CHAPTER

A passive brain–computer interface application for the mental workload assessment on professional air traffic controllers during realistic air traffic control tasks

10

P. Aricò[*,†,‡,2], G. Borghini[*,†,‡,2], G. Di Flumeri[*,†,‡,2], A. Colosimo[*], S. Pozzi[§], F. Babiloni[*,†,1]

[*]*University of Rome "Sapienza", Rome, Italy*

[†]*BrainSigns srl, Rome, Italy*

[‡]*Neuroelectrical Imaging and BCI Lab, IRCCS Fondazione Santa Lucia, Rome, Italy*

[§]*DeepBlue srl, Rome, Italy*

[1]*Corresponding author: Tel.: +39-3287697914, e-mail address: fabio.babiloni@uniroma1.it*

Abstract

In the last decades, it has been a fast-growing concept in the neuroscience field. The *passive brain–computer interface* (p-BCI) systems allow to improve the *human–machine interaction* (HMI) in operational environments, by using the covert brain activity (eg, mental workload) of the operator. However, p-BCI technology could suffer from some practical issues when used outside the laboratories. In particular, one of the most important limitations is the necessity to recalibrate the p-BCI system each time before its use, to avoid a significant reduction of its reliability in the detection of the considered mental states. The objective of the proposed study was to provide an example of p-BCIs used to evaluate the users' mental workload in a real operational environment. For this purpose, through the facilities provided by the *École Nationale de l'Aviation Civile* of Toulouse (France), the cerebral activity of 12 professional *air traffic control officers* (ATCOs) has been recorded while performing high realistic air traffic management scenarios. By the analysis of the ATCOs' brain activity (e*lectroencephalographic* signal—EEG) and the subjective workload perception (*instantaneous self-assessment*) provided by both the examined ATCOs and external air traffic control experts,

[2]These authors contributed equally to this work.

Progress in Brain Research, Volume 228, ISSN 0079-6123, http://dx.doi.org/10.1016/bs.pbr.2016.04.021
© 2016 Elsevier B.V. All rights reserved.

it has been possible to estimate and evaluate the variation of the mental workload under which the controllers were operating. The results showed (i) a high significant correlation between the neurophysiological and the subjective workload assessment, and (ii) a high reliability over time (up to a month) of the proposed algorithm that was also able to maintain high discrimination accuracies by using a low number of EEG electrodes (∼3 EEG channels). In conclusion, the proposed methodology demonstrated the suitability of p-BCI systems in operational environments and the advantages of the neurophysiological measures with respect to the subjective ones.

Keywords

Passive brain–computer interface, Augmented cognition, Air traffic management, Electroencephalogram, Mental workload, Automatic-stop stepwise linear discriminant analysis, Stepwise linear discriminant analysis, Instantaneous self-assessment, Human factor, Neuroergonomic

1 INTRODUCTION

In some operational environments, the safety of the people relies on the attentional and cognitive efforts of the operator(s). In such challenging contexts, human errors could have serious and dramatic consequences. For example, in the transportations domain the safety of the passengers depends on the performance of the pilot(s), of the (eg, air, train, vessel) traffic controller(s), or of the driver(s) of private vehicles.

In general, human error has consistently been identified as one of the main factors in workplaces accidents. In particular, it has been estimated that up to 90% of accidents list human errors as the principal cause (Feyer and Williamson, 1998).

Human error is an extremely common phenomenon: people, regardless of abilities, skills, and expertise, make errors daily. It can be defined as the execution of an incorrect or inappropriate action, or a failure to perform a particular action. The main causes of human errors have to be sought within the internal or psychological factors of the operator (Reason, 2000). In fact, errors could arise from aberrant mental processes, such as inattention, poor motivation, loss of vigilance, mental overload, and fatigue, that negatively affect the user's performance. For example, cognitive psychology literature demonstrated that the mental workload has an "inverted U-shape" relationship with performance. In other words, some levels of mental workload may help the user to reach high-performance level (Calabrese, 2008), since it stimulates positively the user and it keeps him/her awake with high attention level. On the contrary, a period of mental inactivity and "understimulation" can cause a monotonous and boring state (underload), a low level of vigilance and attention, with low cognitive resources demand. Additionally, an operative condition characterized by demanding multitasks can lead the user to an overload condition and to a likely occurrence of errors (Kirsh, 2000). It is interesting to note that all the mentioned causes produced a reduction of the operator's performance and the concomitant change of the spectral properties of his/her cerebral signals.

In this regard, the *augmented cognition* research field aims at developing systems to avoid performance degradation by adapting the user's interface and reducing the task demand/complexity, or by intervening directly on the system (Fuchs et al., 2007). Over the past two decades, researchers in the field of augmented cognition developed novel technologies to both monitor and enhance human cognition and performance. Most of those works were based on research findings coming from cognitive science and cognitive neuroscience (Decades, 2008). On the basis of such findings and technological improvements in measuring human biosignals, it has been possible to evaluate operators' mental states unobtrusively and in realistic contexts. The neurophysiological indexes have then been used as inputs for the interface the operator was interacting with. Such application is usually named *passive brain–computer interface* (p-BCI).

1.1 PASSIVE BRAIN–COMPUTER INTERFACE

In its classical assumption, a BCI is "a communication system in which messages or commands do not pass through the brain normal output pathways of peripheral nerves and muscles" (Wolpaw et al., 2002). More recently, Wolpaw and Wolpaw (2012) defined a BCI as "a system that measures central nervous system (CNS) activity and converts it into artificial output that replaces, restores, enhances, supplements, or improves natural CNS output and thereby changes the ongoing interactions between the CNS and its external or internal environment."

In the BCI community, the possibility of using the BCI systems in different contexts for communication and system control (Allison and Pineda, 2003; Aloise et al., 2010, 2013; Blankertz et al., 2010; Riccio et al., 2015; Schettini et al., 2015), developing applications in realistic and operational environments, is not just a theory but something very close to real applications (Blankertz et al., 2010; Müller et al., 2008; Zander et al., 2009). In fact, in the classic BCI applications the user modulates voluntarily its brain activity to interact with the system. In the p-BCI implementation, the system itself recognizes the spontaneous brain activity of the user related to the considered mental state (eg, emotional state, workload, attention levels) and uses such information to improve and modulate the interaction between the operator and the system.

Systems based on passive BCI technology can provide objective information about covert aspects of the user's cognitive state, since conventional methods, such as behavioral measures, could only detect such mental states with weak reliability (Zander and Jatzev, 2012). The information extracted by the p-BCIs is then employed to improve man–machine interactions and to achieve potentially novel types of skills. Anyhow, the quantification of mental states by using this technology is far from trivial. In fact, it requires a combination of knowledge in different fields (Brouwer et al., 2015), such as neurophysiology (to acquire and manage biosignals), experimental psychology (to find out the right way to assess mental states), machine learning (to develop innovative classification techniques), and human factor (to develop real applications).

Neuroimaging methods and cognitive neuroscience have steadily improved their technical sophistication and breadth of application over the past decade, and there has been growing interest in their use to examine the neural circuits supporting complex tasks representative of perception, cognition, and action as they occur in operational settings. At the same time, many fields in the biological sciences, including neuroscience, are being challenged to demonstrate their relevance to practical real-world problems (Parasuraman, 2003).

In this context, the most studied mental state has been the mental workload due to its strong relationship with the increasing or degrading of user's performance. In fact, mental workload is a complex construct that is assumed to be reflective of an individual's level of attentional engagement and mental effort (Wickens, 1984). Measurement of mental workload essentially represents the quantification of mental activity resulting from performance of a task or set of tasks. As mentioned previously, several empirical investigations have suggested that performance declines at either extreme of the workload demand profile (ie, when event rates are excessively high or extremely low). Consequently, it is important to preserve a proper level of the user's mental workload, avoiding under- or overload state, with the aim to maintain optimal levels of performance and reducing the risk of errors (Borghini et al., 2014a,b, 2015a). For these reasons, the mental workload is an important and central construct in ergonomics and human factor researches.

1.2 MENTAL WORKLOAD: THE MEAN AND ITS NEUROPHYSIOLOGICAL MEASUREMENTS

Various mental workload definitions have been given during the last decades:

- "Mental workload refers to the portion of operator information processing capacity or resources that is actually required to meet system demands" (Eggemeier et al., 1991).
- "Workload is not an inherent property, but rather it emerges from the interaction between the requirements of a task, the circumstances under which it is performed, and the skills, behaviors, and perceptions of the operator" (Hart and Staveland, 1988).
- "Mental workload is a hypothetical construct that describes the extent to which the cognitive resources required to perform a task have been actively engaged by the operator" (Gopher and Donchin, 1986).
- The reason to specify and evaluate the mental workload is to quantify the mental cost involved during task performance "in order to predict operator and system performance" (Cain, 2007).

These definitions show that mental workload may not be a unitary concept because it is the result of different interacting aspects. In fact, several mental processes, such as alertness, vigilance, mental effort, attention, mental fatigue, drowsiness, and so on, can be involved in the meanwhile of a task execution, and they could be affected in each moment by specific task demands.

In general, mental workload theory assumes that: (a) people have a limited cognitive and attentional capacity, (b) different tasks will require different amounts of processing resources, and (c) two individuals might be able to perform a given task equally well, but differently in terms of brain activation (Baldwin, 2003).

Mental workload assessment techniques must be sensitive to cognitive fluctuations in task demands without intruding on primary task performance (O'Donnell and Eggemeier, 1986). In this regard, measuring the mental workload by using subjective measures during the execution of the main task could negatively affect the user's performances. Additionally, it has been widely demonstrated that neurophysiological measurements transcend both behavioral and subjective measures in discriminating cognitive demand fluctuations (Di Flumeri et al., 2015; Mühl et al., 2014; Wierwille and Eggemeier, 1993).

Thus, the online neurophysiological measurements of the mental workload could become very important not only as monitoring techniques but mainly as support tools to the user during operative activities. In fact, as the changes in cognitive activity can be measured in real time, it should also be possible to manipulate the task demand in order to help the user in maintaining optimal levels of mental workload during the work. In other words, the neurophysiological workload assessment could be used to realize p-BCI applications in real operational environments.

Many neurophysiological measures have been used for the mental workload assessment, including *electroencephalography* (EEG), *functional near-infrared* (fNIR) imaging, *functional magnetic resonance imaging* (fMRI), and *magnetoencephalography* (MEG), and other types of biosignals such as *electrocardiography* (ECG), *electrooculography* (EOG), and *galvanic skin response* (GSR) (Borghini et al., 2014b; Ramnani and Owen, 2004; Toppi et al., 2016; Wood and Grafman, 2003). The size, weight, and power constraints outlined above limit the types of neurofeedback that can be used to realize p-BCI applications. For example, fMRI (Cabeza and Nyberg, 2000) and MEG techniques require room-size equipment; thus they would not be portable. EOG, ECG, and GSR activities highlighted correlations with some mental states (stress, mental fatigue, drowsiness), but they were demonstrated to be useful only in combination with other neuroimaging techniques directly linked to the CNS, ie, the brain (Borghini et al., 2014b, 2015b; Ryu and Myung, 2005). Consequently, the EEG and fNIR are the most likely candidates that can be straightforwardly employed to realize passive BCI applications usable in operational environments.

Regarding the EEG measurements, most part of the studies showed that the brain electrical activities mainly considered for the mental workload analysis are the theta and alpha brain rhythms typically gathered from the *prefrontal cortex* (PFC) and the *posterior parietal cortex* (PPC) regions. Previous studies demonstrated as the EEG theta rhythm over the PFC present a positive correlation with the mental workload (Gevins and Smith, 2003; Smit et al., 2005). Moreover, published literature stressed the inverse correlation between the EEG power in the alpha frequency band over the

PPC and the mental workload (Brookings et al., 1996; Gevins et al., 1997; Jaušovec and Jaušovec, 2012; Klimesch et al., 1997; Venables and Fairclough, 2009). Only few studies have reported significant results about the modulation of the EEG power in other frequency bands, ie, the delta, beta, and gamma (Borghini et al., 2014b; Gevins et al., 1997; Smith and Gevins, 2005). More specifically, most of the studies are focalized on the EEG power modulation occurring in theta (4–8 Hz) and alpha (8–12 Hz) frequency bands, usually associated with cognitive processes such as working memory and attention, typically involved in mental workload. Onton et al. (2005) reported that the frontal midline theta rhythm increases with memory load, confirming previous results about the correlation between the frontal theta EEG activity and mental effort (Gevins et al., 1997; Smit et al., 2005). Mental workload is also known to suppress EEG alpha rhythm and to increase theta rhythm during activity of information encoding and retrieval (Klimesch, 1999; Klimesch et al., 1997; Vecchiato et al., 2014).

According to the idea that the higher the mental workload level is, the greater the brain blood oxygenation will be, the functional near-infrared spectroscopy (fNIRs) has been demonstrated to be another reliable mental workload measurement technique (Cui et al., 2011; Derosière et al, 2013). fNIRs is safe, highly portable, user-friendly, and relatively inexpensive, with rapid application times and near-zero run-time costs, so it could be a potential portable system for measuring mental workload in realistic settings. The most common fNIR system uses infrared light introduced in the scalp to measure changes in blood oxygenation. Oxyhemoglobin (HbO_2) converts into deoxyhemoglobin (HbR) during neural activity, that is, the cerebral hemodynamic response. This phenomenon is called *blood-oxygen-level-dependent* (BOLD) signal. fNIRs has been shown to compare favorably with other functional imaging methods (Huppert et al., 2006) and demonstrates solid test–retest reliability for task-specific brain activation (Herff et al., 2013; Plichta et al., 2006). Thus, the primary hypothesis was that blood oxygenation in the PFC, as assessed by fNIR, would rise with increasing task load and would demonstrate a positive correlation with the mental workload. In fact, Izzetoglu et al. (2004) indicated clearly that the rate of changes in blood oxygenation was significantly sensitive to task load variations.

1.3 AN EXAMPLE OF MENTAL WORKLOAD MEASURE IN REALISTIC SETTINGS: THE AIR TRAFFIC MANAGEMENT CASE

In the last 20 years, it has been widely documented that 70% of civil aviation accidents were linked with human errors (Bellenkes, 2007). Recently, the Aviation Safety Network reported 37 accidents with 564 casualties. Moreover, air traffic is growing exponentially, and it has been predicted to double in 2020 (Flight Safety Foundation). It is easy to understand how this factor would increase the work difficulty of *air traffic control officers* (ATCOs). In fact, they have to perform a variety of tasks, including monitoring air traffic, anticipating loss of separation between aircraft, and intervening to resolve conflicts and minimize disruption to air traffic flow (for an extensive compilation of the tasks and goals of en route control, see Rodgers and Drechsler, 1993). In this domain, the ATCO's behavior could already be measured through several human factor tools, such as the explicit measurement of errors committed during the execution of the task, or by using questionnaires related to the subjective workload perception such as the *instantaneous self-assessment* (ISA, Kirwan et al., 2001), *NASA—task load index* (Hart and Staveland, 1988), or the *subjective workload assessment technique* (Reid and Nygren, 1988). Because of their inherently subjective nature, none of such questionnaires allows to have an objective and reliable measure of the actual cognitive demand for the operator in a real environment. All the described questionnaires have pros and cons, but there is not a generally accepted standard (Rubio et al., 2004). Therefore, the need of an objective measure became fundamental for reliable workload evaluations.

Several researches in the air traffic management (ATM) domain treated the neurophysiological measurements of ATCOs' mental workload in realistic settings with the aim of developing *human–machine interaction* (HMI) systems, by using both EEG and fNIRs techniques. In the following examples, it has been discussed how each technique was able to provide reliable estimations of mental workload. The propensity in using EEG or fNIRs techniques in such kind of HMI applications has not been clarified yet. In fact, there are several factors to take into account in real operational scenarios. For example, both EEG and *fast optical signal* (FOS)-based fNIR have similar bandwidth and sample rate requirements, as the FOS appears to directly reflect aggregated neural spike activity in real time and can be used as a high-bandwidth signal akin to EEG (Medvedev et al., 2008). However, EEG and fNIRs systems have different physical interfaces, sizes, weights, and power budgets, thus different wearability and usability in real operational contexts. Specifically, the physical interface merits scrutiny as it is nontrivial to maintain a good contact between the sensors (ie, electrodes or optodes) and the brain scalp in freely moving tasks. It is worth noticing that fNIRs is not affected by motion artifacts and does not require both scalp abrasion and conductive gel. In addition, there is not the necessity to wear a cap but only a headband. Furthermore, unlike EEG, fNIRs recordings are not affected by electroculographic and environmental electrical noise, and less sensitive to facial muscular activity, which are undoubtedly ubiquitous in human–computer interactions. Thus, fNIR technology could appear more suitable in realistic environments (Durantin et al., 2014; Goldberg et al., 2011; Izzetoglu et al., 2004; Owen et al., 2005).

However, in a recent study, Harrison et al. (2014) reported how the BOLD signal showed a lower resolution than the subjective measures (ISA; Kirwan et al., 2001) to evaluate the mental workload of ATCOs involved in the experiment. In particular, while the task was becoming more difficult, the subjective measure was still increasing, and the BOLD signal (neurophysiological index) reached its maximum, lingering on this value. Furthermore, the BOLD signal, used as workload index, was shown to be not reliable over time since the workload measurements performed in different days were significantly different and in discordance with the subjective measures.

In addition, since the presence of hair may impact on both photon absorption (Murkin and Arango, 2009) and the coupling of the probes with the underlying scalp, the fNIRs technique is very reliable only on those unhairy brain areas, like the PFC.

As quoted earlier, the parietal brain sites also play a key role in the mental workload evaluation, and Derosière et al. (2013) pointed out how some fNIRs-measured hemodynamic variables were relatively insensitive to certain changes in mental workload and attentional states.

Due to its higher temporal resolution and usability, in comparison with the fNIRs technique, the EEG technique overcomes such kind of issues. In addition, there are several studies in ATM domain that highlighted the high reliability of EEG-based mental workload indexes (W_{EEG}; Brookings et al., 1996). The results showed that the effects of the task demand were evident on the EEG rhythms variations. EEG power spectra increased in the theta band, while significantly decreased in the alpha band as the task difficulty increased, over central, parietal, frontal, and temporal brain sites. In a recent study, Shou et al. (2012) evaluated the mental workload during an air traffic control (ATC) experiment using a new *time–frequency-independent component analysis* (tfICA) method for the analysis of the EEG signal. They found that "the frontal theta EEG activity was a sensitive and reliable metric to assess workload and time-on-task effect during an ATC task at the resolution of minute(s)." In other recent studies involving professional and trainees ATCOs (Aricò et al., 2014b, 2015c; Di Flumeri et al., 2015), it was demonstrated how it was possible to compute an EEG-based workload index able to significantly discriminate the workload demands of the ATM task by using machine-learning techniques and frontal–parietal brain features. In those studies, the ATM tasks were developed with a continuously varying difficulty levels in order to ensure realistic ATC conditions, ie, starting form an easy level, then increasing up to a hard one and finishing with an easy one again. The EEG-based mental workload indices showed to be directly and significantly correlated with the actual mental demand experienced by the ATCOs during the entire task.

The same EEG-based workload index was also used to evaluate and compare the impact of different avionic technologies on the mental workload of professional helicopter pilots (Borghini et al., 2015b). Furthermore, the machine-learning techniques have been successfully used in other real environments for the evaluation of mental states (Müller et al., 2008) and mental workload (Berka et al., 2004, 2007). Another interesting application of the neurophysiological workload evaluation was proposed by Borghini et al. (2014a), where a neuroelectrical metric was defined and used for the training assessment of subjects while learning to execute correctly a new task.

Even if the main limitation of the EEG is its wearability, technology improvements (Liao et al., 2012) have being developed and tested in terms of dry electrodes (no gel and impedances adaptation issues), comfort, ergonomic, and wireless communications (no cables between EEG sensors and the recording system).

In conclusion, the EEG technique seems to be the appropriate solution to evaluate the mental workload in realistic and operational settings, and to be integrated in passive BCI systems. Such systems will support the operator during his/her working activity in order to improve the works wellness and, most of all, the safety standards of the whole environment.

1.4 PRESENT STUDY

EEG-based p-BCI system potentially appears as the best solution for the user's mental workload estimation in real operational environments. However, this technology could suffer from some practical issues, such as the necessity to be recalibrated each time before its use, reduction of reliability over time (Christensen et al., 2012; Pollock et al., 1991), and intrusiveness due to the high number of EEG electrodes.

The objective of the proposed study was to provide an example of passive BCI-based methodology to evaluate the users' mental workload in operational environment, and to overcome the issues described previously. For such purposes, the brain activity was recorded on 12 professional ATCOs while performing high realistic ATM scenarios. From the EEG signals, a W_{EEG} was computed by means of the *automatic-stop stepwise linear discriminant analysis* (asSWLDA) (Aricò et al., 2015a), a modified version of the standard SWLDA.

To summarize, the proposed study has been organized in order to investigate two important key issues to use neurophysiological measurements in operational environments: the overtime reliability of the measure and the accuracy of the methodology in comparison with the standard (ie, subjective) workload measures.

2 MATERIALS AND METHODS

2.1 EXPERIMENTAL PROTOCOL

2.1.1 Subjects

Twelve professional (40.41 ± 5.54 years old) ATCOs from the *École Nationale de l'Aviation Civile* (ENAC) of Toulouse (France) have been involved in this study. They were selected in order to have a homogeneous experimental group in terms of age and expertise. The ENAC represents one of the most important training schools for pilots and ATCOs in the world. The experimental procedures involving human subjects described in this chapter were approved by the Institutional Review Board.

2.1.2 Experimental task

ATCOs have been asked to perform a series of ATM scenarios in realistic settings. Such particular settings, developed and hosted at ENAC (Fig. 1A), consisted in a functional simulated ATM environment with a 30″ screen (RADAR screen) to display radar image and a 21″ screen (ATM interface) to interact with the radar image (zoom, move, clearances, and information). The experiments have also been attended by two pseudopilots (Fig. 1B) who have interacted with the ATCOs with the aim to simulate real-flight communications.

The complexity of the task could be modulated according to how many aircraft the ATCO had to control the number and type of clearances required over the time and the number/trajectory of other interfering flights. The experiments have taken place in two different sessions, a month on, named hereafter as *Day 1* and *Day 30*.

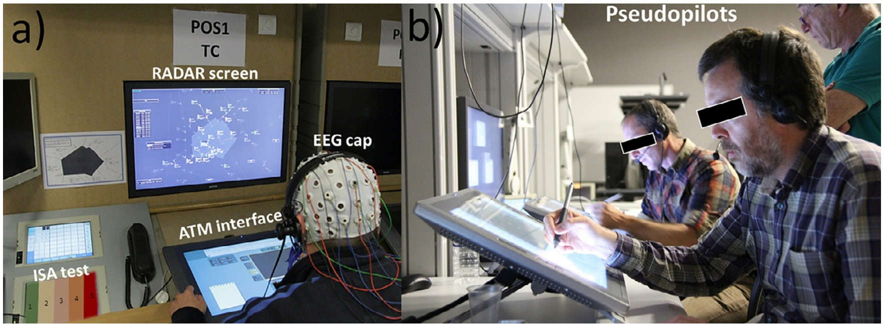

FIG. 1

(A) The ENAC simulator platform, composed of two screens, a 30″ (RADAR) screen to display radar image and a 21″ screen to interact with the radar image (ATM interface). On the little screen on the *left bottom*, the ISA test was proposed every 3 min. (B) Pseudopilots have interacted with the ATCOs with the aim to simulate real-flight communications.

For each session, ATCOs have been asked to perform a different 45-min ATM scenario enclosing three different levels of complexity (15 min for each complexity condition) associated to three different mental workload demands (EASY, MEDIUM, and HARD). For each scenario, the presentation of the difficulty conditions has been randomized. In addition, although the two ATM scenarios were different, in order to avoid any habituation or expectation effect, they have been designed identically in terms of complexity within the same difficulty levels (ie, for instance EASY Day 1 vs EASY Day 30), to make them comparable between sessions, and to avoid any bias in the results. Such scenarios have been validated and tested by a *subject-matter expert* (SME) from ENAC before the experiments.

It has to be stressed that air traffic shape was not constant, and the transitions between the different difficulty levels were smoothly organized, in order to have ATM scenarios as much realistic as possible. Fig. 2 shows a representative scenario's complexity shape.

2.1.3 Collected data related to the mental workload of ATCOs

2.1.3.1 Neurophysiological data

For each ATCO, scalp EEG signals have been recorded by the digital monitoring *BEmicro* system (EBNeuro system) with a sampling frequency of 256 (Hz) by 8 Ag/AgCl passive wet electrodes (Fz, F3, F4, AF3, AF4, Pz, P3, and P4) referenced to both the earlobes and grounded to the Cz electrode, according to the 10–20 standard (Jurcak et al., 2007). In addition, the vertical EOG signal has been recorded concurrently with the EEG, and with the same sampling frequency (256 Hz), by a bipolar channel over the left eye, in order to collect the eyes blink of the subjects during the execution of the task.

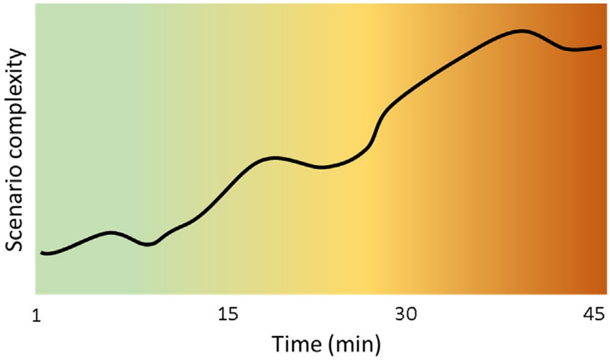

FIG. 2

Representative scenario's complexity shape.

2.1.3.2 Subjective workload assessment (self-assessment)

Simultaneously to the execution of the ATM task, ATCOs have been asked to fill the ISA. In particular, the ISA (Kirwan et al., 2001) is a technique that has been developed to provide immediate subjective ratings of workload demands, from 1 (very easy) to 5 (very difficult), during the execution of a task. The ISA technique has provided a profile of the operator's workload perception throughout the ATM scenarios. The ISA test scale has been presented to the ATCOs every 3 min in the form of a color-coded keypad on a screen placed on the left of the main monitor (Fig. 1A). The keypad flashed and sounded when the workload rating was required, and the participants simply pushed the button related to their workload perception.

2.1.3.3 Subjective Workload Assessment (SME Assessment)

ATC experts (SMEs) seated behind the ATCO, and they have been asked to provide independent rate, from 1 to 5 according to the ISA scale, of the ATCO's mental workload (by filling the paper version of the ISA), in order to have an extra mental workload evaluation experienced by the ATCOs. In particular, SMEs have been asked to express their continuous judgment depending on the ongoing overall performance of the examined ATCO. Such judgment took into account the quality and time of the indications/information provided to the pilots, separation planning strategy, response responsiveness, and general management of the air traffic condition.

ISA scores provided by experimental subjects and by the SMEs are named hereafter, respectively, *SELF-ISA* and *SME-ISA*.

2.1.4 People involved and study organization

The experimental setting consisted in two ATC positions (Fig. 3), two external ATC experts (SMEs), two biosignal measurements experts, and two pseudopilots.

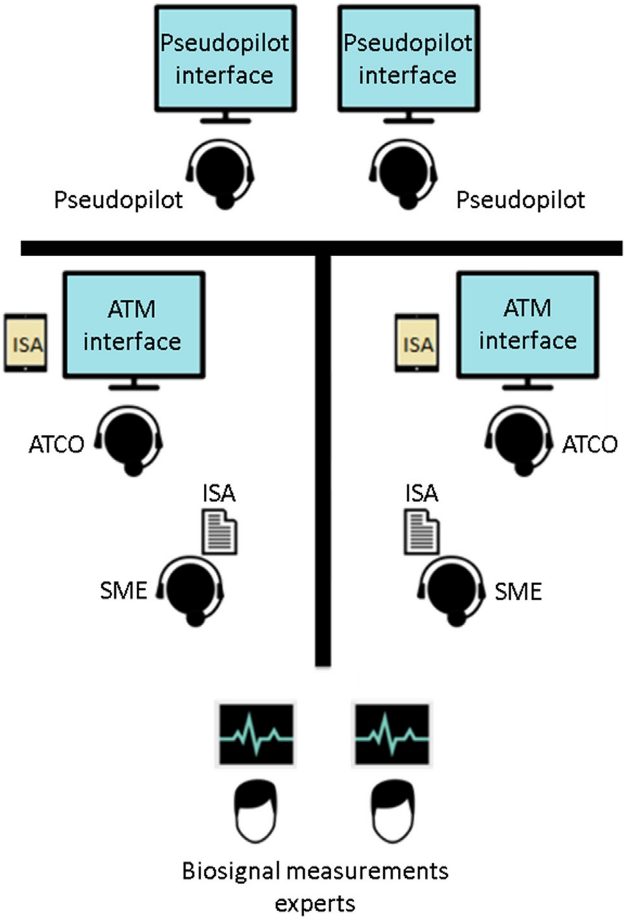

FIG. 3

Experimental settings, involving two experimental subjects (ATCOs), two external ATC Experts (SMEs), two biosignal measurement experts, and two pseudopilots at the same time.

The experimental study has been organized in two main phases:

– *Reliability over time of the neurophysiological workload measure*. The first phase of the study had the aim to test the reliability over time (1 month) of the neurophysiological workload measure, by using two different models (see Section 2.2.2). Five of the 12 mentioned ATCOs gave their availability to take part in both the experimental sessions (Day 1 and Day 30), while the remaining controllers attended only the last experimental session (Day 30).
– *Comparison between neurophysiological and subjective workload assessment*. The second phase of the study had the aim to test the effectiveness of the neurophysiological workload measure in comparison with the subjective

assessment (ISA and SME workload scores). To investigate this point, only the last experimental session (Day 30) has been used, where all the ATOCs participated to the experiment (12 professional ATCOs in total).

2.2 NEUROPHYSIOLOGICAL DATA ANALYSIS

2.2.1 EEG signal processing

For each session (Day 1 and Day 30) and for each difficulty level (EASY, MEDIUM, and HARD), the biosignal data set (EEG and EOG signals) has been segmented in five consecutive parts (named hereafter as "runs") of 3 min each, in order to have five EASY runs (E1, E2, E3, E4, and E5), five MEDIUM runs (M1, M2, M3, M4, and M5), and five HARD runs (H1, H2, H3, H4, and H5), and to have the same time resolution of the ISA scores provided by both the ATCOs and the SMEs (SELF-ISA and SME-ISA), and therefore allowing a more direct comparison between all the collected measures.

The recorded EEG signal has been firstly band-pass filtered with a fourth-order Butterworth filter (low-pass filter cutoff frequency: 30 (Hz), high-pass filter cutoff frequency: 1 (Hz)). The EOG signal, band-pass filtered too with a fourth-order Butterworth filter (low-pass filter cutoff frequency: 7 (Hz), high-pass filter cutoff frequency: 1 (Hz)), has then been used to remove eyes-blink contributions from each epoch of the EEG signal, by using the Gratton et al. (1983) algorithm available on the EEGLab toolbox (Delorme and Makeig, 2004). This last step has been performed because the eyes-blink contribution could affect the frequency bands correlated to the mental workload, in particular the theta EEG band. For other sources of artifacts (ie, ATC-operators normally communicate verbally and perform several movements during their operational activity), specific procedures of the EEGLAB toolbox have been used (Delorme and Makeig, 2004). First, the EEG signal has been segmented into epochs of 2 s (*Epoch length*), shifted of 0.125 s (*Shift*). This windowing have been chosen with the compromise to have both a high number of observations (see Eq. 1), in comparison with the number of variables (see Eq. 2) and to respect the condition of stationarity of the EEG signal (Elul, 1969). In fact, this is a necessary hypothesis in order to proceed with the spectral analysis of the signal. The EEG epochs where the signal amplitude exceeds $\pm 100\ \mu V$ (*threshold criteria*) have been marked as "artifact." Then, each EEG epoch has been interpolated in order to check the slope of the trend within the considered epoch (*trend estimation*). If such slope was higher than 3, the considered epoch was marked as "artifact." The last artifact check has been based on the EEG *sample-to-sample* difference. If such difference, in terms of amplitude, was higher than 25 μV, it meant that an abrupt variation (no-physiological) happened, and the EEG epoch has been marked as "artifact." All the previous values have been chosen following the guidelines reported in Delorme and Makeig (2004). At the end, the EEG epochs marked as "artifact" have been removed from the EEG data set with the aim to have a clean EEG signal from which estimate the brain parameters for the different analyses.

The percentage (with respect to the total number of epochs averaged on all the subjects) of EEG epochs containing artifacts and removed from the EEG data set was 20% (±13%).

From the clean EEG data set, the *power spectral density* (PSD) was calculated for each EEG epoch using a Hanning window of the same length of the considered epoch (2 s length (that means 0.5 Hz of frequency resolution)). The application of a Hanning window helped to smooth the contribution of the signal close to the end of the segment (Epoch), improving the accuracy of the PSD estimation (Harris, 1978).

Then, the EEG frequency bands of interest have been defined for each ATCO by the estimation of the *individual alpha frequency* (IAF) value (Babiloni et al., 2000, 2001; De Vico Fallani et al., 2010; Klimesch, 1999). In order to have a precise estimation of the alpha peak, hence of the IAF, the subjects have been asked to keep the eyes closed for a minute before starting with the experiments. Finally, a spectral features matrix (EEG channels × frequency bins) has been obtained in the frequency bands directly correlated to the mental workload. In particular, only the theta rhythm (IAF − 6:IAF − 2), over the EEG frontal channels (Fz, F3, F4, AF3, and AF4), and the alpha rhythm (IAF − 2:IAF+2), over the EEG parietal channels (Pz, P3, and P4), have been considered as variables for the mental workload evaluation.

In the considered case, the number of observations (#Observations) and number of variables (Variables) are as follows:

$$\#\text{Observations} = \frac{RunDuration - EpochLength}{\text{Shift}} = 1424 \tag{1}$$

where $RunDuration = 180$ s (E_k, M_k, H_k), $k = [1,2, \ldots, 5]$, $EpochLength = 2$ s, and $\text{Shift} = 0.125$ s.

$$\begin{aligned}\#\text{Variables} = {} & (\#\text{FrontalSites} * \#\text{ThetaFrequencyBins}) \\ & + (\#\text{ParietalSites} * \#\text{AlphaFrequencyBins}) = 72\end{aligned} \tag{2}$$

where #FrontalSites = 5 (namely Fz, F3, F4, AF3, and AF4), #ParietalSites = 3 (namely Pz, P3, and P4), $\#\text{ThetaFrequencyBins} = (\text{IAF} - 6{:}\text{IAF} - 2) \cdot 0.5 = 9$, and $\#\text{AlphaFrequencyBins} = (\text{IAF} - 2{:}\text{IAF} + 2) \cdot 0.5 = 9$.

2.2.2 EEG-based mental workload index

Fig. 4 shows the algorithm steps used in this study for the estimation of the user's W_{EEG}.

The effectiveness of two linear classifiers, in particular the standard *stepwise linear discriminant analysis*, SWLDA (Aloise et al., 2012; Draper, 1998), and the asSWLDA (Aricò et al., 2015a), was investigated in terms of reliability over a month. In particular, with respect to the standard SWLDA approach, the asSWLDA algorithm embeds an automatic procedure to select the best number of relevant features to keep into the discrimination model. This property was demonstrated so far to increase the robustness to the under and the overfitting phenomenon, over a week.

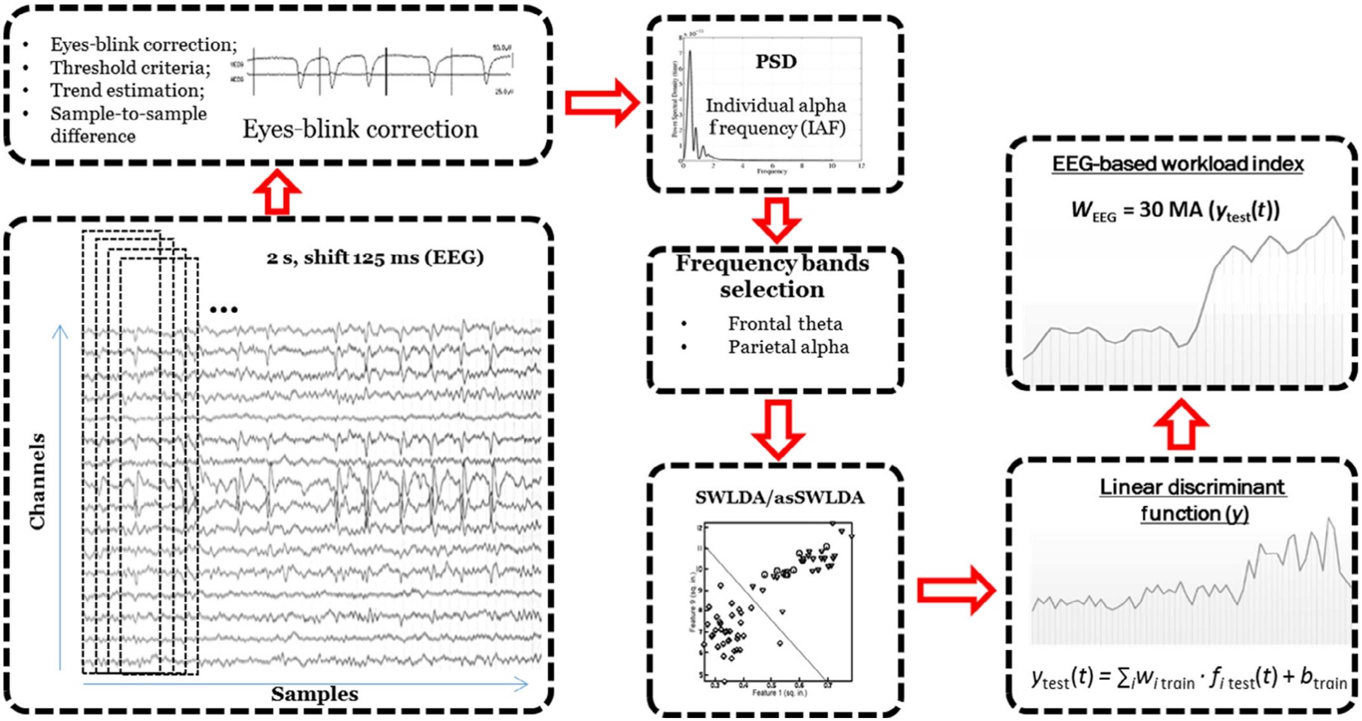

FIG. 4

EEG-based mental workload index (W_{EEG}). The figure explains the algorithm for the EEG-based workload index estimation. The band-pass filtered (1–30 Hz) EEG signal has been segmented into epochs of 2 s, shifted of 0.125 s, and the band-pass filtered (1–7 Hz) EOG signal has been used to remove the eyes-artifact contribution from the EEG signal. Other sources of artifacts have been deleted by using specific algorithms. Then, the *power spectral density* (PSD) has been evaluated for each epoch ($Epoch_{PSD}$), taking into account only the EEG frequency bands and channels correlated with the mental workload variations (frontal theta and parietal alpha bands). Two linear classifiers (SWLDA and asSWLDA) have then been used to select the most relevant brain spectral features for the discrimination of the mental workload levels. The linear discriminant function has been calculated on the EEG testing data sets, and the W_{EEG} has been defined as the moving average of 30 s (30 MA) applied to the linear discriminant function.

In other words, by training the asSWLDA with data acquired on a day, it is possible to maintain high classification performances without any further recalibration of the algorithm on data acquired until a week after (Aricò et al., 2015a). The asSWLDA algorithm is freely available in www.brainsigns.com.

Both the classifiers were used to select the most relevant discriminant features within the different experimental conditions (ie, EASY and HARD), related to the lowest and the highest complexity of the task. Once identified, each classifier assigns to each relevant feature-specific weights $\left(w_{i_{\text{train}}}\right)$, plus a bias ($b_{\text{train}}$). On the contrary, weights related to those features not relevant for the classification model have been set to 0. This step represented the training phase of the algorithm.

Later on, the classifier parameters estimated during the training phase have been used to calculate the *linear discriminant function* ($y_{\text{test}}(t)$) over the testing EEG data set (testing phase, Eq. 3), defined as the linear combination of the testing spectral features $\left(f_{i_{\text{test}}}\right)$ and the classifier weights $\left(w_{i_{\text{train}}}\right)$, plus the bias ($b_{\text{train}}$).

Finally, a moving average of k seconds (kMA) has been applied to the $y_{\text{test}}(t)$ function in order to smooth it out by reducing the variance of the measures, and the result has been named *EEG-based workload index* (W_{EEG}, Eq. 4). The higher is the k value, the less will be the variance of the measure. Accordingly with the SMEs, for a proper evaluation of the mental workload during the execution of ATM tasks, the k value has been set to 30 s:

$$y_{\text{test}}(t) = \sum_i w_{i_{\text{train}}} {}^* f_{i_{\text{test}}}(t) + b_{\text{train}} \tag{3}$$

$$W_{\text{EEG}} = \text{kMA}(y_{\text{test}}(t)) \quad k = 30\,\text{s} \tag{4}$$

2.3 PERFORMED DATA ANALYSES

As stated earlier, the proposed study has been organized in two main sections, in order to describe two important key issues of using passive BCI methodologies in operational environments: the reliability over time of the measure, and the accuracy in comparison with the standard (ie, subjective) workload measures.

2.3.1 Reliability over time of the neurophysiological workload measure

2.3.1.1 Subjective workload assessment

Firstly, the two EEG recording sessions (Day 1 and Day 30) have been compared in terms of subjective workload measure (*SELF-ISA* and *SME-ISA* scores), to check any differences between the workload perception of the ATCOs and of the SMEs. Three two-way ANOVAs (CI = 0.95) have been performed, one for each difficulty level (EASY, MEDIUM, and HARD), on the *SELF-ISA* and the *SME-ISA* scores in order to find out the difference between the two experimental sessions.

2.3.1.2 EEG-based workload assessment

After demonstrating that the two sessions (Day 1 and Day 30) did not differ in terms of perceived workload (see Section 2.3.1.1) for each considered difficulty level (EASY, MEDIUM, and HARD), the stability of the neurophysiological workload measures (W_{EEG}) has been investigated over a month, by using the two classifiers, the SWLDA and the asSWLDA. In particular, two different kinds of cross-validations have been defined: (i) the *Intra*-cross-validation type, where the training and testing data belonged to the same day, and (ii) the *Inter*-cross-validation type, where the training data belonged to Day 1 and the testing data to Day 30.

The third couple of runs (E3 and H3) of the first session (Day 1) has been chosen to calibrate the classifiers. In fact, since the ATM scenarios' profile has been designed without any constant traffic samples or sudden transitions, the easy (E3) and hard (H3) conditions in the middle of each difficulty level have been considered the best choice for training the classifier. Such choice is the best compromise in terms of stable difficulty level to represent the lowest and the highest air traffic complexity condition (and related workload demand), respectively.

To evaluate the reliability of each classifier in discriminating EASY and HARD difficulty levels along the different cross-validation types, *area under curve* (AUC) values of the *receiver operating characteristic* (ROC; Bamber, 1975) have been calculated by considering couple of W_{EEG} distributions (E vs H). An ROC curve is a graphical plot that illustrates the performance of a binary classifier to discriminate two classes. The area under an ROC curve (AUC, which can assumes values comprised between 0.5 and 1) quantifies the overall ability of a binary classifier to discriminate between two conditions (ie, EASY and HARD). If the two conditions are not discriminable, the AUC assumes value of 0.5. On the contrary, if the two conditions are perfectly discriminable, the AUC assumes value of 1.

A two-way repeated measures ANOVA ($\mathrm{CI}=0.95$) analysis has been performed on the AUC values, by considering as *within* factors the "classifiers" (asSWLDA and SWLDA) and the "cross-validation types" (Intra and Inter). Furthermore, Duncan post hoc tests (a multiple comparison procedure, Duncan, 1955) have been performed to assess significant differences between all pairs of levels of the considered factors.

2.3.1.2.1 EEG features selection analysis. As stated earlier, both the SWLDA and the asSWLDA algorithms set to 0 all the weights related to the brain features not significant for the classification model. In addition, the asSWLDA algorithm has the ability to automatically select the lowest number of features from the EEG training data set to optimize the regression model. This property has been demonstrated to make the model able to overcome both the under and the overfitting phenomena (Aricò et al., 2015a). On the contrary, the standard SWLDA has not this property and the algorithm selects, from the EEG training data set, the maximum number of brain features that optimize the regression model until the features selection criteria are satisfied (Draper, 1998).

The hypothesis of the study was that the asSWLDA might be able to achieve high discrimination accuracy, by using a lower number of features (and of EEG channels) than the standard SWLDA algorithm. In this regard, it is easy to realize that for a practical use of p-BCI systems in operational environments, the less is the number of EEG channels, the smaller and less intrusive will be the EEG system.

Two-tailed paired *t*-tests ($\alpha = 0.05$) have been performed: (i) the first to compare the number of total features among the considered algorithms, and (ii) the second one to compare the related number of EEG channels selected by the two models (standard SWLDA and asSWLDA). For the analyses, both the number of features and the EEG channels used in the two sessions have been averaged for each model (SWLDA and asSWLDA).

2.3.2 Comparison between neurophysiological and subjective workload evaluation

As reported earlier, the analyses have been performed by considering only the last session (Day 30), where all the 12 ATCOs participated to the experiment.

2.3.2.1 Self-workload assessment

First, the three difficulty conditions (EASY, MEDIUM, and HARD) have been compared in terms of perceived workload, by using both *SELF-ISA* and *SME-ISA* scores, to assess if the three difficulty conditions had been perceived differently by the ATCOs. In addition, the two subjective scores have been compared for each difficulty condition. In particular, a two-way ANOVA (CI = 0.95) has been conducted on the *SELF-ISA* and *SME-ISA* score, by considering as *within* factors the "difficulty conditions" (EASY, MEDIUM, and HARD) and the "subjective workload scores" (SELF-ISA and SME-ISA), and by averaging the scores across each difficulty level. Furthermore, Duncan post hoc tests have been performed to assess differences between all pairs of levels of the considered factors.

2.3.2.2 EEG-based workload assessment

The E3 and H3 runs of the second session (*Day 30*) have been selected to calibrate the asSWLDA classifier and the W_{EEG} has then been estimated for each ATCO (as ascribed earlier) over the remaining runs (E1, E2, E4, E5, M1, M2, M3, M4, M5, H1, H2, H4, and H5).

A one-way ANOVA (CI = 0.95) has been performed on the W_{EEG} index, by considering as *within* factor the "difficulty conditions" (EASY, MEDIUM, and HARD), by averaging for each ATCO all the W_{EEG} indexes for each difficulty level. Finally, Duncan post hoc tests have been performed to assess differences between all pairs of levels of the considered factor.

2.3.2.3 Accuracy of neurophysiological measurement in comparison with standard workload assessment

In order to assess the accuracy of the proposed methodology for the mental workload assessment, in comparison with standard subjective workload measures (eg, ISA), a *Pearson's correlation* analysis has been done between the W_{EEG} index and both the subjective scores (*SELF-ISA* and *SME-ISA*). Thus, the *Fisher's R-to-Z* transformation (Fisher, 1921) has been performed in order to assess possible differences between the correlation coefficients (W_{EEG} vs *SELF-ISA and* W_{EEG} vs *SME-ISA*).

Before every statistical analysis, the *z*-score transformation (Zhang et al., 1999) has been used to normalize the data.

3 RESULTS

3.1 OVERTIME STABILITY OF THE EEG-BASED WORKLOAD MEASURE

3.1.1 Self-workload assessment

The three two-way ANOVAs (CI = 0.95) did not highlight any significant difference between the two sessions (Day 1 and Day 30) in terms of both *SELF-ISA and SME-ISA* (EASY: $F(1,16) = 0.015$, $p = 0.90$; MEDIUM: $F(1,16) = 2.48$, $p = 0.13$; HARD: $F(1,16) = 1.77$, $p = 0.20$) scores (Fig. 5). The results replicate perfectly the correctness of the ATM scenarios designed by the ATC Experts from ENAC. In fact, the requirements were to define different realistic ATM scenarios with comparable difficulty conditions (EASY, MEDIUM, and HARD), with the aim to avoid habituation and expectation effects.

3.1.2 EEG-based workload assessment

The two-way repeated measures ANOVA (CI = 0.95) highlighted a significant interaction effect between the two (classifiers and cross-validations) factors ($F(1,4) = 10.6$, $p = 0.03$). The post hoc test highlighted a significant decrement ($p = 0.005$)

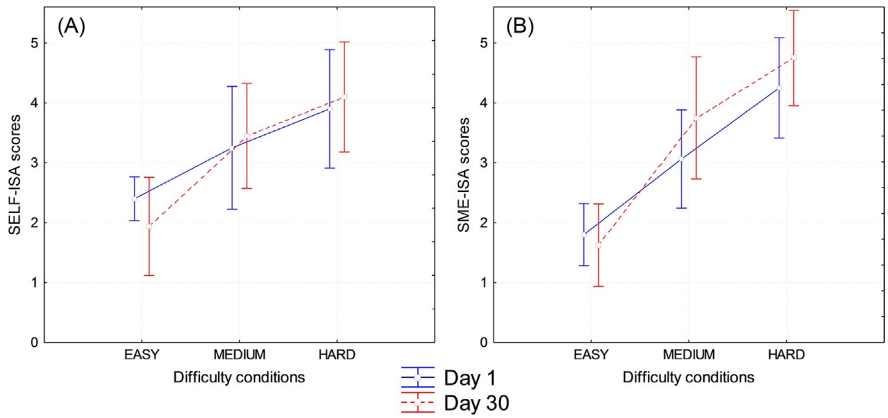

FIG. 5

Error bars related to the *SELF-ISA* and the *SME-ISA* scores for each session (Day 1 and Day 30) and difficulty condition (EASY, MEDIUM, and HARD). The two-way ANOVAs did not show any significant differences between the two experimental sessions for each difficulty level (EASY, MEDIUM, and HARD).

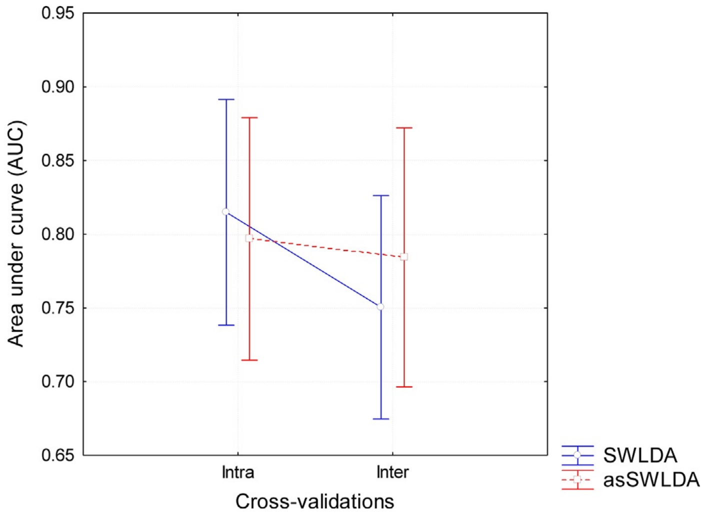

FIG. 6

Error bars (CI = 0.95) related to the AUC values of the SWLDA and asSWLDA classifiers calculated on the EASY and HARD conditions over the two cross-validation types (*Intra* and *Inter*). In particular, regarding the SWLDA, there is a significant AUC decrement (p=0.005) between the *Intra*- and *Inter*-cross-validation types. On the contrary, there is no significant difference (p=0.33) between the two cross-validation types, concerning the asSWLDA. Focusing on the *inter*-cross-validation type, the SWLDA performance (ie, AUC) is significantly (p=0.04) poorer that the asSWLDA one.

of the AUC values related to the SWLDA classifier between the *Intra*- and the *inter*-cross-validation types. On the contrary, no significant differences (p = 0.33) were highlighted for the asSWLDA between the *Intra*- and the *Inter*-cross-validations. In addition, a significant decrement (p = 0.04) of the AUC values related to the *Inter*-cross-validation type was highlighted between the asSWLDA and the SWLDA classifiers. Instead, no significant differences (p = 0.2) between the two classifiers were highlighted regarding the *Intra*-cross-validation type (Fig. 6).

3.1.2.1 EEG features selection analysis

The 2 two-tailed paired t-tests (α = 0.05) highlighted that both the number of features (p = 0.0007) and the related EEG channels (p = 0.0003) used by the asSWLDA model were significantly lower than those used by the standard SWLDA model (Fig. 7). In Fig. 7, both the numbers of features and EEG channels are reported in terms of percentage in respect to the total number of features (72) and EEG channels (8). In particular, the asSWLDA selected the 5.2% of the available features by using the data from the 37% of EEG channels. This means that asSWLDA algorithm selected for each ATCO about four features on three EEG channels. On the contrary, the standard SWLDA used the 44% of the available features by using the 100% of EEG channels.

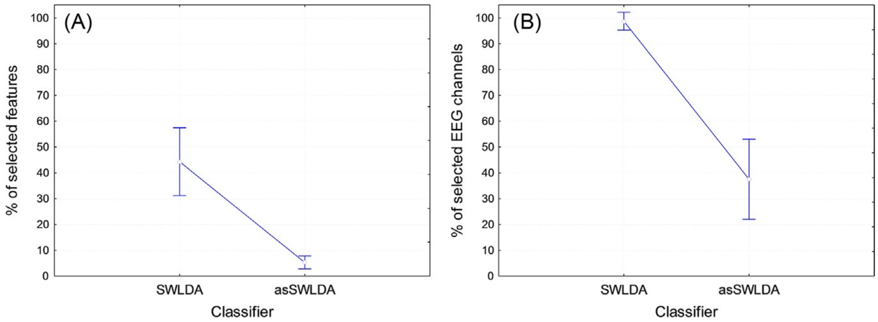

FIG. 7

Error bars related to the number of features (A) and related number of EEG channels (B) selected by the two classification models (SWLDA and asSWLDA). As expected, both the number of features and related number of EEG channels used by the asSWLDA were significantly lower than those used by the standard SWLDA algorithm ($p=0.0007$ and $p=0.0003$, respectively).

This means that the standard SWLDA used for each ATCO 32 features on all the 8 EEG channels. It could be concluded that the asSWLDA algorithm used roughly the 10% of the information employed by the SWLDA, achieving higher performance than the standard SWLDA. This result is important in the perspective of practical usage of such neurophysiological workload assessment in working environment.

3.1.3 Comparison between EEG-based and subjective workload assessment

3.1.3.1 Subjective assessment

The two-way repeated measures ANOVA (CI = 0.95) highlighted a significant interaction effect ($F(2,22)=10.88$, $p=0.005$) between the two factors, difficulty conditions (EASY, MEDIUM, and HARD) and subjective workload scores (*SELF-ISA* and *SME-ISA*). The post hoc test highlighted significant differences (all $p<0.001$) between the difficulty conditions (ie, EASY lower than MEDIUM, MEDIUM lower than HARD, EASY lower than HARD) for both the *SELF-ISA* and the *SME-ISA* scores. In addition, the *SME-ISA* scores over the MEDIUM ($p=0.0007$) and HARD ($p=0.00007$) conditions were significantly higher than those related to the *SELF-ISA* scores (Fig. 8).

3.1.3.2 EEG-based workload assessment

As for the subjective workload assessment, the one-way ANOVA (Fig. 9) on the neurophysiological workload measures (W_{EEG} data) highlighted a significant effect ($F(2,22)=27.4$, $p=0.000001$) between the three levels (EASY, MEDIUM, and HARD). In particular, the post hoc test highlighted significant differences (all $p<0.001$) between the W_{EEG} score related to the difficulty conditions (ie, EASY lower than MEDIUM, MEDIUM lower than HARD, EASY lower than HARD).

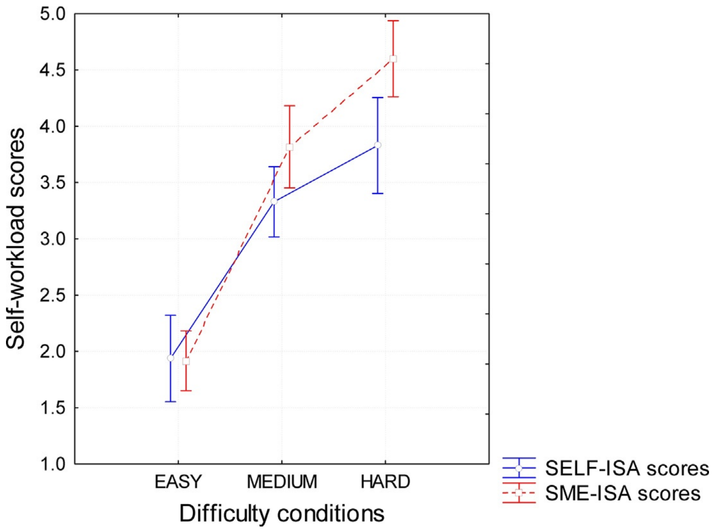

FIG. 8

Error bars (CI = 0.95) related to the *SELF-ISA* and the *SME-ISA* scores along the three difficulty conditions (EASY, MEDIUM, and HARD). Results showed significant differences ($p < 0.003$) between all the difficulty conditions for both the *SELF-ISA* and *SME-ISA* scores. In addition, the *SME-ISA* scores related to the MEDIUM and HARD conditions were significantly higher than those related to the *SELF-ISA* ones ($p = 0.0007$ and $p = 0.00007$, respectively).

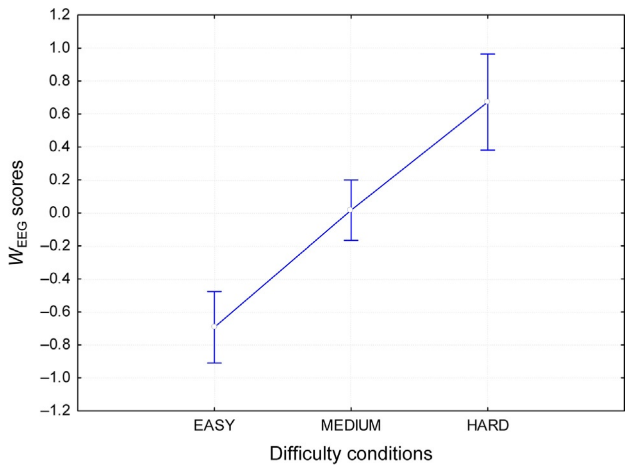

FIG. 9

Error bars (CI = 0.95) related to the W_{EEG} scores along the three difficulty conditions (EASY, MEDIUM, and HARD). Results showed significant differences between all the difficulty conditions ($p < 0.001$).

Table 1 Pearson's Correlation Coefficient (R) and Significance (p) Level Between the Neurophysiological (W_{EEG}) Workload Index and Both the Subjective Measures (ISA and SME Scores)

Statistical Analysis of the Correlation		**Pearson's Correlation Index**	
		R	***p***
12 subjects	W_{EEG} vs SELF-ISA W_{EEG} vs SME-ISA	0.856 0.797	0.0002 0.0011
Statistical Analysis of the Correlation		**Fisher's Transformation**	
		Z	***p***
12 subjects	$R_1 = 0.856$, $R_2 = 0.797$ $n = 13$, 2 tails	0.418	0.676

The Fisher's R-to-Z *transformation showed no significant difference between the two correlation values.*

3.1.3.3 Accuracy of neurophysiological measurement in comparison with standard workload assessment

The correlation analysis (Table 1), by means of the Pearson's correlation coefficient, highlighted a high and positive correlation between the EEG-based workload index (W_{EEG}) and both the ISA (SELF and SME) indexes. In particular the correlation analyses reported $R = 0.856$ and $p = 0.0002$ for the *SELF-ISA data*, and $R = 0.797$ and $p = 0.0011$ for the *SME-ISA* data. In other words, the shape of the three indexes was very similar; that is, the W_{EEG} was able to follow the variation of the mental workload demanded by the ATM scenarios and experienced by the ATCOs during the execution of the task (Fig. 10).

Finally, the Fisher's R-to-Z analysis (Table 1) on the two correlation indexes showed no differences between them ($p = 0.676$).

The scatter plot in Fig. 11 highlighted a high and positive correlation between the neurophysiological and the subjective workload indexes.

4 DISCUSSION

Passive BCI systems can provide reliable information about covert aspects of the user's mental state and overcome the low resolution and limitations of the conventional methods, such as behavioral or subjective measures (Zander and Jatzev, 2012). The information obtained by BCI techniques could then be employed to exploit novel information to improve HMIs. As a consequence, the human performance could be enhanced, and novel operative skills could be potentially achieved. There are many examples in which p-BCIs could be really useful. For example, p-BCI technology can reveal valuable information about the user's mental state in safety-critical applications, such as driving (Borghini et al., 2012; Welke et al., 2009), industrial environments (Venthur et al., 2010), or security surveillance (Marcel and Millan, 2007). With

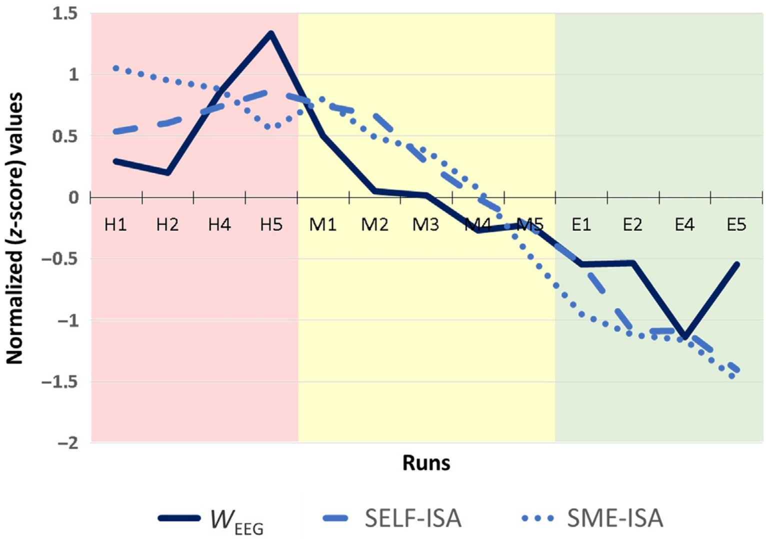

FIG. 10

Shape of the three workload indexes, the neurophysiologic (W_{EEG}), and the subjective (SELF- and SME-ISA) ones, across the experimental task. The three measures were able to follow the mental workload variation along the ATM scenarios.

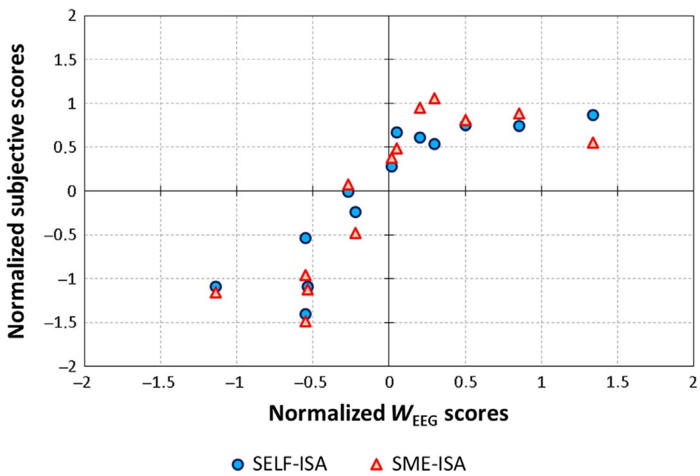

FIG. 11

Scatter plot of the subjective workload measures (*SELF-ISA* and *SME-ISA*) with respect to the neurophysiological workload measure. On *the x-axis* the normalized W_{EEG} index, on the *y-axis* the normalized *SELF-* and S*ME-ISA* indexes. The results of the correlation analyses showed a high and significant correlation among all the workload measures (Table 1).

respect to driving assistance applications, recent studies demonstrated the utility of p-BCI systems during driving simulation for assessing driving performance and inattentiveness (Schubert and Tangermann, 2008), as well as for robustly detecting emergency brakes before braking onset (Welke et al., 2009). In addition, p-BCI systems can potentially be used for real-time cognitive monitoring of the operator's mental workload (Aricò et al., 2014a,b; Kohlmorgen et al., 2007).

Anyhow, the application of the p-BCI technology outside research laboratories (eg, operational environments) has to face several practical issues, such as the reliability over time of the measure (unless a frequent recalibration of the system), or the intrusiveness of the equipment (eg, EEG sensors and recording system).

The purpose of the proposed study was to investigate the effectiveness of a methodology for the user's mental workload assessment in ATM environment, by using neurophysiological measures. The brain activity (EEG) and subjective measures (*SELF-ISA* and *SME-ISA* scores) of 12 professional ATCOs from ENAC (Toulouse, France) have been gathered while managing high-realistic ATM scenarios and analyzed by a machine-learning approach.

The asSWLDA, a modified version of the standard SWLDA (Aricò et al., 2015a), has been used to compute a mental workload index based on the EEG activity of the user (W_{EEG}). The SWLDA model has been chosen because it has been demonstrated to be one of the best outperforming linear classifiers (Aloise et al., 2012; Craven et al., 2006; Krusienski et al., 2006). In fact, in comparison with other linear methods, it has the advantage of having automatic features extraction, so that insignificant terms will be statistically removed from the model.

As stated earlier, the proposed study has been organized in order to investigate two important key issues to use neurophysiological measurements in operational environments: (1) the reliability over time of the measure, and (2) the accuracy in comparison with the standard (ie, subjective) workload measures.

Regarding the first issue, despite the low sample size (only five ATCOs took part to the whole experimental protocol), results demonstrated that the proposed algorithm has been able to maintain a high reliability across a month. It has been already demonstrated the reliability of the model over a week in a previous study (Aricò et al., 2015a). In other words, unlike the standard SWLDA algorithm, it is possible to calibrate the asSWLDA model and then use it every day without accuracy reductions in the mental workload evaluation up to a month after the calibration.

In addition, the asSWLDA model has been able to select a lower number of brain features and EEG channels (37%, that means three EEG channels in the specific study), to be used for the workload assessment, in comparison with the standard SWLDA (100%, that means eight EEG channels). In this regard, in the standard SWLDA algorithm it could be possible to force the model to select less features (that means less EEG channels; Aricò et al., 2015b). However, with the standard SWLDA, it became tricky to empirically (and manually) find out, so that it might change from subject to subject. On the contrary, the asSWLDA is able to automatically select the right number of brain features and to optimize the final model (Aricò et al., 2015a). More important, the results suggested that the selected features for the asSWLDA could be almost an order of magnitude less than those selected by using the standard SWLDA. As a direct consequence, a lower number of EEG channels will be necessary with the asSWLDA than with the standard SWLDA. Since one of the big limitations in using p-BCI systems in operational environments is represented by wearing the EEG cap, it is simple to conclude that the less is the number of EEG channels included in the classification model, the less intrusive will be the system for the final user. In this regard, just moving to a plausible operational showcase, the proposed algorithm will be calibrated once with many electrodes to select the specific subjective features and then used online by assume a lighter EEG configuration (eg, two or three electrodes).

Concerning the second issue, the EEG-based workload measure has been compared with two subjective workload measures, one provided by the ATCOs (*SELF-ISA* scores) and the second provided by two ATC Experts (SMEs), who have been asked to fill the ISA questionnaire at the same time of the ATCOs (*SME-ISA* scores). Results highlighted high and significant correlations between the neurophysiological and both the subjective workload measurements. Such result is very important, since it showed how the W_{EEG} and both the subjective indexes (*SELF-ISA* and *SME-ISA*) were able to follow the actual fluctuations of the mental workload experienced by the ATCOs during the experimental task. In addition, the *SELF-ISA* scores showed a higher workload perception during the MEDIUM and the HARD conditions in comparison with the *SME-ISA* ones. This result highlighted the main limitation of the subjective measures: they are highly operator dependent, and they cannot be used to quantify objectively the operators' mental states (ie, mental workload). On the contrary, the neurophysiological measures can provide, with high resolution, an objective evaluation of the operator's mental workload.

Finally, it has to be underlined that the proposed algorithms do not require a priori information about consecutive data; that is, the calibration of the classifier can be performed on a different day than the testing. As a consequence, the proposed methodologies can also be used for online applications, for example, to improve the HMIs by using information derived by the operator's mental workload states.

To summarize, it has been possible to calibrate the proposed algorithm by using EEG training data and then to evaluate the ATCO's mental workload during workshift across different days (up to a month).

The results reported in the present study do not exclude the possibility to achieve similar or even better results by using other machine-learning techniques, but they have demonstrated the possibility to use p-BCI systems in operational environments to measure the operators' mental workload during the execution of a realistic ATM task.

5 CONCLUSION

In this study, a passive BCI technique for the assessment of ATCOs mental workload has been proposed and tested in a high-realistic operational environment. In particular, the ATM environment imposes multiple concurrent demands on the ATCO, including air traffic monitoring, anticipating loss of separation between aircraft, and intervening to resolve conflicts. The objective assessment of the mental workload associated with operative activities in such a complex environment has long been recognized as an important issue (Gopher and Donchin, 1986).

Results showed that (i) the proposed algorithm maintained a high reliability over time (up to a month), (ii) high and significant correlations between the EEG-based and the subjective (*SELF-ISA* and *SME-ISA* scores) workload measures, and (iii) in comparison with other machine-learning techniques, the proposed algorithm (asSWLDA) was able to reduce the number of brain features and EEG channels, in other words, to reduce the intrusiveness of the p-BCI system.

One of the practical issues of machine learning approaches for their application outside of laboratory setting (ie, operational environments) is the necessity to calibrate the system every time before using it. In addition, it might not be always possible to find the right conditions (eg, EASY and HARD) to calibrate the system in operational environments. Different studies tried to address such issue in laboratory setting and only one in operational environment (Harrison et al., 2014) by using fNIRs technique. However, authors reported conflicting results related to the correlation between neurophysiological and self-assessed mental workload measures. We addressed partially this issue in the present study, since the asSWLDA algorithm can be calibrated once and then used over time (until a month) without requiring any calibration.

Taking into account these limitations, there is the need to perform further experiments to test the possibility to select brain features during the execution of short standard and controlled tasks (that enhance the considered mental state) and then to use such brain features to calibrate the classification algorithm that will be used later in the operative situation.

ACKNOWLEDGMENTS

This work is in part supported by a grant of the European Union through the Horizon2020—HCO funding scheme with the project “SmokeFreeBrain” GA no. 681120. The grant provided by the Italian Minister of University and Education under the PRIN 2012, Grant no. 2012WAANZJ scheme to F.B. is also gratefully acknowledged. The facilities provided by the ENAC, Toulouse, France under the WPE-SESAR research project “NINA”, the precious support of Drs. R. Benhacene, G. Granger, and J.P. Imbert from ENAC and S. Bonelli from DeepBlue srl, Rome are also gratefully acknowledged.

REFERENCES

Allison, B.Z., Pineda, J.A., 2003. ERPs evoked by different matrix sizes: implications for a brain computer interface (BCI) system. IEEE Trans. Neural Syst. Rehabil. Eng. 11, 110–113. http://dx.doi.org/10.1109/TNSRE.2003.814448.

Aloise, F., Schettini, F., Aricò, P., Bianchi, L., Riccio, A., Mecella, M., Babiloni, F., Mattia, D., Cincotti, F., 2010. Advanced brain computer interface for communication and control. In: Proceedings of the International Conference on Advanced Visual Interfaces, pp. 399–400.

Aloise, F., Schettini, F., Aricò, P., Salinari, S., Babiloni, F., Cincotti, F., 2012. A comparison of classification techniques for a gaze-independent P300-based brain-computer interface. J. Neural Eng. 9, 045012. http://dx.doi.org/10.1088/1741-2560/9/4/045012.

Aloise, F., Aricò, P., Schettini, F., Salinari, S., Mattia, D., Cincotti, F., 2013. Asynchronous gaze-independent event-related potential-based brain–computer interface. Artif. Intell. Med. 59, 61–69. http://dx.doi.org/10.1016/j.artmed.2013.07.006 (Special Issue: Brain-computer interfacing).

Aricò, P., Aloise, F., Schettini, F., Salinari, S., Mattia, D., Cincotti, F., 2014a. Influence of P300 latency jitter on event related potential-based brain-computer interface performance. J. Neural Eng. 11, 035008. http://dx.doi.org/10.1088/1741-2560/11/3/035008.

Aricò, P., Borghini, G., Graziani, I., Taya, F., Sun, Y., Bezerianos, A., Thakor, N.V., Cincotti, F., Babiloni, F., 2014b. Towards a multimodal bioelectrical framework for the online mental workload evaluation. Conf. Proc. IEEE Eng. Med. Biol. Soc. 2014, 3001–3004. http://dx.doi.org/10.1109/EMBC.2014.6944254.

Aricò, P., Borghini, G., Di Flumeri, G., Babiloni, F., 2015a. Metodo per stimare uno stato mentale, in particolare un carico di lavoro, e relativo apparato (A method for the estimation of mental state, in particular of the mental workload and its device). P1108IT00.

Aricò, P., Borghini, G., Di Flumeri, G., Colosimo, A., Graziani, I., Imbert, J.P., Granger, G., Benhacene, R., Terenzi, M., Pozzi, S., Babiloni, F., 2015b. Reliability over time of EEG-based mental workload evaluation during air traffic management (ATM) tasks. Conf. Proc. IEEE Eng. Med. Biol. Soc. 2015, 7242–7245. http://dx.doi.org/10.1109/EMBC.2015.7320063.

Aricò, P., Borghini, G., Graziani, I., Imbert, J.P., Granger, G., Benhacene, R., Pozzi, S., Napolitano, L., Di Flumeri, G., Colosimo, A., Babiloni, F., 2015c. ATCO: neurophysiological analysis of the training and of the workload. Ital. J. Aerosp. Med. 12, 18–34.

Babiloni, F., Carducci, F., Cincotti, F., Del Gratta, C., Roberti, G.M., Romani, G.L., Rossini, P.M., Babiloni, C., 2000. Integration of high resolution EEG and functional magnetic resonance in the study of human movement-related potentials. Methods Inf. Med. 39, 179–182.

Babiloni, C., Babiloni, F., Carducci, F., Cincotti, F., Rosciarelli, F., Rossini, P., Arendt-Nielsen, L., Chen, A., 2001. Mapping of early and late human somatosensory evoked brain potentials to phasic galvanic painful stimulation. Hum. Brain Mapp. 12, 168–179.

Baldwin, C.L., 2003. Commentary. Theor. Issues Ergon. Sci. 4, 132–141. http://dx.doi.org/10.1080/14639220210159807.

Bamber, D., 1975. The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. J. Math. Psychol. 12, 387–415. http://dx.doi.org/10.1016/0022-2496(75)90001-2.

Bellenkes, C.A.H., 2007. Contemporary issues in human factors and aviation safety. Ergonomics 50, 963–965. http://dx.doi.org/10.1080/00140130600971093.

Berka, C., Levendowski, D.J., Cvetinovic, M.M., Petrovic, M.M., Davis, G., Lumicao, M.N., Zivkovic, V.T., Popovic, M.V., Olmstead, R., 2004. Real-time analysis of EEG indexes of alertness, cognition, and memory acquired with a wireless EEG headset. Int. J. Hum. Comput. Interact. 17, 151–170. http://dx.doi.org/10.1207/s15327590ijhc1702_3.

Berka, C., Levendowski, D.J., Lumicao, M.N., Yau, A., Davis, G., Zivkovic, V.T., Olmstead, R.E., Tremoulet, P.D., Craven, P.L., 2007. EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks. Aviat. Space Environ. Med. 78, B231–B244.

Blankertz, B., Tangermann, M., Vidaurre, C., Fazli, S., Sannelli, C., Haufe, S., Maeder, C., Ramsey, L., Sturm, I., Curio, G., Müller, K.-R., 2010. The Berlin brain–computer interface: non-medical uses of BCI technology. Front. Neurosci. 4, 198. http://dx.doi.org/10.3389/fnins.2010.00198.

Borghini, G., Vecchiato, G., Toppi, J., Astolfi, L., Maglione, A., Isabella, R., Caltagirone, C., Kong, W., Wei, D., Zhou, Z., Polidori, L., Vitiello, S., Babiloni, F., 2012. Assessment of mental fatigue during car driving by using high resolution EEG activity and neurophysiologic indices. Conf. Proc. IEEE Eng. Med. Biol. Soc. 2012, 6442–6445. http://dx.doi.org/10.1109/EMBC.2012.6347469.

Borghini, G., Aricò, P., Ferri, F., Graziani, I., Pozzi, S., Napoletano, L., Imbert, J.P., Granger, G., Benhacene, R., Babiloni, F., 2014a. A neurophysiological training evaluation metric for air traffic management. Conf. Proc. IEEE Eng. Med. Biol. Soc. 2014, 3005–3008. http://dx.doi.org/10.1109/EMBC.2014.6944255.

Borghini, G., Astolfi, L., Vecchiato, G., Mattia, D., Babiloni, F., 2014b. Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. Neurosci. Biobehav. Rev. 44, 58–75. http://dx.doi.org/10.1016/j.neubiorev.2012.10.003.

Borghini, G., Aricò, P., Di Flumeri, G., Salinari, S., Colosimo, A., Bonelli, S., Napoletano, L., Ferreira, A., Babiloni, F., 2015a. Avionic technology testing by using a cognitive neurometric index: a study with professional helicopter pilots. Conf. Proc. IEEE Eng. Med. Biol. Soc. 2015, 6182–6185. http://dx.doi.org/10.1109/EMBC.2015.7319804.

Borghini, G., Aricò, P., Graziani, I., Salinari, S., Sun, Y., Taya, F., Bezerianos, A., Thakor, N.V., Babiloni, F., 2015b. Quantitative assessment of the training improvement in a motor-cognitive task by using EEG, ECG and EOG signals. Brain Topogr. 29, 149–161. http://dx.doi.org/10.1007/s10548-015-0425-7.

Brookings, J.B., Wilson, G.F., Swain, C.R., 1996. Psychophysiological responses to changes in workload during simulated air traffic control. Biol. Psychol. 42, 361–377.

Brouwer, A.-M., Zander, T.O., van Erp, J.B.F., Korteling, J.E., Bronkhorst, A.W., 2015. Using neurophysiological signals that reflect cognitive or affective state: six recommendations to avoid common pitfalls. Front. Neurosci. 9, 136. http://dx.doi.org/10.3389/fnins.2015.00136.

Cabeza, R., Nyberg, L., 2000. Imaging cognition II: an empirical review of 275 PET and fMRI studies. J. Cogn. Neurosci. 12, 1–47.

Cain, B., 2007. A Review of the Mental Workload Literature. Defence Research and Development Canada Toronto Human System Integration Section, Toronto, Canada. Report No.: RTO-TR-HFM-121-Part-II.

Calabrese, E.J., 2008. Neuroscience and hormesis: overview and general findings. Crit. Rev. Toxicol. 38, 249–252. http://dx.doi.org/10.1080/10408440801981957.

Christensen, J.C., Estepp, J.R., Wilson, G.F., Russell, C.A., 2012. The effects of day-to-day variability of physiological data on operator functional state classification. NeuroImage 59, 57–63. http://dx.doi.org/10.1016/j.neuroimage.2011.07.091.

Craven, P.L., Belov, N., Tremoulet, P., et al., 2006. Cognitive workload gauge development: comparison of real-time classification methods. In: Schmorrow, D., Stanney, K., Reeves, L. (Eds.), Augmented Cognition: Past, Present and Future, Foundations of Augmented Cognition. Strategic Analysis, Inc., Arlington, pp. 66–74.

Cui, X., Bray, S., Bryant, D.M., Glover, G.H., Reiss, A.L., 2011. A quantitative comparison of NIRS and fMRI across multiple cognitive tasks. Neuroimage 54, 2808–2821. http://dx.doi.org/10.1016/j.neuroimage.2010.10.069.

Delorme, A., Makeig, S., 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J. Neurosci. Methods 134, 9–21. http://dx.doi.org/10.1016/j.jneumeth.2003.10.009.

Derosière, G., Mandrick, K., Dray, G., Ward, T.E., Perrey, S., 2013. NIRS-measured prefrontal cortex activity in neuroergonomics: strengths and weaknesses. Front. Hum. Neurosci. 7. http://dx.doi.org/10.3389/fnhum.2013.00583.

De Vico Fallani, F., Nicosia, V., Sinatra, R., Astolfi, L., Cincotti, F., Mattia, D., Wilke, C., Doud, A., Latora, V., He, B., Babiloni, F., 2010. Defecting or not defecting: how to "read" human behavior during cooperative games by EEG measurements. PLoS One 5, e14187. http://dx.doi.org/10.1371/journal.pone.0014187.

Di Flumeri, G., Borghini, G., Aricò, P., Colosimo, A., Pozzi, S., Bonelli, S., Golfetti, A., Kong, W., Babiloni, F., 2015. On the use of cognitive neurometric indexes in aeronautic and air traffic management environments. In: Blankertz, B., Jacucci, G., Gamberini, L., Spagnolli, A., Freeman, J. (Eds.), Symbiotic Interaction. Springer International Publishing, Cham, pp. 45–56.

Draper, N.R., 1998. Applied regression analysis. Commun. Stat. Theory Methods 27, 2581–2623. http://dx.doi.org/10.1080/03610929808832244.

Duncan, D.B., 1955. Multiple range and multiple F tests. Biometrics 11, 1–42.

Durantin, G., Gagnon, J.-F., Tremblay, S., Dehais, F., 2014. Using near infrared spectroscopy and heart rate variability to detect mental overload. Behav. Brain Res. 259, 16–23. http://dx.doi.org/10.1016/j.bbr.2013.10.042.

Eggemeier, F.T., Wilson, G.F., Kramer, A.F., Damos, D.L., 1991. Workload assessment in multi-task environments. In: Multiple-Task Performance. Taylor & Francis, London, pp. 207–216.

Elul, R., 1969. Gaussian behavior of the electroencephalogram: changes during performance of mental task. Science 164, 328–331.

Feyer, A., Williamson, A.M., 1998. Human factors in accident modelling. In: Stellman, J.M. (Ed.), Encyclopedia of Occupational Health and Safety, fourth ed., International Labor Organization, Geneva.

Fisher, R.A., 1921. On the "probable error" of a coefficient of correlation deduced from a small sample. Metron 1, 3–32.

Fuchs, S., Hale, K.S., Stanney, K.M., Juhnke, J., Schmorrow, D.D., 2007. Enhancing mitigation in augmented cognition. J. Cogn. Eng. Decis. Mak. 1, 309–326. http://dx.doi.org/10.1518/155534307X255645.

Gevins, A., Smith, M.E., 2003. Neurophysiological measures of cognitive workload during human-computer interaction. Theor. Issues Ergon. Sci. 4, 113–131. http://dx.doi.org/10.1080/14639220210159717.

Gevins, A., Smith, M.E., McEvoy, L., Yu, D., 1997. High-resolution EEG mapping of cortical activation related to working memory: effects of task difficulty, type of processing, and practice. Cereb. Cortex 1991 (7), 374–385.

Goldberg, D.H., Vogelstein, R.J., Socolinsky, D.A., Wolff, L.B., 2011. Toward a wearable, neurally-enhanced augmented reality system. In: Schmorrow, D.D., Fidopiastis, C.M. (Eds.), Foundations of Augmented Cognition. Directing the Future of Adaptive Systems. Springer, Berlin, Heidelberg, pp. 493–499.

Gopher, D., Donchin, E., 1986. Workload: an examination of the concept. In: Handbook of Perception and Human Performance. Cognitive Processes and Performance, vol. 2. John Wiley & Sons, Oxford, England, pp. 1–49.

Gratton, G., Coles, M.G., Donchin, E., 1983. A new method for off-line removal of ocular artifact. Electroencephalogr. Clin. Neurophysiol. 55, 468–484.

Harris, F.J., 1978. On the use of windows for harmonic analysis with the discrete Fourier transform. Proc. IEEE 66, 51–83. http://dx.doi.org/10.1109/PROC.1978.10837.

Harrison, J., Izzetoglu, K., Ayaz, H., Willems, B., Hah, S., Ahlstrom, U., Woo, H., Shewokis, P.A., Bunce, S.C., Onaral, B., 2014. Cognitive workload and learning assessment during the implementation of a next-generation air traffic control technology using functional near-infrared spectroscopy. IEEE Trans. Hum.-Mach. Syst. 44, 429–440. http://dx.doi.org/10.1109/THMS.2014.2319822.

Hart, S.G., Staveland, L.E., 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In: Hancock, P.A., Meshkati, N. (Eds.), Human Mental Workload. Amsterdam, North-Holland Press.

Herff, C., Heger, D., Fortmann, O., Hennrich, J., Putze, F., Schultz, T., 2013. Mental workload during n-back task-quantified in the prefrontal cortex using fNIRS. Front. Hum. Neurosci. 7, 935. http://dx.doi.org/10.3389/fnhum.2013.00935.

Huppert, T.J., Hoge, R.D., Diamond, S.G., Franceschini, M.A., Boas, D.A., 2006. A temporal comparison of BOLD, ASL, and NIRS hemodynamic responses to motor stimuli in adult humans. NeuroImage 29, 368–382. http://dx.doi.org/10.1016/j.neuroimage.2005.08.065.

Izzetoglu, K., Bunce, S., Onaral, B., Pourrezaei, K., Chance, B., 2004. Functional optical brain imaging using near-infrared during cognitive tasks. Int. J. Hum. Comput. Interact. 17, 211–227. http://dx.doi.org/10.1207/s15327590ijhc1702_6.

Jaušovec, N., Jaušovec, K., 2012. Working memory training: improving intelligence—changing brain activity. Brain Cogn. 79, 96–106. http://dx.doi.org/10.1016/j.bandc.2012.02.007.

Jurcak, V., Tsuzuki, D., Dan, I., 2007. 10/20, 10/10, and 10/5 systems revisited: their validity as relative head-surface-based positioning systems. NeuroImage 34, 1600–1611. http://dx.doi.org/10.1016/j.neuroimage.2006.09.024.

Kirsh, D., 2000. A few thoughts on cognitive overload. Intellectica 30, 19–52.

Kirwan, B., Scaife, R., Kennedy, R., 2001. Investigating complexity factors in UK air traffic management. Hum. Factors Aerosp. Saf. 1, 125–144.

Klimesch, W., 1999. EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. Brain Res. Brain Res. Rev. 29, 169–195.

Klimesch, W., Doppelmayr, M., Pachinger, T., Ripper, B., 1997. Brain oscillations and human memory: EEG correlates in the upper alpha and theta band. Neurosci. Lett. 238, 9–12.

Kohlmorgen, J., Dornhege, G., Braun, M., Blankertz, B., Müller, K.R., Curio, G., Hagemann, K., Bruns, A., Schrauf, M., Kincses, W., 2007. Improving human performance in a real operating environment through real-time mental workload detection. Toward brain-computer interfacing, MIT Press.

Krusienski, D.J., Sellers, E.W., Cabestaing, F., Bayoudh, S., McFarland, D.J., Vaughan, T.M., Wolpaw, J.R., 2006. A comparison of classification techniques for the P300 Speller. J. Neural Eng. 3, 299–305. http://dx.doi.org/10.1088/1741-2560/3/4/007.

Liao, L.-D., Lin, C.-T., McDowell, K., Wickenden, A.E., Gramann, K., Jung, T.-P., Ko, L.-W., Chang, J.-Y., 2012. Biosensor technologies for augmented brain–computer interfaces in the next decades. Proc. IEEE 100, 1553–1566. http://dx.doi.org/10.1109/JPROC.2012.2184829.

Marcel, S., Millan, J.D.R., 2007. Person authentication using brainwaves (EEG) and maximum a posteriori model adaptation. IEEE Trans. Pattern Anal. Mach. Intell. 29, 743–752. http://dx.doi.org/10.1109/TPAMI.2007.1012.

Medvedev, A.V., Kainerstorfer, J., Borisov, S.V., Barbour, R.L., VanMeter, J., 2008. Event-related fast optical signal in a rapid object recognition task: improving detection by the independent component analysis. Brain Res. 1236, 145–158. http://dx.doi.org/10.1016/j.brainres.2008.07.122.

Mühl, C., Jeunet, C., Lotte, F., 2014. EEG-based workload estimation across affective contexts. Front. Neurosci. 8, 114. http://dx.doi.org/10.3389/fnins.2014.00114.

Müller, K.-R., Tangermann, M., Dornhege, G., Krauledat, M., Curio, G., Blankertz, B., 2008. Machine learning for real-time single-trial EEG-analysis: from brain-computer interfacing to mental state monitoring. J. Neurosci. Methods 167, 82–90. http://dx.doi.org/10.1016/j.jneumeth.2007.09.022.

Murkin, J.M., Arango, M., 2009. Near-infrared spectroscopy as an index of brain and tissue oxygenation. Br. J. Anaesth. 103 (Suppl. 1), i3–i13. http://dx.doi.org/10.1093/bja/aep299.

National Research Council (US) Committee on Military and Intelligence Methodology for Emergent Neruophysiological and Cognitive/Neural Research in the Next Two Decades, 2008. Emerging Areas of Cognitive Neuroscience and Neurotechnologies. National Academies Press, Washington, DC.

O'Donnell, R.D., Eggemeier, F.T., 1986. Workload Assessment Methodology. Handbook of Perception and Human Performance. John Wiley and Sons, Inc., New York, NY.

Onton, J., Delorme, A., Makeig, S., 2005. Frontal midline EEG dynamics during working memory. NeuroImage 27, 341–356. http://dx.doi.org/10.1016/j.neuroimage.2005.04.014.

Owen, A.M., McMillan, K.M., Laird, A.R., Bullmore, E., 2005. N-back working memory paradigm: a meta-analysis of normative functional neuroimaging studies. Hum. Brain Mapp. 25, 46–59. http://dx.doi.org/10.1002/hbm.20131.

Parasuraman, R., 2003. Neuroergonomics: research and practice. Theor. Issues Ergon. Sci. 4, 5–20. http://dx.doi.org/10.1080/14639220210199753.

Plichta, M.M., Herrmann, M.J., Baehne, C.G., Ehlis, A.-C., Richter, M.M., Pauli, P., Fallgatter, A.J., 2006. Event-related functional near-infrared spectroscopy (fNIRS): are the measurements reliable? NeuroImage 31, 116–124. http://dx.doi.org/10.1016/j.neuroimage.2005.12.008.

Pollock, V.E., Schneider, L.S., Lyness, S.A., 1991. Reliability of topographic quantitative EEG amplitude in healthy late-middle-aged and elderly subjects. Electroencephalogr. Clin. Neurophysiol. 79, 20–26.

Ramnani, N., Owen, A.M., 2004. Anterior prefrontal cortex: insights into function from anatomy and neuroimaging. Nat. Rev. Neurosci. 5, 184–194. http://dx.doi.org/10.1038/nrn1343.

Reason, J., 2000. Human error. West. J. Med. 172, 393–396.

Reid, G.B., Nygren, T.E., 1988. The subjective workload assessment technique: a scaling procedure for measuring mental workload. In: Hancock, P.A., Meshkati, N. (Eds.), Human Mental Workload. Advances in Psychology, Amsterdam, North-Holland, pp. 185–218.

Riccio, A., Holz, E.M., Aricò, P., Leotta, F., Aloise, F., Desideri, L., Rimondini, M., Kübler, A., Mattia, D., Cincotti, F., 2015. Hybrid P300-based brain-computer interface to improve usability for people with severe motor disability: electromyographic signals for error correction during a spelling task. Arch. Phys. Med. Rehabil. 96, S54–S61. http://dx.doi.org/10.1016/j.apmr.2014.05.029.

Rodgers, M.D., Drechsler, G.K., 1993. Conversion of the CTA, Inc., En Route Operations Concepts Database into a Formal Sentence Outline Job Task Taxonomy.

Rubio, S., Díaz, E., Martín, J., Puente, J.M., 2004. Evaluation of subjective mental workload: a comparison of SWAT, NASA-TLX, and workload profile methods. Appl. Psychol. 53, 61–86. http://dx.doi.org/10.1111/j.1464-0597.2004.00161.x.

Ryu, K., Myung, R., 2005. Evaluation of mental workload with a combined measure based on physiological indices during a dual task of tracking and mental arithmetic. Int. J. Ind. Ergon. 35, 991–1009. http://dx.doi.org/10.1016/j.ergon.2005.04.005.

Schettini, F., Riccio, A., Simione, L., Liberati, G., Caruso, M., Frasca, V., Calabrese, B., Mecella, M., Pizzimenti, A., Inghilleri, M., Mattia, D., Cincotti, F., 2015. Assistive device with conventional, alternative, and brain-computer interface inputs to enhance interaction with the environment for people with amyotrophic lateral sclerosis: a feasibility and usability study. Arch. Phys. Med. Rehabil. 96, S46–S53. http://dx.doi.org/10.1016/j.apmr.2014.05.027.

Schubert, R., Tangermann, M., 2008. Parieto-occipital alpha power indexes distraction during simulated car driving. Abstracts of the 14th World Congress of Psychophysiology. Int. J. Psychophysiol. 69. http://dx.doi.org/10.1016/j.ijpsycho.2008.05.033.

Shou, G., Ding, L., Dasari, D., 2012. Probing neural activations from continuous EEG in a real-world task: time-frequency independent component analysis. J. Neurosci. Methods 209, 22–34. http://dx.doi.org/10.1016/j.jneumeth.2012.05.022.

Smit, A.S., Eling, P.A., Hopman, M.T., Coenen, A.M., 2005. Mental and physical effort affect vigilance differently. Int. J. Psychophysiol. 57, 211–217.

Smith, M.E., Gevins, A., 2005. Neurophysiologic monitoring of mental workload and fatigue during operation of a flight simulator. Proc. SPIE 5797, 116–126. http://dx.doi.org/10.1117/12.602181.

Toppi, J., Borghini, G., Petti, M., He, E.J., Giusti, V.D., He, B., Astolfi, L., Babiloni, F., 2016. Investigating cooperative behavior in ecological settings: an EEG hyperscanning study. PLoS One 11, e0154236. http://dx.doi.org/10.1371/journal.pone.0154236.

Vecchiato, G., Di Flumeri, G., Maglione, A.G., Cherubino, P., Kong, W., Trettel, A., Babiloni, F., 2014. An electroencephalographic peak density function to detect memorization during the observation of TV commercials. Conf. Proc. IEEE Eng. Med. Biol. Soc. 2014, 6969–6972. http://dx.doi.org/10.1109/EMBC.2014.6945231.

Venables, L., Fairclough, S.H., 2009. The influence of performance feedback on goal-setting and mental effort regulation. Motiv. Emot. 33, 63–74. http://dx.doi.org/10.1007/s11031-008-9116-y.

Venthur, B., Blankertz, B., Gugler, M.F., Curio, G., 2010. Novel applications of BCI technology: Psychophysiological optimization of working conditions in industry. 2010 IEEE International Conference on Systems Man and Cybernetics (SMC). Presented at the 2010 IEEE International Conference on Systems Man and Cybernetics (SMC), 417–421. http://dx.doi.org/10.1109/ICSMC.2010.5641772.

Welke, S., Jurgensohn, T., Roetting, M., 2009. Single-trial detection of cognitive processes for increasing traffic safety. In: Proceedings of the 21st (Esv) International Technical Conference on the Enhanced Safety of Vehicles. Held June 2009 Stuttgart, Germany.

Wickens, C.D., 1984. Processing resources in attention. In: Parasuraman, R., Davies, R. (Eds.), Varieties of attention. Academic Press, Orlando, FL, pp. 62–102.

Wierwille, W.W., Eggemeier, F.T., 1993. Recommendations for mental workload measurement in a test and evaluation environment. Hum. Factors J. Hum. Factors Ergon. Soc. 35, 263–281. http://dx.doi.org/10.1177/001872089303500205.

Wolpaw, J., Wolpaw, E.W., 2012. Brain-Computer Interfaces: Principles and Practice. Oxford University Press.

Wolpaw, J.R., Birbaumer, N., McFarland, D.J., Pfurtscheller, G., Vaughan, T.M., 2002. Brain-computer interfaces for communication and control. Clin. Neurophysiol. 113, 767–791.

Wood, J.N., Grafman, J., 2003. Human prefrontal cortex: processing and representational perspectives. Nat. Rev. Neurosci. 4, 139–147. http://dx.doi.org/10.1038/nrn1033.

Zander, T.O., Jatzev, S., 2012. Context-aware brain-computer interfaces: exploring the information space of user, technical system and environment. J. Neural Eng. 9, 016003. http://dx.doi.org/10.1088/1741-2560/9/1/016003.

Zander, T.O., Kothe, C., Welke, S., Roetting, M., 2009. Utilizing secondary input from passive brain-computer interfaces for enhancing human-machine interaction. In: Schmorrow, D.D., Estabrooke, I.V., Grootjen, M. (Eds.), Foundations of Augmented Cognition. Neuroergonomics and Operational Neuroscience. Springer, Berlin, Heidelberg, pp. 759–771.

Zhang, J.H., Chung, T.D., Oldenburg, K., 1999. A simple statistical parameter for use in evaluation and validation of high throughput screening assays. J. Biomol. Screen. 4, 67–73.