SAPIENZA
UNIVERSITÀ DI ROMA
FACOLTÀ DI FARMACIA E MEDICINA

Dottorato di Ricerca in
MORFOGENESI e INGEGNERIA TISSUTALE

XXX Ciclo
(A.A. 2016/2017)

Electroencephalography-based measures of human mental workload in operational environments for the development of passive Brain-Computer Interfaces

Dottorando
Gianluca Di Flumeri

Tutor
Prof. Fabio Babiloni

Coordinator
Prof. Sergio Adamo

Co-Tutor
Ing. Pietro Aricò

ABSTRACT

Is it possible to enhance the performance of an operator by inferring his/her mental workload online from his brain activity? Also, would be possible to use such information to adapt the functionalities of the operative interface he/she is interacting with?

This is the experimental question that my PhD research activity tried to answer.

Even if not quantified, it is widely accepted the assumption that the human brain cognitive resources are limited. Depending on the amount of cognitive resources, i.e. the mental workload, committed to the main task, the human capacity to face additional unexpected events could dramatically decrease. It was demonstrated that humans could achieve their best performance only if maintaining their mental workload within an optimum range, otherwise they will be more prone to commit errors. Unfortunately, the human error is one of the main causes of accidents and no-natural catastrophes. Therefore, neuroscientific research is giving an important contribute in the development of Brain-Computer Interfaces able to recognize the user's mental state covertly (i.e. without interfering with his/her main activity) and to help him if needed.

In this context, my research activity aimed to develop a method able to evaluate, even online, the user's mental workload, on the basis of his/her brain activity measured by Electroencephalography, in operational environments, facing all those issues related to perform reliable neurophysiological measures outside the laboratory controlled conditions.

The developed method (patented) has been successfully validated in three different operational environments, i.e. the Air Traffic Management, the car driving and the robot-assisted surgery. Also, it has been applied online in a real application of Brain Computer Interface, where the operative platform changed its behaviour according on the EEG-based measure of the actual operator's mental workload level.

SOMMARIO

È possibile esaltare le prestazioni di un operatore stimando il suo carico di lavoro mentale dalla propria attività cerebrale, ed usare questa informazione per adattare le funzionalità dell'interfaccia che sta utilizzando?

Questa è la domanda sperimentale a cui la mia attività di ricerca del Dottorato ha provato a dare una risposta.

Anche se in maniera non quantificata, è ampiamente accettata l'idea che le risorse cognitive del cervello umano siano limitate. In funzione della quantità di risorse cognitive, in altre parole del carico di lavoro mentale, dedicate al compito principale, la capacità umana di affrontare ulteriori eventi inaspettati potrebbe diminuire drasticamente. È stato dimostrato che gli uomini possono raggiungere le proprie migliori prestazioni solo mantenendo il proprio carico di lavoro mentale all'interno di un intervallo ottimale, altrimenti aumenta la probabilità che commettano errori. Sfortunatamente, l'errore umano è una delle principali cause di incidenti e catastrofi non naturali. Per tale motivo, la ricerca neuroscientifica sta dando un importante contributo allo sviluppo di Interfacce Cervello-Computer in grado di riconoscere lo stato mentale dell'utente e di aiutarlo se necessario.

In tale contesto, la mia attività di ricercar aveva lo scopo di sviluppare un metodo in grado di valutare online il carico di lavoro mentale dell'utente, sulla base della sua attività cerebrale misurata attraverso Elettroencefalografia, in ambienti operativi, affrontando tutti quei problemi relativi all'eseguire misure neurofisiologiche affidabili al di fuori dei contesti controllati propri dei laboratori.

Il metodo sviluppato (brevettato) è stato con successo validato in tre differenti ambienti operative, ossia il controllo di traffico aereo, la guida di auto e la chirurgia assistita da robot. Inoltre, esso è stato applicato online in una reale applicazione di Interfaccia Cervello Computer, dove la piattaforma operative variava il suo livello di automazione sulla base del carico di lavoro mentale, misurato attraverso tecnica EEG, dell'operatore.

Index

1. INTRODUCTION

1.1 The human brain and its activity

The human brain is the central organ of the human Nervous System, and with the spinal cord makes up the *Central Nervous System* (CNS). In a generical assumption the CNS is not directly in contact either with the body peripheries or the external environment, but all the information is exchanged through the *Peripheral Nervous System* (PNS) (Nieuwenhuys, Voogd, & Huijzen, 2007).

The CNS is so named because it integrates information it receives from the external environment as well as from inside the human body, processes such information, and consequently coordinates and influences the activity of all the parts of the body, including movements and actions, decisions, thoughts, behaviours and organs functionalities. The CNS is contained within the dorsal body cavity, with the brain housed in the cranial cavity and the spinal cord in the spinal canal. In vertebrates, the brain is protected by the skull, while the spinal cord is protected by the vertebrae, both - brain and spinal cord - enclosed in the meninges.

The PNS consists of the nerves and ganglia outside the brain and spinal cord. The main function of the PNS is to connect the CNS to the limbs and organs, essentially serving as a relay between the brain and spinal cord and the rest of the body. Unlike the CNS, the PNS is not protected by the vertebral column and skull, or by the blood–brain barrier, which leaves it more exposed to mechanical injuries and external agents (Nieuwenhuys, Voogd, & Huijzen, 2007).

In Figure 1, a graphical overview on the human Nervous System, with its organization in CNS and PNS, is provided.

Within the following paragraph, briefly it will be put the focus on the brain, its activity from a physiological point of view, and the way to measure such activity. The information presented have been summarized from the book "*Mental states in aviation*" (Gianluca Borghini, Aricò, Di Flumeri, & Babiloni, 2017), whereof I have been co-author.

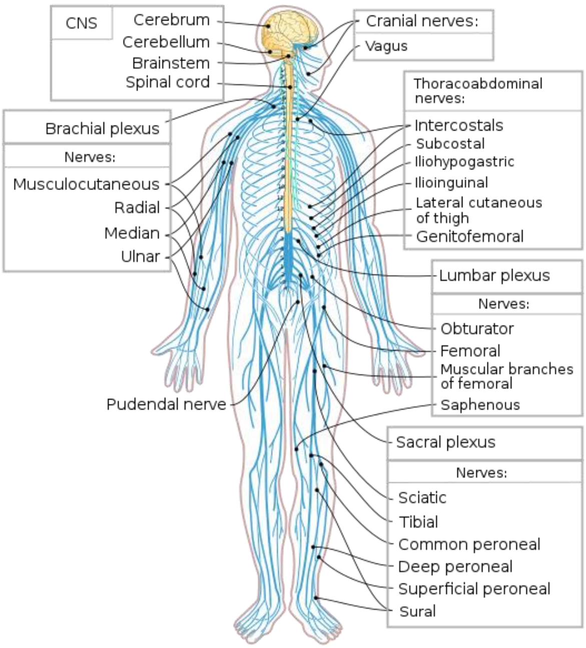


Figure 1. A representation of the human Nervous System. In yellow the Central Nervous System (CNS), in blue the Peripheral (PNS) one.

1.1.1. Brain anatomy

The adult human brain weighs on average about 1.2–1.4 kg, or more in general about 2 % of the total body weight, with a volume of around 1260 cm^3 in men and 1130 cm^3 in women, although there is substantial individual variation (Cosgrove, Mazure, & Staley, 2007).

The brain consists of the *cerebrum*, the *brainstem*, and the *cerebellum*. It controls most of the activities of the body, processing, integrating, and coordinating the information it receives from the sense organs, and making decisions as to the instructions sent to the rest of the body. The brain is contained in, and protected by, the skull bones of the head.

The *cerebrum* is the largest part of the human brain. It is divided into two cerebral hemispheres. The cerebral cortex is an external layer of grey matter, covering the core of white matter. The cortex is split into the neocortex and the much smaller allocortex. The neocortex is made up of six neuronal layers, while the allocortex has three or four. Each hemisphere is conventionally divided into five lobes – the frontal, central, temporal, parietal, and occipital lobes (Figure 2). The frontal lobe is associated with executive functions including self-control, planning, reasoning, and abstract thought, the central lobe is mainly dedicated to movements control and motor imagery, the temporal lobe to language comprehension, the parietal lobe to language processing and long-term memory, while the occipital lobe is dedicated to vision. Within each lobe, cortical areas are associated with specific functions, such as the sensory, motor and association regions. Although the left and right hemispheres are broadly similar in shape and function, some functions are associated with one side, such as language in the left and visual-spatial ability in the right. The hemispheres are connected by nerve tracts, the largest being the corpus callosum (Nieuwenhuys et al., 2007).

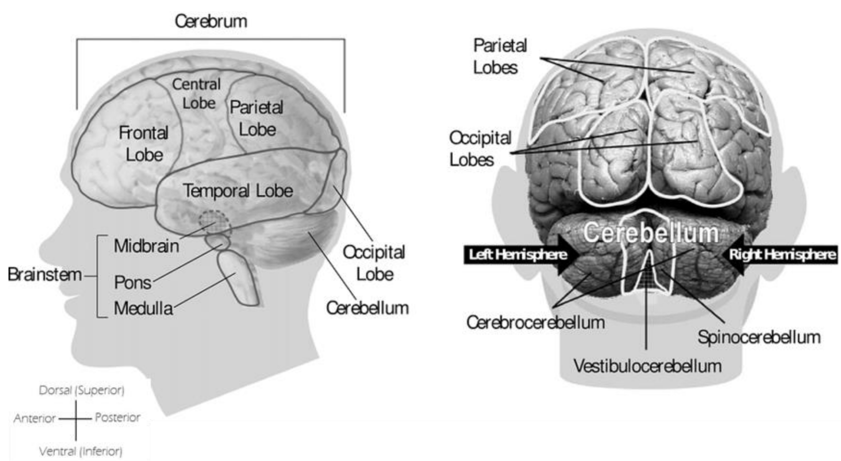


Figure 2. Brain cortex subdivision in lobes and hemispheres.

The cerebrum is connected by the *brainstem* to the spinal cord. The brainstem consists of the midbrain, the pons, and the medulla oblongata. The *cerebellum* is connected to the brainstem by pairs of tracts. Within the cerebrum there is the ventricular system, consisting of four interconnected ventricles in which cerebrospinal fluid is produced and circulated. Underneath the cerebral cortex are several important structures, including the thalamus, the epithalamus, the pineal gland, the hypothalamus, the pituitary gland, and the subthalamus; the limbic structures, including the amygdala and the hippocampus; the claustrum, the various nuclei of the basal ganglia; the basal forebrain structures, and the three circumventricular organs (Figure 3). The cells of the brain include neurons and supportive glial cells. There are more than 86 billion neurons in the brain, and a more or less equal number of other cells. Brain activity is made possible by the interconnections of neurons and their release of neurotransmitters in response to nerve impulses. Neurons form elaborate neural networks of neural pathways and circuits. The whole circuitry is driven by the process of neurotransmission.

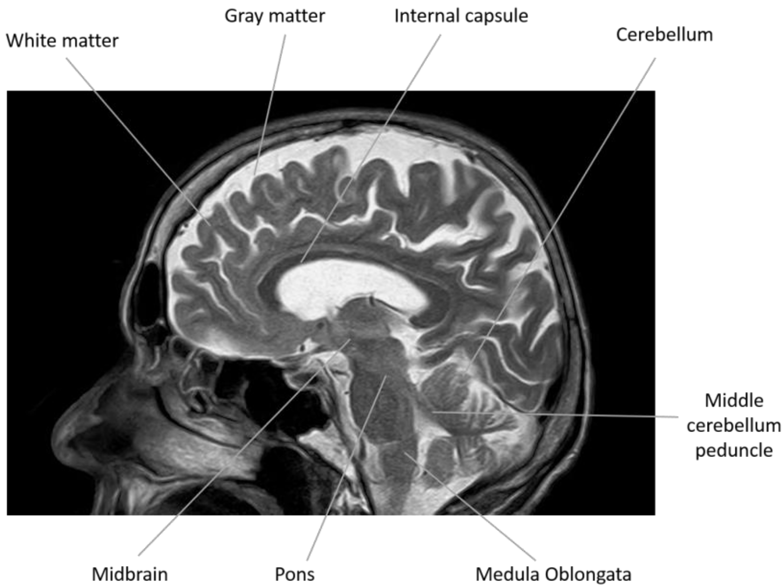


Figure 3. A sagittal brain section in which some of the main structures of the brains are labelled. In particular: gray matter and white matter, cerebellum, midbrain, medulla oblongata, pons, middle cerebellar peduncle and internal capsule.

About 100 billion (1×10^{11}) of *neurons*, as well as many more glial cells, are integrated into the structural and functional structures of the brain (Noback, Strominger, Demarest, & Ruggiero, 2005). They exhibit a wide diversity of shapes and sizes. The neuron is the basic unit of the nervous system and is composed of four regions, structurally defined as follows: *a cell body* (soma) that emits a single nerve process called an *axon,* which ends with the *presynaptic* *terminals,* and a variable number of branching processes called *dendrites* (Figure 4). Each axon, including its collateral branches, usually terminates as an arbor of fine fibres; each fibre ends as an enlargement called *bouton,* which is part of a synaptic junction. At the other end of the neuron, there is a three-dimensional *dendritic field,* formed by the branching of the dendrites. The cell body is the genomic and metabolic centre of the neuron. Dendrites are the main recipients of the neural signals for communication between neurons, and contain critical processing complexes. The axon is the canal apt to conduct messages (action potentials) to the presynaptic terminals, where each neuron is in synaptic contact with other neurons and, thus, is part of the network that constitutes the nervous system. A neuron is designed to react to the stimuli, to rapidly transmit the resulting excitation to the other portions of the nerve cell, and to influence other neurons, muscle cells, and glandular cells. Neurons are so specialized that most are incapable of reproducing themselves, and they lose viability if denied an oxygen supply for more than a few minutes. Dendrites contain the same cytoplasmic organelles (e.g., Nissl bodies and mitochondria) as the cell body of which they are true extensions. The axon is specialized in the transmission of coded information as all-or-none action potentials. The axon arises from the *axon hillock* of the cell body, at a site called the *initial segment,* and extends for a distance of less than 1 mm to as much as 1 m before arborizing into *terminal branches* (Figure 4). The axon hillock, initial segment, and the axon lack Nissl bodies. The branches of an axon can have, potentially, two types of *bouton*. Each branch ends as a *terminal bouton* that forms a synapse with the dendrite, cell body, or axon of another neuron. In addition, along some branches, there are thickenings called *boutons en passage,* which form synapses with another neuron or smooth muscle fibre. The dendrites of many neurons are studded with tiny protuberances called *spines* (e.g., pyramidal neurons of the cerebral cortex): these dendritic spines increase the surface area of the membrane of the receptive segment of the neuron. Located on them, there are over 90% of all the excitatory synapses present in the central nervous system (CNS). Because of their widespread occurrence on neurons of the cortical areas of the cerebrum, they are thought to be involved in learning and memory.

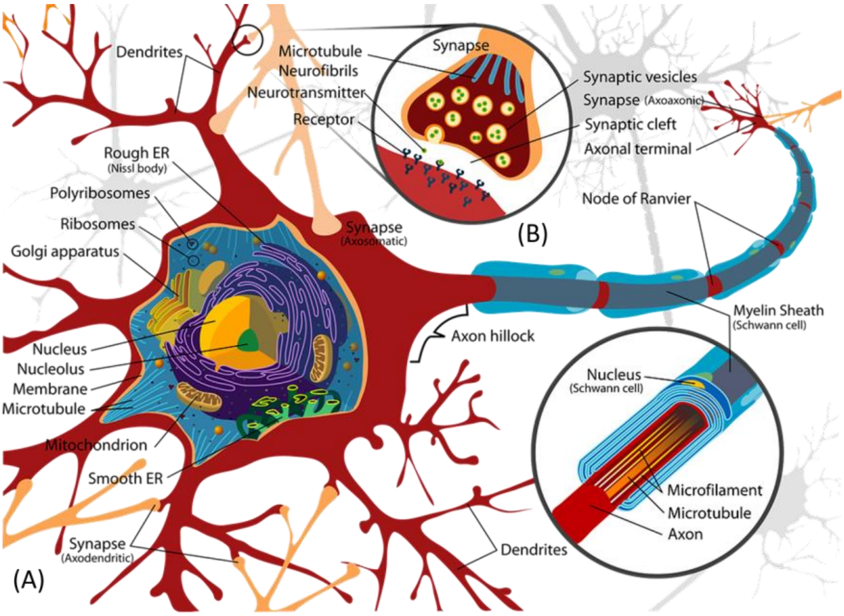


Figure 4. (A) Diagram of a neuron located wholly within the central nervous system. The (B) represents a synapse between two neurons. The myelin sheath of neuron (A) is entirely the product of a glial cell, and the one of neuron is produced by a glial cell inside the central nervous system and by a Schwann (neurolemma) cell in the peripheral nervous system.

The synapse is the contact point between the neurons. A sub-microscopic space, the synaptic cleft, which is about 200 Å, exists between the bouton of one neuron and the cell body of another neuron *(axosomatic synapse)*, between a bouton and a dendrite *(axodendritic synapse)*, and between a bouton and an axon *(axoaxonic synapse)*. In addition, also *dendrodendritic synapses* (between two dendrites). The axon of one neuron might terminate in just a few synapses or up to many thousands of synapses. The dendrite–cell body complex might receive synaptic contacts from many different neurons (up to over 15,000 synapses). The termination of a nerve fibre in a muscle cell (neuromuscular junction) or a glandular cell (neuroglandular junction) is basically similar to the synapse between two neurons.

The synapse of each axon terminal of a motoneuron on a voluntary muscle cell is called a *motor end plate* (Figure 4). The cell membrane of the axon at the synapse is the *presynaptic membrane*, and the cell membrane of dendrite-cell body complex, muscle, or glandular cell is the *postsynaptic membrane*. The *subsynaptic membrane* is that region of the postsynaptic membrane that is juxtaposed against the presynaptic membrane at the synapse. A concentration of mitochondria and *presynaptic vesicles* is present in the cytoplasm of the bouton; none is present in the cytoplasm adjacent to the subsynaptic membrane. Most neurons contain at least two distinct types of vesicle: small vesicles 50 nm in diameter and large vesicles from 70 to 200 nm in diameter.

1.1.2. Basic neurophysiology

Every neuron is said to possess "in miniature, the integrative capacity of the entire nervous system." Neurons can transform information and transmit it to other neurons. In most, the dendrite - cell body unit is specialized as a receptor and integrator of synaptic input from other neurons, and the axon is specialized to convey coded information from the dendrite-cell body unit to the synaptic junctions, where transformation functions take place with other neurons or effectors (muscles and glands). To serve these tasks, the neuron is thus organized into a receptive segment (dendrites and cell body), a conductile segment (axon), and an effector segment (synapse) (Figure 5). Neurons are specialized to generate electrical signals, which are used to encode and convey information. These signals are expressed by alterations in the resting membrane potential. Voltage changes that are restricted to at or near the sites where neurons are stimulated are called *graded potentials.* These can lead to the production of *action potentials (nerve impulses or spikes),* which transmit information for substantial distances along an axon.

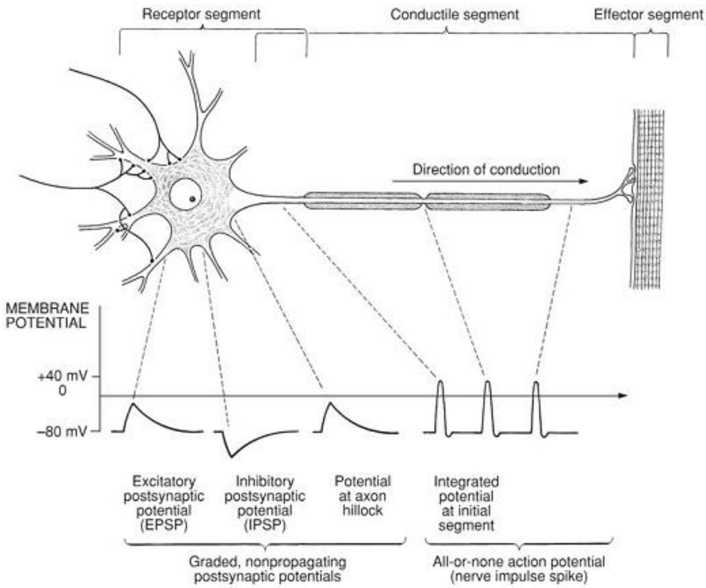


Figure 5. On the surface of the dendrites and cell body are excitatory and inhibitory synapses, which, when stimulated, produce local, graded, non-propagating potentials. These are exhibited as an excitatory or depolarizing postsynaptic potential (EPSP) and as an inhibitory or hyperpolarizing postsynaptic potential (IPSP). These local potentials are summated at the axon hillock and, if adequate, could trigger an integrated potential at the initial segment and an "all-or-none" action potential, which is conducted along the axon to the motor end plate.

Two forms of graded potential are *generator (receptor potentials)* and *synaptic potentials*. Generator potentials are evoked by sensory stimuli from the environment (both inside and outside the body). Information that passes from one neuron to another at synapses produces *synaptic potentials* in the postsynaptic neuron. The activity of either generator or synaptic potentials can elicit action potentials, which, in turn, produce synaptic potentials in the next neuron. Synaptic potentials elicited in effectors (skeletal muscle and glands) at synapses can result in the contraction of the muscle or emission of secretory product from a gland.

Excitability is a property that enables a neuron to respond to a stimulus and to transmit information in the form of electrical signals. The flow of information within a neuron and between neurons is conveyed by both electrical and chemical signals. The electrical signals, called *graded potentials* and *action potentials*, are all produced by temporary changes in the current flow into and out of the neuron. These changes are deviations away from the normal value of the resting membrane potential. Ion channels within the plasma membrane control the inward and outward current flow. The channels possess three features. They (1) conduct ions across the plasma membrane at rapid rates up to 100,000,000 ions per second, (2) can recognize specific ions and be selective as to which can pass through, and (3) selectively open and close in response to specific electrical, chemical, and mechanical stimuli. Each neuron is presumed to have over 20 different types of channel with thousands of copies of each channel. The flux (movement of ions) through the ion channels is passive, requiring no expenditure of metabolic energy. The direction of the flux is determined by the electrochemical driving force across the plasma membrane. The primary role of ion channels in neurons is to mediate rapid signalling. These channels, called *gated channels*, have a molecular "cap" or *gate*, which opens briefly to permit anion species to pass (Figure 6). Gated channels open when a neurotransmitter binds to them; *voltage-gated channels* open and close in response to changes in membrane potential; *modality-gated channels* are activated by specific modalities (e.g., touch, pressure, or stretch). Gating is the process by which a channel is opened or closed during activity. Each channel consists of several plasma membrane-spanning polypeptide subunits (proteins) arranged around a central pore.

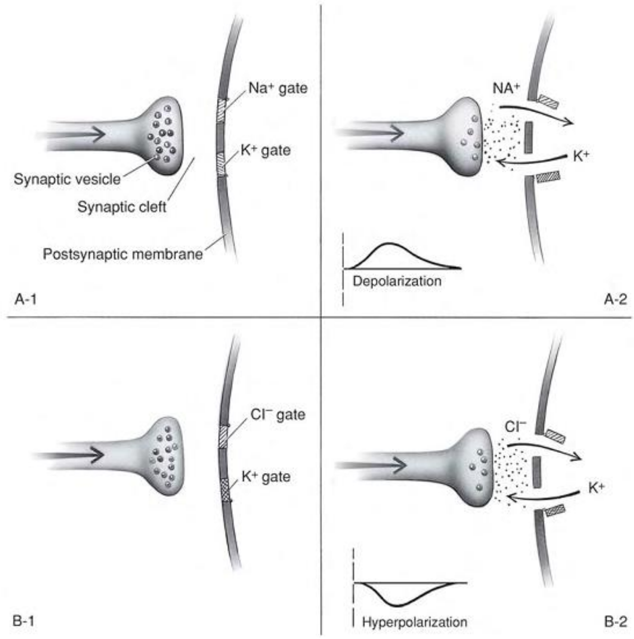


Figure 6. Excitatory synapses (A) and inhibitory synapses (B). *A-1* and *B-1*: Synapses prior to release of neurotransmitter. *A-2*: Excitatory postsynaptic response (EPSP) following release of neurotransmitter with Na+ ion inrush through Na+ gate and K+ ion outrush through K+ gate. *B-2*: Inhibitory postsynaptic response (IPSP) following release of neurotransmitter with Cl– ion in rush through Cl– gate and K+ ion outrush through K+ gate.

Each of these classes of channel belongs to a different gene family. Each member of a family shares common structural and biochemical features, which presumably have evolved from a common ancestral gene of that family. The channels of the voltage-gated gene family are selective for Na+, K+, and Ca2+ ions. The channels for the transmitter-gated channels respond to acetylcholine, gamma amino butyric acid (GABA), and glycine. Most gated-channels are closed with the membrane at rest. They open when activated following the binding of a ligand (ligand gating), a change in the membrane potential (voltage gating), or the stretch of the membrane (modality gating). In the transmitter-gated channel, the transmitter binds to a specific site on the external face of a channel that activates it to open briefly. The energy to open the channels is derived (1) from the binding of the transmitter to the receptor protein in the ligand-gated channels, (2) from the changes in the membrane voltage in the voltage-gated channels, and (3) presumably from the mechanical forces resulting from cytoskeletal interaction at the modality-gated channels. There are two types of membrane response; it could (1) *hyperpolarize* or (2) *depolarize*. During *hyperpolarization*, the membrane becomes more negative on the inside with respect to its outside (i.e., could go from –70 mV to –80 mV). During *depolarization*, the membrane becomes less negative inside with respect to its outside and even might reverse polarity with its inside becoming positive with respect to the outside. This is still called *depolarization* because the membrane potential becomes less negative than the resting potential (e.g., from –70 mV to 0 to +40 mV).

1.1.3. The electroencephalographic signal

The *Electroencephalogram* (EEG) is the electrical biosignal recordable over the scalp as the result of the summation of synchronously postsynaptic potentials. The contribution to the electric field of neurons acting synchronously is approximately proportional to their number, and, for those firing non-synchronously, as a square root of their number (Blinowska & Durka, 2006). The problem of the origins of EEG rhythmical activity has been approached by electrophysiological studies on brain nerve cells and by the modeling of electrical activity of the neural populations. The question emerges whether the rhythms are caused by single cells with pacemaker properties or by the oscillating neural networks. It has been shown that some thalamic neurons display oscillatory behaviour, even in the absence of synaptic input (Jahnsen & Llinás, 1984). Evidence exists that the intrinsic oscillatory properties of some neurons contribute to the shaping of the rhythmic behaviour of networks to which they belong. However, these properties may not be sufficient to account for the network rhythmic behaviour. It is generally accepted that cooperative properties of networks consisting of excitatory and inhibitory neurons connected by feedback loops play the crucial role in establishing EEG rhythms. The frequency of oscillation depends on the intrinsic membrane properties, on the membrane potential of the individual neurons, and on the strength of the synaptic interactions. Bursts of oscillatory activity may constitute a mechanism by which the brain can regulate changes of state in selected neuronal networks and change the route of information (Lopes da Silva, 1991). EEG is usually registered by means of electrodes placed on the brain scalp. They can be secured by an adhesive (like *collodion*) or embedded in a special snug cap. The resistance of the connection should be less than 10 (kΩ), so the recording site is first cleaned with diluted alcohol, and conductive electrode paste applied to the electrode cup. Knowledge of exact positions of electrodes is very important for both interpretation of a single recording as well as comparison of results, hence the need for standardization. The traditional *10–20 International System* (Jasper, 1958) states positions of 19 EEG electrodes (and two electrodes placed on earlobes A1/A2) related to specific anatomic landmarks, such that 10 – 20 % of the distance between them is used as the electrode interval. The first part of derivation's name indexes the array's row—from the front of head: Fp, F, C, P, and O. Such labels are related to the cortial area where the electrode is placed, respectively prefronta, Frontal Central, Parietal and Occipital. The second part is formed from numbers even on the left and odd on the right side, while in the centre ''z''. Progress in topographic representation of EEG recordings brought demand for a larger amount of derivations. Electrode sites halfway between those defined by the standard 10–20 International System were introduced in the extended 10–20 system. EEG is a measure of potential difference: in the referential (or unipolar) setup, it is measured relative to the same electrode for all derivations. This reference electrode is usually placed on the earlobe, nose, mastoid, chin, neck, or scalp centre. No universal consent exists regarding the best position of the reference electrode, because currents coming from bioelectric activity of muscles, heart, or brain propagate all over the human body. In the bipolar setup (montage), each channel registers the potential difference between two particular scalp electrodes. Data recorded in a referential setup can be transformed into any bipolar montage. The common ''average reference'' montage can be obtained by subtracting from each channel the average activity from all the remaining derivations. Contrary to the open question of the reference, the necessity of artefact rejections is universally acknowledged. *Artefacts* are recorded signals that are non-cerebral in origin. They may be divided into one of two categories depending on their origin: *physiological artefacts* or *non-physiological artefacts*. Physiological artefacts can stem from muscle or heart activity (EMG, ECG), eye movement (EOG), external electromagnetic field, poor electrode contact, subject's movement. Corresponding signals (EMG, EOG, ECG, and body movements) registered simultaneously with EEG could be helpful in the visual rejection of artefact-contaminated epochs. Non-physiological artefacts arise from two main sources: external electrical interference (power lines or electrical equipment), and internal electrical malfunctioning of the recording system (electrodes, cables, amplifier).

Furthermore, artefacts may reduce the performance of machine-learning techniques. Several ways of handling physiological artefacts can be found in the literature. Artefacts may be avoided, rejected or removed from the EEG dataset. Artefacts avoidance involves asking users to avoid blinking or moving their body during the experiments (Vigario, Sarela, Jousmiki, Hamalainen, & Oja, 2000). This approach is very simple, because it does not require any computation, as brain signals are assumed to have no artefacts. However, this assumption is not feasible in real environments, i.e. outside the laboratory, since some artefacts, eye and body movements, are not easily avoidable. Artefact rejection approaches suggest discarding the epochs contaminated by the artefacts. Manual artefact rejection is an option to remove artefacts in brain signals, and experts could identify and eliminate all artefact-contaminated EEG epochs. The main disadvantage in using manual rejection is that it requires intensive human work, so this approach is not suitable for real-time evaluations. This issue has been directly treated during my PhD research activity (please see Par. 4.1.2).

In the EEG, the following frequency rhythms are considered characteristic for its analysis (Figure 7): delta (0.5 – 4 Hz), theta (4 – 8 Hz), alpha (8 – 12 Hz), beta (12 – 30 Hz), and gamma (above 30 Hz).

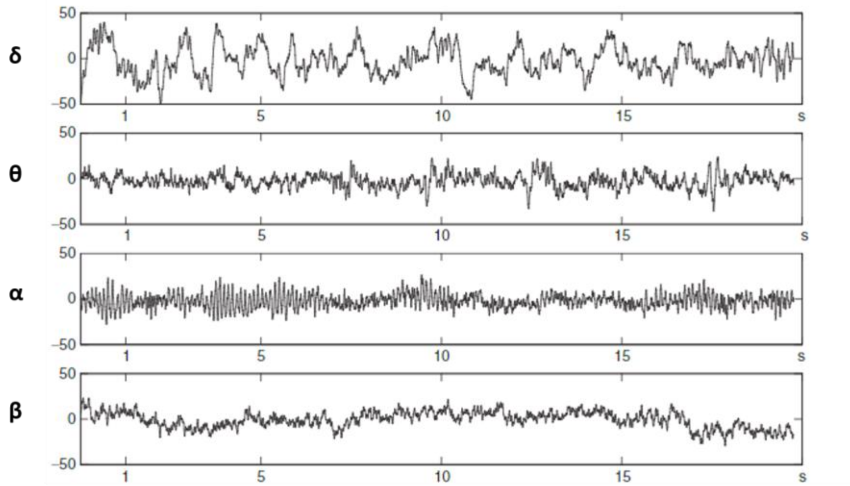


Figure 7. Characteristic EEG rhythms, from the top: δ (0.5 – 4 Hz), θ (4 – 8 Hz), α (8 – 12 Hz), β (12 –30 Hz). The gamma band could reach 100 Hz.

Delta activity is characterized by high amplitude and low frequency. It is usually associated with the slow-wave in psychophysiology of sleep. It is suggested that it represents the onset of deep sleep phases in healthy adults (Hori et al., 2001).

Theta rhythm is generally linked to the hippocampus activity (Buzsáki, 2002) as well as neocortex (Cantero et al., 2003). It is thought to be linked to deep relaxation or meditation, and it has been observed during the transition between wake and sleep (Hagemann, 2008). However, theta rhythms are suggested to be important for learning and memory functions (Sammer et al., 2007), encoding and retrieval (Vecchiato et al., 2014; Ward, 2003), which involve high concentration (Hagemann, 2008). It has also been suggested that theta oscillations are associated with the attentional control mechanism in the anterior cingulated cortex (Smith, Gevins, Brown, Karnik, & Du, 2001), and it is often shown to increase with a higher cognitive task demand (Gundel & Wilson, 1992).

Alpha activity has found in the visual cortex (occipital lobe) during periods of relaxation or idling (eyes closed but awake). In the continuous EEG, alpha band is characterized by high amplitude and regular oscillations, in particular over parietal and occipital areas.

High alpha power has been assumed to reflect a state of relaxation or cortical idling; however, when the operator assigns more effort to the task, different regions of the cortex may be recruited in the transient function network leading to passive oscillation of the local alpha generators, in synchrony with a reduction in alpha power (Smith et al., 2001). Recent results have suggested that alpha is involved in auditory attention processes and the inhibition of task irrelevant areas to enhance signal-to-noise ratio (Gevins et al., 1998; Wolfgang Klimesch, 2012). Additionally, alpha activity may be further divided into sub-bands by means of the frequency corresponding to the alpha peak of the user (Wolfgang Klimesch, 1999), called *Individual Alpha Frequency* (IAF). For instance, *alpha 3* (IAF÷IAF+2 Hz) reflects semantic memory performance, while *alpha 1* and *alpha 2* (respectively, IAF-4÷IAF-2 and IAF-2÷IAF Hz) reflect general task demands and attentional processes.

Beta activity is predominant in wakefulness state, especially in frontal and central areas of the brain. High power in beta band is associated with the increased mental arousal and activity. Dooley (Dooley, 2009) pointed out that beta wave represents cognitive consciousness and active, busy, or anxious thinking. Furthermore, it has been revealed to reflect visual concentration and the orienting of attention. This band can be further divided into *low beta wave* (12.5-15 Hz), *middle beta wave* (15-18 Hz), *high beta wave* (> 18 Hz). Low waves seems to be associated with inhibition of phasic movements during sleep, and high waves with dopaminergic system (Hagemann, 2008).

Finally, *Gamma rhythms* are the fastest activity in EEG and it is thought to be infrequent during waking states of consciousness (Dooley, 2009). Recent studies reveal that it is linked with many cognitive functions, such as attention, learning, and memory (Jensen, Kaiser, & Lachaux, 2007). Gamma components are difficult to record by scalp electrodes, because of their low amplitude, but with *Electrocorticography* (ECoG) components up to 100 Hz, or even higher, may be registered.

The contribution of different rhythms to the EEG depends on the age, psyco-cognitive state of the subject, and level of alertness. Considerable inter-subject differences in EEG characteristics also exist, since EEG pattern is influenced by neuropathological conditions, metabolic disorders, and drug action (Niedermeyer & Silva, 2005).

However, despite the remarkable progresses made in understanding "*how the brain works*", the brain is still not fully understood, and research is ongoing (Van Essen et al., 2012). Neuroscientists, along with researchers from other disciplines, study literally how the human brain works. The boundaries between the specialties of neuroscience, neurology and other disciplines such as psychiatry have faded as they are all influenced by basic research in neuroscience.

Neuroscience research has expanded considerably in recent decades. The "*Decade of the Brain*", an initiative of the United States Government in the 1990s, is considered to have marked much of this increase in research (Jones & Mendell, 1999), and was followed in 2013 by the *BRAIN Initiative* (UnderwoodJun, 2014). The *Human Connectome Project* was a five-year study launched in 2009 to analyse the anatomical and functional connections of parts of the brain, and has also provided much data and contributions in general in this field (Van Essen et al., 2012).

In this context, the aim of my PhD research activity has been the interpretation of the brain activity, in terms of EEG signal, with respect to a specific mental state: the *Mental Workload*.

1.2 The Mental Workload

The *Mental Workload* is a complex construct that is assumed to be reflective of an individual's level of cognitive engagement and mental effort (Wickens, 1984). Measurement of mental workload essentially represents the quantification of mental activity resulting from performing one or more tasks. However, it is difficult to give a unique definition of *Mental Workload*. Various definitions have been given during the last decades, for example:

- "Mental workload refers to the portion of operator information processing capacity or resources that is actually required to meet system demands" (Eggemeier, Wilson, Kramer, & Damos, 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, behaviours, and perceptions of the operator" (Hart & 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 & Donchin, 1986);
- "The reasons 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 how the mental workload may not be a unitary concept because it is the result of different aspects interacting with each other. In fact, several mental processes such as alertness, vigilance, mental effort, attention, stress, mental fatigue, drowsiness and so on, can be involved in the meanwhile of a task execution and they could be affected by specific tasks demand in each moment. In general, mental workload theory assumes that: (i) people have a limited cognitive and attentional capacity, (ii) different tasks will require different amounts (and perhaps different types) of processing resources, and (iii) two individuals might be able to perform a given task equally well, but differently in terms of brain activations (Baldwin, 2003; Wickens, Hollands, Banbury, & Parasuraman, 2012).

1.2.1. A topic in the Human Factors research

In some operational environments, the safety of the people could weight on one or few operators. In such contexts, a human error could have serious and dramatic consequences. For example, in the transports domain the safety of the passengers depends on the performance of the Pilot(s), of the (e.g. Air, Train, Vessel) Traffic-Controller(s) or of the Driver(s).

In general, human error has consistently been identified as one of the main factors in a high proportion of all workplaces accidents. In particular, it has been estimated that up to 90% of accidents exhibits human errors as principal cause(Feyer & Williamson, 2011). For example, in the health care domain, the US Institute of Medicine estimates that there is a high people mortality per year (between 44.000 and 88.000) as a result of medical errors (Helmreich, 2000). It has also been estimated that inappropriate human actions and consequently errors are implicated are the main cause of the 57 % of road accidents and a contributing factor in over 90 % (Aberg & Rimmö, 1998). The Aviation Safety Network reported 19 accidents with 960 casualties during the last years; in many cases factors related to workload, situation awareness and monitoring were a cause or contributing factor (P. Arico et al., 2017; Gianluca Borghini et al., 2017). Additionally, over the past four decades, human error has been involved in a high number of casualty catastrophes, including the Three Mile Island, Chernobyl and Bhopal nuclear power disasters, the Tenerife, Mont St Odile, Papa India and Kegworth air disasters, the Herald of Free Enterprise ferry disaster, the Kings Cross fire disaster, the Lad-broke Grove rail disaster, and many others (Salmon, Regan, & Johnston, 2005). Consequently, the construct (i.e. human error) has received more and more attention, and it has been investigated across a wide range of domains, including military and civil aviation (Shappel & Wiegmann, 2000; Stanton et al., 2006), aviation maintenance (Rankin, Hibit, Allen, & Sargent, 2000), air traffic management (Shorrock & Kirwan, 2002), rail (Lawton & Ward, 2005), road transport (Reason, 2000a; Rumar, 1990), nuclear power and petrochemical reprocessing (Kirwan, 1998), military, medicine (Helmreich, 2000; Sexton, Thomas, & Helmreich, 2000), and even the space travel domain (Nelson, Haney, Ostrom, & Richards, 1998).

Therefore, what are the causes of human errors? Human error is an extremely common phenomenon: people, regardless of abilities, skill level and expertise, makes errors every day. The typical consequence of error-occurrence is the failure to achieve a desired outcome or the production of an undesirable outcome. When it happens in particular working environments, such error can potentially lead to accidents involving injury and fatalities. Human error can be defined as the execution of an incorrect or inappropriate action, or a failure to perform a particular action. According to the scientific literature, there have been numerous attempts at de-fining and classifying the human error. However, a universally accepted definition does not yet exist. Rasmussen (Jens Rasmussen, 1982) pointed out the difficulty in providing a satisfactory definition of human error; in 1987, he suggested that human error represents a mismatch between the demands of an operational system and what the operator does (J. Rasmussen, 1987). The main causes of human errors have to be searched within the internal or psychological factors of the operator (Reason, 2000b). 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.

In particular, cognitive psychology literature demonstrated that the Mental Workload has an ‘‘inverted U-shape’’ relationship with performance (Figure 8). 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. Nevertheless, a period of mental inactivity and “under-stimulation” can cause a monotonous and boring state (underload), a low level of vigilance and attention, with low cognitive resources demand. For example, Warm and colleagues (Warm, Parasuraman, & Matthews, 2008) showed how vigilance requires an important amount of cognitive resources, by using behavioural, neurophysiological and subjective measures. On the contrary, an operative condition characterized by demanding multi-tasks can lead the user to an overload condition (Kirsh, 2000). Both the cases bring to a variation in neurophysiological factors and often to a decrement of performance. Such performance reduction is highly undesired, especially in critical domains, as discussed above.

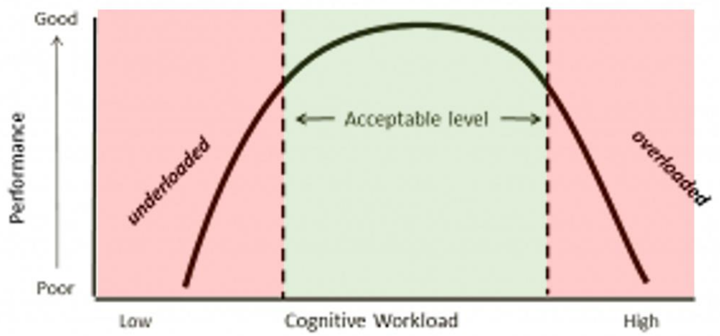


Figure 8. Inverted U-shape relationship between mental workload and performance.

Mental Workload assessment techniques must be sensitive to such fluctuations in task demands without intruding on primary task performance (O'Donnell & Eggemeier, 1986). This level of sensitivity is unobtainable with behavioural and subjective measures alone. In this regard, neurophysiological techniques have been demonstrated to be able to assess mental workload of humans with a high reliability, even in operational environments (Aricò, Borghini, Di Flumeri, Colosimo, Pozzi, et al., 2016; Gianluca Borghini et al., 2016; Di Flumeri et al., 2015; Mühl, Jeunet, & Lotte, 2014; Wierwille & Eggemeier, 1993). Moreover, neurophysiological techniques afford another important advantage: unlike alternative subjective assessment techniques, neurophysiological measures do not require the imposition of an additional task either concurrently (as in secondary task techniques) or subsequently (as in subjective workload assessment techniques) the primary one.

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 his/her 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 (adaptive automations) in order to help the user to keep optimal levels of mental workload under which operating (Prinzel, Freeman, Scerbo, Mikulka, & Pope, 2000). In other words, the neurophysiological workload assessment could be used to realize a *passive Brain Computer Interface* (passive-BCI, please see Par. 1.3) application in real environments.

1.2.2. The Mental workload evaluation in operational environments

In the mental workload related literature, neurophysiological measurements have been and are often used to evaluate the level of cognitive demand induced by a task (P. Arico et al., 2017; Boucsein & Backs, 2000; Desmond & Hancock, 2001).

Many neurophysiological measures have been used for the mental workload assessment, including *Electroencephalography* (EEG), *functional Near-InfraRed* (fNIR) imaging, *functional Magnetic Resonance Imaging* (fMRI), *Magnetoencephalography* (MEG), and other types of biosignals such as *Electrocardiography* (ECG), *Electrooculography* (EOG) and *Galvanic Skin Response* (GSR) (Gianluca Borghini, Astolfi, Vecchiato, Mattia, & Babiloni, 2014; Ramnani & Owen, 2004; Wood & Grafman, 2003). Among all these methods, Aricò and colleagues (Aricò, Borghini, Di Flumeri, Colosimo, Pozzi, et al., 2016) have highlighted the clear advantages of using EEG signal in passive-BCI application, also proposing a cutting-edge algorithm for the online operator's mental workload assessment.

Basically, most part of the EEG research showed that the brain electrical activities mainly considered for the mental workload analysis are the Theta and Alpha rhythms typically gathered from the *Pre-Frontal Cortex* (PFC) and the *Posterior Parietal Cortex* (PPC) regions. In this regard, previous studies demonstrated as the EEG Theta rhythm over the PFC present a positive correlation with the mental workload (Gevins, A & Smith, 2003; Smit, Eling, & Coenen, 2004). Moreover, published literature stressed the inverse correlation between the EEG power in the alpha frequency band over the PPC and the mental workload (G. Borghini, Astolfi, Vecchiato, Mattia, & Babiloni, 2014; Brookings, Wilson, & Swain, 1996; Gevins, Smith, McEvoy, & Yu, 1997; Jaušovec & Jaušovec, 2012; W. Klimesch, Doppelmayr, Pachinger, & Ripper, 1997; Venables & Fairclough, 2009). Only few studies have reported significant results about the modulation of the EEG power in other frequency bands, i.e. the delta, beta and gamma. Therefore, the most accepted evidences about EEG correlates of mental workload could be resumed in an increase of the theta band spectral power, especially on the frontal cortex, and a decrease in alpha band over the parietal cortex, in correspondence of mental workload increasing (G. Borghini et al., 2014; Gianluca Borghini et al., 2017) (Figure 9).

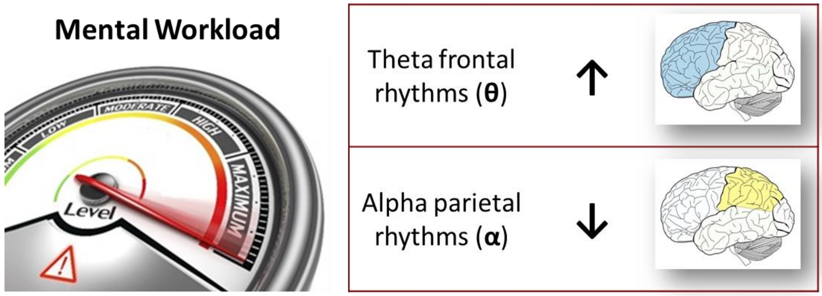


Figure 9. Schematic summary of the main EEG features variations related to the mental workload increasing.

Several studies, in particular in the aviation domain, have highlighted the high reliability of EEG-based mental workload indexes (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. More recently, Shou and colleagues (Shou, Ding, & Dasari, 2012) evaluated the mental workload during an 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 (Pietro Arico et al., 2014; 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, i.e. 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. However, the algorithms proposed were affected by some weaknesses, such as parameters manual settings and performance decreasing over time, that limited their employment in real operational environments (for further information please see Par. 4.1.1).

Moreover, other studies about the mental workload estimation by using neurophysiological measurements, have been performed with other types of transport means, in particular cars (G. Borghini et al., 2014; Göhring, Latotzky, Wang, & Rojas, 2013; Kohlmorgen et al., 2007), and in the military domain (Dorneich, Ververs, Mathan, & Whitlow, 2005).

1.3 Passive Brain-Computer Interfaces and Automation

In this regard, the *Neuroergonomics* research field aims at developing systems that take such limitations of a human's capacity to process information into account and avoid performance degradation by adapting the user's interface to reduce the task demand/complexity or by intervening directly on the system (Fuchs, Hale, Stanney, Juhnke, & Schmorrow, 2007). Over the past two decades, researchers in the field of augmented cognition worked to develop novel technologies that can both monitor and enhance human cognition and performance. Most of this augmented cognition research was based on research findings coming from cognitive science and cognitive neuroscience. On the basis of such findings and the technological improvements, that have allowed to measure human biosignals in a more reliable and no-invasive way, it has been possible to evaluate the actual operator mental states by using neurophysiological indexes, and to use them as input toward the interface the operator is interacting. Such kind of application is called *passive Brain-Computer Interface* (passive-BCI).

In its classical assumption, a Brain-Computer Interface (BCI) is a communication system in which messages or commands that an individual sends to the external world do not pass through the brain's normal output pathways of peripheral nerves and muscles (Jonathan R Wolpaw, Birbaumer, McFarland, Pfurtscheller, & Vaughan, 2002). More recently, Wolpaw and Wolpaw (J. Wolpaw & Wolpaw, 2012) defined a Brain-Computer Interface 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 (Aloise et al., 2012; Riccio et al., 2015), developing also applications in ecological and operative environments, is not just a theory but something very closed to real applications (Blankertz et al., 2010; Müller et al., 2008; T. O. Zander, Kothe, Welke, & Roetting, 2009). In fact, in the classic BCI applications the user can modulate voluntarily its brain activity to interact with the system. In the new BCI concept, i.e. the passive BCI, the system recognizes the spontaneous brain activity of the user related to the considered mental state (e.g. emotional state, workload, attention levels), and uses such information to improve and modulate the interaction between the operator and the system itself (Aricò et al., 2017). Thus, in the context of Adaptive Automation (AA) in operational environments, the passive-BCI perfectly match the needs of the system in terms of *Human-Machine Interaction* (HMI) (Figure 10).

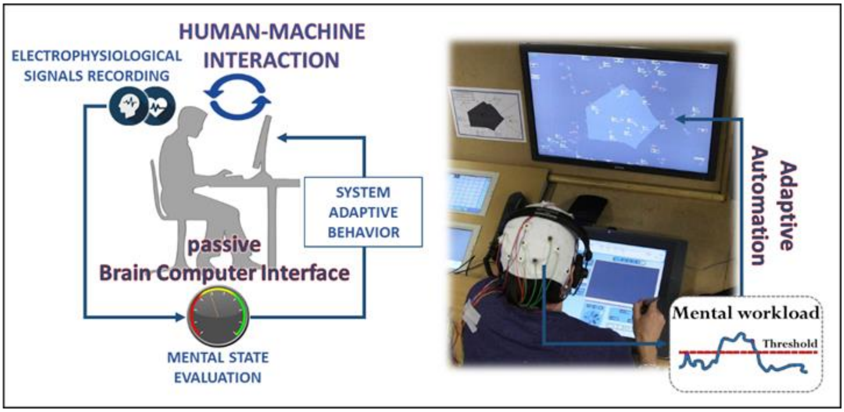


Figure 10. Representation of the passive-BCI concept applied to enhance the Human-Machine Interaction by adapting the automation of an Air Traffic Management workstation (the application is descripted in Par. 3.3**)**

Even if the main limitation of the EEG is its wearability, technology improvements(Chi et al., 2012; 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, EEG-based passive-BCI systems appear the best candidate to be integrated in the development of AA-based systems and dynamically trigger the tasks allocation, on the basis of the user's actual mental state, i.e. his Mental Workload, in order to support him during his working activity consequently improving his performance, thus the safety of the whole environment.

The still open issue is the development of a systematic method, in other words an algorithm, able to assess the user's Mental Workload online, despite all the problems related to operational environments (i.e. no controlled settings, artefacts, time cost in terms of calibration and computation, invasiveness on the subject), and independently of the environment of application.

2. AIMS

Recovering the experimental problem assessed in conclusion to the Introduction chapter (please see Par. 1.3), i.e. the necessity of developing a systematic method, in other words an algorithm, able to assess the user's Mental Workload online, despite all the problems related to operational environments (i.e. no controlled settings, artefacts, time cost in terms of calibration and computation, invasiveness on the subject), and independently of the environment of application, the main aim of my PhD research activity has been:

- The development and the validation of an algorithm able to produce online a robust EEG-based Mental Workload index (see Par. 4.2).

Because of the several issues related to the main experimental problem, my PhD research activity aimed also to:

- Face all the problems due to the fact that such method would be applied not in laboratory controlled settings but in real operational environments, where there are much bigger constrains and obstacles. The mitigation of this issue was fundamental for the robustness and reliability of the algorithm developed (see Par. 3.1 and 4.1);
- Develop a method not specific for a single application but equally reliable and usable in different domains (see Par. 3.2 and 4.3);
- Validate such method in a real application of passive Brain Computer Interfaces in operational environments (see Par. 3.3 and 4.4).

3. RESULTS

This Chapter has been divided in three main sections, in order to guide the reader through the application validation of all the methodology developed. Each paragraph has its correspondent paragraph within the "Material an Methods" Chapter (see Chapter 4), where all the details about experimental hypothesis, protocols and analysis are provided.

In particular, the first section (Par. 3.1) aims to show the results obtained for the validation of two innovative techniques, i.e. the *automatic-stop-StepWise Linear Discriminant Analysis* (asSWLDA) (Aricò et al., 2017; European Patent Nr. EP3143933 A1, 2017), and the *Regressive Eye BLINk Correction Algorithm* (REBLINCA) (Di Flumeri, Arico, Borghini, Colosimo, & Babiloni, 2016), developed to address specific issues related to the mental state evaluation in general, and in particular for online applications in real operational environments.

Once developed and tested the new algorithm for the EEG-based Mental Workload evaluation (Par. 4.2), the second section (Par. 3.2) is dedicated to the validation of such an algorithm in real operational environments: in particular three case studies have been investigated involving Air Traffic Controllers (ATCOs), car drivers in real drive conditions, and surgeons (students) managing robotic technologies in laparoscopic-surgery tasks.

The third and last section (Par. 3.3) is the final fulfilment of such PhD research activity, since it describes the application of the Mental Workload evaluation algorithm in a Passive Brain-Computer Interface with professional ATCOs in their workstation: in the operative interfaces of ATCOs adaptive automation solution have been implemented and their activation was triggered online by the Mental Workload Index of ATCOs themselves.

Also, a specific discussion paragraph is included for each experiment, whilst the Conclusions Chapter (Ch.5) aims to provide a brief overall discussion of the present Thesis work.

3.1 Validation of two innovative techniques

3.1.1. Automatic-stop-StepWise Linear Discriminant Analysis (asSWLDA)

A modified version of the standard SWLDA, the Automatic-stop-StepWise Linear Discriminant Analysis (asSWLDA) (Aricò et al., 2017; European Patent Nr. EP3143933 A1, 2017), has been developed. Such innovative classifier is actually the core of the whole method for the EEG-based Mental Workload evaluation (described in Par. 4.2). In the followings, the results of the validation of such technique on twelve professional Air Traffic Controllers (ATCOs) performing a realistic Air Traffic Management (ATM) scenario are reported.

3.1.1.1. Comparison between neurophysiological and subjective workload evaluation

The one-way ANOVA (Figure 11) 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, HARD). In particular, the post-hoc test highlighted significant differences (all $p<0.001$) between the W_{EEG} score related to all the possible paired comparisons between the difficulty conditions (i.e. EASY vs MEDIUM, MEDIUM vs HARD, EASY vs HARD).

The correlation analysis (Figure 12), 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 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 (Figure 13).

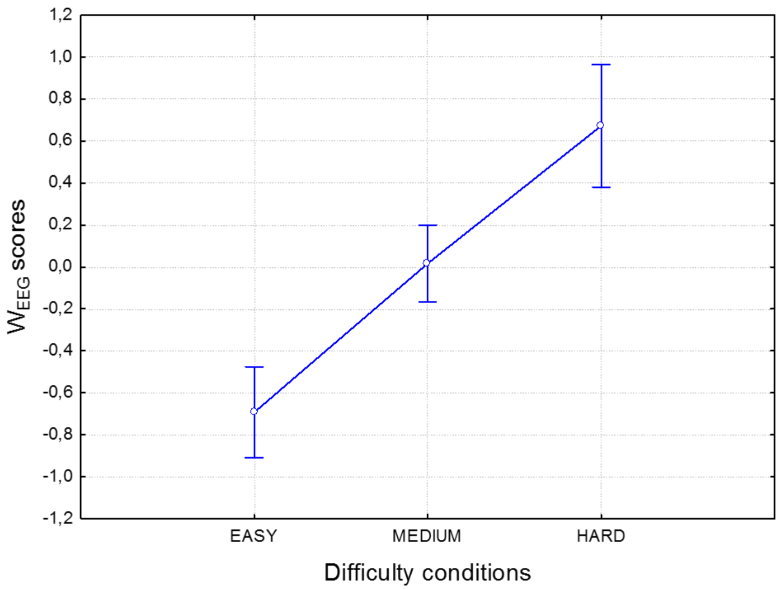


Figure 11. Error bars (CI = 0.95) related to the W_{EEG} scores along the three difficulty conditions (EASY, MEDIUM, HARD). Results showed significant differences between all the difficulty conditions ($p < 0.001$).

Finally, the Fisher's R-to-Z analysis (Figure 12) on the two correlation indexes showed no differences between them ($p = 0.676$).

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

Figure 12. In the table, 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). The Fisher's R-to-Z trasformation showed no significative difference between the two correlation values.

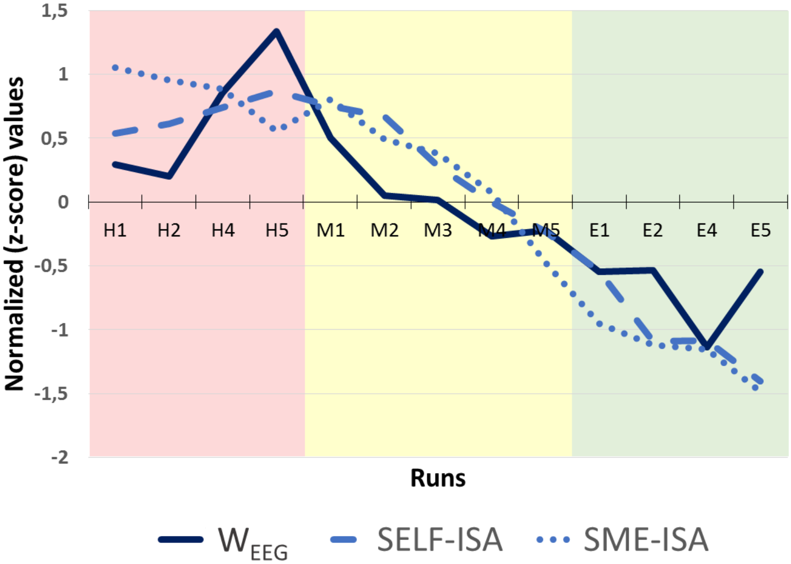


Figure 13. Shape of the three workload indexes, the neurophysiologic (W_{EEG}) and the subjective (SELF- and SME-ISA) ones, across the experimental task. The three measures were able to follow the mental workload variation along the ATM scenarios.

The scatterplot in the Figure 14 highlighted a high and positive correlation between the neurophysiological and the subjective workload indexes.

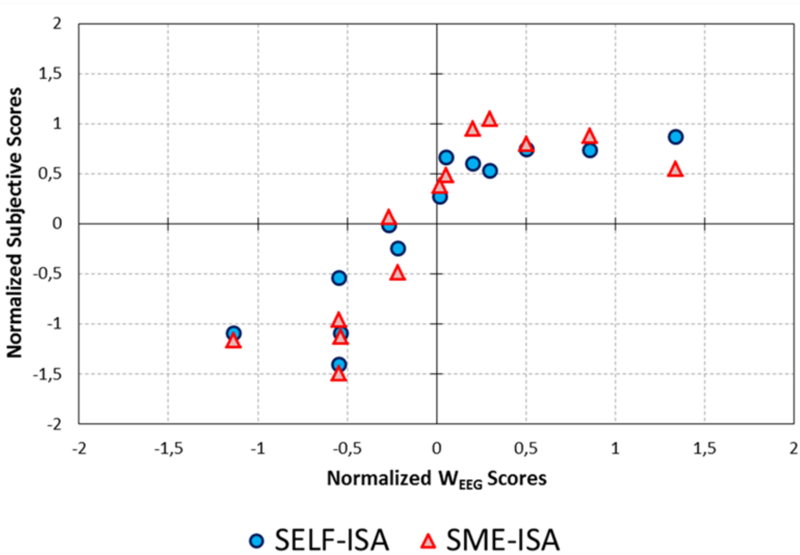


Figure 14. Scatterplot of the subjective workload measures (*SELF-ISA* and *SME-ISA*) with respect to the neurophysiological workload measure. On *the x-axis* the normalized W_{EEG} index, on the *y-axis* the normalized *SELF-* and S*ME-ISA* indexes. The results of the correlation analyses showed high and significant correlation among all the workload measures (Figure 12).

3.1.1.2.Performance stability over-time

The three two–tailed paired t-test (α=0.05) did not highlighted any difference between the two sessions (Day 1 and Day 30), for each difficulty level, both in terms of *SELF-ISA* (EASY: p=0.19; MEDIUM: p=0.63; HARD: p=62) and *SME-ISA* (EASY: p=0.78; MEDIUM: p=0.22; HARD: p=0.11) scores (Figure 15). It means that, although two different scenarios have been used during the two experimental sessions (Day 1 and Day 30), these scenarios were similar in terms of difficulty levels. It was crucial to verify this hypothesis, in order to link eventual classifier performance degradations only to intrinsic limitations of the method itself, and not because of differences in the experimental tasks.

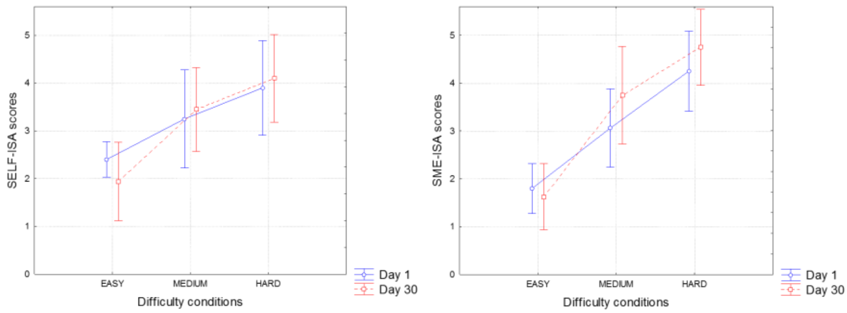


Figure 15. Error bars related to the *SELF-ISA* and the *SME-ISA* scores for each session (Day 1, Day 30) and difficulty condition (EASY, MEDIUM, HARD). Both the *SELF-ISA* (on the left) and the *SME-ISA* (on the right) scores did not show any significant differences between the two experimental sessions for each difficulty condition.

The two-way repeated measures ANOVA (CI = .95) highlighted a significant main 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) 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 SWLDA and the asSWLDA classifiers. Instead, no significant differences (0.2) between the two classifiers were highlighted regarding the *Intra* cross-validation type (Figure 16). In other words, once calibrated both the algorithms, only the asSWLDA is able to maintain its performance stable also after one month.

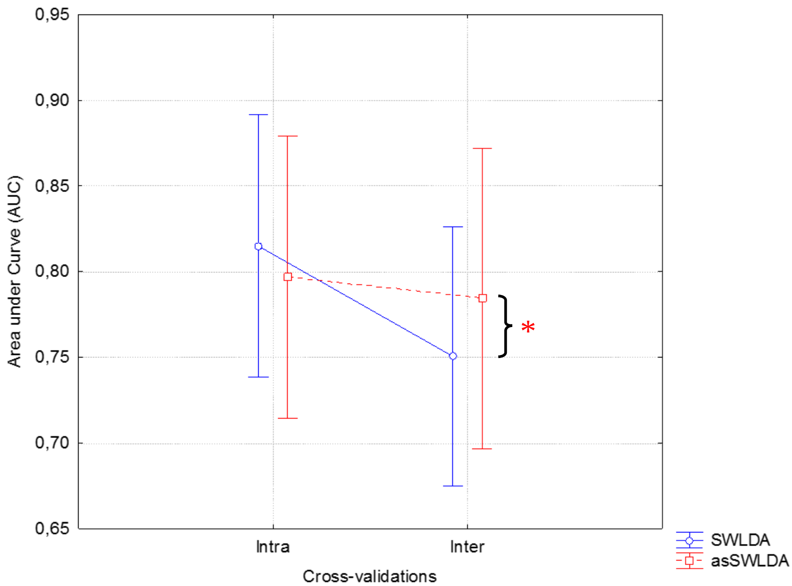


Figure 16. Error bars (CI = .95) related to the AUC values of the SWLDA and asSWLDA classifiers calculated on the EASY and HARD conditions over the two cross-validation types (*Intra* and *Inter*). Focusing on the *Inter* cross-validation type, i.e. after one month, the SWLDA performance (i.e. AUC) is significantly (p = 0.04) poorer that the asSWLDA one.

3.1.1.3.EEG features selection analysis

The 2 two–tailed paired t-test (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. In Figure 17, both the numbers of features and EEG channels are reported in terms of percentage with respect to the total number of features (72) and EEG channels (8). In particular, the asSWLDA selected the 5,2% of the available features, that means about 4 features for ATCO, by using the 37% of EEG channels, that means about 3 EEG channels in the considered study. By contrast, the standard SWLDA used the 100% of EEG channels, i.e. 8 channels, selecting the 44 % of available features, that means about 32 features for ATCO.

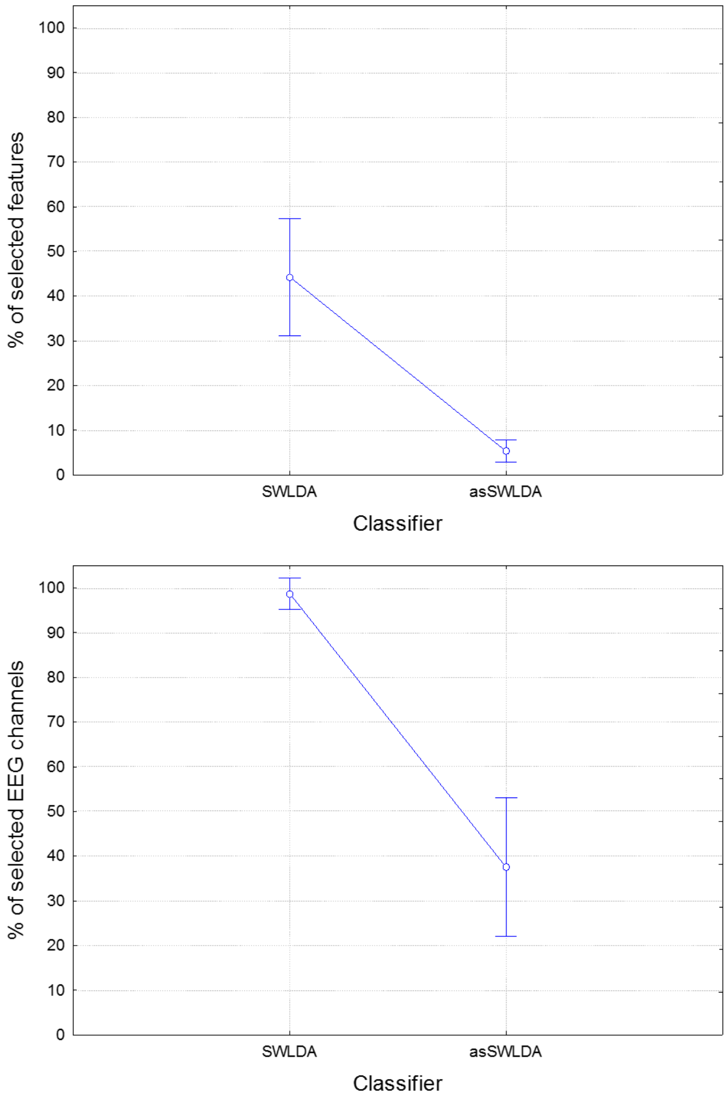


Figure 17. Error bars related to the number of features (a) and related number of EEG channels (b) selected by the two classification models (SWLDA and asSWLDA). As expected, both the number of features and related number of EEG channels used by the asSWLDA were significantly lower than those used by the standard SWLDA algorithm (respectively $p = 0.0007$ e $p = 0.0003$).

3.1.1.4.Discussion

The purpose of this experiment was to investigate the effectiveness of developed method (Par. 4.2) for the user's mental workload assessment in operational environments, in this case Air Traffic Management (ATM), by using neurophysiological measures, i.e. the measure of the operator brain activity through EEG. The brain activity (EEG) and subjective measures (SELF-ISA and SME-ISA scores) of twelve professional Air Traffic Controllers (ATCOs) have been gathered while performing a high-ecological ATM task and analyzed by a machine-learning approach.

In this experiment, the innovative *automatic-stop StepWise Linear Discriminant Analysis* (asSWLDA) (Aricò et al., 2017; European Patent Nr. EP3143933 A1, 2017), a modified version of the standard SWLDA, has been used to compute a mental workload index based on the EEG activity of the user (W_{EEG}).

This experiment in particular aimed to investigate two important key issues to use neurophysiological measurements in operative environments: 1) the reliability over time of the measure, and 2) the accuracy in comparison with the standard (i.e. subjective) workload measures.

Regarding the first issue, 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 standard SWLDA model over a week (Aricò et al., 2015). In other words, unlike the standard SWLDA algorithm, it is possible to calibrate the asSWLDA model and then using 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 3 EEG channels in the specific study), to be used for the workload assessment, in comparison with the standard SWLDA (100%, that means 8 EEG channels). 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., 2015), but it could became tricky to empirically (and manually) find out the best number of features to be used, 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 to optimize the final model. In this way, it has to be stressed that one of the big limitations in using passive BCI systems in operative environments is represented by the EEG cap, that has to be wore by the operator during his/her work-shift. 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 user.

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, because the high similarity, i.e. the correlation, between the shapes of the W_{EEG} and both the subjective indexes (SELF-ISA and SME-ISA), meant that the EEG-based workload index (W_{EEG}) was 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, that is, they are highly operator-dependent and they cannot be used to quantify objectively the operators' mental states (i.e. mental work-load). 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 all the algorithm used in this study did not need any a priori knowledge about data, thus the proposed methodology can also be used online, for example to improve the human-machine interactions 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 work-shift across different days (up to a month). Of course, the results obtained with this experiment do not exclude the possibility to achieve similar or even better results by using other machine learning techniques. However, they have demonstrated the reliability of the developed algorithm to measure the operators' mental workload during a real operational environment.

3.1.2. Regressive Eye BLINk Correction Algorithm (REBLINCA)

The REBLINCA is an innovative method for the blinks artefacts correction on the EEG data I developed during my PhD research activity (Di Flumeri et al., 2016). Because of its particular advantages for the online applications, it is fundamental within the whole method for the EEG-based Mental Workload evaluation (see Par. 4.1.2 and Par. 4.2).

It has been validated by comparing its performance with the best outperforming methods in this field, the Gratton, the extended InfoMax and SOBI, on a EEG dataset of ten subjects.

From a purely qualitative point of view, Figure 18 and Figure 19 show the effects of the correction by using the four methods considered in this study, in correspondence of an "eye blink epoch" randomly selected along one representative EEG channel (i.e. AF3), respectively in time and frequency domain.

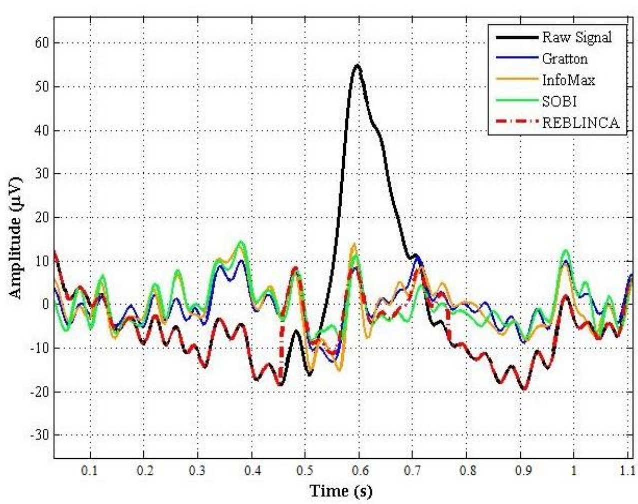


Figure 18. Graphical comparison (time domain) between the effects of the four correction methods on the Raw Signal (black line), in particular the EEG signal

related to the AF3 channel in correspondence of one blink randomly selected. It is evident how the shape of the REBLINCA-corrected signal (dotted red line) is similar to those ones resulting from the application of other methods in correspondence of the eye-blink, and exactly equal to the Raw Signal outside the blink.

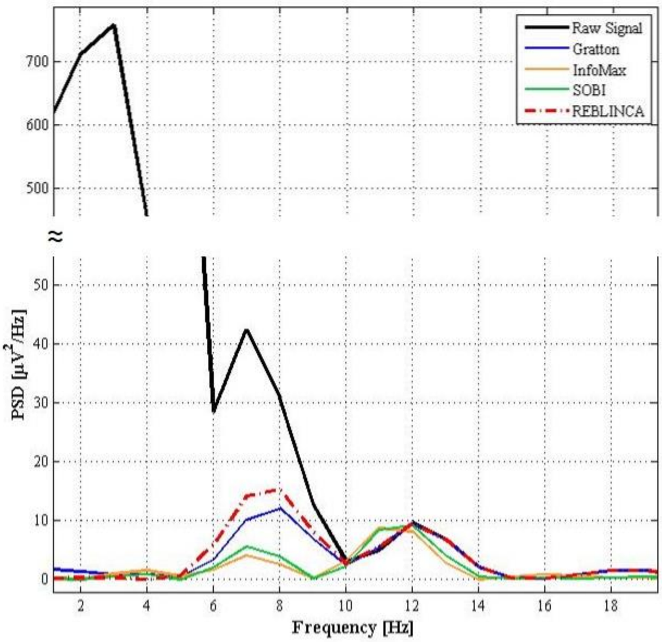


Figure 19. Graphical comparison (frequency domain) between the effects of the four correction methods on the Raw Signal (black line) spectrum, as in Figure 18. It is evident how the shape of the REBLINCA-corrected signal spectrum (dotted red line) is similar to those ones resulting from the application of other methods, with a great reduction of the spectral components related to the blink electrical activity (< 8 Hz).

Figure 20 shows the effects of the performance of the offline REBLINCA algorithm in correcting blinks artefacts on all the channels of an EEG dataset. The red line represents the EEG raw signal, while the blue line the EEG corrected signal. It is important to notice how, except in correspondence of blinks, the blue line is completely overlapped to the red line, i.e. the EEG signal is not manipulated outside the blinks.

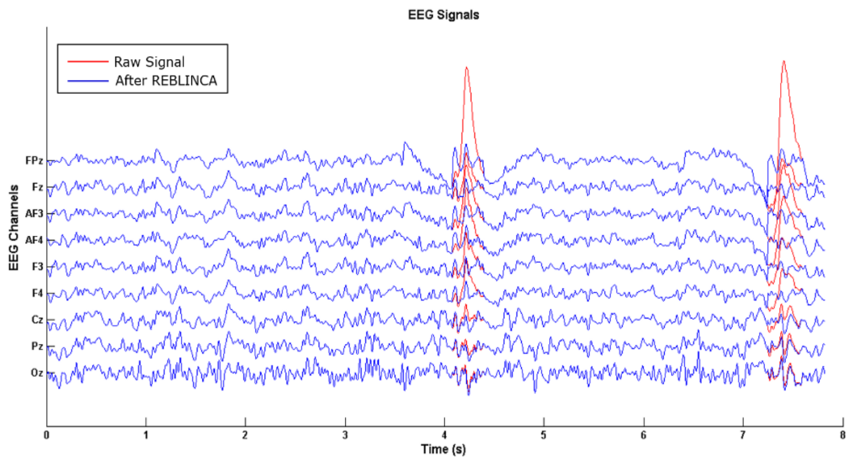


Figure 20. Application of the offline REBLINCA algorithm on an EEG dataset: the red line represents the Raw signal, the blue line is the EEG signal after the artefacts correction.

In the following, the results of the statistical analysis are reported, with the aim to validate the offline and after the online REBLINCA algorithm.

3.1.2.1.REBLINCA offline

The table in Figure 21 reports the results obtained from the 24 ANOVAs on the signal "blink windows" after the correction. The results showed significant ($p<0.05$) differences, in terms of EEG correction by using the four methodologies, in the alpha band for all the frontal electrodes, in the theta band for Cz, Pz and Oz, and in the delta band on Cz. In particular, the post-hoc test highlighted significant ($p<0.05$) differences between, but not within the methodology types (regression-based and the ICA-based algorithms). In other words, the REBLINCA algorithm did not show significant differences with respect to the considered reference method (i.e. Gratton). No significant differences between all the methods, in terms of EEG signal PSDs, have been found in correspondence of the Theta band over the frontal sites, i.e. where the blink artefacts contribute is higher.

		Fz	AF3	AF4	F3	F4	Cz	Pz	Oz
Delta	**F(3,27)**	2,82	2,43	2,04	2,64	2,31	3,25	2,63	1,89
	p	0,051	0,087	0,13	0,07	0,09	0,037	0,07	0,15
Theta	**F(3,27)**	2,45	0,85	0,60	1,60	1,62	7,25	7,34	4,14
	p	0,08	0,48	0,62	0,21	0,21	0,001	0,0009	0,016
Alpha	**F(3,27)**	7,09	8,99	12,375	4,66	8,10	1,16	2,10	1,21
	p	0,001	0,0002	0,00003	0,009	0,0005	0,34	0,12	0,32

Figure 21. The table in figure reports the ANOVAs results on the EEG signal PSD values in correspondence of blinks.

The ANOVAs (Figure 22) performed on the correlation coefficients of the four methodologies were all significant ($p<0.05$). In particular, as confirmed by the post-hoc analyses, the REBLINCA showed a significant higher correlation value than the other methods, very close to 1 for each EEG channel (Pearson's correlation coeff. > 0.985 ± 0.02, Figure 23). This correlation value confirms the efficiency of the threshold method for the eye blinks recognition. Although the trend was the same, for the Cz ($p=0.06$), Pz ($p=0.16$) and Oz ($p=0.11$) EEG channels, such difference was not significant between REBLINCA and InfoMax.

	Fz	AF3	AF4	F3	F4	Cz	Pz	Oz
F(3,27)	13,51	22,71	19,27	11,92	12,97	7,23	4,61	2,97
P	$1*10^{-5}$	$6*10^{-9}$	$5*10^{-9}$	$4*10^{-5}$	$2*10^{-5}$	0,001	0,009	0,049

Figure 22. The table in figure reports the ANOVAs results for the free blinks signal windows. Supported by Figure 23, these results highlight how the REBLINCA preserve the EEG signal significantly better than the other considered methods.

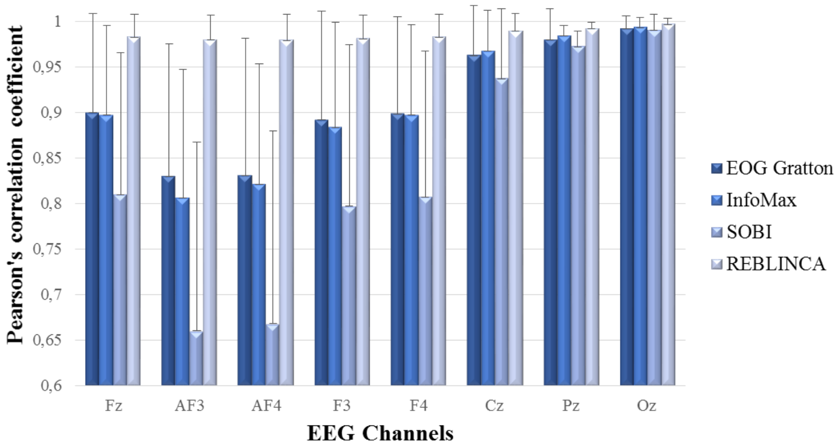
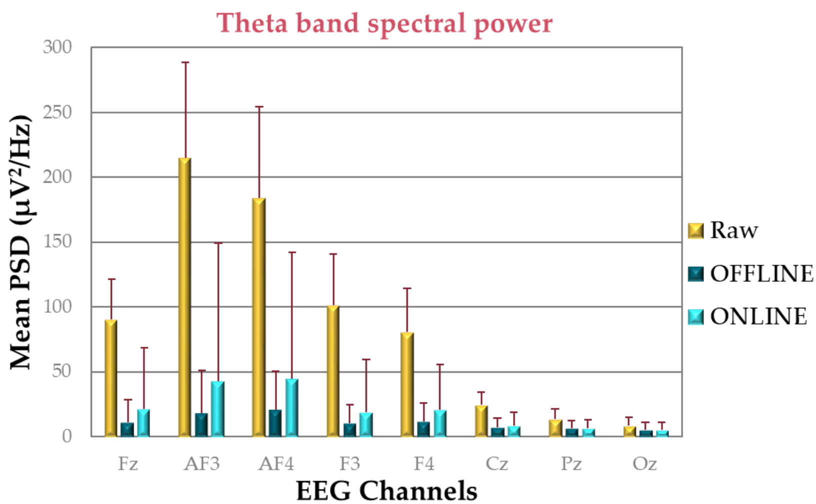


Figure 23. The mean correlation coefficient significantly higher ($p<0.05$) after the REBLINCA application highlights the capability of this method in preserving the EEG information in blink free signal segments, in contrast to the traditional ones. This is verified for all the considered channels.

3.1.2.2.REBLINCA online

Once demonstrated the remarkable performance of the offline REBLINCA algorithm, they are used to evaluate, by comparing them, the performance of the new implementation of the online REBLINCA algorithm. Before any analysis in terms of performance, the Pearson's correlation analysis was performed between the weights used by the REBLINCA online, thus obtained by the Training condition, and offline, thus obtained by the Testing condition, in order to validate the experimental hypothesis (see Par. 4.1.2.2). The values of the two REBLINCA implementations shew a very high and significant correlation, in particular $R = 0{,}984 \pm 0{,}02$, with $p < 10^{-7}$.

In terms of correction performance, i.e. the EEG signal PSDs in correspondence of "blink windows" after the correction, the two-tailed paired T-test analysis have shown no significant differences ($p > 0{,}05$ over all the channels) between the offline and online implementations of REBLINCA (please see Table in Figure 24). Also, the Figure 24 visually highlights that after the correction the PSDs in Theta band, where the blink contribute is higher, are lower than those of the Raw signal, an intrinsic demonstration of the method effectiveness.

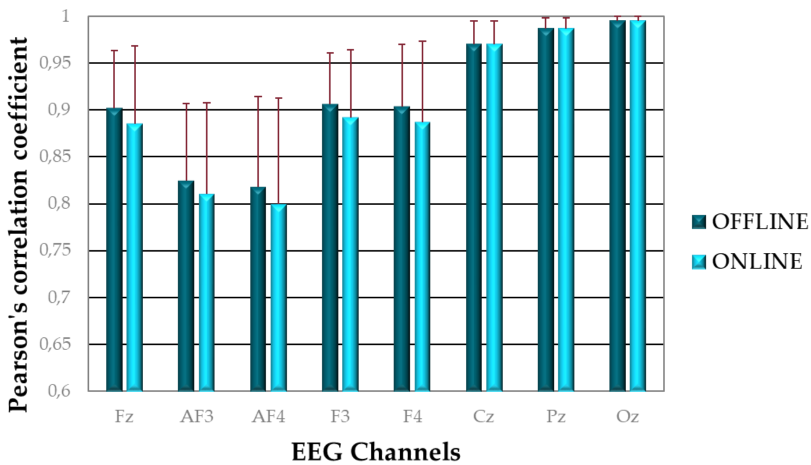


2-tailed paired T-test (p value)			
Channels	**Delta**	**Theta**	**Alpha**
Fz	0,32	0,32	0,26
AF3	0,33	0,35	0,38
AF4	0,30	0,31	0,13
F3	0,34	0,35	0,64
F4	0,28	0,27	0,17
Cz	0,31	0,36	0,30
Pz	0,25	0,54	0,49
Oz	0,16	0,92	0,25

Figure 24. EEG signal PSDs analysis in correspondence of "blink windows". The reported T-test results compared the offline and online REBLINCA implementations.

In terms of preservation of the EEG signal outside the blinks, both the algorithms have been able to achieve Pearson's correlation coefficients, with respect to the raw signal, from 0.8 up to 1, i.e. both the implementations are able to preserve the EEG signal. The two-tailed paired T-test analysis have shown no significant differences (p

> 0,05 over all the channels) between the offline and online implementations of REBLINCA (please see Table in Figure 25).



2-tailed paired T-test	
Channels	**p**
Fz	0,107
AF3	0,091
AF4	0,079
F3	0,109
F4	0,088
Cz	0,99
Pz	0,99
Oz	0,99

Figure 25. EEG signal time-domain correlation analysis in correspondence of "no-blink windows". The reported T-test results compared the offline and online REBLINCA implementations.

3.1.2.3.Discussion

The results demonstrated the high efficiency of the REBLINCA application for eye blinks correction, with respect to three of the best and most used algorithms in scientific literature, i.e. the Gratton, the extended Infomax and the SOBI methods. In particular, results did not show any significant difference for each EEG frequency band by using the REBLINCA and the Gratton method, considered as the outperforming regressive method. This highlights that the two algorithms do not differ in terms of efficiency in removing eye blinks contributions. As expected, the two ICA-based methods showed significantly different behaviours with respect to the two regression-based methods, since they showed a greater reduction on the alpha band PSD over the frontal electrodes, probably a consequence of filtering the EOG/EEG channel used in the regression-based methods for the eye blink component subtraction until 7 (Hz). In addition, the REBLINCA preserved significantly better than all the other considered methods the EEG signal when blinks do not occur. This result is really important, since one of the main disadvantages of the regression-based methods, as discussed in Par. 4.1.2, is the mutual contamination between EEG and the channel used for the regression (i.e. EOG channel). Thus, this REBLINCA feature allows to use an EEG channel for the regression, on the contrary of the Gratton or similar, that in such case could deprive the EEG signal of important neural components. Of course, despite no need of an EOG channel, a frontal EEG channel is preferred, as the eye blink activity has a greater component in such sites than the other brain areas. Finally, REBLINCA does not require any minimum set of EEG channels, great computational effort, and manual components selection to reject, compared to the ICA-based methods. Last but not least, the online implementation of the REBLINCA did not show any significant performance decreasing, thus it appears as the best candidate for online applications in real environments and, for such a reason, it has been included in the whole method for the EEG-based Mental Workload evaluation (Par. 4.2).

3.2 Application in operational environments

In the following, all the results provided have been obtained by using the algorithm for the EEG-based Mental Workload evaluation (Par. 4.2). Therefore, they have to be considered as the validation of such a method in the different operational environments.

3.2.1. Case study 1: Air Traffic Controllers

In this experiment, thirty-seven Air Traffic Controllers (ATCOs), divided in 15 professional (ATCO Experts) and 22 students (ATCO students) were asked to perform a realistic Air Traffic Management (ATM) scenario 45-minutes-long. The task was designed in order to produce a continuously varying difficulty level, from Hard to Medium to Low.

3.2.1.1.Performance analysis

In Figure 26 are reported the results of the analyses, in terms of performance (i.e. the Weighted Mean Reaction Times), for the two different groups. In particular, the blue line shows the WMRT of the Experts, while the red line shows the WMRT of the Students in the three levels of the ATM scenario. The Experts did not report any significant differences ($p > 0.05$) among the difficulty conditions. On the contrary, the Students showed a significant ($p < 0.05$) decrement of performance (increased WMRT) between the Hard condition and the others (Easy and Medium).

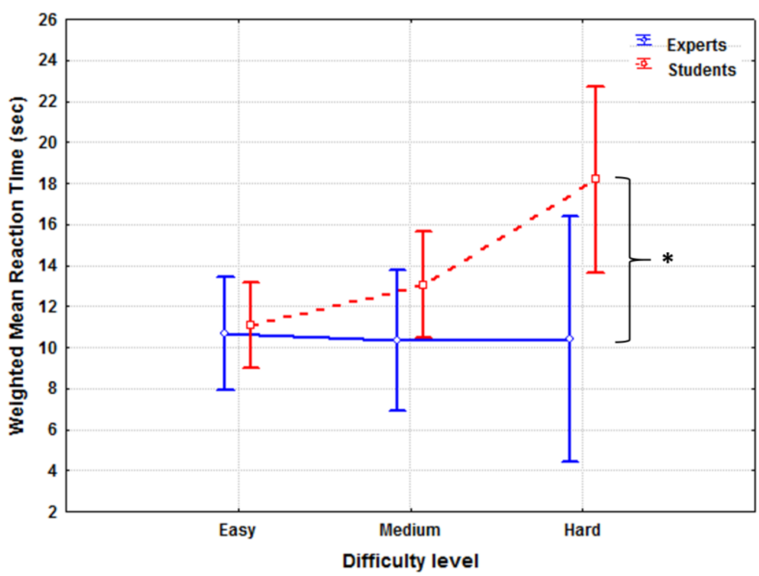


Figure 26. Weighted Mean Reaction Time (WMRT) of the ATCOs across the three difficulty levels of the simulated ATM scenario. The ATC Experts (blue line) did not report any significant differences ($p>0.05$) among the difficulty conditions. On

the contrary, the ATC Students (red line) showed a significant ($p<0.05$) decrement of performance between the Hard condition and the others (Easy and Medium).

The statistical analysis between the groups (Figure 27) showed a significant difference between the two groups ($F(1, 31) = 4.7621$; $p = 0.03679$). In other words, the ATC Experts reacted and provided information to the pilots faster than the ATC Students.

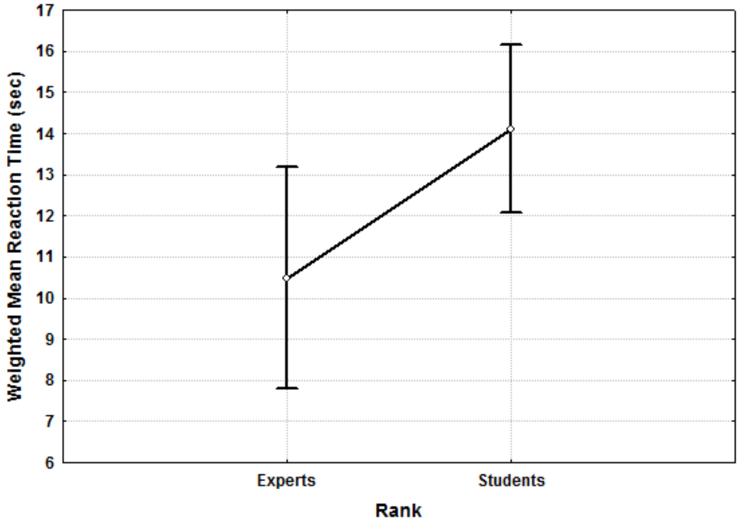


Figure 27. Differences between the ATC Experts and ATC Students in terms of Weighted Mean Reaction Time (WMRT). The ANOVA analysis showed a significant difference ($p=0.037$) between the two groups.

3.2.1.2.Comparison between neurophysiological and subjective workload evaluation

First of all, the discriminability of the three different levels of difficulty has been investigated both in terms of neurophysiological (EEG-based Workload index) and subjective (ISA) measures. In Figure 28 and Figure 29 both the indexes averaged over the whole ATCOs sample are reported.

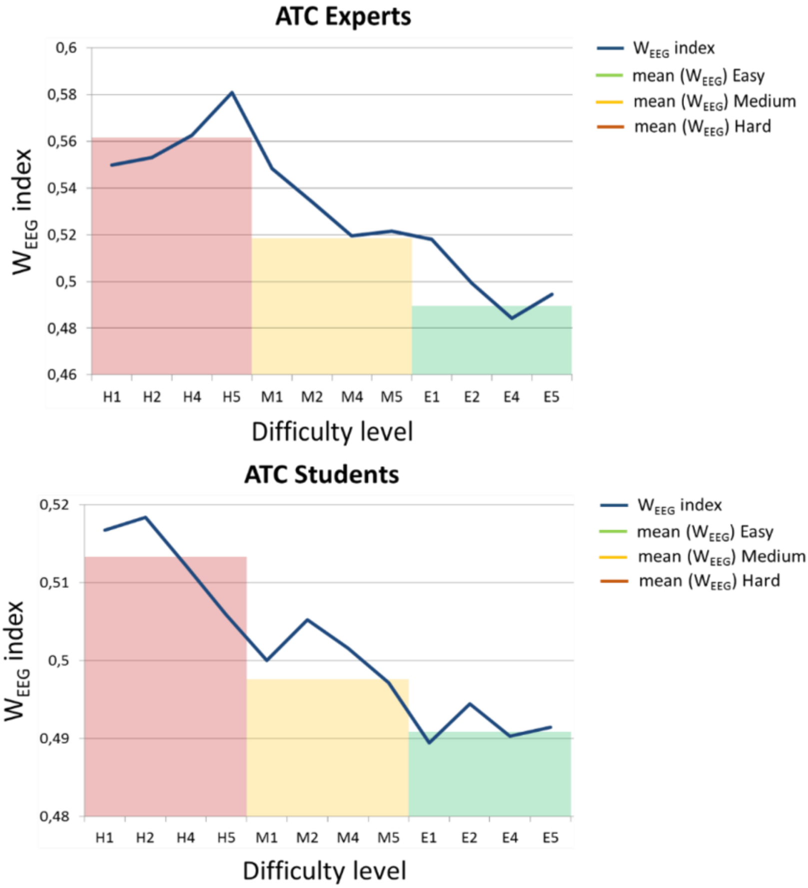


Figure 28. Representation of the W_{EEG} scores across the different difficulty condition runs (E1÷E5 – green bar, M1÷M5 – yellow bar, and H1÷H5 – red bar) and its mean value for the ATC Experts and Students.

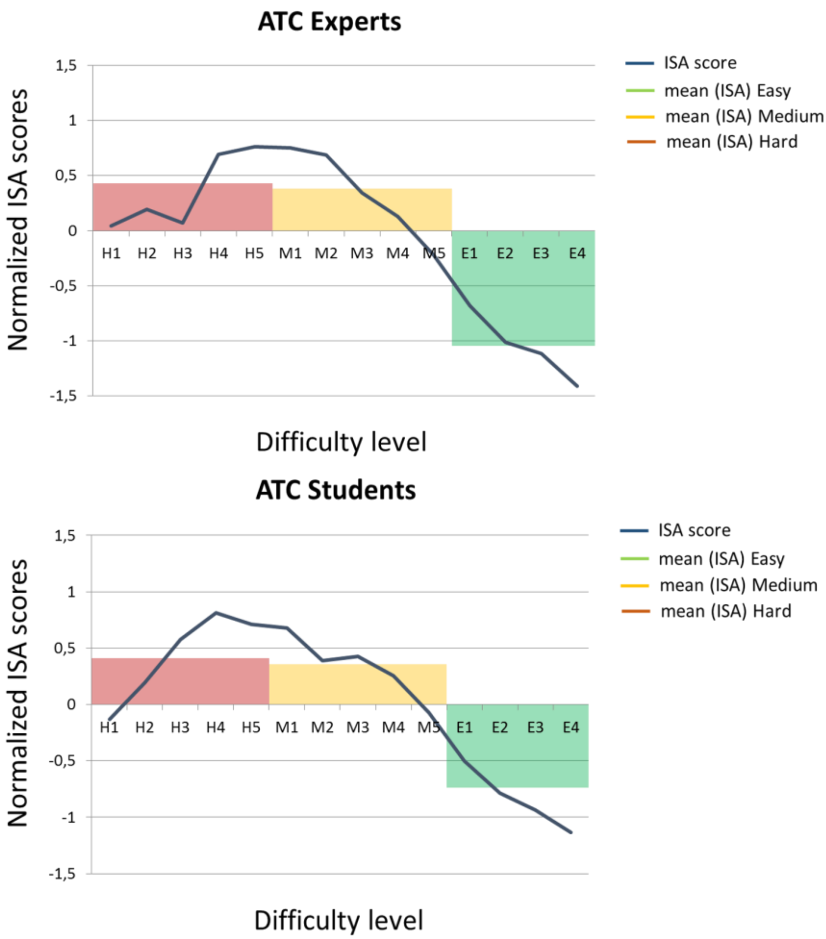


Figure 29. Representation of the ISA scores across the different difficulty condition runs (E1÷E5 – green bar, M1÷M5 – yellow bar, and H1÷H5 – red bar) and its mean value for the ATC Experts and Students.

In Figure 30, the table reporting the p-values of the t-tests on the W_{EEG} and ISA scores distributions related to the different difficulty levels is showed. It is interesting to notice that, while the EEG-based Workload index was able to discriminate significantly the three levels of ATCOs' mental workload, the subjective (ISA) measures were not able to discriminate significantly the high from the medium difficulty level.

		Easy vs Medium	**Medium vs Hard**	**Easy vs Hard**
$\mathbf{W_{EEG}}$	ATC Experts	0.02	0.01	0.003
	ATC Students	0.04	0.04	0.003
ISA	ATC Experts	$5.2*10^{-9}$	0.854	$8.9*10^{-5}$
	ATC Students	$2.6*10^{-5}$	0.593	$7.5*10^{-5}$

Figure 30. The table reports the p-values of the t-tests on the W_{EEG} and ISA scores distributions related to the different difficulty levels for each group of subjects, i.e. ATC Experts and Students.

Figure 31 reports the Table showing the results, in terms of Pearson's correlation index and significance, between the two measures for each group (ATC Experts and Students) and for the whole sample. The results show how, for all the comparisons, the correlation is positive and significant.

	R	**p**
ATC Experts	0.844	0.001
ATC Students	0.640	0.0339
All ATCOs	0.822	0.0019

Figure 31. Pearson's Correlation Coefficient (R) and significance (p-value) level between the neurophysiological (W_{EEG}) and the subjective (ISA) workload measures for the two groups of ATCOs.

Furthermore, the scatterplot in Figure 32 highlights the positive and significant correlation between the two measures for both the ATC groups (ATC Experts in blue circles, and ATC Students in okra triangles), and in particular for the whole sample as underlined by the dashed red tendency line.

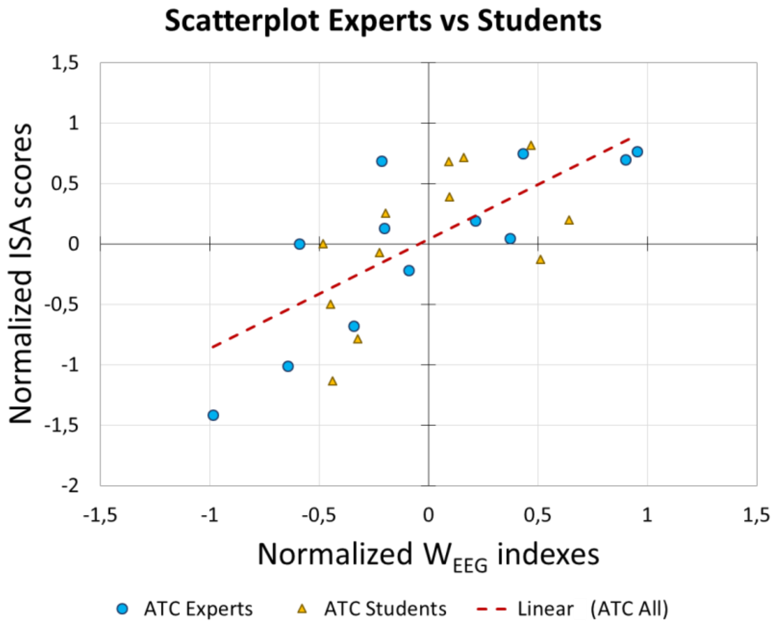


Figure 32. Representation of the correlation between the neurophysiological (W_{EEG}) and subjective (ISA) workload measures for the ATC Experts (blue circles) and ATC Students (okra triangles). The dashed red line is the tendency line of the whole distribution (ATC Experts and Trainees together).

3.2.1.3.Discussion

Detecting a state of mental *under*- or *overload* in working environments is a key issue for i) developing adaptive systems to support the operator in a timely and accurate manner, specifically during periods of high vulnerability, and ii) providing online information to keep the operator engaged along the entire work. Such great effort in improving *Human–Machine* interactions (Aricò, Borghini, Di Flumeri, Colosimo, Bonelli, et al., 2016; Giraudet, Imbert, Bérenger, Tremblay, & Causse, 2015; Thorsten O. Zander & Kothe, 2011) aims to guarantee proper level of safety in the working environment. This approach is known as *Neuroergonomics*, and it was initially proposed by Parasuraman (Parasuraman & Rizzo, 2008) and progressively developed and refined over the last decades. The present experiment aimed to find out if neurophysiologic measures, i.e. the EEG-based workload index (W_{EEG}), could provide additional and reliable information for the evaluation of the mental workload while accomplishing a realistic ATM scenario, and to objectively assess the expertise of two groups of professional ATCOs in terms of experienced workload.

Regarding the expertise assessment, the analysis of the task performance provided important indications regarding the two groups. In particular, the ATC Experts did not show any difference among the three ATM phases, whilst the ATC Students reported a significant performance reduction during the Hard condition. In addition, significant difference has been found between the two groups in terms of reaction time for providing information and command to the pilots, as the Experts needed shorter time ($p=0.01$). On the contrary, no significant differences have been highlighted during the Easy and Medium slots.

In terms of W_{EEG} indexes, the results showed that the proposed index was able to track the workload demand along the entire ATC simulation for both the groups, and to differentiate significantly the three difficulty levels within the ATM scenario, even more than subjective measures (Di Flumeri et al., 2015). Strong correlations (all $R > 0.64$; $p < 0.034$) between the subjective (ISA), neurophysiologic (W_{EEG}) workload measures, and ATM scenario complexity were found for both the ATCO groups. Such correlations confirmed that the three difficulty levels (E, M and H) have impacted differently on the perception, behaviour and cognitive demands of the ATCO, since these measures followed exactly the profile of the ATM scenario.

To sum up, the considered measures allowed to objectively describe each group of ATCOs in terms of experienced mental workload, and demonstrated the additional value of neurophysiologic measures to objectively assess their expertise while dealing with realistic activities. In fact, such a kind of data do not require additional tasks (i.e. no interferences with the main task), they exhibited higher resolution than the subjective and behavioural ones, and they can provide online information concerning the cognitive status of the operator.

3.2.2. Case study 2: Car drivers

In this experiment, twenty male drivers have to perform a route in real traffic conditions. The route was designed in order to identify two different segments characterized by two different difficulties, i.e. Easy and Hard. The eye-tracking data have been used to validate this experimental hypothesis.

3.2.2.1.Results

The analysis of the eye fixations, as reported in Figure 33, showed that during the EASY condition the external environment caught the 12 ± 10.7 % of the drivers' fixations, whilst during the HARD condition only the 5 ± 5.5 %. The paired t-test revealed that such a difference was statistically significant (p = 0.028).

Figure 33. Mean and standard deviation of the eye fixations over the external environments, in other words on all these regions of the scene that do not contain information (e.g. infrastructure, vehicles, signage) useful for the driving. The fixations during the EASY condition (green bar) over the environment were significant higher than during the HARD condition (red bar). For each condition two representative frames are reported on the right: on the top the EASY condition, with the driver's gazed fixed on the external environment; on the bottom the HARD condition, with the driver's gaze fixed on the precedent vehicles.

The AUCs analysis revealed that, by using such approach, it has been possible to achieve a mean AUC value of 0.73 ± 0.13, that is, a classification accuracy of 73 %. In particular, the paired t-test demonstrated that the *Measured AUCs* were significantly higher (p =

0.006) than *Random AUCs* (Figure 34.a). Also, the paired t-test between the WL indexes during the two conditions showed that the WL indexes during the HARD circuit section were significantly higher ($p = 0.0005$) than those during the EASY one (Figure 34.b).

These results confirm the capability of the method in measuring the workload of users during real driving conditions.

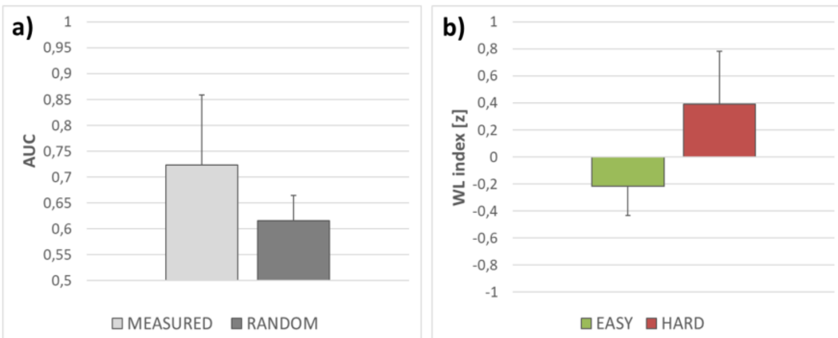


Figure 34. On the left (a), the AUC values, in terms of mean and standard deviation, obtained by using the asSWLDA algorithm on measured (light blue) and random (dark blue) data. On the right (b), the WL indexes (mean and standard deviation of the z-scores) obtained during the EASY (green) and the HARD (red) conditions.

3.2.2.2.Discussion

Also in the automotive domain, the relationship between human errors and driving performance impairment due to a high mental workload has been deeply investigated. Several automated solutions have been introduced to support the driver, by the way there is the need of developing Adaptive Automation (AA) solutions in order to *adapt* the automation intervention to the mental status, i.e. the mental workload, of the driver to keep him/her always "In-the-loop" (Byrne & Parasuraman, 1996).

This experiment aimed to validate the EEG-based workload index for the assessment the driver mental workload even online, thus becoming useful for triggering AA-based system. The Eye-tracking results confirmed the hypothesis about the circuit design, highlighting how the assumed conditions EASY and HARD, were actually perceived coherently with the assumptions by the drivers, since their eye-fixations on the external environment were significantly higher in the EASY than in the HARD condition (de Winter, Happee, Martens, & Stanton, 2014). Therefore, the results about the classifier AUC values demonstrated that the adopted approach achieves considerable performance (AUCs > 0.7), significantly higher than a random classifier ($p < .05$), providing a WL index that correctly assess the driver's mental workload during the different conditions.

On balance, the achieved results highlighted that not only it is possible to assess the driver's workload in real driving, but it is possible to obtain an objective index of his/her experienced workload by applying the same algorithm (Par. 4.2) validated in a different operational field, i.e. the Air Traffic Management.

Certainly, further analyses are necessary in order to validate the method with a larger subjects sample and during different driving conditions.

3.2.3. Case study 3: Robot-assisted surgeons

In this experiment, twenty healthy students of Medicine of the University have been recruited. They had to perform similar laparoscopy-like tasks, under different levels of difficulty (Easy and Hard), by using two different systems for the robot-assisted surgery: the *da Vinci Surgical system* (Intuitive Surgical Inc, USA), leader in the field of robotic surgery, and the *Actaeon Console* (BBZ srl, Italy), basically a cheaper simulator aimed to train students to use the da Vinci system.

3.2.3.1.Validation of EEG-based workload evaluation

In order to validate the algorithm for the mental workload estimation (please see Par. 4.2) in terms of robustness toward the system used by the operator, it has been applied in three ways: in the first two cases, it has been used in the "right" way by using training and testing data from the same system (BBZ simulator and Da Vinci robot). Thirdly, it has been applied also crossing the dataset, i.e. training on the BBZ and testing on Da Vinci data, and vice-versa. In addition, for all the cases the algorithm has been applied randomizing the testing data.

The ANOVAs analysis (Figure 35) revealed no significant differences in terms of algorithm performance independently from the system used: for all the three cases (BBZ, Da Vinci and mixed, blue line in figure) the classification accuracies in distinguishing the Easy from the Hard Tornwire task achieved values in the range between 70 % and 75 %.

In addition, all the three 2-tailed paired t-tests (one for each case: BBZ, Da Vinci and Mixed) have shown always significant differences ($p < 10^{-6}$) with respect to the application on randomized testing data (red line in Figure 35).

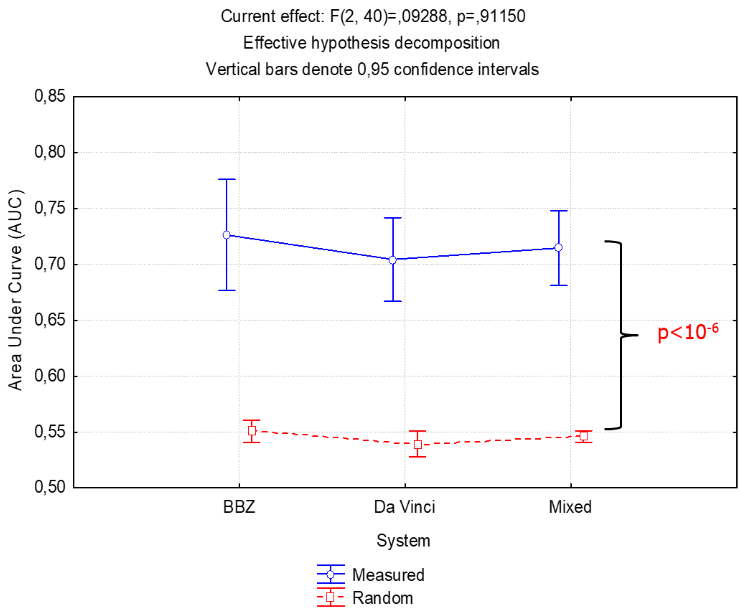


Figure 35. ANOVAs analysis on the AUC values obtained by training and testing the algorithm on the BBZ simulator, on the Da Vinci robot, and crossing the datasets, i.e. training on BBZ and testing on the Da Vinci and vice-versa. Also, the reported significance is related to the three two-tailed paired T-tests performed for each case (BBZ, Da Vinci and Mixed), between the algorithm correct performance (“Measured”, blue line) and those ones obtained randomizing the testing data (“Random”, red line).

3.2.3.2.Comparison of different systems (BBZ vs Da Vinci)

Once validated the method, the EEG-based workload indexes have been used to compare the two systems, i.e. the BBZ simulator and the Da Vinci robot, from a cognitive point of view.

Figure 36 and Figure 37 show the results of the two-tailed paired t-tests respectively on the behavioural (i.e. performance) and subjective (i.e. NASA-TLX) measures of workload. In both the cases, no significant differences arose from the analysis (respectively p = 0.96 an p = 0.31).

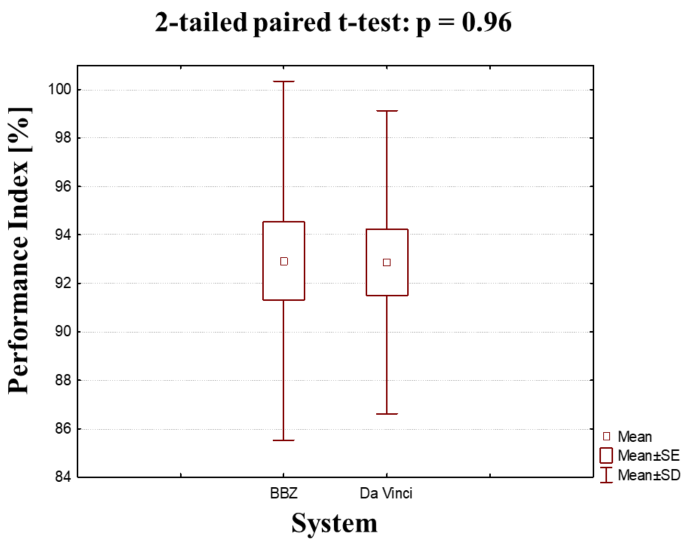


Figure 36. Two-tailed paired t-test on the performance indexes between the two systems, i.e. the BBZ simulator and the Da Vinci robot.

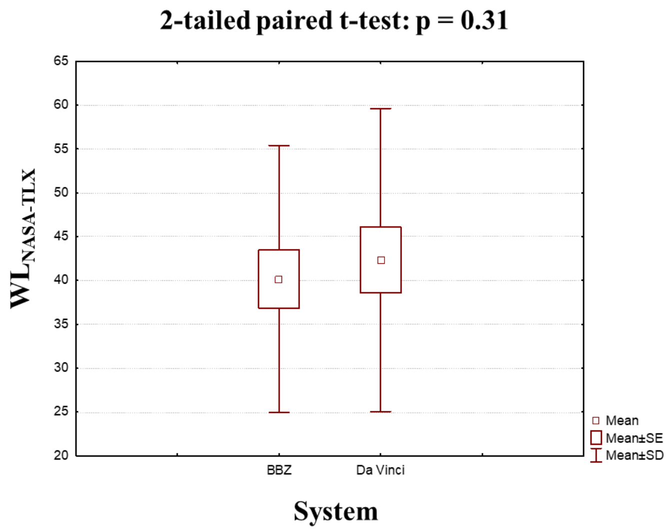


Figure 37. Two-tailed paired t-test on the subjective workload measures (NASA-TLX) between the two systems, i.e. the BBZ simulator and the Da Vinci robot.

The EEG-based workload indexes confirmed these results: the two-tailed paired t-test (Figure 38) did not reveal any significant difference between the two systems (p = 0.66). In addition, the Pearson's correlation analysis (in Figure 39 the scatterplot of such analysis) highlighted a positive and significant correlation between the EEG-based workload indexes obtained on the same subjects managing the two systems. In other words, these results confirm the high similarity of the two systems in terms of cognitive demand.

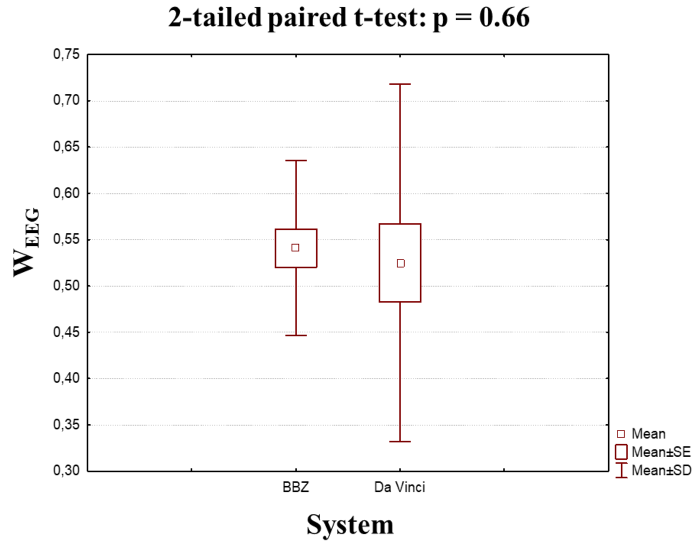


Figure 38. Two-tailed paired t-test on the EEG-based Mental Workload indexes (W_{EEG}) between the two systems, i.e. the BBZ simulator and the Da Vinci robot.

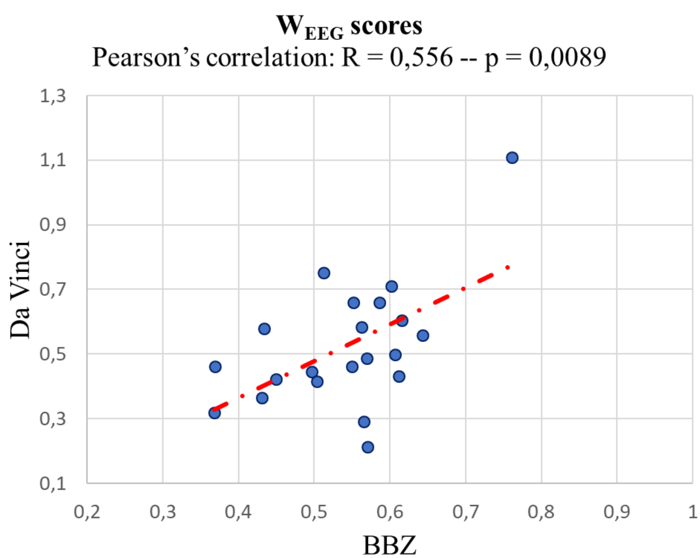


Figure 39. Scatterplot highlighting the positive and significant correlation between EEG-based Mental Workload indexes obtained for each subject (1 dot = 1 subject)

managing the two systems, i.e. the BBZ simulator and the Da Vinci robot. In red the tendency line.

3.2.3.3.Discussion

Also in the medical domain, the relationship between human errors and mental impairment due to a high mental workload has been deeply investigated and documented (Helmreich, 2000). With the exploitation of robotic technology, more and more robotic systems, in particularly in surgery, have been developed with the aim to support the surgeon. Therefore, in the context of my PhD research activity, it appeared as another interesting field where to investigate the possibility of using neurophysiological measures of mental workload, i.e. the EEG-based index, to develop passive Brain-Computer Interfaces.

The results in terms of classification accuracies have demonstrated the reliability of the EEG-based workload evaluation method (Par. 4.2) also in such a kind of application. Furthermore, even more interesting, the robustness of the algorithm has been proven by the fact that also calibrating the method on a system (e.g. the BBZ) and testing it on a different system (e.g. the Da Vinci), the performance in terms of workload classification remained stable. It has to be considered as an implicit demonstration of the fact that the method is really sensitive to the investigated phenomenon, i.e. the Mental Workload, and not to spurious actions that depend from the used system.

Last but not least, this experiment demonstrated that it is also possible to use these EEG-based Workload indexes to compare different devices from a cognitive point of view,

3.3 Online passive Brain Computer Interface application

In this experiment, a real passive-BCI system has been fully integrated with a high realistic ATM simulator able to trigger adaptive solutions in real-time depending on the Mental Workload estimated by means of the ATCO’s brain activity. The expectation was that the passive-BCI system would be able to trigger the AA solutions and to reduce the task workload demand in order avoid both under- and overload conditions.

The data analyses have been performed with the aim to demonstrate the two main experimental questions of my PhD research activity:

- Once validated the EEG-based Mental Workload index, it is possible to trigger the AA by using the online recognition of the actual mental workload of the user?
- Is such a kind of passive-BCI application effective in an operational environment? In other words, does the AA induce a reduction of the mental workload of the operator when it became high and consequently an increasing in performances execution of the task?

The experimental protocol involved twelve student ATCOs, whom have been asked to manage an operational interface that simulates a high realistic ATM scenario (Aricò, Borghini, Di Flumeri, Colosimo, Bonelli, et al., 2016). The complexity of the task could be modulated according to how many aircrafts the ATCO had to control, the number and type of clearances required over the time and the number/trajectory of other interfering flights. Also, for the purposes of this study, specific Adaptive Automation (AA) solutions have been embedded in the ATM interface, with the aim to induce a decreasing in the operators' mental workload during high workload situations. The subjects performed two similar ATM scenarios, one with automation enabled (thus triggered by their online EEG-based Mental Workload index – AA on condition) and one with automation disabled (AA on condition).

3.3.1. Triggering of Adaptive Automation solutions

The t-test results showed that the number of AA activations triggered by the W_{EEG} index (*AA On* condition) was significantly lower ($p = 0.04$) during the Easy period with respect to the Hard one. In other words, the classifier triggered more often the ATM interface when the operator's workload was classified as HIGH (i.e. during the Hard period). Such result is fundamental for appreciating the following ones, because is a demonstration that actually the passive-BCI system was working as hypothesized.

3.3.2. Effectiveness of the passive-BCI system

The t-test results showed that the perceived workload of the user during the *AA Off* condition was not significantly higher ($p = 0.068$) in comparison to the *AA On* condition. It has to be underlined that the workload scores provided by the subjects referred to the whole AA condition, composed by the Easy, the Hard and the transition portion (Figure 40.a).

The t-test results showed that performances of operators during the AA On condition were significantly ($p = 0.045$) higher (WMRT index lower) than performances in the AA Off condition (Figure 40.b). Of course, higher reaction times reflect lower performances, and vice versa.

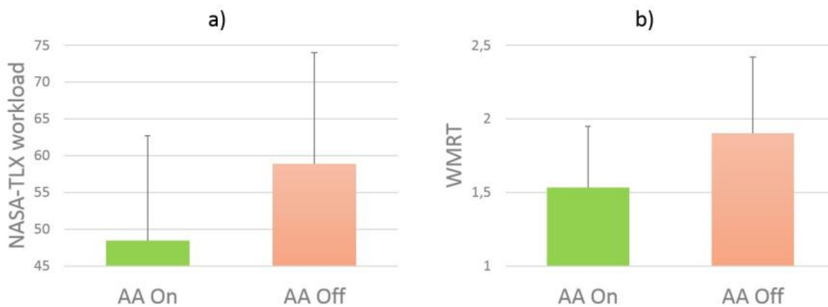


Figure 40. a) Vertical bars related to the subjective measure of the mental workload of the ATCOs, by using the NASA-TLX questionnaire. The results showed a not significant trend ($p=0.068$) between the two conditions (AA On and AA Off). **b)** Vertical bars related to the Weighted Reaction Time index (WMRT), reflecting behavioural performances of operators during the two conditions (AA On and AA Off). The results showed a significant ($p = 0.045$) increasing of task performances execution during AA On condition.

The t-tests showed a significant increasing ($p = 0.03$) of the W_{EEG} indexes distribution between the *Easy* and the *Hard* periods only for the *AA Off* condition. On the contrary no significant differences ($p = 0.65$) have been highlighted between the W_{EEG} indexes related to the *Easy* and the *Hard* slots during the *AA On* condition (Figure 41.a). In addition, a significant increment ($p = 0.04$) of the W_{EEG} indexes distributions related to the *Hard* slot of the *AA Off* condition with respect to the *Hard* slot of the *AA On* condition has been reported. In this regard, Figure 41.b shows the shape of the W_{EEG} distributions related to the *Hard* slot, for both the two conditions (*AA On/Off*). Instead, no significant trends (p = 0.95) have been highlighted between the *Easy* slots of the two AA conditions. In conclusion, Figure 41.c shows the time course of the W_{EEG} index related to the *Easy* and *Hard* slots, in both the two conditions (*AA On/Off*) together with the AA activation segments (Trigger) for a representative subject. The figure suggests that when the AA is activated, the W_{EEG} index related to the *AA On* condition decreases accordingly.

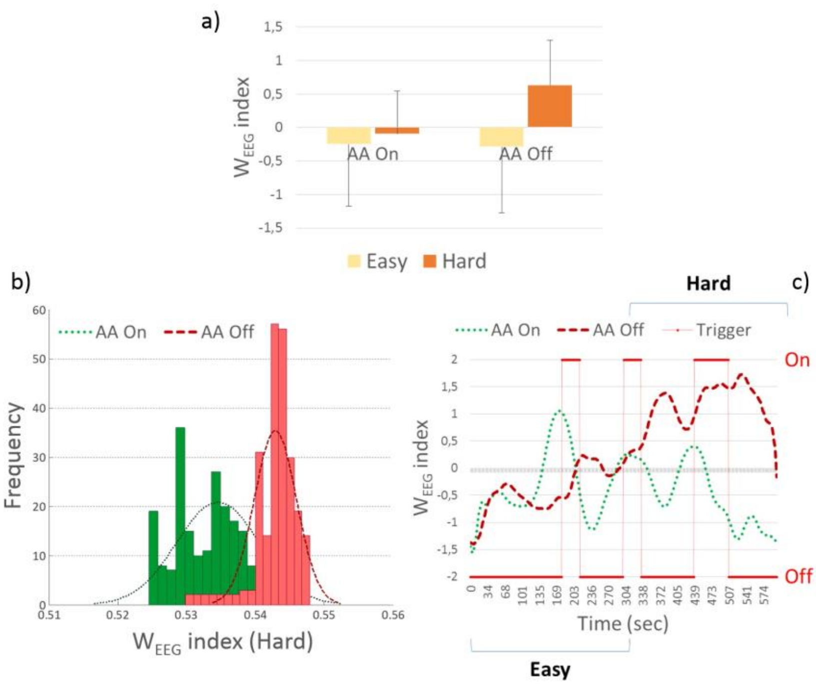


Figure 41. a) Vertical bars of the neurophysiological workload index distributions (W_{EEG}) related to the Easy and the Hard slots, during the conditions AA On and AA Off. **b)** Figure shows the shape of the W_{EEG} distributions related to the Hard slot, for both the two conditions (AA On/Off). **c)** Figure shows the time course of the W_{EEG} index related to the Easy and Hard slots, in both the two conditions (AA On/Off) together with the AA activation segments (Trigger) for a representative subject.

3.3.3. Discussion

Adaptive Automation is intended to help reducing human errors and increase safety. In this regard, AA solutions should be designed to prevent operators' under- and overload situations to ensure operators' attention stays engaged at all times (Scerbo, 1996). Parasuraman and colleagues (Parasuraman, Mouloua, & Molloy, 1996) discussed the role of psychophysiological measures in adaptive automation, presenting some experimental results in a laboratory setting. These results show how the approach can be particularly useful in the prevention of performance deterioration in both underload and overload conditions.

Passive BCI systems can be very useful to realize adaptive solutions, especially in the absence of overt behavioural responses (e.g. performances), since they are able to provide a continuous measure of the mental states, in our case mental workload, of the user by using his/her biosignals (e.g. EEG).

Although it is evident the possibility of evaluate operators' mental workload by using EEG-based measurements and the great potential of applying them through passive-BCI-based systems to trigger AA application, only few studies have been proposed in this regard. In addition, most of such studies have been performed in laboratory settings (Freeman, Mikulka, Prinzel, & Scerbo, 1999; Prinzel et al., 2000) or in real settings but in a very preliminary stage and without using passive-BCI technologies (Abbass, Tang, Amin, Ellejmi, & Kirby, 2014).

With this experiment I finally provided a positive answer to my initial question:

"*Is it possible to enhance the performance of an operator by deducing is mental workload from his brain activity and using the information to adapt the functionalities of the operative interface he is managing?*".

In this experiment a passive-BCI system, able to evaluate and classify online the operators' mental workload by using the EEG activity (W_{EEG}), has been implemented and validated. Depending on the classification result (LOW or HIGH), the system was able to trigger online the operator's interface, changing its behaviour by activating or deactivating few Adaptive Automation (AA) solutions.

The system has been integrated in an already existing ATM experimental platform. In particular, I expected that the proposed system was able (i) to trigger in the right way the ATM interface (i.e. the number of times in which the system activates the AA should be higher during the HARD scenario period in respect to the EASY one in order to prevent overload situation during the former and underload during the latter) and (ii) to induce a decreasing of the mental workload perceived by the operators when the adaptive solutions were activated.

Regarding the first point, results confirmed that the number of AA activations were significantly ($p=0.04$) higher during the hard scenario period in respect to the easy one. The behaviour of activating AA solutions only when the workload of the operator become high is an important issue, in fact if the AA solutions are improperly or even always activated, the workload of the user could decrease too much, inducing the "Complacency" phenomenon (Parasuraman, Molloy, & Singh, 1993) and in general underload conditions, that, as for the overload (e.g. if AA solutions are never activated), could deteriorate performance and decreasing safety. In this regard, results showed no differences in terms of mental workload between the EASY scenarios related to the AA On and AA Off conditions.

For the second point, results revealed a significant decreasing of the mental workload of the operator during the HARD periods in the conditions in which the passive-BCI system was able to activate the AA solutions (*AA On* condition) with respect to the condition without AA solutions (*AA Off*). Furthermore, in the *AA On* condition, the EEG–based mental workload (W_{EEG}) related to the HARD task was not significantly higher than the EASY related workload. On the contrary, when the AA solutions were not activated (*AA Off* condition), the HARD related mental workload was significantly higher than the EASY related workload. This behaviour was confirmed also by the subjective results: in particular, the NASA-TLX questionnaire revealed a perceived mental workload increasing in the scenarios when the AA solutions were not activated (*AA Off*) with respect to the scenarios in which the passive-BCI could activated the AA solutions (*AA On*).

4. MATERIAL AND METHODS

As done for the Results chapter, also this chapter has been divided in four main sections, in order to facilitate the reader in understanding the workflow of the methodology development.

In particular, the first section (Par. 4.1) aims to describe two innovative techniques, i.e. the *automatic-stop-StepWise Linear Discriminant Analysis* (asSWLDA) (Aricò et al., 2017; European Patent Nr. EP3143933 A1, 2017), and the *Regressive Eye BLINk Correction Algorithm* (REBLINCA) (Di Flumeri et al., 2016), developed to address specific issues related to the mental state evaluation in general, and in particular for online applications in real environments.

The second section (Par.4.2) describe the whole algorithm applied in the different applications in order to evaluate the human mental workload of users on the basis of their brain activity, measured by EEG.

The third section (Par. 4.3) is dedicated to the validation of such an algorithm in real operational environments: in particular three case studies have been investigated involving Air Traffic Controllers (ATCOs), car drivers in real drive conditions, and surgeons (students) managing robotic technologies in laparoscopic-surgery tasks.

The fourth and last section (Par. 4.4) is the final fulfilment of such PhD research activity, since it describes the application of the Mental Workload evaluation algorithm in a Passive Brain-Computer Interface with professional ATCOs in their workstation: in the Operative Interfaces of ATCOs adaptive automation solution have been implemented and their activation was triggered by online the Mental Workload Index of ATCOs.

4.1 Validation of two innovative techniques

4.1.1. Automatic-stop-StepWise Linear Discriminant Analysis (asSWLDA)

In neuroscientific field, machine learning approaches for mental states evaluation based on the analysis of neurophysiological data went through a rapid expansion in the last decades since such methodologies are able to provide the means to decode and characterize task relevant cognitive processes and brain states in general (i.e. reducing from multidimensionality to one dimensionality problem), and to distinguish them from non-informative brain signals (i.e. to enhance SNR). Machine learning is an indispensable tool for measuring users' mental states, especially when high dimensional data (i.e. EEG) are used. These techniques had been deeply investigated and successfully applied in different fields, such as image and speech recognition. With particular regard in BCI applications, different machine learning approaches have been developed for the real time decoding of mental activity of the user (e.g. (Lemm, Blankertz, Dickhaus, & Müller, 2011; Lotte, Congedo, Lécuyer, Lamarche, & Arnaldi, 2007; Jonathan R Wolpaw et al., 2002). The first issue to face in approaching to machine learning techniques in order to implement a pBCI application is which kind of algorithm to use. In fact, as stated before, there is a huge amount of techniques that can be employed, and the main objective of the experimenters would be to reach the maximum classification accuracy in evaluating the considered user's mental state. The first aspect to take into account is related to the linearity of the technique, in other words the kind of function (i.e. linear or not linear) that the specific algorithm uses to maximize the differences between classes. In general, it is accepted the assumption that simplicity should be preferred (J. R. Wolpaw et al., 2000). Focusing our interest on Mental Workload, the Human Factor concept investigated during my PhD research activity, it has been demonstrated how linear models outperform the non-linear ones (Berka et al., 2007).

But even choosing the proper models family, another issue to face is the properness of the features selection step. In fact, the model *training phase* is a crucial step, since during this phase the algorithm aims to extract from the neurophysiological data all those features that should be strictly related to the phenomenon, i.e. in our case the user's Mental Workload, we would like to measure. In this regard, the two main causes of poor performances in using machine learning approaches are "*overfitting*" and "*underfitting*" (Aricò et al., 2017). In particular, in the problem of overfitting, the model becomes less specific, including also features that are not strictly related to the phenomenon investigated, but maybe related to spurious differences between classes of the training set. The direct consequence is that the model does not perform well over time, and needs repeated calibration sessions (Vapnik, 2000), reducing its usability in operational environments. On the contrary, underfitting refers to a technique that can neither model the training data and generalize to new data. Intuitively, underfitting occurs when the model or the algorithm does not fit the data well enough (von Luxburg & Schoelkopf, 2008).

At the beginning of my research activity, I investigated the possibility to adopt, within the family of linear methods, the StepWise Linear Discriminant Analysis (SWLDA), since it was demonstrated to be one of the outperforming linear methods in measuring Mental Workload (Craven et al., 2006), and it had been successfully used to measure Mental Workload achieving high results in terms of classification accuracy (≈ 80%) and maintaining its performance stable over one week (Aricò et al., 2015). During my preliminary experiments, the SWLDA was successfully implemented within the algorithm for the Mental Workload estimation (see Par. 4.2), achieving high performance and high correlation (R = 0,9; p = 0,006) with subjective measures of Mental Workload (Di Flumeri et al., 2015).

However, there were still some issues that limited its employment in online applications in operational environments, in particular:

- A manual setting of model training parameters (α_{ENTER}, α_{REMOVE}, number of iterations) was needed;
- Once trained the model, the features remained stable only along one week, thus requiring a frequent recalibration;
- It was tested only in laboratory environments, with very controlled experimental conditions (possible source of artifacts such as subject movements, environmental variables and interferences, etc.), but in case of overfitting such artifacts could degrade the algorithm performance.

The optimum solution to these problems would have been to find out a criterion able to automatically stop the algorithm during the training phase (i.e. without any manual setting) when the best number of features, in other words the best features strictly related to the phenomenon, have been added to the model. In such a way, also the problem of under-/overfitting would be addressed, making the algorithm more efficient and robust, thus able to maintain its performance stable along periods longer than one week and also in applications affected by several artefacts sources, as operational environments.

Consequently, I developed a modified version of the standard SWLDA algorithm, called for obvious reasons the *automatic-stop-StepWise Linear Discriminant Analysis* (asSWLDA). Please let us note that such algorithm has been patented (European Patent EP3143933 A1, 2017)) In the following, the algorithm is described, as well as the experimental protocol used to validate the method, compared with the traditional SWLDA (Aricò, Borghini, Di Flumeri, Colosimo, Pozzi, et al., 2016).

4.1.1.1.The asSWLDA algorithm

The model training parameters tunable in the standard SWLDA algorithm are three: α_{ENTER}, α_{REMOVE}, $Iteration_{MAX}$. In this implementation, the first two parameters are left set at the default value as in the standard SWLDA implementation (i.e. $\alpha_{ENTER} = 0.05$, $\alpha_{REMOVE} = 0.1$). In fact, because of the probability ($pval_{ij}$) associated to each feature is strictly related to the actual iteration (in other words, to all the actual features in the models) this probability changes iteration by iteration, and it would result very difficult to impose a condition by using α_{ENTER} and α_{REMOVE}. In addition, even if no constrains on the α_{ENTER} and the α_{REMOVE} parameters are imposed, the features would be included in the model in order of significance (i.e. the first feature in the model will be the most significant one, and so on). On the contrary, the value of the $Iteration_{MAX}$ parameter will affect the reliability of the classifier over time (optimum classifier, underfitting or overfitting). As we stated before, this parameter should be chosen on the basis of the assumption:

$$\#Features_{UNDERFITTING} << \#Features_{OPTIMUM} << \#Features_{OVERFITTING}$$

In order to make the classifier able to find automatically the best $Iteration_{MAX}$ parameter, I took into account the p-value of the model (pModel), parameter available in the output of the standard SWLDA implementation, that gives information about the global significance of the model at the iteration j^{th}. The more the number of iterations increases (the more features are added to the model), the more the pModel value decreases (tending to zero) with a decreasing exponential shape. First of all, I collected the pModel values for all the iterations (pModel(#iter), Figure 42.a). At this point, we calculated the $\log_{10}$ of the pModel vector ($\log_{10}$(pModel(#iter)), Figure 42.b), and then the first-order differences between adjacent pModel elements (equation 1) that we called Convergence function, or Conv(#iter). We used the $\log_{10}$ function since we would have information about the size of pModel order, and after that about the differences between these pModel orders.

Finally, we plotted this vector as a function of #iter (Figure 42.c).

$$Conv(\#iter) = \log_{10}\left(pModel(\#iter+1)\right) - \log_{10}\left(pModel(\#iter)\right) \tag{1}$$

We identified as the best Iteration$_{MAX}$ value the number of iterations at which the Conv(#iter) assumed the lowest distance from the point (0,0), plus one (because we are working on the first-order differences, equations 2 and 3).

$$\left\| Conv(\#iter_{BEST}) \right\| = \min \left\| Conv(\#iter) \right\| \tag{2}$$

$$Iteration_{MAX} = \#iter_{BEST} + 1 \tag{3}$$

In fact, the best condition would be to have the least possible features and at the same time the convergence of the model (equation 4).

$$\log_{10}(pModel(\#iter+1)) - \log_{10}(pModel(\#iter)) = 0 \tag{4}$$

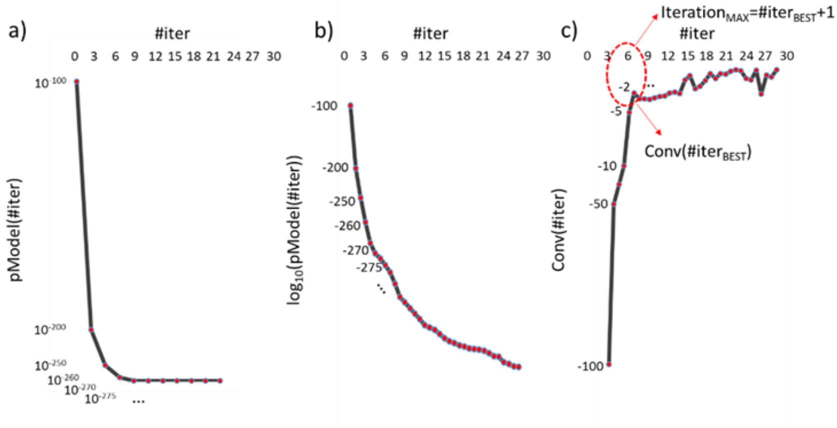


Figure 42. Representation of the a) pModel vector, the b) log10 of the pModel vector and the c) Conv function for each iteration, for a representative subject. In particular, in the figure (c) there are also showed i) the Conv(#iterBEST), in other words the lower distance of the Conv(#iter) function from the point (0,0) and ii) the correspondent Iteration$_{MAX}$, that is #iterBEST.

4.1.1.2.The experimental protocol

Twelve professional (40.41 ± 5.54 years) Air Traffic Controllers (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 experimental procedures involving human subjects described in this paper were approved by the Institutional Review Board.

ATCOs have been asked to perform Air Traffic Management (ATM) tasks in a high ecological setting, that is a functional simulated ATM environment developed and hosted at ENAC (Figure 43). The ATM simulator consisted in two screens, 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 Pseudo-Pilots, hosted in another room, who have interacted with the ATCOs by radio with the aim to simulate real-flight communications.

Figure 43. 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 minutes.

The complexity of the task could be modulated according to how many aircrafts 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*. For each session, ATCOs have been asked to perform a 45-minute ATM scenario enclosing three different levels of complexity (15 minutes for each complexity condition) associated to three different (EASY, MEDIUM, HARD) mental workload demands. The ATM scenarios have been designed with the aim to have the same level of complexity for each difficulty condition, to avoid any habituation or expectation effect, and then they have been validated and tested by a Subject-Matter Expert (SME) from ENAC. It has to be stressed that, because the ATM scenarios had to be as much realistic as possible, air traffic samples have never been constant, and the transitions between the difficulty levels was not sudden.

Neurophysiological data: For each ATCO, scalp EEG signal has been recorded by the digital monitoring BEmicro system (EBNeuro system) with a sampling frequency of 256 Hz by 8 Ag/AgCl electrodes (Fz, F3, F4, AF3, AF4, Pz, P3, P4) referenced to both the earlobes and grounded to the Cz electrode, according to the 10-20 International System. In addition, the vertical EOG signal has been recorded at the same time of 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.

Simultaneously to the execution of the ATM task, ATCOs have been asked to fill the *Instantaneous Self-Assessment* (ISA). In particular, the ISA (Kirwan, Scaife, & Kennedy, 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 task. The appeal of the ISA technique lies in its low intrusiveness and easy use, that is, no interferences with the main task. The ISA test scale has been presented to the ATCOs every 3 minutes in the form of a color-coded keypad on a screen placed on the left of the main monitor (Figure 43). The keypad flashed and sounded when the workload rating was required, and the participants simply pushed the button related to their workload perception. ATC Experts (SMEs) seated behind the ATCO and they have been asked to provide independent rate 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. ISA scores provided by experimental subjects and by the SMEs are named hereafter, respectively, SELF-ISA and SME-ISA.

4.1.1.3.The performed analysis

The analysis could be organized in two main sections, in order to face the two important issues above discussed: the reliability of the method and the accuracy of the measures obtained also in no-controlled settings, i.e. in operational environments, and the stability over time of such measures.

In addition, 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 operative environments, the less is the number of EEG channels, the smaller and less intrusive will be the EEG system. Also, this aspect has been investigated in terms of EEG channels needed by the algorithm.

Comparison between neurophysiological and subjective workload evaluation

This kind of analyses has been performed by considering only the last session (Day 30), where all the twelve ATCOs participated to the experiment.

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

In order to assess the accuracy of the proposed methodology for the mental workload assessment, in comparison with standard subjective workload measures (e.g. 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, W_{EEG} vs SME-ISA).

Performance stability over-time

Firstly, the two EEG recording sessions (Day 1, Day 30) have been comparted, 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–tailed paired T-test (α=0.05) have been performed, one for each difficulty level (EASY, MEDIUM, HARD), on the SELF-ISA and the SME-ISA scores in order to find out the difference between the two experimental sessions. Furthermore, Duncan post-hoc tests have been per-formed to assess significant differences between all pairs of levels of the considered factor.

After proving the compatibility of the two sessions (Day 1, Day 30), 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-validations type, where the training data belonged to Day 1 and the testing data to Day 30. The third couple of runs (E3, H3) of each session has been chosen to train the classifiers. In fact, since the ATM scenarios profile have 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, that 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).

A two-way repeated measures ANOVA (CI = .95) analysis has been performed on the AUC values, by considering as within factors the “classifiers” (asSWLDA, SWLDA) and the “cross-validation types” (Intra, Inter). Furthermore, Duncan post-hoc tests have been performed to assess significant differences between all pairs of levels of the considered factors.

EEG features selection analysis

Two–tailed paired T-tests (α=0.05) have been performed: the first to compare the number of total features among the considered algorithms, and 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).

Before every statistical analysis, the z-score transformation has been used to normalize the data.

4.1.2. Regressive Eye BLINk Correction Algorithm (REBLINCA)

The EEG signal is usually affected by electrical activity arising from sources other than the brain: such recorded activity that is not of cerebral origin is named *artefact*. In many applications, EEG analysis requires to remove the artefacts contributions with the compromise to keep as much information as possible, and to not affect the physiological information contained within the EEG. One of greatest nuisances are those artefacts resulting from oculomotor activity. While in general it is possible to avoid or filter out particular artefactual components, such as electromagnetic interferences, body movements, muscles contractions (Jung et al., 2000), the ocular artefacts are almost inevitable because subjects cannot well control spontaneous eye movements or blinks. Further, the instruction to inhibit eye movements or blinks may seriously distort brain activity (Croft & Barry, 2000). On the other hand, since they affect the EEG signal with a certain frequency, the rejection of the EEG epochs affected by these artefacts could cause a severe data loss. In this context, the goal would be to “clean” as better as possible the EEG data, at the same time both preventing data loss and avoiding interferences with subjects’ mental states (Croft & Barry, 2000). Several methods have been developed to handle this problem. The most popular and reliable approaches are the regression-based techniques and the Indipendent Component Analysis (ICA) (Urigüen & Garcia-Zapirain, 2015). The latter is very recent review, that shows how it is difficult to assess which is the best performing methodology (i.e. regression- or ICA-based one). Among the regression-based techniques, it has been widely demonstrated that the Gratton & Coles method (Gratton, Coles, & Donchin, 1983) is the outperforming one (Pham, Croft, Cadusch, & Barry, 2011; Urigüen & Garcia-Zapirain, 2015). On the other hand, among the ICA-based methods, the extended InfoMax (Lee, Girolami, & Sejnowski, 1999) and the SOBI

(Belouchrani, Abed-Meraim, Cardoso, & Moulines, 1997) appear to be the most performing ones (Romero, Mañanas, & Barbanoj, 2008; Urigüen & Garcia-Zapirain, 2015). Each method has its strength and weakness, and the choice of which method to adopt essentially depends on the specific application, on the basis of the limitations imposed by the method itself. For example, regression-based methods need to record the Electrooculographic (EOG) channel(s). In this regard, it has been demonstrated the mutual contamination between EEG and EOG channels (Romero et al., 2008), thus the regression techniques negatively affect EEG data in the blink-free signal segments. On the contrary, ICA-based techniques need great computational effort and proper sets (sufficient number and type) of EEG channels to reach acceptable performance, i.e. rejecting the component strictly related to eye blinks without causing EEG information loss. In addition, such techniques normally require the manual selection of the components to reject. Although many automatic-selection algorithms have been proposed recently, their performance are not comparable to the ICA-based traditional methods (Urigüen & Garcia-Zapirain, 2015).

In order to overcome the previous limitations, I developed an innovative regression-based algorithm, the Regressive Eye BLINk Correction Algorithm (REBLINCA) (Di Flumeri et al., 2016). The main hypothesis is to correct the eye blinks artefacts by subtracting their related component just in correspondence of blinks themselves, overcoming the EEG and EOG mutual contamination. In addition, the REBLINCA does not require any EOG channel, as the Gratton algorithm, reducing the invasiveness on the subject. Also, the REBLINCA does not require fixed or minimum EEG channels number, high computational effort, and any manual selection of parameters or components, as the ICA-based algorithms.

In order to test its efficiency and advantages, the REBLINCA has been compared with the three best available methods: Gratton, extended InfoMax and SOBI. Secondly, I implemented also an online version of the REBLINCA, whose performance have been compared with the Offline implementation. In the following, the two implementations, i.e. offline and online, of the REBLINCA, are described. Finally, the description of the analysis performed to validate the algorithm are provided.

4.1.2.1.The offline REBLINCA

The REBLINCA represents a modified version of the Gratton algorithm. Two are the main differences: (i) no EOG channel is requested, since the eye blinks template and the regression parameters are derived by an EEG (i.e. Fpz) channel; (ii) a threshold method has been integrated to identify the occurrence of the eye blinks, and to correct the EEG only in the corresponding samples. In this way, the influence of mutual contamination between the EEG and the channel used for the regression (i.e. Fpz) has been extremely reduced, compared to the classical Gratton algorithm, because the EEG is not manipulated apart on the eye blinks occurrence. In addition, as the eye blink electrical activity measured on the scalp is several orders greater than brain activity (Gasser, Sroka, & Möcks, 1985), we can hypothesize that only the blink contribution should be subtracted in correspondence of the blink. Two different components have to be obtained from the Fpz channel (Figure 44), one for the regression and the blink contribution subtraction (*Regr-FPZ*), and the second one for the blink detection (*Thres-FPZ*). *Regr-FPZ* signal is obtained by filtering the Fpz channel by a 5^{th} order Butterworth filter 1÷7 (Hz), as the Gratton algorithm does with the V-EOG channel. *Thres-FPZ* signal is obtained from *Regr-FPZ* (since it contains the components closely related to the blink activity) as follows:

1. The derivative of the *Regr-FPZ* signal is computed, in order to highlight the rapid potential increasing and decreasing typical of the blink electrical activity.
2. The resulting signal is normalized by computing its *z-score*, in order to center it on 0 and have unit variance.
3. The square of the signal is computed to straighten the negative parts of such signal, related to the descending phase of the eye blink electrical activity. Compared with the absolute value, the square has the advantage to enhance greater electrical activities (i.e. blinks).
4. Since 2 consecutive peaks (related to the ascending and descending phases of the blink) will appear for each eye blink,

a moving average is applied to the signal. The resulting signal is called *Thres-FPZ* (Figure1).

At this point, REBLINCA performs the following steps:

1. The B_n weights are estimated, that is the proportion of the *Regr-FPZ* signal present in the *n-th* EEG channel (time domain approach).
2. When the condition *Thres-FPZ >1* is verified (Eq. 1), the subtraction of the B_n-weighted *Regr-FPZ* signal from each n^{th} EEG channel is performed (Eq. 2). In other words, when the *Thres-FPZ* signal (Figure 44, green line) is greater than 1, the *Regr-FPZ* (blue line) is used for the subtraction, otherwise the corrected signal coincides with the raw one (Eq. 3). 1 is equal to the original unit variance of the signal, so it is possible to assess that the related probability of exceeding such threshold is equal to the 31,7%.

$$\textit{If}\ \ \text{Thres-FPZ}(i) > 1 \tag{1}$$

$$cEEG_n(i) = rEEG_n(i) - B_n * \text{Regr-FPZ}(i) \tag{2}$$

$$\textit{else}$$

$$cEEG_n(i) = rEEG_n(i) \tag{3}$$

where *cEEG* is the “corrected EEG”, *rEEG* the “raw EEG”, *i* the *i-th* signal sample and *n* the *n-th* EEG channel.

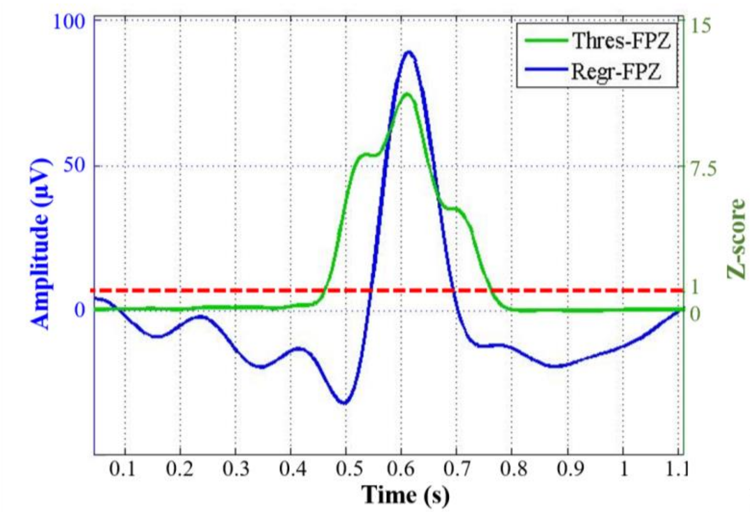


Figure 44. When the *Thres-FPZ* signal (green line, right y-axis) is higher than 1 (dotted red line), the *Regr-FPZ* signal (blue line, left y-axis) will be used for the eye blink contribution subtraction (dashed grey area).

4.1.2.2.The online REBLINCA

The final aim of my research activity was to develop an algorithm to evaluate online the user's Mental Workload. Two main issues impede to employ the REBLINCA in online applications:

1. The B_n weights were estimated by using the *Regr-Fpz* channel during the same recording session;
2. The *Regr-Fpz* channel was normalized by computing its z-score, in order to obtain the *Thres-Fpz* necessary for the blinks recognition threshold criterion.

Because of the assumption that, during the same EEG recording on the same subject (i.e. without any interruptions), some properties of the EEG signal, such as the gradient of the blinks propagation over the scalp, remain constants, or at least without significant variations (Gasser et al., 1985), the solution consist in recording a *Training session* for the algorithm, i.e. 60 seconds in a "Resting state" with open eyes and free to blink. At this point, the previous limitations could be respectively overcome as follows:

1. The weights are estimated by using the *Regr-Fpz* channel during the Training session;

2. The *Regr-Fpz* channel of the Testing session, i.e. during the online application, is normalized by using the *mean* and *standard deviation* values of the signal recorded during the Training session.

4.1.2.3.The experimental protocol

An already existent dataset was used to test the algorithm (Gianluca Borghini et al., 2016). In particular, ten male healthy normal or corrected to normal eyesight subjects (25±3 years old) were involved. Informed consent was obtained from each subject after the explanation of the study, approved by the local institutional ethical committee, and respecting the principles outlined in the Declaration of Helsinki of 1975 (revised in 2000). During these experiments the subjects had to perform the NASA Multi-Attribute Task Battery (Comstock, 1994).

For the validation of the offline REBLINCA, by comparing it with the selected *gold-standard* algorithms, i.e. Gratton, extended InfoMax and SOBI, the "Rest" condition recorded within the experiment was used. It consisted in looking at the center of a black screen for a minute, while seated and without doing anything nor avoiding eye blinks. In particular:

- The MATLAB implementation of the Gratton & Coles algorithm was used to reject the eye blinks contribution from the EEG recordings of each subject. The V-EOG channel was used for the regression. Hereafter, the results obtained by using this method will be labelled *"Gratton"*;
- Both the ICA-based algorithms were applied by using the EEGLAB [11] Matlab toolbox. Both the algorithms were applied taking into account all the 62 available EEG channels. Only one *Indipendent Component* (IC), related to the eye blinks activity, was manually selected for the rejection. The corrected EEG signal was reconstructed from the remaining ICs. Hereafter, the results obtained through such methods will be labelled respectively "*InfoMax*" and "*SOBI*";
- The offline REBLINCA has been applied as previously descripted (see Par. 4.1.2.1).

Secondly, the online REBLINCA performance have been compared with its offline implementation. In this case, the “Rest” condition was used as the *Training condition*, while the same experimental task of 3 minutes was used as the *Testing condition* for the online implementation. Instead, the offline implementation has been applied directly on the *Testing condition.*

4.1.2.4.The performed analysis

REBLINCA offline

Two kinds of analyses were performed: (i) to compare the performances of the four algorithms in removing eye-blinks contributions from the EEG signal; (ii) to test the effects of the four methods on the eye blinks-free EEG signal segments. It was chosen to report results on 8 representative EEG channels: Fz, AF3, AF4, F3, F4, Cz, Pz, Oz.

Eye blinks correction analysis: for each subject, the EEG segments containing eye blinks were manually selected (*“blink windows”*). For each “blink window”, the *Power Spectrum Density* (PSD) was computed for each EEG channel, and three averaged PSD values were considered. In particular, in Delta (1÷4 Hz), Theta (4÷8 Hz) and Alpha (8÷12 Hz) bands, averaged across all the “blink windows” for each subject. 24 one-way repeated-measurements ANOVAs (CI = 0.95) were performed on the PSD values for each EEG channel (8) and for each band (3) (independent variables), with the METHODS (4 levels: Gratton, ICA, SOBI and REBLINCA) as *within* factor. Duncan post-hoc tests were performed to assess any significant differences across the considered factors.

Blink-free EEG signal analysis: for each subject, the blink-free EEG signal was segmented into half second-long windows (“*no-blink windows”*), to be able to select a proper number (at least 10) of windows for each subject (longer is the window, more difficult is to find *blink-free* signal segments). The Pearson’s correlation coefficient between the Raw and the different Corrected EEG signals was estimated. Finally, the average of all the correlation coefficients along all the *no-blink windows* was computed for each channel, for each method. In this case, 8 one-way repeated-measurements ANOVAs (CI = 0.95) were performed on the mean correlation coefficients of each channel (8) (independent variables), with the METHODS as *within* factor (4 levels: Gratton, ICA, SOBI and REBLINCA). Duncan post-hoc tests were performed at the end of ANOVA analysis.

REBLINCA online

First, the Pearson's correlation analysis was performed between the weights used by the REBLINCA online, thus obtained by the Training condition, and offline, thus obtained by the Testing condition, in order to validate the experimental hypothesis (see Par. 4.1.2.2).

Also, the same previous analysis, i.e. the EEG signal PSD analysis on the "blink windows" and the correlation in the time domain on the "no-blink windows", has been performed. Since in this case there were just two methods, i.e. REBLINCA offline and online, on the same subjects, in both the cases a two-tails paired T-test has been performed.

4.2 The algorithm for the estimation of the EEG-based Mental Workload index

In the following, the whole algorithm able to evaluate the user's Mental Workload on the basis of his/her EEG activity is described. The same algorithm has been applied in all the applications described within Paragraphs 4.3 and 4.4.

The recoded EEG signal is firstly band-pass filtered with a fourth-order Butterworth filter (high-pass filter cut-off frequency: 1 Hz, low-pass filter cut-off frequency: 30 Hz). The Fpz channel is generally used to remove eyes-blink contributions from each channel of the EEG signal, by using the REBLINCA algorithm ([(Di Flumeri et al., 2016), please see Par. 4.1.2). This step is necessary 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 artefacts (i.e. operators normally communicate verbally and perform several movements during their operative activity), specific procedures of the EEGLAB toolbox have been used (Delorme & Makeig, 2004). Firstly, the EEG signal is segmented into epochs of 2 seconds (Epoch length), shifted of 0.125 seconds (Shift).

This windowing has been chosen with the compromise to have both a high number of observations, in comparison with the number of variables, 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 exceed ±100 μV (*Threshold criterion*) is marked as “artefact”. 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 10 μV/s, the considered epoch was marked as “artefact”. The last check has been based on the EEG Sample-to-sample difference (*Sample-to-sample criterion*). If such difference, in terms of absolute amplitude, was higher than 25 (μV), i.e. an abrupt variation (no-physiological) happened, the EEG epoch has been marked as “artefact”. At the end, the EEG epochs marked as “artefact” are removed from the EEG dataset with the aim to have a clean EEG signal from which estimate the brain parameters for the different analyses. All the previous mentioned values, have been chosen following the guidelines suggested by Delorme and Makeig (Delorme & Makeig, 2004).

From the clean EEG dataset, the Power Spectral Density (PSD) is calculated for each EEG channel for each epoch using a Hanning window of the same length of the considered epoch (2 seconds length, that means 0.5 Hz of frequency resolution). The application of a Hanning window helps 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 is defined for each ATCO by the estimation of the *Individual Alpha Frequency* (IAF) value (Wolfgang Klimesch, 1999). In order to have a precise estimation of the alpha peak and, hence of the IAF, the subjects is asked to keep the eyes closed for a minute before starting the experimental tasks: this experimental condition is later used for the IAF estimation. Finally, a spectral features matrix (EEG channels x Frequency bins) is obtained in the frequency bands directly correlated to the mental workload. In particular, only the theta band (IAF – 6 ÷ IAF – 2), over the EEG frontal channels, and the alpha band (IAF – 2 ÷ IAF + 2), over the EEG parietal channels, are been considered as variables for the mental workload evaluation (for further information about neurophysiological features of Mental Workload, please see Par. 1.2).

At this point, the features domain is ready to be treated by applying machine learning techniques, in particular the previous described asSWLDA (see Par. 4.1.1). On the basis of the training dataset, the asSWLDA selects the most relevant spectral features to discriminate the Mental Workload of the subjects within the different experimental conditions (i.e. for example EASY and HARD, or in general the two extremes of the mental state to evaluate). Once identified such spectral features, the asSWLDA assigns to each one specific weights ($w_{i\ train}$), plus a bias (b_{train}), such that the asSWLDA discriminant function ($y_{train}(t)$) takes the value *1* in the hardest condition and *0* in the easiest one. This step represents the *Training phase* of the classifier. Later on, the weights and the bias determined during the training phase are used to calculate the Linear Discriminant function ($y_{test}(t)$) over the testing dataset (*Testing phase*). Finally, a moving average of 8 seconds (8MA) is applied to the $y_{test}(t)$ function in order to smooth it out by reducing the variance of the measures, and I defined it as *EEG-based Workload index (W_{EEG})*. In this way, 8 seconds are the maximum time resolution of the *EEG-based Workload index.*

Here below are reported the training asSWLDA discriminant function (Equation 1, where $f_{i\ train}(t)$ represents the PSD matrix of the training dataset for the data window of the time sample t, and of the i^{th} feature), the testing one (Equation 2, where $f_{i\ test}(t)$ is as $f_{i\ train}(t)$ but related to the testing dataset) and the equation of the *EEG-based workload index, W_{EEG}* (Equation 3).

$$y_{train}(t) = \sum_i w_{i\ train} \cdot f_{i\ train}(t) + b_{train} \qquad (1)$$

$$y_{test}(t) = \sum_i w_{i\ train} \cdot f_{i\ test}(t) + b_{train} \qquad (2)$$

$$W_{EEG} = 8MA\ (y_{test}(t)) \qquad (3)$$

The whole workflow of the algorithm is represented in Figure 45.

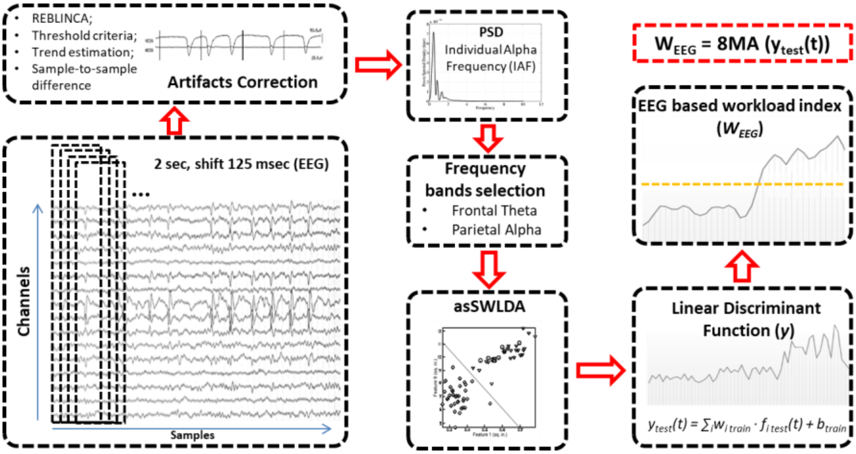


Figure 45. The figure describes the algorithm for the EEG-based Workload index estimation (W_{EEG}).

4.3 Application in operational environments

4.3.1. Case study 1: Air Traffic Controllers

4.3.1.1.The experimental protocol

Thirtyseven professional ATCOs from the *École Nationale de l'Aviation Civile* (ENAC) of Toulouse (France) have been involved in this study. They have been selected in order to have homogeneous experimental groups in terms of age and expertise. In particular, two groups of controllers have been defined, a group of ATC Experts (40.41 ± 5.54), and a group of ATC Students (23 ± 1.95). The experiment was conducted following the principles outlined in the Declaration of Helsinki of 1975, as revised in 2000. ATCOs have been asked to take part in a simulated ATC simulation in which the same ATM realistic scenario, lasting 45 minutes and shifting from a harder to an easier difficulty level, has been proposed on the real ATM platforms normally used at ENAC (Figure 46.a). Since the scenario starts with 15 minutes of very hard air-traffic load, before starting the experiments, 5 minutes have been given to the ATCOs to analyse the air-traffic situation and get ready to take position. In fact, in real ATC turn-over, the outcoming ATCO has to cooperate with the incoming ATCO before leaving the ATC position in order to give enough time to the latter to check the ongoing air-traffic situation and to understand the conditions of the different assumed aircrafts. After the hard condition (15 minutes), the air-traffic load becomes medium (15 minutes) and finally easy for the last 15 minutes. The main way to modulate the task difficulty consists in varying the number of aircrafts to manage: it was, on average, 21 at the beginning (hard condition) and then it starts to decrease, up to 15 in the medium condition and to 10 in the last (easy) condition. Also, the number and type of clearances required over the time, and the number/trajectory of other interfering flights contributed to model the mental demand along the scenario. The experiments have also been attended by two Pseudo-Pilots (Figure 46.b), placed in another room, who have interacted with the ATCOs with the aim to reproduce real communications, thus making the simulation as realistic as possible.

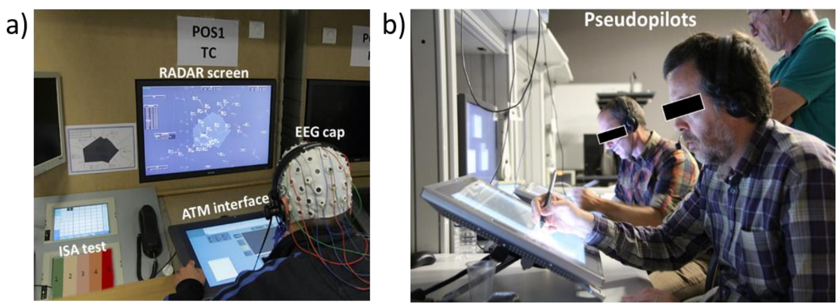


Figure 46. a) The ENAC simulator platform, composed of two screens, a 30" (RADAR) screen to display radar image and a 21" screen to interact with the radar image (ATM interface). On the little screen on the left bottom, the ISA test was proposed every 3 minutes. **b)** Pseudo-Pilots have interacted with the ATCOs with the aim to simulate real-flight communications.

During the task execution, the neurophysiological signals have been recorded by the digital monitoring *BEmicro* system (EBNeuro system). Fifteen electroencephalographic (EEG) channels (Fpz, F3, Fz, F4, AF3, AF4, P3, Pz, P4, P5, P6, POz, O1, Oz, O2), referenced to both the mastoids and grounded to the Cz electrode, according to the 10-20 International System. These data have been used to compute the EEG-based Mental Workload Index, as described in Par. 4.2.

Also, the ATM system recorded the ATCO's reaction times and number of airplanes assumed in each specific task condition (Easy, Medium and Hard). Since the ATCOs might adopt different strategies to manage the air-traffic, their reaction times (RT) have been weighted by taking into account the number of airplanes assumed, a global performance index, *Weighted Mean Reaction Time* (WMRT), has then been defined.

Finally, the ISA assessment (see for example Par. 4.1.1.2) has been included in the experimental protocol in order to have information about the ATCOs' mental workload perception, and to compare it with the neurophysiological (W_{EEG}) workload measure. The ISA score has been filled by the ATCOs every 3 minutes. The last ISA value has been excluded from the analysis because it has not been always presented exactly at the end of the simulation. For each group and for each run, the ISA scores have been averaged across the subjects.

4.3.1.2.The performed analysis

The dataset has been segmented in five parts for each difficulty level in order to have 5 Easy runs (E1, E2, E3, E4, E5), 5 Medium runs (M1, M2, M3, M4, M5) and Hard runs (H1, H2, H3, H4, H5). This segmentation has been used to compare the measures of the workload index provided by the neurophysiological (W_{EEG}) and the subjective (ISA) measurements. In fact, for each difficulty level five ISA scores have been provided by the trainee ATCOs. The third runs (E3, H3) has been chosen to train the classifier, and to test it over the remaining triplets of Easy, Medium and Hard conditions. Before every statistical analysis, the z-score transformation has been used to normalize the data.

<u>Performance analysis</u>

Before analyzing the Workload measures, the two experimental groups (Students and Experts) have been compared in terms of overall performance along the three difficulty levels. Two one-way repeated measures ANOVAs have been performed with the aim to assess the level of performance of the two groups, ATC Experts and ATC Students, across the task conditions (*within* factor: DIFFICULTY LEVEL; 3 levels: Easy, Medium and Hard), and between the groups (*between* factor RANK; 2 levels: Experts and Students). The independent variable was the WMRT in all such statistical analyses.

<u>Comparison between neurophysiological and subjective workload evaluation</u>

The W_{EEG} index has been evaluated for each difficulty level (Easy, Medium, and Hard) and group (ATC Students and ATC Experts) separately. for each group three two-tailed paired t-tests ($\alpha = 0.05$) between the W_{EEG} index distributions related to couple of conditions (Easy vs Medium, Easy vs Hard, Medium vs Hard) have been performed, in order to test the discriminability between different conditions. In order to test the discriminability between the different ATM conditions from a subjective point of view, the same kind of analysis, i.e. for each group three two-tailed paired t-tests ($\alpha = 0.05$) between the ISA score distributions related to couple of conditions (Easy vs Medium, Easy vs Hard and Medium vs Hard) has been performed.

Finally, the correlation between the subjective (ISA scores) and the neurophysiological (W_{EEG}) workload measures (reported above) has been quantified in terms of Pearson's correlation index. for each group (ATC Experts and Students) and for the whole sample.

4.3.2. Case study 2: Car drivers

4.3.2.1.The experimental protocol

Twenty male students (24.9 ± 1.7 years old, licensed from 6 ± 1 years) from the University of Bologna have been recruited, on a voluntary basis, in this study. They were selected in order to have a homogeneous experimental group in terms of age and expertise. The experiment was conducted following the principles outlined in the Declaration of Helsinki of 1975, as revised in 2000. Informed consent was obtained from each subject on paper, after the explanation of the study.

Two equal cars have been used for the experiments, i.e. Fiat 500L 1.3 Mjt, with diesel engine and manual transmission. The subjects had to drive the car along a route going through urban roads. In particular, the route consisted in three laps of a circuit about 2500 m long, to cover with the daylight (Figure 47.a). Although the subjects received a briefing before the experiment, explaining the whole protocol, the stuff and the route in detail, the first lap of all the subjects was considered as a “Lap of adaptation” to the experimental environment, thus it was not taken into account for the following analysis. The circuit was designed in order to have the possibility of identifying two segments of interest, both about 1000 m long, but supposed different in terms of elicited workload (Paxion, Galy, & Berthelon, 2014), thus named hereafter “EASY” and “HARD”: (i) EASY is a secondary road, mainly straight, with an intersection halfway with the right-of-way, one lane and low traffic capacity; (ii) HARD is a main road, mainly straight, with a roundabout halfway, three lanes and high traffic capacity. In this way, each subject performed two times (the first lap was not considered) an EASY driving condition and two times an HARD one. Brain activity by means of EEG techniques, and the eye movements by means of Eye-Tracking device, of each subject were collected, as you can see in Figure 47.b. The rationale was to use the EEG data, because of their specific suitability in evaluating human mental states, to test the EEG-based Workload index (Par. 4.2), while the Eye-Tracker to verify the experimental hypothesis, because of the insights achievable by its offline data analysis (de Winter et al., 2014). In fact, Eye-Tracking data allowed me to verify the properness of the experimental design, i.e. if the two driving conditions, one EASY and one HARD, were actually perceived different by the drivers (de Winter et al., 2014; Lantieri et al., 2015) (see Par. 4.3.2.1). Another added value of this study relies on the investigation of a real driving context. In this regard, all the works related to workload investigation by using EEG have been performed in simulator, or in poor realistic settings. It is important to prove the effectiveness of EEG-based metrics in real contexts, since it has been proven that same experimental tasks are perceived differently, in terms of mental workload, if performed in a simulator or in real environment (de Winter et al., 2014).

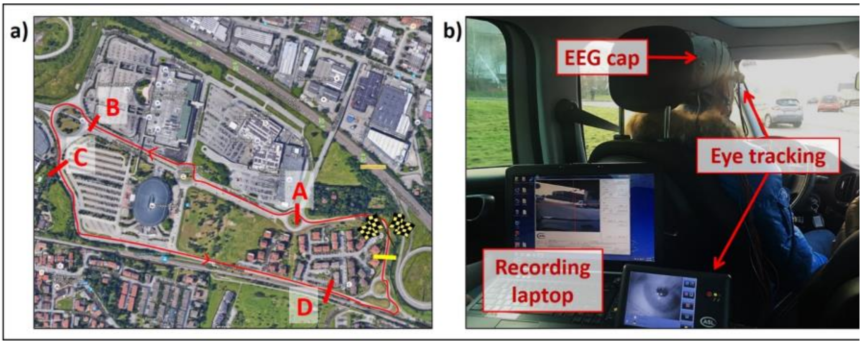


Figure 47. On the left (a) the circuit designed through urban roads, where the HARD segment is that one from A to B, the EASY that one from C to D. On the right (b), a picture within the car during the experimental recordings, showing the used instrumentation.

The EEG signals were recorded using the digital monitoring BEmicro system (EBNeuro, Italy). Twelve EEG channels (FPz, AF3, AF4, F3, Fz, F4, P3, P7, Pz, P4, P8 and P8), placed according to the 10-20 International System, were collected with a sampling frequency of 256 Hz, all referenced to both the earlobes, grounded to the Cz site, and with the impedances kept below 20 kΩ. These data have been used to compute the EEG-based Mental Workload Index, as described in Par. 4.2.

Eye movements of the experimental subjects were recorded through an ASL Mobile Eye-XG device, a system based on lightweight eyeglasses equipped with two digital high-resolution cameras. One camera recorded the scene image and the other the participant's eye. The data were recorded with a sampling rate of 30 Hz (i.e. 33 ms), and a resolution of 0.5-1°. ASL software was used to analyze the data, obtaining information about the drivers' fixation points frame by frame (33 ms). A preliminary calibration procedure was carried out for each subject within the car before starting driving, asking them to fix their gaze on thirty fixed visual points spread across the whole scene, in order to get a good accuracy of the eye-movement recorder. The gazes recorded during the driving task were grouped into three different categories: road infrastructure, traffic vehicles, and external environment. For each subject, each lap (second and third), and each condition (EASY and HARD) the distribution of eye fixations between the three categories was calculated in terms of percentage. In particular, the percentage of fixations over the external environment was investigated, since such indicator has been proven to be inversely correlated with mental workload: the more the experienced workload is, the less the number of fixations over the external environment is, since the driver gaze will mostly focus on infrastructure and vehicles (Costa, 1993; de Winter et al., 2014).

4.3.2.2.The performed analysis

Because of technical issues, the data of 4 subjects have not been used, thus the following analysis have been performed on 16 of the 20 subjects involved.

Analysis of the eye fixations of the subjects: the fixations percentages over the external environment for the EASY and HARD condition were averaged between the two investigated laps for each subject. A one-tailed paired t-test was performed in order to verify that the two segments of the circuit were correctly designed to reproduce an EASY and a HARD condition.

Analysis of the Area Under Curve (AUC) of the Receiver Operator Characteristic (ROC) curve of the classifier (Bamber, 1975). In particular, AUC represents a widely used methodology to test performance of a binary classifier: the classification performance can be considered good with an AUC higher than at least 0.7 (Fawcett, 2006). The classifier was tested shuffling the testing dataset, in order to verify that classifier performance on measure data (M*easured AUC*) were significantly higher than that one obtained on random data (R*andom AUC*). A one-tailed paired t-test was performed between *Measured* and *Random AUC*.

Analysis of the WL index, provided by the classifier: the obtained WL indexes were averaged between the two investigated laps for each condition and each subject. A one-tailed paired t-test was performed in order to verify differences between WL indexes, previously normalized by z-score in order to avoid possible subjective differences, related to EASY and HARD conditions.

4.3.3. Case study 3: Robot-assisted surgeons

4.3.3.1.The experimental protocol

Twenty healthy students (half male) from the Faculty of Medicine of the University of Verona have been recruited, on a voluntary basis, in this study. They were selected in order to have a homogeneous experimental group in terms of age and expertise. The experiment was conducted following the principles outlined in the Declaration of Helsinki of 1975, as revised in 2000. Informed consent was obtained from each subject on paper, after the explanation of the study.

The experimental protocol consisted in executing some laparoscopy-like tasks by using two different systems for the robot-assisted surgery: the *da Vinci Surgical system* (Intuitive Surgical Inc, USA), leader in the field of robotic surgery, and the *Actaeon Console* (BBZ srl, Italy), basically a cheaper simulator aimed to train students to use the da Vinci system (Figure 48).

Figure 48. The same student during the experiments performing the experimental tasks by the *da Vinci Surgical System* (on the left) and by the *BBZ Actaeon Simulator* (on the right).

In particular, after a training period aimed to make the subjects confidents with both the systems, they had to perform two Easy and two Hard versions of the TornWire task. Such a task, generally used to train the subjects to use this kind of devices, consisted in bringing a ring along a twisted wire (the wire passed in the hole of the ring), before in a way and after in the contrary way, in a maximum time of four minutes, without avoiding any contact between the ring and the wire itself. Each contact was considered by the system as error. The difficulty of the task depended on the severity of the wire twisting. The execution order of the four tasks (two easy and two hard tornwire tasks per subject) was randomized between the subjects to avoid any habituation effect.

During the whole experimental protocol, including the previous "Closed Eyes" and "Baseline" conditions, the EEG signals were recorded using the digital monitoring BEmicro system (EBNeuro, Italy). Twelve EEG channels (FPz, AF3, AF4, F3, Fz, F4, P3, P7, Pz, P4, P8 and P8), placed according to the 10-20 International System, were collected with a sampling frequency of 256 Hz, all referenced to both the earlobes, grounded to the Cz site, and with the impedances kept below 20 kΩ. These data have been used to compute the EEG-based Mental Workload Index, as described in Par. 4.2.

In addition, the subjective measures, in terms of perceived workload, have been collected. At the end of each execution, the subjects were asked to fill the NASA–TLX questionnaire, with the aim of collecting data about the perception of the workload across the training sessions. The total workload score was calculated as combination of six factors (Mental Demand, Physical Demand, Temporal Demand, Performance, Effort and Frustration), and it ranges from 0 to 100.

Finally, during the execution of the task, each system collected several parameters related to the control of the robots and correctness of the activities, taking into account also the time of execution, the task completion percentage and the number of errors, and produced a performance index from 0 % (the worst score) to 100 % (the best score).

Of course, the first aim of such experiment was, coherently with my PhD research activity, to validate the possibility of measuring user's mental workload in an operational environment on the basis of his/her brain, i.e. in my case EEG, activity.

Secondly, the experiment aimed to explore the possibility of employing such neurophysiological indicator, i.e. the EEG-based Mental Workload index, to address an open and wide debated issue in the field of ergonomics: the comparison of different systems from a cognitive point of view (Berguer & Smith, 2006; Gianluca Borghini et al., 2015). In this regard, the experimental question is: is it possible to assess if two different systems are perceived similar, in terms of cognitive demand, by the user's brain.

4.3.3.2.The performed analysis

The analyses have been organized in two sections, in order to answer the two experimental questions: the validation of the workload evaluation algorithm, and its application for the comparison of the two robotic systems, i.e. the Da Vinci robot and the BBZ simulator.

<u>Performance of the EEG-based workload evaluation algorithm</u>

Each subject, with each system, performed two Easy and two Hard Tornwire tasks. For each system, the asSWLDA algorithm (please see Par. 4.1.1) has been trained with a couple of Easy-Hard tasks and tested on the remaining couple. Also, in order to test the robustness of the algorithm towards the system used (i.e. the algorithm is sensitive to the task difficulty, or better said the mental workload, independently from the system), the algorithm has been trained crossing the training and testing dataset from the two systems: it was trained on a couple of Easy-Hard tasks performed with Da Vinci robot and tested on BBZ data, and vice-versa (labelled "Mixed" in the results, Par. 3.2.3.1). A repeated measures ANOVA has been performed on the AUC values obtained in the three cases (Da Vinci, BBZ and Mixed).

Also, to validate the algorithm performance, they have been compared, by three two-tailed paired Student's t-tests (one for each case), with the performance obtained by employing the algorithm on the testing dataset randomly shuffled.

<u>Comparison of different systems (BBZ vs Da Vinci)</u>

Once validated the algorithm for the estimation of the EEG-based Workload index (please see Par. 4.2), the neurophysiological scores have been used to explore the possibility to employ them to compare the two systems. Therefore, as well as with the subjective (NASA-TLX) and behavioural (performance) measures, the EEG-based workload scores obtained from the use of the two systems have been compared by three two-tailed paired Student's t-tests, one for each kind of measures (i.e. subjective, behavioural and neurophysiological).

4.4 Online passive Brain Computer Interface application

As previously discussed within the Introduction (please see Par. 1.3), despite the scientific evidences on the possibility of using neurophysiological measures (i.e. by using passive-BCI) to trigger Adaptive Automation (AA) solutions, only few examples have been proposed in this regard, the most of them in laboratory settings. In this context, this experiment, that is actually the final outcome of all my PhD research activity because of the application of all the methodologies developed, is remarkable for its innovation.

In this experiment, a real passive-BCI system has been fully integrated with a high realistic ATM simulator able to trigger adaptive solutions in real-time depending on the Mental Workload estimated by means of the ATCO’s brain activity. The expectation was that the passive-BCI system would be able to trigger the AA solutions and to reduce the task workload demand in order avoid both under- and overload conditions.

4.4.1. The experimental protocol

Twelve Air Traffic Controller (ATC) students (23 ± 2 years old) from the École Nationale de l’Aviation Civile (ENAC, Toulouse, France, one of the most important training schools for ATCOs and Pilots in the World) have been involved in this study. They were selected in order to have a homogeneous experimental group in terms of age and expertise. These students were finishing their three years training at ENAC. The experiment was conducted following the principles outlined in the Declaration of Helsinki of 1975, as revised in 2000. The study involved only healthy, normal subjects, recruiting on a voluntary basis. Subjects were free to accept or not to take part to the experimental protocol. All the recruited subjects accepted to participate to the study. Informed consent was obtained from each subject on paper, after the explanation of the study. No other individual information apart from the cerebral activity was gathered for the purpose of this study. Only aggregate information has been released while no individual information was diffused in any form.

The subjects have been asked to manage a functional interface that simulates a high realistic ATM scenario. The complexity of the task could be modulated according to how many aircrafts the ATCO had to control, the number and type of clearances required over the time and the number/trajectory of other interfering flights. Also, for the purposes of this study, specific AA solutions have been embedded in the ATM interface, with the aim to induce a decreasing in the operators' mental workload during high workload situations. These AA solutions were the result of several brainstorming sessions with subject matters experts (senior ATCOs), human factor and human computer interaction specialists. Briefly, they consisted in reduction or increasing of alerts and messages on the interface screen, highlighting of calling aircrafts, variation of graphical details in order to enhance risk perception.

EEG signal has been recorded and used online to evaluate the mental workload of the controllers. Such mental state has been used to trigger the RADAR screen interface by using the AA solutions described previously, only when the workload of the user become higher than the threshold defined during a specific calibration phase. Triggers depending on the actual mental workload of the user has been sent to the ATC interface by using a dedicated middleware developed at ENAC. Figure 49.a shows the platform architecture realized for the purpose of such experiment.

ATCO students that took part to the experiment were already well trained to use such interface. In particular, controllers have been asked to handle with two ATM scenarios under two different conditions (Easy and Hard). One scenario in which the AA could be triggered by the EEG-based Mental Workload index of the user (AA On), and the other one in which the operators' EEG-based Mental Workload index has been computed and stored, but not used to trigger the ATM interface (AA Off). Each scenario lasted 15 (min), the first 5 and the last 5 (min) have been designed to keep the task difficulty as constant as possible, low (Easy) and high (Hard) respectively. The middle part of the task (5 min) has been designed to simulate a realistic transition between the Easy and Hard segments, but it has not been used in the analysis (Figure 49.b). The two scenarios have been designed comparable in terms of complexity within the same difficulty levels (e.g. Easy of Scenario 1 and Easy of Scenario 2). The combinations of scenarios and conditions (AA On/Off) have been randomized to avoid any habituation effect and bias in the results. In addition, ATCO students have been asked to perform two traffic samples of 3 (min) before the execution of the experimental scenarios, respectively easy (Easy 0) and hard (Hard 0), to be used for the calibration of the EEG-based workload algorithm. Hereafter, these 3-minutes-long tasks will be named "*Calibration scenarios*", while the two consecutive 15-minutes-long scenarios "*Testing scenarios*".

At the end of each Testing scenario, Controllers have been asked to fill the NASA-Task Load indeX (NASA-TLX) questionnaire (Hart & Staveland, 1988), in order to evaluate their perceived mental workload in performing the different conditions (AA On/Off). NASA-TLX is a widely-used, multidimensional assessment tool that rates subjective perceived workload between 0 and 100.

For each subject, scalp EEG signal was recorded (g.USBamp, gTec, Austria) from 9 Ag/AgCl wet electrodes (Fpz, Fz, F3, F4, AF3, AF4, Pz, P3, P4) at 256 (Hz), referenced to both the mastoids and grounded to the Cz electrode, according to the 10-20 International System. These data have been used to compute the EEG-based Mental Workload Index, as described in Par. 4.2.

In addition to the neurophysiological (EEG-based Mental Workload Index) and subjective (NASA-TLX) measures of workload, also behavioural measures, in terms of performance, have been collected by the interface. In fact, the ATM interface recorded information about the reaction time and the number of airplanes the ATCO had assumed in each specific task condition (AA On/Off). Since the ATCOs might adopt different strategies to manage the air-traffic, their reaction times have been normalized on the number of airplanes assumed in the different phases of the simulation. The performance index, *Weighted Mean Reaction Time* (WMRT), has then been defined as the average of the weighted reaction times described previously.

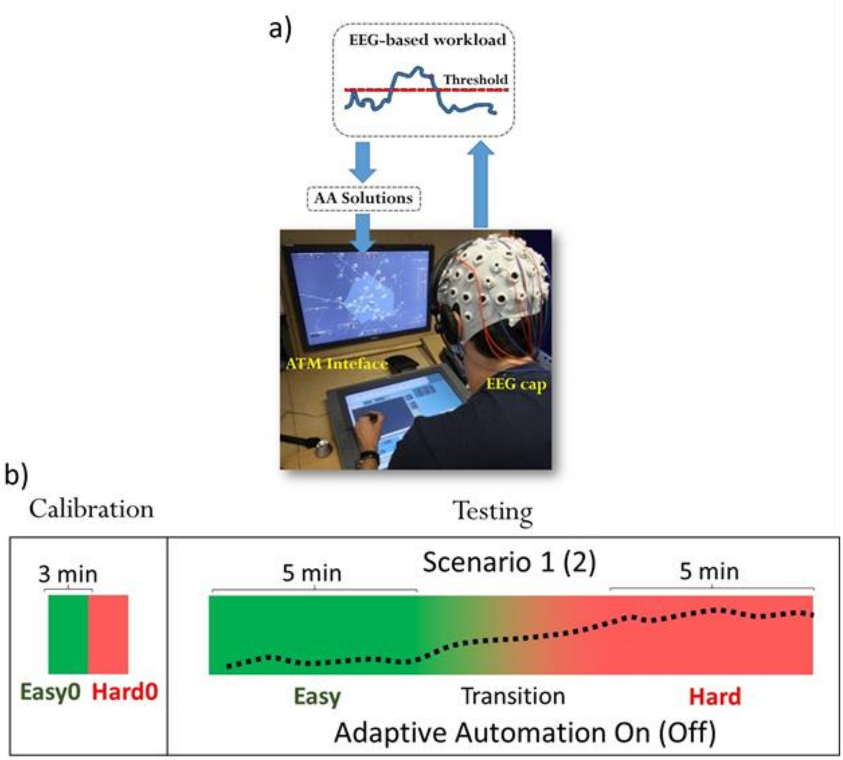


Figure 49. a) figure shows ATCO students wearing the EEG cap during the experiment and managing the ENAC platform, composed of two screens, a 30" (RADAR) screen to display radar image and a 21" screen to interact with the radar image (ATM interface). The mental workload of the user was evaluated online and specific AA solutions changed online the behaviour of the RADAR screen depending on the actual mental workload level. **b)** ATCO students have been asked to perform two ATM scenarios, one in which adaptive solutions could be triggered by the EEG mental workload index (AA On), and the other one in which adaptive automation has been disabled (AA Off). Presentation of each scenario and condition has been randomized to avoid any habituation and expectation effects.

4.4.2. The performed analysis

The data analyses have been performed with the aim to demonstrate the two experimental questions of my PhD research activity:

- Once validated the EEG-based Mental Workload index, it is possible to trigger the AA by using the online recognition of the actual mental workload of the user?
- Is such a kind of passive-BCI application effective in an operational environment? In other words, does the AA induce a reduction of the mental workload of the operator when it became high and consequently an increasing in performances execution of the task?

Triggering of Adaptive Automation solutions

To address the first experimental question, the comparison of the number of times in which the W_{EEG} activated the adaptive solutions in the Easy and Hard slots during the *AA On* condition has been performed by using a two-tailed paired t-test ($\alpha = 0.05$). In fact, the hypothesis was that the mental workload classifier was able to induce an activation of the AA more often during the Hard period with respect to the Easy one.

Effectiveness of the passive-BCI system

To investigate the second experimental question, the analyses on subjective and neurophysiological workload measurements along the two conditions (AA On/Off) and difficulty levels (Easy and Hard), and behavioural performances achieved by the operators in the two conditions, has been performed.

Subjective workload assessment: A two-tailed paired t-test ($\alpha = 0.05$) has been performed on the NASA-TLX scores to investigate differences between the two AA conditions (AA On/Off) in terms of workload perception.

Behavioural performance assessment: A two-tailed paired t-test ($\alpha = 0.05$) has been performed between WMRT indexes over the two conditions (AA On/Off).

Neurophysiological workload assessment: I compared the W_{EEG} indexes referred to the two difficulty levels (Easy, Hard) *within* and *between* the two AA conditions (AA On/Off). In particular, I performed 4 two-tailed paired t-test ($\alpha = 0.05$) to compare difficulty levels *within* the same AA condition (i.e. AA On: Easy vs Hard; AA Off: Easy vs Hard), and to compare the two AA conditions *within* the same difficulty level (Easy: AA On vs AA Off; Hard: AA On vs AA Off).

Before every statistical analysis, the *z-score* transformation has been used to normalize the data.

5. CONCLUSION

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 behavioural or subjective measures. However, the application of the passive-BCI technology outside research laboratories (e.g. operational environments) has to face several practical issues, such as the reliability over time of the measure (unless a frequent recalibration of the system), the intrusiveness of the equipment (e.g. EEG sensors and recording system), the robustness towards different artefacts sources, and the necessity of reliability in general in different operative conditions and, consequently, in different environments.

With the present PhD thesis work, I tried to address the various issues related to the use of EEG-based neurophysiological indicators of Mental Workload to develop a passive-BCI system usable in real operational environments.

First, specific and innovative algorithms, the REBLINCA (published) and the asSWLDA (patented), have been developed to face the problems of artefacts, EEG channels reduction, algorithm robustness and stability over-time.

Second, a method able to provide EEG-based Mental Workload Evaluation of the operator has been developed and successfully validated in three different operational environments, i.e. the Air Traffic Management, the car driving and the robot-assisted surgery.

Finally, as the final fulfilment of such PhD research activity, the method has been implemented in a real passive Brain-Computer Interface with professional ATCOs in their workstation: in the ATCOs' operative interfaces the activation of adaptive automation solutions was triggered online by the Mental Workload Index of ATCOs themselves

Of course, further investigations are needed in order to validate the method also in other operational environments and in different applications, as well as on wider subjects' sample. Furthermore, it would be interesting to apply such method, or better said such approach, with other important mental states in the Human Factors research, such as stress, situation awareness, drowsiness, and so on. However, the results obtained could be considered really innovative and they put this study on the cutting-edge of the applied neuroscience.

REFERENCES

Abbass, H. A., Tang, J., Amin, R., Ellejmi, M., & Kirby, S. (2014). Augmented Cognition using Real-time EEG-based Adaptive Strategies for Air Traffic Control. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *58*(1), 230–234. https://doi.org/10.1177/1541931214581048

Aberg, L., & Rimmö, P. A. (1998). Dimensions of aberrant driver behaviour. *Ergonomics*, *41*(1), 39–56. https://doi.org/10.1080/001401398187314

Aloise, F., Aricò, P., Schettini, F., Riccio, A., Salinari, S., Mattia, D., … Cincotti, F. (2012). A covert attention P300-based brain-computer interface: Geospell. *Ergonomics*, *55*(5), 538–551. https://doi.org/10.1080/00140139.2012.661084

Aricò, P., Borghini, G., Di, F. G., & Babiloni, F. (2017, March 22). *EP3143933 A1*. Retrieved from http://www.google.it/patents/EP3143933A1

Aricò, P., Borghini, G., Di Flumeri, G., Colosimo, A., Bonelli, S., Golfetti, A., … Babiloni, F. (2016). Adaptive Automation Triggered by EEG-Based Mental Workload Index: A Passive Brain-Computer Interface Application in Realistic Air Traffic Control Environment. *Frontiers in Human Neuroscience*, 539. https://doi.org/10.3389/fnhum.2016.00539

Aricò, P., Borghini, G., Di Flumeri, G., Colosimo, A., Graziani, I., Imbert, J. P., … Babiloni, F. (2015). Reliability over time of EEG-based mental workload evaluation during Air Traffic Management (ATM) tasks. *Conference Proceedings: Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference*, *2015*, 7242–7245. https://doi.org/10.1109/EMBC.2015.7320063

Aricò, P., Borghini, G., Di Flumeri, G., Colosimo, A., Pozzi, S., & Babiloni, F. (2016). A passive Brain-Computer Interface (p-BCI) application for the mental workload assessment on professional Air Traffic Controllers (ATCOs) during realistic ATC tasks. *Progress in Brain Research, in Press*.

Aricò, P., Borghini, G., Di Flumeri, G., Sciaraffa, N., Colosimo, A., & Babiloni, F. (2017). Passive BCI in Operational Environments: Insights, Recent Advances, and Future Trends. *IEEE Transactions on Biomedical Engineering*, *64*(7), 1431–1436. https://doi.org/10.1109/TBME.2017.2694856

Arico, P., Borghini, G., Flumeri, G. D., Bonelli, S., Golfetti, A., Graziani, I., … Babiloni, F. (2017). Human Factors and Neurophysiological Metrics in Air Traffic Control: a Critical Review. *IEEE Reviews in Biomedical Engineering*, *PP*(99), 1–1. https://doi.org/10.1109/RBME.2017.2694142

Arico, Pietro, Borghini, G., Graziani, I., Taya, F., Sun, Y., Bezerianos, A., … Babiloni, F. (2014). Towards a multimodal bioelectrical framework for the online mental workload evaluation. In *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 3001–3004). https://doi.org/10.1109/EMBC.2014.6944254

Baldwin, C. L. (2003). Commentary. *Theoretical Issues in Ergonomics Science*, *4*(1–2), 132–141. https://doi.org/10.1080/14639220210159807

Bamber, D. (1975). The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. *Journal of Mathematical Psychology*, *12*(4), 387–415. https://doi.org/10.1016/0022-2496(75)90001-2

Belouchrani, A., Abed-Meraim, K., Cardoso, J. F., & Moulines, E. (1997). A blind source separation technique using second-order statistics. *IEEE Transactions on Signal Processing*, *45*(2), 434–444. https://doi.org/10.1109/78.554307

Berguer, R., & Smith, W. (2006). An Ergonomic Comparison of Robotic and Laparoscopic Technique: The Influence of Surgeon Experience and Task Complexity. *Journal of Surgical Research*, *134*(1), 87–92. https://doi.org/10.1016/j.jss.2005.10.003

Berka, C., Levendowski, D. J., Lumicao, M. N., Yau, A., Davis, G., Zivkovic, V. T., … Craven, P. L. (2007). EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks. *Aviation, Space, and Environmental Medicine*, *78*(5 Suppl), B231-244.

Blankertz, B., Tangermann, M., Vidaurre, C., Fazli, S., Sannelli, C., Haufe, S., … Müller, K.-R. (2010). The Berlin Brain–Computer Interface: Non-Medical Uses of BCI Technology. *Frontiers in Neuroscience*, *4*. https://doi.org/10.3389/fnins.2010.00198

Blinowska, K., & Durka, P. (2006). Electroencephalography (EEG). In M. Akay (Ed.), *Wiley Encyclopedia of Biomedical Engineering*. Hoboken, NJ, USA: John Wiley & Sons, Inc. https://doi.org/10.1002/9780471740360.ebs0418

Borghini, G., Astolfi, L., Vecchiato, G., Mattia, D., & Babiloni, F. (2014). Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. *Neuroscience and Biobehavioral Reviews*, *44*, 58–75. https://doi.org/10.1016/j.neubiorev.2012.10.003

Borghini, Gianluca, Aricò, P., Di Flumeri, G., & Babiloni, F. (2017). *Industrial Neuroscience in Aviation: Evaluation of Mental States in Aviation Personnel*. Cham: Springer International Publishing Springer.

Borghini, Gianluca, Aricò, P., Di Flumeri, G., Salinari, S., Colosimo, A., Bonelli, S., … Babiloni, F. (2015). Avionic Technology Testing by using a Cognitive Neurometric Index: A Study with Professional Helicopter Pilots. *Conference Proceedings: Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference*, *2015*, 6182–6185. https://doi.org/10.1109/EMBC.2015.7319804

Borghini, Gianluca, Aricò, P., Graziani, I., Salinari, S., Sun, Y., Taya, F., … Babiloni, F. (2016). Quantitative Assessment of the Training Improvement in a Motor-Cognitive Task by Using EEG, ECG and EOG Signals. *Brain Topography*, *29*(1), 149–161. https://doi.org/10.1007/s10548-015-0425-7

Borghini, Gianluca, Astolfi, L., Vecchiato, G., Mattia, D., & Babiloni, F. (2014). Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. *Neuroscience and*

Biobehavioral Reviews, *44*, 58–75. https://doi.org/10.1016/j.neubiorev.2012.10.003

Boucsein, W., & Backs. (2000). *Engineering Psychophysiology: Issues and Applications*. CRC Press.

Brookings, J. B., Wilson, G. F., & Swain, C. R. (1996). Psychophysiological responses to changes in workload during simulated air traffic control. *Biological Psychology*, *42*(3), 361–377.

Buzsáki, G. (2002). Theta Oscillations in the Hippocampus. *Neuron*, *33*(3), 325–340. https://doi.org/10.1016/S0896-6273(02)00586-X

Byrne, E. A., & Parasuraman, R. (1996). Psychophysiology and adaptive automation. *Biological Psychology*, *42*(3), 249–268. https://doi.org/10.1016/0301-0511(95)05161-9

Cain, B. (2007). *A Review of the Mental Workload Literature*.

Calabrese, E. J. (2008). Neuroscience and hormesis: overview and general findings. *Critical Reviews in Toxicology*, *38*(4), 249–252. https://doi.org/10.1080/10408440801981957

Cantero, J. L., Atienza, M., Stickgold, R., Kahana, M. J., Madsen, J. R., & Kocsis, B. (2003). Sleep-Dependent θ Oscillations in the Human Hippocampus and Neocortex. *Journal of Neuroscience*, *23*(34), 10897–10903.

Chi, Y. M., Wang, Y.-T., Wang, Y., Maier, C., Jung, T.-P., & Cauwenberghs, G. (2012). Dry and Noncontact EEG Sensors for Mobile Brain–Computer Interfaces. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *20*(2), 228–235. https://doi.org/10.1109/TNSRE.2011.2174652

Comstock, J. R. (1994). MATB - Multi-Attribute Task Battery for human operator workload and strategic behavior research. Retrieved from http://ntrs.nasa.gov/search.jsp?R=19940002837

Cosgrove, K. P., Mazure, C. M., & Staley, J. K. (2007). Evolving Knowledge of Sex Differences in Brain Structure, Function, and Chemistry. *Biological Psychiatry*, *62*(8), 847–855. https://doi.org/10.1016/j.biopsych.2007.03.001

Costa, G. (1993). Evaluation of workload in air traffic controllers. *Ergonomics*, *36*(9), 1111–1120. https://doi.org/10.1080/00140139308967982

Craven, P. L., Belov, N., Tremoulet, P., Thomas, M., Berka, C., Levendowski, D., & Davis, G. (2006). Cognitive Workload Gauge Develpment: Comparison of Real-time Classification Methods. *Augmented Cognition: Past, Present and Future*.

Croft, R. J., & Barry, R. J. (2000). Removal of ocular artifact from the EEG: a review. *Neurophysiologie Clinique/Clinical Neurophysiology*, *30*(1), 5–19. https://doi.org/10.1016/S0987-7053(00)00055-1

de Winter, J. C. F., Happee, R., Martens, M. H., & Stanton, N. A. (2014). Effects of adaptive cruise control and highly automated driving on workload and situation awareness: A review of the empirical evidence. *Transportation Research Part F: Traffic Psychology and Behaviour*, *27*, 196–217. https://doi.org/10.1016/j.trf.2014.06.016

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Desmond, P. A., & Hancock, P. A. (2001). Active And Passive Fatigue States. *Stress, Workload And Fatigue*. Retrieved from /view.aspx?id=683362

Di Flumeri, G., Arico, P., Borghini, G., Colosimo, A., & Babiloni, F. (2016). A new regression-based method for the eye blinks artifacts correction in the EEG signal, without using any EOG channel (pp. 3187–3190). IEEE. https://doi.org/10.1109/EMBC.2016.7591406

Di Flumeri, G., Borghini, G., Aricò, P., Colosimo, A., Pozzi, S., Bonelli, S., … Babiloni, F. (2015). On the Use of Cognitive Neurometric Indexes in Aeronautic and Air Traffic Management Environments. In B. Blankertz, G. Jacucci, L. Gamberini, A. Spagnolli, & J. Freeman (Eds.), *Symbiotic Interaction* (Vol. 9359, pp. 45–56). Cham: Springer

International Publishing. Retrieved from http://link.springer.com/10.1007/978-3-319-24917-9_5
Dooley, C. (2009). The Impact of Meditative Practices on Physiology and Neurology: A Review of the Literature. *Scientia Discipulorum*, *4*(1). Retrieved from http://digitalcommons.plattsburgh.edu/scientia_discipulorum/vol4/iss1/3
Dorneich, M. C., Ververs, P. M., Mathan, S., & Whitlow, S. D. (2005). A joint human-automation cognitive system to support rapid decision-making in hostile environments. In *2005 IEEE International Conference on Systems, Man and Cybernetics* (Vol. 3, p. 2390–2395 Vol. 3). https://doi.org/10.1109/ICSMC.2005.1571506
Eggemeier, F. T., Wilson, G. F., Kramer, A. F., & Damos, D. L. (1991). Workload assessment in multi-task environments. In *Multiple-task performance* (D.L. Damos (Ed.), pp. 207–216). London: Taylor&Francis.
Elul, R. (1969). Gaussian behavior of the electroencephalogram: changes during performance of mental task. *Science (New York, N.Y.)*, *164*(3877), 328–331.
Fawcett, T. (2006). An Introduction to ROC Analysis. *Pattern Recogn. Lett.*, *27*(8), 861–874. https://doi.org/10.1016/j.patrec.2005.10.010
Feyer, A., & Williamson, A. M. (2011). Human Factors in Accident Modelling. In *Encyclopaedia of Occupational health and safety*. Retrieved from about:reader?url=http%3A%2F%2Fwww.iloencyclopaedia.org%2Fpart-viii-12633%2Faccident-prevention%2F92-56-accident-prevention%2Fhuman-factors-in-accident-modelling
Fisher, R. A. (1921). On the “Probable Error” of a Coefficient of Correlation Deduced from a Small Sample. Retrieved from https://digital.library.adelaide.edu.au/dspace/handle/2440/15169
Freeman, F. G., Mikulka, P. J., Prinzel, L. J., & Scerbo, M. W. (1999). Evaluation of an adaptive automation system using three EEG indices with a visual tracking task. *Biological Psychology*, *50*(1), 61–76.

Fuchs, S., Hale, K. S., Stanney, K. M., Juhnke, J., & Schmorrow, D. D. (2007). Enhancing Mitigation in Augmented Cognition. *Journal of Cognitive Engineering and Decision Making*, *1*(3), 309–326. https://doi.org/10.1518/155534307X255645

Gasser, T., Sroka, L., & Möcks, J. (1985). The transfer of EOG activity into the EEG for eyes open and closed. *Electroencephalography and Clinical Neurophysiology*, *61*(2), 181–193. https://doi.org/10.1016/0013-4694(85)91058-2

Gevins, A, & Smith, M. E. (2003). Neurophysiological measures of cognitive workload during human-computer interaction. *Theoretical Issues in Ergonomics Science*, *4*(1–2).

Gevins, A., Smith, M. E., Leong, H., McEvoy, L., Whitfield, S., Du, R., & Rush, G. (1998). Monitoring working memory load during computer-based tasks with EEG pattern recognition methods. *Human Factors*, *40*(1), 79–91.

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. *Cerebral Cortex (New York, N.Y.: 1991)*, *7*(4), 374–385.

Giraudet, L., Imbert, J.-P., Bérenger, M., Tremblay, S., & Causse, M. (2015). The neuroergonomic evaluation of human machine interface design in air traffic control using behavioral and EEG/ERP measures. *Behavioural Brain Research*, *294*, 246–253. https://doi.org/10.1016/j.bbr.2015.07.041

Göhring, D., Latotzky, D., Wang, M., & Rojas, R. (2013). Semi-autonomous Car Control Using Brain Computer Interfaces, 393–408. https://doi.org/10.1007/978-3-642-33932-5_37

Gopher, D., & Donchin, E. (1986). Workload: An examination of the concept. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance, Vol. 2: Cognitive processes and performance* (pp. 1–49). Oxford, England: John Wiley & Sons.

Gratton, G., Coles, M. G., & Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalography and Clinical Neurophysiology*, *55*(4), 468–484.

Gundel, A., & Wilson, G. F. (1992). Topographical changes in the ongoing EEG related to the difficulty of mental tasks. *Brain Topography*, *5*(1), 17–25. https://doi.org/10.1007/BF01129966

Hagemann, K. (2008). The alpha band as an electrophysiological indicator for internalized attention and high mental workload in real traffic driving. [Dissertation]. Retrieved January 8, 2014, from http://docserv.uni-duesseldorf.de/servlets/DocumentServlet?id=8318

Harris, F. J. (1978). On the Use of Windows for Harmonic Analysis With the Discrete Fourier Transform. *Proceedings of the IEEE*, *66*(1), 51–83. https://doi.org/10.1109/PROC.1978.10837

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In Peter A. Hancock and Najmedin Meshkati (Ed.), *Advances in Psychology* (Vol. Volume 52, pp. 139–183). North-Holland. Retrieved from http://www.sciencedirect.com/science/article/pii/S0166411508623869

Helmreich, R. L. (2000). On error management: lessons from aviation. *BMJ (Clinical Research Ed.)*, *320*(7237), 781–785.

Hori, T., Sugita, Y., Koga, E., Shirakawa, S., Inoue, K., Uchida, S., … Fukuda, N. (2001). Proposed supplements and amendments to " *A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects* ", the Rechtschaffen & Kales (1968) standard. *Psychiatry and Clinical Neurosciences*, *55*(3), 305–310. https://doi.org/10.1046/j.1440-1819.2001.00810.x

Jahnsen, H., & Llinás, R. (1984). Electrophysiological properties of guinea-pig thalamic neurones: an in vitro study. *The Journal of Physiology*, *349*(1), 205–226. https://doi.org/10.1113/jphysiol.1984.sp015153

Jasper, H. (1958). The ten twenty electrode system of the international federation. *Electroencephalography and Clinical Neurophysiology*, *10*, 371–375.

Jaušovec, N., & Jaušovec, K. (2012). Working memory training: improving intelligence--changing brain activity. *Brain and*

Cognition, *79*(2), 96–106. https://doi.org/10.1016/j.bandc.2012.02.007

Jensen, O., Kaiser, J., & Lachaux, J.-P. (2007). Human gamma-frequency oscillations associated with attention and memory. *Trends in Neurosciences*, *30*(7), 317–324. https://doi.org/10.1016/j.tins.2007.05.001

Jones, E. G., & Mendell, L. M. (1999). Assessing the Decade of the Brain. *Science*, *284*(5415), 739–739. https://doi.org/10.1126/science.284.5415.739

Jung, T.-P., Makeig, S., Humphries, C., Lee, T.-W., McKEOWN, M. J., Iragui, V., & Sejnowski, T. J. (2000). Removing electroencephalographic artifacts by blind source separation. *Psychophysiology*, *37*(2), 163–178.

Kirsh, D. (2000). A Few Thoughts on Cognitive Overload. *Intellectica*, *30*.

Kirwan, B. (1998). Human error identification techniques for risk assessment of high risk systems--Part 1: Review and evaluation of techniques. *Applied Ergonomics*, *29*(3), 157–177.

Kirwan, B., Scaife, R., & Kennedy, R. (2001). Investigating complexity factors in UK air traffic management. *Human Factors and Aerospace Safety*, *1*(2). Retrieved from http://trid.trb.org/view.aspx?id=717648

Klimesch, W., Doppelmayr, M., Pachinger, T., & Ripper, B. (1997). Brain oscillations and human memory: EEG correlates in the upper alpha and theta band. *Neuroscience Letters*, *238*(1–2), 9–12.

Klimesch, Wolfgang. (1999). EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain Research Reviews*, *29*(2–3), 169–195. https://doi.org/10.1016/S0165-0173(98)00056-3

Klimesch, Wolfgang. (2012). α-band oscillations, attention, and controlled access to stored information. *Trends in Cognitive Sciences*, *16*(12), 606–617. https://doi.org/10.1016/j.tics.2012.10.007

Kohlmorgen, J., Dornhege, G., Braun, M., Blankertz, B., Müller, K.-R., Curio, G., … Kincses, W. (2007). Improving human performance in a real operating environment through real-time mental workload detection [Book Section].

Lantieri, C., Lamperti, R., Simone, A., Costa, M., Vignali, V., Sangiorgi, C., & Dondi, G. (2015). Gateway design assessment in the transition from high to low speed areas. *Transportation Research Part F: Traffic Psychology and Behaviour*, *34*, 41–53. https://doi.org/10.1016/j.trf.2015.07.017

Lawton, R., & Ward, N. J. (2005). A systems analysis of the Ladbroke Grove rail crash. *Accident; Analysis and Prevention*, *37*(2), 235–244. https://doi.org/10.1016/j.aap.2004.08.001

Lee, T.-W., Girolami, M., & Sejnowski, T. J. (1999). Independent Component Analysis Using an Extended Infomax Algorithm for Mixed Subgaussian and Supergaussian Sources. *Neural Computation*, *11*(2), 417–441. https://doi.org/10.1162/089976699300016719

Lemm, S., Blankertz, B., Dickhaus, T., & Müller, K.-R. (2011). Introduction to machine learning for brain imaging. *NeuroImage*, *56*(2), 387–399. https://doi.org/10.1016/j.neuroimage.2010.11.004

Liao, L.-D., Lin, C.-T., McDowell, K., Wickenden, A. E., Gramann, K., Jung, T.-P., … Chang, J.-Y. (2012). Biosensor Technologies for Augmented Brain-Computer Interfaces in the Next Decades. *Proceedings of the IEEE*, *100*(Special Centennial Issue), 1553–1566. https://doi.org/10.1109/JPROC.2012.2184829

Lopes da Silva, F. (1991). Neural mechanisms underlying brain waves: from neural membranes to networks. *Electroencephalography and Clinical Neurophysiology*, *79*(2), 81–93. https://doi.org/10.1016/0013-4694(91)90044-5

Lotte, F., Congedo, M., Lécuyer, A., Lamarche, F., & Arnaldi, B. (2007). A review of classification algorithms for EEG-based brain–computer interfaces. *Journal of Neural Engineering*, *4*(2), R1. https://doi.org/10.1088/1741-2560/4/2/R01

Mühl, C., Jeunet, C., & Lotte, F. (2014). EEG-based workload estimation across affective contexts. *Neuroprosthetics*, *8*, 114. https://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. *Journal of Neuroscience Methods*, *167*(1), 82–90. https://doi.org/10.1016/j.jneumeth.2007.09.022

Nelson, W. R., Haney, L. N., Ostrom, L. T., & Richards, R. E. (1998). Structured methods for identifying and correcting potential human errors in space operations. *Acta Astronautica*, *43*(3–6), 211–222.

Niedermeyer, E., & Silva, F. H. L. da. (2005). *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*. Lippincott Williams & Wilkins.

Nieuwenhuys, R., Voogd, J., & Huijzen, C. van. (2007). *The Human Central Nervous System: A Synopsis and Atlas*. Springer Science & Business Media.

Noback, C. R., Strominger, N. L., Demarest, R. J., & Ruggiero, D. A. (2005). *The Human Nervous System: Structure and Function*. Springer Science & Business Media.

O’Donnell, R. D., & Eggemeier. (1986). *Workload assessment methodology. Handbook of Perception and Human Performance*. (L. K. and K.R. Boff & J. P. Thomas, Eds.) (Vol. 2). John Wiley and Sons, Inc.

Parasuraman, R., Molloy, R., & Singh, I. L. (1993). Performance Consequences of Automation-Induced “Complacency.” *The International Journal of Aviation Psychology*, *3*(1), 1–23. https://doi.org/10.1207/s15327108ijap0301_1

Parasuraman, R., Mouloua, M., & Molloy, R. (1996). Effects of Adaptive Task Allocation on Monitoring of Automated Systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *38*(4), 665–679. https://doi.org/10.1518/001872096778827279

Parasuraman, R., & Rizzo, M. (2008). *Neuroergonomics: The Brain at Work* (1 edition). New York: Oxford University Press.

Paxion, J., Galy, E., & Berthelon, C. (2014). Mental workload and driving. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.01344

Pham, T. T. H., Croft, R. J., Cadusch, P. J., & Barry, R. J. (2011). A test of four EOG correction methods using an improved validation technique. *International Journal of Psychophysiology*, *79*(2), 203–210. https://doi.org/10.1016/j.ijpsycho.2010.10.008

Prinzel, L. J., Freeman, F. G., Scerbo, M. W., Mikulka, P. J., & Pope, A. T. (2000). A Closed-Loop System for Examining Psychophysiological Measures for Adaptive Task Allocation. *The International Journal of Aviation Psychology*, *10*(4), 393–410. https://doi.org/10.1207/S15327108IJAP1004_6

Ramnani, N., & Owen, A. M. (2004). Anterior prefrontal cortex: insights into function from anatomy and neuroimaging. *Nature Reviews Neuroscience*, *5*(3), 184–194. https://doi.org/10.1038/nrn1343

Rankin, W., Hibit, R., Allen, J., & Sargent, R. (2000). Development and evaluation of the Maintenance Error Decision Aid (MEDA) process. *International Journal of Industrial Ergonomics*, *26*(2), 261–276. https://doi.org/10.1016/S0169-8141(99)00070-0

Rasmussen, J. (1987). The definition of human error and a taxonomy for technical system design, 23–30.

Rasmussen, Jens. (1982). Human errors. A taxonomy for describing human malfunction in industrial installations. *Journal of Occupational Accidents*, *4*(2), 311–333. https://doi.org/10.1016/0376-6349(82)90041-4

Reason, J. (2000a). Human error. *Western Journal of Medicine*, *172*(6), 393–396.

Reason, J. (2000b). Human error: models and management. *BMJ : British Medical Journal*, *320*(7237), 768–770.

Riccio, A., Holz, E. M., Aricò, P., Leotta, F., Aloise, F., Desideri, L., … 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. *Archives of Physical Medicine and Rehabilitation*, *96*(3 Suppl), S54-61. https://doi.org/10.1016/j.apmr.2014.05.029

Romero, S., Mañanas, M. A., & Barbanoj, M. J. (2008). A comparative study of automatic techniques for ocular artifact reduction in spontaneous EEG signals based on clinical target variables: A simulation case. *Computers in Biology and Medicine*, *38*(3), 348–360. https://doi.org/10.1016/j.compbiomed.2007.12.001

Rumar, K. (1990). The basic driver error: late detection. *Ergonomics*, *33*(10–11), 1281–1290. https://doi.org/10.1080/00140139008925332

Salmon, P., Regan, M., & Johnston, I. (2005). *Human Error and Road Transport: Phase One - Literature Review*. Monash University Accident Research Centre. Retrieved from http://arrow.monash.edu.au/vital/access/manager/Repository/monash:38304

Sammer, G., Blecker, C., Gebhardt, H., Bischoff, M., Stark, R., Morgen, K., & Vaitl, D. (2007). Relationship between regional hemodynamic activity and simultaneously recorded EEG-theta associated with mental arithmetic-induced workload. *Human Brain Mapping*, *28*(8), 793–803. https://doi.org/10.1002/hbm.20309

Scerbo, M. W. (1996). Theoretical perspectives on adaptive automation. In R. Parasuraman & M. Mouloua (Eds.), *Automation and human performance: Theory and applications* (pp. 37–63). Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.

Sexton, J. B., Thomas, E. J., & Helmreich, R. L. (2000). Error, stress, and teamwork in medicine and aviation: cross sectional surveys. *BMJ : British Medical Journal*, *320*(7237), 745–749.

Shappel, S. A., & Wiegmann, D. A. (2000). *The Human Factors Analysis and Classification System-HFACS* (No. DOT/FAA/AM-00/7). Washington, DC: Federal Aviation Administration.

Shorrock, S. T., & Kirwan, B. (2002). Development and application of a human error identification tool for air traffic control. *Applied Ergonomics*, *33*(4), 319–336.

Shou, G., Ding, L., & Dasari, D. (2012). Probing neural activations from continuous EEG in a real-world task: Time-frequency independent component analysis. *Journal of Neuroscience Methods*, *209*(1), 22–34. https://doi.org/10.1016/j.jneumeth.2012.05.022

Smit, A. S., Eling, P. A. T. M., & Coenen, A. M. L. (2004). Mental effort affects vigilance enduringly: after-effects in EEG and behavior. *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology*, *53*(3), 239–243. https://doi.org/10.1016/j.ijpsycho.2004.04.005

Smith, M. E., Gevins, A., Brown, H., Karnik, A., & Du, R. (2001). Monitoring task loading with multivariate EEG measures during complex forms of human-computer interaction. *Human Factors*, *43*(3), 366–380.

Stanton, N. A., Harris, D., Salmon, P. M., Demagalski, J. M., Marshall, A., Young, M. S., … Waldmann, T. (2006, February 1). Predicting design induced pilot error using HET (human error template) - A new formal human error identification method for flight decks. [Article]. Retrieved November 25, 2015, from https://dspace.lib.cranfield.ac.uk/handle/1826/1158

UnderwoodJun, E. (2014, June 5). A $4.5 Billion Price Tag for the BRAIN Initiative? Retrieved October 23, 2017, from http://www.sciencemag.org/news/2014/06/45-billion-price-tag-brain-initiative

Urigüen, J. A., & Garcia-Zapirain, B. (2015). EEG artifact removal—state-of-the-art and guidelines. *Journal of Neural Engineering*, *12*(3), 031001. https://doi.org/10.1088/1741-2560/12/3/031001

Van Essen, D. C., Ugurbil, K., Auerbach, E., Barch, D., Behrens, T. E. J., Bucholz, R., … Yacoub, E. (2012). The Human Connectome Project: A data acquisition perspective. *NeuroImage*, *62*(4), 2222–2231. https://doi.org/10.1016/j.neuroimage.2012.02.018

Vapnik, V. N. (2000). *The Nature of Statistical Learning Theory*. New York, NY: Springer New York. Retrieved from http://link.springer.com/10.1007/978-1-4757-3264-1

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. *Conference Proceedings: ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference*, *2014*, 6969–6972. https://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. *Motivation and Emotion*, *33*(1), 63–74. https://doi.org/10.1007/s11031-008-9116-y

Vigario, R., Sarela, J., Jousmiki, V., Hamalainen, M., & Oja, E. (2000). Independent component approach to the analysis of EEG and MEG recordings. *IEEE Transactions on Biomedical Engineering*, *47*(5), 589–593. https://doi.org/10.1109/10.841330

von Luxburg, U., & Schoelkopf, B. (2008). Statistical Learning Theory: Models, Concepts, and Results. *arXiv:0810.4752 [Math, Stat]*. Retrieved from http://arxiv.org/abs/0810.4752

Ward, L. M. (2003). Synchronous neural oscillations and cognitive processes. *Trends in Cognitive Sciences*, *7*(12), 553–559. https://doi.org/10.1016/j.tics.2003.10.012

Warm, J. S., Parasuraman, R., & Matthews, G. (2008). Vigilance requires hard mental work and is stressful. *Human Factors*, *50*(3), 433–441.

Wickens, C. D. (1984). Processing resources in attention. In *Varieties of Attention* (pp. 62–102).

Wickens, C. D., Hollands, J. G., Banbury, S., & Parasuraman, R. (2012). *Engineering Psychology & Human Performance* (4 edition). Boston: Psychology Press.

Wierwille, W. W., & Eggemeier, F. T. (1993). Recommendations for Mental Workload Measurement in a Test and Evaluation Environment. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *35*(2), 263–281. https://doi.org/10.1177/001872089303500205

Wolpaw, J. R., Birbaumer, N., Heetderks, W. J., McFarland, D. J., Peckham, P. H., Schalk, G., … Vaughan, T. M. (2000). Brain-computer interface technology: a review of the first international meeting. *IEEE Transactions on Rehabilitation Engineering*, *8*(2), 164–173. https://doi.org/10.1109/TRE.2000.847807

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

Wolpaw, Jonathan R, Birbaumer, N., McFarland, D. J., Pfurtscheller, G., & Vaughan, T. M. (2002). Brain-computer interfaces for communication and control. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, *113*(6), 767–91.

Wood, J. N., & Grafman, J. (2003). Human prefrontal cortex: processing and representational perspectives. *Nature Reviews. Neuroscience*, *4*(2), 139–147. https://doi.org/10.1038/nrn1033

Zander, T. O., Kothe, C., Welke, S., & Roetting, M. (2009). Utilizing Secondary Input from Passive Brain-Computer Interfaces for Enhancing Human-Machine Interaction. In D. D. Schmorrow, I. V. Estabrooke, & M. Grootjen (Eds.), *Foundations of Augmented Cognition. Neuroergonomics and Operational Neuroscience* (Vol. 5638, pp. 759–771). Berlin, Heidelberg: Springer Berlin Heidelberg. Retrieved from http://www.springerlink.com/content/66p475202q80m11k/

Zander, Thorsten O., & Kothe, C. (2011). Towards passive brain–computer interfaces: applying brain–computer interface technology to human–machine systems in general. *Journal of Neural Engineering*, *8*(2), 025005. https://doi.org/10.1088/1741-2560/8/2/025005

LIST OF PUBLICATIONS

PATENTS

Aricò, P., Borghini, G., **Di Flumeri, G.**, & Babiloni, F. (2017, March 22). EP3143933 A1. Method for estimating a mental state, in particular a workload, and related apparatus. Retrieved from http://www.google.it/patents/EP3143933A1.

BOOKS

Borghini, G., Aricò, P., **Di Flumeri, G.**, & Babiloni, F. (2017). Industrial Neuroscience in Aviation: Evaluation of Mental States in Aviation Personnel. Cham: Springer International Publishing Springer.

PAPERS AND CONFERENCE PROCEEDINGS

G. Di Flumeri, P. Aricò, G. Borghini, N. Sciaraffa, AG Maglione, D. Rossi, E. Modica, I. Mascarell, A. Trettel, F. Babiloni, A. Colosimo, MT Herrero. "EEG-Based Approach-Withdrawal Index for the Pleasantness Evaluation During Taste Experience in Realistic Settings". 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2017. *In press*

N. Sciaraffa, G. Borghini, P. Aricò, **G. Di Flumeri**, A. Colosimo, A. Bezerianos, NV Thakor, F. Babiloni. "Brain Interaction during Cooperation: Evaluating Local Properties of Multiple-Brain Network". Brain Sciences, 2017, 7(7): 90, MDPI. DOI: 10.3390/brainsci7070090.

P. Aricò, G. Borghini, **G. Di Flumeri**, N. Sciaraffa, A. Colosimo, F. Babiloni. "Passive BCI in Operational Environments: Insights, Recent Advances, and Future Trends". IEEE Transactions on Biomedical Engineering, 2017; 64: 7, IEEE. DOI: 10.1109/TBME.2017.2694856.

P. Aricò, G. Borghini, **G. Di Flumeri**, S. Bonelli, A. Golfetti, I. Graziani, S. Pozzi, JP Imbert, G. Granger, R. Benhacene, D. Schaefer, F. Babiloni. "Human Factors and Neurophysiological Metrics in Air Traffic Control: a Critical Review". IEEE

Reviews in Biomedical Engineering, 2017; IEEE. DOI: 10.1109/RBME.2017.2694142.

G. Borghini, P. Aricò, **G. Di Flumeri**, N. Sciaraffa, A. Colosimo, MT Herrero, A. Bezerianos, NV Thakor, F. Babiloni. “A New Perspective for the Training Assessment: Machine Learning-Based Neurometric for Augmented User's Evaluation”. Frontiers in Neuroscience, 2017; 11: 325, Frontiers. DOI: 10.3389/fnins.2017.00325.

G. Borghini, P. Aricò, **G. Di Flumeri**, G. Cartocci, A. Colosimo, S. Bonelli, A. Golfetti, JP Imbert, G. Granger, R. Benhacene, S. Pozzi, F. Babiloni. “EEG-based cognitive control behaviour assessment: an ecological study with professional air traffic controllers”. Scientific Reports, 2017; 7: 547, Nature Publishing Group. DOI: 10.1038/s41598-017-00633-7.

P. Aricò, G. Borghini, **G. Di Flumeri**, A. Colosimo, S. Bonelli, A. Golfetti, S. Pozzi, JP Imbert, G. Granger, R. Benhacene and F. Babiloni. Adaptive automation triggered by EEG-based mental workload index: a passive brain-computer interface application in realistic air traffic control environment. Frontiers in human neuroscience, 2016, 10. DOI: 10.3389/FNHUM.2016.00539

Trettel, P. Cherubino, G. Cartocci, D. Rossi, E. Modica, AG Maglione, **G. Di Flumeri** and F. Babiloni. Transparency and Reliability in Neuromarketing Research, In: Ethics and Neuromarketing. Springer International Publishing, 2017. p. 101-111. DOI: 10.1007/978-3-319-45609-6_6.

P. Cherubino, G. Cartocci, A. Trettel, D. Rossi, E. Modica, AG Maglione, M. Mancini, **G. Di Flumeri** and F. Babiloni. Marketing Meets Neuroscience: Useful Insights for Gender Subgroups During the Observation of TV Ads. Applying Neuroscience to Business Practice, 2016, 163. DOI: 10.4018/978-1-5225-1028-4.ch008.

G. Di Flumeri, P. Aricò, G. Borghini, A. Colosimo and F. Babiloni. “A new regression-based method for the eye blinks artifacts correction in the EEG signal, without using any EOG channel”, 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2016. DOI: 10.1109/EMBC.2016.7591406.

G. Borghini, P. Aricò, **G. Di Flumeri**, A. Colosimo, SF Storti, G. Menegaz, P. Fiorini and F. Babiloni. “Neurophysiological Measures for Users’ Training Objective Assessment During Simulated Robot-Assisted Laparoscopic Surgery”, 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2016. DOI: 10.1109/EMBC.2016.7590866.

G. Di Flumeri, M. T. Herrero, A. Trettel, P. Cherubino, AG Maglione, A. Colosimo, E. Moneta, M. Peparaio e F. Babiloni. “EEG frontal asymmetry related to pleasantness of olfactory stimuli in young subjects”, Springer Proceedings in Business and Economic, Selected Issues in Experimental Economics, Springer International Publishing, 2016, pp. 373-381. DOI: 10.1007/978-3-319-28419-4_23.

P. Cherubino, A. Trettel, G. Cartocci, D. Rossi, E. Modica, AG Maglione, M. Mancini, **G. Di Flumeri** e F. Babiloni. "Neuroelectrical indexes for the study of the efficacy of TV advertising stimuli", Springer Proceedings in Business and Economic, Selected Issues in Experimental Economics, Springer International Publishing, 2016, pp. 355-371. DOI: 10.1007/978-3-319-28419-4_22.

G. Cartocci, P. Cherubino, D. Rossi, E. Modica, AG Maglione, **G. Di Flumeri** and F. Babiloni. “Gender and Age Related Effects While Watching TV Advertisements: An EEG Study”, Computational Intelligence and Neuroscience, Hindawi, 2016. DOI: 10.1155/2016/3795325.

P. Aricò, G. Borghini, **G. Di Flumeri**, A. Colosimo, S. Pozzi, and F. Babiloni. “A passive brain–computer interface application for the mental workload assessment on professional air traffic controllers during realistic air traffic control tasks”, Progress in Brain Research, Brain-Computer Interfaces: Lab Experiments to Real-World Applications. 2016, 228: pp. 295-328. DOI: 10.1016/bs.pbr.2016.04.021.

G. Di Flumeri, G. Borghini, P. Aricò, A. Colosimo, S. Pozzi, S. Bonelli, A. Golfetti, W. Kong, F. Babiloni. “On the use of cognitive neurometric indexes in aeronautic and Air Traffic Management environments”, Symbiotic Interaction, Springer

International Publishing, 2015, pp. 45-56. DOI: 10.1007/978-3-319-24917-9.

G. Di Flumeri, G. Borghini, P. Aricò, S. Pozzi, S. Bonelli, A. Golfetti, A. Colosimo, F. Babiloni. "Mental workload evaluation of ATCOs during ecological ATM scenarios", Italian Journal of AeroSpace Medicine, Italian Society of Aerospace Medicine (AIMAS), 2015; (13): pp. 38-48.

G. Cartocci, AG. Maglione, G. Vecchiato, **G. Di Flumeri**, A. Colosimo, A. Scorpecci, P. Marsella, S. Giannantonio, P. Malerba, G. Borghini, P. Aricò e F. Babiloni. "Mental workload estimations in unilateral deafened children.", 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2015, pp. 1654-1657. DOI: 10.1109/EMBC.2015.7318693.

P. Aricò, G. Borghini, **G. Di Flumeri**, A. Colosimo, I. Graziani, JP Imbert, G. Granger, R. Benhacene, M. Terenzi, S. Pozzi, F. Babiloni. "Reliability over time of EEG-based mental workload evaluation during Air Traffic Management (ATM) tasks", 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2015, pp. 7242-7245. DOI: 10.1109/EMBC.2015.7320063.

G. Borghini, P. Aricò, **G. Di Flumeri**, S. Salinari, A. Colosimo, S. Bonelli, L. Napoletano, A. Ferreira, F. Babiloni. "Avionic Technology Testing By Using a Cognitive Neurometric Index: a Study with Professional Helicopter Pilots", 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2015, pp. 1654-1657. DOI: 10.1109/EMBC.2015.7318693.