in Figure 2.



(a) SCL mean distributions during the Stroop test phases.

(b) BPM mean distributions during the Auditory N-Back test phases.

Figure 2: This Figure presents two examples of the K-W test. Figure 2a related to the variation of the SCL mean during the Stroop test, while Figure 2b is related to the BPM mean during the Auditory N-Back test.

- (ii) The second step of our procedure was related to a more precise and deeper analysis to understand how many phases of the tests show statistical differences with respect to the others. In fact, for each dataset and feature that obtained a positive result in the K-W test, we adopted a *Wilcoxon Paired (W-P)* comparison [26] between each possible paired combination between the four test phases. (for example, Rest vs. Stroop 1, Rest vs. Stroop 2, Rest vs. Stroop 3, Stroop 1 vs. Stroop 2, etc.). We considered a statistically significant difference between the two populations when $p < 0.05$.
- (iii) The findings derived from this analysis are presented in Table 2. The initial row displays the features that exhibited significant results from the Kruskal-Wallis test. Subsequent rows indicate the count of characteristics that displayed one, two, and so forth positive outcomes in each pairwise comparison.

Table 2: The outcomes generated by the statistical method outlined are presented in the following table. It displays the number of significant features identified for each dataset, depicted in the columns.

	Stroop	Visual N-Back	Auditory N-Back	Dual N-Back
Test Significance	35	35	35	32
Feature Significance (> 1)	35	35	32	30
Feature Significance (> 2)	32	26	26	29
Feature Significance (> 3)	21	7	6	8
Feature Significance (> 4)	12	3	1	2
Feature Significance (> 5)	2	1	1	1

Moreover, a further step was to perform the same procedure considering in the W-P comparison the combinations related to the rest and the other phases (Rest vs. Stroop 1, Rest vs. Stroop 2, and Rest vs. Stroop 3); the results are reported in Table 3. By observing Table 2, it is possible to notice that from the fourth row, there is a significant drop in the number of features, giving a positive output from the W-P comparison. This behaviour can be explained in Table 3. It is almost a correspondence with the number of features that demonstrated significance in the rest of the comparison. Considering the initial forty-three features, nearly half of them are sensible in discriminating rest and all the other three more cognitively demanding phases. Therefore, a first consideration emerging from this analysis is that a binary classification between rest and cognitive demanding phases is possible through this physiological approach. At the same time, only a few features can distinguish the different mental challenging stages (represented by the last three rows of Table 2). Another consideration is evaluating how many features are sensitive both to stress and MWL. This can be achieved by comparing the relevant features corresponding to the fourth row of Table 3, resulting in 16 characteristics. These features encompass various parameters such as the standard deviation of PPG shape's amplitude, duration, and rise time; BPM standard deviation and pNN50 related to time-domain HRV; standard deviation of LF, HF, and LF/HF related to frequency-domain HRV; mean and standard deviation of SCL; initial, delta, standard deviation, and slope of body temperature; and finally, reaction time.

5. Conclusion

It is gaining increasing attention in aviation due to the possible transition towards Single Pilot Operations. However, to reach this goal, there are some barriers to overcome, including the development of a digital cockpit assistant that is able to replace the figure of today's First Officer. In this context, being conscious of pilots' mental workload and stress levels during the

Table 3: The outcomes derived from the statistical method utilized to explore potential binary classification are presented in the following table. It outlines the number of significant features identified for each dataset, depicted in the columns.

	Stroop	Visual N-Back	Auditory N-Back	Dual N-Back
Test Significance	35	35	35	34
Feature Significance (> 1)	34	34	33	32
Feature Significance (> 2)	29	28	22	20
Feature Significance (> 3)	23	21	19	19

flights is essential. Nowadays, there are three ways to assess these mental conditions: subjective evaluations, behavioural measures, and physiological evaluations. Of these methods, examining signals obtained from the human body has emerged as a dependable option, benefiting from the expansion of the biomedical market in recent years. This growth has led to the availability of more affordable, compact, and trustworthy sensors and technologies. This situation is opening new spaces for research to integrate the physiological approach to the necessity of the aeronautical field in the transition toward SPOs. A reliable solution that links the variation of human body parameters and perceived stress and MWL must still be included. Therefore, this paper aims to enhance the preliminary analysis introduced in [14] by performing a deeper statistical analysis related to the physiological approach to assessing a person's mental workload and stress levels. We demonstrated how a binary classification between a rest and a more cognitively demanding condition is possible, only considering three physiological signals: PPG, EDA, and body temperature. This study highlights the feasibility of the physiological multi-modal approach to evaluating stress and MWL. The next step is to expand the number of involved participants and the acquired physiological signals in order to understand the most significant. It is the first step to bringing this approach to designing the next-generation cockpit for SPO's aircraft.

Acknowledgements

The research presented in this paper is backed by the Future Aircraft Technologies, Autonomy, and HMI Research Unit of Leonardo S.p.A, along with support from European Union funds DM 1061. The authors express their sincere appreciation to both entities for their valuable insights and co-financing of the PhD program, which facilitated the development of this paper. Additionally, we extend our gratitude to the PolitoBIOMed Lab at Politecnico di Torino for granting us access to their facilities and spaces for conducting our tests.

References

- [1] Young M S, Brookhuis C, Wickens C D, Hancock P A 2014 State of science: mental workload in ergonomics *Ergonomics* **58** p 1-17
- [2] EASA 2023 EASA AI Roadmap *European Union Aviation Safety Agency*.
- [3] Li M, Wang M, Ding D, Wang G 2022 Development and Evaluation of Single Pilot Operations with the Human-Centered Design Approach *Aerospace* **9**
- [4] Harris D 2007 A human-centred design agenda for the development of single crew operated commercial aircraft *Aircraft Engineering and Aerospace Technology* **79** p 518-526
- [5] Matessa M, Strybel T, Vu K, Battiste V, Schnell T 2017 Concept of Operations for RCO SPO *NASA*
- [6] Murray G, Heilakka R 2023 The airline pilot shortage will get worse *Technical report Oliver Wyman*
- [7] Faulhaber A K, Friedrich M 2019 Eye-Tracking Metrics as an Indicator of Workload in Commercial Single-Pilot Operations *Communications in Computer and Information Science* p 213-225
- [8] Debie E, Fernandez Rojas R, Fidock J, Barlow M, Kasmarik K, Anavatti S, Garratt M, Abbass H A 2021

- Multimodal Fusion for Objective Assessment of Cognitive Workload: A Review *IEEE transactions on cybernetics* **51** p 1542–1555
- [9] Charles R L, Nixon J 2019 Measuring mental workload using physiological measures: A systematic review *Applied Ergonomics* **74** p 221–232
- [10] Silva M, Santos L, Melicio R, Valério D, Rocha R, Barqueira A, Brito E 2022 Aviation's Approach Towards Pilots' Mental Health: a Review *International Review of Aerospace Engineering (IREASE)* **15**(6) p 924–307
- [11] Giannakakis G, Grigoriadis D, Giannakaki K, Simantiraki O, Roniotis A, Tsiknakis M 2022 Review on Psychological Stress Detection Using Biosignals *IEEE Transactions on Affective Computing* **13**(1) p 440–460
- [12] Luzzani G, Buraioli I, Demarchi D, Guglieri G 2023 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* p 1–22
- [13] Ayres P, Lee J Y, Paas F, van Merriënboer J J G 2021 The Validity of Physiological Measures to Identify Differences in Intrinsic Cognitive Load *Frontiers in psychology* **12**
- [14] Luzzani G, Buraioli I, Demarchi D, Guglieri G 2023 Preliminary Study of a Pilot Performance Monitoring System Based on Physiological Signals *Aerospace Europe Conference 2023 – 10th EUCASS – 9th CEAS*
- [15] Van Acker B, Parmentier D, Vlerick P, Saldien J 2018 Understanding mental workload: from a clarifying concept analysis toward an implementable framework *Cognition Technology Work* **20**
- [16] Endsley M R 1995 Toward a Theory of Situation Awareness in Dynamic Systems *Human Factors* **37**(1) p 32–64
- [17] Stanton N A, Chambers P R G, Piggott J 2001 Situational Awareness and Safety *Safety Science* **39** p 189–204
- [18] Fink G 2010 Stress: Definition and History *Stress Science: Neuroendocrinology* p 3–9
- [19] Martins A P G 2016 A Review of Important Cognitive Concepts in Aviation *Aviation* **20**(2) p 65–84
- [20] Zhang Y, Zheng H, Duan Y, Meng L, Zhang L 2015 An integrated approach to subjective measuring commercial aviation pilot workload *2015 IEEE 10th Conference on Industrial Electronics and Applications (ICIEA)* p 1093–1098
- [21] Brand Y, Schulte A 2021 Workload-adaptive and task-specific support for cockpit crews: design and evaluation of an adaptive associate system *Human-Intelligent Systems Integration* **3**
- [22] Wan H, Zhuang L, Pan Y, Gao F, Tu J, Zhang B, Wang P 2020 *Biomedical sensors* p 51–79.
- [23] Scarpina F, Tagini S 2017 The stroop color and word test *Frontiers in Psychology* **8**
- [24] Janczewski N, Wittmann J, Engeln A, Baumann M, Krauß L 2021 A meta-analysis of the n-back task while driving and its effects on cognitive workload *Transportation Research Part F: Traffic Psychology and Behaviour* **76** p 269–285
- [25] Vargha A, Delaney H D 1998 The Kruskal-Wallis Test and Stochastic Homogeneity *Journal of Educational and Behavioral Statistics* **23**(2) p 170–192
- [26] Rosner B, Glynn R J, Lee M L 2006 The Wilcoxon signed rank test for paired comparisons of clustered data *Biometrics* **62**(1) p 185–192