POLITECNICO DI TORINO
Repository ISTITUZIONALE

A review of physiological measures for mental workload assessment in aviation

Original
A review of physiological measures for mental workload assessment in aviation / Luzzani, G., Buraioli, I., Demarchi, D., Guglieri, G.. - In: THE AERONAUTICAL JOURNAL. - ISSN 0001-9240. - STAMPA. - (2024), pp. 1-22. [10.1017/aer.2023.101]

Availability:
This version is available at: 11583/2983312 since: 2023-10-30T09:06:23Z

Publisher:
Cambridge University Press

Published
DOI:10.1017/aer.2023.101

Terms of use:

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

Publisher copyright
Cambridge University Press postprint/Author's Accepted Manuscript

This article has been published in a revised form in THE AERONAUTICAL JOURNAL http://doi.org/10.1017/aer.2023.101. This version is free to view and download for private research and study only. Not for re-distribution or re-use. © copyright holder.

(Article begins on next page)

19 November 2026

Page 1 of 29. © Royal Aeronautical Society YYYY

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

G. Luzzani[1]
gabriele.luzzani@polito.it
I. Buraioli[2], **D. Demarchi**[2] **and G. Guglieri**[1]

[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

Abstract

The relevant growth of Human-Machine Interaction (HMI) systems in recent years is leading to the necessity of being constantly aware of the cognitive workload level of an operator, especially in a safety-critical context such as aviation. Since the confusion in the definition of this concept, this paper clarifies this terminology, also highlighting its relationship with stress. Thus, we analysed the State-of-the-Art of cognitive workload evaluations, showing three up-to-date methodologies: subjective, behavioural, and physiological. In particular, the physiological approach is increasingly gaining attention in the literature due to today's exponential growth of biomedical sensors. Therefore, a review of the most adopted physiological signals in the workload evaluation is provided, focusing on the aeronautical field. We conclude by highlighting the necessity of a multimodal approach for mental workload assessment as a result of this analysis.

.

NOMENCLATURE

AoI	Area of Interest
BR	Breath Rate
BRO	Blink-Related Oscillation
CWL	Cognitive Workload
DBN	Dynamic Bayesian Network
ECG	Electrocardiogram
EDA	Electrodermal Activity
EEG	Electroencephalography
EMG	Electromyography
EOG	Electrooculography
fNIRS	functional Near Infrared Spectroscopy
GSR	Galvanic Skin Response
HFE	Human Factors and Ergonomics
HR	Heart Rate
HRV	Heart Rate Variability
ICAO	International Civil Aviation Organization
LoA	Level of Autonomy
MCH	Modified Cooper Harper
MUM-T	Manned-Unmanned Teaming
MWL	Mental Workload
NASA TLX	National Aeronautics and Space Association Task Load Index
OTM	One-to-Many
PPG	Photoplethysmography
RCO	Reduced Crew Operations
RPE	Retinal Pigmented Epithelium
SA	Situation Awarness
SC	Skin Conductance
SCL	Skin Conductance Level
SCR	Skin Conductance Response
SPO	Single Pilot Operation
SVM	Support Vector Machines
TSVD	Time Varying Singular Value Decomposition
TV	Tidal Volume
UAS	Unmanned Aircraft System
UAV	Unmanned Aerial Vehicle
USAF	United States Air Force

1.0 Introduction

The achievement of an ever-increasing level of safety has always been a key element of Human-Machine Interaction (HMI) systems. Thanks to the disruptive growth of the technological market in the last few years, it is increasingly common to find contexts where an operator interacts with a machine that exhibits a certain level of automation; hence the necessity of always guaranteeing an adequate safety level during a performance[1]. These scenarios can be found in everyday life, such as when driving a car and in specific jobs, as in the case

of Industry 4.0 operators, air traffic controllers, or pilots[2]. Moreover, the exponential rise of new safety-critical contexts, such as the space market, brings new challenges in developing and designing appropriate HMI systems[3][4].

The concept of safety is particularly important in the aviation sector. Because of the major role played by the air transport industry in the global economy, the key element to maintaining its vitality is to ensure safe, secure, efficient, and environmentally sustainable operations at the global, regional, and national levels. As shown in Fig. 1, this is demonstrated by data presented in the 2021 Safety Report by the International Civil Aviation Organization (ICAO), highlighting a decreasing trend in accident rate that occurred over the last ten years. These results can be explained by the improvement in safety during flight operations[5].

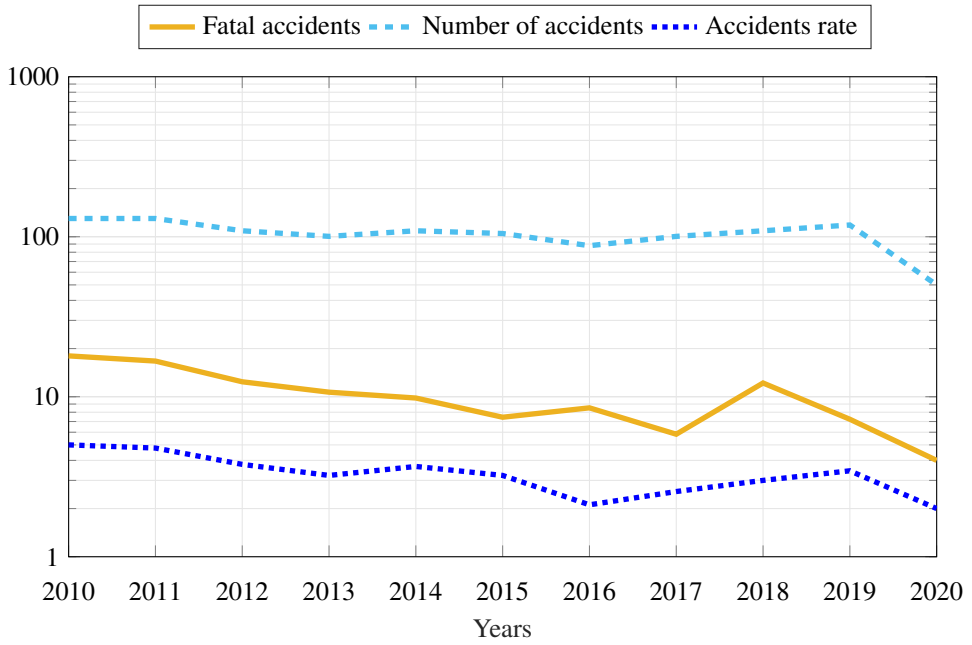


Figure 1: This Figure shows the aeronautical accidents trend from 2010 to 2020[5]. It highlights a decreasing trend in the accidents per year, fatal accidents per year, and the accident rate (defined as the number of accidents per million departures).

According to the Federal Aviation Administration (FAA) Report[6], 60-80% of fatal accidents can be related to human errors, and around half of them occur during the highest workload phases of flight missions[7]. These correspond to the take-off and landing phases when the crew is dealing with the highest workload, and the airplane is in proximity to the ground, reducing manoeuvre margins in case of an emergency. Thus, pilot workload monitoring has always been a central research topic in enhancing civil and military aviation safety levels, especially through subjective assessments, in-flight training, and specific aircraft and cockpit design.

Moreover, despite today's regulations requiring a pilot and a co-pilot on board the aircraft, in recent years, attention to the so-called Single Pilot Operations (SPOs) is growing in the aeronautical field. The term SPO refers to the condition of flying an aircraft with only one

pilot on board, assisted by advanced system automation and/or ground operators providing piloting support services[8]. In order to maintain a high level of safety during flight missions, the need to develop a cockpit assistant to monitor in real-time the cognitive state of health of the pilot represents a key point in both the civil and military aeronautical fields.

As the military aeronautical field is concerned, the next-generation fighters will be provided with the capability of managing a fleet of Unmanned Aerial Vehicles (UAVs). The domain of Manned-Unmanned Teaming (MUM-T), where (multiple) unmanned aerial vehicles are operated from an aircraft cockpit, causes an even broader range of tasks for the cockpit crew. Although several approaches successfully showed that operating multiple UAVs from a cockpit is possible, the challenge of unbalanced and possibly overtaxing workload conditions remains an open question. Hence the need for a cockpit assistant system that is able to understand the pilot's mental workload and consequently modulate the Level of Autonomy (LoA) of the UAVs and the fighter itself. This allows the pilot to always operate in an optimal workload condition and in the most demanding military environments[9].

On the other hand, the push toward the SPOs in civil aviation is led by the necessity of reducing operating costs. Significant cost savings can be produced by halving the number of pilots, especially in smaller regional aircraft operated on shorter routes, which may not be economically viable with higher capacity airliners. In fact, direct operating costs attributable to flight deck crew rise as aircraft size decreases. Moreover, a second driver for this transition is represented by the issues relating to a potential shortage of commercial pilots in the near future. However, the benefits of SPOs in civil aviation can be achieved only by guaranteeing the same (or higher) safety and handling quality level regulated to date in the EASA parts[10]. Therefore, a cockpit assistant with the potential to understand the mental workload of the pilot and his/her ability/inability to operate the airliner is a fundamental need to foster this disruptive transition.

For this reason, understanding how the evaluation of the cognitive state of an operator has been addressed to date is important for developing the next generation of intelligent systems in HMI environments. This paper aims to benefit in this direction by reviewing the State-of-the-Art of physiological approach for an operator's cognitive mental workload monitoring, focusing on aircraft pilots.

This paper is structured as follows. Section 2.0 clarifies the definition of mental workload. Then, Section 3.0 introduces the concept of the physiological approach for the mental workload assessment by describing the most significant signals related to this methodology. Section 4.0 investigates the link between these physiological signals and the variation of an operator's mental workload, analysing twenty-nine papers selected from a State-of-the-Art study. Finally, in Section 5.0, the conclusions are drawn.

2.0 Mental Workload Definition

The concept of mental workload (MWL) has been studied for decades by several influential texts and chapters in the Human Factors and Ergonomics (HFE) field[11]. However, due to its complex and multidisciplinary nature, a unique and universally accepted definition is still missing[31]. For example, from a psychological perspective, Wickens defined MWL in his *Multiple Resource Theory*[14] as the relationship between a task's (quantitative) resource demands and the operator's capacity to meet those demands. Young et al.[1] considered the influence of the external and internal environment, defining MWL as the "product of the de-

mand/s of the task and the capacity/ies of the person performing the task, where demands and capacities may be moderated by context." In general, lots of conceptualizations of MWL consist of a multidimensional idea structured in four main components, which can be represented by Debie et al.[2] in Fig. 2: (1) *Task Load*: the amount of work or external duties that the operator has to perform; (2) *Mental Load*: the level of mental resources an individual is able to supply to maintain a high-performance level while performing a task; (3) *Depletion Factors*: internal and external factors such as stress, fatigue, motivation, attitude, etc.; (4) *Performance*: the outcome the operator is performing.

It is important to highlight that the definition of Mental Load in Fig. 2 represents the component related to the mental load of an operator without considering external factors, while the MWL concept, which this paper aims to address, is represented by the mental condition that Debie et al.[2] referred in their work as Cognitive Load. Therefore, the possible confusion between mental workload (MWL) and cognitive workload (CWL) is evident, seeming to imply two different constructs. However, according to Hancock et al.[11], "the two terms address one issue and should be treated to mean exactly the same thing" therefore, in the following of this text, *MWL and CWL will be treated as synonymous*.

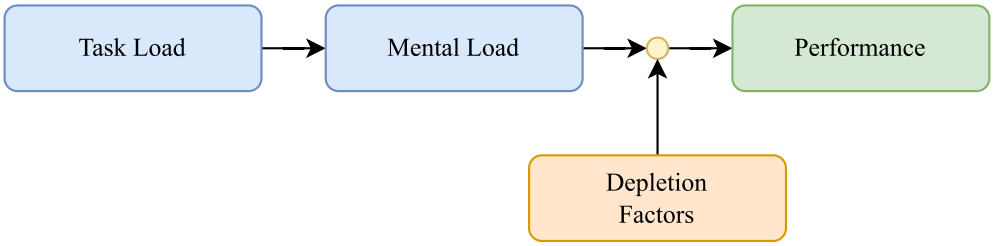


Figure 2: Cognitive mental workload scheme[2]. The figure shows how an operator's performance is influenced by depletion factors (internal or external) and the imposed task load.

Despite the multifaceted characteristic of MWL, some recent studies tried to clarify this concept, providing an implementable framework. An important contribution is given by Van Acker et al.[12], finding the defining attributes of MWL (the features that frequently recur while examining definitions and descriptions and that must exist for the concept to occur) in: task demands and the operator's neural capabilities interaction; the dynamic distribution, supervision, and use of mental resources; and the triggering of a linked subjective response. From this analysis, they defined MWL 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."

Therefore, thanks to the comprehensive approach of this work and its recent formulation, we treat MWL or CWL as referring to this definition of Van Acker[12].

2.1 Stress

A concept that often appears with MWL in the literature is *stress*. Thus, it is necessary to clarify what this term refers to and its link with CWL. The meaning of stress depends on the field of application, and it is different for different people in different situations. There has

been much debate about this theme in the scientific community leading to a not unique definition of it[16]. Regarding the aviation sector, this concept can be associated with "the response of the body to stimuli that affect the normal physiological balance of a person, causing physical, mental or emotional strain"[7]. This condition is not easy to assess because something stressful for one person is not necessarily the same for another. Moreover, although a stressful condition in the same instance may positively affect an operator, focusing his attention and vigilance, most of the time, this leads to unsafe situations, such as neglecting key tasks a pilot.

Stimuli that can lead to a stress condition are called *stressors*, and the most common ones in aviation are noise, uncomfortable temperature, vibrations, and high workload levels[7]. When the pilot has to accomplish too many tasks in too little time, then potential risk conditions arise. Thus, a direct cause-effect link exists between mental workload and the pilot's stress. Referring to Fig. 2, the stress condition can be inserted in the *Depletion Factor* block, representing an internal condition influencing the cognitive workload of an operator.

3.0 Methods for Mental Workload Assessment

An excessive or limited mental workload can lead to situations that are dangerous, especially when managing complex systems like aircraft, cars, or the airport's control tower, causing errors and accidents. Thus, it is clear the necessity to monitor the mental workload of an operator during its performance. The literature shows three up-to-date cognitive workload measurement methodologies: subjective, behavioural, and physiological[15]. In particular, subjective evaluations consist of questionaries submitted to pilots at the end of the performance. Behavioural measures are based on monitoring pilot activities during the flight mission and checking the correspondence between the performed actions and those planned beforehand. Finally, physiological measures are related to monitoring the operator's physiological signals to infer his cognitive workload condition.

Whereas subjective evaluations and behavioural measures are methods studied and adopted for several years, the physiological MWL assessment has been increasingly used in recent years due to the improvement of sensor technologies, leading to unobtrusive measures providing objective and accurate information about MWL[13]. Therefore, this paper focuses on the most adopted physiological signals for MWL evaluation, while providing a general overview of subjective evaluations and behavioural measures.

Subjective Evaluation

As said before, subjective evaluations are post-performance self-reported measures used more frequently than other techniques, mainly because they are inexpensive and simple to administer. The most widely used of these tools is the *National Aeronautics and Space Association Task Load Index* (NASA TLX). The NASA TLX is so frequently adopted that this survey has become synonymous with the concept of mental workload, and it is considered the gold standard in workload measurement in human-computer interaction studies[17]. The NASA TLX is a multi-dimensional rating scale that adopts six workload dimensions to provide diagnostic information about each dimension's nature and relative contribution in influencing overall operator workload. These six drivers are: mental demand, physical demand, temporal demand, performance, effort, and frustration[18]. However, the NASA TLX is not exempt from some limitations. These include how sources of stress are integrated, the tool's lack

of construct validity, its sensitivity to changes in workload over time, and how the NASA TLX can be a source of workload[19]. Due to these drawbacks, behavioural data, physiological measures, and subjective metrics are frequently combined. Other examples of subjective evaluations are the uni-dimensional Bedford scale, designed to identify an operator's spare mental capacity while completing a task, and the uni-dimensional Cooper-Harper and Modified Cooper-Harper (MCH) scales. The former was developed to evaluate the aircraft's handling qualities[20], while the latter assumes a direct relationship between the level of difficulty of aircraft controllability and pilot workload[21].

Behavioural Evaluation

Monitoring pilots' behaviour in the cockpit during a flight mission is another effective way to infer their cognitive workload. This approach is performed in the literature through primary and/or secondary task performance measures. The former consists of monitoring the operator's performance through variables linked with the primary task, represented by the main focal point of the given domain, for which we are attempting to determine the MWL's level[11]. For instance, checking if the pilot pressed a button, fixated on a specific display, or answered a control tower's question can provide useful information about pilots' workload condition compared to a reference, a priori-defined activities plan. The latter consists of a task designed to compete with the same cognitive resources as the primary one. Thus, the operator's MWL can be inferred by analysing the performance decreasing on secondary tasks while maintaining attention on primary tasks[11]. A key contribution in this field is given by A. Schulte from the *Universität der Bundeswehr, München*, thanks to developing a workload-adaptive and task-specific associate system for military helicopters and fighters. This system acts as an additional artificial crew member that can monitor the pilot's activities, confront them with a reference plan and understand their mental workload. Then, by assessing the crew's workload and the existing and upcoming task conditions, it adapts the amount of its support to maintain the pilot's workload always in an optimal state, avoiding overload situations[22]. The behavioural approach is based on[23]: (1) a task model that consists of the characterization of each duty that the crew can operate during a mission, described by properties, and connected to other functions by relations, based on the *Multiple Resource Model* introduced by Wickens in[24][25][26]; (2) a mission plan that represents the pilot's current tasks and those to be performed in the future; (3) activities that are "those elementary tasks that the user performs in parallel at a certain time"[22]. Therefore, by checking during the mission the execution of the planned activities and correlating the missed ones with their quantified workload value, it is possible to determine if the pilot is in a tolerable MWL condition. The MUM-T represents a typical implementation of this concept to improve future military aviation[22]. The same MUM-T application was adopted by Chen et al.[27] to develop a specific mathematical behavioural model that allows the evaluation of the cognitive workload by adopting three factors: (1) pilot utilization factor that quantifies the effort in the interaction process with the UAVs; (2) UAV request rate that is linked with the number of UAV's in the fleet, and (3) the number of humans–robotics interactions.

3.1 Physiological Evaluation

The relationship between the mental workload of an operator and the variation of his physiological signals is an issue that has been investigated for decades. As early as 1979, Moray

described in the proceeding of a NATO conference the different aspects associated with the mental workload by also referring to physiological measures[28]. However, apart from a few studies developed during the years[29], only in the last ten years is it possible to observe an increasing growth of research about this topic. This phenomenon can be easily explained by observing the exponential development in the sensor technologies field. Indeed, the biomedical technology market is growing as fast as smartphones and the Internet[30]. Reliable and cheaper wearable sensors are therefore available on the market, making physiological measures easier and more accurate also in complex conditions like an aircraft cockpit or a car. This allows for a deeper investigation of how the variation of cognitive workload can influence the human body[31]. Furthermore, the literature shows several signals that can be related to the mental workload of an operator during a performance: heart activity, skin activity, eye activity, brain activity, respiration, body temperature, muscle activation, and voice patterns. Nevertheless, based on previous studies' output, only the following signals are the most significant[2][18][31]: *Heart Activity*, *Skin Activity*, *Eye Activity*, *Brain Activity*, and *Respiration*. Before analysing how these signals are related to the variation of mental workload, a physiological description is provided hereinafter.

3.1.1 Heart Activity

Heart activity can be measured with the electrocardiogram (ECG) or photoplethysmography (PPG), and it is the most adopted physiological signal that has ever been related to cognitive workload. In Kramer's review in 1991, the usefulness of this signal to assess mental workload was highlighted[32]. Several features can be obtained from the ECG or the PPG, which adoption varies in the field of application[33]. In particular, Heart Rate (HR) and Heart Rate Variability (HRV) are the most adopted parameters in the literature that can be related to cognitive workload variations. These characteristics can be inferred from the ECG and PPG signals:

- As shown in Fig. 3, ECG represents the electrical activity of the human heart. Its pattern is composed of five waves P, Q, R, S, and T. By evaluating the distance between the QRS complexes (for example, through artificial neural networks, genetic algorithms, adaptive threshold, etc.), it is possible to obtain HR, and the HRV[34].

- Heart activity can also be monitored through the PPG. This consists of measuring blood volume changes in a microvascular bed of the skin based on optical properties, such as absorption, scattering, and transmission properties of human body composition under a specific light wavelength[35]. The peak-to-peak interval of this signal is adopted to infer HR and HRV. Compared to measurements obtained by ECG, the difference is the time it takes the pulse wave to travel from the heart to the spot where the PPG sensor is placed[36].

3.1.2 Eye Activity

Thanks to the recent significant development of ocular technologies, eye activity measures are becoming more and more efficient for assessing an operator's MWL[31]. The monitoring of eye movements has been studied for several years in different fields of application (such as education, user experience, and the study of cognitive learning processes)[51]. However, only with the development of portable and unobtrusive eye-tracker, this technology has been introduced in complex environments, and work conditions[18]. Therefore, eye activity analysis is

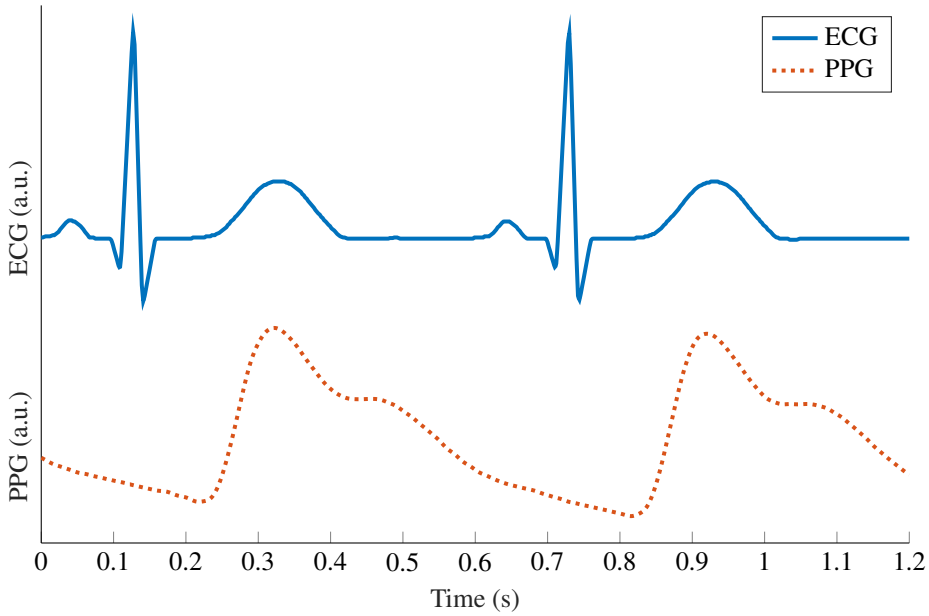


Figure 3: This Figure represents the cardiac cycle measured through both ECG and PPG. The time delay between the peaks of the two signals is due to the blood pulse wave traveling from the heart to the spot where the PPG sensor is placed (in this case, it is represented by the time delay of about 0.2 s, typical of a finger PPG).

related to tracking gaze and pupil measures, which means monitoring the oculomotor events that function as the basis for several eye movement and pupil assessments: fixations, saccades, smooth pursuit, blinks, and vergence[52]. During MWL analysis, only some parameters are usually considered: blinks, fixations, saccades, and pupil dilation. In fact, *fixations* represent the period when our visual gaze remains at a particular location and *saccades* are rapid movements between fixations. Common analysis metrics include fixation or gaze durations, fixation rate, transition rate, saccadic amplitudes, saccadic velocities, and various transition-based parameters between fixations and/or regions of interest[53]. Moreover, according to Ayres et al.[18], many studies have demonstrated that pupil dilation is positively correlated with the cognitive workload imposed by the tasks. The eye blink rate can be measured through Electrooculography (EOG), which consists of "recording the standing corneal–retinal potential arising from hyperpolarisations and depolarisations existing between the cornea and the retina"[54].

3.1.3 Skin Activity

An increasingly important physiological feature in the study of MWL is the monitoring of skin activity, particularly the so-called electrodermal activity (EDA). This represents an electrical manifestation of the sympathetic innervation of the sweat glands, where their function modifies the conductance of an applied current. This modification allows for measuring the changes in electric skin conductance (SC), resulting in an increase in electrical conductance

in accordance with sweating[58]. The interesting characteristic of this signal is that the sweat glands, thus the EDA signal, are more responsive to psychological stimuli rather than thermal stimuli; therefore, emotional information can be inferred[59]. As shown in Fig. 4, the SC signal can be decomposed into a tonic and a phasic component that are respectively called Skin Conductance Level (SCL) and Skin Conductance Response (SCR). These signals have different time scales and relationships to exogenous stimuli and can be expressed as follows[60]:

$$SC = SCR + SCL \tag{1}$$

- SCL represents the tonic component of the EDA signal, and it is a measure related to the slow shifts of the EDA. The SCL is typically computed as the mean of several measurements taken during a specific non-stimulation rest period.
- SCR is the phasic component of the EDA. It reflects the short-time response to the stimulus. The typical shape of the SCR comprises a relatively rapid rise from the conductance level followed by a slower asymptotic exponential decay back to the baseline. This EDA's component is also called Galvanic Skin Response (GSR).

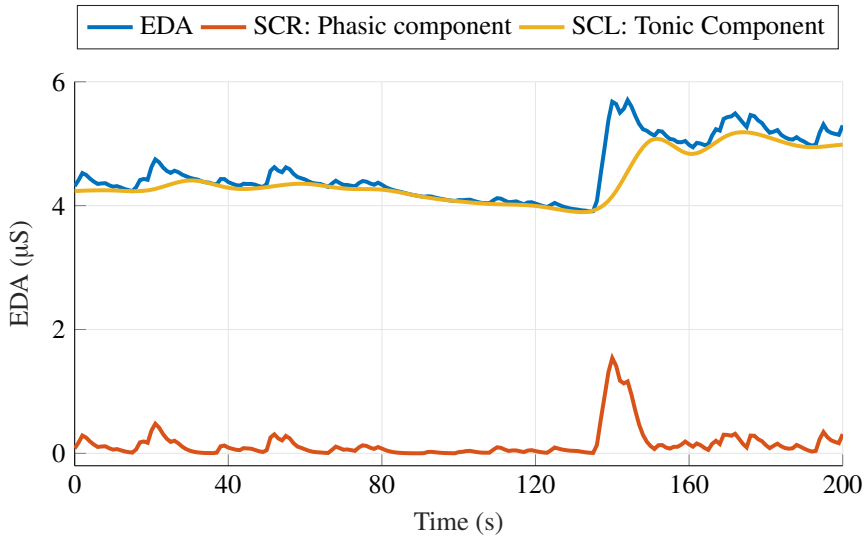


Figure 4: Electrodermal activity (EDA) can be divided into two components. The tonic and slower one is the Skin Conductance Level (SCL), while the phasic and faster component is represented by the Skin Conductance Response (SCR).

3.1.4 Brain Activity

In MWL assessment, monitoring brain activity has always played a key role. Understanding how the brain works during a performance is extremely important because as cognitive workload increases for a specific task, available brain resources for other not primary tasks decrease. The cognitive processing of further tasks may be delayed or hampered if the brain's

resources are depleted below a threshold point[63]. Throughout the years, the assessment of brain activity related to MWL has usually been led with methodologies like electroencephalography (EEG), functional magnetic resonance imaging (fMRI), and functional near-infrared spectroscopy (fNIRS)[18]. However, today the most adopted ones are EEG and fNIRS because of the improvement of sensors, making them portable, easy to use, and reliable[2]. These signals represent two different ways to monitor the MWL:

- EEG is a technique that measures electrical potential from the brain and represents electrically sensed signals over time that can be decomposed in the frequency domain. Usually, this signal is categorized into four frequency-based bands: Delta (< 4 Hz), Theta (4 Hz to 8 Hz), Alpha (8 Hz to 13 Hz), and Beta (13 Hz to 30 Hz)[64]. A key feature of the EEG signal is represented by Event-Related Potentials (ERPs) representing small changes in the brain's electrical activity recorded from the scalp, realised by internal or external events[65].
- fNIRS consists of a non-invasive tool to evaluate regional tissue oxygenation at the bedside continuously. fNIRS exploits the scalp and skull's transparency to infrared light and the differences in absorption spectra between oxyhemoglobin and deoxyhemoglobin to quantify the local oxygen saturation of hemoglobin in the brain respiration, thus highlighting the activation or deactivation of specific brain areas[66] For instance, fNIRS evaluations over the prefrontal cortex (PFC) have a direct link to working memory, decision-making, and executive control, which are all aspects directly related to MWL.

3.1.5 Respiration

In everyday life, whenever a person performs actions involving mental effort, it is evident how breathing is also affected[72]. According to Ayers et al.[18], the relationship between respiration and MWL has been studied for decades, highlighting how behavioural, psychological, and metabolic processes affect respiratory measures. Nevertheless, this signal did not obtain the same attention in MWL evaluation as ocular or cardiac signals; therefore, it is possible to find less research on this topic in the literature. Respiratory monitoring can generally be assessed through non-invasive contact or non-contact methods. The former is related to evaluating one of the following characteristics: respiratory airflow, respiratory sounds, air temperature, air humidity, air components, respiratory-related chest or abdominal movements, and modulation of cardiac activity[77]. The latter, instead, can be obtained by leveraging radar, optical, and thermal technologies[78]. However, the main features that have been extrapolated from the respiratory signal are those based on time and volume (and their variability), in particular, the Breath Rate (BR) and the Tidal Volume (TV).

- Breath Rate represents the number of respirations per minute.
- Tidal Volume is the volume of air that is inhaled and exhaled with each breath[73].

4.0 Biosignals for Physiological Workload Evaluaion

Based on the growing interest in the link between physiological signals and MWL highlighted in the previous section, it is necessary to investigate the State-of-the-Art of biosignals for a physiological workload evaluation. This research started in March 2022 and lasted three months.

The methodology adopted for selecting the papers on which to conduct our analysis consisted in searching the keywords "mental workload" OR "cognitive workload" OR "stress assessment" together with "physiological" OR each of the aforementioned physiological signals on the databases Google Scholar (core collections), Web of Science, and IEEE Xplore Digital Library. A second research was performed by introducing "pilot" OR "aviation" on the same databases. This resulted in 1483 overall papers, as reported in Table 1.

Table 1: This table reports for each database the number of papers obtained during our research by inserting the keywords "mental workload", "cognitive workload", and "stress assessment" together with each of the words reported in the first column of the table. As it is possible to observe, the core results were obtained through the Google Scholar database.

	Google Scholar	**Web of science**	**IEEE Xplore**
Physiological	306	183	26
Heart	182	153	24
Respiration	10	5	2
Skin	33	14	1
Eye	167	71	15
Brain	193	90	8
TOT		1483	

In this analysis, we first eliminated duplicates after acquiring the results reported in Table 1. We then screened the papers by reading the titles and abstracts. Finally, we narrowed the number of selected articles by reading the full text from the subset obtained. The inclusion criteria were the following: (1) peer-reviewed journal articles, book sections, conference papers, and technical reports written in English; (2) papers that assessed the MWL evaluation directly through the physiological signals without introducing time limits; (3) papers that adopted significant and consistent equipment and tests/simulations; (4) papers that performed a structured and consistent analysis of the physiological signals related to evaluating different levels of MWL (for example through statistical or machine learning techniques). We focused on articles that can give us a comprehensive information collection and exclude similar results. We manually searched the reference lists of the publications determined to be pertinent to find additional sources. Moreover, priority was given to papers considering several signals simultaneously. As a result, twenty-nine papers were selected and analysed in the following paragraphs (shown in Table 2).

A subsection is introduced for each of the signals presented in Section 3.0. The aim is to report the goals and results obtained in the research of each paper. In particular, the focus is on investigating the relationship between the aforementioned physiological signals and the variation of MWL with a general perspective, followed by a focus on their aeronautical application. Moreover, the features adopted by the selected papers for each signal and their relationship with the MWL variation are reported in Table 3 to Table 7. The ($\uparrow$) symbol represents an increase of the feature as the MWL increases, the ($\downarrow$) symbol a decrease, and the ($-$) character an uncertain trend. In these tables, the trends that are generally possible to observe from the literature (and therefore in the majority of the papers considered in our study) are reported. As pointed out above, due to the high subjectivity of the topic and the dependence

on the application context, these trends should not be taken as an absolute truth but more as a general trend found in the literature and, in particular, in the papers discussed in this article. Furthermore, the paper's characteristics are summarised in two tables: Table 2 reports: the reference, the name of the authors, and the physiological signals adopted in their research, marked with a (•) sign; Table 8 shows the implemented methods, the kind of test, the involved population, and the year of publication.

Table 2: State-of-the-Art comparison of the physiological signals for mental workload assessment. The analysed papers and relative biosignals are reported and compared.

Ref	Authors	Heart Activity	Eye Activity	Skin Activity	Brain Activity	Respiration	Others
(15)	*McKendrick et al.*				•		
(29)	*Wilson et al.*	•	•	•	•		•
(33)	*Hughes et al.*	•					
(37)	*Miyake et al.*	•		•		•	
(38)	*Gable et al.*	•	•				
(39)	*Morere et al.*	•		•			•
(40)	*Katsis et al.*	•	•	•		•	•
(41)	*Liao et al.*	•	•	•			•
(42)	*Giorgi et al.*	•	•	•			
(43)	*Pongsakornsathien et al.*	•	•		•	•	•
(44)	*Lim et al.*	•	•		•	•	
(45)	*Planke et al.*	•	•		•	•	•
(46)	*Wang et al.*	•	•			•	
(47)	*Mansikka et al.*	•					
(48)	*Wei et al.*	•					
(49)	*Gentili et al.*	•			•		
(50)	*Martin et al.*	•	•				
(55)	*Belkhiria et al.*		•		•		
(56)	*Vanneste et al.*		•	•	•		
(57)	*Dilli Babu et al.*		•				
(61)	*Setz et al.*			•			
(62)	*Ghaderyan et al.*			•			
(67)	*Rebsamen et al.*				•		
(68)	*Svinkunaite et al.*				•		
(69)	*Hajra et al.*				•		
(70)	*Herff et al.*				•		
(71)	*Mohanevelu et al.*				•		
(74)	*Romine et al.*	•		•			•
(76)	*Salomone et al.*						•

4.1 Heart Activity

The literature put in evidence how HR and HRV are highly sensitive physiological features for assessing mental workload (MWL) variations in general applications [18] [37]. In particular, Hughes et al. [33], studying the impact of workload manipulations on various cardiac measurements, showed that the sensitivity of the HR and HRV to cognitive demands reflected relatively stable effects. Similarly, Gable et al. [38] compared the commonly used measure

of heart rate (HR) to that of pupil size (PS) in detecting changes in CWL in a driving environment. They obtained that the HR of drivers while performing an n-back test increased according to the variation of MWL. Different stress levels were successfully discriminated by Zhang et al.[39] considering the participants' HRV and other physiological signals. In this case, the parameters were classified based on a Support Vector Machines (SVM) model. The same classifier was adopted by Katsis et al.[40] that considered HRV to develop an integrated telemedicine platform for assessing affective physiological states. They observed that the system's classification accuracy into five predefined emotional classes (high stress, low stress, disappointment, euphoria, and neutral face) reached 86.0%. Furthermore, a heterogeneous approach considering HRV was also adopted by Liao et al.[41], who modeled a user's stress and mental workload through a Dynamic Bayesian Network (DBN) framework, showing that the inferred user stress level was consistent with that predicted by psychological theories. Another approach was used by Giorgi et al.[42] that considered some physiological measures and HR, obtained with a PPG sensor, to assess users' workload, stress, and emotional states during specific tests. They demonstrated the reliability of today's wearable devices for this kind of application and the capability of discriminating different stress levels.

4.1.1 HR and HRV in the aviation field

Heart measures related to cognitive workload are also considered in the aviation industry when speaking about developing a cockpit assistant for SPOs. Thus, several studies ([43] -[45]) highlight how cardiorespiratory data are essential in the development of Cognitive Human-Machine Interface systems able to understand the cognitive workload of a pilot. In particular, Pongsakornsathien et al.[43] discusses recent advances in sensor networks for aerospace cyber-physical systems, focusing on Cognitive Human-Machine Interface systems. Lim et al.[44] provided a detailed review of Human-Machine Interfaces and Interactions on-board civil and military aircraft. Both of them showed how HR grew as the MWL increased. Then they evaluated HRV through time-domain and frequency-domain metrics specifying the relative trends associated with the variation of MWL (such as the mean of inter-beat intervals decreases as MWL increases).

Moreover, concrete experiments on flight simulators or aircraft have been performed in recent years. Wang et al.[46] studied how HR varied during a simulated mission, showing how this parameter significantly differentiated experienced and inexperienced pilots (the HR was higher for inexperienced pilots). The analysis of HR and HRV was also adopted during instrument flight rules proficiency tests by Mansikka et al.[47]. They put in evidence how HR and HRV could identify differences between mission segments (and consequently to MWL) even when there were no significant performance differences between them. Different results, however, were gained by Wei et al.[48] that, evaluating HR and HRV during some flight simulations, obtained that HRV detection was able to effectively reflect the MWL, while HR did not, probably because the physical effort is the primary effect on the heart rate. The effectiveness of HRV in MWL assessment was also studied by Gentili et al.[49], who simulated aircraft piloting tasks under three progressive levels of challenge. He observed that a decrease in HRV (measured through the mean squared successive differences) closely mirrored the increase in task demands. It is possible to find in the literature also in-flight measurement of MWL through a physiological approach. In 2002, Wilson[29] tried to monitor different physiological signals, including HR and HRV, during an approximately ninety-minutes scenario containing both visual and instrument flight conditions. He found that cardiac and electrodermal mea-

sures were highly correlated and exhibited changes in response to the various demands of the flights. Nevertheless, HRV was less sensitive rather than HR, in contrast with the simulation results ([46] -[49]). Another significant in-flight test that studied the feasibility of assessing the pilot's cognitive workload condition with a physiological approach was provided by the United States Air Force (USAF) with the project reported in [50]. Their test consisted of a C-17 refueling mission while measuring ECG and eye tracker data. These data were compared with those obtained on a simulator. In this case, four scenarios of different difficulties were defined and integrated with a secondary auditory response task. Then, they processed the physiological signals through a specific deterministically nonlinear dynamical classifier to evaluate cognitive workload, to extrapolate a workload index based on a ten-level scale similar to the Bedford scale. The evaluation of this workload index during the refueling mission showed consistency with the most-demanding MWL phases of the flight mission (takeoff, refuel, and landing). In particular, their in-flight results showed that the ECG signal was sensitive to low and high workload conditions, corroborated by the subjective report gained at the end of the tests. Also, the workload index obtained from the ECG monitoring at the simulator demonstrated a direct link with the difficulty of the tasks, highlighting the validity of assessing MWL through cardiac monitoring.

Table 3: Heart Rate and Heart Rate Variability most adopted features for MWL evaluation. It is reported the feature's name, a brief description, the Unit of Measure (U. M.), and the qualitative expected trend with respect to an increasing MWL.

Feature	Description	U. M.	Mental Workload
HR	Time interval between two consecutive heart beats	s	↑
SDNN	Standard Deviation of NN intervals	ms	↓
SDRR	Stadard Deviation of RR interval	ms	↓
RMSSD	Root Mean Square of successive RR interval differences	ms	↑
pNN50	Percentage of successive NN intervals that differ more than 50 ms	%	↓
LF	Absolute power of the low-frequency band (0.04-0.15 Hz)	ms^2	↑
HF	Absolute power of the high-frequency band (0.15-0.4 Hz)	ms^2	↓
LF/HF	LF to HF ratio	%	↑
Poicaré axes	Direction of the Poincarè axes	ms	↓

4.2 Eye Activity

It is possible to find lots of research investigating MWL or stress conditions by exploiting eye activity analysis. For instance, Liao et al.[41] could discriminate different stress levels with a heterogeneous approach, also considering the detection of eye movements. Moreover, Gable et al.[38] found that pupil size measurements were able to detect MWL variations with fewer participants than HR evaluations, suggesting that using PS may be a better way to detect real-time changes in workload. By studying the relationship between cognitive workload (low vs. high) and eye movements (saccades, fixations, and smooth pursuit), Belkhiria and Peysakhovich[55] found that the blink rate and saccades amplitude, measured through the EOG, increased along with the cognitive load increase. Moreover, EOG metrics were also considered by Vanneste et al.[56] that, examining whether and how well experienced cognitive load can be measured through psycho-physiological data (EDA, EEG, EOG), highlighted that the eye blink rate could be related to cognitive load. EOG measures were used by Giorgi et al.[42] to assess the correlation between the eye blink rate (EBR) and the MWL. However, they observed that in their tests, EBR resulted in less sensitivity to cognitive workload variations than other parameters such as HR or HRV.

4.2.1 Eye Activity in the aerospace field

Detecting eye movements is also a key aspect of the MWL assessment in the aerospace field. It is possible to find several examples of flight mission simulations that monitor the pilot's eye activity, especially through eye-tracker technologies, that recently became an efficient and reliable tool to assess mental workload[44]. For instance, during their simulated flight tasks, Wang et al.[46] considered fixation durations, saccade rate, and blink rate (together with other cardiorespiratory variables) to evaluate pilots' MWL. Their results showed that saccade and blink rate could indicate pilots' differences in information access strategy, highlighting the different ranges of task demands. Also, Dilli Babu et al.[57] investigated the use of eye gaze trackers in the military aviation environment to automatically estimate the pilot's cognitive load from ocular parameters. According to their research, different flight conditions substantially impact ocular metrics like the fixation rate. For instance, the rate of descent during an air-to-ground dive training exercise, the normal load factor of the aircraft during constant G-level turn manoeuvres, and the pilot's control inceptor and tracking error during simulation activities were all highly connected with it as well.

Despite eye-tracking studies applied to simulations being the most common in the MWL assessment in aerospace, some examples of in-flight missions adopting this kind of technology exist. In particular, Wilson[29], during his aforementioned tests, monitored the eye blink rate of ten general aviation pilots through EOG data, collected from electrodes placed above and below the right eye and lateral to the outer canthus of both eyes. His results showed that blink rates decreased during the more highly visually demanding segments of the flights. Instead, Martin et al.[50], tried to adopt eye-tracking technologies during a refueling mission. Despite some drawbacks, such as calibration problems or glares and shadows due to the sun's relative position with respect to the aircraft, the eye-tracking data was used to understand how the pilot interacts with the cockpit and which are the more focused areas of interest (AoI). This allowed the authors to understand where the pilots focused their gaze during the refueling mission, providing useful information about their attention.

Table 4: Eye Activity most adopted features for MWL evaluation. It is reported the qualitative expected trend with respect to an increasing MWL

Feature	Description	U. M.	Mental Workload
Fixations	Period of time when our gaze remains at a particular location (duration, frequency, count)	ms Hz -	↑
Saccades	Small, rapid movements between fixations, usually from 20 to 200 ms (duration, frequency, count)	ms Hz -	↓
Blink Rate	Full or partial eye closure measure	-	↑
Pupil Diameter	Pupil size measure	mm	↑

4.3 Skin Activity

Especially in recent years, it is possible to find several studies that investigate the relationship between EDA signals and cognitive workload and stress[18]. Setz et al.[61] investigated the discriminating capability of EDA in distinguishing these two cognitive conditions in an office environment. This context was simulated through specific computer tests with a component of psychologically induced stress. The obtained distributions of the EDA peak height and the instantaneous peak rate provide information about an operator's state of stress, according to data obtained with a wearable device. Another study focused on the EDA analysis is provided by Ghaderyan et al.[62] that implemented a time-varying singular value decomposition (TSVD) approach to extract information about MWL from participants of arithmetic tasks. In this case, the results showed an extremely high accuracy rate demonstrating how this algebraic approach is valuable for the following automatic cognitive load estimation task. Moreover, the aforementioned works of Zhang et al.[39], Katsis et al.[40], and Liao et al.[41] also considered EDA to feed their modes to infer physiological states, obtaining significant results in terms of classification accuracy. They highlight how heterogeneous sources can be reliably handled to recognize stress and MWL. Similar results were also obtained by Vanneste et al.[56] that monitored the duration of SCR and other parameters and were able to correlate these features with the results obtained through the subjective report. A different approach was adopted by Giorgi et al.[42] analysing only the EDA's tonic component (SCL) to assess the emotional state of an operator. Finally, Miyake et al.[37], studying the test/retest consistency of physiological responses induced by mental tasks, found that the SCL component significantly correlated with their tests, although highlighting large differences between individuals.

4.3.1 Skin Activity in the aerospace sector

In the aerospace sector, the analysis of the EDA signal is not frequently introduced in MWL assessment research. In fact, the recent improvement in EDA sensor capability makes it possible to find references to monitoring skin activity only in a few works. An example is given by Wilson[29] that, during the aforementioned in-flight tests, took EDA signals and observed that

cardiac and electrodermal measures were highly correlated and showed variations according to the different workload conditions of the flights.

Table 5: Skin activity most adopted features for MWL evaluation. It is reported the qualitative expected trend with respect to an increasing MWL

Feature	Description	U. M.	Mental Workload
SCL Mean	Mean value of SCL	μS	↑
SLC Standard Deviation	Standard Deviation of SCL	μS	-
SCR Mean	Mean value of SCR	μS	↑
SCR Standard Deviation	Standard Deviation of SCR	μS	-
Number of SCR peaks	Number of peaks of the SCR component of the EDA	-	↑
Magnitude of SCR	Magnitude of the SCR manifestations	μS	-
Duration of SCR	Duration of the SCR manifestations	ms	-

4.4 Brain Activity

Several studies investigate the relationship between these signals and cognitive load. Rebsamen et al.[67] extracted EEG during a mental arithmetic task with five different difficulty levels and trained a classifier to discriminate three conditions: relaxed, low, and high. Their results showed an average classification accuracy of 62%, highlighting a significant link between EEG and MWL. Instead, Belkhiria and Peysakhovich[55] investigated the correlation between EEG theta/alpha ratio and EOG signal obtained during specific MWL tests. Their analysis highlighted how an increase in theta/alpha ratio could be predicted by ocular features such as saccades and blinks during high workload phases due to the direct link between EEG and EOG characteristics. Given that, a caveat to consider is that EEG is frequently contaminated with EOG signals and that blinks contain significant delta and theta band activity, leading to the findings that could be likely a spurious correlation. Moreover, the relationship between EEG and EOG was also assessed by the aforementioned work of Vanneste et al.[56], which obtained weak evidence in the association between a lower EEG alpha power, a higher alpha peak frequency, and an increase in cognitive load. The link between ocular movements and brain effects was also investigated by Hajra et al.[69]. In particular, they examined Blink-Related Oscillations (BROs), which represent brainwave responses that follow spontaneous blinking, linked to the environmental monitoring and awareness processing, in an N-Back working memory scenario with simultaneous visual inputs. They obtained that BRO responses were present during the n-back task and that the response amplitudes were dynamically modified by the difference in workload between the 0-back and 3-back situations.

The fNIRS signal was considered by McKendrick et al.[15], exploring three different ways of labeling cognitive states, two algorithms for supervised learning, and three techniques for handling class imbalances. They found that labels that consider individual characteristics are extremely beneficial when used with supervised learners for neurophysiological mental

state prediction. They also observed that when states are defined using individual differences modeling, it may be able to construct classifiers of certain mental states which are person- and task-independent. On the other hand, Svinkunaite et al.[68] analysed fNIRS to assess mental workload from a different perspective. They aimed to understand whether including heart rate and respiration features contained in the fNIRS signal improved the accuracy of the classification level of mental workload. Their results showed how this approach outperformed the highest mean accuracies of previous studies by around 10%. A similar approach was adopted by Herff et al.[70] that measured fNIRS in the PFC during n-back tests. They obtained an accuracy of 78% in single-trial discrimination of three workload levels from each other. The results highlighted how even low workload tasks could be easily discriminated from a resting condition through their specific fNIRS analysis.

4.4.1 Brain activity in the aerospace sector

The monitoring of brain activity through EEG or fNIRS signals is a discussed aspect of the MWL evaluation in the aerospace field[43][44]. In the work mentioned above by Planke et al.[45], they considered the EEG signal, together with eye tracking measurement, to monitor MWL during their simulation of One-to-Many (OTM) Unmanned Aircraft Systems (UASs). In particular, they compared a specific EEG index with subjective data, measures obtained through a behavioral approach (task index), and eye-tracking data. The results showed a good correlation between the physiological data and the task index. EEG-derived ERP was considered by Gentili et al.[49] to select physiological and brain biomarkers of MWL during a simulated aircraft piloting task under three progressive levels of challenge. They obtained that specific parameters extrapolated from ERPs and EEG varied according to the increased workload, highlighting the link between MWL and ERPs. Mohanavelu et al.[71] extracted the spectral features of EEG to assess the dynamic workload of an aircraft's pilot during a simulation of a flight mission, discriminating four different MWL conditions: normal, moderate, high, and very high. They obtained the engagement of specific brain regions during different flying conditions under different cognitive loading. Moreover, the EEG signal was taken into account by Wilson[29] in his in-flight tests, obtaining that alpha and delta bands of the brain activity exhibited significant variations to the varying demands of the scenarios.

4.5 Respiration

It is possible to find some studies investigating the relationship between respiration and MWL, despite their amount being fewer concerning the previously analysed signals. Katsis et al.[40] considered respiration measures in their heterogeneous approach because they considered that the rapidity and depth of breathing could indicate emotional arousal and physical activity. They succeeded in developing a system that could estimate a person's emotional state through a classification vector of features also gained from respiratory measures. Moreover, Miyake et al.[37] monitored the BR and other parameters to examine the consistency of physiological output during specific test and retest sessions. Their results showed that the respiration measures are not the most reliable to assess the MWL of an operator if compared to EDA.

Table 6: Brain Activity most adopted features for MWL evaluation. It is reported the qualitative expected trend with respect to an increasing MWL

Feature	Description	U. M.	Mental Workload
fNIRS deoxygenated hemoglobin (Hb) level	Variation of concentration levels of the deoxygenated hemoglobin related to the baseline	μmol	↓
fNIRS oxygenated hemoglobin (HbO2) level	Variation of concentration level of the oxygenated hemoglobin related to the baseline	μmol	↑
EEG theta band absolute spectral power	Spectral power content of the EEG theta band	μV^2	↑
EEG alpha band absolute spectral power	Spectral power content of the EEG alpha band	μV^2	↓
EEG theta/alpha ratio	Ratio between the theta and alpha spectral power content	-	-
EEG event related potentials (ERPs)	ERPs are the measured brain response to a specific sensory, cognitive or motor event	V	↓

4.5.1 Respiration measures in the aerospace field

Aerospace respiratory parameters for MWL assessment are usually extracted from the pilot with other physiological parameters. This signal is traditionally analysed through multimodal approaches, studying its variation with those of cardiac, ocular, etc. signals[44]. For instance, Pongsakornsathien et al.[43] highlighted a decreasing trend of the BR by the increase in workload and HR, with the disadvantage of the slow response of these signals to the rapid changes in cognitive states. Instead, Wang et al.[46] included respiratory measures in their work to assess different levels of pilot experience. In particular, they observed that the BR and the respiration amplitude (that can be linked to the TV) were lower in experienced pilots thanks to their expertise and practice.

Table 7: Respiration most adopted features for MWL evaluation. It is reported the qualitative expected trend with respect to an increasing MWL

Feature	Description	U. M.	Mental Workload
Breath Rate	Number of breath per minute	min^{-1}	↓
Tidal Volume	Air inspired in one respiratory cycle	mL	↓
Minute Ventilation	Air inspired in one minute	$\mathrm{L\,min}^{-1}$	↓

4.6 Other measures

Despite the most employed measures in MWL evaluation being cardiorespiratory, eye-tracking, skin, and brain signals, in the literature, it is possible to find also other physiological parameters adopted for assessing the cognitive load of an operator during a performance, such as skin temperature and Electromyography (EMG). Therefore, the relationship between skin

temperature and MWL is discussed in the following paragraph, followed by a similar analysis based on the EMG. A paragraph about the application of these two signals in the aviation field concludes the section.

4.6.1 Skin Temperature

The evaluation of the effect of different cognitive loads on skin temperature was considered in a few studies as a non-intrusive measure[74]. The idea is to find a relationship between the temperature trend and the variation of MWL. The most common methods for assessing this biosignal trend are conductive and infrared devices[79]. The conductive devices (such as thermocouples, thermistors, and telemetry sensors) base their functioning on the heat energy transfer into the device directly contacting the subject, while infrared devices (such as infrared thermometers and infrared thermal imaging cameras) are adopted as non-contact technologies.

The aforementioned work of Liao et al.[41] also considered skin temperature data to feed his DBN framework and successfully monitor human stress. Similarly, Romine et al.[74] managed body temperature with EDA and HR to develop a machine learning algorithm and understand how to develop a real-time wearable device that can track MWL. Their results showed how their multimodal methodology was able to consistently distinguish high, medium, and low levels of cognitive load in a learning context.

4.6.2 Electromyography

The EMG signal "measures electrical currents generated in muscles during its contraction representing neuromuscular activities", where the muscles contraction and relaxation activities are controlled by the nervous system[75]. Due to this link between EMG and the nervous system, some research tried to correlate this signal to the variation of MWL in critical situations, extracting this signal from different muscles. For instance, Salomone et al.[76] acquired EMG activity of the flexor pollicis brevis from both hands to analyse cognitive fatigue during a specific computer test. They aimed to assess whether cognitive fatigue increases the capture of the incorrect automatic response or if it impairs its suppression. They correlated EMG measures with subjective fatigue and perceived effort, obtaining that this physiological signal can provide significant results in evaluating cognitive fatigue. The EMG signal was also adopted in the aforementioned work of Katsis et al.[40]. In this study, facial EMG, with respiration, EDA, and ECG parameters, was considered to obtain the classifying vector of features necessary to perform the emotional state evaluation tool. Moreover, Zhang et al.[39] placed EMG sensors on the shoulder (trapezius muscle) to discriminate different stress levels during their heterogeneous approach.

4.6.3 Other measures in the aviation sector

In aviation, it is rarer to find studies linking physiological signals other than those reported above with MWL. Indeed, when talking about high-level systems that allow the monitoring of the pilot's health status and cognitive load, numerous other possible signals are assumed to be taken into account (such as body temperature, voice patterns, EMG, etc.)[43][45], however concrete examples of this kind of application are very infrequent. An example is provided by Wilson[29], that recorded the EMG signal from the calf of the right leg during his aforementioned in-flight tests. He aimed to verify whether the leg movements associated with

controlling the aircraft artifactually influenced the EDA responses from the foot. The results showed that EMG signal patterns did not show the same characteristics of EDA, highlighting how the EMG data are not responsible for the EDA effects.

5.0 Conclusions

In the context of HMI and autonomous systems, this paper aimed to provide an overview of the cognitive workload evaluation based on physiological measures. Due to the risk of confusion on what the term MWL refers to, a clarification on its meaning was introduced, highlighting its multifaceted nature. As the literature shows, there are mainly three ways to assess this mental condition: subjective, behavioral, and physiological. Notably, among these, the physiological approach for MWL evaluation is gaining more and more attention in scientific research, leveraging the disruptive growth of the biomedical sensor market in recent years. Thus, a State-of-the-Art deep analysis on this topic was performed, focusing our attention on the aeronautical context. In particular, the objectives, results obtained, methods implemented, tests performed, and year of publication were reported for each selected article.

The summary of the twenty-nine reviewed articles in Table 2 shows that an operator's cognitive load assessment cannot be obtained merely by observing a single signal. Our work highlights how, despite the scientific community's great effort in recent years to solve this problem, no solution that allows real-time MWL evaluation and monitoring with sufficient reliability to be implemented on the next-generation aircraft has yet been found. The possible explanation emerging from this paper may be that, as presented in Figure 2, multiple internal and external factors influence working conditions stimulating different psycho-physiological areas, simultaneously manifesting on different biosignals. By focusing on Table 2 and Table 8, it is possible to observe that no existing solutions comprehend all the involved physiological signals together (heart, eye, skin, brain, and respiration activities), with the test performed on a large population in an operative environment. Considering the complexity and variety of the potential application fields of HMI systems (such as aeronautics, transport, or controls), it becomes evident that the only successful solution could be using a multimodal approach.

Furthermore, a fruitful approach could merge the different evaluation methods presented in section 3.0. If subjective evaluations are often adopted together with the other assessments as post-performance questionnaires, a synergy between behavioral and physiological approaches still needs to be investigated. The features of these two methods could overcome the unawareness of the former of the operator's physical condition. On the other hand, it fosters a more accurate analysis of the latter, thanks to the knowledge of the task plan.

For what concerned MWL in aeronautics, research has mainly been conducted on simulators, apart from a few rare cases, pointing out the need for designing an efficient autonomous system to foster the development of SPO in civil aviation and MUM teaming in the military field. In conclusion, the availability of smaller, cheaper, and more reliable wearable sensors allows for investigating technologies that could not be realized so far to enhance safety and push the aviation sector to the next generation of aircraft.

ACKNOWLEDGEMENTS

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 sin-

Table 8: State-of-the-Art comparison of the physiological signals for mental workload assessment. This table shows the adopted methods, the implemented tests, the involved sample, and the year of publication for each paper.

Ref	Methods	Tests	Sample	Year
(15)	Labeling Methods: Difficulty split Rasch model Stress-strain curve Classification algorithms: Elastic Net Random Forest	Spatial memory tasks	34	2021
(29)	Statistical analysis (ANOVA test)	Live in-flight	10	2002
(33)	Morris and DeShon's procedure	-	-	2019
(37)	Test/retest	Multi-attribute task battery trial	15	2009
(38)	Statistical analysis (ANOVA test)	Driving test with N-Back test	8	2015
(39)	Support Vector Machine (SVM)	Stroop test and acoustic induction	9	2016
(40)	Support Vector Machine (SVM)	-	-	2006
(41)	Dynamic Bayesian Network (DBN)	Math task Audio task	5	2005
(42)	Statistical analysis (Wilcoxon signed-rank and ANOVA test)	N-Back test Doctor game test Webcall task	17	2021
(43)	-	-	-	2019
(44)	-	-	-	2018
(45)	Statistical analysis (ANOVA test and Correlation)	One-to-Many UAS scenario	5	2020
(46)	Statistical analysis (ANOVA test)	Simulated flight tasks	24	2016
(47)	Statistical analysis (ANOVA test)	Simulated flight tasks	26	2016
(48)	Bayesian Fisher discrimination and classification	Simulated flight tasks	16	2014
(49)	Statistical analysis (ANOVA test)	Simulated flight tasks	38	2014
(50)	-	Live flight Simulated flight	7	2019
(55)	Statistical analysis (ANOVA test)	Auditory noise Arithmetical count Working memory load	10	2021
(56)	Statistical analysis (pairwise comparison)	Woking memory test	46	2020
(57)	Statistical analysis (ANOVA test)	Simulated flight Live flight	14	2019
(61)	Linear Discriminating Analysis (LDA) Support Vector Machine (SVM) Nearest Class Center (NCC) algorithm	Arithmetic tasks	33	2010
(62)	Time Varying Singular Value Decomposition	Arithmetic tasks	35	2018
(67)	Quadratic discriminant analysis classifier (PyMVPA)	Mental arithmetic task	16	2011
(68)	Random Forest Classification	N-Back test	14	2021
(69)	Statistical Analysis (ANOVA)	N-Back test	26	2019
(70)	Linear Discriminant Analysis (LDA) classifier	N-Back test	10	2014
(71)	Statistical analysis (Friedman and Wilcoxon signed-rank test)	Simulated flight tasks	16	2020
(74)	Constant Logistic Regression Naïve Bayes K Nearest Neighbors Decision Tree Support Vector Machine Neural Networks AdaBoost Random Forest	Playing Sudoku Browsing the Internet Sitting and doing nothing	7	2020
(76)	Statistical analysis (ANOVA test)	Time Load Dual Back Task Simon Task Dual Psychomotor Vigilance Task	24	2021

cerely thank both for their insights and for co-financing the PhD program from which this paper was born.

References

1. YOUNG, M. S., BROOKHUIS, C., WICKENS, C. D. and HANCOCK, P. A. State of science: mental workload in ergonomics, *Ergonomics*, 2014, **58**, (1), pp 1-17.
2. DEBIE, E., ROJAS, R. F., FIDOCK, J., BARLOW, M., KASMARIK, K., ANAVATTI, S. and GARRATT, M. Multimodal Fusion for Objective Assessment of Cognitive Workload, *IEEE TRANSACTION ON CYBERNETICS*, 2021, **51**, (3), pp 1542-1555.
3. DE FILIPPIS, L., GAIA E., GUGLIERI, G., RE, M. and RICCO, C. Filippis, Luca de, Enrico Gaia, Giorgio Guglieri, Marco Re and Claudia Ricco. Cognitive Based Design of a Human Machine Interface for Telenavigation of a Space Rover, *Journal of Aerospace Technology and Management*, 2014, **6**, pp 415-430.
4. DE FILIPPIS, L., GUGLIERI, G., RICCO, C. and SARTORI, D. Remote Control Station Design and Testing for Tele-Operated Space-Missions, *International Journal of Aerospace Sciences*, 2013, **2**, (3), pp 92-105.
5. ICAO, 2021 Safety Report, 2021, International Civil Aviation Organization.
6. SHAPPELL, S., DETWILER, C., HOLCOMB, K., HACKWORTH, C., BOQUET, A. and WIEGMANN, D. Human Error and Commercial Aviation Accidents: A Comprehensive, Fine-Grained Analysis Using HFACS, 2006, Federal Aviation Administration.
7. MARTINS, A. P. G. A Review of Important Cognitive Concepts in Aviation, *Aviation*, 2016, **20**, (2), pp 65-84.
8. MATESSA, M., STRYBEL, T., VU, K., BATTISTE, V. and SCHNELL, T. Concept of Operations for RCO SPO, 2017, NASA.
9. HONECKER, F. and SCHULTE, A. Automated Online Determination of Pilot Activity Under Uncertainty by Using Evidential Reasoning, *Engineering Psychology and Cognitive Ergonomics: Cognition and Design*, 2017, **10276**, pp 231-250.
10. AEROSPACE TECHNOLOGY INSTITUTE The single Pilot commercial Aircraft, 2019, INSIGHT
11. HANCOCK, G. M., LONGO, L., YOUNG, M. S. and HANCOCK, P. A. Mental Workload, *Handbook of Human Factors and Ergonomics*, Fifth Edition, 2021, **7**, pp. 203-226, John Wiley & Sons, Inc.
12. VAN ACKER, B., PARMENTIER. D., VLERICK, P. and SALDIEN, J. Understanding mental workload: from a clarifying concept analysis toward an implementable framework, *Cognition, Technology & Work*, 2018, **20**, pp. 351-365.
13. TAO, D., TAN, H., WANG, H., ZHANG, X., QU, X. and ZHANG, T. A Systematic Review of Physiological Measures of Mental Workload. *Int J Environ Res Public Health*, 2019, **16**, (15)
14. WICKENS, C. D. Multiple resources and performance prediction, *Theoretical Issues in Ergonomics Science*, 2002, **3**, (2), pp 159-177.
15. MCKENDRICK, R., FEEST, B., HARWOOD, A. and FALCONE, B. Theories and Methods for Labeling Cognitive Workload: Classification and Transfer Learning, *Frontiers in Human Neuroscience*, 2019, **13**, (259), pp 1-20.

16. Fink, G. Stress: Definition and History, *Stress Science: Neuroendocrinology*, 2010, pp 3-9.

17. Hart, S. G. and Staveland, L. E. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research, *Human Mental Workload. Advances in Psychology*, 1988, **52**, pp 139-183.

18. Ayres, P., Lee, J.Y., Paas, F. and van Merriënboer, J. The Validity of Physiological Measures to Identify Differences in Intrinsic Cognitive Load, *frontiers in Psychology*, 2021, **12**, 1-16.

19. McKendrick, R. D. and Cherry, E. A Deeper Look at the NASA TLX and Where It Falls Short, *Proceedings of the Human Factors and Ergonomics Society 2018 Annual Meeting*, 2018, **62**, (1), pp 44-48.

20. Cooper, G. H. and Harper, R. P. The Use of Pilot Rating in the Evaluation of Aircraft Handling Qualities, 1969, NASA.

21. Casali, J. and Wierwillie, W. A Comparison of Rating Scale, Secondary-Task, Physiological, and primary-Task Workload Estimation techniques in a simulated Flight Task Emphasizing Communications Load, *Human Factors*, 1983, **25**, (6), pp 623-641.

22. Brand, Y. and Schulte, A. Workload-adaptive and task-specific support for cockpit crews: design and evaluation of an adaptive associate system, *Human-Intelligent System Integration*, 2021, **3**, pp 187-199.

23. Brand, Y. and Schulte, A. Model-based Prediction of Workload for Adaptive Associate Systems, *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2017.

24. Wickens, C. D. Multiple resources and performance prediction, *Theoretical Issues in Ergonomics Science*, 2002, **3**, (2), pp 159-177.

25. Wickens, C. D. Multiple Resources and Mental Workload, *human factors*, 2008, **50**, (3), pp 449-455.

26. Horrey, W. and Wickens, C. D., Multiple resource Modeling of Task Interference in Vehicle Control, Hazard Awareness and In-Vehicle Task Performance, *Driving Assessment Conference 2003*, 2003.

27. Chen, J., Zhang, Q. and Hou, B. An assessment method of pilot workload in manned/unmanned-aerial-vehicles team, *2017 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)*, October 2017, pp 1-5.

28. Moray, N. Mental Workload: Its Theory and Measurement, *Proceedings of the NATO Symposium on Theory and Measurement of Mental Workload*, 1979.

29. Wilson, G. An Analysis of Mental Workload in Pilots During Flight Using Multiple Psychophysiological Measures, *The International Journal of Aviation Psychology*, 2002, **12**, (1), pp 3-18.

30. Wan, H., Zhuang, L., Pan, Y., Gao, F., Tu, J., Zhang, B. and Wang, P. Biomedical sensors, *Biomedical Information Technology*, 2020, pp 51-79.

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

32. Kramer, A. F. Physiological metrics of mental workload: A review of recent progress, *Multiple-task performance*, 1991, Taylor & Francis, pp 279-328.

33. HUGHES, A. M., HANCOCK, G. M., MARLOW, S. L., STOWERS, K. and SALAS, E. Cardiac Measures of Cognitive Workload: A Meta-Analysis, *Human Factors*, 2019, **61**, (3), pp 393-414.

34. PARAK, J. and HAVLIK, J. ECG Signal Processing and Heart Rate Frequency Detection Methods, *Technical Computing 2011*, 2011.

35. PARK, J., SEOK SEOK, H., KIM, S. ans SHIN, H. Photoplethysmogram Analysis and Applications: An Integrative Review, *frontiers in Physiology*, 2022, **12**, pp 1-23.

36. Yoo, K. S. and LEE, W. H. Mental Stress Assessment Based on photoplethysmography, *2011 IEEE 15th International Symposium on Consumer Electronics*, 2011.

37. MIYAKE, S., YAMADA, S., SHOJI, T., TAKAE, Y., KUGE, N. and YAMAMURA, T. Physiological responses to workload change. A test/retest examination, *Applied Ergonomics*, 2009, **40**, (6), 987-996.

38. GABLE, T. M., KUN, A. L., WALKER, B. N. and WINTON, R. J. Comparing heart rate and pupil size as objective measures of workload in the driving context: initial look, *Adjunct Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '15)*, 2015.

39. ZHANG, B., MORERE, Y., SIELER, L. and LENGLET, C. Stress Recognition from Heterogeneous Data, *Journal of Image and Graphics*, 2016, **4**, (2),pp 116-121.

40. KATSIS, C. D., GANIATAS, G. and FOTIADIS, D. I. An integrated telemedicine platform for the assessment of affective physiological states, *Diagnostic Pathology*, 2006, **1**, (16).

41. LIAO, W., ZHANG, W., ZHU, Z. and JI, Q. A Real-Time Human Stress Monitoring System Using Dynamic Bayesian Network, *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Workshops*, 2005.

42. GIORGI, A., RONCA, V., VOZZI, A., SCIARAFFA, N., DI FLORIO, A., TAMBORRA, L., SIMONETTI, I., ARICÒ, P., DI FLUMERI, G., ROSSI, D. and BORGHINI, G.Wearable Technologies for Mental Workload, Stress, and Emotional State Assessment during Working-Like Tasks: A Comparison with Laboratory Technologies, *Sensors*, 2021, **21**, (7).

43. PONGSAKORNSATHIEN, N., LIM, Y., GARDI, A., HILTON, S. PLANKE, L., SABATINI, R., KISTAN, T. and EZER, N. Sensor Networks for Aerospace Human-Machine Systems, *Sensors*, 2019, **19**, (16).

44. LIM, Y., GARDI, A., SABATINI, R., RAMASAMY, S., KISTAN, T., EZER, N., VINCE, J. and BOLIA, R. Avionics Human-Machine Interfaces and Interactions for Manned and Unmanned Aircraft, *Progress in Aerospace Sciences*, 2018, **102**, pp 1-46.

45. PLANKE, L. J., LIM, Y., GARDI, A., SABATINI, R., TREVOR, K. and EZER, N. A Cyber-Physical-Human System for One-to-Many UAS Operations: Cognitive Load Analysis, *Sensors*, 2020, **20**, (19).

46. WANG, Z., ZHENG, L., LU, Y. and FU, S. Physiological Indices of Pilots ' Abilities Under Varying Task Demands, AEROSP MED HUM PERFORM, 2016, **87**, (4), pp 375-381.

47. MANSIKKA, H., VIRTANEN, K., HARRIS, D. and SIMOLA, P. Fighter pilots' heart rate, heart rate variation and performance during an instrument flight rules proficiency test, *Applied Ergonomics*, 2016, **56**, pp 213-219.

48. WEI, Z., ZHUANG, D., WANYAN, X., LIU, C. and ZHUANG, H. A model for discrimination and prediction of mental workload of aircraft cockpit display interface, *Chinese Journal*

of Aeronautics, 2014, **27**, (5), pp 1070-1077.

49. Gentili, R. J., Rietschel, J. C., Jaquess, K. J., Lo, L. C., Prevost, M., Miller, M. W., Mohler, J. M., Oh, H., Tan, Y. Y. and Hatfield, B. D. Brain biomarkers based assessment of cognitive workload in pilots under various task demands, *Annu Int Conf IEEE Eng Med Biol Soc 2014*, 2014, pp 5860-5863.
50. Martin, p., Calhoun, P., Schnell, T. and Thompson, C. Objective Measures of Pilot Workload, 2019.
51. Lai, M. L., Tsai, M. L., Yang, F. Y., Hsu, C. Y., Liu, T. C., Lee, S. W. Y., Lee, M. H., Chiou, G. L., Liang, J. C. and Tsai, C. C. A review of using eye-tracking technology in exploring learning from 2000 to 2012, *Educational Research Review*, 2013, **10**, pp 90-115.
52. Mahanama, B., Jayawardana, Y., Rengarajan, S., Jayawardena, G., Chukoskie, L., Snider , J. abd Jayarathna, S. Eye Movement and Pupil Measures. A Review, *frontiers in Computer Science*, 2022, **3**.
53. Salvucci, D. D. and Goldberg, J. H. Identifying Fixations and Saccades in Eye-Tracking Protocols, *Proceedings of the 2000 symposium on Eye tracking research & applications (ETRA '00)*, 2000.
54. Barea, R., Boquete, S., Ortega, S., Lopez, E. and Rodríguez-Ascariz, J. M. EOG-based eye movements codification for human computer interaction, *Expert Systems with Applications*, 2012, **39**, (3), pp 2677-2683.
55. Belkhiria, C. and Peysakhovich, V. EOG metrics for cognitive workload detection, *25th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems*, 2021, **192**, pp 1875-1884.
56. Vanneste, P., Raes, A., Morton, J., Bombeke, K., Van Acker, B. B., Larmuseau, C., Depaepe, F. and Van den Noortgate, W. Towards measuring cognitive load through multimodal physiological data, *Cognition, Technology & Work*, 2021, **23**, pp 567-585.
57. Dilli Babu, M., Jeevitha, S., Gowdham, P., Kamalpreet, S., Abhay, P. and Pradipta, B. Estimating Pilots' Cognitive Load From Ocular Parameters Through Simulation and In-Flight Studies, *Journal of Eye Movement Research*, 2019, **12**, (3).
58. Posada-Quintero, H. F. and Chon, K. H. Innovations in Electrodermal Activity Data Collection and Signal Processing: A Systematic Review, *Sensors*, 2020, **20**, (2).
59. Cacioppo, J., Tassinary, L. and Berntson, G. Handbook of psychophysiology, 2007.
60. Greco, A., Valenza, G., Lanata, A., Scilingo, E. P. and Citi, L. cvxEDA: A Convex Optimization Approach to Electrodermal Activity Processing, *IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING*, 2016, **63**, (4), pp 797-804.
61. Setz, C., Arnrich, B., Schumm, J., Marca, R., Tröster, G. and Ehlert, U. Discriminating stress from cognitive load using wearable EDA device, *IEEE Transactions on Information Technology in Biomedicine*, 2010, **14**, pp 410-417.
62. Ghaderyan, P., Abbasi, A. and Ebrahimi, A. Time-varying singular value decomposition analysis of electrodermal activity: A novel method of cognitive load estimation, *Measurement*, 2018, **126**, pp 102-109.
63. Ghani, U., Signal, N., Niazi, I. K. and Taylor, D. ERP based measures of cognitive workload: A review, *Neuroscience and Biobehavioral Reviews*, 2020, **118**, pp 18-26.

64. KUMAR, S. J. and BHUVANESWARI, P. Analysis of Electroencephalography (EEG) Signals and Its Categorization - A Study, *Procedia Engineering*, 2012, **38**, pp 2525-2536.

65. HANDY, T. C. Event-related potentials: A methods handbook, 2005, MIT Press.

66. WELCH, T. L. and PASTERNAK, J. J. Chapter 56 - Recent Advances in Neuroanesthesiology, *Essentials of Neuroanesthesia*, 2017, Academic Press, pp 897-905.

67. REBSAMEN, B., KWOK, K. and PENNEY, T. B. Evaluation Of Cognitive Workload From EEG Durting A Mental Arithmetic Task, *PROCEEDINGS of the HUMAN FACTORS and ERGONOMICS SOCIETY 55th ANNUAL MEETING*, 2011, **55**, (1), pp 1342-1345.

68. SVINKUNAITE, L., HORSCHIG, J. M. and FLOOR-WESTERDIJK, M. J. Employing cardiac and respiratory features extracted from fNIRS signals for mental workload classification, *SPIE BiOS*, 2021.

69. HAJRA, S. G., LIU, C., LAW, A. Neural responses to spontaneous blinking capture differences in working memory load: Assessing blink related oscillations with N-back task, *International Neuroergonomics Conference*, 2019.

70. HERFF, C., HEGER, D., FORTMANN, O., HENNRICH, J., PUTZE, F. and SCHULTZ, T. Mental workload during n-back task—quantified in the prefrontal cortex using fNIRS, *Frontiers in Human Neuroscience*, 2014, **7**.

71. MOHANAVELU, K., POONGUZHALI, S., ADALARASU, K., RAVI, D., VIJAYAKUMAR, C., VINUTHA, S., RAMACHANDRAN, K. and SRINIVASAN, J. Dynamic cognitive workload assessment for fighter pilots in simulated fighter aircraft environment using EEG, *Biomedical Signal Processing and Control*, 2020, **61**.

72. GRASSMANN, M., VLEMINCX, E., VON LEUPOLDT, A., MITTELSTÄDT, J. M. and VAN DEN BERGH, O. Respiratory Changes in Response to Cognitive Load: A Systematic Revie, *Neural Plast*, 2016.

73. MONACO, V. and STEFANINI, C. Assessing the Tidal Volume through Wearables: A Scoping Review, *Sensors*, 2021, **21**, (12).

74. ROMINE, W., SCROEDER, N., GRAFT, J., YANG, F., SADEGHI, R., ZABIHIMAYVAN, M., KADARIYA, D. and BANERJEE, T. Using Machine Learning to Train a Wearable Device for Measuring Students' Cognitive Load during Problem-Solving Activities Based on Electrodermal Activity, Body Temperature, and Heart Rate: Development of a Cognitive Load Tracker for Both Personal and Classroom Use, *Sensors*, 2020, **20**, (4883).

75. RAEZ, M. B., HUSSAIN, M. S. and MOHD-YASIN, F. Techniques of EMG signal analysis: detection, processing, classification and applications, *Biol Proced Online*, 2006, **8**, (163).

76. SALOMONE, M., BURLE, B., FABRE, L. and BERBERIAN, B. An Electromyographic Analysis of the Effects of Cognitive Fatigue on Online and Anticipatory Action Control, *Frontiers in Human Neuroscience*, 2021, **14**.

77. MASSARONI, C., NICOLÒ, A., LO PRESTI, D., SACCHETTI, M., SILVESTRI, S. and SCHENA, E. Contact-Based Methods for Measuring Respiratory Rate, *Sensors*, 2019, **19**, (4).

78. AL-KHALIDI, F., SAATCHI, R., BURKE, D. and ELPHICK, H. Respiration Rate Monitoring Methods: A Review, *Pediatric Pulonology*, 2011, **46**, pp 523-529.

79. BACH, A. J. E., STEWART, I. B., MINNETT, G. M. and COSTELLO, J. T. Does the technique employed for skin temperature assessment alter outcomes? A systematic review, *Physi-*

ological Measurement, 2015, **36**, (9), pp 27-51