DOI: 10.13009/EUCASS2023-349
Aerospace Europe Conference 2023 – 10TH EUCASS – 9TH CEAS

Preliminary Study of a Pilot Performance Monitoring System Based on Physiological Signals

Gabriele Luzzani[⋆†], *Irene Buraioli*[•], *Danilo Demarchi*[•] *and Giorgio Guglieri*[⋆]
[⋆]*Politecnico di Torino, Department of Mechanical and Aerospace Engineering*
Corso Duca degli Abruzzi 24, 10129 - Turin, Italy
[•]*Politecnico di Torino, Department of Electronics and Telecommunication*
Corso Duca degli Abruzzi 24, 10129 - Turin, Italy
gabriele.luzzani@polito.it · giorgio.guglier@polito.it
[†]Corresponding author

Abstract

The rising operating costs and the decreasing availability of pilots push the aviation market towards Single Pilot Operations. Due to the related need for a cockpit assistant able to understand pilots' cognitive workload and stress and the lack of robust solutions, the relationship of these conditions with variations in physiological parameters is being studied. In this paper, initial computer cognitive tests were performed on thirty healthy volunteers, providing the physiological parameters of PPG, EDA, and temperature under four mental workloads and stress conditions. A statistical approach was performed to identify these cognitive states' 16 out of 43 most representative characteristics.

1. Introduction

The aviation sector has always been at the forefront of technical advancement, continuously pushing the envelope to improve the accessibility, effectiveness, and safety of air travel. This field has undergone numerous technological revolutions since the first pioneer flights up until the present, each of which has advanced the development of safer air travel. In particular, the disruptive growth of Artificial Intelligence (AI) is currently leading to significant changes in the aviation market. According to the EASA AI Roadmap,[9] this technology has the potential to transform this world by enabling us to develop intelligent systems capable of providing advanced assistance solutions to human end users, optimizing aircraft performance, improving air traffic management, and thus increasing safety in previously unimaginable ways. The concept of safety in aviation is fundamental, leading to continuously finding new technologies and systems to enhance surer flights in the future. According to the 2021 Safety Report by the International Civil Aviation Organization (ICAO),[13] shown in Figure 1, this effort is represented by the decreasing trend in the yearly accidents rate of commercial flights in the last ten years.

Maintaining this trend by integrating new technologies, such as AI, opens new spaces for innovation in this continuously evolving context. In particular, there has recently been an increasing focus on the so-called Single Pilot Operations (SPOs).[16] Despite today's regulations for most airline flights requiring a pilot and a co-pilot on board the aircraft, the SPO's term is related to the condition of flying an aircraft with only one pilot, assisted by advanced systems and/or ground operators providing piloting support services. However, this technological variation in aviation can be achieved only by guaranteeing the same (or higher) safety and handling quality level regulated to date in the EASA parts. SPOs would permit solving two key challenges facing the aviation market today, which are represented by the decreasing trend in the availability of pilots and the rising operating costs, especially on short and medium-haul flights.[17] On the one hand, due to the current market trend, the most recent forecasting highlights how we have to expect a nearly 80000 pilot shortage in global aviation by 2032 if something will not change in the meantime.[19] On the other hand, it was evaluated that the transition towards only one pilot on board the aircraft allows nearly half the cockpit crew costs. To reduce the number of pilots from two to one, it is necessary to implement a cockpit assistant with the potential to understand their cognitive workload and their ability/inability to operate the aircraft by leveraging the disruptive trend of AI applications in aviation.

In fact, many studies have been carried out on the assessment of the pilot's mental workload in a Human-Machine Interface (HMI) context by exploiting the capabilities of subjective questionnaires, behavioural analysis (through primary and secondary tasks evaluation methods), and physiological measures. In particular, among these methodologies, thanks to the disruptive growth of the biomedical market in recent years, the analysis of the physiological response to

DOI: 10.13009/EUCASS2023-349

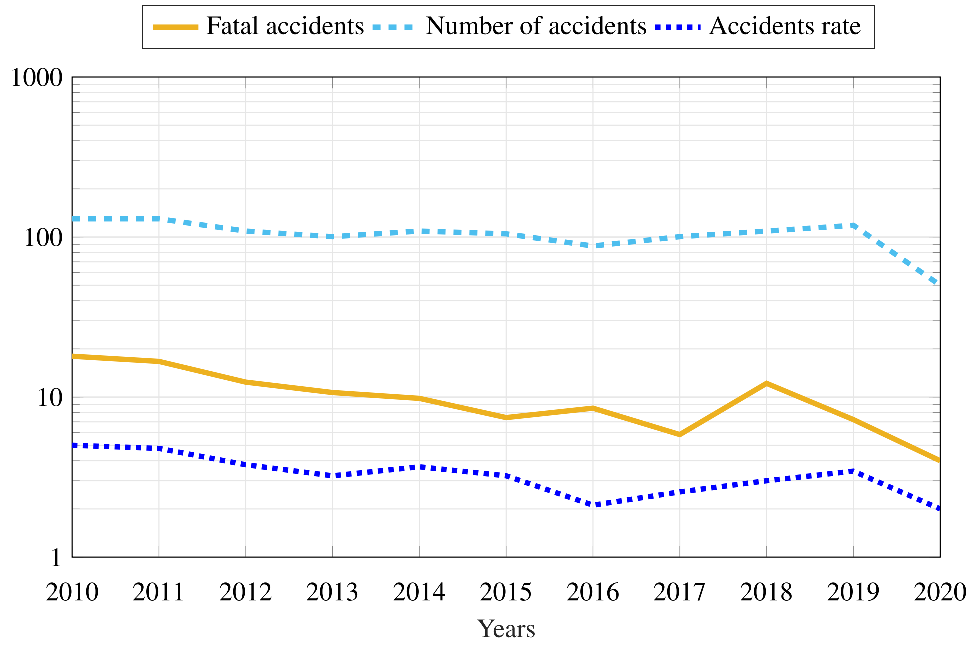


Figure 1: This Figure shows the aeronautical accidents trend from 2010 to 2020.[13] It highlights a decreasing trend in the accidents per year, fatal accidents per year, and the accident rate (defined as the number of accidents per million departures).

cognitive load variation is gaining increasing importance. However, due to the complex and multifaceted dimension of the cognitive workload, a comprehensive approach to assessing this mental condition has not been found, leading to the necessity of new solutions to tackle this challenge.[25] Several physiological signals were linked with this mental condition's variation, such as cardiorespiratory measures, brain and electrodermal activity, body temperature, and eye parameters monitoring. Nevertheless, a robust solution demonstrating the connection between these signals and the mental workload is still missing.

This paper aims to provide a preliminary step into this field, analyzing the relationship between cardiac activity, electrodermal activity, and body temperature, with the variation of stress and cognitive workload. A population of 31 volunteers was measured during the execution of two specific cognitive tests, Stroop and N-Back. As explained in the following paper, the former is adopted to stimulate three external stress conditions and the latter to different mental workload situations. We decided to start from a subset of the aforementioned physiological signals due to the strong connection of those signals with the human body's response to these external stimuli. The following section presents the background of studying stress and MWL in aviation by introducing their definition and the state-of-the-art technologies adopted in this field. The methods of our study are reported in Section 3, followed by the presentation of the results in Section 4. The conclusions are drawn in Section 5.

2. Background

Mental workload and *stress* are two concepts that are gaining increasing importance in the aeronautical sector due to their key role in the implementation of the next generation of Human-AI interactive cockpits.[9] Therefore, it s necessary to clarify what these terminologies refer to.

Mental Workload The mental workload (MWL) concept has been studied in the Human Factors and Ergonomics (HFE) context for several years. Due to its complex and multifaceted nature, this term has no unique definition. In fact, over the years, several researchers tried to find common ground for MWL by providing different definitions related to their application fields.[82726] Nevertheless, some recent studies attempted to clarify this, introducing the following definition of mental workload as *a subjectively experienced physiological processing state, revealing the interplay between one's limited and multidimensional cognitive resources and the cognitive work demands being exposed to.*[23] Moreover, it is important to highlight that, in the literature, there is some debate about the difference between mental and cognitive workload. Nevertheless, according to Hancock et al.,[10] the two terms can be referred to as the same condition; therefore, the mental and cognitive workload will be treated as synonymous in this paper.

DOI: 10.13009/EUCASS2023-349

Stress It is often found in the literature confusion about the definition of the term stress in relation to mental workload. Thus, a common definition of this term is difficult to find also in this case. In the aeronautical context, it can be defined as *the body's reaction to stimuli that interfere with a person's normal physiological balance and cause physical, mental, or emotional stress*.[4]

However, these two conditions could not be considered as two separate entities. In fact, according to the Debie et al. model,[8] a direct cause-effect link exists between stress and MWL, by seeing the former as a depletion factor that affects the MWL of an operator.

It is possible to find in the literature mainly three ways to assess pilots' stress, and MWL adopted up to date: subjective evaluations, behavioural analysis, and physiological measures.[18] In particular:

- **Subjective evaluations** are essentially represented by post-performance questionnaires submitted to pilots, asking their subjective perception of mental workload and stress through specific ad-hoc questions. This tool is often adopted due to its ease and low cost of implementation.[28] The most used subjective questionnaires in aviation are the NASA TLX,[12] the Cooper-Harper or Modified Cooper-Harper scale,[7] and the Bedford one's.[5]
- **Behavioural measures** are based on observing the pilots' actions during the flight mission and comparing them with a defined mission plan. The cognitive workload is therefore inferred by checking the number of wrong or missed actions. This approach represents the base of the so-called primary or secondary tasks performance analysis.[10] According to Schulte et al., this method could be successfully adopted in developing a digital cockpit assistant to build the next generation of fighters.[3]
- **Physiological measures** refer to analysing and processing physiological signals to infer a pilot's cognitive condition during the flight mission. Because of the market's recent significant growth in the biomedical sector, the study of this topic in the aeronautical field is growing. The availability of cheaper, smaller, and more reliable biomedical sensors allowed the possibility of the implementation of new technologies.[24] Although a complete, reliable physiological-based solution is still missing, the literature shows several signals that are sensitive in different ways to the variation of MWL and stress. In particular, they are reported in Table 1 and are represented by heart activity, skin activity, eye activity, brain activity, respiration, body temperature, muscle activation, and voice patterns.[8][6][27][1]

Table 1: The state-of-the-art methods adopted to assess pilots' MWL and stress levels are reported in this table. For each of them is provided with a general description, and the most adopted techniques related to them (for the physiological assessment, the involved human body's signals are reported).

Method	Techniques	Description
Subjective questionnaire	NASA TLX Cooper-Harper scale Modified Cooper-Harper scale Bedford scale	They represent post-performance subjective questionnaires submitted to pilots
Behavioural measure	Primary task performance Secondary task performance	It consists of the comparison between an a priori defined mission plan and the currently activities of the pilot
Physiological analysis	Heart activity Respiration monitoring Eye activity Skin activity Body temperature Brain activity Muscle activation Voice pattern	This tool is based on the acquisition and processing of human body's signals. The most important features are evaluated in order to infer from them the pilots' MWL.

2.1 Involved physiological signals

The analysis of subjective perceptions and behavioural observations are two ways to assess MWL and stress that have been studied since the last decades of the twentieth century. Nevertheless, the physiological response to these external

DOI: 10.13009/EUCASS2023-349

stimuli has accelerated recently. The current technological development has yet to define which signals are the most significant, opening up new spaces for research in this area. Therefore, this is the goal of our activity, and this work will focus on the most significant physiological signals according to the literature: the heart and skin activity[22] and the body temperature. It is important to underline that we aim to consider all those signals that could be implemented in a future SPO's cockpit. In particular, we decided to monitor all the signals that could be measured on a single hand, considering the unobtrusiveness of the sensors in an optic of a possible future application in the aeronautical field.

Skin activity The monitoring of skin activity is known as electrodermal activity (EDA). This signal is related to the functioning of the sweat glands and the skin blood vessels, which are exclusively led by the sympathetic nervous system. Its activation is linked to stress and mental workload reactions[22] by modifying the sweating of the skin and, consequently, the conductance of on applied current. Thus, it is possible to infer a person's stress or mental workload enhancement by measuring the variations in the electric skin conductance.[20] The key characteristic of EDA is represented by the possibility of dividing it into a slow tonic component named Skin Conductance Level (SCL) and a fast phasic part defined as Skin Conductance Response (SCR). The former is linked with the slow reaction to external stimuli, while the latter represents the short-time response.

Heart activity The heart activity can be measured with an electrocardiogram (ECG) or photoplethysmography (PPG).[2] The ECG represents the heart's electrical functioning and is composed of a well-defined pattern represented by the five waves P, Q, R, S, and T. The PPG, instead, consists of measuring the variations in blood volume through optical analysis under a specific light wavelength. Both of these measures allow us to evaluate the peak-to-peak interval of heart activity in order to assess the heart rate (HR) and the heart rate variability (HRV) that contains significant information about the changes in stress and cognitive workload conditions.[1] It is important to highlight that, during the execution of our tests, we decided to measure HR and HRV through the PPG analysis due to the sensor's unobtrusiveness and reliability and its possibility to be placed on a single finger, as shown in the next section.

Body temperature Body temperature is one of the most well-known biosignals since it provides important information about a person's health. The significance of this signal in this instance arises from the fact that blood flow in superficial vessels, which comprise the majority of the microcirculation, greatly influences body temperature. Therefore, it is assumed that a person's skin temperature will change due to vasoconstriction and decreased blood flow, as it does, for instance, during times of stress.[15]

3. Methods

This section aims to present all the characteristics of our study, first introducing the description of our test, the involved population, and the adopted equipment. Then, an explanation of the developed procedure and the feature extraction process will be exploited.

3.1 Test

Stress and cognitive load variations are simulated through two specific computerised tests widely studied in the literature: the Stroop and N-Back tests.

The Stroop test,[21] shown in Figure 2a, is adopted to simulate an external stress condition. It consists of projecting on display a series of words related to the names of one colour in the language of the country where they were performed (in this case Italian), coloured with other colours, and asking the user to click the button corresponding to the colour of the word itself. This test will be repeated three times, in which the degree of difficulty will be increased by including certain distractors such as background noise and the variation of the position of the buttons.

The N-Back test,[14] shown in Figure 2b, is adopted to induce different levels of mental workload. It consists of a square grid projection containing nine boxes on the screen. A grey square will move within this grid and change its position every 2.25 s. The user is asked to press a button named *Position* when the square returns to the same position as 1/2/3 previous steps. This test is also repeated from an auditory point of view, in which the users are made to listen to a sequence of letters and, again, are asked to press the *Audio* button when they hear the same letter as 1/2/3 of previous steps. The test concludes with the last series of tests in which both auditory and visual N-Back are combined, and the user must be able to manage both.

DOI: 10.13009/EUCASS2023-349

(a) Stroop test graphical interface.

(b) N-Back test graphical interface.

Figure 2: This Figure provides the computer graphical interface adopted during our tests. Figure 2a is related to the Stroop test to stimulate external stress in the participant. Figure 2b shows the N-Back screen implemented to engage different mental workload levels.

3.2 Participants

Our procedure was authorized by the Politecnico di Torino ethics committee (protocol number 1606), allowing us to engage the population selected on a voluntary base. In particular, we obtained 32 participants, composed of 64% males and 36% of females. The age range is from 23 to 41, with a mean of 26 years. However, we had to exclude 4 subjects from our population due to some problems that happened during the acquisition session. Therefore we obtained a final sample of 28 people.

3.3 Equipment

We decided to employ the hardware and software products provided by the company *g.tec* in collaboration with the *PolitoBIOMed Lab* of the Politecnico di Torino. In particular, we decide to adopt the *g.HIAMP 144 Biosignal Amplifier* as a synchronized multi-channel signal acquisition device. It is a high-performance biosignal amplifier with channels for invasive and non-invasive measurements. The sensors used in this experiment are: *g.GSRsensor2* placed on the second phalanx of fingers 2 and 3 to measure the EDA signal; *g.SENSOR Oxygen Saturation* on the tip of the index finger with the LED positioned above to measure the PPG; *g.SENSOR temperature* to measure the peripheral external body temperature, with the thermo-sensor on the fingertip of the fifth finger (in order to measure the greater temperature variations than other positions). The g.HIAMP device interacts directly with the ad-hoc developed *g.Recorder* software allowing the setting of the sampling frequency and digital filters and real-time monitoring of the biosignals acquired. In particular, we decide to fix a sampling frequency of 1200 Hz, a notch filter at 50 Hz to all of them, and the digital filters reported in Table 2.

Table 2: This table shows the low and high pass filter's cut-off frequencies inserted in the g.Recorder software during the acquisition of the biosignals. The symbol - indicates that the filter was not implemented.

Signal	**Low-pass filter cut-off frequency (Hz)**	**High-pass filter cut-off frequency (Hz)**
EDA	30	-
PPG	30	0.1
Temperature	30	-

3.4 Procedure

This section will explain the procedure followed for acquiring the physiological signals of each participant. The overall duration of the test was around an hour to complete. Before the beginning, the volunteers were asked to remove watches, rings, and bracelets from the opposite hand than the one taking the test, to set the mobile phone to silent mode without vibration, and not to wash their hands before the test to avoid superficial skin conditions far from the natural. Depending on the subject, an interval of time was chosen to acclimatize to the room, both in terms of environment

DOI: 10.13009/EUCASS2023-349

and temperature. After this, the three sensors were connected to the off-hand of the subject (in this case, for all 28 participants on the left hand). Then, the volunteer was asked to sit at the desk in front of the computer and relax. In the meantime, the tests were explained, and a short example was provided for practice. This period was necessary to let the physiological signals reach their baseline values. As soon as the subject was confident with the functioning of the test, the following data acquisition session was started.

The starting test was the Stroop test, divided into three sub-phases. The first one was with the meaning of the word congruent with the colour; the second introduced an inconsistency between the meaning of the word and its significance, changing the position of the buttons; the last introduced a voice to disturb the subjects by pronouncing random colour names. All three phases presented a background noise to stress the people. At the end of this part, a five minutes rest phase was introduced in order to restore the physiological signal at their baseline. Then, the N-Back test could start. It was divided into a visual, an auditory, and a dual phase (where the visual and auditory parts were overlapped), as explained in Section 3.1. Each phase was divided into three sub-phases corresponding to the 1, 2, and 3 back concepts. The test concluded with a final rest phase where we asked the subjects to compile a self-assessed questionnaire about their subjective perception of the mental workload and stress level on a five-level scale.

Figure 3 shows the entire procedure followed during the test performance in a flow chart. The green boxes represent the rest phases, the orange blocks are related to the Stroop test, and the purple ones are the three different N-Back test steps. Moreover, the duration of each sub-phase is reported, showing how the entire test lasts around one hour, considering that the overall time depends on the answer velocity of the subject.

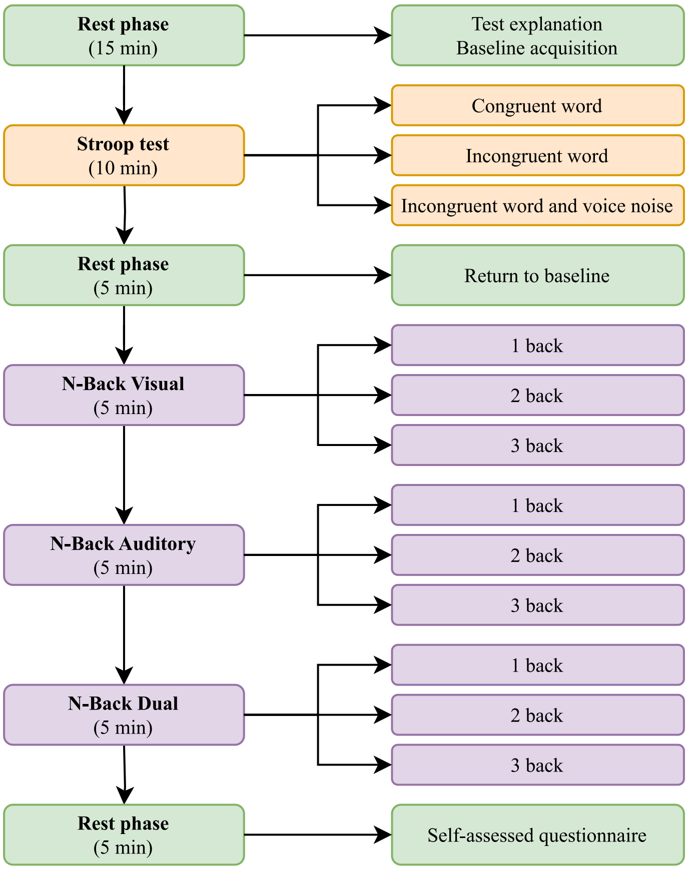


Figure 3: This Figure represents the test procedure adopted for each participant. The different colours highlight the Rest, Stroop, and N-Back phases. The duration of each subphase is also reported.

3.5 Feature Extraction

The next step in the pipeline of our work was represented by processing the acquired signal to extract all the features that, from a literature point of view, showed a relationship between the variation of stress and cognitive workload. The code was developed in *Matlab* environment. The selected PPG, EDA, and body temperature features are explained in

DOI: 10.13009/EUCASS2023-349

the following Sections. In particular, the processing of the three physiological signals is described separately, and, for each Section, an introduction to the selected features, followed by an explanation of the steps taken in their evaluation, is reported. In fact, the PPG's characteristics are written in Table 3, the EDA's ones in Table 4, and the body temperature attributes in Table 5. Moreover, these tables also show the literature's expected trend of these features that, in general, it is possible to observe. The (↑) symbol represents an increase of the feature as the MWL increases, the (↓) symbol a decrease, and the (−) character an uncertain trend. Finally, it is important to highlight that each signal was divided into each of the fifteen phases presented in the previous section in Figure 3, leading to fifteen different values of each feature for each participant.

3.5.1 PPG processing

As mentioned in Section 2, heart activity was assessed by analysing the PPG signal. Thanks to this physiological signal's well-defined characteristics, we extracted 19 features during this post-processing part of the experiment. These are presented in Table 3. They referred to the *PPG shape*'s amplitude, duration, and rise time, as reported in Figure 4b. In particular, for each of them, the mean, median, and standard deviation in a window of 10 s were evaluated. The mean, median, standard deviation, and pNN50 of the beats per minute (BPM) trend were related to the analysis of the heart rate *HR* and heart rate variability *HRV* in the time domain. Moreover, a frequency analysis was also performed by extracting the low (PLF 0.04 Hz to 0.15 Hz) and high-frequency (PHF 0.15 Hz to 0.4 Hz) content from the BPM trend, as explained in the following. Finally, the ratio between PLF and HLF was evaluated as a key indicator of stress and mental workload. The first step in processing the PPG was selecting a single starting point to which subsequent signals

Table 3: This table shows the PPG features adopted in our analysis related to the shape of the signal, the analysis of the heart rate, and the heart rate variability. Their literature expected trend is also reported.

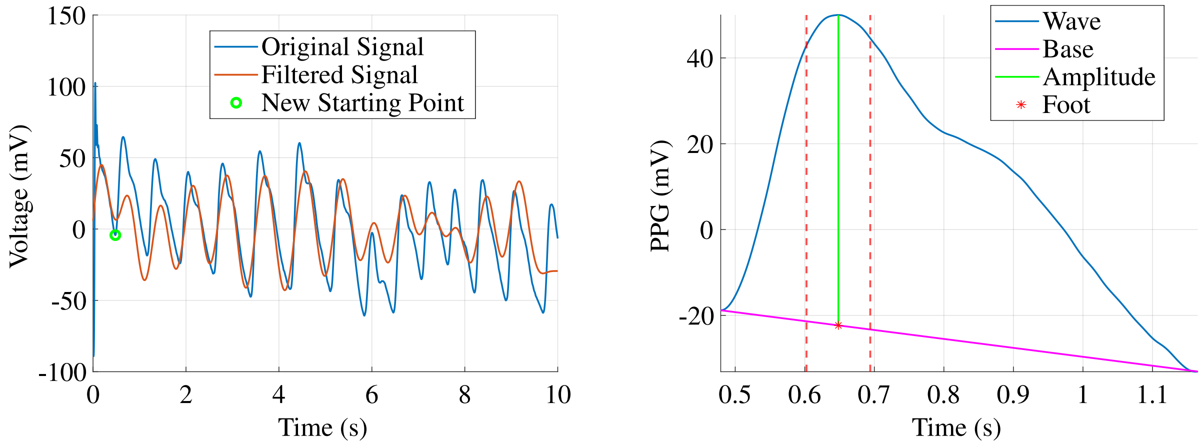
Signal	Feature	Expected trend
PPG	Mean amplitude	↑
	St. Dev. amplitude	-
	Median amplitude	↑
	Mean duration	↓
	St. Dev duration	-
	Median duration	↓
	Mean rise time	↓
	St. Dev. rise time	-
	Median rise time	↓
HR	Mean BPM	↑
	St. Dev. BPM	↓
	Median BPM	↑
HRV	pNN50	↓
	Mean PLF IBI	↑
	St. Dev. PLF IBI	-
	Mean PHF IBI	↓
	St. Dev. PHF IBI	-
	Mean PLF/PHF IBI	↑
	St. Dev PLF/PHF IBI	-

of EDA and temperature were aligned to remove noise at the beginning of the recording. To do so, the first 10 s of the PPG registration was filtered with a fourth-order Chebyshev Type I passband filter from 0.5 Hz to 1.5 Hz. The second minimum was identified and chosen as the reference starting point (see Figure b4a).

Then, the PPG raw signal was taken again in order to remove the noise with a fourth-order Chebyshev Type I passband filter 0.5 Hz to 10 Hz; the output was processed through a windowing method with a 10 s length, in which every single wave of the PPG was identified, as shown in Figure 4b, to extract all the features reported in Table 3 related to the *PPG shape*. In particular, identifying the wave peak represented a key factor because, as a consequent step, the temporal distance between them was evaluated to assess the Inter-Beat Intervals (IBIs). Starting from them, it was possible to delete the outliers and calculate the BPM trend by adopting the following equation:

$$BPM\ (-) = \frac{f_s\ (s^{-1})}{IBI\ (samples)} \times 60\ (s) \tag{1}$$

DOI: 10.13009/EUCASS2023-349



(a) Starting point selection.

(b) PPG's single wave.

Figure 4: This Figure provides three steps to process the PPG signal. Figure 4a shows the method adopted to obtain an initial starting point. Figure 4b reports the features related to the shape of the PPG signal.

where f_s represented the sampling frequency of the test. From the BPM trend, shown in Figure 5, it was possible to obtain the features reported in Table 3 related to the *HR* and *HRV* in the time domain. The final step of the PPG processing was implementing a frequency analysis of the BPM trend through a Power Spectral Density. In particular, Burg's method was used with windows of 168 heartbeats, from which the features in Table 3 related to the frequency domain analysis of the *HRV* were evaluated.

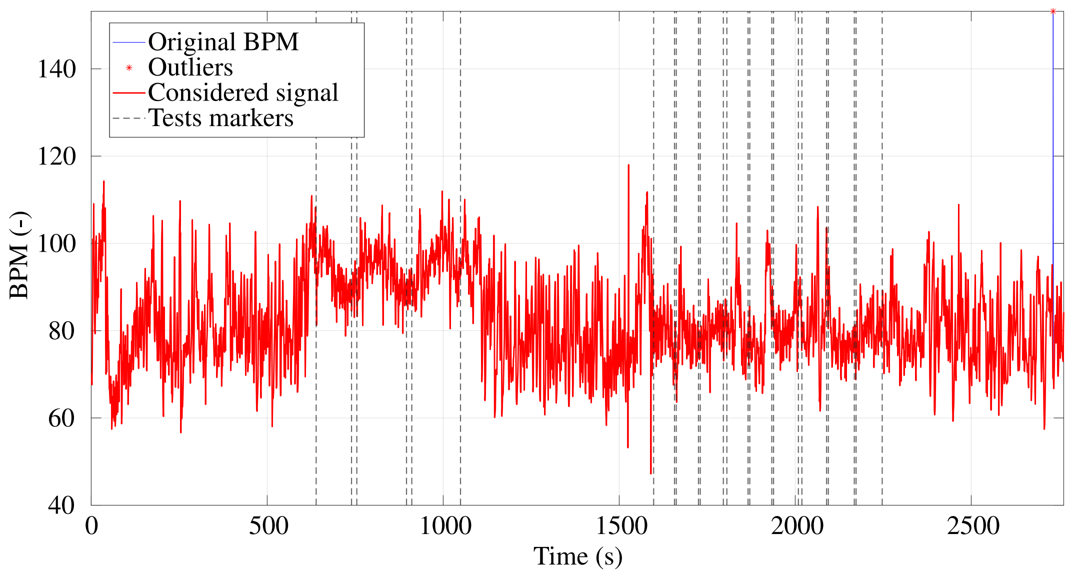


Figure 5: This Figure shows the BPM output obtained from the PPG process.

3.5.2 EDA processing

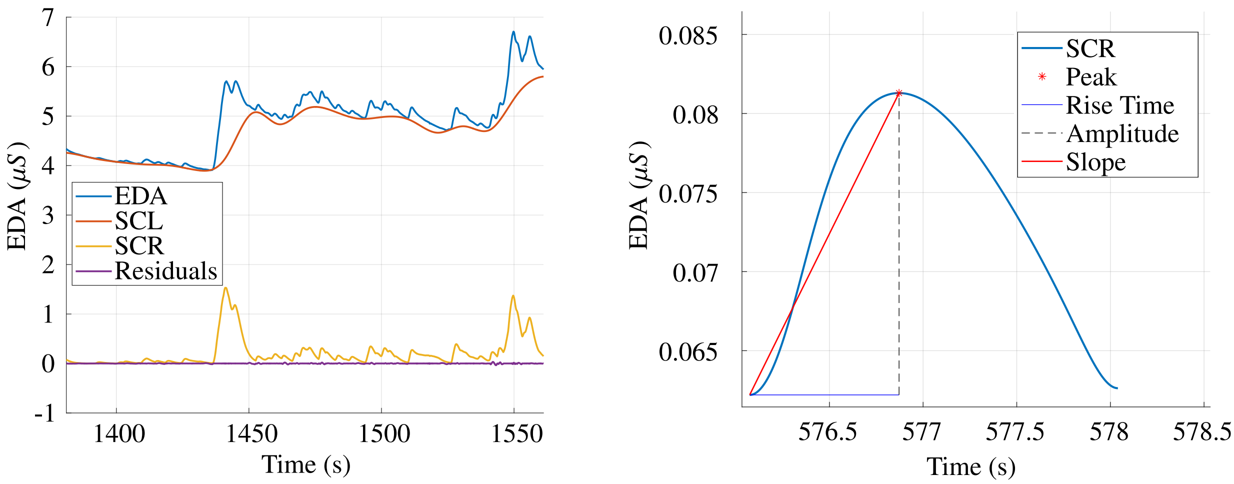
Table 4 reports the features adopted to study the trend of the EDA. As mentioned in the previous Section, this physiological signal can be divided into slow and fast components. The former, named Skin Conductance Level (*SCL* in the following Table), was studied by assessing its mean, standard deviation, and slope for each test phase. The latter, specifically Skin Conductance Response (*SCR*), was assessed by evaluating the averaged mean and standard deviation of the amplitude and rise time of the SCR's peaks and their averaged number per phase. Also in this case, Table 4 provides the literature's expected trend for each feature of the EDA.

DOI: 10.13009/EUCASS2023-349

Table 4: This table shows the EDA features adopted in our analysis related to the SCL and SCR components. For each feature, the literature's expected trend is also reported.

Signal	Feature	Expected trend
SCL	Mean	↑
	St. Dev.	↑
	Slope	↑
SCR	Mean amplitude	↓
	St. Dev. amplitude	-
	Mean rise time	-
	St. Dev. rise time	-
	Average peaks number	↓

The EDA analysis began with the alignment with the starting point obtained during the PPG analysis described in the previous Section. Then, both the noise removal and the separation between SCL and SCR were performed through the cvxEDA method introduced by A. Greco et al.,[11] giving us the output shown in Figure 6a. It is possible to observe the original EDA signal in blue, the slow tonic component in red, the fast component in yellow, and the residuals of the process in purple. On the output provided by this algorithm, the *SCL* and *SCR* features shown in Table 4 were evaluated. In particular, to obtain the phasic component's characteristics, a 90 s window and a seventh-order Butterworth low-pass filter at 5 Hz were implemented.



(a) EDA decomposition output.

(b) EDA single wave features.

Figure 6: Figure 6a provides an example of the output obtained from the EDA signal processing algorithm presented in.[11] The raw EDA in blue is recognizable, the SCL's slow component in green, the SCR's fast component in light blue, and the residuals in red. The SCR's characteristics are described in Figure 6b.

3.5.3 Body temperature processing

The features selected for the body temperature analysis are reported in Table 5. Since this physiological signal is not so widely studied in the literature, we decided to assess some features related to its first derivative, never introduced in the analysis of the MWL and stress level. In particular, the following values were evaluated related to the body temperature raw signal and its first derivative: initial value, final value, delta (final less initial value), mean, standard deviation, variation over time (delta value over the selected time interval), and variation over time slope (first coefficient of the linear regression of the temperature or its first derivative in the relative phase).

As presented in the previous chapter, the first step in processing the body temperature was represented by the alignment with the aforementioned starting point. To extract all the features presented in Table 5, a second-order Butterworth low-pass filter at 10 Hz was realised to remove the noise. In particular, Figure 7 shows a graphical example of the raw signals of the body temperature and the study about the sign of its first-time derivative.

DOI: 10.13009/EUCASS2023-349

Table 5: This Table shows the body temperature features adopted in our analysis. For each feature, the literature's expected trend is also reported.

Signal	Feature	Expected Trend
Temperature	Initial value	↓
	Final value	↓
	Delta value	↑
	Mean	↓
	St. Dev.	-
	Variation over time	↑
	Variation over time slope	↑
Temperature First Derivative	Initial value	-
	Final value	-
	Delta	-
	Mean	-
	St. Dev.	-
	Variation over time	-
	Variation over time slope	-

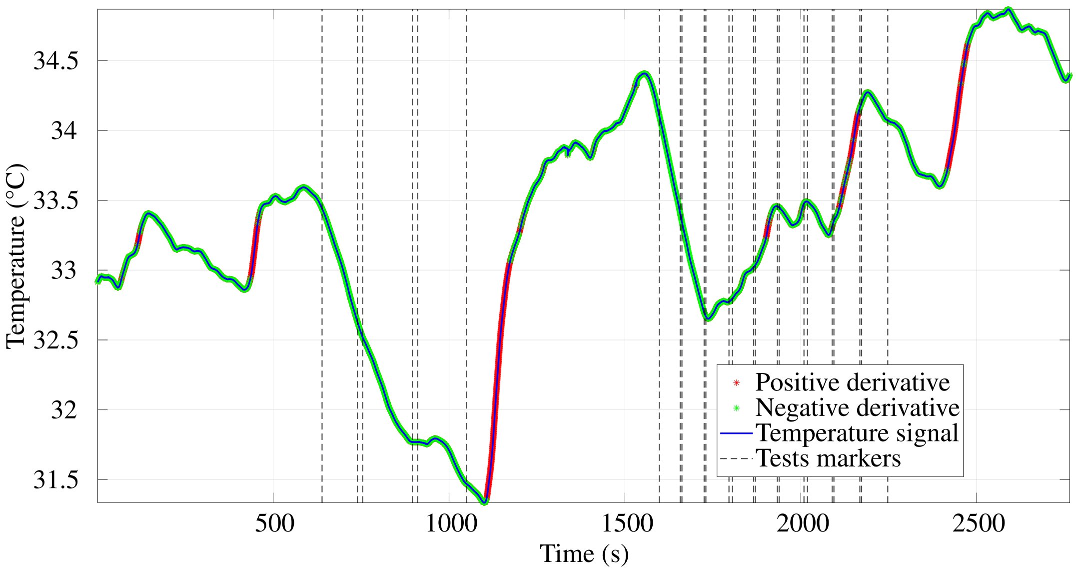


Figure 7: This Figure provides an example of the raw body temperature trend and displays the positive and negative values of the first-time derivative of the signal.

4. Results

The signal-processing strategy described in the previous Section resulted in 43 overall features per participant extracted for each test phase. We decided to consider the tests separately to analyse the obtained results, therefore considering four different datasets: Stroop, N-Back Visual, N-Back Auditory, and N-Back Dual. As shown in Figure 3, it is possible to observe that each of the four different tests was divided into three sub-phases. Thus, considering an addition *rest phase* calculated on the physiological data acquired during the initial Rest phase, it is possible to observe that each dataset was composed of the same amount of data divided into the four parts of the tests:

$$n\_data = n\_participants \times n\_phases \times n\_features = 4816 \quad (2)$$

where n_data is the overall number of data of each test dataset, $n\_participants = 28$ is the number of participants, $n\_phases = 4$ is the number of phases for each test, and $n\_features = 43$ is the number of features. Once obtained this result, a maximum-minimum normalization process per subject was performed on each feature for each phase to

DOI: 10.13009/EUCASS2023-349

hold a homogeneous dataset and perform the analysis described in the following.

4.1 Statistical analysis

We implemented a classical statistical approach to our four datasets separately to verify if the tests resulted in significant statistical differences in the aforementioned features based on the procedure presented in Figure 8. Therefore, the same process was repeated four times to analyse the features' significance in the Stroop and the Visual, Auditory, and Dual N-Back tests. This will allow us to understand if the variation of cognitive workload and perceived stress has a reflection on the physiological signals.

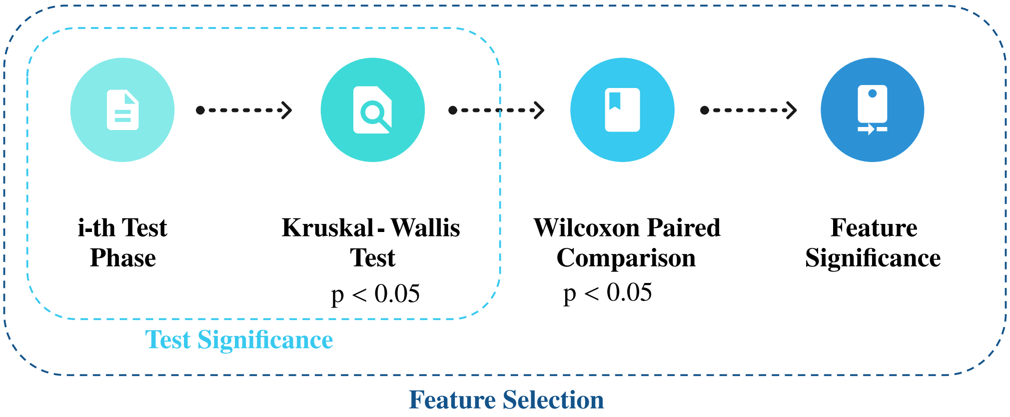
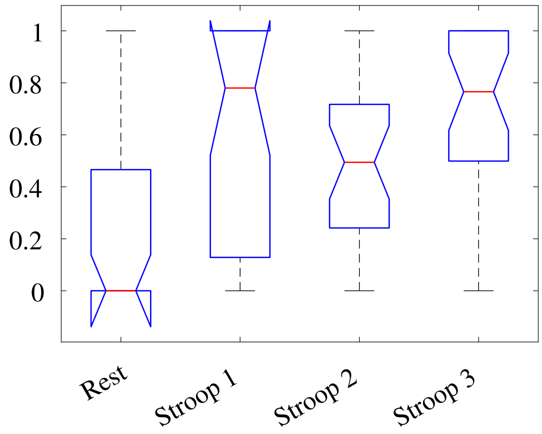


Figure 8: Data analysis statistical approach.

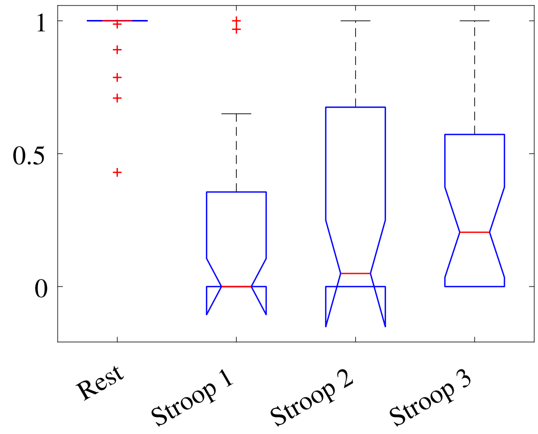
As shown in Figure 8, the first step was to evaluate if, for a specific feature, the considered test had significance in the variation of the data distribution between the rest phases and the three corresponding phases. Since it is not possible to assume a normal distribution for the data in each phase of the dataset, it was decided to implement a Kruskal-Wallis test. An example of the output provided by this method is reported in Figure 9a, comparing the distributions of the BPM mean between the four considered phases of the Stroop test. It is evident from this graph that the median of the distribution increases in the different phases in line with the increase in our external stress condition. More specifically, if the *p value* given by this test resulted in $p < 0.05$, we considered the test significant, and we could deeper analyse the effect of our test. The second step of our procedure was related to a deeper analysis of the relevance of each feature by specifically understanding which of the 43 considered features was indeed sensitive to the variation of the external cognitive workload or stress with respect to a rest condition. To do so, we adopted a Wilcoxon Paired comparison between the population of the Rest and each of the other three phases (for example, Rest vs Stroop 1, Rest vs Stroop 2, and Rest vs Stroop 3). In this analysis, we want to point out which feature could provide information between a condition of rest and a more stressed and mental demand. Also, in this case, we considered a statistically significant difference between the two populations when $p < 0.05$. The results obtained from this analysis are reported in Table 6. The first row notes the features that resulted significantly from the Kruskal-Wallis test. The second, third and fourth ones show the number of characteristics that demonstrated at least one, two or three respectively positive outputs in the comparison between the rest and the other three phases. This method allows us to define which of the initial 43 features is more sensible to the variation of external stress and visual, auditory and visual-auditory mental workload. In particular, we observed that the Stroop test produced the highest number of significant features for the other mental workload ones.

Moreover, noticing how many common characteristics are in the four datasets is interesting. Thus, by comparing the relative features corresponding to the fourth row of Table 6, we could obtain that 16 characteristics demonstrated to our tests a sensibility between the rest and each test phase. More specifically, these features are the standard deviation of the amplitude, duration and rise time of the *PPG shape*; the BPM standard deviation and the pNN50 related to the *time-domain HRV*; the standard deviation of the LF, HF, and LF/HF related to the *frequency-domain HRV*; the mean and standard deviation of the *SCL*; the initial, delta, standard deviation, and slope of the *body temperature*; and finally the *reaction time*.

DOI: 10.13009/EUCASS2023-349



(a) BPM mean population comparison. It is evident an increase in the feature according to the increase in external stress.



(b) pNN50 population comparison. It is evident that the feature decreases according to the increase in external stress.

Figure 9: Figure 9a represents the output of the Kruskal-Wallis test related to the BPM's mean in the four different phases of the Stroop dataset. Figure 9b shows another example of the implemented statistical comparison related to the pNN50 characteristics of the same dataset.

Table 6: The results obtained through the statistical procedure shown in Figure 8 are shown in this Table. For each dataset (represented in the columns) are reported the number of features that resulted significant through our test.

	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

5. Conclusions

Two key challenges facing the aviation market today are rising operating costs, especially on short- and medium-haul flights, and the decreasing availability of pilots. Despite today's regulations for most airline flights requiring a pilot and a co-pilot on board the aircraft to address these problems, there has recently been an increasing focus on the so-called Single Pilot Operations (SPOs). The SPOs in aviation can be achieved only by guaranteeing the same (or higher) safety and handling quality level regulated up to date in the EASA parts. Therefore, a cockpit assistant with the potential to understand the cognitive workload of pilots and their ability/inability to operate the aircraft is a fundamental need to foster this disruptive transition. Three main ways to assess these conditions are subjective, behavioural, and physiological evaluations. In this context, this paper investigates a multimodal approach based on three physiological signals: PPG, EDA, and body temperature. The idea is to estimate individual capacity limits by correlating the variation of the features extracted from these signals with the mental load and stress's increase or decrease during an operator's performance by considering the possible future application in SPOs. Because the sensors are unobtrusive, we specifically chose to monitor physiological values that might be measured on a single hand in order to be comfortable for pilots.

Due to the recent push in this field, a robust and well-established strategy to assess mental workload and stress still needs to be included. Therefore, initial computer cognitive tests, focusing specifically on these states (Stroop and N-Back test), have already been completed on 28 healthy volunteers, providing the aforementioned parameters under three mental workloads and stress conditions. Consequently, the most significant features have been extracted from the signals, providing a representative data set for studying the correlation between these physiological signals with stress and mental workload. These datasets were separately investigated through a statistical approach to evaluate the most sensitive features for each mental condition. Finally, 16 out of 43 most significant features were found commonly in each group. This result represents only the starting point for a deeper analysis based on AI algorithms. Still, it aims to validate the sensitivity of the considered physiological signals to the variation of external stress and cognitive workload. The next steps are to expand the population involved to validate our result and implement an AI algorithm

DOI: 10.13009/EUCASS2023-349

that can predict an operator's mental workload and stress starting from these physiological data. Once verified the functioning of this structure is, it will be possible to start the simulation and real operative environment tests. Indeed, we aim to develop a system that has versatility as the key principle to have the possibility to integrate this tool not only in an aircraft cockpit but also in the control tower to monitor the mental workload of air traffic controllers.

To sum up, we investigated the physiological multimodal approach to foster the transition toward SPOs. This is a crucial topic for the aviation industry, which is continuously growing thanks to the exponential growth of the biomedical sensor market in the last few years. The availability of smaller, cheaper, and more reliable wearable sensors allows for investigating and developing technologies that could not be realized so far to enhance safety and push the aviation sector to the next generation of aircraft.

6. Acknowledgments

This research work is supported by the Future Aircraft Technologies, Autonomy, and HMI Research Unit of *Leonardo s.p.a* and the European Union funds DM 1061. The authors sincerely thank both for their insights and for co-financing the PhD program from which this paper was born. We want to thank the *PolitoBIOMed Lab* of the Politecnico di Torino for allowing us to access the equipment and spaces for our tests.

References

[1] Paul Ayres, Joy Lee, Fred Paas, and Jeroen J. G. Van Merrienboer. The validity of physiological measures to identify differences in intrinsic cognitive load. *Frontiers in Psychology*, 12, 09 2021.

[2] Win-Ken Beh, Yi-Hsuan Wu, and An-Yeu Wu. Robust ppg-based mental workload assessment system using wearable devices. *IEEE Journal of Biomedical and Health Informatics*, 27(5):2323–2333, 2023.

[3] Yannick Brand and Axel Schulte. Workload-adaptive and task-specific support for cockpit crews: design and evaluation of an adaptive associate system. *Human-Intelligent Systems Integration*, 3, 06 2021.

[4] Joan Cahill, Paul Cullen, and Keith Gaynor. Interventions to support the management of work-related stress (wrs) and wellbeing/mental health issues for commercial pilots. *Cognition, Technology Work*, 22, 08 2020.

[5] Stephen Casner and Brian Gore. Measuring and evaluating workload: A primer, 07 2010.

[6] Rebecca L. Charles and Jim Nixon. Measuring mental workload using physiological measures: A systematic review. *Applied Ergonomics*, 74:221–232, 2019.

[7] G. E. Cooper and R. P. Harper Jr. The use of pilot rating in the evaluation of aircraft handling qualities. NASA Technical Note (TN) 69N22539, National Aeronautics and Space Administration (NASA), April 1969.

[8] Essam Debie, Raul Fernandez Rojas, Justin Fidock, Michael Barlow, Kathryn Kasmarik, S.G. Anavatti, Matt Garratt, and Hussein Abbass. Multimodal fusion for objective assessment of cognitive workload: A review. *IEEE Transactions on Cybernetics*, PP:1–14, 09 2019.

[9] EASA. Easa ai roadmap. Technical report, European Union Aviation Safety Agency, 2023.

[10] M. S. Young G. M. Hancock, L. Longo and P. A. Hancock. Mental workload. In *Handbook of Human Factors and Ergonomics*, volume 7, pages 203–226. 2021.

[11] Alberto Greco, Gaetano Valenza, Antonio Lanata, Enzo Scilingo, and Luca Citi. cvxeda: A convex optimization approach to electrodermal activity processing. *IEEE Transactions on Biomedical Engineering*, 2016:797–804, 04 2016.

[12] Sandra G. Hart and Lowell E. Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. In Peter A. Hancock and Najmedin Meshkati, editors, *Human Mental Workload*, volume 52 of *Advances in Psychology*, pages 139–183. North-Holland, 1988.

[13] ICAO. 2021 safety report. Technical report, International Civil Aviation Organization, 2021.

[14] Nikolai Janczewski, Jennifer Wittmann, Arnd Engeln, Martin Baumann, and Lutz Krauß. A meta-analysis of the n-back task while driving and its effects on cognitive workload. *Transportation Research Part F: Traffic Psychology and Behaviour*, 76:269–285, 01 2021.

DOI: 10.13009/EUCASS2023-349

[15] A Kistler, C Mariauzouls, and K von Berlepsch. Fingertip temperature as an indicator for sympathetic responses. *International journal of psychophysiology : official journal of the International Organization of Psychophysiology*, 29(1):35â41, June 1998.

[16] Min Li, Miao Wang, Dongjin Ding, and Guoqing Wang. Development and evaluation of single pilot operations with the human-centered design approach. *Aerospace*, 9(10), 2022.

[17] Michael Matessa, Thomas Z. Strybel, Kim-Phuong L. Vu, Vernol Battiste, and Thomas Schnell. Concept of operations for rco spo. 2017.

[18] Ryan Mckendrick, Bradley Feest, Amanda Harwood, and Brian Falcone. Theories and methods for labeling cognitive workload: Classification and transfer learning. *Frontiers in Human Neuroscience*, 13, 09 2019.

[19] Geoff Murray and Rory Heilakka. The airline pilot shortage will get worse. Technical report, Oliver Wyman, 2023.

[20] Hugo Posada-Quintero and Kaye Chon. Innovations in electrodermal activity data collection and signal processing: A systematic review. *Sensors*, 20:479, 01 2020.

[21] Federica Scarpina and Sofia Tagini. The stroop color and word test. *Frontiers in Psychology*, 8:557, 04 2017.

[22] Cornelia Setz, Bert Arnrich, Johannes Schumm, Roberto Marca, Gerhard Tröster, and Ulrike Ehlert. Discriminating stress from cognitive load using a wearable eda device. *IEEE Transactions on Information Technology in Biomedicine*, 14:410–417, 01 2010.

[23] Bram Van Acker, Davy Parmentier, Peter Vlerick, and Jelle Saldien. Understanding mental workload: from a clarifying concept analysis toward an implementable framework. *Cognition, Technology Work*, 20, 08 2018.

[24] Hao Wan, Liujing Zhuang, Yuxiang Pan, Fan Gao, Jiawei Tu, Bin Zhang, and Ping Wang. *Biomedical sensors*, pages 51–79. 01 2020.

[25] Zongmin Wei, Damin Zhuang, Xiaoru Wanyan, Chen Liu, and Huan Zhuang. A model for discrimination and prediction of mental workload of aircraft cockpit display interface. *Chinese Journal of Aeronautics*, 27(5):1070–1077, 2014.

[26] Christopher Wickens. Multiple resources and performance prediction. *Theoretical Issues in Ergonomic Science*, 3:159–177, 01 2002.

[27] Mark Young, Karel Brookhuis, Christopher Wickens, and Peter Hancock. State of science: mental workload in ergonomics. *Ergonomics*, 58:1–17, 12 2014.

[28] Yu Zhang, Hongtao Zheng, Yunshan Duan, Meng Li, and Lu Zhang. An integrated approach to subjective measuring commercial aviation pilot workload. pages 1093–1098, 06 2015.