ALMA MATER STUDIORUM
UNIVERSITÀ DEGLI STUDI DI BOLOGNA

DOTTORATO DI RICERCA IN BIOINGEGNERIA

CICLO XXVIII

SETTORE CONCORSUALE DI AFFERENZA: 09/G2

SETTORE SCIENTIFICO DISCIPLINARE: ING-INF/06

DEVELOPMENT OF A SYSTEM FOR THE TRAINING ASSESSMENT AND MENTAL WORKLOAD EVALUATION

STUDENTE**: GIANLUCA BORGHINI**

RELATORE:

Prof. **Fabio Babiloni**

REVISORI:

Prof. **Mauro Ursino**
Dr. **Ricardo Chavarriaga**

CO-RELATORE:

Prof.ssa **Serenella Salinari**

COORDINATORE DOTTORATO:

Prof.ssa **Elisa Magosso**

ESAME FINALE ANNO 2016

Summary

Tables

Figures

Preface

During the 3 years of my PhD, I had the opportunity to enlarge my professional and personal knowledge and experience, and to deeply investigate the concepts of workload and cognitive training, focusing on the operational environments of Pilots and Air-Traffic-Controllers (ATCOs). The experimental activities have been ran in several Labs across Europe and Asia. In particular, I spent 3 months at the *Cognitive Engineering Lab of the Institute for Neurotechnology* (SINAPSE) in Singapore (*National University of Singapore* – NUS), where several experiments for the mental workload and training assessments have been designed and performed. I have been involved in the European project "NINA: Neurometrics INdicators for ATM" (*www.nina-wpe.eu*), in which professional ATCOs have been recruited from the *École Nationale de l'Aviatione Civile* (ENAC) of Toulouse (France), and from the *Ente Nazionale di Assistenza al Volo* (ENAV) of Rome (Italy) to take part in realistic experiments. Also, experiments regarding the mental workload evaluation have been done at the Research Centre of *AgustaWestland* (Yeovil, UK) involving professional Pilots.

The results of the PhD have brought to idea of realizing a device, based on the metrics and methodologies developed during these 3 years of research, able to objectively measure the mental workload, during the execution of operative tasks, and to quantitatively estimate the learning progress, across the training sessions, both in order to better manage the training courses, and to support the Instructors in the organization, training personalization and selection phases.

In this regard, a H2020-SMEINST project has been submitted, and different Companies (ENAC (France); Thales (Italy); Alenia Aermacchi (Italy); NATS (UK); CIRA (IT) and Flight Schools (School of Aeronautics and Astronautic, Zeijang University, Hangzhou (China); Civil Aviation University, Tijanin (China); Pilot school Guanchan, Sichuan (China)) have already stated, by means of letters of interest, their interest in testing the device or developing further research collaborations. The phase 1 has been

founded and already developed. The phase 2 is under evaluation by the H2020 Commission.

The PhD results have been presented in many International and National conferences (2 prizes for the best research studies in the aeronautical medicine field - AIMAS 2014 and 2015), International and National Summer School (first IEEE EMBS International Summer School on Neural Engineering (ISSNE), Shanghai, China; IV Convegno Nazionale di Bioingegneria, GNB, Pavia) and published on 16 conferences, and 7 journal papers (until December 2015). Furthermore, currently there are two patents pending on the algorithms and methodologies developed during the PhD and described in details in the Thesis.

Abstract

Several studies have demonstrated that the main cause of accidents are due to *Human Factor* (HF) failures. Humans are the least and last controllable factor in the activity workflows, and the availability of tools able to provide objective information about the user's cognitive state should be very helpful in maintain proper levels of safety. To overcome these issues, the objectives of the PhD covered three topics. The first phase was focused on the study of machine-learning techniques to evaluate the user's mental workload during the execution of a task. In particular, the methodology was developed to address two important limitations: i) over-time reliability (no recalibration of the algorithm); ii) automatic brain features selection to avoid both the *underfitting* and *overfitting* problems. The second phase was dedicated to the study of the training assessment. In fact, the standard training evaluation methods do not provide any objective information about the amount of brain activation\resources required by the user, neither during the execution of the task, nor across the training sessions. Therefore, the aim of this phase was to define a neurophysiological methodology able to address such limitation. The third phase of the PhD consisted in overcoming the lack of neurophysiological studies regarding the evaluation of the cognitive control behaviour under which the user performs a task. The model introduced by Rasmussen was selected to seek neurometrics to characterize the *skill*, *rule* and *knowledge* behaviours by means of the user's brain activity.

The experiments were initially ran in controlled environments, whilst the final tests were carried out in realistic environments. The results demonstrated the validity of the developed algorithm and methodologies (2 patents pending) in solving the issues quoted initially. In addition, such results brought to the submission of a H2020-SMEINST project, for the realization of a device based on such results.

Keywords: **Mental Workload, Training, Brain Activity, Cognitive Control Behaviour, Human Factor, Pilot, Air Traffic Controller, Learning, SWLDA, Machine Learning, EEG, ECG, EOG, Cortical Maps, Self Assessment, Cognitive Spare Capacity.**

1. Introduction

In the operational environments, the safety of the people rely on the work and efficiency of 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 Factors* (HFs) have 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 HF as principal cause (Feyer and 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), and it has also been estimated that inappropriate human actions are implicated in as much as 95 % of road traffic crashes (Aberg and Rimmö, 1998). Additionally, over the past four decades, HF has been involved in high number of casualty catastrophes (Salmon et al., 2005). Consequently, the HF construct has received more attention, and it has been investigated across a wide range of domains, including military and civil aviation (Shappel and Wiegmann, 2000; Stanton et al., 2006), aviation maintenance (Rankin et al., 2000), air traffic management (Shorrock and Kirwan, 2002), rail (Lawton and Ward, 2005), road transport (Reason, 2000; Rumar, 1990), nuclear power and petrochemical reprocessing (Kirwan, 1992, 1998), military, medicine (Helmreich, 2000; Sexton et al., 2000), and even the space travel domain (Nelson et al., 1998). Human error is an extremely common phenomenon, since 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. According to the scientific literature, there have been numerous attempts at defining and classifying the human error. However, a universally accepted definition does not exist yet. 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. The main causes of human errors have to be sought within the internal or psychological factors of the operator (Reason, 2000). In fact, errors could arise from aberrant mental processes, such as inattention, poor motivation, loss of vigilance, mental overload and fatigue that negatively affect the user's performance. For example, cognitive psychology literature demonstrated that the mental workload has an ''inverted U-shape'' relationship with performance. In other words, some levels of mental workload may help the user to reach high performance level (Calabrese, 2008; Warm et al., 2008), since it stimulates positively the user and it keeps him/her awake with high attention level. On the contrary, a period of mental inactivity and "under-stimulation" can cause a monotonous and boring state (*underload*), a low level of vigilance and attention, with low cognitive resources demand. Additionally, an operative condition characterized by high demanding multitasks can lead the user to an *overload* condition and to a likely occurrence of errors (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. In this regard, the *augmented cognition* research field aims at developing systems to avoid performance degradation by adapting the user's interface and reducing the task demand/complexity, or by intervening directly on the system (Fuchs et al., 2007). Over the past two decades, researchers in the field of augmented cognition developed novel technologies to both monitor and enhance human cognition and performance. Most of those works were based on evidences coming from cognitive science and cognitive neuroscience (Decades, 2008). On the basis of such findings and technological improvements in measuring human biosignals, it has been possible to evaluate operators' mental states unobtrusively and in realistic contexts. The neurophysiological indexes have then been used as inputs for the interface the operator was interacting with. Such kind of application is called *passive Brain-Computer Interface* (p-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 (Wolpaw et al., 2002). More recently, Wolpaw and Wolpaw (2012) defined a BCI as *"a system that*

measures Central Nervous System (CNS) activity and converts it into artificial output that replaces, restores, enhances, supplements, or improves natural CNS output and thereby changes the ongoing interactions between the CNS and its external or internal environment". In the BCI community, the possibility of using the BCI systems in different contexts for communication and system control (Aloise et al., 2013), developing applications in realistic and operational environments, is not just a theory but something very closed to real applications (Blankertz et al., 2010; Müller et al., 2008; Zander et al., 2009). In fact, in the classic BCI applications the user modulates voluntarily its brain activity to interact with the system (Borghini et al., 2015c). In the new BCI concept, the system itself 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 (i.e. *p-BCI*). Systems based on passive BCI technology can provide objective information about covert aspects of the user's cognitive state, since conventional methods, such as behavioral measures, could only detect such mental states with weak reliability (Zander and Jatzev, 2012). The information extracted by the p-BCIs are then employed to improve *Human - Machine Interactions* (HMI) and to achieve potentially novel types of skills. Anyhow, the quantification of mental states by using BCI technology is far from trivial. In fact, it requires combination of knowledge in different fields (Brouwer et al., 2015), such as neurophysiology (to acquire and manage biosignals), experimental psychology (to find out the right way to assess mental states), machine learning (to develop innovative classification techniques), and human factor (to develop real applications). Mental states monitoring is of particular interest in safety-critical applications where the human factor is often the least controllable factor. In this regard, there are many examples in which p-BCIs could be very useful. For example, p-BCI technology can reveal valuable information about the user's mental state in safety-critical applications, such as driving (Borghini et al., 2012a, b; Welke et al., 2009), industrial environments (Venthur et al., 2010) or security surveillance (Marcel and Millan, 2007). With respect to driving assistance applications, recent studies demonstrated the utility of p-BCI systems during driving simulation for assessing driving performance and inattentiveness (Schubert

and Tangermann, 2008), as well as for robustly detecting emergency brakes before braking onset (Welke et al., 2009). In addition, p-BCI systems can potentially be used for real-time cognitive monitoring of the operator's mental workload (Aricò et al., 2014a, 2014b; Kohlmorgen et al., 2007). These and other forms of automation have yielded several benefits in terms of improved capacity, efficiency and safety. There is evidence that, *"although unconstrained handover of automated tasks to operators would increase workload and defeat the purpose of automation, there are situations during which temporary allocation to human control is desirable"* (Parasuraman et al., 1996). They highlighted the superiority of adaptive task allocation (i.e. automation) over static automation, at least for the monitoring aspect of performance, during a multitasking flight simulation. (Parasuraman and Rizzo, 2008) reported a list of new adaptive solutions able to improve the human performance in various environments, thanks to the improvement of the technology in the biosensors domain. Neuroimaging methods and cognitive neuroscience have steadily improved their technical sophistication and breadth of application over the past decade, and there has been growing interest in their use to examine the neural circuits supporting complex tasks representative of perception, cognition, and action as they occur in operational settings. At the same time, many fields in the biological sciences, including neuroscience, are being challenged to demonstrate their relevance to practical real-world problems (Parasuraman, 2003). Researchers in HF and ergonomics sectors studied human capabilities and limitations, both cognitive and physical, and used such knowledge to design technologies and work environments to be safer and more usable, efficient, and enjoyable for people to interact with (Norman and Berkrot, 2011; Wickens et al., 2012). In today's technology-driven environment, where human capabilities are struggling to keep up with technology offerings, techniques for augmenting human performance are becoming the critical gap to preclude realizing the full benefits that these technology advances offer. The concept of human performance augmentation is not so recent. The idea was developed during the past decade (Blankertz et al., 2010; Zander et al., 2009), and, at the same time, the concept of Augmented Cognition (*AugCog*) was borne out of the *Defense Advanced Research Projects Agency's* (DARPA) pushing for technologies that enhanced the Warfighter's communication skills and those

technologies that involved biosensors for medical applications (Schmorrow, 2002). In this context, the most studied mental states, due to their strong relationship with the increasing or the degrading of user's performance, have been the mental workload and cognitive training level. The techniques for the mental workload and training assessment must be sensitive to fluctuations in task demands without intruding on primary task performance (O'Donnell and Eggemeier, 1986). This level of sensitivity is not obtainable with behavioral (task performance) and subjective measures (e.g. workload perception) alone. In fact, subjective measures concurrently the execution of the considered activity require additional tasks (self-assessment itself). Additionally, it has been widely demonstrated that neurophysiological measurements transcend both behavioral and subjective measures in discriminating cognitive demand fluctuations (Di Flumeri et al., 2015; Mühl et al. 2013; Wierwille and Eggemeier, 1993). Therefore, the neurophysiological measurements 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. In other words, the neurophysiological measures could be used to realize p-BCI applications in real operational environments. Many neurophysiological measures have been used for the mental workload and training 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) (Borghini et al., 2012a; Ramnani and Owen, 2004; Wood and Grafman, 2003). The size, weight, and power constraints outlined above limit the types of neurofeedbacks that can be used to realize p-BCI applications. For example, fMRI (Cabeza and Nyberg, 2000) and MEG techniques require room-size equipment that are not portable. EOG, ECG and GSR activity measurements highlighted a correlation with some mental states (Stress, Mental Fatigue, Drowsiness), but they were demonstrated to be useful only in combination with other neuroimaging techniques directly linked to the *Central Nervous System* (CNS), i.e. the brain (Borghini et al.,

2012 a, b; Ryu and Myung, 2005). Since the presence of hair may impact on both photon absorption (Murkin and Arango, 2009) and the coupling of the probes with the underlying scalp, the fNIRs technique is very reliable only on those un-hairy brain areas, like the frontal brain areas. In addition, because of the key role of other brain areas, Derosière et al. (2013) pointed out how some fNIRs-measured hemodynamic variables were relatively insensitive to certain changes in the workload and attentional states. Consequently, the EEG potentially appears to be the best solution for the user's workload and training estimation in operational environments, since it also has high temporal resolution, proper usability and it can be straightforwardly employed to realize passive-BCI applications.

To overcome the issues and limitations described above, I focused my PhD on the definition and validation of metrics and methodologies able to address such problems and to provide objective and quantitative information about the user's mental workload and cognitive training level in operational environments. For such purposes, specific experimental protocols have been designed. The first phase of the research has been developed and tested in controlled settings (Laboratory), while the consecutive phases have been carried out in realistic environments (e.g. air traffic control room) involving professional personnel. During the execution of the experimental tasks, the neurophysiological signals (EEG, ECG and EOG) of the users have been recorded continuously, and both the behavioral (task performance) and subjective (workload perception) data have been gathered as well, with the aim to analyze the considered cognitive phenomena from the three different points of view (Triangulation methodology, Bekhet and Zauszniewski, 2012). From the biosignals, specific metrics have been defined to investigate the considered cognitive phenomenon. Both standard spectrum EEG analysis and techniques coming from the BCI field (i.e. machine - learning approach) have been used to analyze the physiological data. In particular, the *StepWise Linear Discriminant Analysis* (SWLDA) has been chosen since it has been demonstrated to be one of the best outperforming linear classifiers for the mental workload evaluation (Craven et al., 2006; Krusienski et al., 2006; Berka et al., 2007; Rabbi et al., 2009). Before its application, the use of the machine-learning approach for the training assessment and workload evaluation in operational environments has been studied carefully and

deeply, and, as a result, a modified version of the SWLDA (*automatic stop SWLDA*, asSWLDA) has been developed during the PhD and then patented (Patent I and Patent II, 2015). The asSWLDA has been able to solve important key issues in using neurophysiologicalmeasurements in operational environments: i) over - time reliability of the measure (no recalibration of the algorithm is necessary within a month), ii) automatic features selection to avoid both the *underfitting* and *overfitting* problems (no empiric and manual selection of the analysis parameters), and iii) it provided higher accuracy (resolution) in comparison with the standard (i.e. subjective) workload measures.

Training not only could result in the acquisition of new skills (Satterfield and Hughes, 2007; Hill and Lent, 2006), but also in improved declarative knowledge, enhance strategic knowledge, defined as knowing when to apply a specific procedure or skill, in particular during unexpected events (Kozlowski et al., 2001). Despite the time goes on, there might be the need to assess if the operator is still able to work ensuring high performance levels, hence, a proper level of safety. For such a reason, another issue is the necessity of objectively monitoring and assessing operators' performance (Leape and Fromson, 2006), in terms of cognitive control behaviours. In fact, different operators could achieve the same performance results, but involving a different amount of cognitive resources, thus different expertise. Nowadays, there are no tools able to provide such objective and quantitative information to better manage the training program and personnel selection. As a reference to such purpose, the Skill – Rule – Knowledge (S-R-K) framework (still used in aeronautic field) introduced by Rasmussen (1983) has been considered. The aim of the framework is to explain human behavior and describe the wide range of mental capabilities used in everyday situations (e.g. working environments). Rasmussen's model is characterized by three dynamic and parallel cognitive levels of expertise: s*kill-based behavior* (the sensory-motor activity, based on the repetition of automated and integrated behavioral routines that takes place without conscious attention, control or cognitive effort. The response is rapid, flexible and resistant to changes), *rule-based behavior* (the procedural behavior, characterized by the use of rules and procedures to select a course of actions in a familiar situation. "Control at rule-based level requires a conscious preparation of the sequence beforehand and the conscious mind

operates infrequently in synchrony with the environment."), and the *knowledge-based behavior* (usually occurs in situations where no rules are available from previous experiences. It involves problem solving based on the "knowledge" considered as a complex process of collection and integration of different information that, once interpreted, result in the comprehension of the current situation, and in the planning and execution of a new appropriate plan of actions. This level is mentally demanding and requires high attention. The response is slow and prone to error).The three levels are concerned as a dynamic system that generally works in parallel and are integrated, where the control of behavior continuously shifts from a level to another one. Such a structure allows to react quickly to environmental changes, to deal with ambiguous situations, to solve familiar or unfamiliar problems and to set new problems in an efficient and flexible way. When expertise evolves, cognitive control shifts, for example when a subject is learning to drive, his/her attention and cognitive resources are located to the set of actions of "driving a car". These rule-based behavior, with practice, requires less attention and control, until it becomes automated, shifting to the skill-based level.

2. Preliminary Concepts

2.1. Nervous System

Gross Anatomy of the Brain

The nervous system essentially exhibits a bilateral symmetry with those structural features and pathways located on one side of the midline (Noback et al., 2005). It is subdivided anatomically into the *Central Nervous System* (CNS) and the *Peripheral Nervous System (PNS),* and functionally into the *Somatic Nervous System* (SNS) and the *Autonomic (*visceral*) Nervous System* (ANS). The CNS comprises the brain and spinal cord. The brain is encapsulated within the skull and the spinal cord is at the centre of the vertebral column (Figure 1).

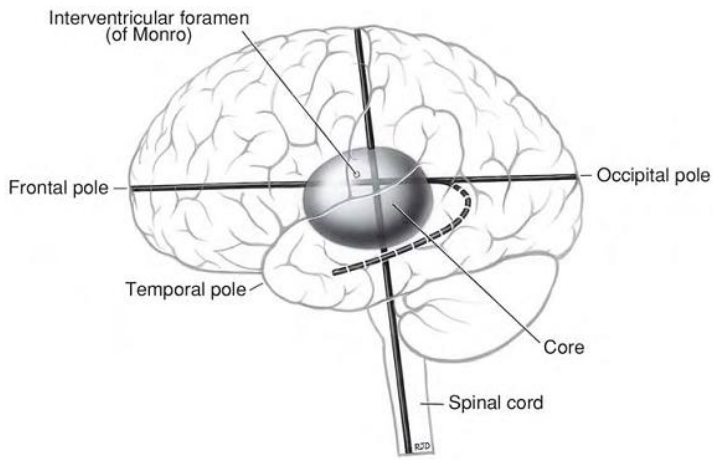

Figure 1. Geometry of the brain. Vertical axis is parallel to the long axis of the brainstem and spinal cord. Horizontal axis is parallel to the cerebrum from frontal pole to occipital pole.

The PNS consists of the nerves emerging from the brain (called *cranial nerves*) and from the spinal cord (called *spinal nerves*). The peripheral nerves convey neural messages from (1) the sense organs and sensory receptors in the organism inward to the CNS, and from (2) the CNS outward to the muscles and glands of the body. The SNS consists of those neural structures of the CNS and PNS responsible for (1) conveying and processing

conscious and unconscious sensory *(afferent)* information, vision, pain, touch, unconscious muscle sense from the head, body wall, and extremities to the CNS and (2) motor *(efferent)* control of the voluntary (striated) muscles. The ANS is composed of the neural structures responsible for (1) conveying and processing sensory input from the visceral organs (e.g., digestive system and cardiovascular system) and (2) motor control of the involuntary (smooth) and cardiac musculature, and of glands of the viscera. Sensory signals originating in sensory receptors are transmitted through the nervous system along sensory pathways, e.g., pain and temperature pathways and visual pathways. These signals may reach consciousness or may be utilized at unconscious levels. Neural messages for motor activity are conveyed through the nervous system to the muscles and glands along motor pathways. Both the sensory (ascending) and motor (descending) pathways include processing centres (e.g., ganglia, nuclei, laminae, cortices) for each pathway located at different anatomic levels of the spinal cord and brain. The processing centres are the computers of the complex high-speed systems within the brain. Differences in the basic sequence are present in some ascending systems. In a general way, the motor systems are organized to receive stimuli from the sensory systems, at all levels of the spinal cord and brain, and to convey messages via motor pathways to neuromuscular and neuro-glandular, endings at muscle and gland cells in the head, body, and extremities. The motor pathways comprise sequences of processing centres and their fibres conveying neural influences to other processing centres within the CNS, and the final linkages extending from the CNS via motor nerves of the PNS to muscles and glands. The CNS comprises *gray matter* and *white matter* (Figure 2). Gray matter consists of neuronal cell bodies, dendrites, axon terminals, synapses, and glial cells, and is highly vascular. White matter consists of bundles of axons, many of which are myelinated, and oligodendrocytes; the white colour is imparted by the myelin. It lacks neuronal cell bodies and is less vascular than gray matter. Groupings of neuronal cell bodies within the gray matter are variously known as a nucleus, ganglion, lamina, body, cortex, centre, formation, or horn. A cortex is a layer of gray matter on the surface of the brain. Two major cortices are recognized: *cerebral* and *cerebellar* cortices. The superior and inferior colliculi of the midbrain and the hippocampal formation also form cortex-like structures.

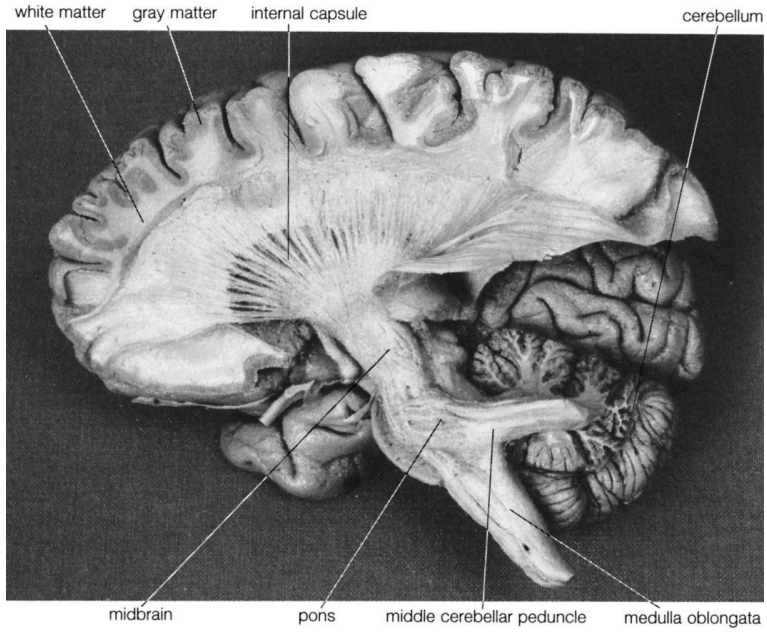

Figure 2. Sagittal brain section in which some of the main structures of the brains are labelled. In particular, the gray and white matters, the cerebellum, the midbrain, the medulla oblongata, the pons, the middle cerebellar peduncle and the internal capsule.

The *cerebrum* includes the paired cerebral hemispheres, a small median segment (derived from the telencephalon) and the diencephalon. The *cerebral hemispheres* consist of the cerebral cortex (gray matter), underlying white matter, corpus striatum, corpus callosum, anterior commissure, hippocampal formation and the amygdala (Figure 3). The brain hemispheres are marked on the surface by slit-like incisures called *sulci*. The raised ridge between two sulci is a *gyrus*. The cortex lining a sulcus is considered part of the adjacent gyrus. The hemispheres are separated from one another in the midline by the longitudinal *fissure* (particularly deep and constant sulcus). Each hemisphere is conventionally divided into six *lobes*: frontal, parietal, occipital, temporal, central (insula), and limbic (Figure 4). The portion of the frontal, parietal and temporal lobes that overlie the insula is called the *operculum.* The lobes are delineated from each other by several major sulci, including the lateral sulcus of Sylvius, central sulcus of Rolando, cingulate sulcus, and parieto-occipital sulcus. The *lateral sulcus* is a deep furrow that extends posteriorly from the basal surface of the brain along the lateral surface of the hemisphere, to terminate

usually as an upward curve within the inferior part of the parietal lobe. The *central sulcus* of Rolando extends obliquely from the region of the lateral sulcus across the dorsolateral cerebral surface and, for a short distance, onto the medial surface. The *cingulate sulcus* is a curved cleft on the medial surface extending parallel to the curvature of the corpus callosum. The *parieto-occipital sulcus* is a deep cleft on the medial surface located between the central sulcus and the occipital pole.

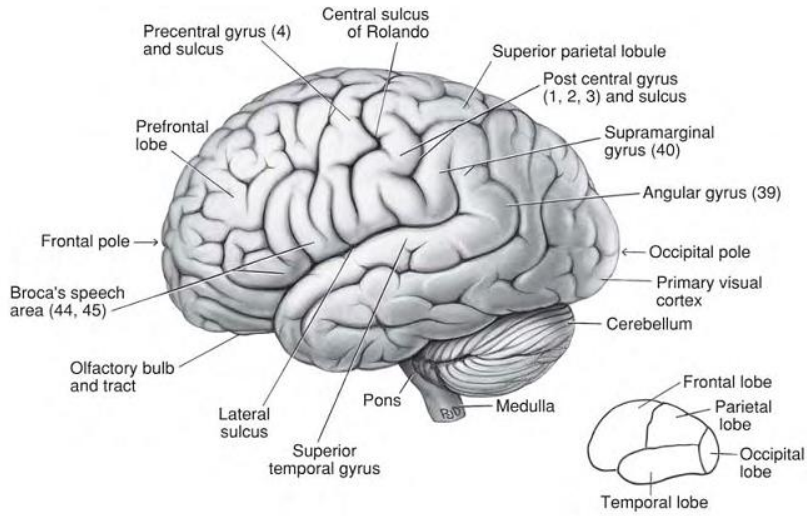

Figure 3. Lateral surface of the brain. Numbers refer to Brodmann's areas.

The boundaries of the lobes on the lateral cerebral surface are as follows: (1) the frontal lobe is located anterior to the central sulcus and above the lateral sulcus; (2) the occipital lobe is posterior to an imaginary line parallel to the parietooccipital sulcus, which is on the medial surface; (3) the parietal lobe is located posterior to the central sulcus, anterior to the imaginary parieto-occipital line, and above the lateral sulcus and a projection toward the occipital pole before it takes an upward curve; (4) the temporal lobe is located below the lateral sulcus and anterior to the imaginary parietooccipital line; and (5) the central lobe is located at the bottom (medial surface) of the lateral sulcus of Sylvius, which is actually a deep fossa (depression). It can be seen only when the temporal and frontal lobes are reflected away from the lateral sulcus (Figure 4). Furthermore, the boundaries of the lobes on the medial cerebral surface are as follows: (1) the frontal lobe is located rostral to a line formed by the central sulcus; (2) the

parietal lobe is between the central sulcus and the parieto-occipital sulcus; (3) the temporal lobe is located lateral to the para-hippocampal gyrus; (4) the occipital lobe is posterior to the parieto-occipital sulcus; and (5) the limbic lobe is a synthetic one formed by parts of the frontal, parietal, and temporal lobes. It is located central to the curved line formed by the cingulate sulcus and the collateral sulcus (the latter is located lateral to the para-hippocampal gyrus). The limbic lobe is the ring (limbus) of gyri bordered by this line; it includes the subcallosal area, cingulate gyrus, para-hippocampal gyrus, hippocampus, dentate gyrus, and uncus.

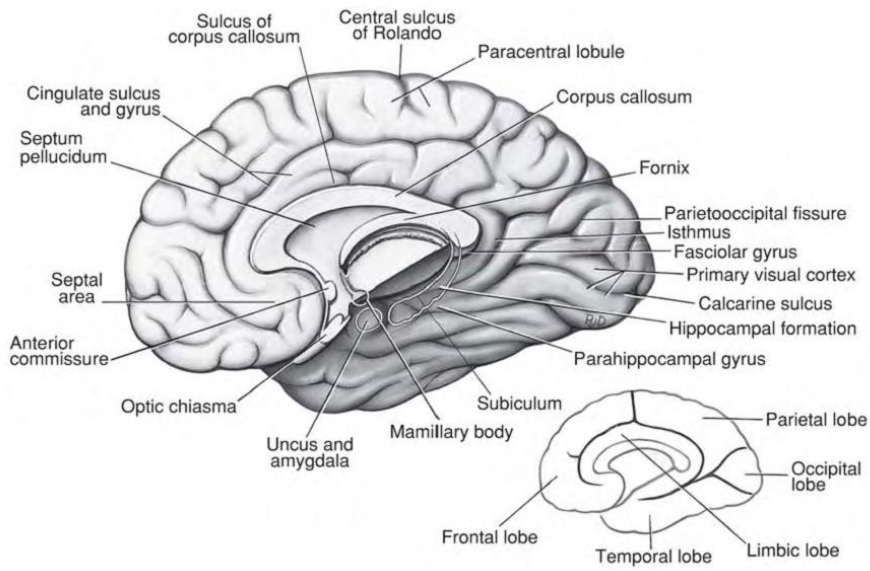

Figure 4. Medial surface of the cerebral hemisphere. The limbic lobe consists of the cingulate gyrus, isthmus, and para-hippocampal gyrus. White lines represent the amygdala and the hippocampal formation. The amygdala is located within the uncus. The hippocampal formation (hippocampus and dentate gyrus) is located in the floor of the temporal horn of the lateral ventricle.

Basic Unit of the Nervous System: Neurons

Some 100–200 billion ($[1–2] \times 10^{11}$) *neurons*, as well as many more glial cells, are integrated into the structural and functional structures of the brain (Noback et al., 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 structurally defined regions: *a cell body* (soma) that emits a single nerve

process called an *axon,* which ends at *presynaptic terminals,* and a variable number of branching processes called *dendrites* (Figure 5). Each axon, including its collateral branches, usually terminates as an arbor of fine fibres; each fibre ends as an enlargement called a *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 neural signals for communication between neurons and contain critical processing complexes. The axon is the conduit for conducting 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 stimuli, to transmit the resulting excitation rapidly to 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 for 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 5). The axon hillock, initial segment, and the axon lack Nissl bodies. The branches of an axon could have 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 are over 90% of all the excitatory synapses 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 (see Paragraph 4).

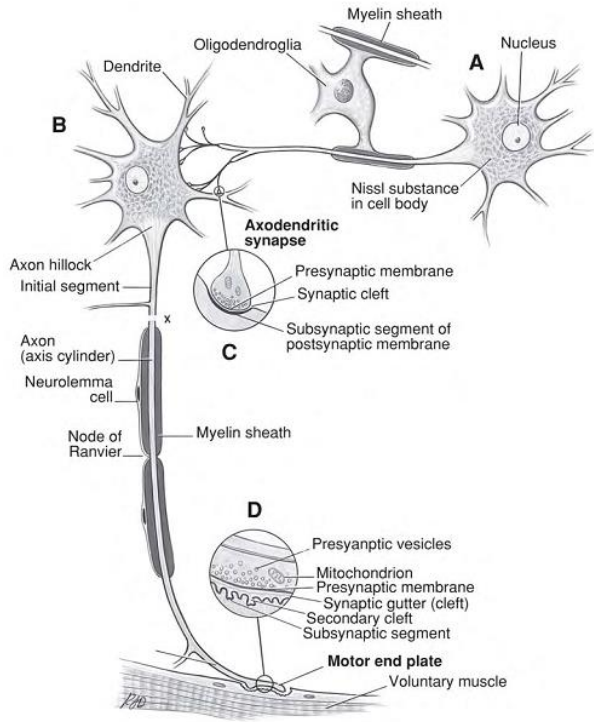

Figure 5. Diagram of (A) a neuron located wholly within the central nervous system and (B) a lower moto-neuron located in both the central and peripheral nervous systems. The latter synapses with a voluntary muscle cell to form a motor end plate. Note the similarities, as reconstructed from electron micrographs, between (C) a synapse between two neurons and (D) a motor end plate. The X represents the border between the central nervous system (above the X) and the peripheral nervous system (below the X). The myelin sheath of neuron (A) is entirely the product of a glial cell, and that of neuron (B) 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 site of contact of one neuron with another. A submicroscopic 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, *dendrodendritic synapses* (between two dendrites) have been noted (e.g., in the olfactory bulb and retina). The axon of one neuron might terminate in only 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 5). 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.

2.1.1. 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 6). 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. 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.

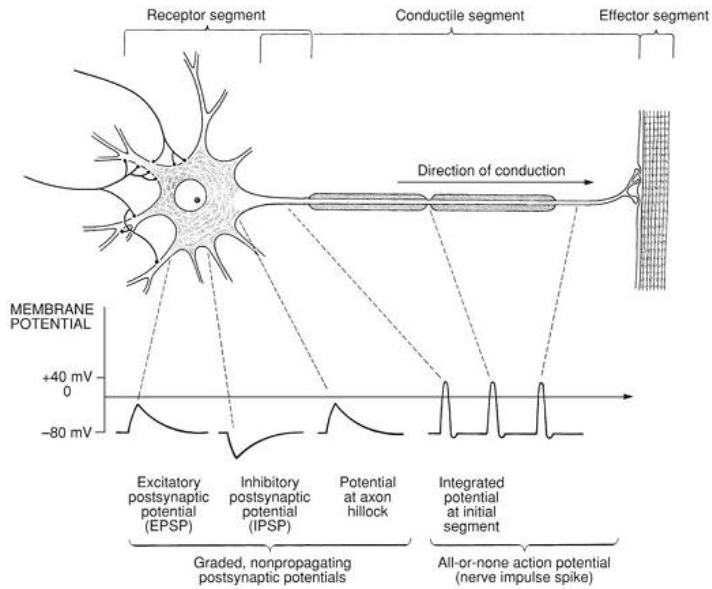

Figure 6. On the surface of the dendrites and cell body are excitatory and inhibitory synapses, which, when stimulated, produce local, graded, non-propagating potentials. These are exhibited as an excitatory or depolarizing postsynaptic potential (EPSP) and as an inhibitory or hyperpolarizing postsynaptic 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.

Resting Potential of the Neuron

The resting neuron is a charged cell that is not conducting a nerve impulse. The plasma membrane, which acts as a thin boundary between the extracellular (interstitial) fluid outside the neuron and the intracellular fluid (neuroplasm) inside the neuron (Figure 7), is critical for maintaining this charged state or resting potential. The electric charge across the plasma membrane results form a thin film of positive and negative ions, unequally distributed across the membrane. These are sodium (Na+) and chloride (Cl–) ions, which are in higher concentration in the interstitial fluid, and potassium (K+) and protein (organic) ions that are in higher concentration in the neuroplasm. A tendency exists for the Na+, K+, and Cl– ions to diffuse across the membrane from regions of high to low concentration (along concentration gradients), through Na+, K+, and Cl– channels, respectively.

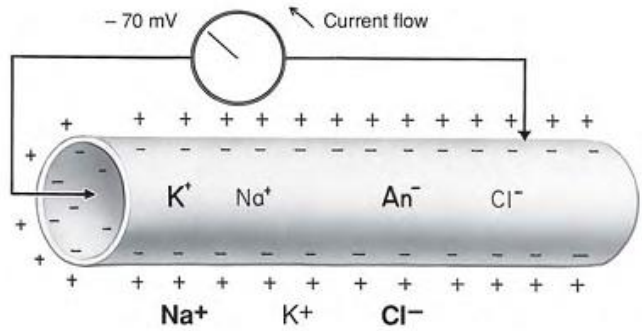

Figure 7. Resting potential. The intracellular neuroplasm potential of the normal nerve fibre "at rest" is negative to the extracellular potential. Sodium (Na+) and chloride (Cl–) ions are in high concentration in the extracellular fluid, and potassium (K+) ions and protein (An–) are in high concentrations in the neuroplasm. The potential across the plasma membrane is –70 to –80 (mV).

The passage of ions across the membrane is known as *conductance*. Thus, the semipermeable plasma membrane is selectively permeable through non-gated open channels to Na+, K+, and Cl– ions and impermeable to large protein ions. These channels, which are always open, are important in determining the resting potential. The ionic concentrations on either side of the membrane are produced and maintained by a system of membrane pumps called the *sodium–potassium pump* requiring metabolic energy released by adenosine triphosphate (ATP). The sodium–potassium exchange pump is an integral membrane protein that utilizes ATP as an energy source for its role in *active transport*. This transport is an energy-dependent process in which the movement of Na+ and K+ ions is "uphill" against a concentration gradient. The activity of the pump results in the passage of three Na+ ions out of and two K+ ions into the neuron. This causes the restoration of a concentration of K+ 30 or more times higher within the neuroplasm than in the interstitial fluid and in a concentration of Na+ that is 10 times and Cl– that is 14 times higher in the interstitial fluid than in the neuroplasm. Most neurons do not have a Cl– pump; hence, Cl– ions diffuse passively across the membrane. These are the ionic concentrations responsible for establishing an electric potential across the membrane. The transmembrane potential, known as the *resting potential*, is about –70 to –80 (mV) (millivolts) inside the neuron (Figure 7).The resting potential is in a steady state (*dynamic equilibrium*) requiring metabolic

energy to maintain the ionic gradients across the membrane. When the neuron is "at rest," its membrane potential is the result of a balance (involving Na+ and K+ ions) between the active fluxes (movements) of ions metabolically driven by *pumps* and the passive fluxes caused by *diffusion*. The active fluxes result from the pump extruding three Na+ ions for every two K+ ions it brings into the neuron. The passive fluxes of ions take place through non-gated channels. The outward flux of positive charges by the pump tends to hyperpolarize the membrane. The greater the hyperpolarization, the greater the inward electrochemical force driving Na+ into the neuron and the smaller the force driving K+ out. The steady state for the neuron is attained when the resting potential is reached at the point when the net passive inward current (movement of electrical charge) through the ion channels exactly counterbalances the active outward current driven by the pump. The steady state is not basically the result of *passive diffusion*, which is the diffusion of a solute down a concentration gradient without the expenditure of energy.

Excitability of the Neuron

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 8). 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.

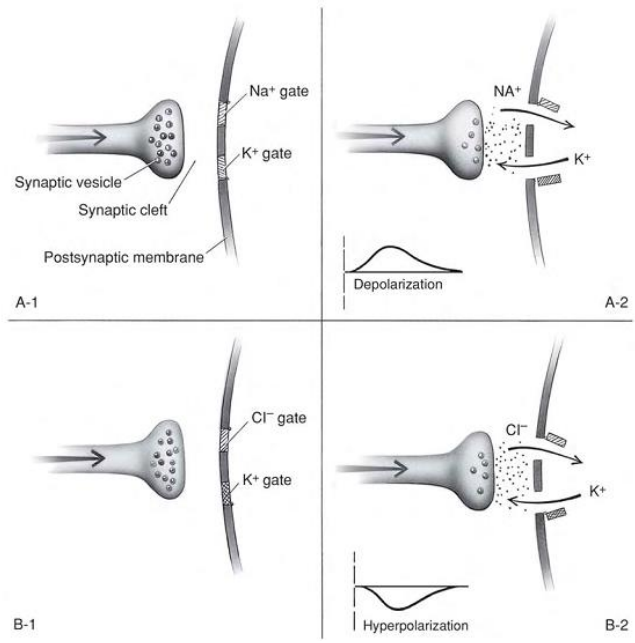

Figure 8. Excitatory synapses (A) and inhibitory synapses (B). *A-1* and *B-1*: Synapses prior to release of neurotransmitter. *A-2*: Excitatory postsynaptic response (EPSP) following release of neurotransmitter with Na+ ion inrush through Na+ gate and K+ ion outrush through K+ gate. *B-2*: Inhibitory 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)).

2.1.2. EEG Generation

The *Electroencephalogram* (EEG) comes from 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 and 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 (Lopez da Silva, 1996; Freeman, 1991). 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 and Linas, 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 (Lopez da Silva, 1996). 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 electrode 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. The second part is formed from numbers even on the left and odd on the right side, in the centre ''z'' or ''0''. 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 system were introduced in the extended 10 – 20 system (Pivik et al., 1993). 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. The *Hjorth* transform references each electrode to the four closest

neighbours, which is an approximation of the *Laplace transform* (LT). LT is calculated as a second spatial derivative of a signal, offering information about vertical current density. For best performance, it needs an adequate spatial sampling-inter electrode distance around 20 (mm) (e.g., 128 electrodes on the scalp). The estimates obtained by means of LT for the electrodes lying at the scalp periphery are biased and have to be excluded. Contrary to the open question of the reference, the necessity of artifact rejections is universally acknowledged. *Artifacts* are recorded signals that are non-cerebral in origin. They may be divided into one of two categories depending on their origin: *physiological artifacts* or *non-physiological artifacts*. Physiological artifacts 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 artifact-contaminated epochs. Non-physiological artifacts 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, artifacts may reduce the performance of machine-learning techniques. Several ways of handling physiological artifacts can be found in the literature. Artifacts may be avoided, rejected or removed from the EEG dataset. Artifacts avoidance involves asking users to avoid blinking or moving their body during the experiments (Vigário, 1997). This approach is very simple, because it does not require any computation, as brain signals are assumed to have no artifacts. However, this assumption is not feasible in operational environments, since some artifacts, eye and body movements, are not easily avoidable. Artifact rejection approaches suggest discarding the epochs contaminated by the artifacts. Manual artifact rejection is an option to remove artifacts in brain signals, and experts could identify and eliminate all artifact-contaminated EEG epochs. The main disadvantage in using manual rejection is that it requires intensive human labor, so this approach is not suitable for real-time evaluations. In the EEG, the following frequency rhythms are considered charachteristic for its analysis (Figure 9): delta (0.5 – 4 (Hz)), theta (4 – 8 (Hz)), alpha (8 – 12 (Hz)), beta (12 – 30 (Hz)), and gamma (above 30 (Hz)).

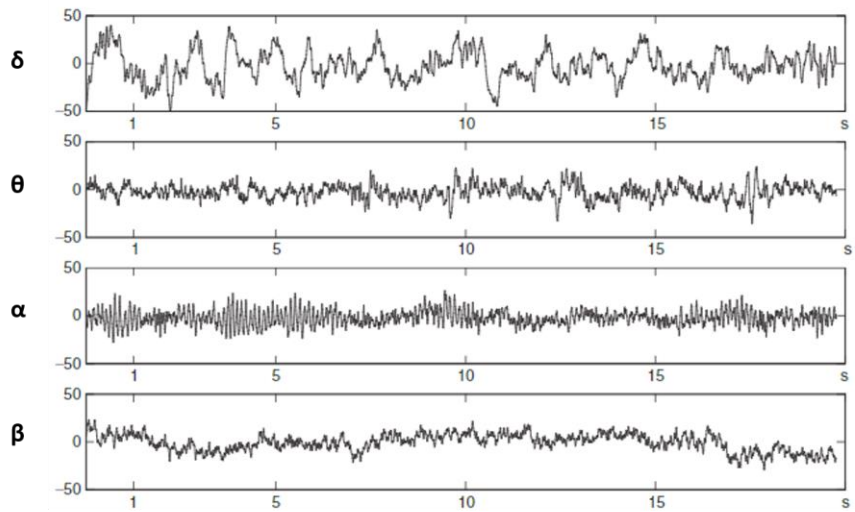

Figure 9. Characteristic EEG rhythms, from the top: δ (0.5 – 4 (Hz)), θ (4 – 8 (Hz)), α (8 – 12 (Hz)), β (12 –30 (Hz)). The gamma band could reach 100 (Hz).

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 (Rechtschaffen and Kales, 1968). 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 (Kubota at al., 2001), 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 (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 (Kubota at al., 2001; Smith et al., 2001), and it is often shown to increase with a higher cognitive task demand (Gundel and 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; Klimesch et al., 2007). Additionally, alpha activity may be further divided into sub-bands by means of the frequency corresponding to the alpha peak of the user (Klimesch et al., 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 (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 (Birbaumer and Schmidt, 1996). 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 is 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 et al., 2007). Gamma components are difficult to record by scalp electrodes, because of their low amplitude, but with *Electrocorticography* (ECoG) components upto 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 (Niedermayer and Lopes da Silva, 1993).

2.2. Cognitive Processes

The use of neuroimaging techniques allows the analysis of cognitive processes underlying cognition and human behavior. A precise identification of neural responses across specific brain regions can provide information by which defining metric for the objective evaluation of the considered cognitive phenomena. Among the main cognitive processes, the working memory, attention and decision making can be considered as elementary parts for the study of the learning and mental workload (Harris, 2013), and they are discussed in details in the following sections.

2.2.1. Working Memory

Working memory (WM) is considered to be one of the building blocks for higher cognitive processes (Den Bosch et al., 2014), such as decision making (Toth and Lewis, 1997), reasoning (Ruff et al., 2003), and recognition (Bledowski et al., 2012), since it is in charge of holding and simultaneously manipulating a small amount of information stored in the mind for a limited period of time for further processing (Protopapa et al., 2014). It also provides an essential interface between perception, attention, memory and action (Baddeley, 1996, 2003). WM involves three primary processes: encoding information, actively maintaining this information on-line in memory (storage), and finally, using the information to guide behaviour (retrieval). During encoding, individuals actively attend and construct an internal representation of the information in memory. This mental representation of the information is maintained during a delay period, during which the information is actively prevented from decaying due to interfering or competing stimuli. Finally, the information is retrieved from the memory buffer and conveyed through a motor response (e.g. verbal, oculomotor or manual response). It is known from behavioural studies that WM performance continues to improve from childhood, through adolescence and into early adulthood (Huizinga et al., 2006; Luciana and Nelson, 2000; Luna et al., 2004). The functional neurophysiological structure subserving the WM remains still controversial, but substantial progress has been made in the last years towards characterizing the neural basis of human WM (Yoon et al., 2006; Gevins,

1997). The anatomical substrate of working memory does not appear to depend on the corticolimbic circuitry that is necessary for retrieval of information when a delay of more than a few seconds is interposed between study and test trials of a memory task (Wickelgren, 1968; Cave and Squire, 1992). Rather, lesion studies have indicated that the frontal lobes often play a critical role in this function (Petrides and Milner, 1982; Frisk and Milner, 1990; Owen et al., 1990). Similarly, recent neuroimaging studies have found metabolic increases in regions of the frontal lobes as well as in other areas of association cortex during working memory tasks (McCarthy et al., 1994; Courtney et al., 1996; Smith et al., 1996). Electrophysiological methods have also been used to measure activity related to working memory. Recordings of neuronal activity in primates have indicated that the neural representation of information over short delays is associated with transient activation of widespread populations of association cortex neurons (Chelazzi et al., 1993; Wilson et al., 1993; Miller and Desimone, 1994). In humans, EEG studies have focused on Evoked Potential (EP) measures in an effort to track the subsecond time course and distribution of working memory processes (Lang et al., 1992; Ruchkin et al., 1995; Gevins et al., 1996). Gevins and colleagues (1996) identified phasic, sub-second changes in brain electrical signals over frontal and parietal cortex, that, in agreement with the studies of neuronal activity in primates, provided a dynamic picture of momentary changes in the spatial distribution of WM effects over the course of individual trials, as well as evidence for differences in the activity elicited between trials that have different decision requirements. Two prominent EEG features are highly sensitive to variations in WM. Several studies reported that the power of the frontal theta increased during active maintenance of working memory information until the information is retrieved (Jensen and Tesche, 2002b; Raghavachari et al., 2001; Klimesch et al., 2005). Fewer studies showed that the parietal and occipital theta are associated with the working memory task (Kawasaki and Yamaguchi, 2012; Raghavachari et al., 2006). Several studies also reported that posterior alpha suppression is associated with working memory tasks (Klimesch et al., 2007; Meltzer et al., 2008). Sauseng et al. (2005) and Huang et al. (2013) reported an increase in frontal alpha during working memory tasks. The posterior regions also underwent theta augmentation during a working

memory task, though this was followed by an alpha decrease (Lai et al., 2012).

2.2.2. Attention

The ability to suitably allocate processing resources is fundamental for the efficient processing of incoming sensory signals and for the generation of appropriate behaviour in complex environments. The brain continuously receives a large amount of sensory signals that cannot be fully processed. Attention selection allows preferential processing of some signals and filtering out of other inputs. Traditionally, attention research has distinguished between two types of mechanisms that contribute to the selection process: endogenous top-down factors, that primarily relate to the subject's "internal" goals and intentions, and exogenous bottom-up factors, that primarily relate to the characteristics "external" stimulus-driven control and they are thought to take place automatically (Macaluso and Doricchi, 2013). The distinction between these two types of control can be rather intuitive: when searching for a friend in a crowd, we will voluntary shift attention from one face to another until the external input (i.e., what we see) matches our internal knowledge about the physical appearance of our friend. By contrast, in a crowd of people all dressed with dark cloths, we will quickly notice a person dressed in bright red, who will catch attention automatically. Many functional neuroimaging studies have investigated the neural basis of endogenous and exogenous visuo-spatial attention control. These highlighted the central role of the frontal and parietal lobes, with a notable distinction between the response patterns in dorsal and ventral regions. A variety of tasks requiring endogenous control of spatial attention highlighted the activation of those brain regions (Corbetta et al., 1993; Gitelman et al., 1999; Yantis et al., 2002). The activation pattern typically includes the *Intraparietal Sulcus* (IPS) and/or the *Posterior Parietal Cortex* (PPC) plus the *Frontal Eye-Field* (FEF) in the premotor cortex. These areas are found in serial search tasks (Gitelman et al., 2002; Himmelbach et al., 2006; Fairhall et al., 2009), and many other non-spatial attention tasks also requiring the voluntary allocation of processing resources (Wojciulik and Kanwisher, 1999). Furthermore, specific experimental designs allowed segregating brain activity associated with the different phases of attention

tasks, which often comprise multiple control processes (Hahn et al., 2006; Doricchi et al., 2010; Simpson et al., 2011). Sensory aspects of the incoming signals may also play an important role, particularly so in posterior parietal regions. Indeed, while showing increased activation even in the absence of any sensory input (Kastner et al., 1999; Hopfinger et al., 2000), the parietal cortex shows some residual spatially-specific responses also to unattended stimuli (Saygin and Sereno, 2008). The co-occurrence of sensory-driven responses and top-down internally - controlled activity, also associated with motor planning, has triggered an intense debate about the format of the spatial representations in parietal cortex (Andersen et al., 1993; Colby and Goldberg, 1999). Previous studies considered the role of the dorsal-fronto-parietal cortex for the processing of bottom-up signals within two main frameworks. On one hand, it has been proposed that activation of the dorsal system "follows" the detection of relevant stimuli by the ventral attention system (Geng and Mangun, 2011; Vossel et al., 2012). Accordingly to this view the dorsal system generates top-down signals that modulate processing in visual cortex, which require re-setting when a relevant/infrequent "bottom-up" stimulus is presented outside the current focus of attention (Corbetta and Shulman, 2002; Shulman et al., 2009). A different prospective entails a more "direct" activation of the dorsal system due to forward input from the visual cortex. In this context, differences between the involvement of the posterior IPS and the anterior FEF nodes of the dorsal fronto-parietal system have been proposed. For example, using concurrent *Transcranial Magnetic Stimulation* (TMS) - fMRI over either IPS or FEF, Ruff et al. (2008) showed that the inter-regional influence of FEF over visual cortex does not depend on the visual input, while the influence of IPS can change according to the current visual context. These findings demonstrated that in complex and naturalistic experimental situations that involve high levels of sensory competition, the dorsal fronto-parietal system responds to bottom-up salient signals. In addition, despite the common assumption that mechanisms of attentional reorienting are predominantly - if not exclusively - homed in the right hemisphere (Shulman et al., 2010), a number of studies have reported bilateral rather than right unilateral activation of the *Temporo-Parietal Junction* (TPJ) in response to unexpected and invalid targets (Bartolomeo et al., 2007; He et al., 2007; Asplund et al., 2010; Doricchi et al., 2010; Verdon et al., 2010). Proper level of attention is crucial in complex

and operational environments. It has been demonstrated that under high workload conditions, attention appears to channel or tunnel, reducing focus on peripheral information and tasks and centralizing focus on main tasks (Staal, 2004). This tunnelling of attention can result in either enhanced performance or reduced performance, depending on the nature of the task and the situation. For instance, when peripheral cues are irrelevant to task completion the ability to tune them out is likely to improve performance. On the other hand, when these peripheral cues are related to the task and their incorporation would otherwise facilitate success on the task, performance suffers when they are unattended. Furthermore, high workload might cause a reduction in environmental sampling (visual scanning) and that this results in distortions in the perception of space and time (Hancock and Weaver., 2002). Several groups have studied the effects of time pressure on attentional processes. Entin and Serfaty (1990) examined the decision-making of military personnel under the stress of time pressure in a combat simulation study. Their results suggested a reduction in the frequency and/or amount of information sought by decision makers as well as a reduction in the accurate detection of friend or foe submarines under high time pressure conditions. Ozel (2001) examined how people process environmental information for exiting (fire fighting scenarios) under time pressure and threat of fire. Based on the existing research literature, Ozel theorized that when slightly stressed, one's ability to determine the best time to exit was likely to be enhanced, while under high workload demand there would probably be a restriction in the range of cues attended to and a distortion of information processing. He asserted that the result was likely to be a decrement in performance. In addition, several studies support the idea that attentional tunnelling can be reproduced and reversed pharmacologically. Caldwell and his colleagues (2001; Caldwell and Gilreath, 2002) have examined this line of research extensively on U.S. Army aviation personnel. For example, performance on several flight measures was enhanced after the introduction of dextroamphetamine (Dexedrine R): airspeed, heading, roll, and turning control. Furthermore, self ratings of mood and physiologic measures of alertness and attention (predominantly arousal measured by EEG) also improved with pharmacological intervention. Similar findings have been obtained by the use of modafinil (Provigil R). This is certainly not a new assertion and many researchers have reported such findings in

military aviation. For example, dextroamphetamine has been used in sustained and continuous flying operations in the military with significant success in the past (Cornum, 1992; Enmonson and Vanderbeek, 1995). High workload-induced cognitive tunnelling or narrowing of attention has also been causally linked to airline accidents and crash sequences (Kornovich, 1992). Due to the high workload and inherent stressors associated with flying, the design of aviation systems has been a priority of various human factors engineers and cognitive scientists. Several authors have explored the relationship between HUD (Head Up Display) symbology used in aircraft operations and attentive processes (Dowell et al., 2002; Foyle et al., 2001). The results of these investigations suggest that such symbology is so compelling that it fosters cognitive tunnelling under the workload of flying. The result is a decrease in overall situational awareness and a greater vulnerability to error (Wickens et al., 1998). These authors have also suggested that various design modifications and compensatory strategies appeared to be available to reduce this risk. For instance, in a similar investigation, Wilson et al. (2002) found that situation-guided symbology in HUDs led to increased situational awareness, increased taxi speeds, and less workload. They surmised that this was at least in part due to a reduction in cognitive tunnelling. Beilock et al. (2002) explored the effect of attention on sensorimotor skills. They found that well-learned skills do not require deliberate, conscious control of attention and as a result dual-task performance is easier. Given that well-learned skills tend to require fewer mental resources for their performance, more resources are available to devote to additional tasks accordingly. When prompted to focus attention to a particular component of the well-learned task, performance was degraded. It appears that the step-by-step attention to tasks is beneficial during initial learning stages but that this tends to be detrimental once skills are well-learned.

2.2.3. Decision Making

Decision -making (DM) can be defined as a problem-solving activity resulting in the selection of a belief or action among alternative possibilities based on the values and preferences of the decision-maker. It is therefore a process which can be more or less rational or irrational, and can be based on explicit knowledge or tacit knowledge (Staal, 2004). When facing with a decision, individuals are wired to employ the most adaptive heuristic available (Gigerenzer and Selten, 2001). One of the most famous heuristic is the *Take-the-Best* (TTB). This model proposes that decision makers search their memories for criterion-linked probability information. Gigerenzer et al. (1991) and Broder (2003) suggested that this search is bounded by cognitive economy. In other words, the search for information is quick, streamlined, and tends to rely on the most valid probabilistic cue separating alternatives. In other words, the authors challenged the notion that individuals reasoned from "multiple conditional probabilities," instead opting for a single discriminatory cue. When individuals are engaged in a high working memory loaded task, their capacity for other information processing operations is likely to be compromised. Broder (2003) conjectured that, if this is so, then individuals engaged in a decision making task would be more likely to rely on a simple heuristic than a complex one. They would be content with limiting their scan of alternatives in the name of resource management. On the contrary, Dougherty and Hunter (2003) suggested that individuals tend to make probability judgments by comparing a focal hypothesis with relevant alternatives retrieved from long-term memory. They also concluded that individuals with a large working memory span tend to include more alternatives in their comparison process, and that time constraints probably truncate the alternative generation process so that fewer alternatives are recalled from long-term memory for comparison. In general, decision making is degraded under stressful conditions. For example, Kastner et al. (1989) employed a naval combat simulation task in their examination of the effects of task difficulty on decision making processes. These authors designated subjects as anti-submarine warfare commanders charging them with the task of friend or foe determinations. Decision makers were provided access (at a cost) to either a human consultant or a probe sensor. The results indicated that subjects

tended to seek more information than was optimal and that such information tended to be preferentially sought from consultants who had already processed the data themselves. In an extension of this work, Entin and Serfaty (1990) explored a very similar paradigm. In their experiment, subjects were under the dual "stress" of time pressure and a secondary task that intruded on their primary decisions. The Authors found that decision makers tended to seek out more information in difficult versus easy discrimination tasks, and they typically followed a rational cost-benefit rule in their decisions. This was particularly the case as time pressure and workload increased. Speier et al. (1999) investigated the effects of interruptions on individual decision making noting that interruptions make information overload worse by reducing the amount of time one has to spend working on the problem, which in turn leads to feeling time pressured. This creates both capacity, too much information to process, and structural, inputs that are occupying the same physiological channel, interference (i.e., monitoring two visual displays at once). Moreover, interruptions also place greater demand on the cognitive processing system. For example, information that is forgotten due to overload requires further resources to re-process or simply never gets encoded. Smith (1979) and Wickens et al. (1991) investigated the effects of increased workload on pilots. The results indicated that periods of increased workload, even those considered to be part of normal flight operations, 1) increased the frequency and volume of errors, 2) reduced the cue sampling; 3) reduced the resource-limited capacity of working memory, and 4) when time was reduced, a speed-accuracy trade-off in performance outcome was adopted. Stokes (1995), Wiggins et al. (1995) and Li et al. (2001) further examined pilot performance on an aeronautical decision making task. They found that novice pilots made poorer decisions than expert pilots and their decision making was further degraded under stress while expert pilots decision making was not. Additionally, experienced pilots tended to visually search fewer screens of information, return to previously scanned information screens less often, and spent less time examining these screens than novice pilots. Thus, expertise in and of itself helps, but it may be that task-specific experience is most relevant to time-critical decision making. This finding highlights the potential importance of realistic emergency and abnormal situation training, for in no other way can novice operators (regardless of

the industry they represent) safely garner emergency decision making expertise. Sperandio (1971) explored this strategy in his examination of Air Traffic Controllers (ATCOs). He found that controllers tend to exercise a regulating effect on their workload through strategy shifting. Observing ATC operations, Sperandio found that controllers under increasing workload conditions tended to shift from a direct approach strategy (data needed to verify and possibly achieve separation between aircraft) to a standard approach strategy (data concerning aircraft performance not needed for separation of aircraft). Sperandio concluded that controllers economized their workload with a shift in strategy or method, reducing any redundant information or non-essential information from being processed. These findings are consistent with previously reported research on expert versus novice decision making (Stokes, 1995; Lansdown, 2001).

From the neuro – physiological perspective, several brain regions situated in the frontal lobes have been demonstrated to be related to decision making processes. Specifically, there are evidences of both the *Ventromedial Prefrontal Cortex* (VMPFC) and the *Orbitofrontal Cortex* (OFC) being involved in the processing of different alternatives and potential outcomes through the assessment of their (perceived) value (Tremblay and Schultz, 1999; Daw et al., 2006). In particular, the OFC is associated with the evaluation of trade-off and the expected capacity of outcomes to satisfy the user's needs (Wallis, 2007). It also plays a central role in making choices about appropriate behaviors, especially when one is faced with unpredictable situations (Elliott et al., 2000). Finally, the *Dorsolateral Prefrontal Cortex* (DLPFC) also plays a critical role in decision-making as it is known for being involved in cognitive control over emotions (Rilling et al., 2008). In particular, it is involved in the control of impulses for complying with social norms while the *Ventrolateral Prefrontal Cortex* (VLPFC) could play a role in motivating this social norm compliance by representing the threat of punishment from others (Rilling and Sanfey, 2011). Interestingly, the cognitive effort in the PFC appears to be lower when a sure gain is expected compared with risky decisions (Gonzalez et al., 2005).

2.3. Machine-Learning

Machine learning is a discipline focused on two interrelated questions: *How can one construct computer systems that automatically improve through experience?* and *What are the fundamental statistical computational-information-theoretic laws that govern learning systems?* The study of machine learning is important both for addressing these fundamental scientific and engineering questions and for the highly practical computer software it has produced and fielded across many applications (Jordan and Mitchell, 2015). Machine learning has progressed dramatically over the past two decades, from laboratory curiosity to a practical technology in widespread commercial use. Within artificial intelligence (AI), machine learning has emerged as the method of choice for developing practical software for computer vision, speech recognition, natural language processing, robot control, and other applications. Many developers of AI systems now recognize that, for many applications, it can be far easier to calibrate a system by showing it examples of desired input - output behaviour than to program it manually by anticipating the desired response for all possible inputs. The effect of machine learning has also been felt broadly across computer science and across a range of industries concerned with data - intensive issues, such as consumer services, the diagnosis of faults in complex systems, and the control of logistics chains. There has been a similarly broad range of effects across empirical sciences, from biology to cosmology to social science, as machine-learning methods have been developed to analyze high throughput experimental data in novel ways. In particular, under the recognition of the user's mental states (e.g. mental workload), such techniques should be able to extract from the big amount of physiological data the most significant characteristics closely related to the examined mental state, and then it can be used to assess the examined user's mental state on different working – days.

A diverse array of machine-learning algorithms has been developed to cover the wide variety of data and problem types exhibited across different machine-learning problems (Murphy, 2012; Hastie et al., 2011). Conceptually, machine-learning algorithms can be viewed as searching through a large space of candidate programs, guided by training experience, to find a program that optimizes the performance metric. Machine-learning

algorithms vary greatly, in part by the way in which they represent candidate programs (e.g., decision trees, mathematical functions, and general programming languages) and in part by the way in which they search through this space of programs (e.g., optimization algorithms with well-understood convergence guarantees and evolutionary search methods that evaluate successive generations of randomly mutated programs).

Many algorithms focus on function approximation problems, where the task is embodied in a function (e.g., given an input transaction, output a "fraud" or "not fraud" label), and the learning problem is to improve the accuracy of that function, with experience consisting of a sample of known input-output pairs of the function. In some cases, the function is represented explicitly as a parameterized functional form; in other cases, the function is implicit and obtained via a search process, a factorization, an optimization procedure, or a simulation-based procedure. Even when implicit, the function generally depends on parameters or other tunable degrees of freedom, and training corresponds to finding values for these parameters that optimize the performance metric. Whatever the learning algorithm, a key scientific and practical goal is to theoretically characterize the capabilities of specific learning algorithms and the inherent difficulty of any given learning problem: *How accurately can the algorithm learn from a particular type and volume of training data? How robust is the algorithm to errors in its modeling assumptions or to errors in the training data? Given a learning problem with a given volume of training data, is it possible to design a successful algorithm or is this learning problem fundamentally intractable?* Such theoretical characterizations of machine-learning algorithms and problems typically make use of the familiar frameworks of statistical decision theory and computational complexity theory. In fact, attempts to characterize machine-learning algorithms theoretically have led to blends of statistical and computational theory in which the goal is to simultaneously characterize the sample complexity (how much data are required to learn accurately) and the computational complexity (how much computation is required) and to specify how these depend on features of the learning algorithm such as the representation it uses for what it learns (Chandrasekaram and Jordan, 2013; Shalev-Shwartz et al., 2012; Decatur, 2000). As a field of study, machine learning sits at the crossroads of computer science, statistics and a variety of other disciplines concerned with

automatic improvement over time, and inference and decision-making under uncertainty. Related disciplines include the psychological study of human learning, the study of evolution, adaptive control theory, the study of educational practices, neuroscience, organizational behavior, and economics. Although the past decade has seen increased crosstalk with these other fields, we are just beginning to tap the potential synergies and the diversity of formalisms and experimental methods used across these multiple fields for studying systems that improve with experience.

Supervised Learning Systems

The most widely used machine-learning methods are supervised learning methods. Supervised learning systems, including spam classifiers of e-mail, face recognizers over images, and medical diagnosis systems for patients, all exemplify the function approximation problem where the training data take the form of a collection of (x, y) pairs and the goal is to produce a prediction y^* in response to a query x^*. The inputs x may be classical vectors or they may be more complex objects such as documents, images, DNA sequences, or graphs. Similarly, many different kinds of output y have been studied. Much progress has been made by focusing on the simple binary classification problem in which y takes on one of two values (for example, "spam" or "not spam"), but there has also been abundant research on problems such as multiclass classification, where y takes on one of K labels. Supervised learning systems generally form their predictions via a learned mapping $f(x)$, which produces an output y for each input x (or a probability distribution over y given x). Many different forms of mapping f exist, including decision trees, decision forests, *linear regression* (LDA), *support vector machines* (SVM), *neural networks* (NN), kernel machines, and Bayesian classifiers.

Unsupervised Learning Systems

While much of the practical success in deep learning has come from supervised learning methods, efforts have also been made to develop deep learning algorithms that discover useful representations of the input without the need for labelled training data (Hinton and Salakhutdinov, 2006). The general problem is referred to as unsupervised learning, a second paradigm in machine-learning research (Murphy, 2012). Broadly, unsupervised

learning generally involves the analysis of unlabelled data under assumptions about structural properties of the data (e.g., algebraic, combinatorial, or probabilistic). For example, one can assume that data lie on a low- dimensional manifold and aim to identify that manifold explicitly from data. Dimension reduction methods, including principal components analysis, manifold learning, factor analysis, random projections, and auto-encoders (Murphy, 2012; Hastie et al., 2011) make different specific assumptions regarding the underlying manifold (e.g., that it is a linear subspace, a smooth nonlinear manifold, or a collection of submanifolds).

Reinforcement Learning System

A third major machine-learning paradigm is reinforcement learning (Mnhi et al. 2015; Sutton and Barto, 1998). Here, the information available in the training data is intermediate between supervised and unsupervised learning. Instead of training examples that indicate the correct output for a given input, the training data in reinforcement learning are assumed to provide only an indication as to whether an action is correct or not; if an action is incorrect, there remains the problem of finding the correct action. More generally, in the setting of sequences of inputs, it is assumed that reward signals refer to the entire sequence; the assignment of credit or blame to individual actions in the sequence is not directly provided.

Although these three learning paradigms help to organize ideas, much current research involves blends across these categories. For example, *semi - supervised* learning makes use of unlabelled data to augment labelled data in a supervised learning context, and discriminative training blends architectures developed for unsupervised learning with optimization formulations that make use of labels.

2.3.1. Machine - Learning Models

Machine learning model selection does not mean only to choose the model that better fits the calibrating data, but also to select from a family of models. Many issues influence the design of learning algorithms across all of the previous paradigms, including whether data are available in batches or arrive sequentially over time, how data have been sampled, requirements that learned models be interpretable by users, and robustness issues that arise when data do not fit prior modelling assumptions. Additionally, from a neuro – physiological perspective, different cognitive processes and mental states may result in different patterns of brain activations. Machine – learning algorithms are seen as pattern recognition techniques that classify each pattern into a class according to its features. The features are measured or derived from the properties of the signals which contain the discriminative information needed to distinguish their different types. The design of a suitable set of features is a challenging issue. The information of interest in brain signals is hidden in a highly noisy environment, and brain signals comprise a large number of simultaneous sources. A signal that may be of interest could be overlapped by multiple signals depending on different brain processes or artifacts. Therefore, one of the critical point of machine – learning techniques is the selection of the relevant features within the available ones. High dimensional feature vectors are not desirable due to the "curse of dimensionality" in calibrating the algorithms. In other words, the feature selection may be attempted examining all possible subsets of the features, thus it might take a long time and computational demand. However, the number of possibilities grows exponentially, making an exhaustive search impractical for even a moderate number of features.
The aim of BCI applications is the recognition n of a user's intentions on the basis of a feature vector that characterizes the brain activity provided by the feature dataset. Either regression or classification algorithms can be used to achieve this goal (Lotte et al., 2007). Regression algorithms employ the features extracted from EEG signals as independent variables to predict user intentions. In contrast, classification algorithms use the features extracted as independent variables to define boundaries between the different classes in feature space. McFarland and Wolpaw (2005) illustrated the differences between the two alternatives. For a two-data classes case, both the

regression and the classification approach require the parameters of a single function to be determined. In a four-data classes case, assuming that the targets are distributed linearly, the regression approach still requires only a single function. In contrast, the classification approach requires the determination of three functions, one for each of the three boundaries between the four classes. Therefore, the classification approach might be more useful for two-target applications and the regression approach may be preferable for greater numbers of targets, when these targets can be ordered along one or more dimensions. Classification algorithms can be developed via either offline, online or both of conditions (Nicolas – Alonso and Gomez – Gil, 2012). The offline session involves the examination of dataset. The statistics of the data may be estimated from observations across entire sessions and long-term computations may be performed. The results can be reviewed by the analyst with the aim of fine-tuning the algorithms. In contrast, online sessions provide a means of evaluation in a real-world environment. The data are processed in a causal manner and the algorithms are tested in an environment in which the users change over the time as a result. Although some researchers test new algorithms with only offline data, both offline simulation and online (or simulated online) experiments are necessary for effective algorithm design in closed-loop systems. In other words, offline simulation and cross-validation can be valuable methods to develop and test new algorithms, but only online (or simulated online) analysis can yield solid evidence of machine - learning system performance (Townsend et al., 2010; Daly and Wolfgang, 2008; McFarland et al., 2006). Traditional classification algorithms have been calibrated through labelled dataset. It is assumed that the classifier is able to detect the patterns of the brain signal recorded in online sessions with feedback. The design of the classification step involves the choice of one or several classification algorithms from many alternatives. Linear and nonlinear classification algorithms have been proposed in the literature, such as *k*-nearest neighbour classifiers, linear classifiers, support vector machines, and neural networks, among others (Table 1). Linear methods assume a linear separabilty of the data classes. The separability of the classess is measured by two quantities: how far are the projected class means apart (should be large), and how big is the variance of the data in this direction (should be small). This can be achieved by maximizing the so-called *Rayleigh coefficient* of between and

within class variance (Fisher, 1936; Fukunaga, 1990). On the contrary, when the classes are not separable linearly, or the source of the data to be classified is not well understood, methods for finding good nonlinear transformations are required. For example, the classification in input space requires some complicated nonlinear (multiparameter) ellipsoid classifier, and appropriate feature space representation.

Finally, certain inherent dangers of classification algorithm usage should be pointed out. Although classification algorithms have clearly helped to characterize task relevant brain states, several pitfalls may occur when these algorithms are used. Bias and variance of the estimated error of the algorithms, and their overfitting are the main source of problems (Lemm et al., 2011). If a classifier is overfitted, then it will only be able to classify classes in calibration or similar data. Classification error is estimated by means of cross - validation. Once a classification algorithm is calibrated, the algorithm is validated on a testing dataset, which should be independent on the calibration dataset. This procedure is usually repeated several times, using different partitions of the calibration data. The resulting validation errors are averaged across multiple rounds (cross – validations). This approach presents some inherent dangers that must be prevented, because some elements of the partition may not be independent of each other (*auto-validation*) or may not be identically distributed, among other reasons (Lemm et al., 2011).

Linear Discriminant Analysis (LDA)

Linear Discriminant Analysis (LDA) is usually applied to classify patterns into two classes, although it is possible to extend the method to multiples classes (Garrett et al., 2003). For a two-class problem, LDA assumes that the two classes are linearly separable. According to this assumption, LDA defines a linear discrimination function $y(x)$ which represents a hyperplane in the feature space in order to distinguish the classes. The class to which the feature vector belongs will depend on the side of the plane where the vector is found. In the case of an N-class problem ($N > 2$), several hyperplanes are used.

Table 1: Summary of the most popular machine learning techniques.

Generative model	Approach	Properties
Linear	Bayesian analysis	− Assigns the observed feature vector to the labeled class to which it has the highest probability of belonging. − Produces nonlinear decision boundaries. − Not very popular in the BCI systems.
	LDA	− Simple classifier with acceptable accuracy. − Low computation requirements. − Fails in the presence of outliers or strong noise. Regularization required. − Usually two class. Extended multiclass version exits. − Improved LDA versions: BLDA, FLDA.
	SWLDA	− Extention of the LDA that performs feature space reduction by selecting suitable features to be included in the discriminant analysis. − Combination of forward and backward LDA. − Multi – classes analysis.
NonLinear	SVM	− Linear and non-linear (Gaussian) modalities. − Binary or multiclass method. − Maximizes the distance between the nearest training samples and the hyperplanes. − Fails in the presence of outliers or strong noise. − Regularization required. − Speedy classifier.
	k - NNC	− Uses metric distances between the test feature and their neighbors. − Multiclass. − Efficient with low dimensional feature vectors. − Very sensitive to the dimensionality of the feature vectors.
	ANN	− Very flexible classifier. − Multiclass. − Multiple architectures (PNN, Fuzzy ARTMAP ANN, FIRNN, PeGNC)

The decision plane can be represented mathematically as:

$$y(\boldsymbol{x}) = w^T \boldsymbol{x} + w_0 \quad (2.1)$$

where, *w* is known as the weight vector, **x** is the input feature vector and w_0 is a bias. The input feature vector is assigned to one class or the other on the basis of the sign of y(**x**).
There are many methods to compute *w*. For example, *w* may be calculated as (Fisher, 1936):

$$w = \sum_c^{-1} (\mu_2 - \mu_1) \quad (2.2)$$

where, μ_i is the estimated mean of the class *i*, and $\sum_c = \frac{1}{2} (\sum_1 + \sum_2)$ is the estimated common covariance matrix. The estimators of the covariance matrix and of the mean are calculated as:

$$\sum_c = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \mu)(x_i - \mu)^T \quad (2.3)$$

$$\mu = \frac{1}{n} \sum_{i=1}^{n} x_i \quad (2.4)$$

where, **x** is a matrix containing *n* feature vectors $x_1, x_2, \ldots, x_n \in \mathbb{R}^d$. The estimation of the covariance defined in Equation (2.3) is unbiased and has good properties under usual conditions. Nevertheless, it may become imprecise in some cases where the dimensionality of the features is too high compared to the number of available trials. The estimated covariance matrix is different from the true covariance matrix, because the large eigenvalues of the original covariance matrix are over estimated and the small eigenvalues are under estimated. It leads to a systematic error which degrades LDA performance (Blankertz et al., 2011). For this reason, a new procedure has been proposed to estimate the covariance, improving the standard estimator defined in the Equation (2.3).

The new standard estimator of the covariance matrix is given by:

$$\sum_c(\gamma) = (1 - \gamma)\sum_c + \gamma\nu I \quad (2.5)$$

The γ value is referred to as a shrinkage parameter and is tunable between 0 and 1. ν is defined as $\mathrm{trace}(\sum_c)/\mathrm{d}$ with d being the dimensionality of the features space. The selection of a shrinkage parameter implies a trade-off and is estimated on the basis of the input data (Vidaurre et al., 2009). Some improved algorithms have been introduced based on LDA (Hoffmann et al., 2008), such as *Fisher LDA* (FLDA) and *Bayesian LDA* (BLDA). In the first example, performance was improved by projecting the data to a lower dimensional space, in order to achieve larger intervals between the projected classes and, simultaneously, to reduce the variability of the data in each class. However, FLDA does not work well when the number of features becomes too large in relation to the number of calibrating examples. This is known as the small sample size problem (Hoffmann et al., 2008). The second modification can be seen as an extension of the FLDA. BLDA solves the small sample size problem by introducing a statistical method known as *regularization*. The regularization is estimated through Bayesian analysis of calibrating data and is used to prevent *overfitting* of high dimensional and possibly noisy datasets In comparison to FLDA, the BLDA algorithm provides higher classification accuracy and bit-rates, especially in those cases where the number of features is large (Hoffmann et al., 2008). Additionally, BLDA requires only slightly more computation time, which is a crucial requirement in real BCI systems.

Stepwise Linear Discriminant Analysis (SWLDA)

The SWLDA is an extension of the FLDA which performs a reduction of the features' space by selecting the most significant ones. Particularly, the SWLDA consists of combination of forward selection and backward elimination analyses, where the input features are weighted using ordinary least-squares regression to predict the target class label. The method starts by creating an initial model of the discriminant function with the most statistically significant feature for predicting the target labels (e.g. p-value < 0.5). Then at each new step a new term is added to the model and a backward stepwise analysis is performed to remove the least statistically

significant feature (e.g. p-value > 0.1). This process goes on until the predefined number of significant features is reached or until there are no features satisfying the entry/removal condition (Draper and Smith, 2014).
Regarding the mental workload evaluation, it has been demonstrated that linear classifiers outperformed the nonlinear ones (Craven et al., 2006; Krusienski et al., 2006). In particular, the SWLDA has been chosen since it was shown to be one of the best classifiers for the real-time mental workload evaluation (Rabbi et al., 2009; Berka et al., 2007). Generally, several cognitive processes interact each other during the execution of a task. Since the aim of my PhD is to track the variation of only a specific phenomenon (i.e. mental workload or training), the SWLDA has been selected due to the possibility to set its parameters, hence statistic criteria, for the features selection (see paragraph 3.1 for more details). In fact, the SWLDA is able to weight the contribution of the features to the classification model, and to find out if each feature has to be inserted into or removed from the model. As consequence, by properly acting on the SWLDA's features selection parameters (see paragraph 3.1), if the considered phenomenon is sufficiently emphasized by the task, it will be possible to hook and track the variation of such phenomenon ongoing the task execution. In fact, the SWLDA will consider (into the model) only those brain features able to describe univuquely the examined cognitive phenomenon, and which will be always turned up by the execution of the proposed task, therefore stable over time.

2.4. Experimental Tasks

2.4.1. NASA - Multi Attribute Task Battery (MATB)

The NASA - *Multi Attribute Task Battery* (MATB) is a computer-based task designed by the NASA to evaluate pilots and operator performance and workload (Comstock and Arnegard, 1992). It could be freely download from the NASA Website at the following address: http://matb.larc.nasa.gov/. It is a standard platform for the evaluation of the cognitive operational capability, since it could provide different tasks that have to be attended by the subject in parallel, as well as each task could be modulated greatly in difficulty (Figure 10).

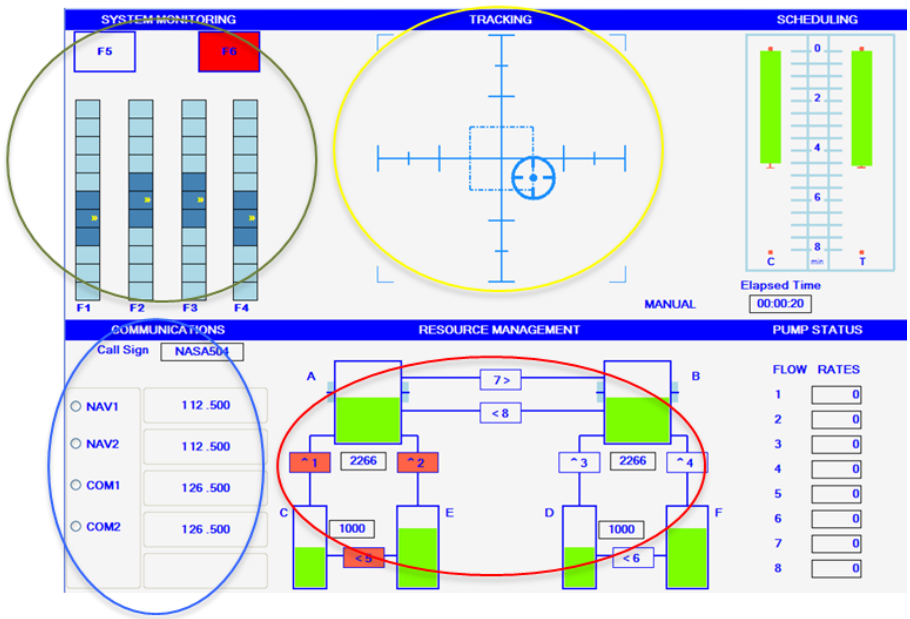

Figure 10. NASA - *Multi Attribute Task Battery* (MATB) interface. On the top left corner (green circle), there is the emergency lights sub-task; on the top, in the center (yellow circle), there is the cursor tracking task; on the left bottom corner (blue circle), there is the radio communication task and, finally, in the center on the bottom (red circle), there is the fuel managing task.

In particular, the experimental software requires the simultaneous performance of tasks that are generalizations of piloting tasks: tracking (TRCK), auditory monitoring (COMM), resource management (RMAN),

and response to event onsets (SYSM).The demand of the system monitoring task is monitoring the gauges and the warning lights by responding to the absence of the Green light, the presence of the Red light, and monitors the four moving pointer dials for deviation from midpoint. The demands of manual control are simulated by the tracking task. The subject keeps the target at the centre of the window by moving the joystick. This task can be automated to simulate the reduced manual demands of autopilot. Subjects are also required to respond to a communication task. This task presents pre-recorded auditory messages to the operator at selected intervals during the simulation. However, not all of the messages are relevant to the operator. The goal of the COMM task is to determine which messages are relevant and to respond by selecting the appropriate radio and frequency on the communications task window. The demands of fuel management are simulated by the resource management task. The goal is to maintain the main tanks at 2500 (lbs). This is done by turning On or Off any of the eight pumps. Pump failures occur when they are red colored. Four performance indexes (one for each subtask) have been defined by means of the log files provided by the MATB at the end of the task execution. An example of log file is reported in Figure 11. The index for the TRCK task has been defined by considering the complement of the ratio between the cursor's distance got by the subject and the maximum of this distance (fixed) from the center of the screen. The indexes of the COMM and SYSM tasks have been defined as a linear combination of accuracies in terms of correct answers (e.g., correct radio or frequency selected) and the complement of the ratio between the subject's reaction time and the maximum time for answering; then, the results have been multiplied for "100" in order to obtain a percentage. Finally, the index for the RMAN task has been defined as the mean value of the fuel's levels in the tank A and B, multiplied by "100". In order to get a global performance index, the average of the previous indexes has been calculated, as the single index was a percentage.

```
# 10-16-2012     13.38.49     COMM_2012_10161338.txt
#
# Events Filename: MATB_EVENTS.xml
#
# Timeout (in seconds) = 30
#
# RT = Response Time (in seconds)
#
# _T = Target, _S = Selected
# R_ = Radio,  F_ = Frequency
#
#-TIME-      -RT-    -SHIP-  -RADIO_T-  -FREQ_T-  -RADIO_S-  -FREQ_S_  -R_OK-   -F_OK-   -REMARKS-
#-------------------------------------------------------------------------------------------------------
 00:00:39.0                                                                              - Event Timer Reset
 00:00:47.3  02,1     OWN     COM1       127.550   COM1       127.550   True     True
 00:01:05.0                                                                              - Event Timer Reset
 00:01:17.1  02,3     OWN     COM2       125.575   COM2       125.675   True     False
 00:01:45.5                                                                              - Event Timer Reset
 00:01:58.1  01,7     OWN     NAV1       110.650   NAV1       110.650   True     True
 00:02:25.0                                                                              - Event Timer Reset
 00:02:35.5  03,9     OWN     NAV1       113.500   NAV1       113.500   True     True
 00:03:17.9   +20     OTHER   NAV1       117.650   NULL       NULL      True     True     - Appropriate Subject Response
 00:03:30.7  03,3     OWN     COM2       124.450   COM2       124.450   True     True
 00:04:05.5                                                                              - Event Timer Reset
 00:04:25.0                                                                              - Event Timer Reset
 00:04:33.5  01,3     OWN     NAV1       111.500   NAV1       111.500   True     True
```

Figure 11. Example of log file provided by the MATB at the end of the task execution. The performance indexes have been defined as combination of *Reaction Time* (RT) and correctness of the subject's responses.

2.4.2. ATM Environment

Air traffic management (ATM) is a term largely used to encompass all systems that assist aircraft to depart from an aerodrome, transit airspace, and land at a destination aerodrome. It includes different areas such as *Air Traffic Control* (ATC), *Air Traffic Flow Management* (ATFM) and *Aeronautical Information Services* (AIS). Rather than discussing all these aspects of the Air Traffic Management Systems, the following studies will focus on the area of ATC, in which an *Air Traffic Controller* (ATCO) monitors the movement of a number of aircraft in his area of responsibility, typically via flight progress strips and a radar screen and provides control instructions to the pilots if need be, e.g. in case of risk of violation of the legal safety separations between aircraft. Compared to other safety critical and high-hazard domains, such as nuclear power, air traffic control is characterized by the key role played by human actors. Despite advances in technology, the ATM system is still human-centred. As a matter of fact, safety relevant decisions are made mostly by humans, whereas computer systems are supporting tools, assisting the controller in monitoring and communication tasks. In the review about the levels of automation, Parasuraman et al. (2000) argued that safety critical systems, such as air traffic control system, are typically less automated. Automation might be

higher in information acquisition and information analysis tasks (provided the automation is reliable) and lower for decision selection and moderately high for action implementation. Indeed, in ATM system the controller's work is very cognitively demanding. Activities for managing traffic, such as solving conflicts, maintain separations between aircraft and coordinating traffic involve cognitive processes such as visual scanning, managing situation awareness, decision making and attention control. Controllers have a key-role in facing system complexity, because their main objective is to anticipate and manage unpredictable situations affected by multiple elements. Complexity does not regard only the environment. Radio communications, phone communications, radar displays and computers are all system elements, adding to its complexity. A peculiar characteristic of the ATM work is that ATCOs usually perform few recurring tasks – apart perhaps from routine tasks such as welcoming aircraft on the frequency and handing them off to adjacent sectors. Anyway, even if these tasks are well-known, their order remains largely unpredictable because of the dynamic nature of weather, traffic demands, operational conditions, and so on. ATC research could benefit greatly from psyco-physiological approaches, i.e. the ability to measure certain physiological indicators known to be linked to underlying mental and physiological processes (since the work of air traffic controllers involves little physical activity mental processes will be the main concern, but physiological artifacts are of some concern, e.g. eye blink artefacts on certain EEG signals). Amongst these psychophysiological approaches, neurometrics, i.e. measuring the brains electrical activity and inferring from these on the activity of the brain is certainly very promising. In this regard, specific experiments and *ad-hoc* ATM scenarios have been designed by expert Controllers from the Training Centre of the *École Nationale de l'Aviatione Civile* (ENAC) of Toulouse (France). In addition, professional and students *en-route* Controllers have been recruited and involved in the experiments at ENAC. The creation of the ATM tasks has started from a scenario already used for training and it has been modified according to the objectives and needs of the experiments. The adjustments mostly impacted the routes and air-traffic within the air-space sector (adding or changing them), while the borders of the sectors have not been changed. The air-traffic adjustments has consisted in flights rescheduling and flight

levels (FLs) changing. Three ATM scenarios have been defined according to the *Number of aircraft*, *Traffic geometry*, and *Number of conflicts*.
The scenarios have been created with different air-traffic geometries but similar difficulty levels, and each of them has been defined as combinations of three air-traffic level, *Easy* (E), *Medium* (M) and *Hard* (H), with the aim to simulate realistic ATC conditions. The total duration of the scenario is 45 minutes, while the easy, medium and hard level lasts 15 minutes. During the execution of the ATM task, the controller may assume an aircraft before it enters the sector and could hand over an aircraft before it actually leaves the sector. In other words, the ATCO could anticipate the heavy traffic load when they are still in the easy period and, consequently, change the duration of the next condition. The estimated level of difficulty, traffic evolution and the potential conflicts are reported in the following for each of the considered ATM scenario. The reported time of conflict is the time of the closest point of approach of two (or three) aircrafts. It means that the de-conflicting activity of the controller is ahead of such time. The real time can not be provided as it might be different for each controller, but the presented values still represent the difficulty of the scenario with a shift in time that can not be evaluated.

Scenario #1: The scenario starts with an easy air-traffic condition of 15 minutes and then it passes through a medium air-traffic (5 minutes) before reaching the hard condition (15 minutes). Then, the last 10 minutes presents a medium air-traffic load (Figure 12).

Figure 12. Time schedule of the ATM Scenario #1. The air-traffic load transition is easy (15 minutes), medium (5 minutes), hard (15 minutes) and, finally, 10 minutes of medium air-traffic.

Figure 13 reports the number of airplanes across the Scenario #1. In the first 15 minutes, easy conditions, there are only 2 - 4 airplanes, while in the medium and hard conditions there are, on average, 8 and 15 aircrafts. As quoted previously, since the ATM simulation had to be as much realistic as possible, the air-traffic transitions have not been abrupt but, on the contrary,

gradual. Table 2 presents the potential flight conflicts during the Scenario #1. Most part of the conflicts have been inserted within the hard phase (8:10 – 8:25) in order to further increase the difficulty of the condition with respect to the other ones.

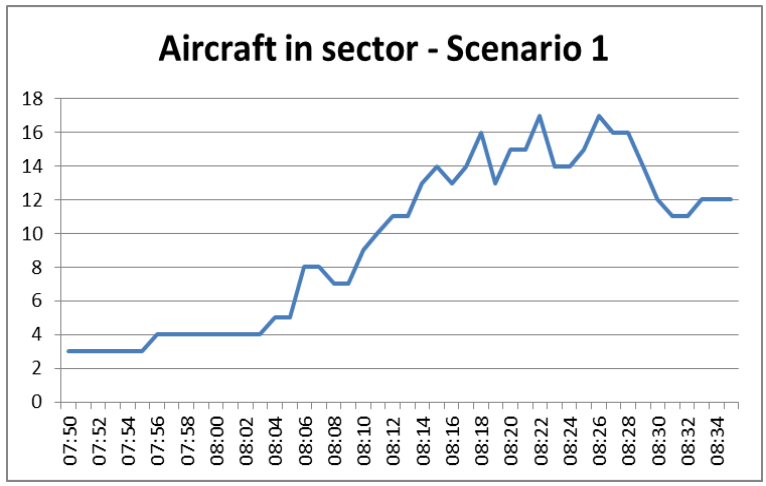

Figure 13. Number of airplanes across the Scenario #1. In the first 15 minutes, easy conditions, there are only 2 - 4 airplanes, while in the medium and hard conditions there are, on average, 8 and 15 aircrafts.

Table 2: Potential flight conflicts within the Scenario #1.

Time	Id1	Id2	Id3	Flight Level	Comment
08:06	AIB003				
08:07	DLH891	RYR1219		FL370	
08:10	KLM1222	AFR1732		FL340	
08:15	RAE569	EZY835	BMM716	FL360	
08:15	TAP537	RYR1219		FL370	
08:17	RYR517	(JKK657 could be conflictual)			Speed is challenging
08:22	NJE017	FGHDX		FL370	
08:22	SWR199				
08:26	AMC614	IBE724		FL360	
08:29	COA071	RAM2654		FL380	
08:32	AMB522	BER716	AZA418	FL380	Not real conflict but it will request surveillance

Scenario #2: The scenario starts with 15 minutes of medium air-traffic load. Then, the air-traffic condition becomes easy (15 minutes) and it ends with a hard condition in the last 15 minutes (Figure 14).

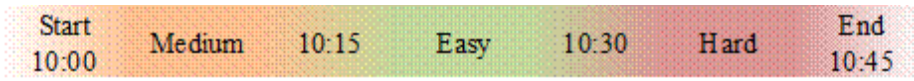

Figure 14. Time schedule of the ATM Scenario #2. The air-traffic load transition is medium (15 minutes), easy (15 minutes) and hard (15 minutes).

Figure 15 shows the number of aircrafts within the Scenario #2. In the first 15 minutes, medium conditions, there are about 12 airplanes, in the easy conditions there are on average 6 airplanes and, in the hard condition, the number of airplanes reaches upto 25 units (20 airplanes on average).

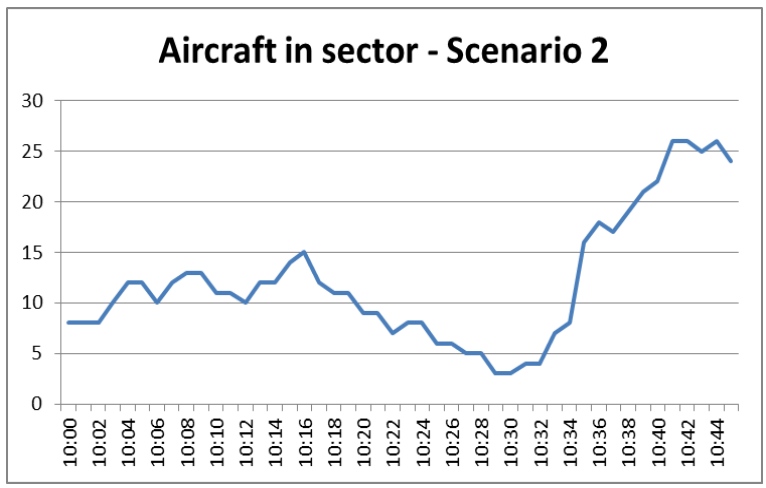

Figure 15. Number of airplanes across the Scenario #2. In the first 15 minutes, medium conditions, there are about 12 airplanes, in the easy conditions there are on average 6 airplanes and, in the hard condition, the number of airplanes reaches upto 25 units (20 airplanes on average)**.**

Table 3 presents the potential flight conflicts during the Scenario #2. Most part of the conflicts (14 potential conflicts) has been inserted within the hard phase (10:30 – 10:45). In fact, only 3 potential conflicts have been inserted in the medium condition (10:00 – 10:15) and 2 in the easy condition (10:15 – 10:30).

Table 3: Potential flight conflicts within the Scenario #2.

Time	Id1	Id2	Id3	Flight Level	Comment
10:08	THY472	DAN562		FL380	
10:08	THY472	FVID		FL340	Gaining on it
10:09	DLH586	AF4873		FL390	
10:13	VRG715	IBE354	AER284	FL360	
10:18	CSA912	AF4873		FL390	
10:20	AZA589	JKK5432		FL360	
10:35	KLM4321	VAL829		FL390	
10:37	CYP437	TAP264		FL390	
10:37	BAW525	ACA443	AIC532	FL390	
10:37	SWR29G	AFR065X		FL350	
10:38	KLM729	SAA324		FL380	
10:38	AFR527				Make it descending
10:39	LOT313	DLH663		FL360	
10:41	CTM1317	FDVID	IBE692	FL340	
10:41	BAW846				Make it climbing
10:42	CSA219	MSR521		FL360	
10:42	AUA205	NJE370		FL410	
10:45	IBE467	ACA443		FL390	
10:47	AIC532	MAH618		FL390	

Scenario #3: Since the scenario starts with 15 minutes of 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, the air-traffic load becomes medium (15 minutes) and finally easy for the last 15 minutes (Figure 16).

Figure 16. Time schedule of the ATM Scenario #3. The air-traffic load transition is hard (15 minutes), medium (15 minutes) and easy (15 minutes).

The number of aircrafts within the Scenario #3 is, on average, 21 at the beginning (hard condition) and then it starts to decrease to 15 in the medium condition and to 10 in the last (easy) condition (Figure 17).

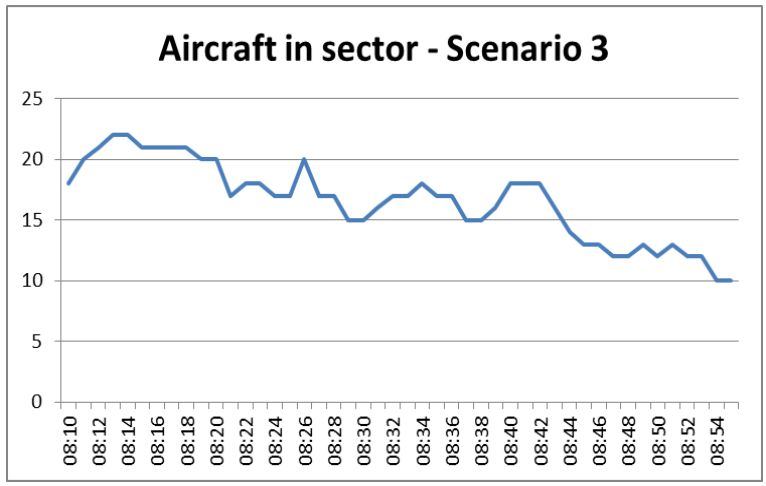

Figure 17. The number of aircrafts within the Scenario #3 is, on average, 21 at the beginning (hard condition) and then it starts to decrease to 15 in the medium condition and to 10 in the last (easy) condition.

In Table 4, the potential flight conflicts during the Scenario #3 are reported. Most part of the conflicts (10 potential conflicts) has been inserted within the hard phase (8:10 – 8:25), while the remaining have been inserted within the medium condition (3 potential conflicts) and easy condition (1 potential conflicts).

Table 4: Potential flight conflicts within the Scenario #3.

Time	Id1	Id2	Id3	Flight Level	Comment
08:13	AUA				Make it climbing
08:15	SAS618	AFR825		FL360	Not real conflict but it will request surveillance
08:15	DAL241	KLM587		FL370	
08:17	CSA148	BAW363		FL390	
08:18	FIN467	AZA339		FL350	
08:20	IBE346	LOT219		FL380	
08:24	ADR258	DAL241		FL370	
08:24	TAP674	RCHDQ1		FL390	
08:25	CCA711	BER330	CTM228	FL340	Speed is challinging
08:25	MAH1128				Make it descending
08:30	DAH248	AFR028		FL360	
08:38	ICE873	FIN557		FL390	
08:39	IBE692	IBE612		FL340	
08:42	AEE734	TAP369		FL380	

The considered air-space has been featured by an *en-route* sector where the air-traffic has been handled and transferred according to the standard protocols currently used in the Area Control Center (ACC) of Toulouse. Figure 18 shows the sector features. The RU (dark gray) is the considered air-sector with routes and beacons (light gray). The radio frequency is FM 125,305 and the controlled altitude range starts from FL335 to above. The RS sector is under RU sector, and air-traffic between those sectors has to respect specific procedures, in terms of altitude levels, when flying through the RS sector depending on departure or arrival flights. Specifically, departing flights through the RS sector have to climb to FL330 before crossing RU, while arrival flight through the RS sectors have to descent to FL340 before entering the RU air-space. Also, *feeder sectors* have been simulated with the aim to handover traffics within the considered sectors, and they are labelled in Figure 18 as *OS, OM, N3, G2, M2, I3, I2, S3* and *S2*. The red and green arrows show the traffic flow direction, respectively, incoming and outcoming traffic.

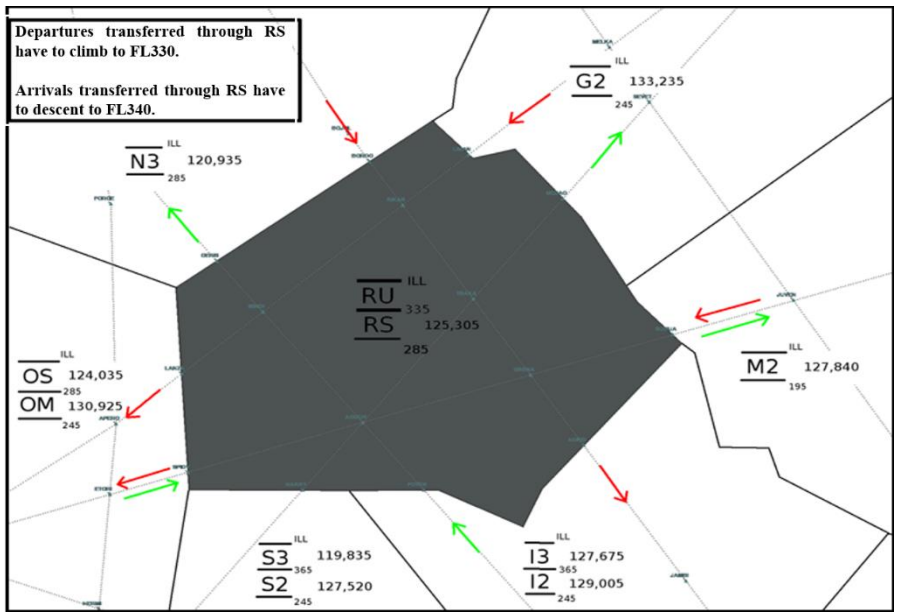

Figure 18. The RU (dark gray) is the considered air-sector with routes and beacons (light gray). The controlled altitude range starts from FL335 to above. The RS sector is under RU sector and air-traffic between those sectors is described in the box in the top left corner. The feeder sectors are OS, OM, N3, G2, M2, I3, I2, S3 and S2. The red and green arrows show the traffic flow direction, respectively, incoming and outcoming traffic.

The operational ATC position (left side in Figure 19) consisted in two screens. One of 30" to display radar image, and a second of 21" (WACOM) to interact with the radar image (zoom, move, clearances and information). The simulated ATC position was very similar to the French *Control Working Position* (CWP) in real ATM environments, except for the interface. The WACOM interface was a prototype where the ATCO could interact with the RADAR by means of a touch-screen and capacitive-pen. Before the experiments, the ATCOs have been trained to use it correctly in order to avoid learning effects during the recordings of brain activity.

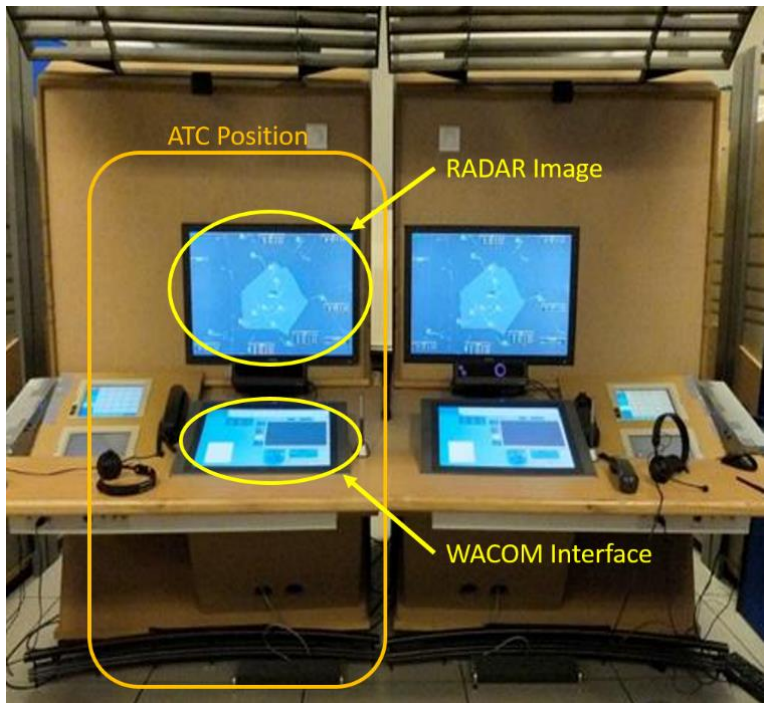

Figure 19. Operational position of the *en-route* ATCOs. It consisted in two screens. One of 30" (RADAR Image) to display radar image, and a second of 21" (WACOM Interface) to interact with the radar image (zoom, move, clearances and information).

Screenshots of the radar screen and of the prototyping interface are reported in Figure 20 and Figure 21. In the radar picture are reported the air-sector (light gray), routes, waypoints and flights displayed according to their status (the white airplanes are the flights assumed by the ATCO). Information of neighbour flights are listed under the relative sector.

To interact with the airplanes, the ATCO has to click on the *indirection pointing area* (gray area on the right side) of the WACOM interface and then a pie menu will pop up (Figure 20). From such menu, the controller is able to give commands and to look for specific airplane information.

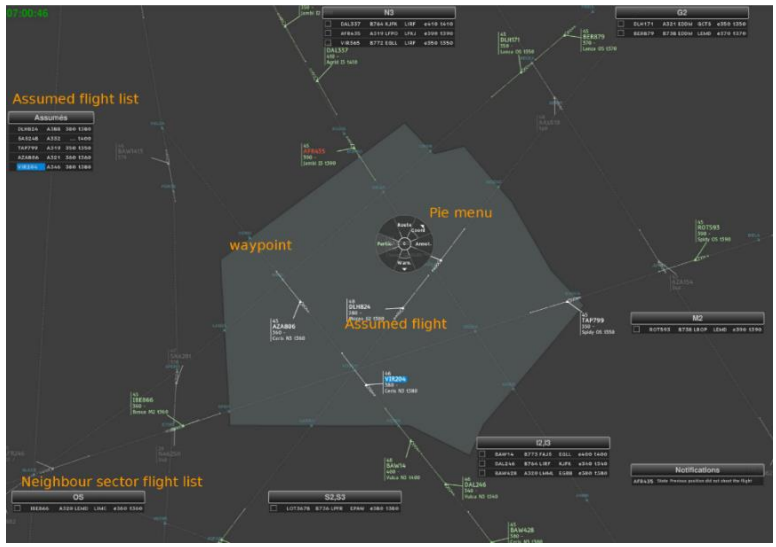

Figure 20. The picture reports the air-sector (light gray), routes, waypoints and flights displayed according to their status (the white airplanes are the flights assumed by the ATCO). Information of neighbour flights are listed under the relative sector.

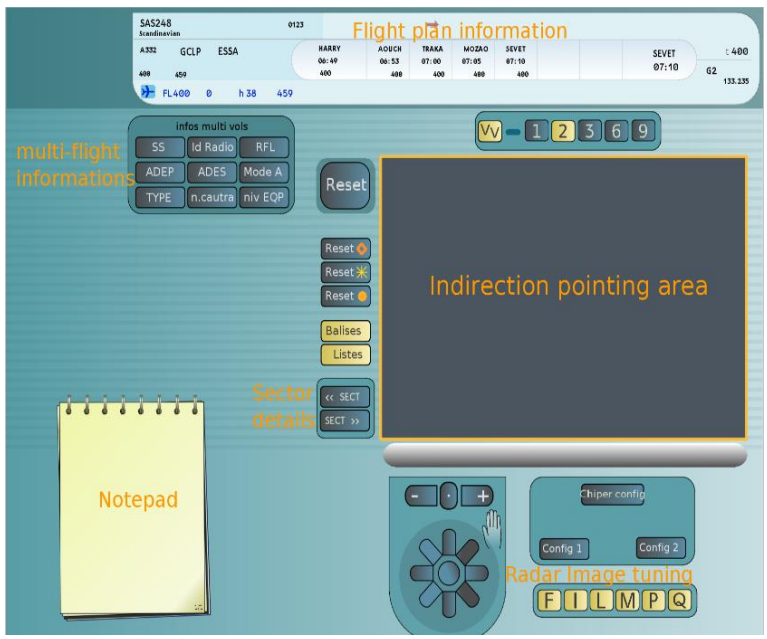

Figure 21. To interact with the airplanes, the ATCO used the prototypal WACOM interface. By cliccling on the *indirection pointing area* (gray area on the right side), a pie menu will pop up. From such menu, the Controller is able to give commands and to look for specific airplane information.

To make the ATC simulation realistic, two Pseudo Pilots have been involved in the experiments with the aim to have real communications with the ATCOs. Headsets (headphone and microphone) have been used by the ATCOs and Pseudo Pilots. The interface of the pilots is reported in Figure 22 and it provided information about all the airplanes on the RADAR and those assumed by the ATCOs. In particular, the main view shows the selected aircraft navigation display. On the left side, the managed aircraft are listed, while the second list shows the flights that will be managed next, with the currently selected aircraft highlighted. On the bottom of the screen, there is the schedule of requested actions for the different flights. Furthermore, the Pseudo Pilots could modify the airplane upset depending on the ATCOs instructions or intentionally to produce particular events (e.g. conflicts).

Figure 22. The interface of the Pseudo Pilots provided information about all the airplanes on the RADAR and those assumed by the ATCOs. Furthermore, the Pseudo Pilots could modify the airplane upset depending on the ATCOs instructions or intentionally to produce particular events (e.g. conflicts).

During the ATM simulation, the system recorded different parameters in order to define an index of the ATCO performance. In particular, the ATM simulator provided information on the reaction times and number of airplanes the ATCO assumed in each specific task condition (Easy, Medium and Hard). The parameters considered as performance metric are the following:

- *Time Shoot*: the interaction time needed for delivering an aircraft to the next sector;
- *Time Assume*: the interaction time for accepting an aircraft in the sector;
- *Time Route*: the interaction time to display the graphic route of an aircraft;
- *Time Cancel*: the interaction time between triggering write recognition box and pressing of the cancel button;
- *Time Annul*: the interaction time between the opening of a pie menu or write recognition box and clicking on the radar image background;
- *Time Heading*: the interaction time between triggering the menu and the validation of the inserted value in the write recognition box;
- *Time Turn*: the interaction time between triggering the menu and the validation of the write recognition box;
- *Time Flight Level*: the interaction time between triggering the menu and the validation of the write recognition box;
- *Time Direct*: the interaction time between triggering the menu and selecting the waypoint in the flight plan list.

Since each ATCO might adopt different strategies to manage the air-traffic, their *Reaction Times* (RT) have been weighted by taking into account the number of airplanes (nb) 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 RTs described previously. An example of log file provided by the ATM simulator is reported in Figure 23.

RANK	exofile	difficulty	meantimeshoot	nbshoot	meantimeassume	nbassume	meantimeroute	nbroute	meantimecancel	nbcancel
STUDENT	EXO3	EASY	4.684	23	2.222	7	1.251	16	2.212	2
STUDENT	EXO3	EASY	3.345	18	1.904	8	0.991	21	3.382	9
STUDENT	EXO3	EASY	4.796	21	1.484	9	0.901	16	2.936	8
STUDENT	EXO3	EASY	3.672	18	4.061	9	0.957	43	2.352	12
STUDENT	EXO3	EASY	3.099	21	1.589	9	1.277	11	3.113	6
STUDENT	EXO3	EASY	4.579	21	4.942	9	1.051	10	3.121	7
STUDENT	EXO3	EASY	6.162	21	3.385	9	0.779	17	2.296	14
STUDENT	EXO3	EASY	3.65	17	1.609	8	0.686	17	1.193	4
STUDENT	EXO3	EASY	3.431	17	2.34	8	0.689	15	4.104	6
STUDENT	EXO3	EASY	3.598	17	2.615	9	1.163	38	1.691	12
STUDENT	EXO3	EASY	4.274	24	2.529	8	0.669	13	1.356	1
STUDENT	EXO3	EASY	1.972	15	1.58	8	17.26	7	1.748	9
STUDENT	EXO3	EASY	4.218	18	1.686	9	0.968	27	2.809	10
STUDENT	EXO3	EASY	4.457	21	1.669	8	0.853	28	3.056	19
STUDENT	EXO3	EASY	4.278	20	3.943	8	1.022	12	1.909	15
STUDENT	EXO3	EASY	4.493	26	2.017	9	35.017	27	2.366	3
STUDENT	EXO3	EASY	2.101	16	2.583	8	1.406	14	2.658	11
STUDENT	EXO3	EASY	3.271	25	1.549	14	0.768	12	1.431	3
STUDENT	EXO3	EASY	4.681	20	3.112	8	1.031	29	1.371	5
STUDENT	EXO3	EASY	5.321	21	2.938	8	1.048	20	3.71	3
STUDENT	EXO3	EASY	2.96	17	1.973	8	0.882	38	2.284	3
EXPERT	EXO3	EASY	3.053	19	1.483	9	30.539	13	4.111	7
EXPERT	EXO3	EASY	4.036	20	2.31	8	0.905	14	3.518	3
EXPERT	EXO3	EASY	1.657	17	1.765	8	115.356	12	2.324	4
EXPERT	EXO3	EASY	3.04	17	1.576	9	1.808	10	2.851	10
EXPERT	EXO3	EASY	3.012	16	1.267	8	0.959	23	0	0
EXPERT	EXO3	EASY	2.723	18	1.085	12	0.831	26	1.108	5
EXPERT	EXO3	EASY	2.768	14	1.086	8	0.978	13	1.854	1
EXPERT	EXO3	EASY	3.348	19	1.539	9	0.649	12	6.329	2
EXPERT	EXO3	EASY	1.267	19	2.326	9	1.086	14	3.242	7
EXPERT	EXO3	EASY	4.303	18	1.041	9	0.859	26	3.501	12
EXPERT	EXO3	EASY	2.144	19	2.129	8	0.864	13	1.786	2
EXPERT	EXO3	EASY	4.084	21	1.44	10	0.857	15	5.294	4
EXPERT	EXO3	MEDIUM	2.742	17	1.467	16	2.898	2	2.582	5
EXPERT	EXO3	MEDIUM	9.002	21	1.579	17	0	0	1.688	6
EXPERT	EXO3	MEDIUM	1.914	21	1.061	18	0	0	3.14	3

Figure 23. Example of log file provided by the ATM simulator. For each session and task difficulty, the reaction times and number of assumed airplanes have been collected in order to define an index by which quantifying the ATCO performance.

2.5. Subjective Questionnaires

2.5.1. NASA - Task Load Index (NASA – TLX)

The NASA- *Task Load Index* (NASA-TLX) is a widely-used (Colligan et al., 2015), subjective, multidimensional assessment tool that rates perceived workload in order to assess a task, system, or team's effectiveness or other aspects of performance. It was developed in the 80’ by the Human Performance Group at NASA's Ames Research Centre over a three-year development cycle that included more than 40 laboratory simulations (Hart and Staveland, 1988). It has been used in a variety of domains, including aviation, healthcare and other complex socio-technical domains.

The total workload score is calculated by the combination of six factors:

- *Mental Demand*: How much mental and perceptual activity was required (e.g. thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?
- *Physical Demand*: How much physical activity was required (e.g. pushing, pulling, turning, controlling, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?
- *Temporal Demand*: How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?
- *Performance*: How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?
- *Effort*: How hard did you have to work (mentally and physically) to accomplish your level of performance?
- *Frustration*: How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?

At the end of the task or during the execution of the task, the NASA-TLX is presented and the subjects have to rate each factor (from 0 to 100).
Later, the factors are compared in pairs to each others, and the number of times each factor is chosen is the weight by which the previous rates will be multiplied. The linear combination of the weighted rates is then divided by 15 (number of total comparisons) and the NASA-TLX total score (value from 0 to 100) will be provided. The NASA-TLX can be administered using paper sheets or digital interfaces (NASA, 2013). An example of interface is reported in Figure 24. While there are multiple ways to administer the NASA-TLX, some may change the results of the test. One study showed that the paper sheets version led to less cognitive workload than the digital one (Noyes and Bruneau, 2007). To overcome the delay in administrating the test, the digital version can be used to submit the test immediately after the task is completed in order to capture accurate workload perception.

Figure 24. Example of NASA - TLX interface. The total workload index is a weighted combination of six factors: mental demand, physical demand, temporal demand, performance, effort and frustration.

2.5.2. Instantaneous Self Assessment (ISA)

Instantaneous Self Assessment (ISA) is a technique that has been developed as a measure of workload to provide immediate subjective ratings of work demands during the performance of primary work tasks (Brennen, 1992; Jordan, 1992). The ISA was originally developed by the *National Air Traffic Service* (NATS) in UK for use in the assessment of air traffic controller's mental workload during the design of future ATM systems (Kirwan et al 1997). It has then been used by other UK organisations, including the *Air Traffic Control Evaluation Unit* (ATCEU) at Bournemouth airport, the *Defence Evaluation and Research Agency* (DERA) in Portsdown and the *Royal Navy's research organisation* (now part of *QinetiQ*), and also European organizations, the *Eurocontrol Experimental Centre* (EEC), the *Centre d'études de la navigation aérienne* (CENA) and by the *Rijksluchtvaartdienst* (RLD) of Netherland (now *Civil Aviation Authority* –

CAA). ISA involves participants self-rating their workload during a task (normally every two minutes) on a scale of 1 (low) to 5 (high). Kirwan et al (1997) used the following (Figure 25) ISA definitions to assess air traffic controllers (ATC) workload.

Level	Workload Heading	Spare Capacity	Description
5	Excessive	None	Behind on tasks; losing track of the full picture
4	High	Very Little	Non-essential tasks suffering. Could not work at this level very long.
3	Comfortable Busy Pace	Some	All tasks well in hand. Busy but stimulating pace. Could keep going continuously at this level.
2	Relaxed	Ample	More than enough time for all tasks. Active on ATC task less than 50% of the time.
1	Under-Utilised	Very Much	Nothing to do. Rather boring.

Figure 25. Example of ISA scale. When the ISA scale is presented to the user, he\she has to select a value from 1 (very low) to 5 (very high) in order to score the perception of the workload on that moment.

Typically, the ISA scale is presented to the participants in the form of a colour-coded keypad. The keypad flashes and sounds when the workload rating is required, and the participant simply pushes the button that corresponds to their perception of workload (Figure 26). Alternatively, the workload ratings can be requested and acquired verbally. The appeal of the ISA technique lies in its low resource usage and its low intrusiveness. In fact, the output of the ISA (values from 1 to 5) allows a workload profile for the task under analysis to be constructed and it is very quick in its application as data collection occurs during the trial. Whilst the technique might be a bit obtrusive to the primary task, it is probably the least intrusive of the on-line workload assessment techniques.

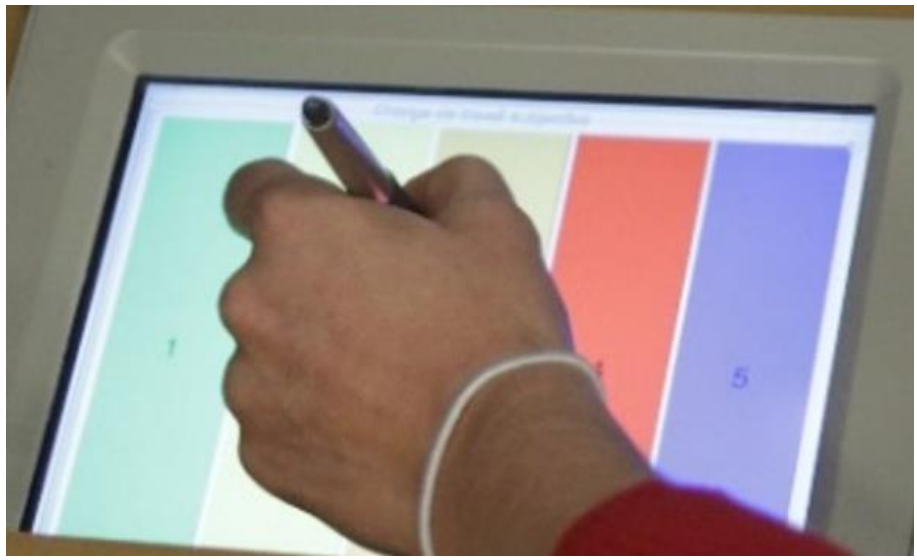

Figure 26. The ISA test is typically presented to the participants in the form of a colour-coded keypad. The keypad flashes and sounds when the workload rating is required, and the participant simply pushes the button that corresponds to their perception of workload.

3. The Workload

The Workload is a complex construct that is assumed to be reflective of an individual's level of attentional engagement and mental effort (Wickens, 1984). Measurement of mental workload essentially represents the quantification of mental activity resulting from performing a task or set of tasks. Several empirical investigations have indicated that performance declines at either extremes of the workload demand continuum , that is when the event rate is excessively high as well as when the event rate is extremely low (Borghini et al., 2012a). Consequently, it is important to preserve a good level of the user's mental workload, avoiding mental under- or over-load state, with the aim to maintain an optimal level of performance and reducing the probability of errors commission (Borghini et al., 2014; Parasuraman and Hancock, 2001). For these reasons, the mental workload is an important and central construct in ergonomics and human factor researches.

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

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

Such definitions show that the mental workload may not be a unitary concept because it is the result of different interacting aspects. In fact, several mental processes, such as alertness, vigilance, mental effort, attention, mental fatigue, drowsiness, and so on, can be involved in the meanwhile of a task execution and they could be affected 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 activation (Baldwin, 2003; Wickens, 1984).

In the last 20 years, it has been widely documented that the 70% of civil aviation accidents were linked with human factors (Bellenkes, 2007). Recently, the Aviation Safety Network reported 37 accidents with 564 casualties, and 8 accidents with 295 casualties, respectively, during the 2013 and the first months of the 2014. Moreover, the air traffic is growing exponentially and it has been predicted to double in the 2020 (Flight Safety Foundation). This factor would increase the work difficulty of Pilots and Air Traffic Controllers (ATCOs). In particular, the latter ones have to perform a variety of tasks, including monitoring air traffic, anticipating loss of separation between aircrafts, and intervening to resolve conflicts and minimize disruption to air traffic flow (for an extensive compilation of the tasks and goals of *en-route* control, see Rodgers and Drechsler, 1993). In this domain, the Pilots and ATCOs' behaviour could already be measured through several human factor tools, such as the explicit measurement of errors committed during the execution of the task, or by using questionnaires related to the subjective workload perception, such as the *Instantaneous Self Assessment* (ISA, Kirwan et al., 2001), *NASA – Task Load IndeX* (NASA-TLX, Hart and Staveland, 1988) or the *Subjective Workload Assessment Technique* (SWAT, Reid and Nygren, 1988) questionnaires. Because of their inherently subjective nature, none of such questionnaires allows to have an objective and reliable measure of the actual cognitive demand in a real environment. All the described questionnaires have pros and cons, but there is not a generally accepted standard (Rubio et al., 2004). Therefore, the need of an objective measure became fundamental for reliable workload evaluations.

Many neurophysiological measures have been used for the mental workload evaluation, 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) (Borghini et al., 2012a, b; Ramnani and Owen, 2004; Wood and Grafman, 2003). The size, weight, and power constraints, outlined in the Introduction, limit the use of some techniques in real operational environments. In fact, fMRI (Cabeza and Nyberg, 2000) and MEG techniques require room-size equipment that are not portable, as appropriate structures, electro-magnetic interference-shields. EOG, ECG and GSR activity measurements highlighted correlations with some mental states, but they have been demonstrated to be useful only in combination with measures directly linked to the brain activity (Borghini et al., 2012a; Ryu and Myung, 2005). Consequently, the EEG and fNIR are the most likely candidates that can be straightforwardly employed in operational environments.

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

oxygenation was significantly sensitive to task load variations. Regarding the EEG measurements, most part of the studies shown that the brain electrical activities mainly considered for the mental workload evaluation are the theta and alpha EEG rhythms, in particular on the *Pre-Frontal Cortex* (PFC) and the *Posterior Parietal Cortex* (PPC) regions. The theta (4 – 8 Hz) rhythm, especially over the PFC, presents a positive correlation, i.e. increases when the mental workload increases (Gevins and Smith, 2003; Smit et al., 2005), while the alpha (8 – 12 Hz) rhythm, especially over the PPC, presents an inverse correlation, i.e. decreases, (Brookings et al., 1996; Gevins et al., 1997; Jaušovec and Jaušovec, 2012; Klimesch et al., 1997; Venables and Fairclough, 2009). Only few studies have reported results about other EEG bands, i.e. the delta, beta and gamma EEG rhythms. Onton et al. (2005) reported that the frontal midline theta rhythm increases with memory load, confirming previous results about the correlation between the frontal theta EEG activity and mental effort (Gevins et al., 1997; Smit et al., 2005). Mental workload is also known to suppress EEG alpha rhythm and to increase theta rhythm during activity of encoding and retrieval of information (Klimesch et al., 1997; Klimesch, 1999). Several researches in the ATM domain treated the neurophysiological measurements of ATCOs' mental workload in realistic settings with the aim of developing HMI systems. These studies were based on both EEG and fNIRs techniques. It has been discussed how each technique is able to provide a reliable estimation of mental workload. The propensity in using EEG or fNIRs techniques in such kind of applications has not been clarified yet. In fact, there are several factors to take into account about real operational scenarios. For example, both EEG and *Fast Optical Signal* (FOS)-based fNIR have similar bandwidth and sample rate requirements, as the FOS appears to directly reflect aggregated neural spike activity in real-time and can be used as a high-bandwidth signal akin to EEG (Medvedev et al., 2008). However, EEG and fNIRs systems have different physical interfaces, sizes, weights and power budgets, thus different wearability and usability in real operational contexts. Specifically, the physical interface merits scrutiny, as it is non-trivial to maintain a good contact between the sensors (i.e. electrodes or optodes) and the brain scalp in freely-moving tasks. The use of fNIRs eliminates motion artifacts and the need of both scalp abrasion and conductive gel. In addition, there is not the necessity to wear a cap but

only a headband. Furthermore, unlike EEG, fNIRs recordings are not affected by electroculographic or facial electromyographic activity and environmental electrical noise, which are undoubtedly ubiquitous in human-computer interactions. Thus, this technology could appear more suitable in realistic environments (Goldberg et al., 2011). In two recent studies (Ayaz et al. 2012, 2013), the use of fNIRs technique to assess the mental workload has been investigated in the ATM environment. Based on the hypothesis (Durantin et al., 2014; Izzetoglu et al., 2004; Owen et al., 2005) that the hemodynamic response over the dorsolateral and ventrolateral prefrontal cortex was positively responsive to mental workload, Ayaz et al. (2012) demonstrated how the BOLD signal was a reliable workload indicator. In fact, by using this index, it was possible to discriminate significantly the different workload requested by two different communication types (Data and Voice) on 24 professional ATCOs. Furthermore, Ayaz et al. (2012, 2013) and Bunce et al. (2011) showed that fNIR could be successfully used in realistic environments to assess: i) mental workload levels of ATCOs performing standardized n-back task and operative ATM tasks, and ii) expertise development when learning a complex cognitive and visuo-motor tasks on Unmanned Aerial Vehicles (UAV) Pilots. However, in a recent study, Harrison et al. (2014), reported how the BOLD signal showed a lower resolution than the subjective measures (ISA, Kirwan et al., 2001) to evaluate the mental workload of ATCOs involved in the experiment. In particular, while the task was becoming more difficult, the subjective measure was still increasing, and the BOLD signal (neurophysiological index) reached its maximum, lingering on this value. Furthermore, the BOLD signal, used as workload index, was shown to be not reliable over time since the workload measurements performed in different days were significantly different and in discordance with the subjective measures. Indeed, also this technique has its weaknesses. Since the presence of hair may impact on both photon absorption (Murkin and Arango, 2009) and the coupling of the probes with the underlying scalp, the fNIRs technique is very reliable only on those un-hairy brain areas, like the PFC. As quoted above, also the parietal brain sites play a key role in the mental workload evaluation, and Derosière et al. (2013) pointed out how some fNIRs-measured hemodynamic variables were relatively insensitive to certain changes in mental workload and attentional states. Due to its higher

temporal resolution and usability, in comparison with the fNIRs technique, the EEG technique overcomes such kind of issues.

In addition, there are several studies in ATM domain that highlighted the high reliability of EEG-based mental workload indexes (Brookings et al., 1996). The results showed that the effects of the task demand were evident on the EEG rhythms variations. EEG power spectra increased in the theta band, while significantly decreased in the alpha band as the task difficulty increased, over central, parietal, frontal and temporal brain sites. In a recent study, Shou et al. (2012) evaluated the mental workload during an ATC experiment using a new *time-frequency Independent Component Analysis* (tfICA) method for the analysis of the EEG signal. They found that "the frontal theta EEG activity was a sensitive and reliable metric to assess workload and time-on-task effect during an ATC task at the resolution of minute(s)". In other recent studies involving professional and trainees ATCOs (Aricò et al., 2015a,b, 2014c; Di Flumeri et al., 2015; Borghini et al., 2014), it has been demonstrated how it was possible to compute, by machine-learning techniques and specific brain features, an EEG-based Workload Index able to significantly discriminate the workload demands of the ATM task. In those studies, the ATM task was 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 index showed to be directly and significantly correlated with the actual mental demand experienced by the ATCOs during the entire task. The same EEG-based Workload Index was also used to evaluate and compare the impact of different avionic technologies on the mental workload of professional helicopter pilots (Borghini et al., 2015a) and to. Furthermore, the machine-learning techniques have been successfully used in other real environments for the evaluation of mental states (Müller et al., 2008) and mental workload (Berka et al., 2004, 2007). Even if the main limitation of the EEG is its wearability, technology improvements (Liao et al., 2012) have being developed and tested in terms of dry electrodes (no gel and problems of lowering the impedances), comfort, ergonomic and wireless communications (no cables between the EEG cap and the recording system. In conclusion, the EEG technique seems to be the appropriate solution to evaluate the mental workload in realistic and operational settings, and to be

integrated in passive BCI systems. Such systems will support the operator during his/her working activity in order to improve the works wellness and, most of all, the safety standards of the whole environment.

Workload and Autonomic Signals: Heart and Eyes Activities

Heart Rate (HR) has been demonstrated to be considered a reliable measure of mental workload by many studies. Generally, there is a direct correlation between workload and heart rate, so when the first increases the second increases as well (Hankins and Wilson, 1998; Wilson et al., 1999; Costa, 1993; Jorna, 1992; Roscoe, 1993; Veltman and Gaillard, 1996; Wilson et al., 1994). Nevertheless, some authors are critical respect the use of HR for evaluating workload, because of the various psychological, environmental and emotional factors that can influence the response (Jorna, 1992; Roscoe, 1993; Lee and Park, 1990). For example, uncertainty and anxiety can significantly raise heart rate (Jorna, 1993). Gravitational forces may also affect this technique (Wilson, 1993). Furthermore, heart rate does not measure absolute levels of workload, but relative levels (Roscoe, 1992; Roscoe, 1993). This may be a benefit in real-world rather than simulated situations because there is less control over variables (Roscoe, 1993). In a study on multitasking performance, (Fairclough, 2005) explored the interaction between learning and task demand on psychophysiological reactivity. These authors used EEG activity, cardiac activity, respiration rate and to evaluate the impact of task demand and learning and found that the sustained response to task demand was characterized by a reduction of parasympathetic inhibition (reduced vagal tone and increased heart rate), reduced eye blink duration. This sustained response may represent the state of focused concentration that was necessary to perform a multitasking activity.

In a flight experimental scenario, (Wilson, 1993) evaluated ten pilots who were required to fly a 90-minute to test the reliability of psychophysiological measures of workload. Cardiac, electrodermal and electrical brain activity measures were highly correlated and exhibited changes in response to the demands of the flights.

Using the eye activity measurement, he found that blink rates decreased during the more highly visually demanding segments of the flights.

Another important technique to assess the mental workload is the analysis of the *Eye Blinks Rate* (EBR). In effect, eye blinks and blink duration decrease when visual workload intensifies (Hankins and Wilson, 1998; Brookings et al., 1996; De Waard, 1996; East, 2000; Van Orden, 1999). It has been shown that, although ocular measures are sensitive to mental demands, they are affected by other factors, such as fatigue, body or head moves and environmental changes. A change in light or air quality may influence eye blinks rate (De Waard, 1996). Some studies (Sirevaag et al., 1993) presented contradictory results caused by different methods used for the experiments. Furthermore, as reported by (Cain, 2007), blinks measures can be context dependent. Blinks rate decrease with increased workload with the processing of visual stimuli, while increases with the augment of workload in memory tasks (Wilson, 1993). Effectiveness, according to Castor (Castor, 2003), the connection between blinks rate and workload seems tenuous. Blink closure duration decreases with increasing of visual stimuli workload or acquisition of data from a wide field of view, while blink latency increases with memory and response demands (Castor, 2003). Stern and Skelly (1984) found that blinking is "inhibited during the acquisition of information, whether such information is presented visually or auditory. Once a decision is made whether it requires action or requires the inhibition of action, a blink is likely to occur. The non-inhibition of blinking during that time periods is associated with a higher probability of occurrence of missed signals and erroneous responses". In the field of flight and driving research, some authors tried to discriminate between visual and mental workload to make the eye blinks (Van Orden, 1999) and eye blinks duration (Sirevaag, 1993) as a good index of estimation in some aspects of mental workload (Miller, 2001). Some authors found eye blinks as a good measure for visual workload (East, 2000; Brookings et al., 1996; Hankins and Wilson, 1998). Brookings et al. (1996) in a study of Air Traffic Control simulation (description below in Air Traffic Control section), found that blinks rate decreased significantly when the task became more difficult.

3.1. Automatic Stop StepWise Linear Discriminant Analysis (asSWLDA)

Nowadays, the effects of day-to-day fluctuations in the operator's physiology have not been thoroughly assessed while operators are engaged in complex tasks. Different studies showed that the performance of classifiers in evaluating the different mental workload levels of the user dramatically decrease over days (Burgess and Gruzelier, 1993; Christensen et al., 2012; McEvoy et al., 2000; Pollock et al., 1991; Salinsky et al., 1991). Recent attempts have been made to test the stability of different classifiers for assessing the mental workload of the user across different days. It has been found that the performance of three different classifiers were significantly negatively impacted by the features time-stability (Christensen et al., 2012). In this regards, authors considered all the frequency bands (0.1 – 100Hz) in the calibration stage of the classifier, and probably the classifiers could not be able to select the features strictly related to the mental workload because of the big amount of information. Reliability of the system is of great importance for a practical usability of such approach in real working contexts. In fact, the necessity to re-calibrate the system every day make such kind of approach unusable in operational environments. In this study, I propose a new classification approach able to automatically optimize the features selection process, in order to make the classification performances of the considered mental state (e.g. mental workload) stable over time. In particular, I propose a new implementation of the *Stepwise Linear Discriminant Analysis* (SWLDA), the *automatic-stop SWLDA* (asSWLDA). The SWLDA is an extension of the *Linear Discriminant Analysis* (LDA, Fisher, 1936) and it is one of the best outperforming linear classifiers (Rabbi et al., 2009; Berka et al., 2007; Craven et al., 2006; Krusienski et al., 2006) which performs a reduction of the features' space by selecting the most significant ones. In fact, with respect to other linear methods, it has the advantage of having automatic features extraction in order to statistically remove the insignificant terms from the model. Particularly, the SWLDA regression consists in the combination of *forward* and *backward* stepwise analyses, where the input

features are weighted by using ordinary least-squares regression to predict the target class label. The method starts by creating an initial model of the discriminant function in which the most statistically significant feature is added to the model for predicting the target labels ($pval_{ij} < \alpha_{ENTER}$), where $pval_{ij}$ represents the p-value of the *i-th* feature at the *j-th* iteration (in this case the first iteration). Then, at each new iteration, a new term is added to the model (if $pval_{ij} < \alpha_{ENTER}$). If there are not more features that satisfy this condition, a *backward* elimination analysis is performed to remove the least statistically significant feature (if $pval_{ij} > \alpha_{REMOVE}$). The standard implementation of the SWLDA algorithm uses $\alpha_{ENTER} = 0.05$ and $\alpha_{REMOVE} = 0.1$, and no constrains on the $Iteraction_{MAX}$ (predefined number of iterations) parameter are imposed, that is, the feature selection keeps going unless there are no more features satisfying the entry (α_{ENTER}) and the removal (α_{REMOVE}) conditions (Draper and Smith, 1998). Normally, it is possible to optimize a SWLDA regression by tuning all or some of the three parameters available in the algorithm (α_{ENTER}, α_{REMOVE} and $Iteraction_{MAX}$). There are not standard procedures to choose these parameters, and they could be manually gauge. Despite the strength of the method, it is easy to realize how it would be difficult to set properly the parameters in order to optimize the algorithm.
In fact, since the probability associated to each feature is strictly related to the actual iteration and to the other features, this probability changes iteration by iteration, and it would result very difficult to impose a condition by manually trimming α_{ENTER} and α_{REMOVE}. In addition, even if no constrains on the α_{ENTER} and the α_{REMOVE} parameters are imposed, the features are included in the model in order of significance (i.e. the first feature is the most significant one, and so on). As consequence, the value of the $Iteration_{MAX}$ parameter will affect the reliability of the classifier over time (optimum classifier, *underfitting* or *overfitting*). The expected result is that the more general the classification calibration would be, the higher the reliability of the algorithm over time will be (Vapnik, 2000). For example, if the value of the parameters are too selective, that is α_{ENTER} and/or α_{REMOVE}, and/or $Iteration_{MAX}$ values very low, the features added to the model will not be sufficient for predicting the target labels (*underfitting*, von Luxburg and Schoelkopf, 2008). On the contrary, if α_{ENTER} and/or α_{REMOVE}, and/or $Iteration_{MAX}$ values are too high, most of the features added to the final model will be related to spurious differences between the classes of the

training dataset, that are obviously not generalizable, so that, the reliability of the algorithm will decrease over time (*overfitting*, Vapnik, 2000). The optimum solution to these problems would be to find out a criteria able to automatically stop the algorithm when the best number of features, $\#Features_{OPTIMUM}$, are added to the model. The idea to fix such issues, for the proposed applications, was to define a criteria such that the number of the SWLDA interations (*IterationMAX*) is: $\#Features_{UNDERFITTING} < \#Features_{OPTIMUM} < \#Features_{OVERFITTING}$. In other words, I wanted to make the SWLDA algorithm able to automatically select the right number of features ($\#Features_{OPTIMUM}$) to avoid both the *underfitting* and *overfitting* issues. In order to achieve such goal, I analyzed the *p-convergence* function of the SWLDA classification model (*Conv(#iter)*). Since more the number of iterations increases (more features are added to the model), more the p-value of the model (*pModel*) decreases (tending to zero) with a decreasing exponential shape (convergence of the model), the idea of the best trade-off between the lowest number of features and the convergence of the model was to automatically-stop the SWLDA (asSWLDA) algorithm in correspondence of the minimum distance from the (0,0)-point, that is, when the *Conv(#iter)* function assumed the lowest distance from the origin. In Figure 27a is reported the trend of the *Conv(#iter)* function for a representative subject.

The *pModel* values were returned by the MATLAB function of the SWLDA, and they have been collected for all the SWLDA iterations (*pModel(#iter)*). Then, the *log10* of the *pModel* vector (*log10(pModel(#iter)*) has been applied to gather information about the *pModel* order. Figure 27b shows how the order of the model decreases with the number of interactions. The *pModel* parameter is available in the output of the standard SWLDA implementation and it provides information about the global significance of the model at the *j-th* iteration.

To estimate the p-convergence function (*Conv(#iter*)) and the information about the size of *pModel* order, the *log10* of the first-order differences between adjacent *pModel* elements has been calculated by the formula (3.1).

$$Conv(\#\text{iter}) = log_{10}(pModel(\#\text{iter} + 1)) - log_{10}(pModel(\#\text{iter})) \qquad (3.1)$$

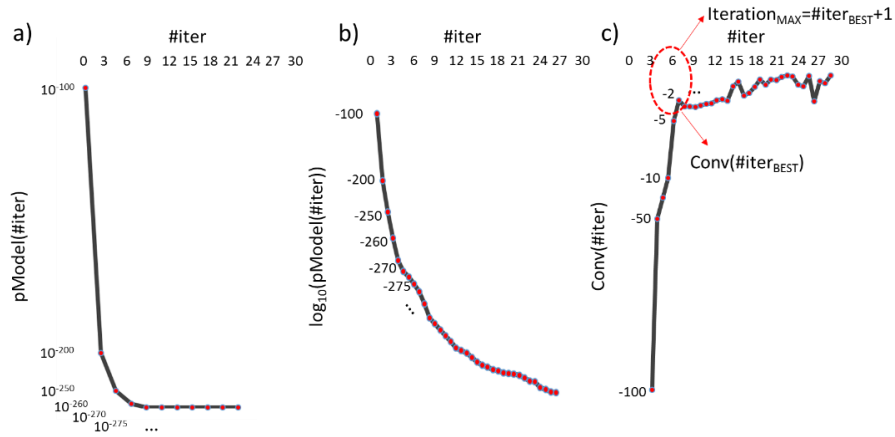

Figure 27. Representation of the a) *pModel vector*, the b) l*og10* of the *pModel* vector and the c) *Conv* function for each iteration, for a representative subject. In particular, in the figure (c) there are also showed i) the *Conv(#iter$_{BEST}$)*, in other words the lower distance of the *Conv(#iter)* function from the point (0,0) and ii) the correspondent *Iteration$_{MAX}$*, that is *#iter$_{BEST}$*.

Finally, the best *Iteration$_{MAX}$* value has been identified as the number of iterations at which the *Conv(#iter)* assumed the lowest distance from the origin (*Iteration$_{BEST}$*), plus one because of the first-order differences:

$$\|Conv(\#iter_{BEST})\| = min\|Conv(\#iter)\| \tag{3.2}$$

$$Iteration_{MAX} = \#iter_{BEST} + 1 \tag{3.3}$$

In fact, the best condition for the aim of the considered studies (training assessment and workload evaluation over time) is to have a low numbers of features and, at the same time, the convergence of the model:

$$log_{10}(pModel(\#\text{iter} + 1)) - log_{10}\big(pModel(\#\text{iter})\big) = 0 \tag{3.4}$$

Figure 27c shows the graphical meaning of the *Iteration$_{BEST}$* selection.
In order to assess the feasibility and reliability over-time of the asSWLDA algorithm in discriminating the workload and the differences with respect to the standard-SWLDA, a proper EEG dataset has been selected and the entire study is described in the following section.

3.2. Over Time Reliability and Workload Evaluation: SWLDA vs asSWLDA

In this study, the performances of the asSWLDA and of the standard SWLDA have been compared to assess the reliability of the two algorithms over time (one week) in terms of different workload levels discrimination. The methods have been tested on users while performing the MATB (see paragraph 2.3.1) under three different difficulty levels (Easy, Medium and Hard). The MATB has been chosen due to its flexibility in modulating workload demand in a very controlled and standardized way. Finally, the algorithms have been tested in a simulated operational environment.

3.2.1. Experimental Protocol

Subjects

Ten healthy volunteers (students of the *National University of Singapore* - NUS) have given their informed consent for taking part in the experiments. Each of them has been paid SG$200 for the entire experimental period. The study protocol has been approved by the local Ethics Committee. The selection of the subjects has been done accurately in order to ensure the same mental and physical status (homogeneity of the experimental sample). The subjects (25±3 years old) have been instructed to maintain a specific kind of lifestyle. In particular, they have been asked to avoid alcohol, caffeine and heavy meals before the experiments, and to avoid extreme physical activity over the entire experimental protocol (homogeneity of the "internal conditions" of the subjects). The Lab environment has been kept under control (lights intensity, room temperature, seat position) across the different days of the experiments (homogeneity of the "external conditions" during the experiment). In addition, in order to have low sources of variances, the experimental group has been composed only by males.

Experimental Design

The experimental protocol (Figure 28) consisted in six recording sessions, two sessions per day, one in the morning and one in the afternoon. The first four sessions have been performed in two consecutive days, named

hereafter as *Day 1* and *Day 2*. The last two sessions have been performed after a week from the last one (*Day 9*). In each session, the subjects performed the three MATB (see paragraph 2.3.1) conditions (Easy, Medium and Hard) twice in a randomized sequence, in order to avoid any habituation and expectation effects. Each condition lasted 2.5 minutes. In summary, the whole dataset has been composed of twelve triplets of conditions (four triplets of Easy, Medium and Hard conditions for each of the three experimental days, two in the morning and two in the afternoon). At the end of each task condition, subjects have been asked to fill the NASA-TLX (see paragraph 2.4.1) with the aim to gather information about the perceived workload during the different difficulty conditions and experimental sessions.

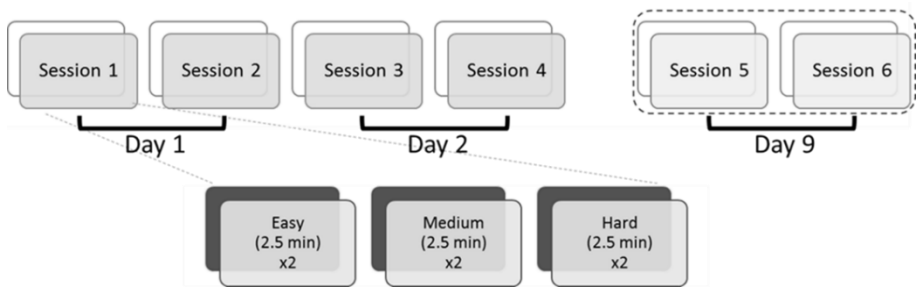

Figure 28. Experimental protocol scheme: each subject performed 6 recording sessions in three days, two sessions per day. The first four sessions were performed within two consecutive days (*Day 1* and *Day 2*), and the others after a week from the last one (*Day 9*). In each session, the subjects performed the three MATB difficulty levels (Easy, Medium and Hard) twice, randomly proposed. The time duration of each run was 2.5 minutes.

EEG Analyses for Mental Workload Evaluation

Scalp EEG has been recorded by the Waveguard amplifier (ANT-neuro, Netherlands) with a sampling frequency of 256 (Hz) by 13 EEG electrodes (*FPz, F3, Fz, F4, AF3, AF4, P3, Pz, P4, POz, O1, Oz* and *O2*) referenced to both the earlobes and grounded to the AFz electrode. Simultaneously, the vertical EOG signal has been acquired by bipolar electrodes over the left eye, in order to collect the eyes blink of the subjects during the execution of the task. Figure 29 shows the steps of the EEG analysis to evaluate the mental workload. Firstly, the EEG signal has been band-pass filtered between 1 and 30 (Hz) with a fourth-order Butterworth filter. Then, the EOG signal has been used to remove eyes-blink contributions from each

epoch of the EEG signal, by using the Gratton and Coles (Gratton et al., 1983). Other specific algorithms, for different sources of artifacts, have been applied by using the EEGLAB toolbox (Delorme and Makeig, 2004). The segmentation of the EEG has been done in according with Elul (1969), that is, with the aim to have a stationary EEG signal within the considered epoch, on which it is then possible to apply the methods for the analysis. In particular, the EEG signal has then been segmented into epochs of 2 seconds, shifted of 0.125 seconds. These parameters have been chosen also in according with the task duration (2.5 minutes), since, in the classification methods is really important to have a high number of observations to calibrate the algorithm. In particular, one of the working hypothesis was that the number of observations should be greater than the number of variables. The *Power Spectral Density* (PSD) has been calculated for each EEG epoch using a *Hanning* window of 2 seconds length (that means 0.5 (Hz) of frequency resolution). Then, the EEG frequency bands have been defined for each subject by the estimation of the *Individual Alpha Frequency* (IAF) value (Klimesch, 1999). In order to have a better estimation of the alpha peak and, hence of the IAF, the subjects have been asked to keep the eyes closed for a minute. Then, this condition has been used for the IAF estimation. In this way, a spectral features matrix (EEG channels x Frequency bins) has been obtained in the frequency bands directly correlated to the mental workload. In particular, only the theta band, over the frontal brain sites, and the alpha band, over the parietal brain, sites have been considered for the analyses.

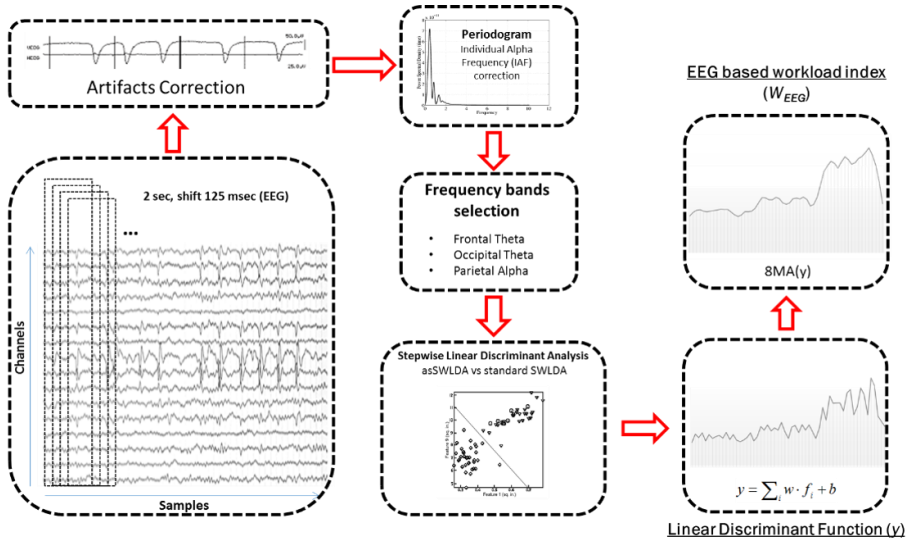

Figure 29. EEG based workload index (W_{EEG}). The figure explains the algorithm for the EEG-based workload index evaluation. The band-pass filtered (0.1-30 Hz) EEG signal was segmented into epochs of 2 seconds, shifted of 0.125 seconds. The EOG signal was used to remove the eyes-artifact contribution from the EEG signal. Other sources of artifacts have been deleted by using specific algorithms. Then, the power spectral density (PSD) was evaluated for each EEG channel, taking only the frequency bands involved in the mental workload estimation (frontal theta and parietal alpha bands). After the asSWLDA (standard SWLDA) was used to select the most relevant spectral features for the discrimination of the mental workload levels. A moving average of 8 seconds (8MA) was applied to the linear discriminant function (y) in order to reduce the variability of the index.

Mental Workload Index Based on EEG Activity

A three-classes SWLDA regression has been used to select within the training dataset the most relevant EEG spectral features to discriminate the three workload conditions (Easy, Medium and Hard). The *linear discriminant function* ($y_{test}(t)$) for each window has been computed, by using the coefficients (weights: wi_{train} and bias: b_{train}) returned by the SWLDA function (equation 3.5, where $fi_{test}(t)$ represents the PSD matrix of the testing dataset at the time sample *t*, and of the *i-th* feature). Finally, a moving average of 8 seconds (*8MA*) has been applied to the $y_{test}(t)$ function in order to smooth it out by reducing the variance of the measure, and the *EEG-based workload index* (W_{EEG} , equation 3.6) has been defined. The formula of the SWLDA discriminant function (3.5) and the *EEG-based workload index* (3.6) equations are reported in the following:

$$y_{test}(t) = \sum_i w_{i\,train} \cdot f_{i\,test}(t) + b_{train} \quad (3.5)$$

$$W_{EEG} = 8MA\left(y_{test}(t)\right) \quad (3.6)$$

Cross-Validations Across the Days

For each subject, different cross-validations have been performed by calibrating the classifier with a triplet of Easy, Medium and Hard conditions, and then by testing it over the remainng triplets. In particular, to investigate the stability of the measure across the different days, I considered three types of cross-validations. The *Intra* cross-validation type, where the training and testing triplets belonged to the same day; the *Short term* cross-validation type, where the training triplets belonged to *Day 1* (*Day 2*), whilts the testing triplets to *Day 2* (*Day 1*); and finally, the *Medium term* cross-validation type, where the training triplets belonged to *Day 1* or *Day 2* (*Day 9*), and the testing triplets to *Day 9* (*Day 1* or *Day 2*).

Subjective, Behavioral and Neurophysiological Data Analyses

Different analyses have been performed: i) NASA-TLX analyses have been used to assess if the neurophysiological measures were consistent with the perception of the workload; ii) task performance, have been analyzed to assess if the subjects reached the saturation of the *Learning Curve* (Ritter and Schooler, 2001), that is, they could maintain the same performance level across the sessions; iii) regression performance analyses have been used to assess the stability of the neurophysiological workload measure over time; iv) neurophysiological workload distributions (W_{EEG}) analyses have been used to assess if the three MATB conditions could be discriminated correctly by the algorithms; v) the features selected within the same session and in the medium term (after a week) have been considered to assess the existence of brain features stable over time. Specifically:

i) NASA-TLX: Subjective workload perception has been obtained by asking the subjects to fill the NASA-TLX questionnaire (see paragraph 2.4.1) at the end of the execution of each task condition. An one-way ANOVA (CI = 0.95) has been performed on the NASA-TLX total scores (independent variable) and *within* factor DIFFICULTY LEVEL (3 levels: Easy, Medium, Hard). Duncan post-hoc test has been performed to find out the differences between the three levels of difficulty.

ii) MATB performance: The MATB performance have been collected and evaluated for each task condition. A two-way repeated-measures ANOVA analysis (CI = 0.95) has been performed on the MATB performance (independent variable) with the SESSION NUMBER and the DIFFICULTY LEVELS (Medium and Hard) as *within* factors.
iii) Regression performance: I have performed analysis by using the two classifiers (asSWLDA and standard SWLDA). For each testing triplet, I calculated the W_{EEG} indexes. The *Area Under Curve* (AUC) values of the *Receiver Operating Characteristic* (ROC, Bamber, 1975) have been calculated by considering couple of W_{EEG} distributions (Easy vs Hard, Medium vs Hard and Easy vs Hard). The area under a ROC curve (which can assume values comprised from 0.5 to 1) quantifies the overall ability of a binary classifier to discriminate the two classes. If the two classes are not discriminable, the AUC will assume the value of 0.5. If the two classes are perfectly discriminable, the AUC will assume the value of 1. For the AUC analysis, I have averaged the AUC values related to the three couples of W_{EEG} distributions (Easy vs Medium, Easy vs Hard, Medium vs Hard) considering as *between* factor the "Classifiers" (asSWLDA or standard SWLDA), and as *within* factor the "Cross-validation types" (Intra, Short term, Medium term). In particular, I performed five one-way ANOVAs (CI=.95). In the first three ones, I compared the AUC values of the two classifiers, for each cross-validation type. After, I performed other two one-way ANOVA in which I compared the AUC values of the three cross-validation types, for each classifier. Duncan post-hoc tests have been performed to assess significant differences between all pairs of levels of the considered factors.
iv) Neurophysiological workload distribution (W_{EEG}): The W_{EEG} distributions have been calculated over the testing conditions by considering only the best classifier derived by the previous analysis (*iii*). In particular, I performed three one-way ANOVAs (CI=.95) by which the W_{EEG} distributions related to the three difficulty levels (Easy, Medium, Hard) have been compared for each cross-validation type (Intra, Short term, Medium term). In addition, Duncan post-hoc tests have been performed to highlight differences between couples of levels of the considered factors. Before every statistical analysis, I used the *z-score* correction (Zhang et al., 1999) on the data to normalize the different behaviors of the subjects.

v) For each subject, the selected brain features in the different experimental sessions, by the asSWLDA and SWLDA, have been analyzed. In particular, two analyses have been performed. The aim of the first analysis was to assess the existence of brain features stable over time, selected both at the beginning and at the end of the experimental protocol. In other words, I investigated if the algorithms were able to select the brain features directly linked to the considered cognitive process (i.e. mental workload). In particular, it has been counted how many times the single feature was selected, during the algorithm calibration phase, in the different EEG channels, for each subject and considered session (T6 and T12). To this purpose, a zeros-matrix of *n* rows per *m* columns (*n* = EEG channels, *m* = subjects number) has been initialized. When a feature was selected from an EEG channel, "1" was summed in the corresponding position of the matrix. At the end, the percentages of selection of the EEG channels, across the experimental sessions, have been calculated for each subject. The choice of considering EEG channels, rather than frequency bins (*features*), has been taken because differences could appear only due to the selection of adjacent bins in the same EEG channel.

The second analysis has been performed to evaluate the capability of the brain features, selected by the asSWLDA and SWLDA, to discriminate the testing classes. In this regards, the *signed r–squares* between the Hard and Easy conditions have been calculated in the first (T6) and last (T12) testing dataset (*RSTest*). The *Coefficient of Determination* (r^2), or *r-square*'s value is a statistical measure computed over a pair of sample distributions, giving a measure of how strongly the means of the two distributions differ in relation to variance (Steel and Torrie, 1960). For two datasets *x* and *y*, the *r-square* is calculated by the following formula:

$$r^2 = \frac{\mathrm{cov}(x, y)^2}{\mathrm{var}(x)\,\mathrm{var}(y)} \tag{3.7}$$

where *cov(x,y)* is the covariance between the two datasets, and *var(•)* is the variance of the single dataset. As $0 < r^2 < 1$, by definition, a *signed r-square* has been derived by multiplying the coefficient of determination by the sign of the slope of the corresponding linear model of the regression analysis. In this way, it has been possible to obtain information not only about if the two

datasets were different, but also about the direction of such difference. The *RSTests* values, related to the features selected by the asSWLDA\SWLDA algorithm during the calibration phase, have been multiplied by the corresponding weights assigned in the T6 (T12) session. Then, the *Weighted Mean r-square* (Wrq) has been estimated for the different testing conditions (T6 and T12) and algorithms (asSWLDA and standard SWLDA). In particular, two conditions have been defined. The INTRA condition refers to the Wrq estimated in correspondence to the features selected within the same session, while the INTER condition means that the Wrq has been calculated by taking into account the r-square calculated in correspondence to the features selected in the other session.

Paired t-test ($\alpha = 0.05$) has been performed on the Wrqs estimated by the asSWLDA (SWLDA) in the two testing sessions (T6 and T12). The hypothesis is that the features selected by the asSWLDA algorithm do not differ significantly over time. In other words, there is not significant difference between the Wrq in T6 and T12. On the contrary, features selected by the standard SWLDA algorithm will differ over time, and this will induce a significant reduction in the discriminability of the classes (i.e AUC) when the features selected in T6 are used in the testing phase of T12.

3.2.2. Results

NASA-Task Load Index (NASA-TLX)

Figure 30 shows changes in the perceived workload estimated by the analysis of the NASA-TLX scores for the different conditions (Easy, Medium and Hard). The workload perception increased as the difficulty of the task increased. The ANOVA revealed a main effect of the difficulty levels ($F(2, 27) = 6.84$; $p = 0.004$). The post-hoc test highlighted that the hard condition was perceived significantly higher than the other two (all $p < 10^{-2}$). Although the perceived workload for the Medium task level was higher than the Easy one, there was no significant difference between them ($p = 0.32$).

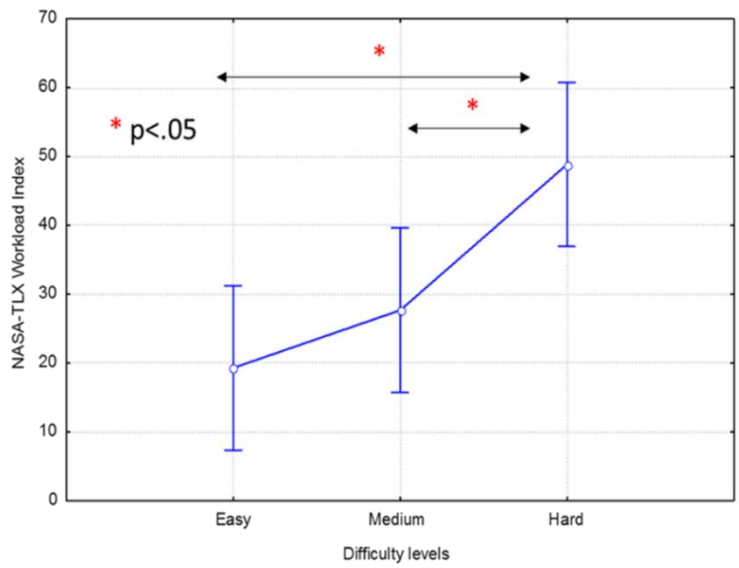

Figure 30. ANOVA results about the perceived workload estimated by the NASA-TLX questionnaire over the different task conditions (Easy, Medium and Hard).

MATB Performances

The ANOVA revealed no significant differences between the MATB performance levels (Mean values: 92.7 ± 2.4) achieved by the subjects within the different experimental sessions ($F(1, 9) = 0.51$; $p = 0.49$). In other words, the subjects maintained their skills after the first week of training, and they did not show any variations across the considered experimental sessions.

Stability Across the Days

Figure 31 represents the mean AUC values (z-score correction) of the two classifiers (standard SWLDA, blue line, and asSWLDA, red line) and the three cross-validation types (Intra, Short term and Medium term). The ANOVAs related to the comparison of the two classifiers along the three cross-validations type, highlighted no significant differences in the Intra condition ([asSWLDA vs standard SWLDA]: $F(1, 18) = 0.32$; $p = 0.57$). On the contrary, significant differences in the Short term and the Medium term conditions have been found (Short term [asSWLDA vs standard SWLDA]: $F(1, 18) = 4.22$; $p = 0.05$; Medium term [asSWLDA vs standard SWLDA]: $F(1,18) = 6.7$; $p = 0.02$). The ANOVAs on the AUC values of the three cross-validation types for each classifier reported the following results. No significant differences between the cross-validation types have been highlighted for the asSWLDA classifier (asSWLDA [Intra, Short term, Medium term]: $F(2, 27) = 0.62$; $p = 0.54$). On the contrary, for the standard SWLDA a significant effect between the cross-validation types has been highlighted (standard SWLDA [Intra, Short term, Medium term]: $F(2, 27) = 3.57$; $p = 0.04$). In particular, the post-hoc test highlighted a significant decrement of the AUC values between the Intra and the Short term and the Medium term cross-validations (all $p < 0.05$). These results demonstrated that the asSWLDA classifier allowed to maintain the performance of the measure stable across days (no significant difference between the cross-validation types). On the contrary, the standard SWLDA classifier suffered of a significant decrement of the performances across days. In fact, since the second day (short term) the performance of the asSWLDA (red line) outperformed the standard SWLDA (blue line).

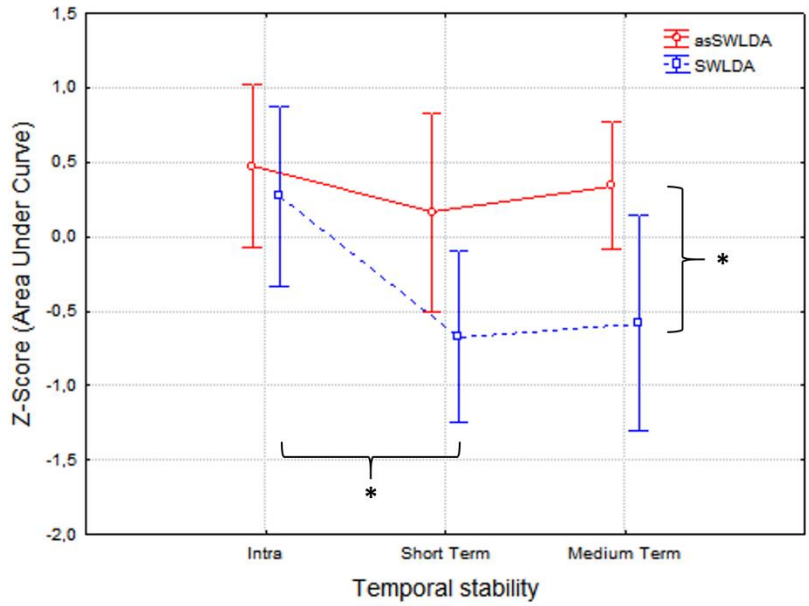

Figure 31. Error bars (CI = .95) related to the AUC values (z-score corrected) of the classifier by using the two classifiers (asSWLDA – red line, standard SWLDA – blue line) and the three cross-validation types (Intra, Short term, Medium term).

Mental Workload Discrimination

As consequence of the results described above, I have chosen to evaluate how well the asSWLDA was able to discriminate the three MATB conditions (Easy, Medium and Hard) across the different days. Figure 32 shows the time representation of the W_{EEG} index for a representative subject. Figure 33 reports the W_{EEG} values measured for each difficulty level (Easy, Medium and Hard) for the different cross-validation types (Intra, Short term and Medium term).

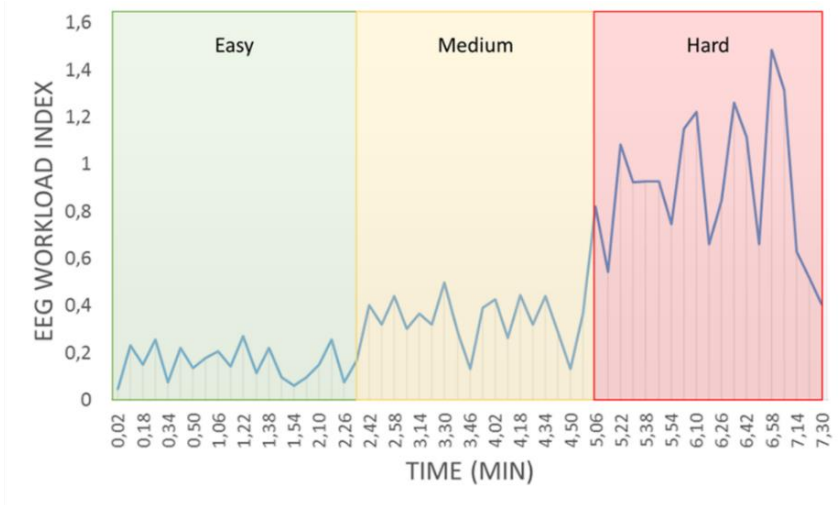

Figure 32. Time representation (time resolution of 8 (sec)) of the W_{EEG} index evaluated for the three task conditions. From left to right: Easy, Medium and Hard condition for a representative subject.

ANOVA analyses showed that the W_{EEG} values related to the three difficulty levels (Easy, Medium and Hard) were significantly different for each cross-validation type (Intra: $F(2, 27) = 417.65$; $p = 10^{-6}$; Short term: $F(2, 27) = 360.9$; $p = 10^{-6}$; Medium term: $F(2, 27) = 459.78$; $p = 10^{-6}$). The post-hoc test confirmed that the W_{EEG} distributions related to the three difficulty levels have been significantly different each other at each cross-validation type (all $p<10^{-4}$). The Figure 34 shows the average W_{EEG} values distributions for each subject across the cross-validation types.

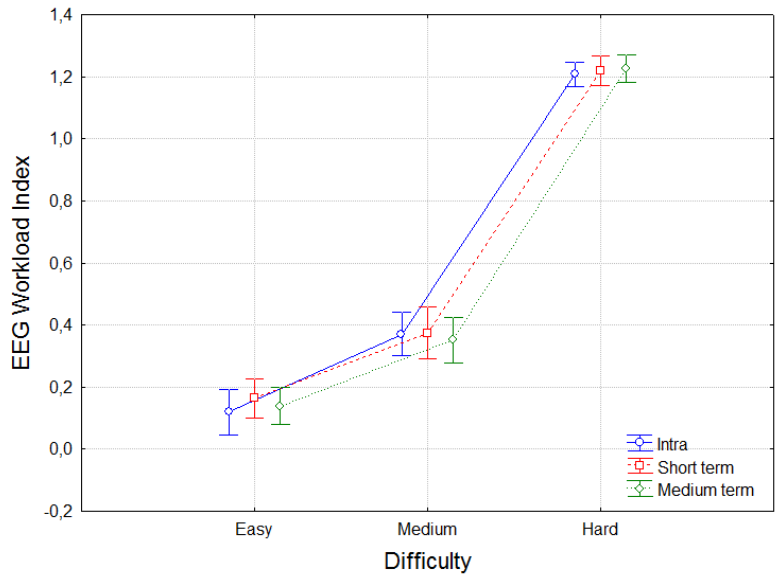

Figure 33. Error bars (CI = .95) related to the W_{EEG} values measured for each difficulty level (Easy, Medium and Hard) for the different cross-validation types (Intra – blue line, Short term – red line, and Medium term – green line).

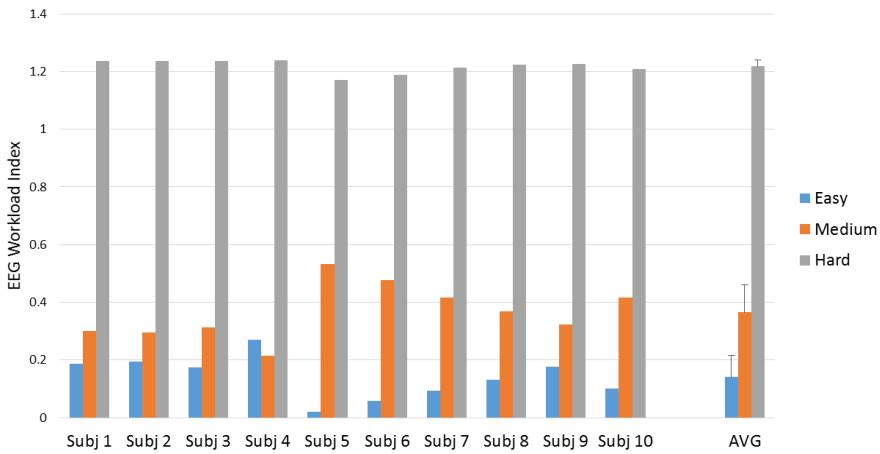

Figure 34. Averages of W_{EEG} values distribution in the different task condition (Easy – blue bar, Medium – orange bar, and Hard – gray bar) for each subject.

Brain Features Stability Over Time

Figure 35 reports the percentages related to the EEG channels most commonly selected by the asSWLDA (on the left) and by the SWLDA (on the right). In the asSWLDA case, for each subject there are at least 2 EEG channels who have been selected both at the beginning (T6) and at the end (T12) of the experimental protocol. Specifically, for 7 subjects out of 10, frontal and parietal channels have been always selected, respectively, in the theta and alpha EEG frequency bands.

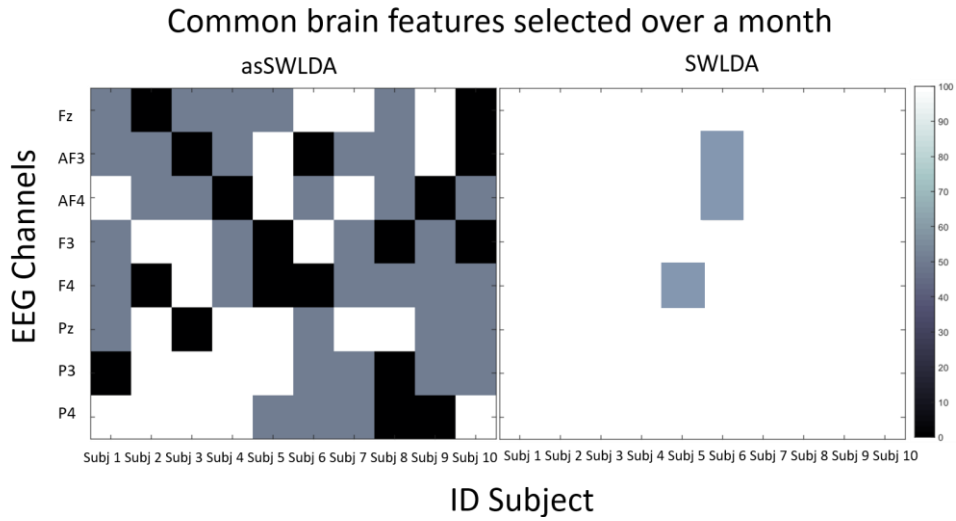

Figure 35. Percentages related to the EEG channels most commonly selected by the asSWLDA (on the left) and by the SWLDA (on the right). White color means that brain features have been selected in all the considered sessions. On the contrary, the EEG channel is black colored if no features have been selected from it. Gray color means that the considered EEG channel has been selected only once.

In the SWLDA case, all the EEG channels have been selected in all the EEG bands and considered sessions. The paired t-test on the Wrq corresponding to the features selected by the asSWLDA, within a week (T6 and T12), did not show any significant ($p = 0.1$) difference between the INTRA and INTER conditions (Figure 36). On the contrary, the standard SWLDA reported a significant reduction ($p = 0.01$) in terms of Wrq, between the INTRA and INTER condition.

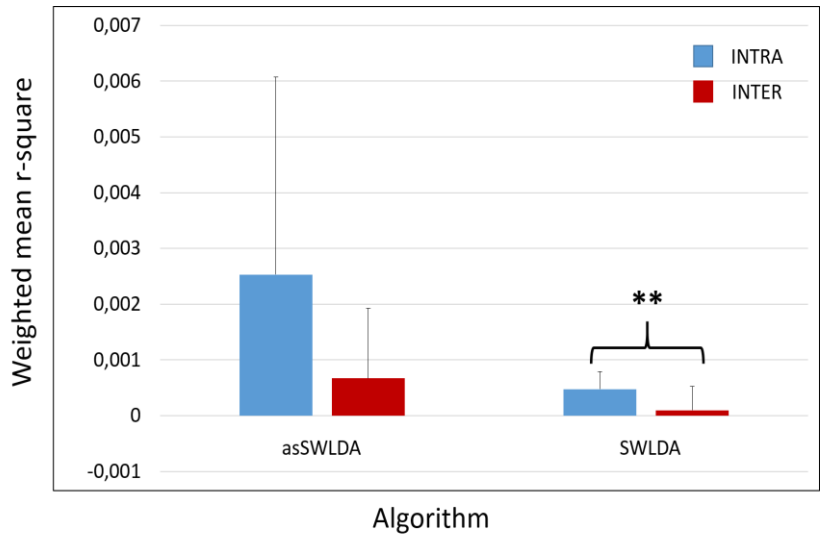

Figure 36. Wrqs corresponding to the brain features selected by the asSWLDA and the standard SWLDA in the INTRA (blu bars) and INTER (red bars) condition. The asSWLDA did not show any significant difference ($p = 0.1$) between the INTRA and INTER conditions. On the contrary, the standard SWLDA reported a significant reduction ($p = 0.01$) of the Wrqs calculated in the two conditions.

3.2.3. Discussions

A new implementation of the SWLDA to optimize the features selection and make the workload measure stable across different days (a week) has been proposed. In particular, the hypothesis was that, as long as the subject is trained in performing correctly a task, the requested cognitive processes will be almost the same over time (Borghini et al., 2015b). Moreover, there are literature evidences that demonstrated a linear correlation between the mental workload and specific EEG rhythms (theta and alpha bands). Based on these assumptions, it could be possible to identify the EEG features directly related to the mental workload, who remain stable across the days. The proposed algorithm (asSWLDA) was able to identify those features and to maintain higher workload evaluation performance than the standard SWLDA, across the different days (Figure 35). In addition, although the asSWLDA generally selected less features than the SWLDA, the Wrqs did not differ between the INTRA and INTER condition (Figure 36). In other words, the asSWLDA was able to recognize the brain features

corresponding to cognitive processes stable over time (T6 and T12), and there were no differences in calculating the Wrq with the features selected in T6 or T12, that is within a week. On the contrary, significant reduction was found on the Wrqs estimated in the INTER condition, with respect to the INTRA, with the brain features selected by the standard SWLDA. The regression performance analyses (AUC) and the workload distributions (W_{EEG}) analyses showed that if the asSWLDA is trained with the features related to the mental workload (frontal theta and parietal alpha EEG rhythms), it will be possible to make the workload measure stable across the days (no significant decrement across Intra, Short term and Medium term cross-validations), so, even after a week, it will not be necessary to recalibrate the algorithm with new EEG data of the user. On the contrary, the standard implementation of the SWLDA was not capable to keep such stability over time, since the classification was too specific to the calibration dataset of the considered day. This aspects are highly important for the practical usability of the system, especially in operational environments. In fact, it should be enough to calibrate the algorithm with the specific operator EEG parameters only once and then use it without further adjustments, maintaining high reliability and stability over a week. In addition, the algorithm was able to differentiate significantly ($p < 0.05$) the mental workload levels over the three experimental conditions (Easy, Medium and Hard). The results also highlighted the higher resolution of the neurophysiological measure than the subjective ones, since the Easy and the Medium tasks were not discriminable from the NASA-TLX scores.

3.2.4. 3.2.4. Conclusions

A new classification algorithm (asSWLDA) for the evaluation of the mental workload of the user has been proposed and validated. I have demonstrated that 1) the asSWLDA significantly outperformed the standard implementation of the SWLDA; 2) the performance of the asSWLDA algorithm remained high across different days (a week) without any recalibrations, so it is reliable over time; 3) the algorithm was able to significantly differentiate three mental workload levels. The innovation of the study, with respect to the actual literature, is that the proposed algorithm has been showed to be reliable over a week without the necessity of

recalibrate it. Further experiments will be performed to test and extend the period of stability ("Long-term"). Also, the present study, carried out in a laboratory environment, will be replicated on a larger sample size (more than ten subjects) and in a more realistic scenario, involving professional operators. Finally, the reliability of the asSWLDA classifier will be tested also in other fields (e.g. active BCI), where the decreasing of the system performances across days represents also a big issue, and it is not always possible to calibrate the system.

3.3. asSWLDA Testing in Realistic ATM Environment: Stability Over a Month

The objective of the study was to provide a methodology for the evaluation of Air-Traffic Controllers' mental workload in operational environment, and to overcome the issues described previously (see paragraph 1). For such purposes, the brain activity has been recorded on twelve professional *Air Traffic Controllers* (ATCOs) while performing high realistic *Air Traffic Management* (ATM) scenarios (see paragraph 2.3.2). From the EEG signals, a workload index has been computed by means of a machine - learning approach (Patent II, 2015). In particular, the asSWLDA (Patent I, 2015) has been used to compute such *EEG-based Mental Workload Index* (W_{EEG}). The study has been designed to investigate two important key issues: the over-time reliability of the neurophysiological workload measure, and the accuracy of the methodology in comparison with the standard (i.e. subjective) workload measures:

- *Reliability over time of the neurophysiological workload evaluation.* The first phase of the study had the aim to test the reliability over time (a month) of the neurophysiological workload measure, by using two different models (see Paragraph 3.1). Five out of the twelve ATCOs gave their availability to take part in both the experimental sessions (*Day 1* and *Day 30*), while the remaining controllers attended only the last experimental session (*Day 30*).
- *Comparison between neurophysiological and subjective workload measurements.* The second phase of the study aimed to test the effectiveness of the neurophysiological workload measure in comparison with the subjective assessment (ISA and SME workload scores). For this purpose, only the last session (*Day 30*) has been considered, where all the ATOCs participated to the experiment (twelve professional ATCOs in total).

3.3.1. Experimental Protocol

Subjects

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

Experimental Setup

The ATCOs have been asked to perform *Air Traffic Management* (ATM) scenarios (see paragraph 2.3.2) in a high realistic setting, that is, a functional simulated ATM environment developed and hosted at ENAC (Figure 37a). 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 (Figure 37b) who have interacted with the ATCOs with the aim to reproduce real communications.

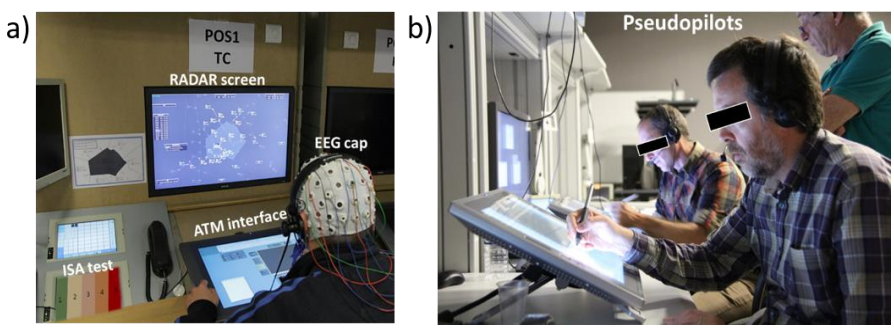

Figure 37. a) The ENAC simulator platform, composed of two screens, a 30" (RADAR) screen to display radar image and a 21" screen to interact with the radar image (ATM interface). On the little screen on the left bottom, the ISA test was proposed every 3 minutes. b) Pseudo-Pilots have interacted with the ATCOs with the aim to simulate real-flight communications.

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 *Day1* and *Day30*. 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 mental workload demands (EASY, MEDIUM, HARD). For each scenario, the presentation of the difficulty conditions has been randomized. In addition, although the two ATM scenarios were different, in order to avoid any habituation or expectation effect, they have been designed identically in terms of complexity within the same difficulty levels (i.e. for instance EASY Day1 vs. EASY Day30), to make them comparable between sessions, and to avoid any bias in the results. Such scenarios have been validated and tested by a *Subject-Matter Expert* (SME) from ENAC before the experiments. It has to be stressed that air traffic shape was not constant, and the transitions between the different difficulty levels was smoothly organized in order to have ATM scenarios as much realistic as possible. Figure 38 shows a representative scenario's complexity profile.

Collected Data for the Mental Workload Evaluation

Neurophysiological Data.

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

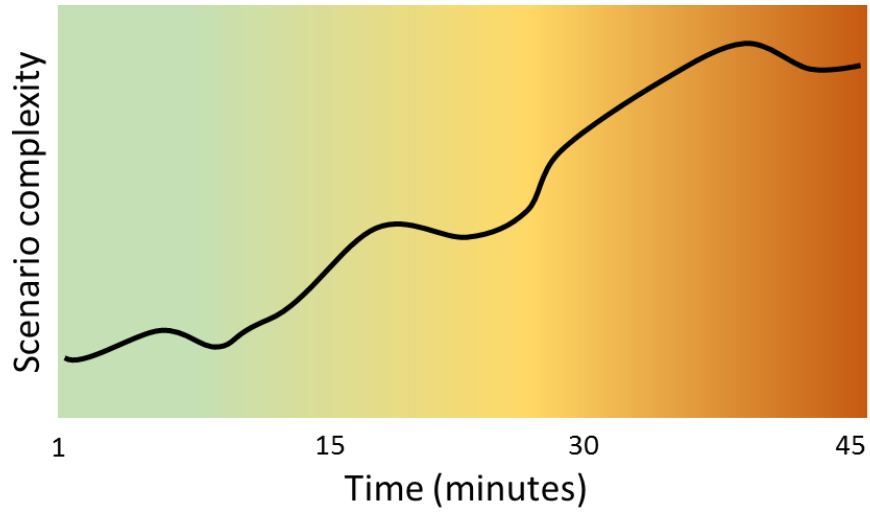

Figure 38. Representative ATM scenario's complexity profile. The complexity of the task has been modulated gradually in order to obtain realistic ATM scenarios.

Subjective workload Assessment (Self-assessment).
The ATCOs have been asked to fill the ISA (see paragraph 2.4.2) with the aim to define a profile of the operator's workload perception during the execution of the ATM task. The ISA 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 37a). The keypad flashed and sounded when the workload rating was required, and the participants simply pressed the button related to their workload perception. ATC Experts (SMEs), seated behind the ATCO, have been asked to provide independent rates of the workload experienced by the ATCO by filling the paper version of the ISA. ISA scores provided by experimental subjects and by the SMEs are named, hereafter, respectively, *SELF-ISA* and *SME-ISA.*

Subjective Workload Assessment (SME assessment).
ATC Experts (SMEs) provided independent rates, from 1 to 5 accordingly to the ISA scale, of the workload experienced by the ATCO. In particular, SMEs have been asked to express their continuous judgment depending on the ongoing overall performance of the examined ATCO. Such judgment took into account the quality and time of the indications/information provided to the pilots, separation planning strategy, responsiveness, and general management of the air-traffic situation.

Neurophysiological Data Analysis

For each session (*Day 1*, *Day 30*) and difficulty level (EASY, MEDIUM, HARD), the biosignal dataset (EEG, EOG signals) has been segmented in five consecutive parts (named hereafter as "runs") of 3 minutes each, in order to define 5 EASY runs (E1, E2, E3, E4, E5), 5 MEDIUM runs (M1, M2, M3, M4, M5), 5 HARD runs (H1, H2, H3, H4, H5), and to have the same time resolution of the ISA scores (SELF-ISA, SME-ISA), and therefore allowing a more direct comparison between all the collected measures. The recorded EEG signal has been firstly band-pass filtered with a fifth-order Butterworth filter (low-pass filter cut-off frequency: 30 (Hz), high-pass filter cut-off frequency: 1 (Hz), whilst the EOG signal has been band-pass filtered with a low-pass filter (cut-off frequency: 7 (Hz)), and high-pass filter (cut-off frequency: 1 (Hz)), has then been used to remove eyes-blink contributions from each EEG epoch, by using the Gratton and Coles (1983) algorithm, available on the EEGLab toolbox (Delorme and Makeig, 2004). For other sources of artifacts (i.e. ATC-operators normally talk and perform several body movements during ATM activities), specific procedures of the EEGLAB toolbox have been used (Delorme and Makeig, 2004). Firstly, the EEG signal has been segmented into epochs of 2 seconds (*Epoch length*), shifted of 0.125 seconds (*Shift*). This windowing have been chosen with the compromise to have both a high number of observations (see equation 3.7), in comparison with the number of variables (see equation 3.8), and to respect the condition of stationarity of the EEG signal (Elul, 1969). In fact, this hypothesis is necessary to proceed with the spectral analysis of the signal. The EEG epochs where the signal amplitude exceed ±100 (μV) have been marked as "artifact" (*Threshold criteria*). Then, each EEG epoch has been interpolated in order to check the slope of the trend within the considered epoch (*Trend estimation*). If such slope was higher than 3, the considered epoch was marked as "artifact". The third check was based on the EEG *Sample-to-sample* difference. If such difference, in terms of amplitude, was higher than 25 (μV), it meant that an abrupt variation (non-physiological) happened, and the EEG epoch was marked as "artifact". All the previous values have been chosen following the guidelines reported in Delorme and Makeig, (2004). At the end, the EEG epochs marked as "artifact" have been removed with the aim to have a clean EEG dataset from which estimating the brain parameters for the analyses. The percentage

(with respect to the total number of epochs averaged on all the subjects) of EEG epochs containing artifacts and removed from the EEG dataset was 20% (± 13%). From the clean EEG dataset, the *Power Spectral Density* (PSD) has been calculated for each EEG epoch using a of 2-second Hanning window (that means 0.5 (Hz) of frequency resolution). The application of a Hanning window helped to smooth the signal at the edges of the epochs, improving the accuracy of the PSD estimation (Harris, 1978). Then, the EEG frequency bands have been defined for each ATCO by the estimation of the *Individual Alpha Frequency* (IAF) value (Klimesch, 1999). In order to have a precise estimation of the alpha peak and, hence of the IAF, the subjects have been asked to keep the eyes closed for a minute before starting with the experiments. Finally, a spectral features matrix (EEG channels x Frequency bins) has been obtained in the frequency bands directly correlated to the mental workload. In particular, only the theta rhythm (IAF-6 ÷ IAF-2), over the EEG frontal channels (Fz, F3, F4, AF3, AF4), and the alpha rhythm (IAF-2 ÷ IAF+2), over the EEG parietal channels (Pz, P3, P4), have been considered as parameters for the mental workload evaluation.
In the considered case, the number of observations (*#Observations*) and number of variables *(#Variables*) were:

$$\#Observations = \frac{Run\ duration - Epoch\ length}{Shift} = 1424 \quad (3.7)$$

where *RunDuration* = 180 seconds (E_k, M_k, H_k), k = [1, 2,…,5], *EpochLength* = 2 seconds, *Shift* = 0.125 seconds.

#Variables=(#Frontal sites*#Theta frequency bins)+(#Parietal sites*#Alpha frequency bins)= 72 (3.8)

where *#FrontalSites* = 5 (Fz, F3, F4, AF3, AF4), *#ParietalSites* = 3 (Pz, P3, P4), *#ThetaFrequency bins* = (IAF-6 ÷IAF-2)·0.5 = 9, *#AlphaFrequency bins* = (IAF-2 ÷ IAF+2)·0.5 = 9.

EEG-Based Mental Workload Index (W_{EEG})
Figure 39 shows the algorithm steps used in this study for the evaluation of the user's EEG-based mental workload. The effectiveness of two linear classifiers (standard SWLDA and asSWLDA), in terms of reliability over a

month, has been investigated. Both the classifiers have been used to select the most discriminating features of the different experimental conditions (i.e. Easy and Hard). Once identified, each classifier assigned to each relevant feature specific weights ($w_{i\,train}$), plus a bias (b_{train}). On the contrary, weights related to not relevant features for the classification model have been set to 0. This step represented the calibration phase of the algorithm. The parameters estimated during the calibration phase have been then used to calculate the *Linear Discriminant Function* ($y_{test}(t)$) over the testing EEG dataset (testing phase, equation 3.9), defined as the linear combination of the testing spectral features ($f_{i\,test}$) and the classifier weights ($w_{i\,train}$), plus the bias (b_{train}). Finally, a moving average of *k* seconds (kMA) has been applied to the $y_{test}(t)$ function in order to smooth it out by reducing the variance of the measures, and the result has been named *EEG-based Workload Index* (W_{EEG}, equation 3.10). The higher is the *k* value, the less will be the variance of the measure. Accordingly with the SMEs, for a proper evaluation of the mental workload during the execution of the ATM task, the *k* value has been set to 30 seconds.

$$y_{test}(t) = \sum_i w_{i_{train}} * f_{i_{test}}(t) + b_{train} \tag{3.9}$$

$$W_{EEG} = kMA\big(y_{test}(t)\big), \quad k = 30\ (sec) \tag{3.10}$$

The third couple of runs (E3, H3) of each session has been chosen to calibrate 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 to calibrate the classifier, that is the best compromise in terms of stable difficulty level to represent the lowest and the highest air-traffic complexity condition. To evaluate the reliability of each classifier in discriminating the workload 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).

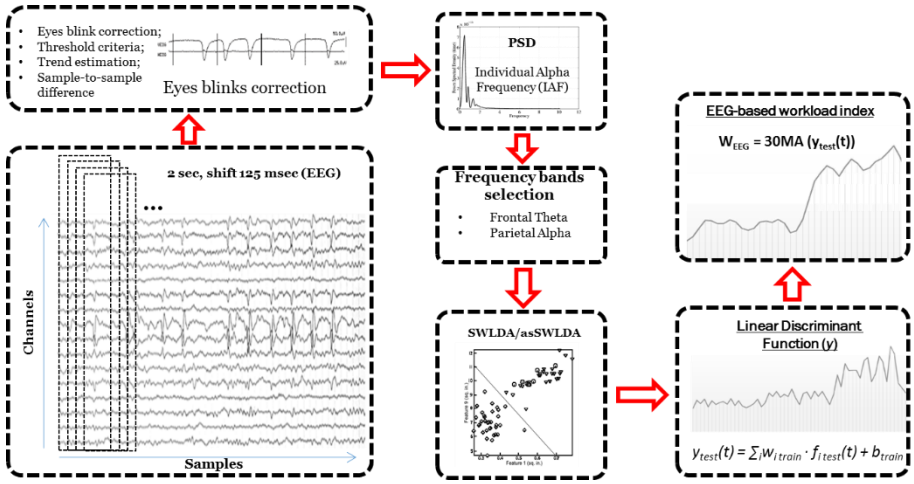

Figure 39. EEG-based workload index estimation. The filtered (1-30 (Hz)) EEG has been segmented into epochs of 2 (sec), shifted of 0.125 (sec), and the filtered (1-7 (Hz)) EOG signal has been used to remove the eyes-artifact contribution. Other sources of artifacts have been deleted by using specific algorithms. Then, the PSD has been evaluated for each epoch ($Epoch_{PSD}$) taking into account only the EEG frequency bands and channels correlated with the mental workload (frontal theta and parietal alpha bands). The SWLDA (asSWLDA) has then been used to select the most relevant brain spectral features for the discrimination of the mental workload levels. The linear discriminant function has been calculated on the EEG testing dataset, and the W_{EEG} has been defined as the moving average of 30 seconds (30MA) applied to the linear discriminant function ($y_{test}(t)$).

Statistical Analyses

Subjective Workload Assessment

The two EEG recording sessions (*Day 1*, *Day 30*) have been compared in terms of subjective workload perception (*SELF-ISA* and *SME-ISA* scores) by the ATCOs and SMEs. Three two–tailed paired t-tests ($\alpha = 0.05$) 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 performed to assess the differences between all pairs of levels of the considered factor.

EEG-Based Workload Assessment

The reliability 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* (*Day 30*) and the testing data to *Day 30* (*Day 1*). A two-way repeated measures ANOVA (CI = 0.95) 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.

Brain Features Selection and Stability Analyses

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

Furthermore, for each controller, the existence of stable brain features and importance of the selected ones, separately by the asSWLDA and SWLDA, have been assessed as described in detail in the paragraph 3.2.1. In particular, two conditions have been defined. The INTRA condition refers to the Wrq estimated in correspondence to the features selected within the same day, while the INTER condition means that the Wrq has been calculated by taking into account the r-square calculated in correspondence to the features selected in the other day. Paired t-test ($\alpha = 0.05$) has been performed on the Wrqs estimated by the asSWLDA (SWLDA) in the two days (*Day1* and *Day30*). The hypothesis is that the features selected by the asSWLDA algorithm do not differ significantly over time. In other words, there is not significant difference between the Wrq in *Day1* and *Day30*. On the contrary, features selected by the standard SWLDA algorithm will differ over time, and this will induce a significant reduction in the discriminability of the classes (i.e AUC) when the features selected in *Day1* are used in the testing phase of *Day30*.

Comparison Between Neurophysiological and Subjective Measures

Self - Workload Assessment

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

EEG-Based Workload Assessment

An 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.

Accuracy of Neurophysiological Measurement with respect to Standard Workload Assessment

In order to assess the accuracy of the 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*).

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

3.3.2. Results

Self-Workload Assessment

The three two–tailed paired t-tests (α = 0.05) did not highlight any difference between the two sessions (*Day1* and *Day30*), both in terms of *SELF-ISA* (Easy: p = 0.19; Medium: p = 0.63; Hard: p = 0.62) and *SME-ISA* (Easy: p = 0.78; Medium: p = 0.22; Hard: p = 0.11) scores (Figure 40). In fact, as described in paragraph 3.3.1, the different ATM scenarios have been desingned with the aim to maintain comparable levels of complexity, and they were tested by Expert ATCOs before starting the experiments.

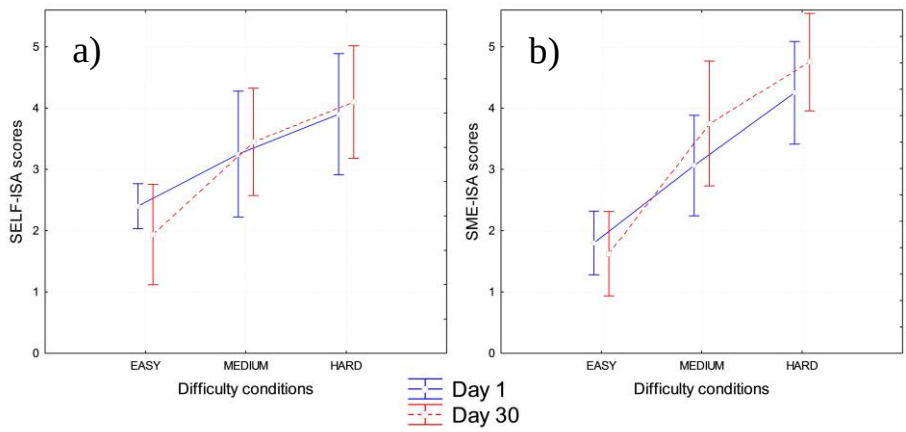

Figure 40. ANOVAs related to the *SELF-ISA* and the *SME-ISA* scores for each session (*Day1* – blue line, *Day30* – red line) and difficulty condition (Easy, Medium, Hard). Both the *SELF-ISA* (panel a) and the *SME-ISA* (panel b) scores did not show any significant differences between each difficulty condition.

EEG-Based Workload Assessment

The two-way repeated measures ANOVA (CI = 0.95) highlighted a significant (F(1, 4) = 10.6; p = 0.03). main effect between the two factors (CLASSIFIERS and CROSS-VALIDATIONS). 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 (Figure 41). 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 (p = 0.2) between the two classifiers were highlighted regarding the *Intra* cross-validation type (Figure 41).

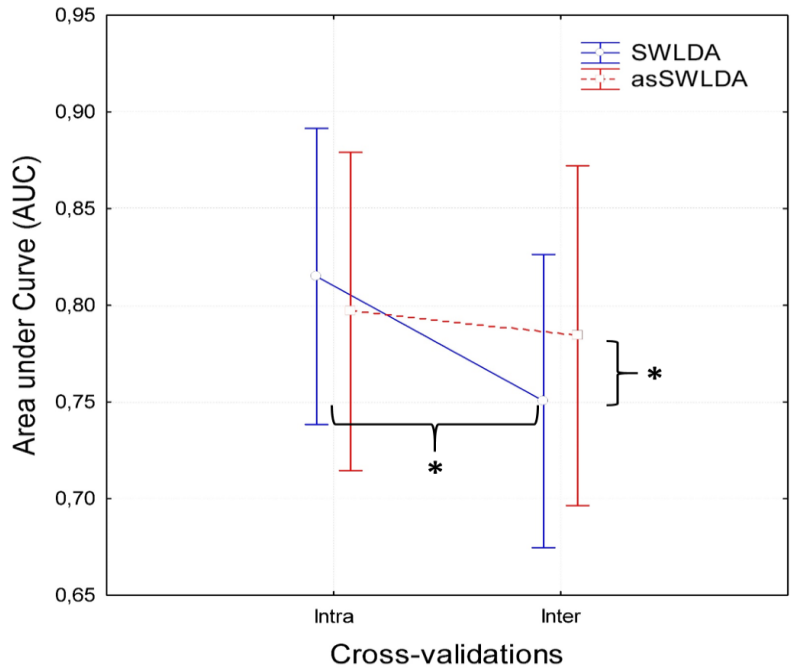

Figure 41. ANOVA related to the AUC values of the SWLDA (blue line) and asSWLDA (red line) calculated on the "Easy vs Hard" conditions over the two cross-validation types (*Intra* and *Inter*). In particular, a significant AUC decrement (p = 0.005) between the *Intra* and *Inter* cross-validation types was found for the standard SWLDA. On the contrary, there was not significant difference (p = 0.33) between the two cross-validation types, concerning the asSWLDA. Focusing on the *Inter* cross-validation type, the SWLDA performance was significantly (p = 0.04) lower that the asSWLDA.

Brain Features Selection Analysis

The two–tailed paired t-tests (α = 0.05) highlighted that both the number of features (p = 0.0007) and the related EEG channels (p = 0.0003) used by the asSWLDA model were significantly lower than those used by the standard SWLDA model (Figure 42). In Figure 42, both the numbers of features and EEG channels are reported in terms of percentage with respect to the total number of features, and EEG channels. In particular, the asSWLDA selected the 5.2% of the available features by using the data from the 37% of EEG channels. In other words, the asSWLDA algorithm selected for each

ATCO about 4 features on 3 EEG channels. On the contrary, the standard SWLDA used the 44% of the available features by using the 100% of EEG channels, that is, the standard SWLDA used for each ATCO 32 features on all the 8 EEG channels. Generally, the asSWLDA algorithm used roughly the 10% of the information employed by the standard SWLDA, achieving higher performance.

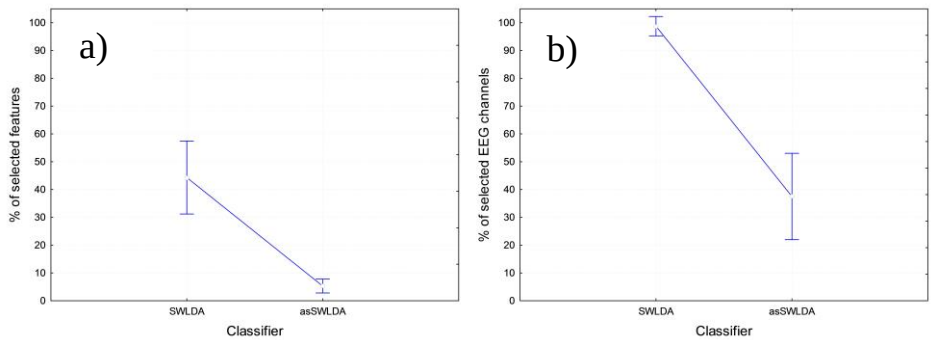

Figure 42. ANOVAs related to the number of features (panel a) and related number of EEG channels (panel b) selected by the two classification models (SWLDA and asSWLDA). As expected, both the number of features and related number of EEG channels used by the asSWLDA were significantly lower than those used by the standard SWLDA algorithm, respectively, $p = 0.0007$ and $p = 0.0003$.

Brain Features Stability Over Time

In Figure 43 are reported the percentages related to the EEG channels most commonly selected by the asSWLDA (on the left) and by the SWLDA (on the right). In the asSWLDA case, at least one EEG channel has been selected, for all the subject, both at the beginning (*Day 1*) and at the end (*Day 30*) of the experimental protocol.

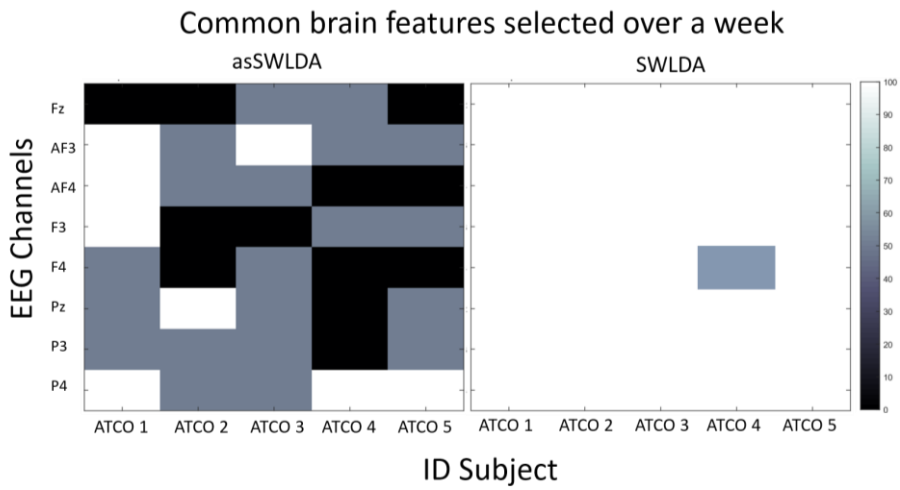

Figure 43. Percentages related to the EEG channels most commonly selected by the asSWLDA (on the left) and by the SWLDA (on the right). White color means that brain features have been selected in all the considered sessions. On the contrary, the EEG channel is black colored if no features have been selected from it. Gray color means that the considered EEG channel has been selected only once.

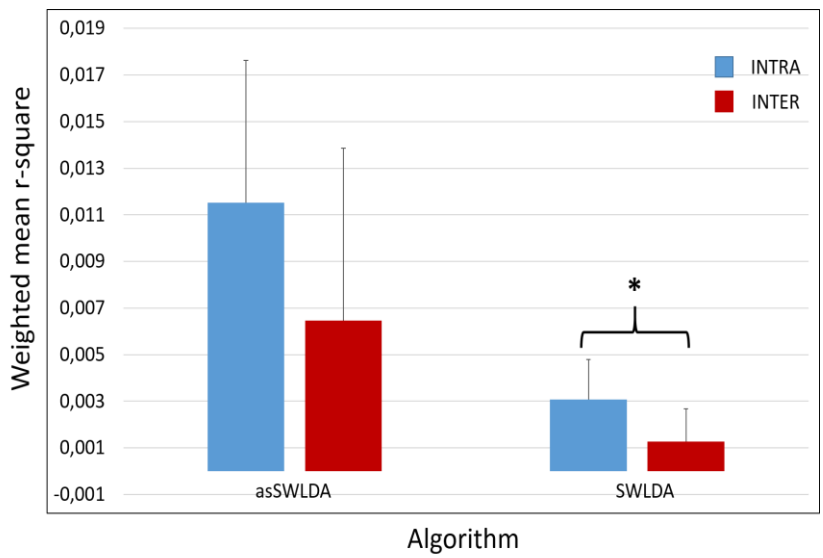

Figure 44. *Weighted mean r–squares* (Wrqs) estimated in correspondence of the brain features selected by the asSWLDA and the standard SWLDA in the INTRA (blu color) and INTER (red color) condition. The asSWLDA did not show any significant difference ($p = 0.17$) of Wrq between the INTRA and INTER conditions. On the contrary, the standard SWLDA reported a significant reduction ($p = 0.04$) of the Wrqs calculated in the two conditions.

The paired t-test on the Wrq corresponding to the features selected by the asSWLDA, within a month (*Day1* and *Day30*) and T12), did not show any significant ($p = 0.17$) difference between the INTRA and INTER conditions (Figure 44). On the contrary, the standard SWLDA reported a significant reduction ($p = 0.04$) in terms of Wrq, between the INTRA and INTER condition.

Comparison Between Neurophysiological and Subjective Measures

Subjective Assessment

The two-way repeated measures ANOVAs (CI = 0.95) highlighted a significant effect ($F(2, 22) = 10.88$; $p = 0.005$) between the two factors, DIFFICULTY CONDITIONS (Easy, Medium, Hard) and SUBJECTIVE WORKLOAD SCORES (*SELF-ISA* and *SME-ISA*). The post-hoc test highlighted significant differences (all $p < 0.001$) between the difficulty conditions (i.e. Easy vs Medium, Medium vs Hard, Easy vs Hard) for both the *SELF-ISA* and the *SME-ISA* scores. In addition, the *SME-ISA* scores over the Medium ($p = 0.0007$) and Hard ($p = 0.00007$) conditions were significantly higher than those related to the *SELF-ISA* scores (Figure 45).

EEG-Based Workload Assessment

The one-way ANOVA (Figure 46) 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 the difficulty conditions (i.e. Easy vs Medium, Medium vs Hard, Easy vs Hard).

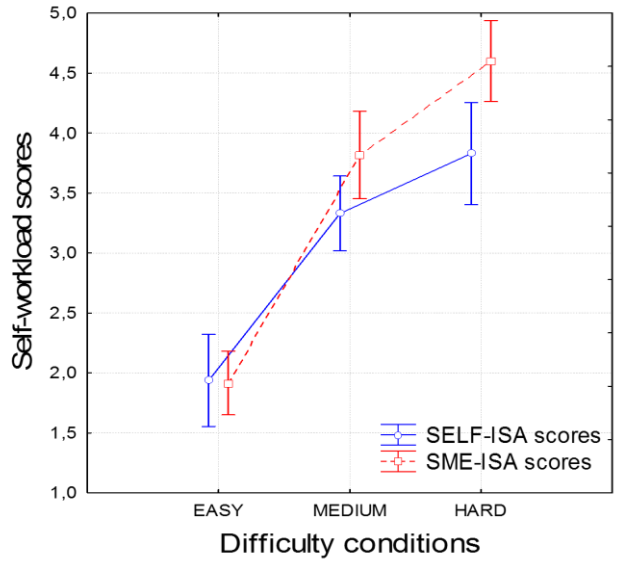

Figure 45. ANOVA related to the *SELF-ISA* (blue line) and the *SME-ISA* (red line) scores along the three difficulty conditions. Results showed significant differences ($p < 0.003$) between all the difficulty conditions for both the *SELF-ISA* and *SME-ISA* scores.

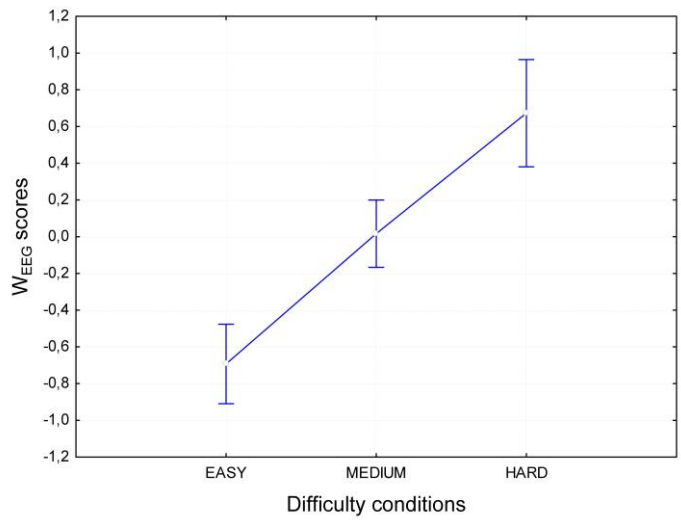

Figure 46. ANOVA related to the W_{EEG} scores along the three difficulty conditions. Results showed significant differences between all the difficulty conditions (all $p < 0.001$).

Accuracy of Neurophysiological measurement with respect to Standard Workload Assessment

The correlation analysis (Table 5), by means of the Pearson's correlation coefficient, highlighted a high and positive correlation between the EEG-based workload index (W_{EEG}) and 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, and they followed the variation of the mental workload demanded by the ATM scenarios and experienced by the ATCOs during the execution of the task (Figure 47).

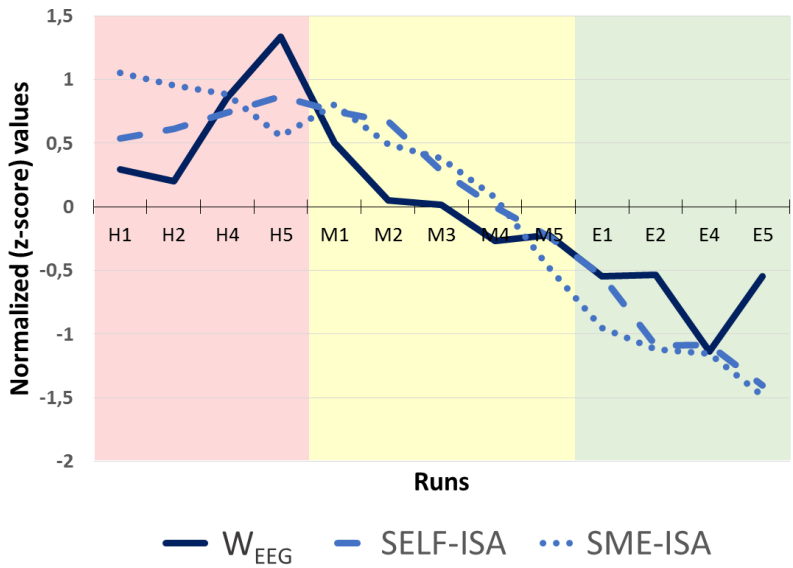

Figure 47. Shape of the three workload indexes: the neurophysiological (W_{EEG} – dark blue line) and the subjective ones (*SELF-ISA* – solid light blue line, and *SME-ISA* – dashed light blue line). The three measures were able to follow the workload profile during the execution of the ATM scenarios.

Finally, the Fisher's R-to-Z analysis (Table 5) on the two correlation indexes showed no differences between them ($p = 0.676$).

The scatterplot in the Figure 48 highlighted a high and positive correlation between the neurophysiological and the subjective workload indexes.

Table 5: Pearson's Correlation Coefficient (R) and significance (p) level between the neurophysiological (WEEG) workload index and both the subjective measures (ISA and SME scores). The Fisher's R-to-Z transformation showed no significant difference between the two correlation values.

Statistical analysis of the correlation	**Pearson's correlation index**	
	R	**p**
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

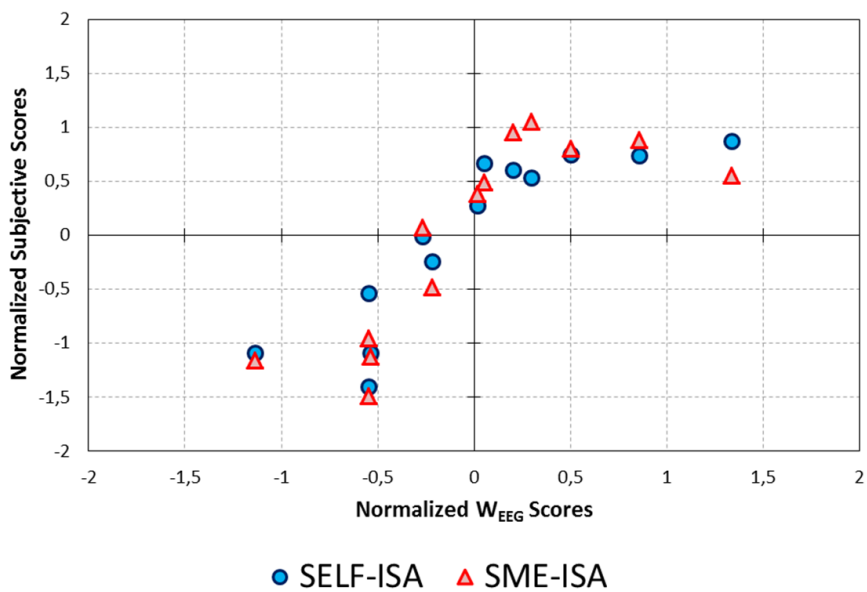

Figure 48. Scatterplot of the subjective workload measures (*SELF-ISA* – blue dots and *SME-ISA* – red dots) with respect to the neurophysiological workload measure. On *the x-axis* the normalized W_{EEG} index, on the *y-axis* the normalized *SELF-* and S*ME-ISA* indexes. The results of the correlation analyses showed high and significant correlation among all the workload measures (Table 4).

3.3.3. Discussions

The purpose of the proposed study was to investigate the effectiveness of a methodology for the mental workload evaluation in an operational *Air Traffic Management* (ATM) environment, by using neurophysiological measures. The brain activity (EEG) and subjective measures (*SELF-ISA* and *SME-ISA* scores) of 12 professional ATCOs from ENAC (France, Toulouse) were gathered while performing high-realistic ATC tasks and analyzed by a machine-learning approach. The asSWLDA (see paragraph 3.1) was used to compute a mental workload index based on the EEG activity of the user (W_{EEG}). The proposed study was designed with the aim to investigate two important key issues of the use of neurophysiologicalmeasurements in operational environments: 1) the reliability of the workload evaluation over time (a month), and 2) the accuracy in comparison with the standard (i.e. subjective) workload measures. Concerning the first issue, the results demonstrated that the proposed algorithm was able to maintain a high reliability across a month (Figures 40 and 41). It had been already demonstrated the reliability of the model over a week (see paragraph 3.2). In other words, it is possible to calibrate the asSWLDA model and then using it for the mental workload evaluation, with neither any recalibration nor any accuracy reductions over a month. In addition, the asSWLDA model was able to select a lower number of brain features (Figure 42) and EEG channels (37% of the channels, that means 3 EEG channels in this specific study) in comparison with the standard SWLDA (100%, of the available channels that means 8 EEG channels), to be used for the workload assessment. 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., 2015a), but it could become tricky to empirically (and manually) find out the proper 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 fact, the asSWLDA showed the existence of specific brain features stable over a month, since they were selected both in the *Day1* and *Day30* (Figure 43), and the corresponding Wrqs did not differ between the INTRA and INTER condition. On the contrary, significant differences were found between the INTRA and

INTER condition on the Wrqs calculated in the correspodence to the brain features selected by the standard SWLDA (Figure 44). More important, the results suggested that the selected features for the asSWLDA could be almost an order of magnitude less than those selected by using the standard SWLDA. As direct consequence, a lower number of EEG channels will be necessary with the asSWLDA than with the standard SWLDA. Since one of the most important limitations in using p-BCI systems in operational environments is represented by wearing the EEG cap, if the number of EEG channels used to define the classification model is low, the intrusiveness of the recording system will be low too. In this regard, just moving to a plausible operational showcase, the proposed algorithm will be calibrated once with many electrodes (e.g. 64 channels) to select the specific subjective features, and then used online with the bunch of EEG channels selected by the asSWLDA (e.g. 2 or 3 electrodes).

Concerning the second issue, the EEG-based workload measure was compared with two subjective workload measures, one provided by the ATCOs (*SELF-ISA* scores), and the second provided by two ATC Experts (SMEs), who were 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 (Figures 47 and 48). Such result is very important, since it showed how the W_{EEG} and both the subjective indexes (*SELF-ISA* and *SME-ISA*) were able to follow the actual fluctuations of the mental workload experienced by the ATCOs during the experimental task. In addition, the *SELF-ISA* scores showed a higher workload perception during the MEDIUM and the HARD conditions in comparison with the *SME-ISA* ones. This result highlighted the main limitation of the subjective measures: they are highly operator-dependent and cannot be used to quantify objectively the operators' mental states (i.e. mental workload). On the contrary, the neurophysiological measures can provide, with high resolution, objective evaluations of the operator's mental state (Figure 46). Finally, it has to be underlined that the proposed algorithms do not require *a-priori* information about consecutive data, that is, the calibration of the classifier can be performed on a different day than the testing. As a consequence, the proposed methodologies can also be used for online

applications, for example to improve the *Human-Machine Interactions* (HMI) by using information derived by the operator's mental workload states.

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

3.3.4. Conclusions

In this study, a passive BCI technique for the assessment of ATCOs mental workload has been proposed and tested in a high-realistic operational environment. Results showed that i) the proposed algorithm maintained a high reliability over time (up to a month), ii) high and significant correlations between the EEG-based and the subjective (*SELF-ISA* and *SME-ISA* scores) workload measures, and iii) in comparison with other machine-learning techniques, the proposed algorithm (asSWLDA) was able to reduce the number of brain features and EEG channels, in other words, to reduce the intrusiveness of the p-BCI system. One of the practical issues, partially addressed in this study, is related to the necessity to calibrate the system every time before using it. In fact, I demonstrated that by using the asSWLDA the measure of the workload will be reliable over a month. However, it might not be always possible to find the right conditions (e.g. EASY and HARD) to calibrate the system in operational environments. Taking into account these limitations, there is the need to perform further experiments to test the possibility to select brain features during the execution of short standard and controlled tasks (that enhance the considered mental state), and then to use such brain features to calibrate the classification algorithm that will be used later in the operative situation.

3.4. Professional ATCOs' Workload Evaluation in Realistic Settings

The aim of this study was to test if the asSWLDA algorithm, the neurometric (frontal theta and parietal alpha EEG rhythms), and the methodology developed in the previous studies for the workload evaluation were able 1) to evaluate the workload levels in real ATM settings, and 2) to highlight the differences between groups of ATC Students and ATC Experts, in terms of workload. In this regard, two groups of professional ATCOs have been recruited to execute the same ATM scenario, and their brain activity has been recorded during the simulation. In order to compare the groups, a workload index has been defined as the ratio between the frontal theta and the parietal alpha PSDs (*ThetaF/AlphaP Index*). Furthermore, the asSWLDA algorithm has been used to compute an *EEG-based Mental Workload Index* (W_{EEG}) able to track the workload during the execution of the ATC task. In addition, both the subjective (ISA) and behavior data (task performance) have been collected to define a more complete profiles of the two groups, thus to better compare them (*Triangulation methodology*, Bekhet and Zauszniewski, 2012).

3.4.1. Experimental Protocol

Subjects

Thirty - seven 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). ATCOs have been asked to take part in a simulated ATC simulation in which the same ATM scenario (Scenario #3, see paragraph 2.3.2) has been proposed.

Neurophysiological Signals Recording and Processing

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), and one electrocardiographic (ECG) channel have been collected simultaneously with a sampling frequency of 256 (Hz). In particular, the ECG electrode has been placed on the Erb's point. In addition, in order to make the system less invasive, differently from the previous experimental sessions, the EOG channel has not been used, but an algorithm able to remove eyes-blink artifacts from the EEG data by means of the Fpz EEG channel has been developed. This algorithm is going to be published soon. All the electrodes (EEG and ECG) have been referenced to both earlobes, and the impedances of the EEG electrodes have been kept below 10 (kΩ). The EEG signals have been digitally band-pass filtered by a 5th order Butterworth filter (low-pass filter cut-off frequency: 30 (Hz), high-pass filter cut-off frequency: 1 (Hz)). The EEG signal has then been segmented in periods of 2 seconds, 0.125 seconds – shifted. These values have been chosen in order to have both a high number of observations (see equation 3.7) in comparison with the number of variables (see equation 3.8), and to respect the condition of stationarity of the EEG signal (Elul, 1969). In fact, the latter one is a necessary hypothesis in order to proceed with the spectral analysis of the signal. After, for each segment the *Power Spectral Density* (PSD) has been estimated by using the *Fast Fourier Transform* (FFT) in the EEG frequency bands involved in the mental workload estimation (theta and alpha bands), defined for each subject by the estimation of the *Individual Alpha Frequency* (IAF) value (Klimesch, 1999). The asSWLDA model has been used to select the most relevant features from the acquired biosignals, to discriminate the mental workload of the ATCOs within the different experimental conditions (e.g. Easy, Medium and Hard). Furthermore, a classification threshold has been estimated by the state classifier, by using the *Receiving Operating Characteristic* (ROC) theory (Bamber, 1975). This threshold have been used during the online assessment of the ATCO's workload. In particular, if the workload index was higher than the threshold, the workload would be classified as "High". On the contrary, the workload would be classified as "Low". The output of this classifier model represents the

neurophysiological measure of the mental workload (W_{EEG}). The asSWLDA has been extensively described in the paragraph 3.1. Regarding the ECG signal, it has not been considered in the analyses because the ATCOs assumed particular body postures and produced high muscular contraction on the chest, thus several artifacts during the execution of task. The main artifacts source has been the ocular activity, that is, the eye-blinks. This kind of artifact has been easily recognized and corrected by using one of the several methods proposed in literature, for example, algorithms based on the regression of an EOG channel (Gratton and Coles, 1983), as done in the previous studies. In this phase of the study, I have defined a new algorithm based on the regression of the *Fpz* EEG channel. Naturally, the *Fpz* channel contains EEG information (the frontal brain sites are very important in the analysis), thus applying regression techniques, important EEG information could be lost. Nevertheless, to prevent the loss of EEG information and, at the same time, to reject the ocular blinks, I have digitally band-pass filtered a copy of the *Fpz* channel (indicated in the following as *Fpz-EOG channel*) by a 5th order Butterworth filter (low-pass filter cut-off frequency: 5 (Hz), high-pass filter cut-off frequency: 1 (Hz)), in order to emphasize the eye-blink signal component, with respect to the signal components related to the brain activity. Then, by means of a threshold criteria, the time windows in which an eye-blink occured have been detected. Only in such windows, the regression of the filtered *Fpz-EOG* channel has been applied to all the EEG channels, by means of the Gratton algorithm (Gratton and Coles, 1983). Despite a possible, but very little loss of EEG information (just in correspondence of the eye-blink), this approach produced a great improvement in terms of system invasiveness and feasibility for real-time applications. For the other artifacts, specific procedures of the EEGLAB toolbox have been used (Delorme and Makeig, 2004). In particular, three methods have been selected, that is *Threshold criteria*, *Trend estimation*, and *Sample-to-sample difference*. In the threshold criteria the EEG epoch are marked as "artifact" if the EEG amplitude is higher than ±100 (μV). In the trend estimation, the EEG epoch is interpolated in order to check the slope of the trend within the considered epoch. If such slope is higher than 3, the considered epoch will be marked as "artifact". The last check calculates the difference between consecutive EEG samples. If such difference, in terms of amplitude, is higher than 25

(μV), it means that an abrupt variation (non - physiological) happened, thus it will be marked as "artifact". At the end, the EEG epochs marked as "artifact" have been removed from the EEG dataset with the aim to have an EEG artifacts-free dataset from which estimate the brain parameters for the different analyses.

Task Performance

During the execution of the simulated ATM scenarios, the 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 (nb) assumed, a global performance index, *Weighted Mean Reaction Time* (WMRT), has then been defined as described in the paragraph 2.4.2.

Power Spectrum Density (PSD) analysis

In order to compare the level of mental workload between the two groups, a proper workload index has been defined as the ratio between the frontal theta (4 – 8 (Hz)) and the parietal alpha (8 – 12 (Hz)) brain activity (*ThetaF/AlphaP Index*). Such EEG rhythms have been demonstrated to be correlated to vary cognitive processes involved in the concept of mental workload. Many experiments have found correlations with mental workload, particularly concerning the enhancement of the frontal theta (Gevins et al., 1998; Klimesch, 1999; Wilson et al., 1999; Yamada, 1998) and suppression of parietal alpha brain rhythms (Pfurtscheller and Klimesch, 1991; Sterman et al., 1994). Different studies have demonstrated that the increment of task difficulty (increased working memory load) is associated with an increase of frontal theta activity and a suppression of parietal alpha activity. In particular, Klimesch (1999) and Klimesch et al. (2005) found that, in an alert brain, EEG oscillations in the alpha and theta band are linked to cognitive and memory performance. Synchronization in the theta band is possibly related to working memory function, lower alpha desynchronization may be reflect attentional processes, while upper alpha may be related to long-term memory function (Klimesch, 1999). According to this author, good performance is characterised by two phenomena: "a tonic increase in alpha but a decrease in theta power, or a large phasic

decrease in alpha but increase in theta, depending on the type of memory demands". "After sustained wakefulness and during the transition from waking to sleeping, upper alpha power decreases, whereas theta increases. Event-related changes indicate that the extent of upper alpha desynchronization is positively correlated with (semantic) long-term memory performance, whereas theta synchronization is positively correlated with the ability to encode new information. The encoding of new information is reflected by theta oscillations in hippocampo-cortical feedback loops, whereas search and retrieval processes in (semantic) long-term memory are reflected by upper alpha oscillations in thalamo-cortical feedback loops." In a research developed to clarify the interaction between learning and task demand, Fairclough et al. (2005) found that the sustained response to a demanding task produceed alpha suppression (state of electrocortical activation) and theta increasing in a multitasking performance activity. Recently, Jaušovec and Jaušovec (2012) investigated the influence of working memory (WM) training on intelligence and brain activity. They found that the influence of WM training on patterns of neuroelectric brain activity was most pronounced in the theta and alpha bands. Theta and lower-1 alpha band synchronization was accompanied by increased lower-2 and upper alpha desynchronization and concluded that WM training increased individuals performance on tests of intelligence. In a recent review, Borghini et al. (2012a) have exposed that there is an increase of EEG power in theta band and a decrease in alpha band with high mental workload.

Workload Index (W_{EEG})

The algorithm for the workload (W_{EEG}) estimation has been extensively explained in the paragraph 3.1. In particular, the asSWLDA has been used to select the most relevant spectral features to discriminate the mental workload of the subjects within the different experimental conditions (Easy, Medium and Hard). As quoted in the paragraph 3.3.1, 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 then 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 triplet (E3, H3) has been chosen to train the classifier, and to test it over the remaining triplets of Easy, Medium and Hard conditions.

Instantaneous Self-Assessment of Workload (ISA)

As explained previously (paragraph 2.4.2), the ISA assessment 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. In order to make comparable the scores of each subject, I previously normalized them along the subjects, by using the *z-score* definition (Zhang et al., 1999).

3.4.2. Results

Task Performance

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. In Figure 49 are reported the results of the analyses 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).

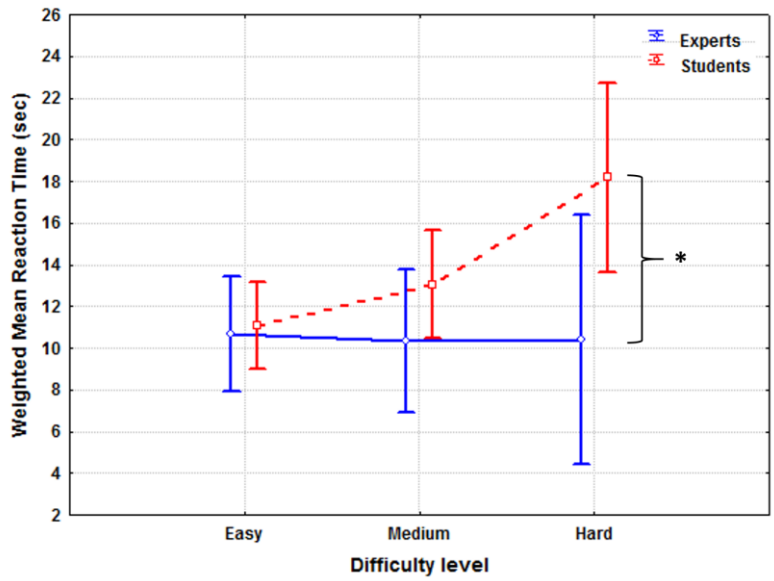

Figure 49. Weighted Mean Reaction Time (WMRT) of the ATCOs across the three difficulty levels of the simulated ATM scenario. The ATC Experts (blue line) did not report any significant differences ($p>0.05$) among the difficulty conditions. On the contrary, the ATC Students (red line) showed a significant ($p<0.05$) decrement of performance between the Hard condition and the others (Easy and Medium).

The statistical analysis between the groups (Figure 50) 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.

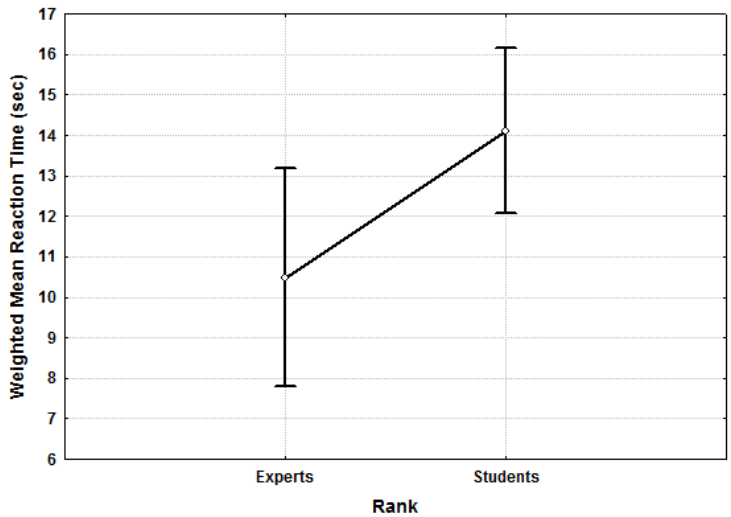

Figure 50. Differences between the ATC Experts and ATC Students in terms of Weighted Mean Reaction Time (WMRT). The ANOVA analysis showed a significant difference ($p=0.037$) between the two groups.

Power Spectrum Density (PSD) analysis

The one-way ANOVA (*between* factor RANK; 2 levels: Experts and Students) has been performed on the workload index (*ThetaF/AlphaP*) to assess the existence of difference between the group of ATC Experts and ATC Students. The results showed that the mental workload of the Students during the whole ATM scenario was significantly higher ($F(1, 35) = 8.8145$; $p = 0.00536$) than the workload of the Experts (Figure 51).

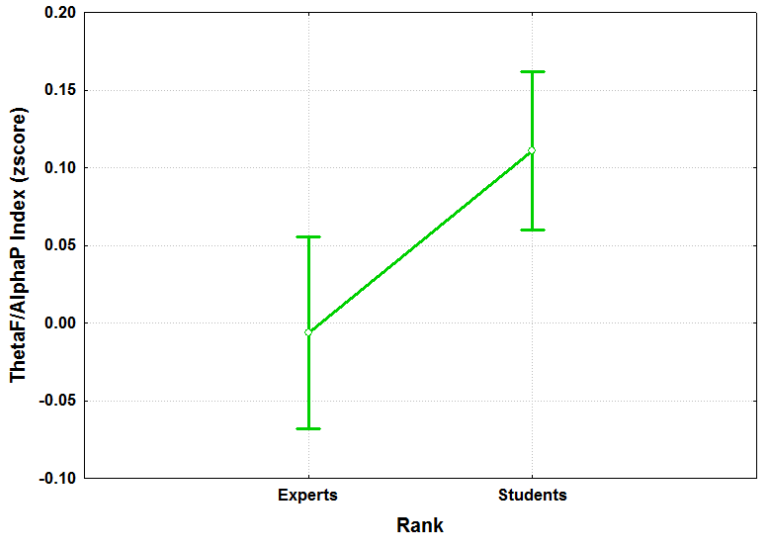

Figure 51. The ANOVA on the *ThetaF/AlphaP Index* reported a significant difference between the group of ATC Experts and Students. In fact, the *ThetaF/AlphaP Index* of the ATC Students was significantly (p = 0.005) higher than the ThetaF/AlphaP Index of the Experts.

EEG-based Mental Workload Index (W_{EEG})

The W_{EEG} index has been evaluated for each difficulty level (Easy, Medium, and Hard) and group (ATC Students and ATC Experts). In Figures 52 and 53, the W_{EEG} of the ATC Experts (on the top) and of the ATC Students (on the bottom) calculated along the experimental session are reported. In particular, the solid blue lines represent the W_{EEG} index for each run (i.e. ~3 minutes average), and the colored rectangles represent the W_{EEG} index averaged over each difficulty level (Easy, Medium and Hard, respectively in green, yellow and red). The figures show that for each group, the W_{EEG} index followed the profile of the ATM scenario.

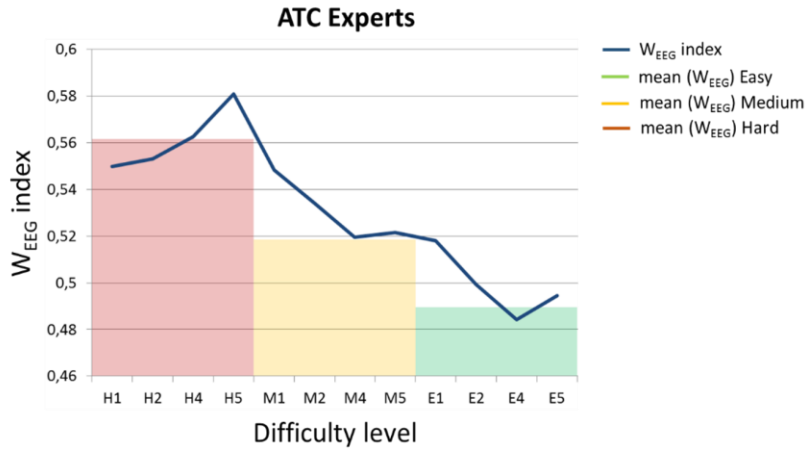

Figure 52. Representation of the W_{EEG} index across the different difficulty condition (E1÷E5 – green bar, M1÷M5 – yellow bar, and H1÷H5 – red bar) and its mean value for the ATC Experts.

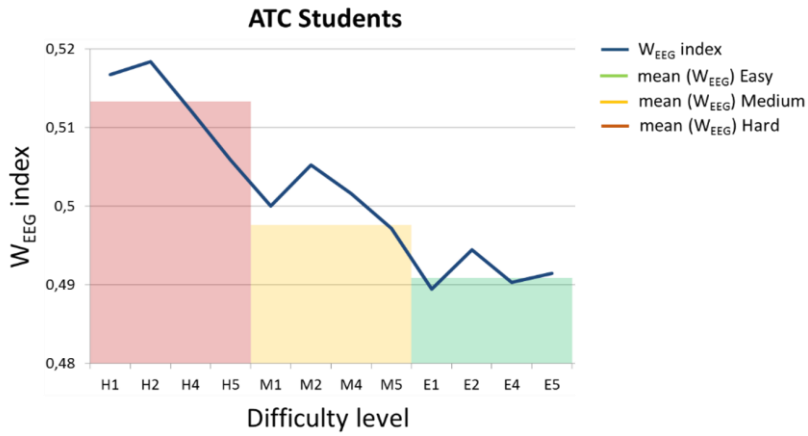

Figure 53. Representation of the W_{EEG} index across the different difficulty condition (E1÷E5 – green bar, M1÷M5 – yellow bar, and H1÷H5 – red bar) and its mean value for the ATC Students.

Furthermore, 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. The Table 6 reports the p-values of the t-tests on the W_{EEG} distributions related to the different difficulty levels. The asSWLDA was able to discriminate significantly the three levels of ATCOs' mental workload.

Table 6: p-values of the two-tailed paired t-tests ($\alpha = 0.05$) between the W_{EEG} index distributions related to couple of conditions (E vs M; M vs H and E vs H).

	Easy vs Medium	**Medium vs Hard**	**Easy vs Hard**
ATC Experts	0.02	0.01	0.003
ATC Students	0.04	0.04	0.003

Instantaneous Self-Assessment of Workload (ISA)

Figures 54 and 55 show, respectively for the ATC Experts and ATC Students, how the ISA scores (z-score normalized) followed the profile of the ATM scenario, except that in the first hard run (H1).

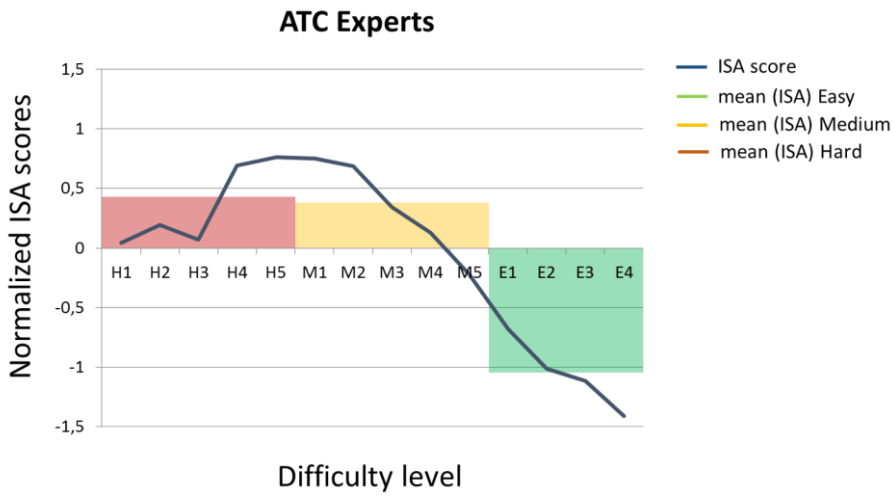

Figure 54. Representation of the ISA score across the different difficulty condition runs (E1÷E5 – green bar, M1÷M5 – yellow bar, and H1÷H5 – red bar) and its mean value for the ATC Experts.

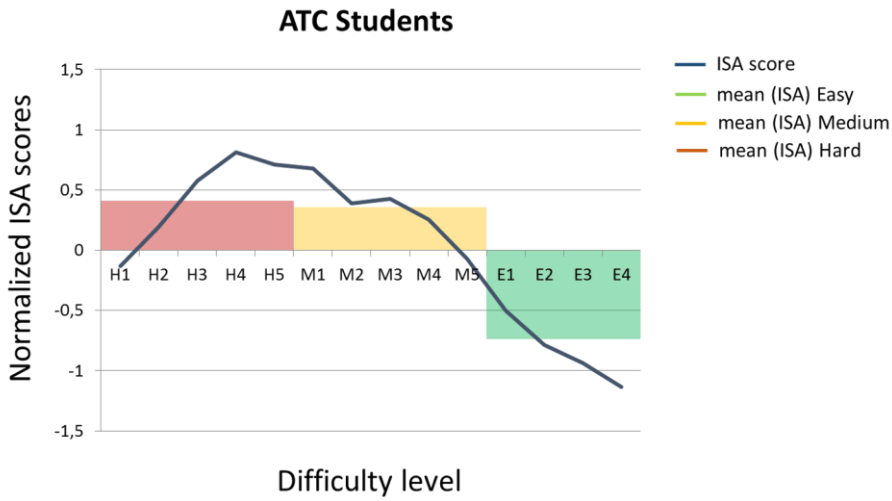

Figure 55. Representation of the ISA score across the different difficulty condition runs and its mean value for the ATC Students.

In order to test the discriminability between the different ATM conditions, 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) have been performed. The Table 7 shows that the ISA scores could differentiate significantly only the couples Hard vs. Easy and Medium vs. Easy ($p < 0.05$). On the contrary, for both the groups the Hard and Medium conditions were not significantly different.

Table 7: p-values of the two tailed paired t-tests ($\alpha = 0.05$) between the ISA scores distributions related to couple of conditions (E vs M; M vs H and E vs H).

	Easy vs Medium	**Medium vs Hard**	**Easy vs Hard**
ATC Experts	$5.2*10^{-9}$	0.854	$8.9*10^{-5}$
ATC Students	$2.6*10^{-5}$	0.593	$7.5*10^{-5}$

Correlation Between Workload and ISA Indexes

The correlation between the subjective (ISA scores) and the neurophysiological (W_{EEG}) workload measures (reported above) has been quantified. Firstly, the W_{EEG} index has been normalized in order to make it comparable with the ISA score. Therefore, the W_{EEG} has been *z-score* normalized along each subject, and then averaged across each run. Table 8

shows 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.

Table 8: 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.

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

Furthermore, the scatterplot in Figure 56 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), in particular, for the whole sample as underlined by the dashed red tendency line.

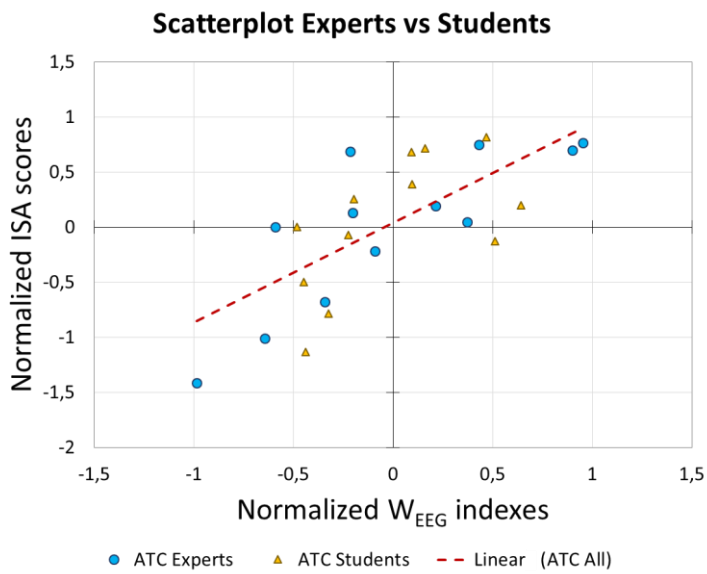

Figure 56. Representation of the correlation between the neurophysiological (W_{EEG}) and subjective (ISA) workload measures for the ATC Experts (blue circles) and ATC Students (okra triangles). The dashed red line is the tendency line of the whole distribution (ATC Experts and Trainees together).

3.4.3. Discussions

The analysis of the task performance, in terms of reaction time, related to the execution of the ATM scenario provided important indications of the two groups. The ATC Experts did not show any difference among the three difficulty levels (Easy, Medium and Hard), whilst the ATC Students reported a significant performance reduction during the Hard scenario (Figure 49). In addition, significant difference has been found between the two groups (Figure 50), as the Experts needed shorter time to react and provide information to the pilots than the ATC Students ($F(1, 31) = 4.7621$; $p = 0.03679$). In terms of mental workload, a synthetic index has been defined with the aim to check the difference between the two groups. The results (Figure 49) showed a significant difference, in terms of mental workload, between the ATC Experts and ATC Students ($F(1, 35) = 8.8145$; $p = 0.005$). Concerning the machine-learning analysis, the asSWLDA was able to track the workload index of the ATCOs (both ATC Experts and Students) along the experimental session (Figures 52 and 53), and to differentiate significantly the three difficulty levels proposed in the ATM scenario (Table 5). It should be underlined that in a realistic environment, different sources of artifacts could affect the recorded neurophysiological signals, and this could produce wrong estimations and misclassifications of the physiological workload measures. ATCOs normally communicate verbally and perform several movements during their operational activity. Despite this, the "speaking" and the other artifacts were not an issue, because the analyses have been performed 1) by frequency filtering the neurophysiological signals, 2) by considering only the EEG channels strictly correlated to the examined cognitive phenomenon, and 3) by using a methodology (the machine-learning approach) able to reject automatically those features that are not important for and related to the observed cognitive phenomenon (i.e. workload). The analyses of the ISA scores confirmed that the three workload levels were perceived differently by the ATCOs. In particular, the Hard and Medium levels were significantly more difficult that the Easy one (Figures 54 and 55), but the ISA values could not discriminate the Medium and Hard conditions, as, on the contrary, the W_{EEG} did for all the workload levels (Table 6). In other words, the resolution of the W_{EEG} index was higher than the ISA measure. A strong correlation (all

$p < 0.05$) between the subjective (ISA) and the neurophysiological (W_{EEG}) workload measures was highlighted (Figure 56). In addition, the results showed that the correlation was significant for both the experimental groups, but it was higher for the ATCO Experts than for the ATC Students (Table 7). The greater expertise and experience of the ATC Experts could explain why such self-report was more reliable than those of the ATC Students. This result could be seen as a further demonstration of the poor suitability of subjective workload measures.

3.4.4. Conclusions

The results, in terms of workload evaluation and correlation between neurophysiological (W_{EEG}) and subjective workload perception (ISA), highlighted the efficiency and reliability of the proposed algorithm and methodology. Both the neurophysiological indexes, *ThetaF/AlphaP* and W_{EEG}, were able to provide useful information regarding the workload levels required by the ATC tasks and about the difference between the groups of ATCO. In addition, the results underlined the higher resolution, thus the higher reliability of the neurophysiological measures than of the subjective questionnaire.

4. Cognitive Training

Inappropriate training might have high social costs, especially for personnel that operates in complex operational environments, such as air-traffic controllers or pilots. One of the current limitations of the standard training assessment procedures is the leak of objective information about the cognitive resources requested by the user for the correct execution of the proposed task across the different training sessions. The training level and related performance are generally evaluated by the supervision of experts and it is easy to understand how this approach is highly operator–dependent. In fact, even if a task is accomplished perfectly, by means of the standard assessment methods will not be possible to estimate information about the user's cognitive capacity and expertise. Therefore, it becomes difficult to quantify the cognitive resources available to the user for accomplish unexpected events or an emergency conditions. Thus, unappropriated training evaluation could bring to high probability to commit mistakes and the consequences might have important social and economic impacts. In the 1928, Spearman (1928) stated that psychological writings "crammed" a lot of allusions to *Human Factor* (HF) in an "incidental manner", but "they hardly arrived at considering such concept systematically and profoundly". The last decades have seen an increasing interest in studying the HF in working environments, especially its causes and possible prevention approaches. For example, in the aviation field, aircraft accident investigations had revealed that 80% of accidents were based on human error, and further investigation indicated that a significant portion of human error was attributable to HF failures primarily associated with inadequate communication and coordination within the crew (Taggart, 1994). Several researches have continued to seek solutions to improve safety, and many of them have identified the personnel training as a significant strategy in achieving such a goal (Barach and Small, 2000; Barach and Weingart, 2004; Hamman, 2004; Leonard et al., 2004; Leonard and Tarrant, 2001). *Training* refers to a systematic approach to learning and development to improve individual, team, and organizational effectiveness (Goldstein and Ford, 2002). The importance and the interest in the concept of training in operational environments (e.g. aviation, hospital and public transport) is

reflected by the regular publication of scientific reviews in the *Annual Review of Psychology* since 1971 (Campbell, 1971; Goldstein, 1980; Wexley, 1984; Tannenbaum and Yukl, 1992; Salas et al., 2001; Aguinis and Kraiger, 2009). An objective quantification of cognitive resources required by the operator during the execution of a task is very important, since the human brain resources are limited. The difference between the available cognitive resources and the amount of those involved for the task execution is called *Cognitive Spare Capacity* (Rudner and Lunner, 2014). The higher the cognitive spare capacity during a normal working activity is (i.e. the operator is involving a low amount of cognitive resources), the greater the operator ability to perform secondary tasks or to react to unexpected - emergency events will be. It is a common experience that at the first time subjects face a task, they attempt to understand its requests in a trial and error fashion; eventually, the subjects gain experience up to the full comprehension of the task that it is quickly followed by an increase of their task-related performance. For example, when we drive a car for the first time, we pay high attention to many details, but after a while, we are able to drive the car much better than at the beginning, and being also able to perform other tasks in parallel with the driving itself. It was showed that the learning is due to the acquisition and to the memorization of solutions to specific situations (Logan, 1988) while individual information of general meaning can be grouped together for creating more specific and larger knowledge by a practice (Newell and Rosenbloom, 1993). In addition, it was sustained that the automation progress is due to two consecutive phenomena: at the beginning, the memorization of the procedures allows to substitute a general knowledge with a specific (procedural) knowledge; later, a decrease of the involvement of the cerebral attention system is observed (Anderson, 1993) since performed procedures will become automatically performed. Such learning process is well described by the *Learning Curve* (Ritter and Schooler, 2001), where the performance enhances and the time necessary for the task execution decreases by increasing the training sessions. A focused and progressive training program makes the people able to gain experience and knowledge for facing and reaching appropriate standard of performance (Parasuraman and McKinley, 2014; Erickson et al., 2007; Gopher 1993; Kramer et al., 1999). In addition, training programs in which performance on a task is methodically improved

through adaptive performance feedback (instructional strategies and individualized program) over multiple sessions have been shown to reliably reduce costs associated with performing multiple concurrent tasks (Kramer et. al, 1995; Kramer, et al., 1999; Schumacher et al., 1999; Glass et al., 2000; Bherer et al., 2006).

In this context, the possibility to objectively assess the training level of the operator, both in terms of task performance and requirement of cognitive activation, appears very useful. Such concern is based on the hypothesis that, during a training period the execution of the task become more automatic and less cognitive resources are required, thus higher amount of cognitive resources will be available. In fact, after having being sufficiently practiced, the long-term memory will contain huge amount of domain-specific knowledge structures that can be described as hierarchically organized schemas that allow us to categorize different problems states and decide the most appropriate solution moves. These schemas can operate under automatic rather than controlled processing (Kotovsky et al., 1985; Schneider and Shiffrin, 1977; Shiffrin and Schneider, 1977). Automatic processing of schemas requires minimal working memory resources and allows problem solving to proceed with minimal effort. In the *Cognitive Load Theory* (CLT), it has been highlighted how the proper allocation of available cognitive resources is essential to learning (Sweller, 1999; Sweller et al., 1998). The effects of practice at a cognitive level must be carefully considered in the interpretation of any study, as the changes in cognitive processes underlying task performance can be a key to understanding the changes in functional activations observed. Many researches have revealed that both developmental change and changes in response to experience can occur at multiple levels of the *Central Nervous System* (CNS), from changes at the molecular or synaptic level, to changes in cortical maps and large-scale neural networks (Kelly and 2005). Across studies, three main patterns of practice-related activation change can be distinguished. Practice may result in an *increase or a decrease in activation* in the brain areas involved in task performance, or it may produce a *functional reorganization* of brain activity, which is a combined pattern of activation increases and decreases across a number of brain areas (Petersen et al., 1998; Poldrack, 2000; Raichle et al., 1994; Poldrack and Gabrieli, 2001). *Increased neural efficiency* may correspond to a sharpening of the response in a particular

neural network so that only a minority of neurons now fire strongly in response to a particular task or stimulus (Poldrack, 2000). Practice-related activation decreases are therefore the result of more efficient use of specific 'neuronal circuits' (Petersen et al., 1998). In topographically organized cortex, this will be seen as a reduction in the spatial extent of activation, or, in areas with distributed representations, as a reduction in activation (Poldrack, 2000). The term *increases in activation* refers to both practice-related expansions in cortical representations and increases in the strength of activations. Expanded representations may be reflected in increased strength of activations (Poldrack, 2000). On a neural level, increases in activation are suggested to reflect recruitment of additional cortical units with practice, seen in topographically organized cortex as an increase in the spatial extent of activation, or a strengthening of response within a region, observed as an increase in the level of activity within that region (Poldrack, 2000). Practice-related reorganization of the functional anatomy of task performance may also be distinguished into two types, one constituting a *redistribution*, the other a *'true' reorganization*. The first might be considered as a pseudo-reorganization, or redistribution of functional activations. It constitutes a combination of increases and decreases in activation such that the task activation map generally contains the same areas at the end as at the beginning of practice, but the levels of activation within those areas have changed. The functional anatomy of the task therefore remains basically the same but the contribution of specific areas to task performance changes as a result of practice. It has been discussed previously by Petersen et al. (1998) in terms of a 'scaffolding storage' framework. According to this framework, a 'scaffolding' set of regions is used to cope with novel demands during unskilled, effortful performance. After practice, processes or associations are more efficiently stored and accessed and the scaffolding network falls away, evinced by decreased activation in those 'scaffolding' attentional and control areas. Activations seen earlier in practice therefore involve generic attentional and control areas — *prefrontal cortex* (PFC), *anterior cingulate cortex* (ACC) *and posterior parietal cortex* (PPC) are the main areas considered to perform the 'scaffolding' role, consistent with theories of Shadmehr and Holcomb (1997). The PFC, which has been previously implicated in goal-based and planning processes (e.g., (Shallice, 1982), is most readily associated with

‘‘goal processor’’, whose primary function is to guide novice behavior by instantiating the task context (e.g., the instructions) and then structuring the sequential execution of task-relevant operations (Chein et al., 2005). Correspondingly, Cohen and colleagues (O’Reilly et al., 1999; Cohen et al., 1996; Braver and Cohen, 2001) have argued that the prefrontal cortex, in particular the dorsolateral part, serves to maintain goal-related information that biases subsequent processing according to task demands. By this account, practice-related decreases observed in the *Dorsolateral Prefrontal Cortex* (DLPFC) may occur as practice strengthens the stimulus–response associations between paired items, correct responses are potentiated, and the need for monitoring and biasing from task context (and thus the prefrontal cortex) is reduced. From a working memory perspective, the activity in these regions during encoding may reflect material non-specific rehearsal processes to maintain information on-line, while it is being more permanently embedded into long-term memory (Buckner, 1996).

The empirically observed domain-general network also included an extensive PPC where activity decreased with practice. This region, which is widely viewed as an attentional center of the brain (Vandenberghe and Gillebert, 2013; Vandenberghe et al., 2012; Corbetta et al., 1995; Desimone, 1992; Posner and Petersen, 1990), subserves a control process that allocates attention to task-specific information processing regions. Such attentional modulation is imposed by an ‘‘attention controller’’ that guides information flow by regulating the output of distributed information processing modules during novice performance.

Practice also led to decreased activation of the ACC for both verbal and nonverbal paired associates. The function of this region corresponds to an ‘‘activity monitor’’ in control system. The Activity Monitor acts to track and record recent activity levels in task- relevant areas and to establish a decision threshold for identifying strongly activated areas to the goal processor. Indeed, much of the recent literature on ACC function has emphasized its role in communicating a need for greater cognitive control to other executive brain regions (Carter at al., 1999). Accordingly, this region’s decrease in activation with practice may correspond with the decreased sampling time needed to select a decision threshold as familiarity with the correct pairings increases.

The second type of functional reorganization is considered to reflect a *'true' reorganization* of activations. It is observed as a change in the location of activations and is associated with a shift in the cognitive processes underlying task performance. This shift ('process switching': see Poldrack 2000; Dayan and Cohen, 2011) means that neuro-biologically and cognitively, different tasks are being performed at the beginning and end of practice, resulting in a coordinated increase and decrease of activation in separate brain regions, and is consistent with studies demonstrating a reorganization of activation as a result of explicit shifts or differences in task strategy (e.g. Sakai et al. 1998; Ashby et al., 2010; Doyon and Benali, 2005 Bernstein et al., 2002; Glabus et al., 2003; Hikosaka et al., 2002). The results of these studies are also consistent with those of Seidler et al. (2002) who saw learning related activations in premotor, motor, prefrontal, cingulate and parietal cortices, and in the thalamus, during practice on a serial reaction time (SRT) task. Activation in these areas was observed to dissipate with practice even though a distractor task prevented the behavioral expression of learning. On the other hand, cerebellar activation increased during the expression of performance improvements, once the distractor had been removed. Hikosaka et al. (2002) have also demonstrated that automatic execution of a motor adaptation task produces long-term plastic changes in the *Cortico Cerebellar* (CC) instead of the *Cortico Striatal* (CS) system. In this study, subjects were scanned on the first day of practice (early learning) and after they reached automatic performance (i.e. after 21 days of practice on average) on a joystick, target reaching task. Automatic execution was assessed by verifying that subjects had reached complete asymptotic performance and by testing their performance in a dual-task condition. Although activations in the *putamen* and other motor areas were observed in the early learning phase, increased activity in the *cerebellum* and parietal cortex were seen after training, suggesting that the *cerebellum* and associated cortical regions are sufficient to mediate automatic adapted movements. Both these studies demonstrated the 'scaffolding' role of frontal and parietal areas, activation which is subject to 'pruning' as performance improves and becomes increasingly reliant on task-specific, performance-related areas located posteriorly in the brain. Furthermore, they demonstrated how learning-related changes in activation

may occur in the absence of performance change, emphasizing the complexity of the brain's response to repeated task experience.

As quoted above, the literature dealing with the effect of practice on the functional anatomy of task performance is extensive and complex, comprising a wide range of papers from disparate research perspectives. An interesting question is then if it is possible to follow such changes of the cerebral activity during the training, using neuroimaging methodologies. This capability would provide additional and objective evidences related to the acquisition of an ability during the training phase, beside the overt behavioural measurements. The neuroimaging methodologies would be able to track reallocation of such cognitive resources during the training phase, providing an independent metric on the quality of the performed task. Neuroimaging methods such as *Positron Emission Tomography* (PET) and *functional Magnetic Resonance Imaging* (fMRI) are excellent tools in this endeavor, enabling the examination of how the brain adapts itself in response to practice or repeated exposure to particular tasks. However, their limitations in terms of cost, space and invasiveness make them not suitable for real working environment settings, where a less invasive approach would be preferable and the costs for its implementation and usage has to be limited. In fact, PET and fMRI techniques require expensive devices, big customized rooms and high maintenance costs. In order to deal with these requirements and limitations, it has been investigated the possibility to track the training progress of the operator, by recording its brain activity by using the *Electroencephalography* (EEG) technology, both in controlled environments (Labs) and in realistic operational settings.

4.1. Training Assessment in Lab Settings

The aim of this study was to derive robust neurophysiological signature related to the learning progress of a motor-cognitive task, by using biosignals such as *Electroencephalogram* (EEG), *Electrooculogram* (EOG) and the *Electrocardiogram* (ECG).

All these variables have been reported (see paragraph 1) to be correlated with the levels of mental effort generated during the execution of a task (Elliot et al., 2011; Fukuda, 2001; Palomba et al., 1997; Shultz et al., 2011). In particular, it has been suggested that an increased *Heart Rate* (HR) could be related with an increased mental effort, while *Eye Blinks Rate* (ERB) are inversely correlated with the increase of the mental effort (Lal and Craig, 2002). Finally, it has also been shown that one of the most prominent neurophysiological event linked to the increase of the cognitive effort and decision making, in a strategy selection process during a complex task, is the increase of the EEG *Power Spectral Density* (PSD) in the theta frequency band (Berka et al., 2007; Berka, 2011; Galán and Beal, 2012; Jaušovec and Jaušovec, 2012; Luu and Tucker, 2001; Mackie et al., 2013; "Fielder, 2011"; Ridderinkhof et al., 2004), and the corresponding decrease of the EEG PSD in the alpha frequency band (Borghini et al., 2012a, b; Doppelmayr et al., 2008; Luu and Tucker, 2001, Kelly et al, 2003; Klimesch, 2012, 1999; Klimesch et al., 1998; Rihs et al., 2007; Sadaghiani et al., 2010). Such variations typically occur over the prefrontal and frontal areas, for the theta rhythm, and over the parietal areas, for the alpha rhythm.

The hypothesis of the study was that at the first time subjects face the task (first training session), they attempt to understand its requests in a trial and error fashion, and the neurophysiological parameters would show certain changes. Eventually, the subjects gain experience up to the full comprehension of the task that it is followed by a significant increase of their task-related performance, thus significant variations of the neurophysiological parameters (e.g. frontal theta and parietal alpha PSDs, HR and EBR). Once trained, the subjects will not show further performance improvement (saturation of the learning curve) and, from a cognitive perspective, they will report less mental effort in coping with the proposed task and reaching high performance levels.

Such trends could be taken as indexes of the correct acquisition of skills across the training sessions. Such hypotheses have been tested on a group of 10 healthy students while they have been daily learned to perform a motor-cognitive task that they had never practiced before, along a training period of a week (5 consecutive days).

4.1.1. 4.1.1 Experimental Setup

Subjects

Ten healthy volunteers (students of the *National University of Singapore* - NUS) have given their informed consent for taking part at the experiment and each of them has been paid SG$200 for the entire experimental period. The study protocol has been approved by the local Ethics Committee. The selection of the subjects has been done accurately in order to ensure the same mental and physical state (homogeneity of the experimental sample). The subjects (25±3 years old) have been instructed to maintain a specific kind of lifestyle. In particular, they have been asked to avoid alcohol, caffeine, heavy meals right before the experiments, and to avoid extreme physical activity over the entire experimental protocol (homogeneity of the "internal conditions" of the subjects during the experiments). The Lab environment has been kept under control (lights intensity, room temperature, seat position) across the different days of the experiments (homogeneity of the "external conditions" during the experiment). In addition, in order to have low sources of variances, the experimental group has been composed only by males.

Experimental Protocol

The subjects performed the NASA *Multi Attribute Task Battery* (MATB, see paragraph 2.3.1 for more details) for five consecutive days (SESSIONS T1÷T5), for a total training period of 30 minutes per day. In order to investigate possible trends and changes of the neurophysiological parameters across the experimental sessions, the EEG, the ECG and the EOG have been recorded in the first (T1), in the third (T3), and in the fifth (T5) training session, while the behavioural and subjective (NASA-TLX) data have been collected after each training condition. Instructions have been provided to each subject on the first day of training (T1), since the

subjects had never performed the proposed task before. To be efficient, the instructional design has been tailored to the level of the subject, in order to avoid that the effectiveness of the training was likely random (Kalyuga et al., 2003). At the end of each experimental condition of any training session, the subjects filled the NASA-TLX questionnaire. Yet, I considered the factors "Frustration" of the NASA-TLX as indicator of the *Emotive Engagement*. In fact, this factor is defined as *how much the subject felt insecure, discouraged, stressed and annoyed versus secure, gratified, confident and relaxed with the task* (Hart S. G. and L. E. Staveland, 1988).

Neurophysiologicaldata Analysis

Electroencephalogram (EEG) and physiological signals, vertical electrooculogram (EOG) and electrocardiogram (ECG) have been recorded by a digital monitoring system (ANT Waveguard system). The 64 EEG channels, the ECG and the EOG channels have been collected simultaneously during the experiment with a sampling frequency of 256 (Hz). All the EEG electrodes have been referenced to both earlobes, grounded to the AFz channel and their impedances were kept below 10 (kΩ). The bipolar electrodes for the heart activity have been placed on the left pectoral muscle, while the bipolar electrodes for the EOG have been positioned vertically over the left eye. The acquired EEG signals have been digitally band-pass filtered by a 4th order Butterworth filter (low-pass filter cut-off frequency: 30 (Hz), high-pass filter cut-off frequency: 1 (Hz)) and the *Independent Component Analysis* (ICA) (Hyvärinen and Oja, 2000) has been run in order to remove eye-blinks and eyes-gaze artifacts from the EEG data. For other source artifacts, a specific procedure for artifact removal, based on the approach involving the Riemman geometry theory has then been applied (Barachant et al., 2013a,b). The EEG *Power Spectral Density* (PSD) has then been estimated by using the periodogram with Hanning window (2 second-length, overlapped of 1 second). These values have been chosen in order to have both a high number of observations (see equation 3.7) for a proper statistical analysis, and to respect the condition of stationarity of the EEG signal (Elul, 1969). In fact, the latter one is necessary to proceed with the spectral analysis of the signal. The EEG frequency bands have been defined for each subject by the estimation of the *Individual Alpha* Frequency (IAF) value (Klimesch, 1999). In order to have a better

estimation of the alpha peak and, hence of the IAF, the subjects have been asked to keep the eyes closed for a minute. Then, this condition (OC condition) has been used for the IAF estimation and then for the EEG bands definition. The *Heart Rate* (HR) and the *Eyes Blink Rat*e (EBR) have been estimated by calculating the distance between consecutive peaks occurring in the ECG and in the EOG signals. In particular, the R-peaks and the Blink-peaks have been used for the HR and EBR estimation, respectively. All the analyses of the neurophysiological parameters, the EEG PSDs estimated in the different frequency bands, the HR and the EBR have been done by estimating the signed r-square (see paragraph 3.2.1) between the considered experimental condition and the reference one. Such reference conditions consisted in looking at a black screen for 2 minutes before starting with the execution of the different task difficulties. The results proposed in the following sections will then be represent as r-square values, thus if the r-square value (y-axis) is zero, it will mean that there were no differences between the considered experimental condition with respect to the reference one.

Estimation of Cortical Source Current Density

Cortical activity has been estimated from EEG scalp recordings by employing the high-resolution EEG technologies (Astolfi et al., 2007, 2006) with the use of the average head model from McGill University (Ding et al., 2005). Electrode positions over the scalp have been obtained individually for each subject by using the *Photomodeler software* (Eos System Inc). The cortical model consisted of about 8,000 dipoles uniformly disposed on the cortical surface. The estimation of the current density strength for each dipole has been obtained by solving the *electromagnetic linear inverse problem* (He et al., 2006; Nunez, 1995). In particular, the solution of the linear system,

$$\mathbf{Ax} = \mathbf{b} + \mathbf{n} \tag{4.2}$$

at a particular time instant *t* provides an estimation of the dipole source configuration $\boldsymbol{x}$ at time *t* that generates the measured EEG potential distribution **b** in the same instant. The system also includes the measurement noise **n**, assumed to be normally distributed. **A** is the lead field matrix,

where each *j-th* column describes the potential distribution generated on the scalp electrodes by the *j-th* unitary dipole. The current density solution vector of Equation (4.1) has been obtained as:

$$\xi = \arg\min_{x}\left(\left\|\mathbf{A}x - b\right\|_{\mathbf{M}}^{2} + \lambda^{2}\left\|x\right\|_{\mathbf{N}}^{2}\right) \tag{4.3}$$

where **M**, **N** are the matrices associated to the metrics of the data and of the source space, respectively, λ is the regularization parameter and $||\ x\ ||_N$ represents the *N* norm of the vector *x* (Babiloni et al., 2005; Grave de Peralta Menendez and Andino, 1999). The solution of Equation (4.2) is given by the inverse operator ***G***:

$$\boldsymbol{\xi}(t) = \mathbf{G}\mathbf{b}(t), \text{ where } \mathbf{G} = \mathbf{N}^{-1}\mathbf{A}'\left(\mathbf{A}\mathbf{N}^{-1}\mathbf{A}' + \lambda\mathbf{M}^{-1}\right)^{-1} \tag{4.4}$$

The optimal determination of the regularization term λ of this linear system has been obtained by the *L-curve approach* (Hansen, 1992; He et al., 2006). As a metric in the data space, the identity matrix has been used, while as a norm in the source space, the following metric was adopted:

$$\left(\mathbf{N}^{-1}\right)_{ii} = \left\|\mathbf{A}_{\cdot i}\right\|^{-2} \tag{4.5}$$

where $(\mathbf{N}^{-1})_{ii}$ is the *i-th* element of the inverse of the diagonal matrix **N** and all the other matrix elements N_{ij} are set to 0. The L_2 norm of the *i-th* column of the lead field matrix **A** is denoted by $||A_i||$. Using the relations described above, an estimation of the signed magnitude of the dipolar moment for each cortical dipoles has been obtained for each time point *t*. As the orientation of the dipole has been defined to be perpendicular to the local cortical surface in the head model, the estimation process returned a scalar rather than a vector field. The spatial average of the signed magnitude, of all the dipoles belonging to a particular ROI at each time sample, was used to estimate the waveforms of cortical ROI activity in that ROI, indicated as **ρ**(t) to highlight their time-dependence. Spatial averaging can be expressed in terms of matrix multiplication by a matrix **T**. This matrix is sparse and has as many rows as ROIs, and as many columns as the number of dipole sources.

ROI cortical current density waveforms can then be expressed as:

$$\boldsymbol{\rho}(t) = \mathbf{Tx}(t) = \mathbf{TGb}(t) = \mathbf{G}_{ROI}\mathbf{b}(t) \text{ , where } \mathbf{G}_{ROI} = \mathbf{TG} \quad (4.5)$$

where $\mathbf{b}(t)$ is the array of the waveforms recorded from the scalp electrodes and $\mathbf{x}(t)$ is the array of the cortical current density waveforms estimated at the cortical surface. The $\mathbf{G}_{ROI}$ matrix only depends on geometrical factors, and can thus be computed and stored off-line. The matrix multiplication can be interpreted as a spatial filtering of the scalp potential $\mathbf{b}(t)$, using the elements of $\mathbf{G}_{ROI}$ as weights. In this way, we could obtain time-varying waveforms at the level of different cortical areas. In the next paragraph, such cortical areas will be described as coincident with particular Brodmann areas for all the subjects involved in the present study.

Generation of the Regions of Interest (ROIs)

Cortical *regions of interest* (ROIs) have been drawn on the computer-based reconstruction of the average cortical model. The ROIs were segmented automatically on the basis of the *Brodmann areas* (BA) classification (Brodmann, 2006). Particular ROIs considered in this study were those suggested by the previous literature relative to the mechanism of the "decision-making" processes, then involving *dorsolateral prefrontal cortical* (DLPFC) and parietal regions, as well as the cingulate cortices (*Anterior Cingulate Cortex*; ACC and the *Cingulate Motor Area*; CMA). The DLPFC regions including the BAs 8, 9/46 and 10 have been then selected, together with the ACC defined by comprising the BAs 24, 25 and 33. In particular, in the following it will be presented in detail the variation of the PSD estimated over the frontal cortical areas (composed by the BA6, BA8, BA9 and BA46), in the theta frequency bands, and of the PSD for the parietal cortical areas (composed by the BA5, BA7, BA19 and BA40), in the alpha frequency band, by solving the *linear inverse problem*. Such focus of interest is justified by the involvement of these cortical regions in the insurgence mental effort (reviewed in Borghini et al., 2012a). In order to generate a cortical distribution of the statistically significant increase or decrease of EEG PSD across the training sessions (T1, T3 and T5), series of univariate t-tests, for each considered voxel of the adopted cortical model, have been performed at a nominal level of significance of $\alpha = 0.05$.

However, since the execution of multiple univariate tests could raise the Type I errors (the so called alpha-inflaction; Zar, 1999), the *False Discovery Rate* (FDR) methodology was adopted to correct for the multiple comparisons (Strimmer, 2008).

Statistical Analysis

Repeated measures ANOVAs (Confidence Interval, CI = 0.95) have been used for the statistical validation of the results, derived by the different analyses, by using the STATISTICA software (Statsoft). In particular, seven one-way repeated measures ANOVAs have been performed separately for frontal theta PSD, parietal alpha PSD, Heart Rate, Eyeblinks Rate, Performance, NASA-TLX total score and "Frustration" score, as independent variables, with the *within* factor SESSION at 3 levels (T1, T3 and T5 - the 3 recording-training sessions). In addition, three two-way repeated-measures ANOVAs, with the *within* factors SESSION (3 levels) and SUBJECTS (10 levels) have been performed for the single-subjects analyses where the frontal theta PSD, parietal alpha PSD and Performance were the independent variables. Yet, a two-way repeated measures ANOVA, with the factors SESSION (3 levels) and CONDITIONS (4 levels) has been performed on the NASA-TLX total score (independent variable) throughout the training sessions and task conditions. Duncan post-hoc tests have been performed to highlight the effects between all the employed factors and levels.

4.1.2. Results

Tasks Performance

Throughout the training sessions, the performance of the subjects increased continuously in terms of mean performance level and accuracy. Figure 57 shows the simultaneous increase of the performances level and the slight decrease of the amplitude of the standard deviations across the training sessions. Since the second day of training, all the subjects reached a good level of performance (above the 90% of correct responses). The repeated measure ANOVA performed on the Global MATB Performance index showed a statistical significant difference among the sessions ($F(4, 40) = 8.1356$; $p = 7 \bullet 10^{-5}$). The post-hoc Duncan test showed that the first session was statistically different from all the others ($p < 10^{-3}$). On the contrary, no difference was found between T4 and T5. Overall, the average final performance of the experimental group was significantly improved (92%) with respect to the beginning of the training (88%).

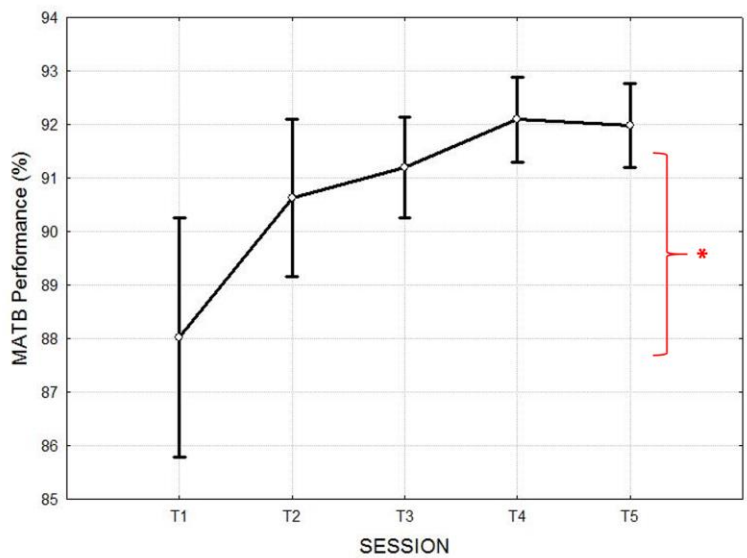

Figure 57. MATB performance index across the five different training sessions (T1÷T5) of the experimental population. A significant increase was obtained since the third day (T3) of training when compared to the first (T1) and second (T2) day. At the end of the training period (T5), the learning curve reached the saturation point, as there was no statistical difference among the third (T3) and fifth (T5) training session, in terms of performance level.

Perceived Workload: NASA-TLX Scores

In Figure 58, the total scores of the subjective workload evaluation (NASA-TLX) across the different sessions (different coloured style lines) as well as across the different difficulty levels (*x-axis*) are presented. Each single task condition (hyper-easy, easy, medium and hard) has been perceived easier after any training session, with respect to the previous one (coloured lines), until the last one (black line), in which all the conditions were perceived as the easiest ones. The repeated measure ANOVA showed significant results among the training sessions ($F(4, 56) = 19.1$; $p < 10^{-5}$). The post-hoc test showed that the average of the NASA-TLX scores at T1 was statistically different from all the others and the average NASA-TLX score at T5 was statistically lower than all the first three sessions (T1, T2 and T3); NASA-TLX scores at T4 and T5 were not significantly different from each other.

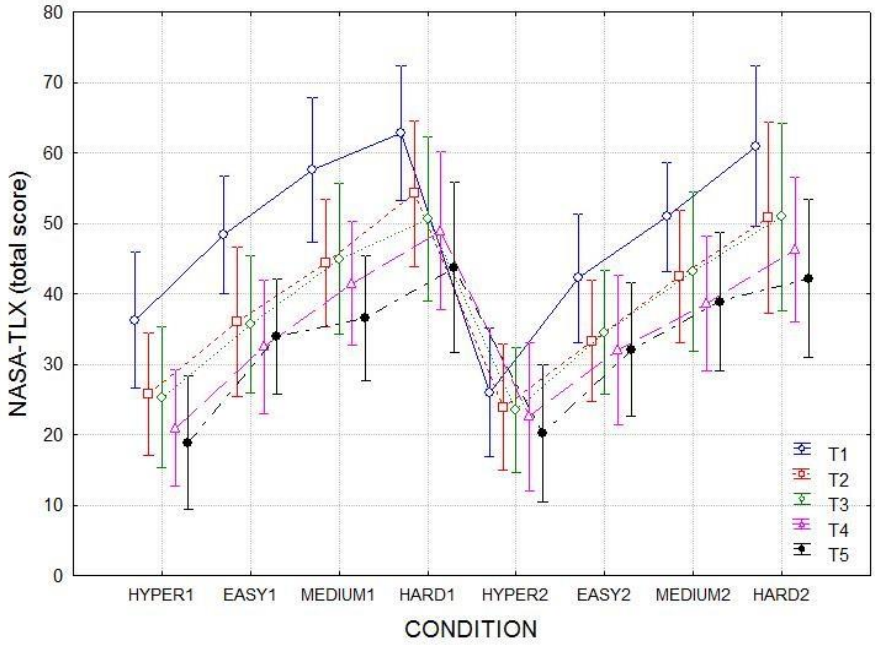

Figure 58. NASA–TLX total scores obtained at the end of each training condition, *x-axis*, (hyper-easy, easy, medium and hard) in the five training sessions (different colour and shape lines). The single task condition was perceived easier after each training session up to the last one (T5). In fact, the NASA-TLX scores of the first session (T1) were the highest (blue colour – solid line), while those of the last training session (T5) were the lowest (black colour – chain line).

Frontal Theta PSD

Figure 59 reports the averaged r-square values of the EEG theta PSD estimated over the frontal channels (EEG channels: AF7, AF3, AF8, AF4, F7, F5, F3, F1, Fz, F2, F4, F6 and F8). The results from the repeated measure ANOVA showed a statistical significant modulation of the signed r-square of EEG PSD in theta band (Figure 53) over the frontal areas across the different training sessions ($F(2, 18) = 16.21$; $p < 10^{-5}$).

Parietal Alpha PSD

In the Figure 60, the trends of the signed r-square of the parietal EEG PSD in alpha frequency band estimated during the experimental sessions is showed. In particular, the figure shows the trend of the parietal EEG PSD in alpha band over the parietal scalp EEG channels (CP1, CP3, CP5, P1, P3, P5, P7, CP2, CP4, CP6, P2, P4 and P6). The repeated measure ANOVA showed significant ($F(2, 18) = 10.39$, $p < 10^{-3}$) differences across the training sessions.

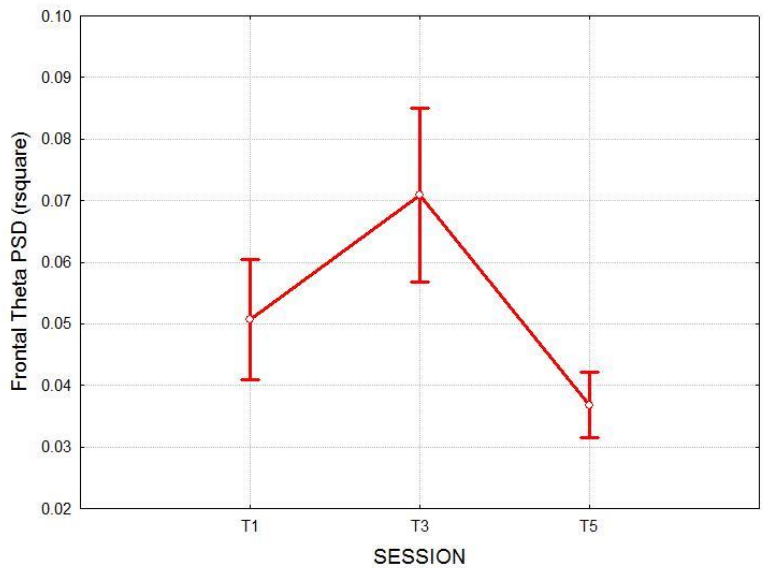

Figure 59. Frontal EEG PSD (r-square) in theta band over all the frontal EEG channels. The repeated measures ANOVA returned a significant value of the factor SESSION ($p < 10^{-5}$). The frontal theta PSD peaks at the T3, and then drops down at T5, with a value lower than T1.

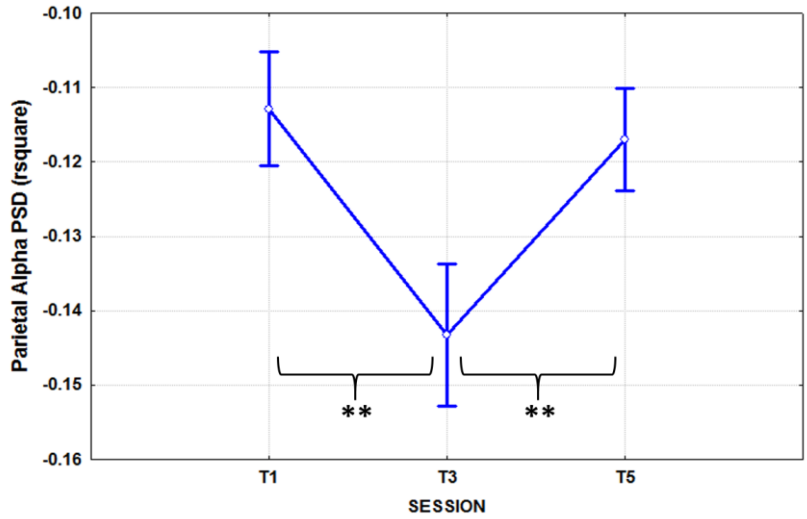

Figure 60. Parietal EEG PSD (r-square) in alpha frequency band during the training period. The Parietal Alpha PSD has a V-trend, similar to those reported for the theta band, but with an opposite sign. In fact, in the last training session (T5) it decreased, but less than in the central session (T3) and with a mean value close to that in the first session (T1).

Single-Subject PSD and Performance Analysis

The analysis of the learning curve at the single subject level reveal a similar behaviour of the whole group (Figure 57). Two of the 10 subjects (circled in orange in Figure 61) showed a lower starting point, in terms of mean performance value, when compared to the others. Subject GUOZHA (green line) and WANJAS (purple line) gained more than the 4% after the T2, and about the 10% in T3, reaching the group since the central training session. These subjects started the training period with a mean performances of 77% and 83% and reached a level of 88% and 92%, respectively, on the third day (T3).

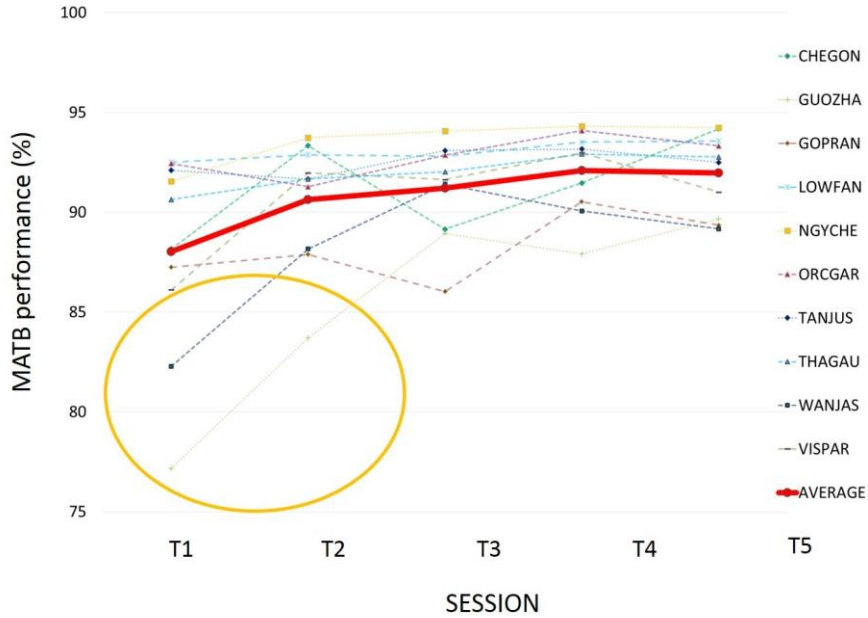

Figure 61. Single-subject MATB mean performance. All the experimental group completed correctly the training period as there were no statistical differences among the last training sessions. By this analysis is possible to assert that a couple of subjects (GUOZHA and VISPAR) started their training period with a mean performance level lower than the rest of the group (about 77% and 83%, respectively) but they increased their performance quite quickly and reached the group since the third day of training (T3).

The mean value of the frontal theta PSD showed a clear trend across the training sessions. Figure 62a reports the frontal EEG PSD in theta band for each subject in the first (T1), the central (T3) and final (T5) training session. Almost all the subjects showed the trend described by the average in Figure 57. In agreement with the different learning trends described in Figure 57, the two subjects who previously had the lowest performance starting points (GUOZHA and WANJAS) showed the highest frontal theta PSD values (light green and dark blue), of the considered experimental group, in particular for the central session (T3) of training (circled in orange in Figure 60a).

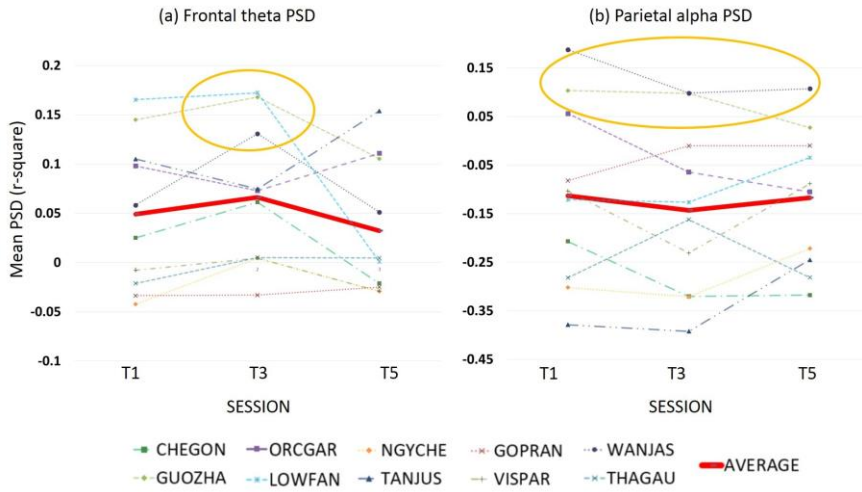

Figure 62. On the panel a) are presented the mean frontal theta PSD values, while in the panle b) are presented the mean parietal alpha PSD values estimated for each subject in the three recording-training sessions (T1, T3 and T5). In the orange circles, the two subjects who showed a slow performance improvement have been highlighted.

Figure 62b presents the individual trends of the variations of the parietal EEG PSD in the alpha band for each subject of the considered experimental group. The majority of the subjects showed the trend suggested at the beginning of the study, that is the parietal alpha PSD reached the lowest value in the central training session (T3). Three subjects out of 10 showed a continuous decrease of the parietal alpha PSD across the training sessions. These trends could be due to the fact that these subjects had to keep more attention during the execution of the task respect to the others, which probably had automatized many processes for the correct execution of the task. Two subjects (GUOZHA and WANJAS), considered previously, were the only ones for which the parietal alpha PSDs did not decrease with respect to the reference condition (circled in orange in Figure 62b).

Cortical Maps

Figure 63 shows how the cortical PSD, estimated for the entire population, varied across the different training sessions: the first row is relative to the cortical PSD in the theta band, while the second row is relative to the alpha band. No statistical differences between the training session (T1, T3 or T5) with respect to the reference condition are represented by gray voxels. On the contrary, a red colour is used to represent voxels which presented

significant increase ($p < 0.05$) of EEG PSDs, when compared to the reference condition, and in blue colour for the significant decrement ($p < 0.05$). Circles on the first row of the Figure 63 highlight the frontal areas with the statistical increase of EEG PSD in the theta band (represented as a red color). The cerebral activity within the frontal areas reached the maximum at T3 and then decrease slightly at T5 when compared to the T1 condition. The yellow circles in the second row of Figure 61 highlight the parietal areas in which significant decreases (represented as blue color) of the alpha PSD were found across the sessions. The highest desynchronization happened on the central session of the training (T3), in correspondence of the highest increment of the frontal theta PSD.

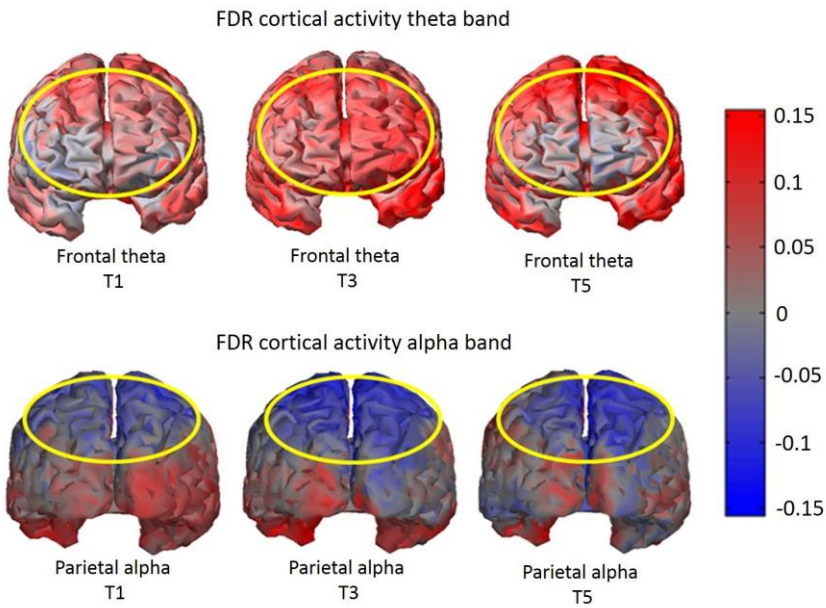

Figure 63. Cortical maps (frontal view) of the EEG PSD for theta band (top) and of the EEG PSD in alpha band (bottom). Only the FDR corrected significant t-values are plotted in colour. It is possible to note a trend of the EEG PSD changes throughout the successive training sessions analyzed (T1, T3, and T5). The red colour means that the EEG PSD estimated on the cortical surface during the task performed at a particular training day was higher than the EEG PSD estimated during the reference conditions(vice versa for blue colour).

The quantitative analysis of the increase or decrease of the frontal theta and parietal alpha EEG PSDs have been investigated also by focusing on the analysis over particular ROIs identified by the Brodmann areas (BAs). The

frontal channels have been modelled with the cortical voxels that belong to the BA6, BA8, BA9 and BA46, whereas the parietal channels have been modelled with those belong to the BA5, BA7, BA19 and BA40. In these areas, the number of cortical dipoles showing statistically significant difference during the different training sessions with respect to the reference condition has been estimated. The results reported in Figures 64 and 65 confirmed the analysis of the scalp EEG and are helpful for the interpretation of the previous results. In particular, over the frontal BA areas, the theta PSD (Figure 64) had the highest peak in the number of cortical activated voxels in the third training session (T3), while such number decreased in the last session (T5). Over the parietal areas, the desynchronization of the alpha rhythm (Figure 65) reached the maximum number of the statistically significant activated voxels in T3 session, while such number was lower than T3 in the T5 training session.

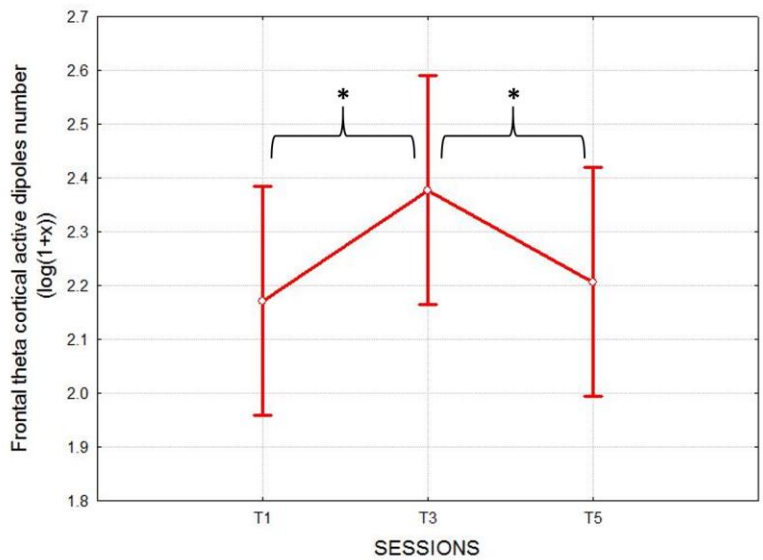

Figure 64. Frontal Brodmann ROIs quantitative analysis related to the numbers of statistically significant cortical voxels elicited by the tasks, when compared to the reference condition, across the training sessions (T1, T3 and T5) for the theta frequency band.

The repeated measure ANOVAs performed on the number of activated voxels in the frontal and parietal areas, and the Duncan's post-hoc tests showed how in correspondence of T3 the number of dipoles was the highest across all the training period, both in the theta ($F(2, 18) = 9.97$; $p = 0.001$) and in the alpha bands ($F(2, 18) = 4.48$; $p = 0.02$). These results confirmed also the trends, found by the EEG scalp data analyses, in Figures 59 and 60.

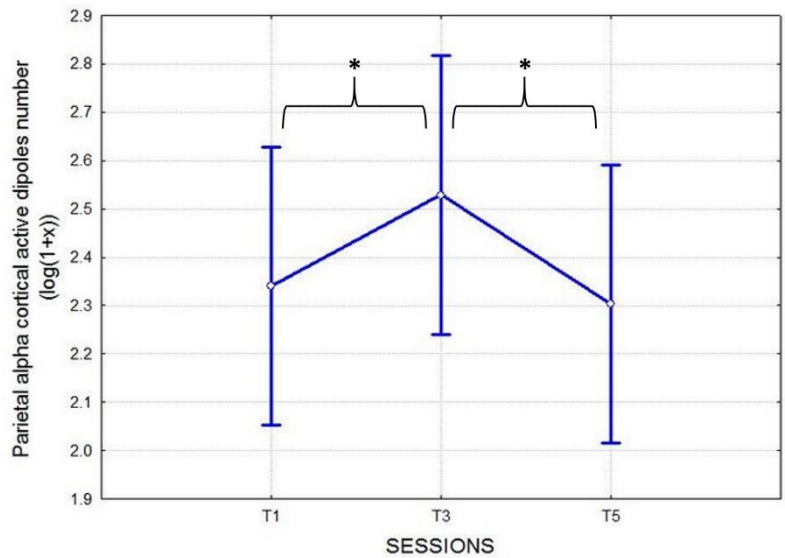

Figure 65. Parietal Brodmann ROIs quantitative analysis related to the numbers of statistically significant cortical voxels elicited by the tasks, when compared to the reference condition, across the trainig session (T1, T3 and T5) for the alpha frequency band.

Autonomic Signals

Figure 66 shows the results of the HR, presented as averaged r-square values over all the subjects. The HR showed that the experimental group was emotively engaged in correspondence of the first training session (T1), and then the HR continuously decreased across the other training sessions (T3, T5). The difference between T1 and T3 was not significant ($F(2, 18)= 32.23$; $p = 10^{-4}$), but the Duncan's post-hoc test, showed that T5 was statistically lower than the others ($p < 0.03$).

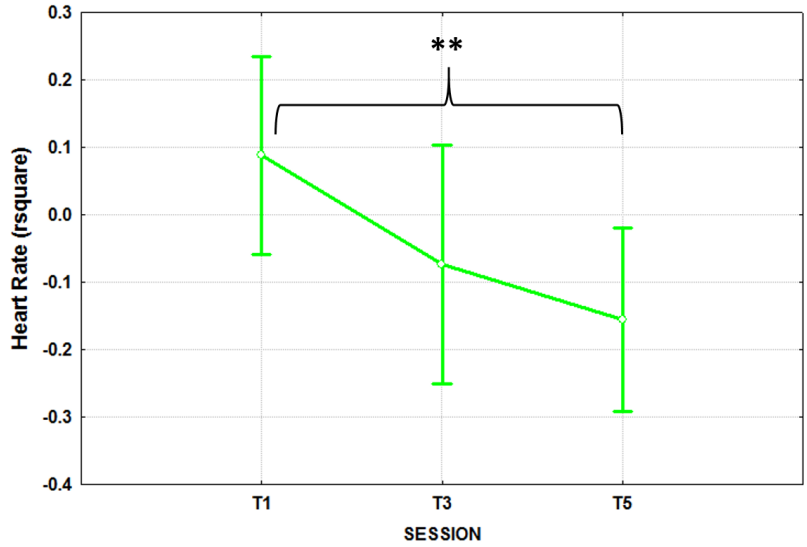

Figure 66. Heart Rate (r-square) values across the recording training sessions (T1, T3 and T5). The trend showed how the subjects felt more confident session after session, as the HR decreased continuously from the first (T1) to the final (T5) training session.

Figure 67 presents the changes of the EBR index across the different training sessions. The statistical analysis showed no significant results across the sessions ($F(2, 18) = 1.09$, $p = 0.36$), but the results suggest that in the first session (T1) the value of the EBR was lower than in the other two tasks (T3 and T5). A possible interpretation of this result could be that the subjects paid more attention to the task at the beginning (T1), as the EBR decreased respect to the reference task, than in the central and last ones (T3 and T5). This interpretation derives from the fact that in literature high EBR is associated with low attentional states (reviewed in Borghini et al., 2012a).

Frustration Analysis

The one-way repeated measures ANOVA performed on the "Frustration" factor of the NASA-TLX questionnaire showed a significant decrement across the training sessions (Figure 68). This factor is defined as how much stressed or comfortable the subjects felt in the various sessions. The Duncan post-hoc test confirmed how the subjects were significantly ($p<.05$) more confident with the task at the end of the training session (T5) than at the beginning (T1).

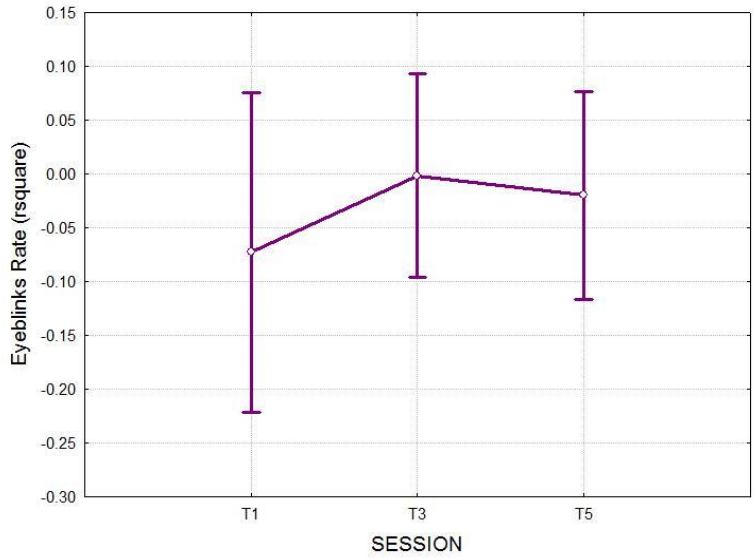

Figure 67. Eyesblink Rate (r-square) values across the recording training sessions (T1, T3 and T5). The differences among the sessions are not statistically significant, but we can hypothize that the subjects probably paid more attention in T1 than in the others, as the mean EBR in T1 is lower than in T3 and T5.

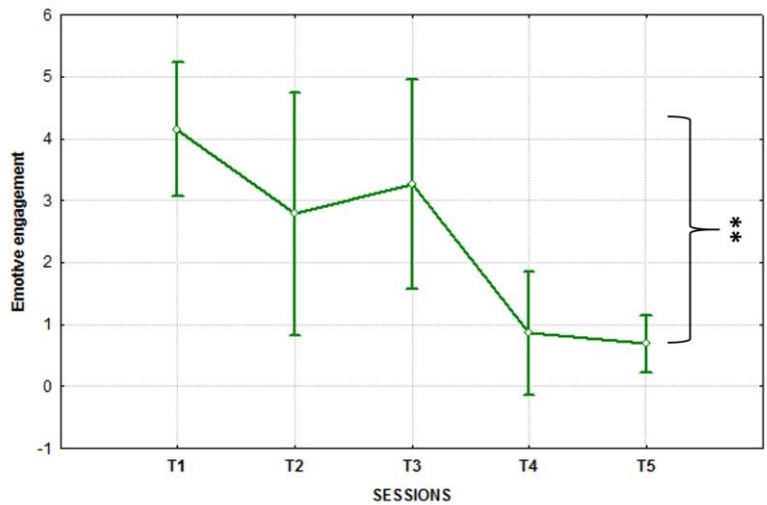

Figure 68. Emotive engagement of the subjects across the training sessions (T1÷T5). The results show how the subjects felt significantly ($p<.05$) more confident with the task, session after session.

4.1.3. Discussions

The aim of this work was to find out if it was possible to obtain quantitative information about the degree of the learning process throughout a training period by analyzing the variations of the EEG, ECG and EOG signals. The results suggested that the neurophysiological signals changed consistently across the training sessions, and they correlated with the overt behaviour of the subjects. Such overt behavioural scores and the experienced workload, reported by NASA-TLX, describe a story in which the subjects quickly found their own strategies and got confident with the execution of the proposed tasks (Figures 57, 58 and 68). In fact, the performance improvement was characterized by an asymptotic increasing where more than 90% of the final score was reached within the third training session (Figure 57), reaching a plateau until the last training sessions.

The gathered EEG, EOG and ECG data suggested a chain of slightly different cerebral events throughout the training period. Figure 59 shows a clear increase of the frontal EEG PSD in theta band from the first (T1) to the third day (T3) of training, and a high decrease from the third (T3) to the fifth day (T5). The same trend, but with opposite sign, was observed for the parietal alpha PSD (Figure 59). These results were also confirmed by the cortical maps (Figure 61). The increase of the frontal cerebral activity in theta band, and the decrease of the parietal alpha band in the third day (T3) were quantified by the evidence that the number of activated voxels in the frontal and parietal areas (Figures 64 and 65) decreased significantly in the first (T1) and fifth day (T5), when compared to the third one (T3). As the modulation of EEG PSD over the frontal areas, in theta band, and over the parietal areas, in the alpha band, were associated with the concept of "mental workload" (Borghini et al., 2012a, b; Klimesch, 1999), it might be concluded that the subjects had a peak of mental efforts in accomplishing the MATB task on the central session (T3), by returning to a moderate effort at the end of the training period (T5), due to the memorization and automatization of many cognitive processes for the correct execution of the task.

This overview is also supported by the trend of the HR and of the factor "Frustration" across the training period. Both the parameters continuously decreased from T1 to T5, and it showed how the subjects felt more confident

with the task on the last day (T5), as showed in Figures 66 and 67. From a single-subject point of view, it was possible to assess which subjects had a fast learning, a slow learning or who still needed to train. In fact, by comparing the trends of the brain activities of each subjects, it was possible to assess these differences and, then, to plan the training for each subject. In fact, by Figure 61, it was possible to find two subjects which started the training period with lower mean performance values than the rest of the group, but they reached the same performance of the other subjects since the third training session (T3) then. This fast-learning has been assessed by the trends of the single-subject performance and reflected by the highest values of the frontal theta PSD in the same session (T3) for these two subjects. What has then changed from T1 to T5 at the cerebral level? The EEG data suggested that the subjects during the training period generated a lot of mental effort up to the T3, with an accompanying increase of the overt performance. This mental effort decreased after T3, but the high level of performance was sustained until the end of the training period, where they did not significantly changed the mean performance level (saturation of the learning curve). The fewer cognitive resources needed by the subjects and the higher performance level on the last day of training (T5) confirmed how the automatic schemas, gained after being practiced for 5 consecutive days, made the subjects requiring minimal working memory resources and allowed problem solving to proceed automatically and with minimal effort.

4.1.4. Conclusions

The present study showed that the integration of information derived by the brain activity, through the EEG, and the physiological signals of ECG and of EOG, with the supervision of experts, could be used as possible innovative neurometric for evaluating the degree of the learning process and the training progress of trainees, e.g. pilots and air traffic controllers, throughout their periods of professional formation. Also, this method could be applied when the comparison between subjects is required. In fact, after a fixed period of training it could be possible 1) to quantify how well the subjects can complete a task, in terms of cognitive resources necessary to the correct execution, and 2) to compare subject's cognitive performances by estimating the neurophysiological EEG, HR and EBR parameters

presented in this study. Although the MATB is a benchmark for many studies in the human factor domain (Prinzel et al., 2003; Smith et al., 2001; Trejo and Shensa, 1999; Wilson and Russell, 2003), the observed modulation effects of the cerebral activity are not examined for other kinds of motor cognitive tasks. This limitation has been addressed in the next studies employing different motor-cognitive tasks, more subjects and different setup. In fact, the next study consisted in taking into account the entire experimental paradigm, in which the 10 students accomplished the MATB for a period of 3 weeks (SESSIONS T1 ÷ T12) with the aim to assess if the students were trained for real, or if some learning processes were still happening.

4.2. Machine Learning Approach for Training Assessment

As described previously, practice-related reorganization of the functional anatomy of task performance may be distinguished into a *redistribution* and a *'true' reorganization* (Kelly and Garavan, 2005). The first case constitutes a combination of increases and decreases in activation such that the brain activation map generally contains the same areas at the end as at the beginning of practice, but the levels of activation within those areas have changed. The second type of functional reorganization is considered to reflect a change in the location of activations and it is associated with a shift in the cognitive processes underlying task performance (Poldrack, 2000; Glabus and Horwitz, 2003). According to this framework, a 'scaffolding' set of regions is used to cope with novel demands during unskilled, effortful performance. After practice, processes or associations are more efficiently stored and accessed and the scaffolding network falls away, evinced by decreased activation in those '*scaffolding'* attentional and control areas. Activations seen earlier in practice therefore involve generic attentional and control areas - prefrontal cortex (PFC), anterior cingulate cortex (ACC) and posterior parietal cortex (PPC) are the main areas considered to perform the '*scaffolding'* role, consistent with theories of PFC function and the involvement of these areas in the distributed working memory system (Seidler et al., 2002). In order to deal with these phenomena across a training period of 3 weeks, the asSWLDA (see paragraph 3.1) has been used to select

the most important brain features able to provide the proper resolution to track and quantify the evolution of the brain processes throughout the different sessions. The hypothesis was that, even if the EEG is not a stationary process, there could be some brain features closely related to the observed physiological phenomenon (i.e. learning process) enough stable over the time that can be used for a reliable classification of such phenomenon. In other words, if the subject is not trained his/her brain features and activations will change across consecutive training sessions, and this will be reflected on a decreasing of the asSWLDA performance between consecutive sessions. On the contrary, if the subject is well trained his/her brain features and activations will be stable over time, and, as consequence, the performance of the classifier will not change over consecutive training sessions. Such hypotheses have been tested on a group of 10 students who have been asked to learn the execution of a task that they had never practiced before (MATB, see paragraph 2.3.1). The results of the proposed methodology (machine-learning analysis of EEG data) have also been validated and supported by the results of the behavioural data (task performance) and subjective data (workload perception) analyses.

4.2.1. Experimental Setup

Subjects

Ten healthy volunteers (students of the *National University of Singapore* - NUS) have given their informed consent for taking part at the experiment and each of them has been paid SG$200 for the entire experimental period. The study protocol has been approved by the local Ethics Committee. The selection of the subjects has been done accurately in order to ensure the same mental and physical state (homogeneity of the experimental sample). The subjects (25±3 years old) have been instructed to maintain a specific kind of lifestyle. In particular, they have been asked to avoid alcohol, caffeine, heavy meals right before the experiments, and to avoid extreme physical activity over the entire experimental protocol (homogeneity of the "internal conditions" of the subjects during the experiments). The Lab environment has been kept under control (lights intensity, room temperature, seat position) across the different days of the experiments (homogeneity of the "external conditions" during the experiment). In

addition, in order to have low sources of variances, the experimental group has been composed only by males.

Experimental Protocol

The subjects have been asked to practice and to learn to execute correctly the MATB (see paragraph 2.3.1) for three consecutive weeks (WEEK_1, WEEK_2 and WEEK_3). In total, they have taken part in 12 training sessions (T1 ÷ T12). with a duration of 30 minutes each (Figure 69). Each training session consisted in the execution of the MATB under different difficulty conditions (EASY and HARD) proposed twice in a random sequence in order to avoid any expectation and habituation effects. To investigate the trends and changes of the EEG parameters throughout the training period (3 weeks), as signs of learning progress, the behavioral, subjective and physiological data have been analyzed in specific training sessions (Kelly and Garavan, 2005), that is, the first day of WEEK_1 (T1), the last day of WEEK_1 (T5), the first repetition of the last day of WEEK_2 (T8), the first repetition of the WEEK_3 (T11) and the last repetition of the WEEK_3 (T12).

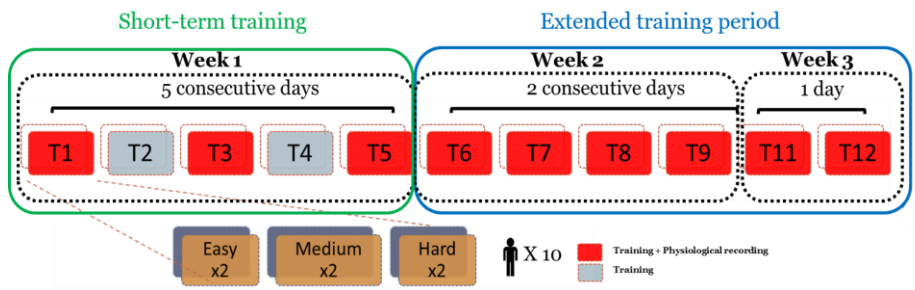

Figure 69. Extended training protocol. The subjects were asked to take part in a training period of 12 sessions (T1 ÷ T12) in which the different task conditions have been presented twice in random sequence in order to avoid habituation and expectation effects. In the red sessions, the EEG, ECG and EOG signals have been recorded, while in the gray ones, only the behavioural and subjective data have been collected.

As the subjects had never done the MATB, instructions about the execution of the task have been provided on the first day of training (T1) before starting with the experiments. To be efficient, the instructional design has to be tailored to the level of the subject, in order to avoid that the effectiveness of the training was likely random (Kalyuga et al., 2003). At

the end of each experimental condition of each training session, the subjects have filled the NASA-TLX questionnaire (Hart and Staveland, 1988) to gather the subjective workload perception of the proposed task.

Brain Activity Recording

Electroencephalogram (EEG) and vertical electrooculogram (EOG) have been recorded by a digital monitoring system (ANT Waveguard system). The 64 EEG channels and the EOG channels have been collected with a sampling frequency of 256 (Hz). All the EEG electrodes referred to both earlobes, grounded to the AFz channel and their impedances were kept below 10 (kΩ). The bipolar electrodes for the EOG have been positioned vertically over the left eye. The EOG signal has then been used to remove eyes-blink contributions from each epoch of the EEG signal, by using the Gratton and Coles (1983) algorithm, whilst for other sources of artifacts, specific procedures of the EEGLAB toolbox have been used (Delorme and Makeig, 2004). The EEG signal has been firstly band-pass filtered with a fourth-order Butterworth filter (low-pass filter cut-off frequency: 30 (Hz), high-pass filter cut-off frequency: 1 (Hz), and then it has been segmented into epochs of 2 seconds (*Epoch length*), shifted of 0.125 seconds (*Shift*). The epochs where the EEG signal amplitude exceed ±100 (μV) (*Threshold criteria*) have been marked as *artifact*. Then, each EEG epoch has been interpolated in order to check the slope of the trend within the considered epoch (*Trend criteria*). If such slope was higher than 3, the considered epoch was marked as *artifact*. The last check has been based on the EEG Sample-to-sample difference (*Sample-difference criteria*). If such difference, in terms of amplitude, was higher than 25 (μV), it meant that an abrupt variation (non-physiological) happened and the EEG epoch has been marked as *artifact*. At the end, all the EEG epochs marked as *artifact* have been removed with the aim to have a clean EEG dataset from which estimate the brain parameters for the different analyses. All the previous mentioned values have been chosen following the guidelines reported in Delorme and Makeig (2004). From the clean EEG dataset, the *Power Spectral Density* (PSD) has been calculated for each EEG 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 helped to smooth the contribution of the signal close to the end of the segment

(Epoch), improving the accuracy of the PSD estimation (Harris, 1978). The segmentation of the EEG signal has been done in order to have both a high number of observations (see equation 3.7) in comparison with the number of variables (see equation 3.8), and to respect the condition of stationarity of the EEG signal (Elul, 1969). In fact, the latter one is necessary to proceed with the spectral analysis of the signal. Then, the EEG frequency bands have been defined accordingly with the *Individual Alpha Frequency* (IAF) value estimated for each subject (Klimesch, 1999). Since the alpha peak is mainly prominent during rest conditions, the subjects have been asked to keep the eyes closed for a minute, and then, such condition the IAF has been estimated. Finally, a spectral features matrix (EEG channels x Frequency bins) has been obtained in the frequency bands directly correlated to learning processes. In particular, only in the theta band (IAF-6 ÷ IAF-2), over the EEG frontal channels (*Fz, F3, F4, AF3* and *AF4*), and in the alpha band (IAF-2 ÷ IAF+2), over the EEG parietal channels (*Pz, Pz* and *P4*), have been considered as variables for the mental workload evaluation.

Machine Learning Approach

A two-classes asSWLDA regression has been used to select within the training EEG dataset the most relevant EEG spectral features to discriminate the different task conditions. For each subject, 10 fold-cross validations, 5 considered training sessions (T1, T5, T8, T11, T12) per 2 task condition repetitions (TX-1 and TX-2, where X = 1, 5, 8, 11, 12), have been performed, by calibrating the classifier with the "EASY – HARD" conditions couple of a session repetition, and then by testing it over the following repetitions. For each couple of conditions, the *Linear Discriminant Function* has been calculated and the *Area Under Curve* (AUC) values of the *Receiver Operating Characteristic* (ROC, (Bamber, 1975)) have been estimated. In particular, five cross – validations groups have been defined (T1', T5', T8', T11' and T12'), as reported in Table 9. Within each group, an AUC value has been calculated as average of the cross-validations considering the training session of that group (e.g. T1 for T1' group) and testing it over the consecutive ones, and *vice versa*. Specifically, for the first four cross-validations groups (i.e. T1', T5', T8', T11'), the asSWLDA has been calibrated with the EEG dataset of each repetition of the considered session (i.e. group TX', thus TX-1 or TX-2,

where X = 1, 5, 8, 11), and then tested on the following repetitions (i.e. TY-1 and TY-2, where Y = 5, 8, 11, 12, with Y > X). For each combination, the roles of training and testing dataset have been successively inverted (e.g. considering group TX', the AUC values were computed training the classifier with TX-1/2 and testing it on TY-1/2 and *vice versa*). In the case of the last cross-validations group, i.e. T12', since there were not following sessions, the asSWLDA has been trained with one repetition and tested on the remaining one. In total, the number of cross – validations *(#Cross-validations*) for each one of the considered sessions has been 32 for T1', 24 for T5', 16 for T8', 8 for T11' and 2 for the last cross-validations group (T12').

Table 9: Cross-validations scheme.

Cross-validation	**Calibration dataset**	**Testing dataset**	**#Cross-validations**
T1'	T1-1, T1-2	T5-1, T5-2, T8-1, T8-2, T11-1, T11-2, T12-1, T12-2	32
	T5-1, T5-2, T8-1, T8-2, T11-1, T11-2, T12-1, T12-2	T1-1, T1-2	
T5'	T5-1, T5-2	T8-1, T8-2, T11-1, T11-2, T12-1, T12-2	24
	T8-1, T8-2, T11-1, T11-2, T12-1, T12-2	T5-1, T5-2	
T8'	T8-1, T8-2	T11-1, T11-2, T12-1, T12-2	16
	T11-1, T11-2, T12-1, T12-2	T8-1, T8-2	
T11'	T11-1, T11-2	T12-1, T12-2	8
	T12-1, T12-2	T11-1, T11-2	
T12'	T12-1	T12-2	2
	T12-2	T12-1	

The expected results have been based on the following hypotheses: i) if the subject is well trained, the AUC of such cross-validation (i.e. T1', T5', T8', T11'), and the consecutive ones, will show comparable value to the *gold standard* condition (no significant difference, $p > 0.05$), that in the considered study is the last cross-validation (T12'); ii) on the contrary, if

the subject is not well trained, the AUC of the considered cross-validation will show a significant lower AUC value than the *gold standard* condition; iii) since the subject is trained, the brain features might remain almost stable within the same session, especially after 3 weeks of training, therefore the T12' cross-validation should show the highest AUC value.

Statistical Analyses

Repeated measures ANOVAs (Confidence Interval, CI = 0.95) have been used for the statistical validation of the results derived by the different analyses, by using the STATISTICA software (Statsoft). In particular, two one-way repeated-measures ANOVAs have been performed on the task performance data and NASA-TLX total score (independent variables) with the *within* factors SESSION (five levels: T1, T5, T8, T11 and T12). The same ANOVA has been performed by considering as *within* factor the CROSS-VALIDATION (five levels: T1', T5', T8', T11' and T12') on the AUC values (independent variable). Duncan post-hoc tests have been performed to assess significant differences between all pairs of levels of the considered factors.

4.2.2. Results

Task Performance

The ANOVA highlighted significant differences ($F(4,36)=9.92$; $p= 0.00002$) between the task performance values (Figure 70) across the different training sessions. In particular, it showed that at the end of the WEEK_1 (T5), the subjects improved significantly the level of their task performance ($p < 0.05$), and then they kept such a high level of performance stable (no significant differences, $p > 0.05$) for the rest of the training period, that is, until the last session (T12). In fact, the Duncan post–hoc test did not find any differences, in terms of task performance, between the last day of WEEK_1 (T5) and the remaining training sessions (T8, T11 and T12).

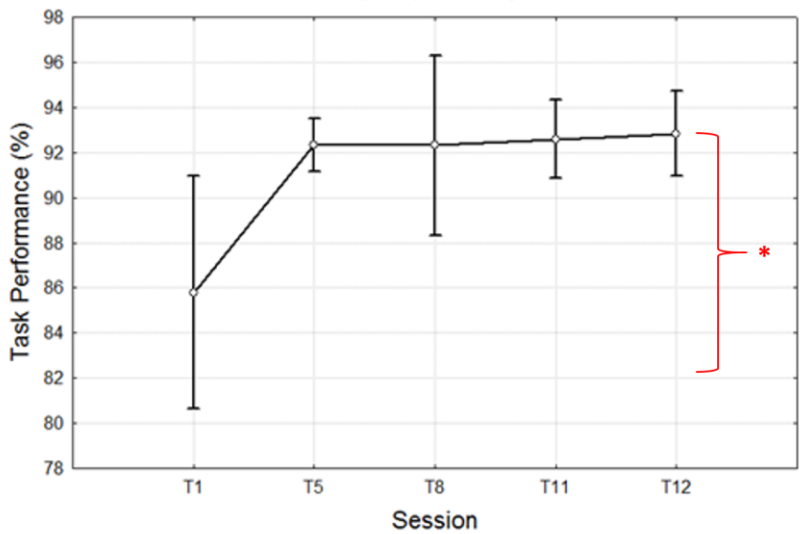

Figure 70. Task performance values over 3 weeks of training. The ANOVA showed a significant ($p < 0.05$) improvement of performance from T1 to T5 and then no differences were found between the consecutive training sessions (T5÷T12). Such results indicated that since T5 the subjects reached the saturation area in terms of task performance.

Workload Perception: NASA-TLX

In Figure 71, the total scores of the subjects' workload perceptions across the considered training sessions are reported. The ANOVA analysis showed a significant effect on the factor SESSION ($F(4,36)=4.14$; $p=.00731$). The Duncan post–hoc test showed significant differences ($p < 0.03$) in the

perception of the task's workload between WEEK_1 (T1 and T5) and WEEK_3 (T11 and T12). In the WEEK_2 (T8), the perception of the workload was slight lower than in the WEEK_3, but not significantly different from the last session of WEEK_1 (T5).

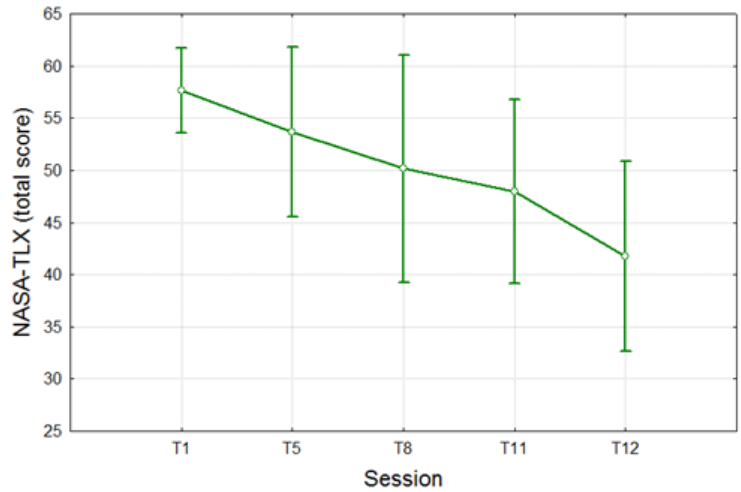

Figure 71. The figure reports the ANOVA on the NASA–TLX total scores across the considered training sessions. The results showed that the perception of the workload decreased significantly session after session. In other words, the subjects were feeling more confident and familiar with the task after each training session.

Machine Learning Results

The ANOVA (Figure 72) highlighted significant differences ($F(4,36) = 4.77$; $p = 0.003$) between the cross-validations (T1'÷T12'). In particular, the Duncan post-hoc test showed that the AUC value related to the first cross-validation (T1') was significantly lower (all $p < 0.02$) than those related to the others (T5', T8', T11', T12'). On the contrary, from the second cross-validation (T5'), no significant differences on the AUC values have been highlighted among the remained cross-validations (T8', T11' and T12').

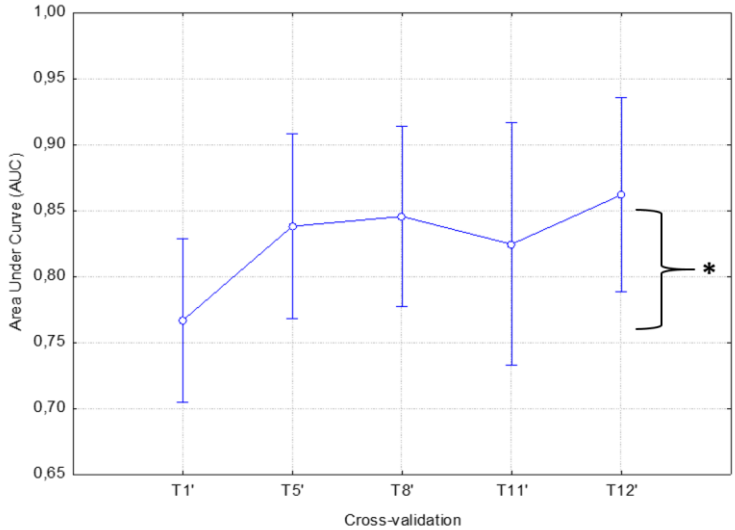

Figure 72. The ANOVA on the AUCs highlighted significant differences across the training sessions. In particular, the AUCs related to the first cross-validation (T1') was significantly lower (all $p < 0.02$) than those related to the others (T5', T8', T11', T12'). On the contrary, from the second cross-validation (T5'), no significant differences have been highlighted among the remained cross-validations (T8', T11' and T12').

In Figure 73, the single cross-validation has been plotted with the aim to better illustrate the differences along the training sessions. In particular, the AUCs of the cross-validations by calibrating the asSWLDA on T1 (red bars), on T5 (yellow bars), on T8 (blue bars), on T11 (green bars) and on T12 (black bar) have been reported. The dash black line indicates the trend of the averaged AUCs from T1 to the last training session (T12). As illustrated above, there was a significant increment of the AUCs from T1 to T5, and then no further differences were found among the remaining sessions.

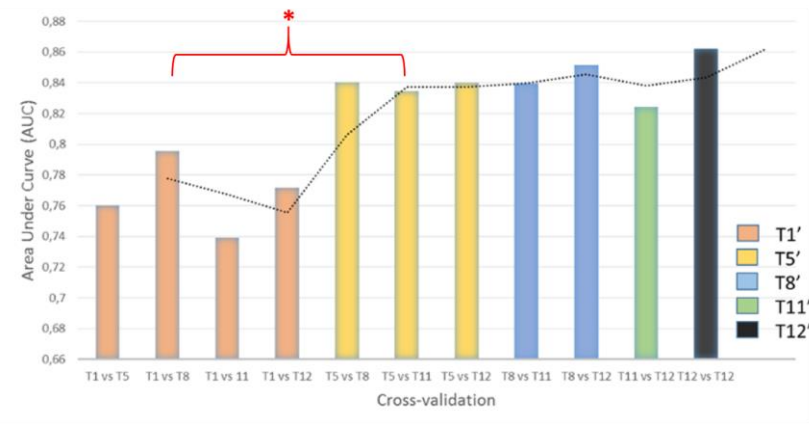

Figure 73. In the figure, the single cross-validations has been plotted with the aim to better illustrate the differences among the training sessions. In particular, the AUCs of the cross-validations by calibrating the asSWLDA on T1 (red bars), on T5 (yellow bars), on T8 (blue bars), on T11 (green bars) and on T12 (black bar) have been reported. The dash black line indicates the trend of the averaged AUCs and it highlights the stability (no significant differences) of the selected brain features from T5 to the last training session (T12).

4.2.3. Discussions

The analysis of the task performance (Figure 70) highlighted the existence of a training effect, as since the last day of WEEK_1 (T5) the subjects kept their performance level significantly higher ($p < 10^{-4}$) than the beginning of the training (T1), and no differences were found among the rest of the sessions (T5 ÷ T12). In terms of workload perception, in the last week of training (WEEK_3) the subjects perceived a significant lower ($p = 0.007$) workload demand than in the previous weeks (WEEK_1 and WEEK_2), but no differences have been found between WEEK_2 and WEEK_3 (Figure 68). In terms of AUC (Figures 72 and Figure 73), the asSWLDA showed a significant increment ($p = 0.03$) from T1 to T5 and then no differences among the rest of training sessions (T5 ÷ T12). Such results suggested that from T1 to T5 the subjects improved their training level with the aim to accomplish correctly the MATB and to reach high performance. As consequence, the most important brain features selected by the classifier in the T1 session could be hypnotized significantly different from those selected in T5. Successively, since in T5 no significant changes were found, in terms of classifier performance (Figures 72 and 73), up to the end of the

training period (T12), it could be reasonable to retain that the brain features remained stable over all the remaining training sessions (T5 ÷ T12). In other words, the subjects became trained at the end of the first week (T5) and then no significant cognitive changes happened in the next two weeks (WEEK_2 and WEEK_3). Two main limitations of the study have to be considered at the moment. The first limitation is that the experimental group is just sufficient to highlight some significant statistical patterns, but it needs to be increased to explore if other stable neuroelectrical pattern could emerge from the analysis of the EEG signals during the training. A second limitation is the proposed task (MATB) provided by the NASA. While this is good for the analysis of the brain reaction while handling with multiple tasks, it could be reasonable to retain that brain activities occurring during specific training inside an airplane cockpit, or air-traffic room could be different from that elicited by MATB tasks. Therefore, the results presented in this study have to be considered as a promising step for further studies.

4.2.4. Conclusions

In this study, I demonstrated that a machine-learning approach could be used to objectively assess if and when the subjects are well trained, especially in terms of cognitive resources. In fact, there were no differences in calibrating the classifier with the brain features extracted from one training session rather than the others, once the subjects became trained. The results stressed the importance of the proposed approach because different subjects could achieve the same performance level and, by the only task performance analysis, it would not be possible to obtain information in terms of cognitive activations and to assert possible differences between them. In other words, by the standard analysis (task performance or self-reports), it would not be possible to find out the subjects who require fewer cognitive resources than the others, hence who show higher cognitive spare capacity to handle with unexpected or emergency events. The results highlighted the wide applicability and general validity of the proposed method for the training assessment in operational environments. In fact, the proposed approach could be used in flight schools to select the most skilled pilots for a certain kind of activity or in hospitals for the training assessment of operators during (simulated) surgery interventions.

5. Cognitive Control Behaviour: Skill, Rule and Knowledge

Training not only could result in the acquisition of new skills (Satterfield and Hughes, 2007; Hill and Lent, 2006)) but also in improved declarative knowledge, enhance strategic knowledge, defined as knowing when to apply a specific procedure or skill, in particular during unexpected events (Kozlowski et al., 2001). Furthermore, despite the time passed from the last training session, there might be the need to assess if the operator is still able to work ensuring a high performance level, hence, a proper level of safety. For such a reason, another issue is the necessity of objectively monitoring and assessing operators' performance (Leape and Fromson, 2006), in terms of cognitive control behaviours. Nevertheless, although the results in terms of performance should be the same, the cognitive demand for the same operator could not. In other words, after a certain time period the operator is still able to execute the same task by achieving the same performance level, but it might require different amount of cognitive resources. Therefore, different operators could achieve the same performance results, but involving a different amount of cognitive resources, thus different expertise, and nowadays there are no tools able to provide such objective and quantitative information in order to better manage the training program and personnel selection. As a reference to my study, I consider the S-R-K framework introduced by Rasmussen (1983), still used in the aeronautic field. The aim of the framework is to explain human behavior and describe the wide range of mental capabilities used in everyday situations (e.g. working environments).

During recent decades, attempts have been made to extend these models to higher level human decision making to conform with the increasing levels of automation in operational environments (e.g. aviation, hospitals, public transport), and to transfer such models for process control applications. In this effort, it has to be considered that humans are not simply deterministic input-output devices but *goal-oriented* creatures who actively select their goals and seek the relevant information (Rosenbluth et al., 1943; Polanyi, 1958). At a higher level of conscious planning, most human activity

depends upon a rather complex sequence of activities, and feedback correction during the course of behaviour from mismatch between goal and final outcome will therefore be too inefficient, since in many cases it would lead to a strategy of blind search (Rasmussen, 1986, 1989). Human activity in a familiar environment will not be goal-controlled; rather, it will be oriented towards the goal and controlled by a set of rules which has proven successful previously. In unfamiliar situations when proven rules are not available, behaviour may be goal-controlled in the sense that different attempts are made to reach the goal, and a successful sequence is then selected. The efficiency of humans in coping with complexity is largely due to the availability of a large *repertoire* of different mental representations of the environment from which rules to control behaviour can be generated *ad-hoc.* When we distinguish categories of human behaviour according to basically different ways of representing the constraints in the behaviour of a deterministic environment or system, three typical levels of performance emerge: *skill-*, *rule-*, and *knowledge*-based performance (Rasmussen, 1983). These levels and a simplified illustration of their interrelation are shown in Figure 74. The *skill-based behaviour* represents sensory-motor performance during acts or activities which, following a statement of an intention, take place without conscious control as smooth, automated, and highly integrated patterns of behaviour. In most skilled sensory-motor tasks, the body acts as a multivariable continuous control system-synchronizing movements with the behaviour of the environment. Performance is based on feedforward control and depends upon a very flexible and efficient dynamic internal world model. Characteristically, skilled performance rolls along without conscious attention or control. The total performance is smooth and integrated, and sense input is not selected or observed: the senses are only directed towards the aspects of the environment needed subconsciously to update and orient the internal map. In general, human activities can be considered as a sequence of such skilled acts or activities composed for the actual occasion. The flexibility of skilled performance is due to the ability to compose, from a large *repertoire* of automated subroutines, the sets suited for specific purposes.

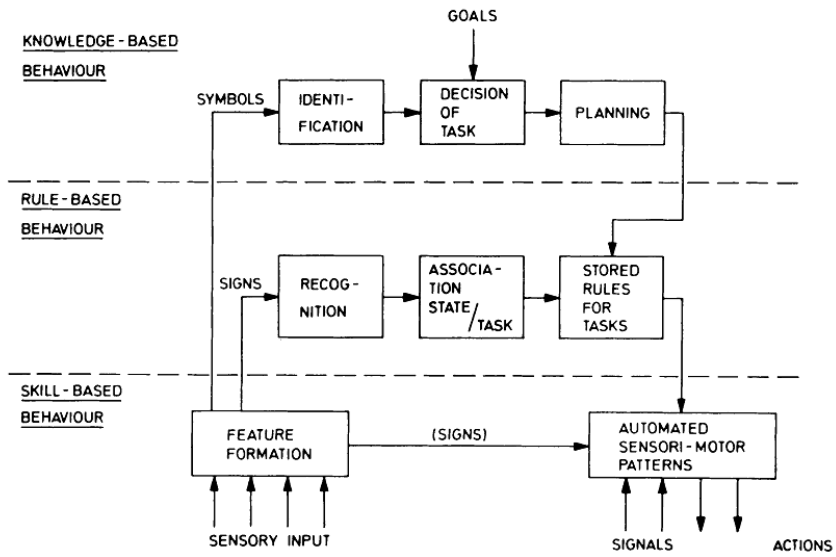

Figure 74. Simplified illustration of three levels of the human performance model proposed by Rasmussen (1983). Note that skill, rule and knowledge levels are not alternatives, but interact in a way only rudimentarily represented in the diagram.

At the next level of *rule-based behaviour*, the composition of such a sequence of subroutines in a familiar work situation is typically controlled by a stored rule or procedure, which may have been derived empirically during previous occasions or it may be prepared on occasion by conscious problem solving and planning. In this case, the performance is goal oriented but structured by "feedforward control" through a stored rule. Very often, the goal is not even explicitly formulated, but it is found implicitly in the situation releasing the stored rules, that is, selected from previous successful experiences. Feedback correction during performance will require functional understanding and analysis of the current response of the environment, which may be considered an independent concurrent activity at the next higher level (*knowledge-based*). The boundary between skill-based and rule-based performance is not quite distinct, and much depends on the level of training and on the attention of the person. In general, the skill-based performance rolls along without the person's conscious attention, and he/she will be unable to describe how he/she controls and on what information he/she bases the performance. The higher level, rule-based coordination, is generally based on explicit know-how, and the rules

used can be reported by the person. During unfamiliar situations, faced with an environment for which no know-how or rules for control are available from previous encounters, the control of performance must move to a higher conceptual level, in which performance is goal-controlled and *knowledge-based.* In this situation, the goal is explicitly formulated, based on an analysis of the environment and the overall aims of the person. Then a useful plan is developed -by selection - such that different plans are considered, and their effect tested against the goal, physically by trial and error, or conceptually by means of understanding the functional properties of the environment and prediction of the effects of the plan considered. At this level of functional reasoning, the internal structure of the system is explicitly represented by a "mental model" which may take several different forms. An example is learning to use the T9-keyboard system on the mobile phone. The first time is knowledge-based to diagnose the way for using T9-keyboard system and produce a rule. When the procedure to select letters and words is known, the control shifts to the rule-based behavior for applying the rules learned. Finally, after practice, the procedure may turn automated, therefore skill-based. Similar distinctions between different categories of cognitive control behaviour have been proposed elsewhere. For example, Fitts and Posner (1962) distinguishes between three phases of learning a skill: the early or cognitive phase, the intermediate or associative phase, and the final or autonomous phase. If we consider that in real life a person will have a varying degree of training when performing a task depending on variations and disturbances, the correspondence with the three levels in the present context is clear. One aspect of the categorization of human performance in skill/rule/knowledge-based behaviour is the role of the information observed from the environment, which is basically different in the different categories. This is the case even though major problems during unfamiliar situations may be caused by the fact that the same indication may be perceived in various different roles and that it is a well-known psychological phenomenon that shift between different modes of perception is difficult (Rasmussen, 1986).

Towards a Possible Integration Between SRK and Neurophysiological Variables

Several interesting issues about the neuroanatomical and functional models may be related to the level of attentional control of SRK framework, for a possible integration between SRK and neurophysiological variables. One of the first models of attentional control was proposed by Posner and Petersen (1990) and consists of two separate systems, but closely interrelated: an anterior attentional system (linked to environmental monitoring and detection of target stimuli) and a posterior attentional system (linked to the orientation of attention). More recent versions of this model describe three systems (Posner and Rothbart, 2007; Raz and Buhle, 2006; Kochanska et al., 2000):

- *Alert/vigilance system*, connected to frontal and parietal regions, in particular of right hemisphere and its function consists in maintaining a state of activation;
- *Orienting system* (posterior attentional system, PAS), consists of posterior parietal and frontal cortex, temporo-parietal junction (TPJ), thalamic nuclei such as pulvinar and reticular nuclei, and superior colliculus. Its functions include anchoring, disanchoring and shift of attention, selection of specific information from multiple sensory stimuli;
- *Executive system* (anterior attentional system, AAS), consists of prefrontal medial cortex, anterior cingulate cortex and supplementary motor area included. Its function include voluntary control of behavior, conscious elaboration of experience, handling novel situations, monitoring, and resolution of conflicts. These conflicts may include planning or decision-making, error detection execution of new responses, inhibitory control, self-regulation, involvement in stressful conditions. It is proposed that anterior cingulate cortex (ACC) has a crucial role in monitoring the environment, detection of target stimuli appearance and resolution of conflicts between stimuli.

Furthermore, Corbetta and Shulman (2002), by incorporating several recent studies, have proposed that cortical control of attention is divided into two functionally divided but interacting systems:

- *Dorsal frontoparietal network* (top-down, endogenous attention), is guided by cognition and is involved in control of goal-driven attention, preparation and application of relevant stimuli selection. In dorsal frontal regions, this system includes frontal eye fields (FEF) and medial and lateral frontal cortex while, in dorsal parietal regions, includes the intraparietal sulcus (IPS) and superior parietal lobule (SPL).
- *Ventral frontoparietal network* (bottom-up, exogenous attention), is guided by perception and is a stimulus-driven attentional system and is involved in disengagement and re-orienting of attention towards salient or unexpected stimuli. The attentional shifts are automated. The areas involved are temporoparietal junction (TPJ), inferior parietal lobule (IPL), superior temporal gyrus (STG) and ventral frontal cortex (VFC) in particular inferior and middle frontal gyrus (IFG, MFG).

The S-R-K framework can be connected to the models mentioned above that provide an anatomical and functional base to do a step towards the investigation of neurophysiological correlates of Rasmussen's taxonomy. In fact, among the cognitive neuropsychologists, in terms of limited capacity, it has been hypothesized a mechanism with the function of programming and control cognitive processes in relation to priorities, goals and external conditions. This mechanism called *central executive* provides a higher-level processing in a hierarchical organization. Since the central executive has a limited capacity, many routine operations should be delegated to mechanisms that operate automatically, regardless of the voluntary attention costly in terms of effort. An automatic process (Shiffrin and Schneider, 1977) consists in the activation, on the base of appropriate inputs, of a learned sequence of elements, which proceeds without intentional attention that is, without the active control by the subject. A controlled process is, instead, limited capacity: requires continuous effort and monitoring by the subject. It works in serial mode, drawing on the stock of short-term memory.

The automation (or proceduralisation) is facilitated by practice, which makes the process faster, parallel, with less demand for mental load. An example is driving a car, in which processes before "controlled", subsequently become largely automated, because various operations are performed without the use of conscious attention. Anyway, in special cases (use of a new car, writing a text in a foreign language) the return to control of different operations, that in usual conditions are automated, is needed. However, it should be remembered that attention does not always coincide with controlled process: the automatic process also focuses attention on certain stimuli, selecting among other not-relevant stimuli for the task, and keeps on focusing them for the necessary time. Neuman (1984) pointed out that the automatic process is not completely uncontrolled, but rather control is below the threshold of consciousness. Norman and Shallice (1986) distinguished between processes totally automatic and others partially automated, based on selection (in the absence of conscious or voluntary control) between the established schemes, competing (defined *contention scheduling*); the selection is based on immediate priorities dictated by learning or by context. Thus, it seems appropriate to distinguish between *non-intentional* attentional processes, involving minimal effort and self-perception of the subject, and *intentional attention*, conscious and limited capacity. The relationship between awareness and attention is therefore extremely complex. Several empirical studies (Kellogg and Ronald, 1980; Marcel, 1980; Davidson et al., 1986) demonstrate how information receiving not consciously attention are processed in relation to their meaning, and how the latter can also be analyzed in absence of intentional control and therefore of consciousness. Starting from the point of view that attentional mechanisms control the access to awareness (Laberge, 1995; Baars, 1997), different models tend to identify the function of attentional control with conscience (as central operative system of activation and self-monitoring). According to Allport (1988) and Schmidt (1990, 2001), the intentionality, attention, control and awareness are the constitutive dimensions of consciousness. Other authors distinguish between consciousness of external reality and self-awareness or awareness of their own cognitive processes (meta-cognition), their emotions and their choices (Figure 75).

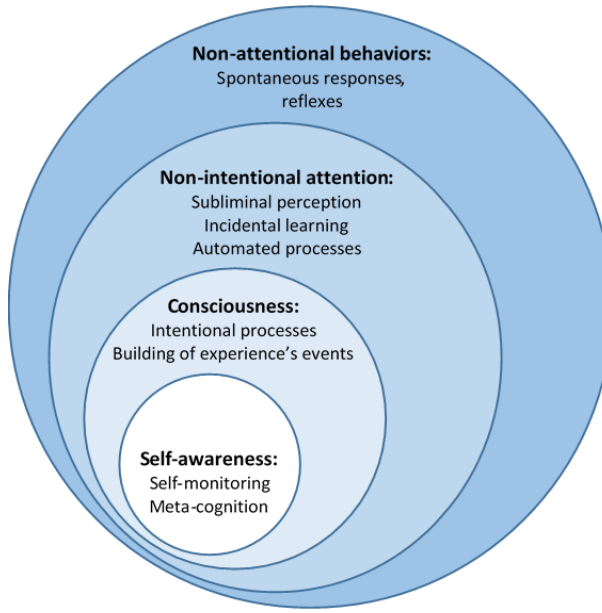

Figure 75. Several studies pointed out that the automatic process is not completely uncontrolled, but rather control is below the threshold of consciousness. Automatic process can be distinguished between processes totally automatic and others partially automated. Thus, it seems appropriate to distinguish between *non-intentional* attentional processes, involving minimal effort, and self-perception of the subject, and *intentional attention*, conscious and limited capacity.

5.1. Expertise Estimation of Professional ATCOs

According to the *Skill-Rule-Knowledge* (S-R-K) framework proposed by Rasmussen (1983, 1986, 1989), the human behavior can be controlled at different levels of conscious control, depending on the degree of familiarity with the task and the environment. SRK behaviors represent three dynamic and parallel cognitive levels of expertise, where the control of behavior continuously shifts from a level to another one. The S-R-K framework is still lacking a quantitative, reliable and validated way to be measured. Assessing the level of cognitive control of real professionals working in their environments would enable a better comprehension of the cognitive process actually activated in real work activities. The objective of this study was to assess if it was possible to differentiate the three degrees of cognitive controls proposed by the S-R-K model by the analysis of the Air Traffic Controllers' brain activity and to define an objective method to overcome the actual limitation in measuring and quantifying the S-R-K cognitive behaviors mentioned previously. In particular, the goal is to find out the brain features able to define a metric able to discriminate the S-R-K level independently by those used for the mental workload (see paragraph 3) evaluation and to compare ATC Experts and ATC Students in terms of SRK behaviours from a neurophysiological point of view. Firstly, a literature review has been done specifically to find out brain features mainly correlated to cognitive processes involved in the different S-R-K cognitive behaviours. Among all the possible brain features, four have been initially selected and tested if they were able to discriminate the S-R-K levels. Because their correlation with working memory, information processing, decision making, two of those features were the frontal theta and parietal alpha EEG rhythms (see paragraph 3). The other two brain features were the parietal theta and frontal alpha EEG rhythms. The former is correlated to the procedural memory and the latter is correlated to the attention level (see following sections). Then, the asSWLDA (see paragraph 3.1) has been calibrated with the selected brain features to characterize and then discriminate the different cognitive control levels during the execution of realistic ATC activities by professional Controllers.

Parietal Theta EEG Rhythm and Procedural Memory.

The potential role of theta rhythms engaged in the hippocampal/PFC (prefrontal cortex) interplay in consolidation of memory and report empirical results on the effect of enhanced theta oscillations on memory consolidation. It is widely accepted that memory consolidation process happen off-line, after the initial hippocampus encoding event. Consolidation relies on the re-activation of neuronal circuits that were implicated in the initial encoding (Albouy et al., 1995). Experimental evidence suggests that motor memory formation occurs in two subsequent phases (Albouy et al., 1995; Dudai, 2004; Karni et al., 1994; Luft & Buitrago, 2005). The first is initial encoding of experience during training that occurs within the first minutes-to-hours after training, and is characterized by rapid improvement in performance. The second phase is memory consolidation, and involves a series of systematic changes at the molecular level, that occur after training. This second phase requires longer time. During consolidation, memories are reorganized and hippocampus-dependent initial memories may become hippocampal independent (Albouy et al., 1995; Maquet, 2008). Processes of reactivation of memories lead to renewed consolidation each time reactivations occur, enhancing the first consolidated memory representation, and converting it into a long-lasting stable memory trace (Dudai, 2004). Delayed additional gains occur after the second phase, even without additional practice (Karni et al.,1994). Neural oscillations, in general, have been assumed to play a central role in cognitive processes and specific states of phase synchronization are considered a mechanism of increased communication between regions (Fell & Axmacher, 2011; Varela, Lachaux, Rodriguez, & Martinerie, 2001; Womelsdorf et al., 2007). Several lines of evidence suggest that theta oscillations play an important role in formation of memory: theta oscillations are typical of hippocampal activity, upon memory encoding, generating oscillations which can propagate to other brain structures (even relatively distant), supporting memory consolidation and are thought to play a critical role in the induction of long-term plasticity, associated with memory consolidation (Chauvette,2013; Kropotov, 2008). Theta rhythms are correlated with episodic and semantic memory (Buzsáki, 2005; Guderian & Duzel, 2005; Kahana, 2003) and are involved in learning and memory within the mPFC (medial prefrontal cortex) and hippocampal system (Anderson, Rajagovindan, Ghacibeh,

Meador, & Ding, 2010, Benchenane et al., 2010; Steinvorth, Wang, Ulbert, Schomer, & Halgren, 2010). Several studies point to theta synchronization as a mechanism underlying communication between the hippocampus, the ventromedial prefrontal cortex and remote memory areas, during consolidation. The underlying mechanism is still not clear. One attempt to explain the mechanism of memory consolidation is known as the "system-level memory consolidation theory" (Nieuwenhuis & Takashima, 2011). This model suggests that the hippocampus is strongly activated in the first stages of memory related neocortical formations, but gradually new forms of memory become independent of hippocampal activations, and consolidation correlates with increased activation in the human subgenual ventromedial prefrontal cortex (vmPFC). The vmPFC, similar to the anterior cingulate cortex, seems to link the neocortical representational areas in remote memory (Maquet, 2008; Nieuwenhuis & Takashima, 2011). The system-level memory consolidation view implies exchange of information in a network of brain areas. The center is the hippocampus and the communicating areas include the neocortex and structures such as the amygdala and the striatum (Battaglia et al., 2011; Maquet, 2008). The interaction between the hippocampus and striatum resembles the interaction between the hippocampus and neocortex (Battaglia et al., 2011). This exchange is theorized to be linked to theta oscillations: hippocampal cells fire preferentially at a specific theta phase (Mizuseki et al., 2009; Klausberger et al., 2003), and so do areas in the medial temporal, parietal lobes and other areas that exchange information with the hippocampus. Thus theta is assumed to regulate information exchange between the hippocampus and striatum (for a review see Battaglia et al., 2011). This exchange of information extends to relatively distant sensory and associative areas of parietal cortex, which are also entrained by theta oscillations (Sirota et al., 2008). Exchange of information is based on a dynamical evolving schema, in which synchronized discharge of cell assemblies across brain structures are orchestrated by theta to encode information. Recent results further support the central role of the hippocampus–striatum exchange in motor memory consolidation, and suggest that the interplay between the striatum and the hippocampus during motor training conditions subsequent motor sequence memory consolidation, which is further supported by reorganization of cerebral

activity in hippocampo-neocortical networks after sleep (Albouy et al., 2013). The ventral striatum is involved in learning beyond memory consolidation and was found to be related to individual variations in learning performance (Vink, Pas, Bijleveld, Custers, & Gladwin, 2013). We hypothesize that enhanced theta supports exchange of information between the hippocampus and neocortical areas during consolidation of memory, hence will be reflected in indicators of memory consolidation. (Gruzelier, 2009; Gruzelier et al., 2006). Gruzieler (2009) associated increased parietal theta (recorded from the Pz channel) with increased activity in the hippocampus. Parietal theta synchronization was also found to be correlated with retrieval (Jacobs, Hwang, Curran, & Kahana, 2006; Sauseng, Klimesch, Schabus, & Doppelmayr, 2005). In addition, the system level theory predicts that synchronized elevated theta power leads to consolidation of memory. To test the role of theta in memory consolidation, Reiner et al, (2014) asked whether enhanced theta during awake-hours, affects consolidation of memory. Indeed, results show that enhanced theta is correlated with behavioural changes indicating consolidation of memory.

Frontal Alpha EEG Rhythm and the Attention.

Concerning the alpha EEG rhythm, many evidences demonstrated how the structures of the thalamus are involved in exerting attentional bias (e.g., Crick 1984; Laberge 2001). A measure of support for this notion comes from imaging studies reporting attentional modulations in the pulvinar nucleus (Laberge and Buchsbaum 1990; Petersen et al. 1987) and in the lateral geniculate nucleus (O'Connor et al. 2002; Vanduffel et al. 2000). However, far less is known regarding the dynamical character of the control mechanisms exerted by such a network. A phenomenon that has repeatedly been linked with thalamo-cortical interplay is the human alpha rhythm, an oscillation within the 8 to 12 (Hz) frequency band observable in the scalp EEG (Lopes da Silva 1991). It was hypothesized that oscillations in this range embody the mechanism by which gating might occur in the thalamus (Lopes da Silva 1991). Perhaps not surprisingly given these findings, effects of alpha power have been observed during selective attention tasks that require the gating of distracting information (Foxe et al. 1998; Fu et al. 2001). As outlined in Neuper and Pfurtscheller (2001), the event-related desynchronization (ERD) of EEG activity in the alpha band presumably

reflects an increased excitability level of neurons in the involved cortical areas, which could be related to an enhanced information transfer in thalamocortical circuits (see also Pfurtscheller & Lopes da Silva, 2005). In contrast, event-related synchronization (ERS) of alpha activity (i.e., increases in alpha activity from the pre-stimulus reference to the task performance interval) is thought to reflect a reduced state of active information processing in the underlying neuronal networks (Pfurtscheller & Lopes da Silva, 2005) or 'cortical idling' (Pfurtscheller, 1999; Pfurtscheller, Stancak, & Neuper, 1996). However, recent evidence in this field of research also suggests that synchronization of alpha activity can be viewed as a functional correlate of active cognitive task performance presumably involving cognitive inhibition processes (for a review see Klimesch et al., 2007). Contrary to the usual finding that alpha power decreases when individuals become engaged in the performance of cognitively demanding tasks, Klimesch et al. (1999) reported a 'paradoxical' synchronization of alpha activity during the retention period in a short term memory task. Moreover, the amount of alpha activity has been shown to increase with memory load (Jensen et al., 2002a) and during manipulation of memory content as compared to simple retention of information (Sauseng et al., 2005). Cooper et al. (2003) showed that alpha synchronization is also related to internally versus externally directed attention. They presented sequences of stimuli in the visual, acoustic and haptic domain and then trained participants to imagine these stimulus sequences. They found that alpha activity was consistently higher during the imagination of stimulus sequences (i.e., internally directed attention) than during their presentation (i.e., externally directed attention). In these studies the observed synchronization of alpha activity has been interpreted to reflect selective inhibition of task irrelevant brain areas or inhibition of interfering external input (Klimesch et al., 2000, 2007; Rihs et al., 2007), and to reflect internal information processing involving top-down control on internally represented information (Sauseng et al., 2005; Von Stein & Sarnthein, 2000). EEG alpha activity has also been found to be sensitive to creative cognition in a series of studies employing a variety of methodological approaches (Arden et al., 2010; Dietrich & Kanso, 2010). These studies include the investigation of ERD/S in divergent thinking tasks (i.e., task commonly employed in the assessment of creativity), which require

participants to generate many original ideas to open problems (e.g., a typical example is the alternate uses task, which asks to think of many unusual uses of everyday objects such as a brick). Taken together, there is evidence that alpha synchronization, especially in frontal and posterior parietal brain regions, is related to creative task demands (Fink et al., 2007, Jausovec, 2000, Jausovec & Jausovec, 2000, Martindale & Hasenfus, 1978, Razumnikova, 2000) and the subjective experience of insight (Jung-Beeman et al., 2004, Sandkühler & Bhattacharya, 2008, Bowden et al., 2005). These alpha effects associated with creativity were sometimes interpreted in terms of low cortical arousal reflecting states of defocused attention and highly associative thinking (Martindale, 1999). Another line of interpretation stresses that alpha synchronization during creative task performance probably indicates high internal processing demands (knowledge and unusual events) and states of heightened internal attention facilitating the re-combination of distantly related semantic information (Fink et al., 2007, 2009a, b). The available evidence on alpha synchronization during active cognitive task performance can thus suggest that alpha synchronization generally reflects high internal processing demands; or, in considering the extensive evidence on alpha synchronization and creativity, it could also be assumed that alpha synchronization indicates cognitive or neural processes specifically related to creative cognition.

5.1.1. Experimental Protocols

Subjects

Thirty-seven 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 subjects have been defined, a group of ATC Experts (40.41 ± 5.54) and a group of ATC Students (23 ± 1.95).

ATM Scenario

The ATCOs have been asked to perform an ATM simulation using a research simulator hosted at ENAC (Figure 76). The experiments have also

been attended by two Pseudo-Pilots (Figure 77) who have interacted with the ATCOs with the aim to simulate and reproduce real communications and to modulate the S-R-K events. The ATM scenario enclosed different levels of difficulty (see paragraph 2.3.2). The traffic complexity has been modulated in terms of number of aircrafts in the controlled sector, number of conflicts and geometry complexity.

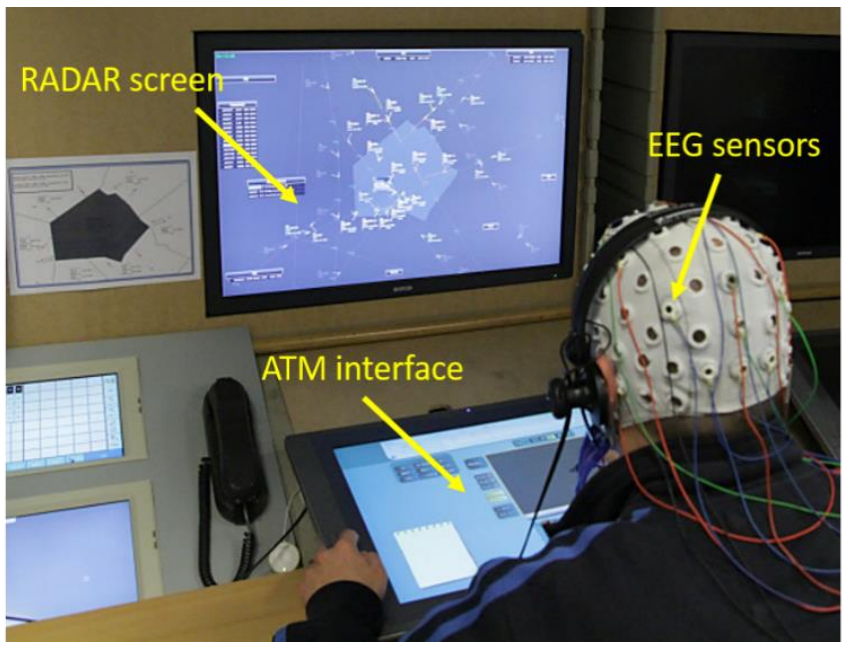

Figure 76. Experimental setup: ATCO working positions developed and hosted at ENAC (Toulouse, France). The ATCO's brain activity has been recorded continuously during the execution of the ATM scenario, and the S-R-K events have been marked in order to recognize them within the EEG recording.

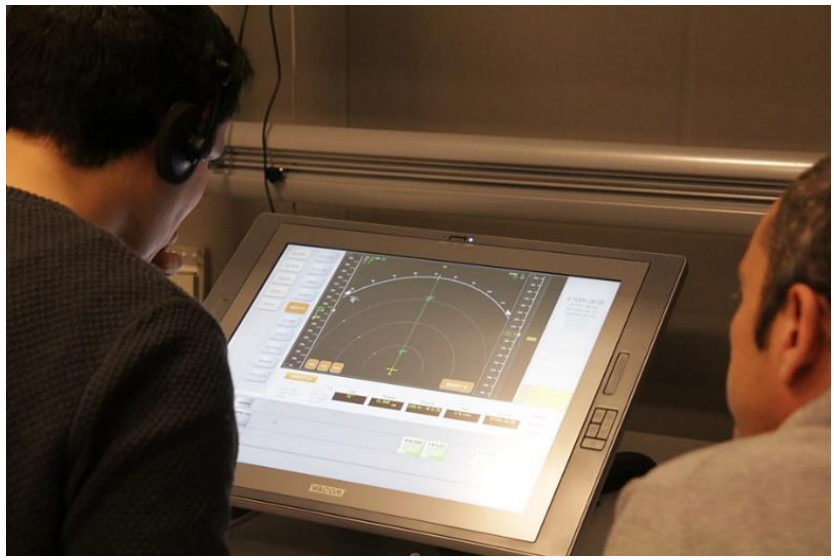

Figure 77. Two professional Pilots have been recruited as *Pseudo-Pilots* with the aim to simulate real flights communications and to reproduce specific S-R-K events within the ATM scenario.

The entire simulation lasted 45 minutes and two triplets of S-R-K events (S1, R1, K1, S2, R2 and K2) have been inserted into the ATM scenario within coherent difficulty conditions (Figure 78).

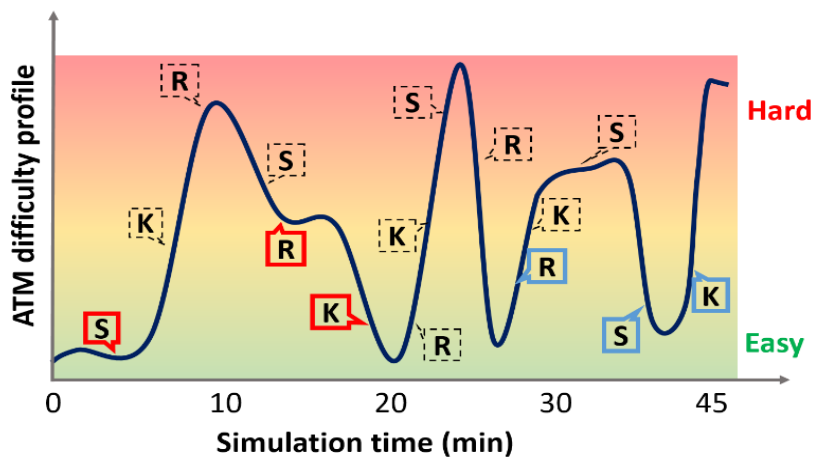

Figure 78. ATM simulation time-line as a function of traffic complexity showing S-R-K events. The simulation lasted 45 minutes and the triplets of S-R-K events have been inserted within coherent difficulty conditions (red, blue and black squares).

S-R-K Events

A *Subject Matter Expert* (SME) from the *Ente Nazionale di Assistenza al Volo* (ENAV, Rome, Italy) has been involved in order to create realistic and not disruptive S-R-K events during the simulation (to be limited in time and not changing the realism of the scenario). The events designed in the air-traffic sample scenario represented an attempt to induce ATCO behaviors associated with S-R-K levels during usual air-traffic conditions. Figure 79 reports the integration of the S-R-K events within the ATM scenario.

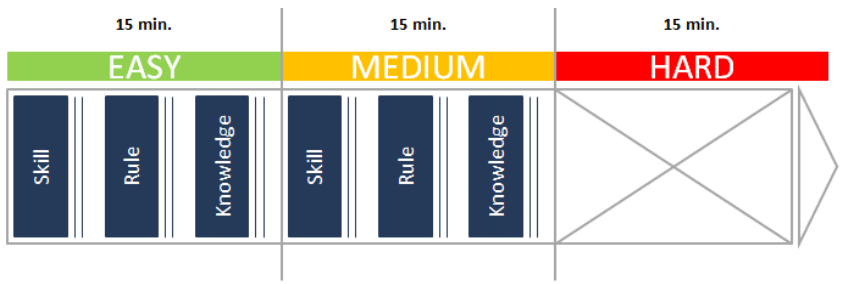

Figure 79. Example of SRK events distribution along the considered ATM scenario. The triplets of S-R-K events have been inserted within coherent difficulty conditions. No events have been inserted in the Hard condition in order to maintain the realism of the simulation.

In particular, six S-R-K events have been generated, two for each level of cognitive control behaviour. The events have not been inserted in the Hard part of the ATM scenarios, since the introduction of external events could increase the difficulty to unacceptable level, generating too much disruption on the controlling activity and possibly invalidating part of the recording and of the realism of the task. On the contrary, in the Easy and Medium parts of the scenarios, three S-R-K events have been inserted, and each of them lasted approximately 30 seconds. The design of the S-R-K events has been the following.

The *Skill* events are basically interactions with the interface, during the task execution. Controllers have been asked to visualize the distance between two aircraft (Distance event) or to display the *Flight Plan Level* (FPL) trajectory of each aircraft present in the controlled sector (Display FPs).

The *Rule* events are mainly control-tasks and conflicts-resolutions, during which controllers are also performing skill-events (interaction with the interface). In the “conflict event”, Controllers had to detect and solve a

conflict by using the menu of the interface and assigning new altitudes and headings. The hypothesis was that conflict detection task represents a familiar situation for ATCOs. Therefore, Controllers should recognize the correct procedures and familiar solutions and then to apply them to solve the conflict.

The *Knowledge* events integrated in the scenario represent unfamiliar and unusual situations. This uncertainty led the Controllers to require time to analyze the situation and to find out the proper procedure to cope with the unexpected event. In other words, the ATCOs initially had to analyze the unusual air-traffic (Knowledge level) and then came back to Rule level for adopting the proper procedures. In the first Knowledge-event, "deviation event", Controllers have been supposed to detect and understand that an aircraft was not following the route filled in the flight plan (FPL). In particular, an alarm has been displayed on the *Human-Machine-Interface* (HMI) Radar with the aim to make the ATCO focusing on the aircraft (a/c), to check its maneuvers and to detect if something was wrong. In the second Knowledge-event (*Unidentified Flying Object* - UFO), the Pseudo-Pilot should report an unknown-traffic detected by the *Traffic Collision Avoidance System* (TCAS) and a TCAS resolution advisory to avoid a mid-air collision. This unknown aircraft has not been displayed on the Controller's radar image, who had been supposed to understand the situation and to ask additional information to the pilot of the considered aircraft. After the avoidance manoeuvre (descent), the Pseudo-Pilot had to ask for his previous flight-level, which never changed on Controller's HMI. The Controller could not observe neither the aircraft responsible of the TCAS advisory nor the implementation of the avoidance manoeuvre. A summary of the S-R-K events is reported in Table 10.

Table 10: Short description of the S-R-K events.

Name of the event	Type of event
	Skill based
Distance	Controllers are asked to measure the distance between two aircraft using the validate tool.
Display FPs	The SME expert triggered the event giving to the controller a paper with the instruction to visualized the Flight Plan (FPL) trajectory of each aircraft assumed. Controller have to use the pie menu and select the route option for each aircraft.
	Rule based
Conflict	Controllers have to detect and solve a conflict, using the pie menu to assign new altitudes and headings.
	Knowledge based
Deviation	Controllers are asked to face with unexpected situation, in which an aircraft changes its route respect to the cleared FP. Controllers have to detect this manoeuvre and check the a/c flight plan on the HMI, to evaluate if it is correct or if something is going wrong. .
UFO	Controllers are asked to verify an unknown traffic detected by the Traffic Collision Avoidance System (TCAS), to avoid a mid air collision. After the avoidance manoeuvre pilots report a different FL respect to what provided by the HMI

Physiological Signals Recording and Pre-Processing

The neurophysiological signals have been recorded by the digital monitoring *BEmicro* system (EBNeuro system). The thirteen EEG channels (FPz, F3, Fz, F4, AF3, AF4, P3, Pz, P4, POz, O1, Oz, O2) and the EOG channel have been collected with a sampling frequency of 256 (Hz). All the EEG electrodes have been referenced to both the earlobes, grounded to the mastoids, and the impedances of the electrodes were kept below 10 (kΩ). The bipolar electrodes for the EOG have been positioned vertically over the left eye. The acquired EEG signals have been digitally band-pass filtered by a 5th order Butterworth filter (low-pass filter cut-off frequency: 30 (Hz), high-pass filter cut-off frequency: 1 (Hz)) and the EOG signal has been used to remove eyes-blink artifacts from the EEG by using the Gratton (1983)

method. For other sources of artifacts, specific procedures of the EEGLAB toolbox, based on threshold methods have been used (Delorme and Makeig, 2004). In particular, three methods have been used: the threshold criteria, the trend estimation and the sample-to-sample difference. In the threshold criteria the EEG epoch has been marked as "artifact" if the EEG amplitude was higher than ±100 (μV). In the trend estimation, the EEG epoch has been interpolated in order to check the slope of the trend within the considered epoch. If such slope was higher than 3 (non-physiological variation), the considered epoch has been marked as "artifact". The last check calculated the difference between consecutive EEG samples. If such difference, in terms of amplitude, was higher than 25 (μV), it meant that an abrupt variation (non-physiological) happened, thus it was marked as "artifact". At the end, the EEG epoch marked as "artifact" have been removed with the aim to have a clean EEG dataset from which estimating the brain parameters for the different analyses. The EEG signal has been then segments in 2 second-epochs, shifted of 0.125 seconds, with the aim to have both a high number of observations (see equation 3.7) in comparison with the number of variables (see equation 3.8), and to respect the condition of stationarity of the EEG signal (Elul, 1969). In fact, the latter one is necessary in order to proceed with the spectral analysis of the signal. The *Power Spectral Density* (PSD) has then been estimated by using the *Fast Fourier Transform* (FFT) in the EEG frequency bands defined for each subject by the estimation of the Individual *Alpha Frequency* (IAF) value (Klimesch, 1999). Furthermore, the *Baselines* (brain activity during rest conditions, that is, closed and opened eyes) and the *Reference* (ATCOs looked at the radar screen without reacting, where two non-colliding airplanes have been presented) conditions have been recorded before starting with the ATM simulations.

Power Spectrum Density Analysis

The review of the literature provided useful cues for the definition of a metric to be used for the S-R-K discrimination. The PSD has been estimated for the different brain features (frontal theta, parietal theta, frontal alpha and parietal alpha EEG rhythms), and the analysis of their spectral information has been performed in order to assess which brain features can be used to achieve the proposed goal. The one-way repeated measures ANOVAs have

been performed for each brain rhythm with the *within* factor S-R-K, (3 levels: Skill, Rule and Knowledge), and with the PSD as independent variable. Two different kind of analysis have been ran. The first one-way ANOVAs have been done with the aim to assess the difference between the two groups (*between* factor RANK; 2 levels: Experts and Students). The aims of the two-way ANOVAs were to find out the differences between the corresponding S, R and K levels of the two groups (*within* factor RANK*SRK; 6 levels: Skill, Rule and Knowledge * Experts and Students).

S-R-K Discrimination: Machine Learning Analysis

The classification algorithm asSWLDA (see paragraph 3.1) has been used to select the most relevant brain spectral features to discriminate the three S-R-K cognitive levels. In particular, the algorithm has been calibrated by using the brain areas and rhythms found in the scientific literature described previously (see paragraph 5). In this way, the algorithm has been calibrated with brain features extracted from one triplet of S-R-K events (S1, R1, K1) and then tested on the remaining triplet (S2, R2, K2) and vice-versa. For each testing triplet, the *Area Under Curve* (AUC) values have been calculated of the *Receiver Operating Characteristic* (ROC, Bamber, 1975) by considering couples between S-R-K distributions. The AUC values related to the discrimination accuracy between the three couples of conditions (S vs R, S vs K, R vs K) have been calculated and analyzed for each ATCO. In order to test the effectiveness of the algorithm, for each couple of conditions (S vs R, R vs K, S vs K), I have compared the AUC distributions obtained from the experimental data of all the ATCOs (*Measured AUC*), with the same distributions centered in 0.5 (*Random AUC*), situation corresponding to the *chance level*. An AUC of 0.5 means that the algorithm is not able to discriminate the considered conditions. (S vs R, S vs K, R vs K). On the contrary, if the AUC is higher than 0.5 and lower than 0.7, the classification is good, while if it is higher than 0.7, the classification is optimum, that is, the conditions can be discriminated well. The *Random AUC* distributions have been compared with the *Measured AUC*, by using three two tailed student t-tests ($\alpha = 0.05$), in order to demonstrate the reliability of the algorithm. For the ATC Students, I was unable to perform a three classes analysis, because most of them missed the first Knowledge event (Deviation). In this regard, I have analysed only the

Rule and Skill conditions. In particular, the asSWLDA has been calibrated by using one couple (S1, R1) and the discrimination accuracy (AUC values) has been tested on the remaining couple (S2, R2), and *vice-versa*. As consequence, the AUC distributions (*Measured AUC* and *Random AUC*) have been calculated for the Skill vs Rule comparison.

Comparison Between ATC Experts and Students

Unpaired t-tests ($\alpha = 0.05$) have been performed to compare the discrimination accuracy between the Skill and the Rule conditions (the knowledge event was missed as quoted above) between the two groups (ATC Experts and ATC Students). The hypothesis was that the ATC Students were not skilled as the ATC Experts, so that the Skill events (that should be characterised by completely automated activities) could require a brain activation and attention as for the Rule events. The state classifier has been trained by using one couple (S1, R1) of conditions, and the discrimination accuracy (AUC values) has been tested on the remaining couple (S2, R2) and *vice-versa*, both for the ATC Experts and Students.

5.1.2. Results

Power Spectral Densities (PSDs)

The results of the ANOVA on the frontal theta PSDs (Figure 80) showed that this brain features changed significantly ($p < 10^{-5}$) among the S, R and K events, thus it could be used for the S-R-K discrimination. In particular, the post-hoc test reported the high discriminability of the Skill ($p < 0.00006$) and Knowledge ($p < 0.0002$) events, while the Rule event was similar ($p = 0.3$) to the Knowledge one.

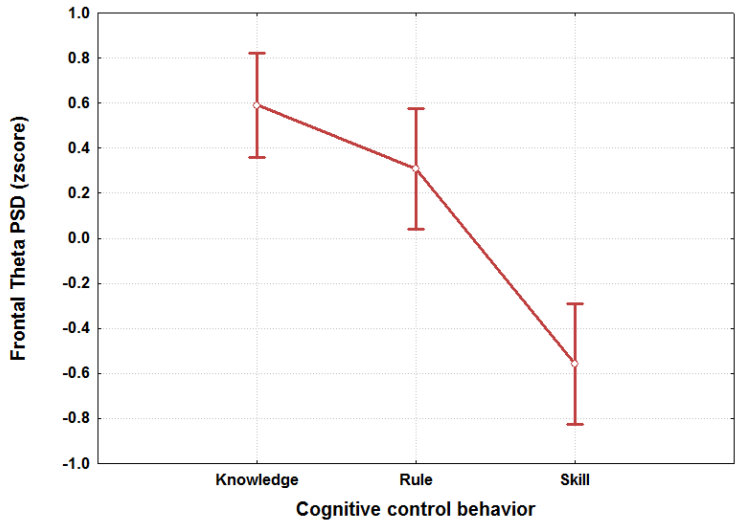

Figure 80. The figure reports the results of the ANOVA analysis on the frontal theta PSD with the factor "SRK" of 3 levels (Skill, Rule and Knowledge) The results showed that such brain feature could be used as S-R-K discriminant brain feature, as its PSD values were significantly different ($p = 0.000001$) between the S, R and K levels.

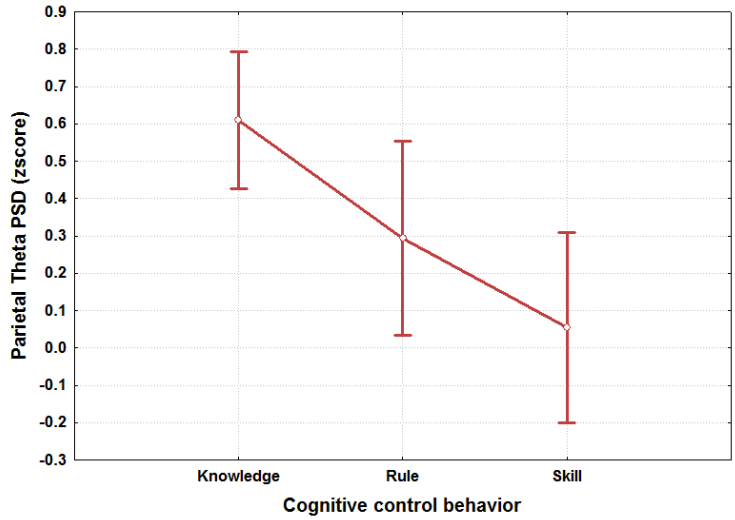

Figure 81. The figure reports the results of the ANOVA analysis on the parietal theta PSD with the factor "SRK" of 3 levels (Skill, Rule and Knowledge) The results showed that such brain feature could be used as S-R-K discriminant brain feature, as its PSD values were significantly different ($p = 0.02092$) between the S, R and K levels.

The ANOVA performed on the parietal theta PSD (Figure 81), demonstrated that it was able to discriminate the S-R-K levels ($F(2, 64) = 4.1104$; $p = 0.02092$). The post-hoc confirmed that the parietal theta PSD could be used to discriminate accurately the Skill level from the Knowledge level ($p < 0.002$). The ANOVA on the frontal alpha PSD reported a significant effect in the S-R-K discrimination ($F(2, 64) = 11.475$; $p = 0.00006$). The post-hoc test demonstrated that the frontal alpha rhythm could be used to differentiate the K and R events from the S one ($p < 0.0002$). On the contrary, the parietal alpha (Figure 83) feature did not show any significant differences between the S, R and K conditions ($F(2, 64) = 2.4478$; $p = 0.9454$), therefore it was not considered in the definition of the S-R-K discrimination metric.

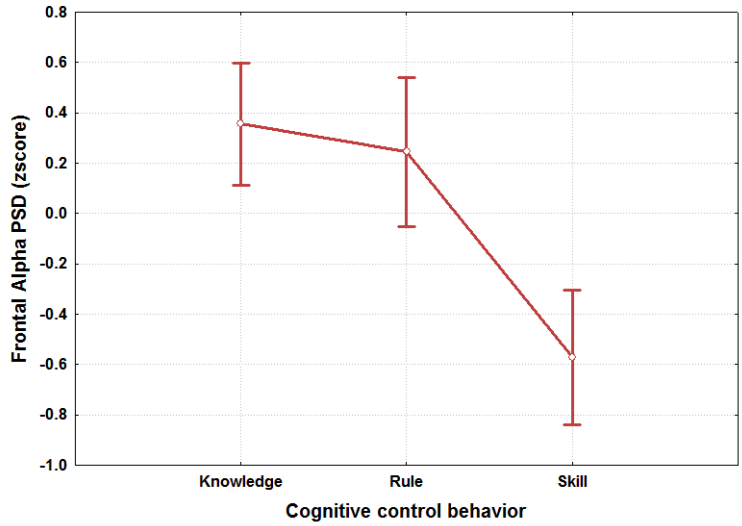

Figure 82. The figure reports the results of the ANOVA analysis on the frontal alpha PSD with the factor "SRK" of 3 levels (Skill, Rule and Knowledge). The results showed that such brain feature could be used as S-R-K discriminant brain feature, as its PSD values were significantly different ($p = 0.00006$) between the S, R and K levels.

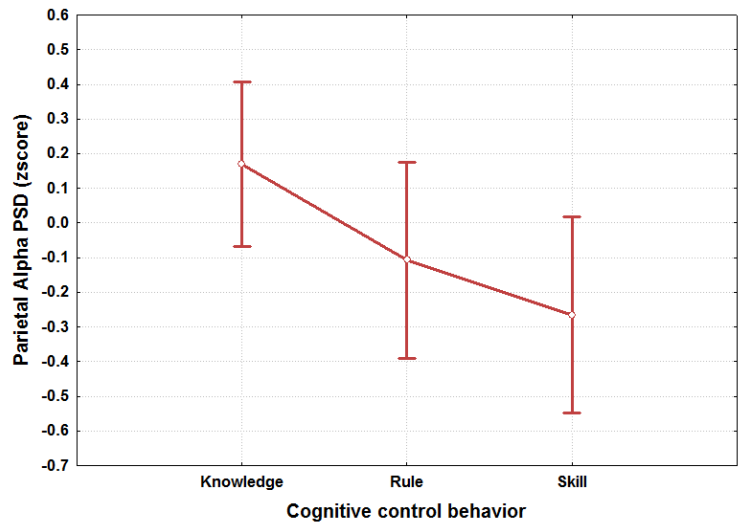

Figure 83. The figure reports the results of the ANOVA analysis on the parietal alpha PSD with the factor "SRK" of 3 levels (Skill, Rule and Knowledge). The results showed that such brain feature could not be used for the S-R-K discriminant, as no significant differences were found ($p = 0.9454$).

Comparison Between ATC Experts and ATC Students

The ANOVAs showed significant differences between the group of ATC Experts and ATC Students for the parietal theta and frontal alpha brain features, both in terms of general activation of the brain rhythms (Figures 84 and 85), and of brain activation during the execution of the same S-R-K events (Figures 86 and 87).

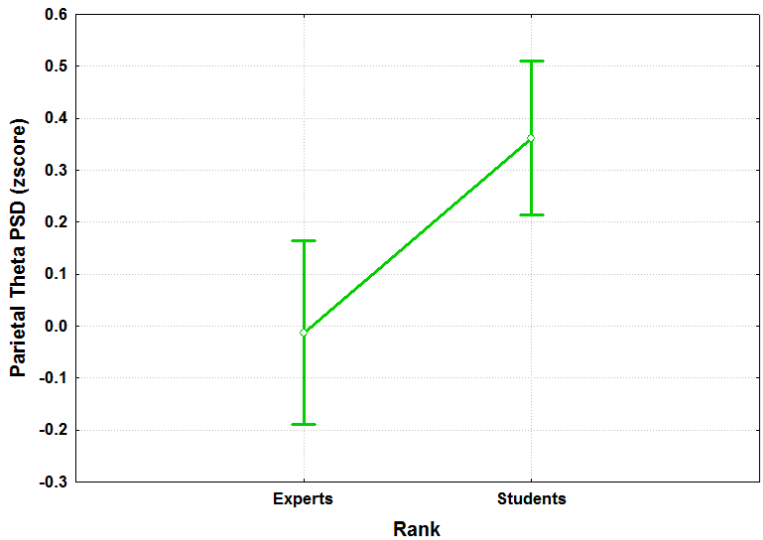

Figure 84. The figure reports the results of the ANOVA analysis on the parietal theta PSD with the factor "RANK" of 2 levels (Experts and Students). The results showed that the two groups were statistically different ($p = 0.0023$) in terms of activation of the parietal theta rhythm when facing the same S, R and K events.

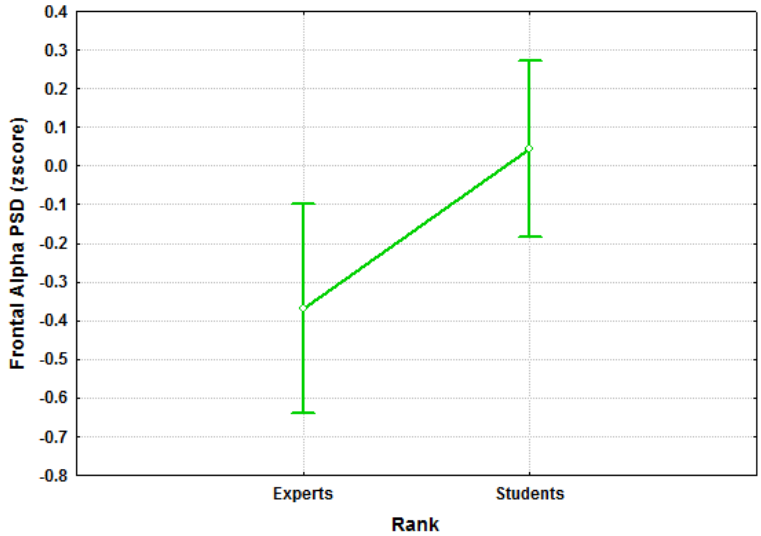

Figure 85. The figure reports the results of the ANOVA analysis on the frontal alpha PSD with the factor "RANK" of 2 levels (Experts and Students). The results showed that the two groups were statistically different ($p = 0.02389$) in terms of activation of the frontal alpha rhythm when facing the same S, R and K events.

In particular, the examined brain features could be used to distinguish the level of expertise (Expert vs Student) with significant reliability (all $p < 0.03$). In fact, the ATC Students showed higher ($F(1, 32) = 10.971$; $p = 0.0023$) parietal theta synchronization and a lower ($F(1, 32) = 5.6248$; $p = 0.02389$) frontal alpha desynchronization than the ATC Experts.

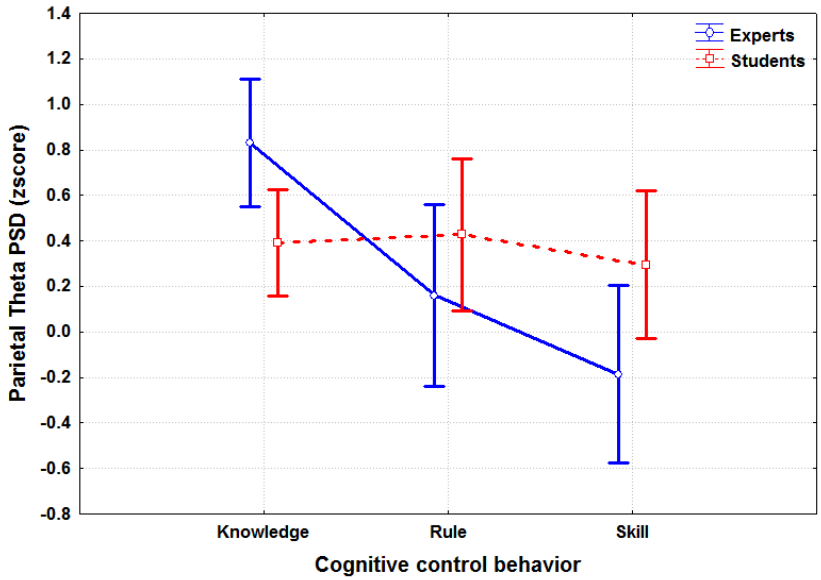

Figure 86. The figure reports the results of the ANOVA analysis on the parietal theta PSD with the factor "RANK*SRK" of 6 levels (Skill, Rule and Knowledge * Experts and Students). The results showed that the same S-R-K events were statistically different ($p < 0.05$), in terms of activation of the parietal theta rhythm, between the group of ATC Experts and ATC Students.

The two-way ANOVAs furnished important information about the different spectral features activations in the different S-R-K events between the ATC Experts (blue line) and ATC Students (red line) in the Figures 86 and 87. In particular, the post-hoc tests showed that the synchronization of the ATC Experts' parietal theta rhythm achieved the lowest value in the Skill event and the highest in the Knowledge event (Figure 86). On the contrary, no differences were found among the S, R and K events for the ATC Students in terms of parietal theta activation (Figure 87). The comparison of the groups highlighted that the *rule* events did not differ ($p > 0.3$), but all the other comparisons were significantly different ($p < 0.05$), and they demonstrated how the ATC Experts required lower activation of the procedural memory than for the ATC Students for all the cognitive control behaviours. To be noted that, since the ATC Students were taught to solve the designed knowledge event (*TCAS event*) by a specific procedure, the difference with respect to the K-Experts was evident, and it achieved a Rule-like activation level. In Figure 87 are reported the results of the two-way

ANOVA performed on the frontal alpha PSDs both of the ATC Experts (blue line) and for the Students (red line). The post–hoc tests highlighted the same trends and differences found for the parietal theta PSDs, that is, significant desynchronization of the ATC Experts' frontal alpha rhythm across the S, R and K events with the highest desynchronization in the S event and the lowest in the K event. The frontal alpha activation between the two groups was not different in the S events ($p > 0.17$), but there were significant differences ($p < 0.05$), in terms of frontal alpha desynchronization, in the R and K events.

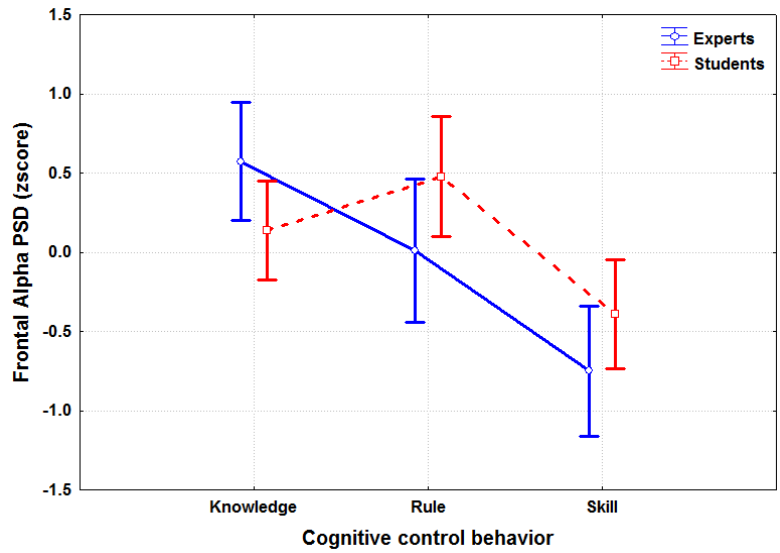

Figure 87. The figure reports the results of the ANOVA analysis on the frontal alpha PSD with the factor "RANK*SRK" of 6 levels (Skill, Rule and Knowledge * Experts and Students). The results showed that the same S-R-K events were statistically different ($p < 0.05$), in terms of activation of the parietal theta rhythm, between the group of ATC Experts and ATC Students.

As quoted above, the ATC Students adopted a procedural resolution for the Knowledge event, therefore the desynchronization of the frontal alpha rhythm was similar to the activation during the Rule event.
No significant differences were found between the two groups for the frontal theta PSD ($p = 019147$).

Machine - learning analysis

As mentioned above, the asSWLDA was trained by using one triplet (S1, R1, K1) and the discrimination accuracy (AUC) values were calculated by testing it on the remain triplet (S2, R2, K2) and *vice-versa*. Referring on the results reported in the previous sections, the parietal theta and frontal alpha EEG rhythms were defined the frequency domain in which the classification model had to select the most significant features.

ATC Experts

The t-tests ($\alpha = 0.05$) showed that all the measured AUC distributions were significantly higher (Figure 88) with respect to the random AUC distributions (all $p < 10^{-5}$).

ATC Students

The tests showed that, as reported for the ATC Experts, also for the ATC Students all the measured AUC distributions were significantly higher (Figure 89) with respect to the random AUC distributions (all $p < 10^{-3}$).

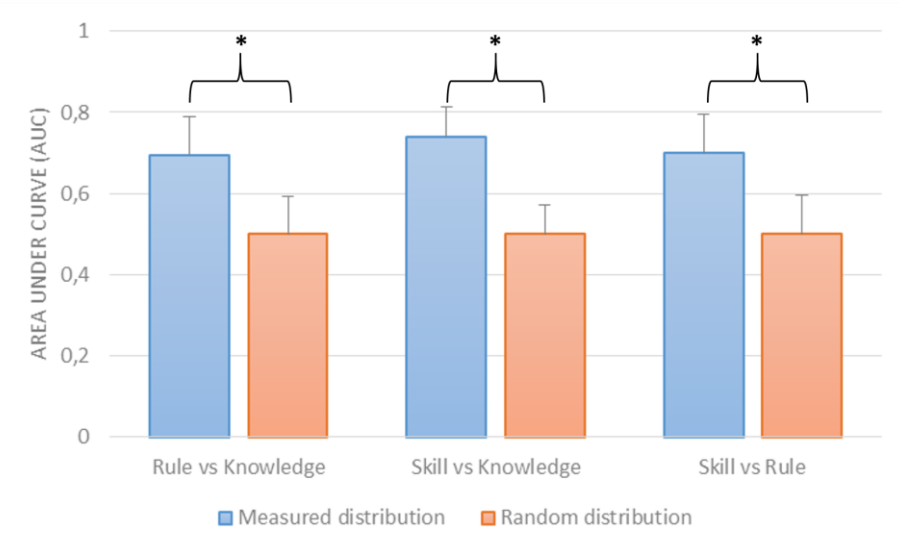

Figure 88. Error bars (CI=.95) related to the Measured AUC distributions and the Random AUC distributions, achieved by the ATC Experts, referred to the discrimination accuracy between the three couples of conditions (S vs R, S vs K, R vs K).

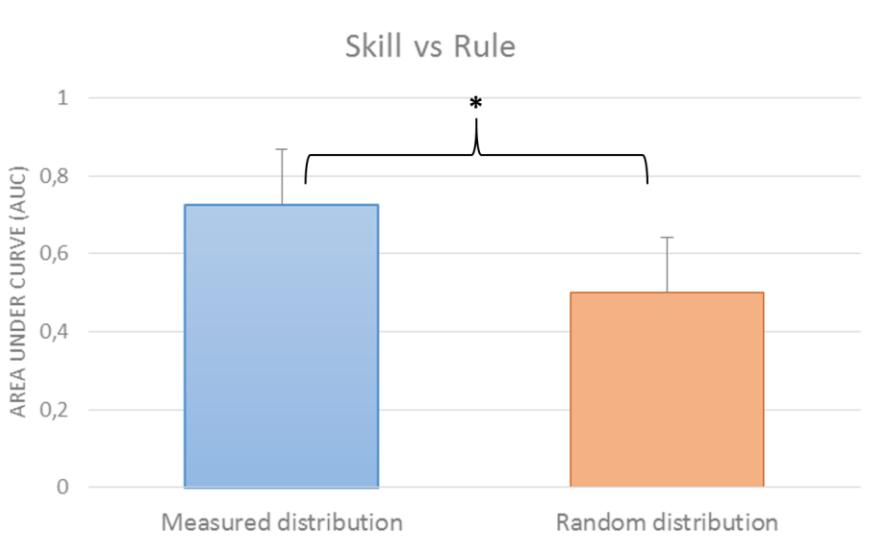

Figure 89. Error bars (CI=.95) related to the Measured AUC distributions and the Random AUC distributions, achieved by the ATC Students, referred to the discrimination accuracy between the Skill and Rule conditions.

Comparison Between ATC Experts and Students

In Figure 90 are reported the AUC values of the considered ATCO groups. The ATC Experts (blue bar) exhibited a higher discrimination accuracy (AUC = 0.8) with respect to the ATC Students (red bar, AUC = 0.72), but not significant (p=0.1). The trend may indicate how the three cognitive control behaviorus were better discriminable for the Experts than for the Students. For example, the *Skill* behaviour of the Students could not be very different from the *Rule* one.

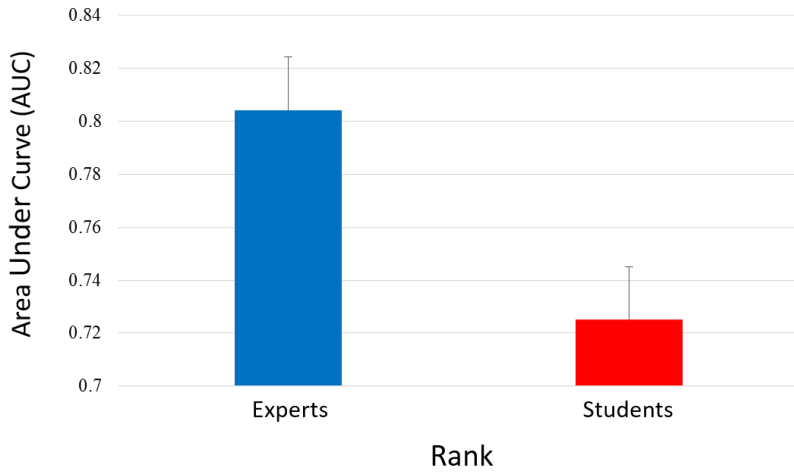

Figure 90. Error bars (CI=.95) related to the discrimination accuracy of the Skill vs Rule condition achieved by the ATC Experts (blue bar) and ATC Students (red bar), referred to the discrimination accuracy between the Skill and Rule conditions.

5.1.3. Discussions

Several studies tried to model the different information processing procedures based on a Skill, Rule or Knowledge control cognitive levels from a human factor point of view. Despite such studies, there are not evidences in which these control cognitive levels have been taken into account from a neurophysiological point of view. For example, by considering the EEG variations between the different S–R-K levels. An extensive literature review has been done in order to identify all the cognitive processes related to the S, R and K processes, and the related EEG variations. The selected brain features for the definition of the metric for the S-R-K discrimination were the parietal theta and frontal alpha EEG rhythms. The former is correlated to the procedural memory, and more is the memory consolidation, more its synchronization is over the parietal brain areas, while the latter is correlated to the attention level, and it has been demonstrated that more is the attention, higher is the desynchronization of the alpha activity over the frontal brain areas. The strength of such the metric is that it could be applied to assess simultaneously the S-R-K levels and the mental workload, since they use

different brain features (see paragraph 3 and 5). In fact, the hypothesis was that, if the classifier is calibrated by using brain features different from those used for the workload assessment (frontal theta, parietal alpha), and the SRK conditions classification remains high, it would be possible to classify different workload levels and SRK conditions at the same time during the execution of specific tasks. The analysis of the parietal theta and frontal alpha PSDs across the S-R-K levels confirmed that these brain features were able to discriminate significantly both the S-R-K levels (Figures 80 and 81) and the differences between the group of ATC Experts and ATC Students, in terms of expertise (Figures 83 and 84). Those brain features were also used to calibrate the asSWLDA classification algorithm, and to test it about the possibility to differentiate in real-time the three cognitive control levels during the execution of a realistic ATM scenario executed by professional ATCOs. The asSWLDA was able to discriminate significantly the S-R-K conditions for the ATC Experts ($p < 10^{-5}$), and the S-R conditions for the ATC Students ($p<10^{-3}$), and all the measured AUC distributions were significantly higher (Figure 87 and 88) than the random AUC distributions (all $p < 10^{-3}$).

5.1.4. Conclusions

In this study, it has been demonstrated that it was possible to assess, with a high reliability, the ATCO's cognitive control level (S-R-K) by using a specific metric estimated from the brain activity. The proposed method has been proven to be reliable in realistic ATM settings. The results showed that: 1) the proposed metric (parietal theta and frontal alpha) should be a valid solution to investigate and objectively analyze the different cognitive control behaviours under which the user is performing a task; and 2) the asSWLDA was able to discriminate significantly the S-R-K conditions for both the groups of ATCOs. Due to the need of inserting S-R-K events into realistic ATM tasks, it was not possible to create "pure" S-R-K behaviours. The aim of future S-R-K studies is to design experimental protocols by using controlled tasks in order to better simulate and induce pure skill, rule and knowledge reaction to the subjects.

6. General Conclusions

The main objectives of the PhD project were the definition and validation of neurometrics and methodologies for the workload and training assessment in operational environments. In fact, objective investigations and deeper comprehension and evaluation of human factors, such as workload, training and cognitive control behavior, in working operational environments can avoid and limit the occurrence of errors and improve the level of safety.

Several experimental protocols have been designed in order to achieve such results, both in controlled setting (Lab) and in realistic environment (ATC control room) with professional personnel (*Air Traffic Controllers*).

A modified version of the SWLDA algorithm was defined and validated in realistic environment: the asSWLDA (Patent I, 2105). This algorithm has the capability to include in the classification model only those features strictly related to the considered cognitive phenomenon, thus avoiding the overfitting. In addition, it does not suffer of any decreasing in discrimination accuracy without recalibratingthe algorithm within a month. Another important advantage, with respect to the standard SWLDA implementation, is that the asSWLDA uses a significant lower number of EEG channels (among the available ones) to select the most significant features necessary to track and quantify the considered cognitive phenomenon. Consequently, the advantages of the asSWLDA are very practical, since both the number of sensors (EEG channels), thus the invasiveness of the recording system, and the time needed to run the experiments are significantly reduced (no calibration upto a month). In this way, the proposed method could be easily employed in working operational environments.

Furhthermore, such characteristics make the proposed algorithm and methodologies very useful to assess the learning progress of the user across a specific training period (i.e. 3 weeks), and the cognitive control behaviour under which the user is executing the considered tasks. In particular, as described in the introduction of chapter 4, the main limitations in the training assessment are the lack of objective information regarding the

cognitive activations within and between consecutive training sessions. The results demonstrated that the proposed methodologies could overcome such limitations and provide important information to better evaluate the user's learning progress, and efficiently tailoring the training programms.

Concerning the evaluation of cognitive control behaviour (chapter 5), no neurophysiological metrics have been proposed in the scientific literature. During the selection of personnel, it might be important to know how the users can perform the proposed task. If they can achieve high performance (100%), it will not mean they can do it automatically or by involving few brain resources. By the standard evaluation methods, it is not possible to have this kind of information. The results of my study, demonstrated that the proposed neurophysiological metric can be used to discriminated the cognitive control behaviour under which the user is accomplishing the task. Therefore, the users evaluation can be done more accurately.

To sum up, the work arose from the PhD returned metrics and methodologies which overcame most of the limitations affecting the use of neurophysiological measures and analysis of the workload and training in real operational environments. In particular, the results showed that:

I. The asSWLDA significantly outperformed the standard implementation of the SWLDA, in fact:
    i. the asSWLDA reduced significantly the EEG channels required for a proper analysis of the considered cognitive phenomenon;
    ii. No manual triggering and settings of the parameters of the algorithm to obtain the correct behaviour and functionality of it;
    iii. the performance of the asSWLDA algorithm remained stable and reliable upto a month without any recalibration of the algorithm.

II. The asSWLDA was able to significantly differentiate three mental workload levels, both in controlled environment (Lab) and in operational settings (ATC room).

III. The proposed methodology could supply quantitative and objective information regarding the training level of the subjects in order to support the Instructor for a better management of the training schedule and program.

IV. The neurometric defined for the characterization of the cognitive control behaviours (skill, rule and knowledge) was a valid solution to objectively investigate and analyze the expertise of the subjects, and it could be employed for innovative personnel selections or customization of training program.

These evidences represent a promising step forward in the analysis of human behaviours and they demonstrate the possibility of developing a device able to evaluate, also in real-time, the cognitive engagement of the operators during the working activities in real operational environments. This tool could be used to improve complex systems design and to enhance the capability to anticipate errors. For example, *Realistic Interface Design* (EID, Vicente and Rasmussen, 1992; Pawlak and Vicente, 1996; Vicente, 2002) uses the SRK framework as an analytical tool for the design of interfaces in complex sociotechnical, real-time, and dynamic systems. Another possible use might be an online tool for triggering *Adaptive Automations* (AA, Byme and Parasuraman, 1996; Parasuraman, 1996; Uebbing-Rumke et al., 2012; Wilson et al., 2000), in which the systems behaves depending on the Operator's current level of cognitive control.
In this regard, an H2020 - SMEINST project has been submitted with the aim to realize and to commercialize a device with the characteristics developed during my PhD and described above.

References

Aberg, L., Rimmö, P.A., 1998. Dimensions of aberrant driver behaviour. Ergonomics 41, 39–56. doi:10.1080/001401398187314.

Aguinis H. and Kraiger K. (2009): «Benefits of Training and Development for Individuals and Teams, Organizations, and Society», *Annu. Rev. Psychol.*, vol. 60, n. 1, pagg. 451–474, 2009.

Albouy, G., Sterpenich, V., Vandewalle, G., Darsaud, A., Gais, S., Rauchs, G., et al. (2013). Interaction between hippocampal and striatal systems predicts subsequent consolidation of motor sequence memory. PLoS ONE, 8(3), e59490.

Albouy, G., Sterpenich, V., Balteau, E., Vandewalle, G., Desseilles, M., Dang-Vu, T., et al. (1995). The distribution of EEG frequencies in REM and NREM sleep stages in healthy young adults. Sleep, 18(5), 334.

Allport A. (1988): What concept of consciousness? In A. Marcel and E. Bisiach eds Consciousness in contemporary science Oxford: Oxford University Press.

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

Anderson JR (1993): Rules of the Mind. Psychology Press. 320 p.

Anderson, K. L., Rajagovindan, R., Ghacibeh, G. A., Meador, K. J., & Ding, M. (2010). Theta oscillations mediate interaction between prefrontal cortex and medial temporal lobe in human memory. Cerebral Cortex, 20(7), 1604–1612.

Arden, R., Chavez, R. S., Grazioplene, R., & Jung, R. E. (2010). Neuroimaging creativity: A psychometric review. Behavioral Brain Research, 214, 143–156.

Aricò, P., Borghini, G., Di Flumeri, G., Colosimo, A., Graziani, I., Imbert, J.P., Granger, G., Benhacene, R., Terenzi, M., Pozzi, S., Babiloni, F., 2015a. Reliability over time of EEG-based mental workload evaluation during Air

Traffic Management (ATM) tasks. Conf. Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Conf. 2015, 7242 – 7245. doi:10.1109/EMBC.2015.7320063.

Aricò, P., Borghini, G., Graziani, I., Imbert, J.P., Granger, G., Benhacene, R., Pozzi, S., Napolitano, L., Di Flumeri, G., Colosimo, A., Babiloni, F., 2015b. ATCO: Neurophysiological Analysis Of The Training And Of The Workload. Ital. J. Aerosp. Med. 1.

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

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

Ashby, F.G., Turner, B.O., and Horvitz, J.C. (2010). Cortical and basal ganglia contributions to habit learning and automaticity. Trends Cogn. Sci. (Regul. Ed.) 14, 208–215.

Astolfi L, De Vico Fallani F, Cincotti F, Mattia D, Marciani MG, Bufalari S, Salinari S, Colosimo A, Ding L, Edgar JC, Heller W, Miller GA, He B, Babiloni F (2007): Imaging functional brain connectivity patterns from high-resolution EEG and fMRI via graph theory. Psychophysiology 44: 880–893.

Astolfi L, Cincotti F, Mattia D, Marciani MG, Baccalà LA, De Vico Fallani F, Salinari S, Ursino M, Zavaglia M, Babiloni F (2006): Assessing cortical functional connectivity by partial directed coherence: simulations and application to real data. IEEE Trans. Biomed. Eng 53: 1802–1812.

Ayaz, H., Onaral, B., Izzetoglu, K., Shewokis, P.A., McKendrick, R., Parasuraman, R., 2013. Continuous monitoring of brain dynamics with functional near infrared spectroscopy as a tool for neuroergonomic research: empirical examples and a technological development. Front. Hum. Neurosci. 7. doi:10.3389/fnhum.2013.00871.

Ayaz, H., Shewokis, P.A., Bunce, S., Izzetoglu, K., Willems, B., Onaral, B., 2012. Optical brain monitoring for operator training and mental workload assessment. NeuroImage 59, 36–47. doi: 10.1016/j.neuroimage.2011.06.023.

Baars, B. J. (1997). Some essential differences between consciousness and attention, perception, and working memory. *Consciousness and cognition*, *6*(2), 363-371.

Babiloni F, Cincotti F, Babiloni C, Carducci F, Mattia D, Astolfi L, Basilisco A, Rossini PM, Ding L, Ni Y, Cheng J, Christine K, Sweeney J, He B (2005): Estimation of the cortical functional connectivity with the multimodal integration of high-resolution EEG and fMRI data by directed transfer function. NeuroImage 24: 118–131.

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

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

Barachant A, Andreev A, Congedo M (2013a): The Riemannian Potato: an automatic and adaptive artifact detection method for online experiments using Riemannian geometry. In: Proceedings of TOBI Workshop IV. Sion, Suisse. p. 19–20.

Barachant A, Bonnet S, Congedo M, Jutten C (2013b): Classification of covariance matrices using a Riemannian-based kernel for BCI applications. Neurocomputing 112: 172–178.

Barach P. and Weingart M. (2004): «Trauma team performance», in *Trauma: Resuscitation, anesthesia, surgery, and critical care New York: Dekker, Inc*, 2004.

Barach P. and Small S. D. (2000): «Reporting and preventing medical mishaps: lessons from non-medical near miss reporting systems», *BMJ*, vol. 320, n. 7237, pagg. 759–763, mar. 2000.

Battaglia, F. P., Benchenane, K., Sirota, A., Pennartz, C., & Wiener, S. I. (2011). The hippocampus: Hub of brain network communication for memory. Trends in Cognitive Sciences, 15(7), 310–318.

Bekhet, A. K., & Zauszniewski, J. A. (2012). Methodological triangulation: an approach to understanding data. *Nurse Researcher, 20*(2), 40-43.

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

Benchenane, K., Peyrache, A., Khamassi, M., Tierney, P. L., Gioanni, Y., Battaglia, F. P., et al. (2010). Coherent theta oscillations and reorganization of spike timing in the hippocampal–prefrontal network upon learning. Neuron, 66(6), 921–936.

Bernstein LJ, Beig S, Siegenthaler AL, Grady CL (2002) The effect of encoding strategy on the neural correlates of memory for faces. Neuropsychologia 40:86--98.

Berka C (2011): Exploring Subjective Experience during Simulated Reality Training with Psychophysiological Metrics. Marine Corp Warfighting Laboratory Workshop. Physiological Metrics of Immersion. Advanced Brain Monitoring, Inc. 12 p.

Berka, C., Levendowski, D.J., Lumicao, M.N., Yau, A., Davis, G., Zivkovic, V.T., Olmstead, R.E., Tremoulet, P.D., Craven, P.L., 2007. EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks. Aviat. Space Environ. Med. 78, B231–244.

Berka, C., Levendowski, D.J., Cvetinovic, M.M., Petrovic, M.M., Davis, G., Lumicao, M.N., Zivkovic, V.T., Popovic, M.V., Olmstead, R., 2004. Real-Time Analysis of EEG Indexes of Alertness, Cognition, and Memory Acquired With a Wireless EEG Headset. Int. J. Hum.-Comput. Interact. 17, 151–170. doi:10.1207/s15327590ijhc1702_3.

Birbaumer N, Schmidt RF. (1996). Biologische Psychologie. Springer-Lehrbuch (German Edition).

Blankertz, B., Tangermann, M., Vidaurre, C., Fazli, S., Sannelli, C., Haufe, S., Maeder, C., Ramsey, L., Sturm, I., Curio, G., Müller, K.-R., 2010. The Berlin Brain–Computer Interface: Non-Medical Uses of BCI Technology. Front. Neurosci. 4. doi:10.3389/fnins.2010.00198.

Blinowska, K., & Durka, P. (2006). Wiley Encyclopedia of Biomedical Engineering, chapter Electroencephalography (EEG).

Borghini, G., Aricò, P., Di Flumeri, G., Salinari, S., Colosimo, A., Bonelli, S., Napoletano, L., Ferreira, A., Babiloni, F., (2015a). Avionic Technology Testing by using a Cognitive Neurometric Index: A Study with Professional Helicopter Pilots. Conf. Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Conf. 2015, 6182 – 6185. doi:10.1109/EMBC.2015.7319804.

Borghini, G., Aricò, P., Graziani, I., Salinari, S., Sun, Y., Taya, F., Bezerianos, A., Thakor, N.V., Babiloni, F., (2015b). Quantitative Assessment of the Training Improvement in a Motor-Cognitive Task by Using EEG, ECG and EOG Signals. Brain Topogr. doi:10.1007/s10548-015-0425-7.

Borghini, G., Vecchiato, G., Aricò, P., Graziani, I., Maglione, A. G., Cherubino, P., & Babiloni, F. (2015c). Investigation of the effect of EEG-BCI on the simultaneous execution of flight simulation and attentional tasks. *Medical & biological engineering & computing*.

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

Borghini, G., Astolfi, L., Vecchiato, G., Mattia, D., Babiloni, F., (2012a). Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. Neurosci. Biobehav. Rev. 44, 58–75. doi:10.1016/j.neubiorev.2012.10.003.

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

Borghini G, Isabella R, Vecchiato G, Toppi J, Astolfi L, Caltagirone C, Babiloni F (2012c): A Flight history from a cognitive point of view: Novices versus Experts. Ital. J. Aerosp. Med 5: 34-47.

Bowden, E. M., Jung-Beeman, M., Fleck, J., & Kounios, J. (2005). New approaches to demystifying insight. Trends in Cognitive Sciences, 9, 322–328.

Braver T.S., Cohen J.D. (2001). Working memory, cognitive control, and the prefrontal cortex: computational and empirical studies, Cogn. Process. 2 (2001) 25–55.

Brennen.S.D. (1992). An Experimental Report on Rating Scale Descriptor Sets for the Instantaneous Self Assessment (ISA) Recorder, DRA/TM/CAD5/92017. Defence Research Agency. Portsmouth.

Brodmann K (2006). Localisation in the Cerebral Cortex. Springer (Ed.). 150 p.

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

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

Buckner R.L. (1996). Beyond HERA: contributions of specific prefrontal brain areas to long-term memory retrieval, Psychon. Bull. Rev. 3 (1996) 149–158.

Bunce, S.C., Izzetoglu, K., Ayaz, H., Shewokis, P., Izzetoglu, M., Pourrezaei, K., Onaral, B., (2011). Implementation of fNIRS for Monitoring Levels of Expertise and Mental Workload, in: Schmorrow, D.D., Fidopiastis, C.M. (Eds.), Foundations of Augmented Cognition. Directing the Future of Adaptive Systems. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 13–22.

Buzsáki, G. (2005). Theta rhythm of navigation: Link between path integration and landmark navigation, episodic and semantic memory. Hippocampus, 15, 827–840.

Buzsáki G. (2002). Theta oscillations in the hippocampus. Neuron., 33:325-40.

Byme E. A., Parasuraman R. (1996). “Psychophysiology and Adaptive Automation”, Bilogical psychology, Vol. 42, No. 3, 1996, pp 249-268.

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

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

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

Campbell J. P. (1971). «Personnel Training and Development», *Annu. Rev. Psychol.*, vol. 22, n. 1, pagg. 565–602, 1971.

Cantero JL, Atienza M, Stickgold R, Kahana MJ, Madsen JR, Kocsis B. (2003). Sleep-dependent theta oscillations in the human hippocampus and neocortex. J Neurosci., 23:10897-903.

Carter C.S., Botvinick M.M., Cohen J.D. (1999). The contribution of the anterior cingulate cortex to executive processes in cognition, Rev. Neurosci. 10 (1999) 49– 57.

Castor M., Hanson E., Svensson E., Nählinder S., LeBlaye P., MacLeod I., Wright N., Alfredson J., Agren L., and Berggren P. (2003). «GARTEUR Handbook of mental workload measurement», Group for Aeronautical Research and Technology in Europe, Flight Mechanics Action Group FM AG13, Technical chapter 145, 2003.

Chandrasekaran V., Jordan M. I. (2013). Proc. Natl. Acad. Sci. U.S.A. 110, E1181–E1190 (2013).

Chauvette, S. (2013). Slow-wave sleep: Generation and propagation of slow waves, role in long-term plasticity and gating. Doctoral dissertation, Université Laval.

Chein, J. M., & Schneider, W. (2005). Neuroimaging studies of practice-related change: fMRI and meta-analytic evidence of a domain-general control network for learning. *Cognitive Brain Research*, *25*(3), 607-623.

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

Cohen J.D., Braver T.S., O'Reilly R.C. (1996). A computational approach to prefrontal cortex, cognitive control and schizophrenia: recent developments and current challenges, Philos. Trans. R. Soc. Lond., B Biol. Sci. 351 (1996) 1515– 1527.

Colligan L, Potts HWW, Finn CT, Sinkin RA (2015). Cognitive workload changes for nurses transitioning from a legacy system with paper documentation to a commercial electronic health record. International Journal of Medical Informatics, in press, accepted manuscript. doi:10.1016/j.ijmedinf.2015.03.003.

Comstock JR, Arnegard RJ (1992). MATB - Multi-Attribute Task Battery for human operator workload and strategic behavior research. NASA Technical Memorandum 104174.114 p.

Cooper, N. R., Croft, R. J., Dominey, S. J. J., Burgess, A. P., & Gruzelier, J. H. (2003). Paradox lost? Exploring the role of alpha oscillations during externally vs. internally directed attention and the implications for idling and inhibition hypotheses. International Journal of Psychophysiology, 47, 65–74.

Corbetta, M. and Shulman, G. L. (2002). Controls of Goal-Directed and Stimulus-Driven Attention in the Brain. Nature Neuroscience, 3(3), 201-215.

Corbetta M., Shulman G.L., Miezin F.M., Petersen S.E. (1995). Superior parietal cortex activation during spatial attention shifts and visual feature conjunction, Science 270 (1995) 802–805.

Costa G. (1993). «Evaluation of workload in air traffic controllers», *Ergonomics*, vol. 36, n. 9, pagg. 1111–1120, set. 1993.

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. Augment. Cogn. Past Present Future.

Crick F. (1984). Function of the thalamic reticular complex: the searchlight hypothesis. Proc Natl Acad Sci USA 81: 4586–4590, 1984.

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

Davidson, R. J., Schwartz, G. E., & Shapiro, D. (Eds.). (2013). *Consciousness and Self-regulation: Advances in Research and Theory* (Vol. 4). Springer Science & Business Media.

Dayan, E., & Cohen, L. G. (2011). Neuroplasticity subserving motor skill learning. *Neuron*, *72*(3), 443-454.

Decades, N.R.C. (US) C. on M. and I.M. for E.N. and C.R. in the N.T., 2008. Emerging Areas of Cognitive Neuroscience and Neurotechnologies.

Decatur S., Goldreich O., Ron D. (2000). SIAM J. Comput. 29, 854–879 (2000).

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

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

Desimone R., Duncan J. (1995). Neural mechanisms of selective visual attention, Annu. Rev. Neurosci. 18 (1995) 193–222.

De Waard, (1996). «The measurement of drivers' mental workload», University of Groningen, Traffic Research Centre, Netherlands, 1996.

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

Dietrich, A., & Kanso, R. (2010). A review of EEG, ERP, and neuroimaging studies of creativity and insight. Psychological Bulletin, 136, 822–848.

Ding L, Lai Y, He B (2005). Low resolution brain electromagnetic tomography in a realistic geometry head model: a simulation study. Phys Med Biol 50: 45–56.

Dooley C. (2009). The Impact of Meditative Practices on Physiology and Neurology: A Review of the Literature. Scientia Discipulorum, p. 35-59.

Doppelmayr M, Finkenzeller T, Sauseng P (2008). Frontal midline theta in the pre-shot phase of rifle shooting: differences between experts and novices. Neuropsychologia 46: 1463–1467.

Doyon, J., and Benali, H. (2005). Reorganization and plasticity in the adult brain during learning of motor skills. Curr. Opin. Neurobiol. 15, 161–167.

Draper, N. R., and Smith, H. (2014). Applied regression analysis. John Wiley & Sons.

Draper, N.R., (1998). Applied regression analysis. Commun. Stat. - Theory Methods 27, 2581–2623. doi:10.1080/03610929808832244.

Dudai, Y. (2004). The neurobiology of consolidations, or, how stable is the engram? Annual Review of Psychology, 55, 51–86.

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

East J. A. (2000): «Feature Selection for Predicting Pilot Mental Workload», mar. 2000.

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

Elliot AJ, Payen V, Brisswalter J, Cury F, Thayer JF (2011). A subtle threat cue, heart rate variability, and cognitive performance. Psychophysiology 48: 1340–1345.

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

Erickson, K. I., Colcombe, S. J., Wadhwa, R., Bherer, L., Peterson, M. S., Scalf, P. E., & Kramer, A. F. (2007). Training-induced functional activation changes in dual-task processing: an FMRI study. *Cerebral Cortex*, *17*(1), 192-204.

Fairclough S. H., Venables L., and Tattersall A. (2005). «The influence of task demand and learning on the psychophysiological response», *International Journal of Psychophysiology*, vol. 56, n. 2, pagg. 171–184, mag. 2005.

Fell, J., & Axmacher, N. (2011). The role of phase synchronization in memory processes. Nature Reviews Neuroscience, 12(2), 105–118.

Feyer, A., Williamson, A.M., (2011). Human Factors in Accident Modelling. Encycl. Occup. Health Saf.

Fielder J (2011): Electroencephalogram (EEG) Study of Learning Effects across Addition Problems.PEBL Technical Report Series.

Fink, A., Grabner, R. H., Benedek, M., Reishofer, G., Hauswirth, V., Fally, M., et al. (2009a). The creative brain: Investigation of brain activity during creative problem solving by means of EEG and fMRI. Human Brain Mapping, 30, 734–748.

Fink, A., Graif, B., & Neubauer, A. C. (2009b). Brain correlates underlying creative thinking: EEG alpha activity in professional vs. novice dancers. NeuroImage, 46, 854–862.

Fink, A., Benedek, M., Grabner, R. H., Staudt, B., & Neubauer, A. C. (2007). Creativity meets neuroscience: Experimental tasks for the neuroscientific study of creative thinking. Methods, 42, 68–76.

Fisher, R. A. (1936). The Use of Multiple Measurements in Taxonomic Problems. Annals of Eugenics 7 (2): 179–188. doi:10.1111/j.1469-1809.1936.tb02137.x. hdl:2440/15227.

Fisher, R.A., (1921). On the "Probable Error" of a Coefficient of Correlation Deduced from a Small Sample.

Fitts P. M. and Posner M. I. (1962). *Human Performance.* Brooks Cole, 1962.

Foxe, J.J., Simpson, G.V., and Ahlfors, S.P. (1998). Parieto-occipital approximately 10 Hz activity reflects anticipatory state of visual attention mechanisms. Neuroreport 9: 3929–3933.

Freeman W. J. (1991). The physiology of perception. Scientif. Amer. 1991; 26:78–85.

Fu, K.M., Foxe, J.J., Murray, M.M., Higgins, B.A., Javitt, D.C., and Schroeder, C.E. (2001).Attention-dependent suppression of distracter visual input can be crossmodally cued as indexed by anticipatory parieto-occipital alpha-band oscillations. Brain Res Cogn Brain Res 12: 145–152.

Fuchs, S., Hale, K.S., Stanney, K.M., Juhnke, J., Schmorrow, D.D., (2007). Enhancing Mitigation in Augmented Cognition. J. Cogn. Eng. Decis. Mak. 1, 309–326. doi:10.1518/155534307X255645.

Fukuda K (2001). Eye blinks: new indices for the detection of deception. Int J Psychophysiol Off J Int Organ Psychophysiol 40: 239–245.

Fukunaga, K. (2013). *Introduction to statistical pattern recognition*. Academic press.

Galán FC, Beal CR (2012). EEG Estimates of Engagement and Cognitive Workload Predict Math Problem Solving Outcomes. In: Masthoff, J., Mobasher, B., Desmarais, M.C., Nkambou, R. (Eds.), User Modeling, Adaptation, and Personalization, Lecture Notes in Computer Science. Springer Berlin Heidelberg. p. 51–62.

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

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», *Hum Factors*, vol. 40, n. 1, pagg. 79–91, mar. 1998.

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

Glabus MF, Horwitz B, Holt JL, Kohn PD, Gerton BK, Callicott JH, MeyerLindenberg A, Berman KF (2003) Interindividual differences in functional interactions among prefrontal, parietal and parahippocampal regions during working memory. Cereb Cortex 13:1352--1361.

Glass JM, Schumacher EH, Lauber EJ, Zurbriggen EL, Gmeindl L, Keiras DE, Meyer DE, (2000). Aging and the psychological refractory period: task-coordination strategies in young and old adults. Psychol Aging 15(4):571--595.

Goldberg, D.H., Vogelstein, R.J., Socolinsky, D.A., Wolff, L.B., (2011). Toward a Wearable, Neurally-Enhanced Augmented Reality System, in:

Schmorrow, D.D., Fidopiastis, C.M. (Eds.), Foundations of Augmented Cognition. Directing the Future of Adaptive Systems. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 493–499.

Goldstein I. L. and Ford J. K. (2002). Training in Organizations: Needs Assessment, Development, and Evaluation. Wadsworth, 2002.

Goldstein I. L. (1980). «Training in Work Organizations», *Annu. Rev. Psychol.*, vol. 31, n. 1, pagg. 229–272, 1980.

Gopher D. (1993). The skill of attention control: acquisition and execution of attention strategies. In: Meyer D, Kornblum S, editors. Attention and performance XIV. Hillsdale, NJ: Lawrence Erlbaum. p 299--322.

Gopher, D., Donchin, E., (1986). Workload: An examination of the concept, in: Handbook of Perception and Human Performance, Vol. 2: Cognitive Processes and Performance. John Wiley & Sons, Oxford, England, pp. 1–49.

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

Grave de Peralta Menendez R, Andino SG (1999). Distributed Source Models: Standard Solutions and New Developments. In: Uhl, C. (Ed.), Analysis of Neurophysiological Brain Functioning. Springer Berlin Heidelberg, Berlin, Heidelberg. p. 176–201.

Gruzelier, J. (2009). A theory of alpha/theta neurofeedback, creative performance enhancement, long distance functional connectivity and psychological integration. Cognitive Processing, 10, 101–109.

Gruzelier, J., Egner, T., & Vernon, D. (2006). Validating the efficacy of neurofeedback for optimising performance. Progress in Brain Research, 159, 421–431.

Guderian, S., & Duzel, E. (2005). Induced theta oscillations mediate large-scale synchrony with mediotemporal areas during recollection in humans. Hippocampus, 15(7), 901–912.

Gundel A, Wilson GF. (1992). Topographical changes in the ongoing EEG related to the difficulty of mental tasks. Brain Topogr., 5:17-25.

Hagemann K. (2008). The alpha band as an electrophysiological indicator for internalized attention and high mental workload in real traffic driving. PhD Thesis. University of Düsseldorf, Germany.

Hamman W. (2004). «The complexity of team training: what we have learned from aviation and its applications to medicine», *Qual. Saf. Health Care*, vol. 13, n. Suppl 1, pagg. i72–i79, oct. 2004.

Hankins and Wilson (1998). «A comparison of heart rate, eye activity, EEG and subjective measures of pilot mental workload during flight.», *Aviat Space Environ Med*, vol. 69, n. 4, pagg. 360–367, apr. 1998.

Hansen PC (1992). Analysis of Discrete Ill-Posed Problems by Means of the L-Curve. SIAM Rev. 34: 561–580.

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

Harrison, J., Izzetoglu, K., Ayaz, H., Willems, B., Hah, S., Ahlstrom, U., Woo, H., Shewokis, P.A., Bunce, S.C., Onaral, B., (2014). Cognitive Workload and Learning Assessment During the Implementation of a Next-Generation Air Traffic Control Technology Using Functional Near-Infrared Spectroscopy. IEEE Trans. Hum.-Mach. Syst. 44, 429–440. doi:10.1109/THMS.2014.2319822.

Hart, S.G., Staveland, L.E., (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research, in: Human Mental Workload. North-Holland, pp. 139–183.

Hastie T., Tibshirani R., Friedman J. (2011). The Elements of Statistical Learning: Data Mining, Inference, and Prediction (Springer, New York, 2011).

He B, Hori J, Babiloni F (2006). Electroencephalography (EEG): Inverse Problems, in: Wiley Encyclopedia of Biomedical Engineering. John Wiley & Sons, Inc. p. 1355-1363.

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

Herff, C., Heger, D., Fortmann, O., Hennrich, J., Putze, F., Schultz, T., (2013). Mental workload during n-back task-quantified in the prefrontal

cortex using fNIRS. Front. Hum. Neurosci. 7, 935. doi:10.3389/fnhum.2013.00935.

Hikosaka, O., Nakamura, K., Sakai, K., & Nakahara, H. (2002). Central mechanisms of motor skill learning. *Current opinion in neurobiology*, *12*(2), 217-222.

Hill C. and Lent R. (2006): «A narrative and meta-analytic review of helping skills training: time to revive a dormant area of inquiry», *Psychother. Theory Res. Pract.*, vol. 43, pagg. 154–172, 2006.

Hinton G. E., Salakhutdinov R. R. (2006). Science 313, 504–507 (2006).

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

Hyvärinen A, Oja E (2000). Independent component analysis: algorithms and applications. Neural Netw. 13: 411–430.

Izzetoglu, K., Bunce, S., Onaral, B., Pourrezaei, K., Chance, B.,(2004). Functional Optical Brain Imaging Using Near-Infrared During Cognitive Tasks. Int. J. Hum.-Comput. Interact. 17, 211–227. doi:10.1207/s15327590ijhc1702_6.

Jacobs, J., Hwang, G., Curran, T., & Kahana, M. J. (2006). EEG oscillations and recognition memory: Theta correlates of memory retrieval and decision making. Neuroimage, 32(2), 978–987.

Jahnsen H. and Linas R. (1984). Electrophysiological properties of guinea-pig thalamic neurons: an in vitro study. J. Physiol. 1984; 349:205–226.

Jasper H. (1958). Report of the Committee on Methods of Clinical Examination in Electroencephalography. Electroenceph. Clin. Neurophysiol. 1958; 10:370–375.

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

Jaušovec, N. (2000a). Differences in cognitive processes between gifted, intelligent, ˇ creative, and average individuals while solving complex problems: An EEG Study. Intelligence, 28, 213–237.

Jaušovec, N., & Jau ˇ sovec, K. (2000b). EEG activity during the performance of complex ˇ mental problems. International Journal of Psychophysiology, 36, 73–88.

Jensen O, Kaiser J, Lachaux J-P. (2007). Human gamma-frequency oscillations associated with attention and memory. Trends Neurosci.; 30:317-24.

Jensen, O., Gelfand, J., Kounios, J., & Lisman, J. E. (2002). Oscillations in the alpha band (9–12 Hz) increase with memory load during retention in a short-term memory task. Cerebral Cortex, 12, 877–882.

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects.*Science*, *349*(6245), 255-260.

Jordan.C.S. (1992). Experimental Study of the Effect of An Instantaneous Self Assessment Workload Recorder on Task Performance, DRA/TM/CAD5/92011. Defence Research Agency. Portsmouth.

Jorna P. G. (1993). «Heart rate and workload variations in actual and simulated flight», *Ergonomics*, vol. 36, n. 9, pagg. 1043–1054, set. 1993.

Jorna P. G. (1992). «Spectral analysis of heart rate and psychological state: a review of its validity as a workload index», *Biol Psychol*, vol. 34, n. 2–3, pagg. 237–257, nov. 1992.

Jung-Beeman, M., Bowden, E. M., Haberman, J., Frymiare, J. L., Arambel-Liu, S., Greenblatt, R., et al. (2004). Neural activity when people solve verbal problems with insight. PLOS Biology, 2, 500–510.

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

Kahana, M. J. (2003). Human theta oscillations related to sensorimotor integration and spatial learning. Journal of Neuroscience, 23, 4726–4736.

Kalyuga S, Ayres P, Chandler P, Sweller J (2003): The Expertise Reversal Effect. Educ. Psychol. 38: 23–31.

Kandel E, Schwartz J, Jessell T (2000). Principles of Neural Science, Fourth Edition. McGraw-Hill Companies, Incorporated. 1414 p.

Karni, A., Tanne, D., Rubenstein, B. S., Askenasy, J. J., & Sagi, D. (1994). Dependence on REM sleep of overnight improvement of a perceptual skill. Science, 265(5172), 679–682.

Kellogg, R. T. (1980). Is conscious attention necessary for long-term storage? *Journal of Experimental Psychology: Human Learning and Memory*, *6*(4), 379.

Kelly, A. C., and Garavan, H. (2005). Human functional neuroimaging of brain changes associated with practice. *Cerebral Cortex*, *15*(8), 1089-1102.

Kelly SP, Dockree P, Reilly RB, Robertson IH (2003). EEG alpha power and coherence time courses in a sustained attention task. In: First International IEEE EMBS Conference on Neural Engineering, 2003. Conference Proceedings. pp. 83–86.

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

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

Kirwan, B., Evans, A., Donohoe, L., Kilner, A., Lamoureux, Atkinson, T., & MacKendrick, H. (1997) Human Factors in the ATM System Design Life Cycle. FAA/Eurocontrol ATM R&D Seminar, Paris, France. Internet source.

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

Kirwan, B., (1992). Human error identification in human reliability assessment. Part 2: Detailed comparison of techniques. Appl. Ergon. 23, 371–381.

Klausberger, T., Magill, P. J., Márton, L. F., Roberts, J. D. B., Cobden, P. M., Buzsáki, G., et al. (2003). Brain-state-and cell-type-specific firing of hippocampal interneurons in vivo. Nature, 421(6925), 844–848.

Klimesch W (2012): Alpha-band oscillations, attention, and controlled access to stored information. Trends Cogn. Sci. 16: 606–617.

Klimesch, W., Sauseng, P., & Hanslmayr, S. (2007). EEG alpha oscillations: The inhibition-timing hypothesis. Brain Research Reviews, 53, 63–88.

Klimesch W., Schack B., and Sauseng P. (2005): «The functional significance of theta and upper alpha oscillations», *Exp Psychol*, vol. 52, n. 2, pagg. 99–108, 2005.

Klimesch, W., Doppelmayr, M., Röhm, D., Pöllhuber, D., & Stadler, W. (2000). Simultaneous desynchronization and synchronization of different alpha responses in the human electroencephalograph: A neglected paradox? Neuroscience Letters, 284, 97–100.

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

Klimesch W, Doppelmayr M, Russegger H, Pachinger T, Schwaiger J (1998). Induced alpha band power changes in the human EEG and attention. Neurosci. Lett. 244: 73–76.

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

Kochanska, G., Murray, K. T., & Harlan, E. T. (2000). Effortful control in early childhood: continuity and change, antecedents, and implications for social development. *Developmental psychology*, *36*(2), 220.

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

Kotovsky K, Hayes JR, Simon HA (1985). Why are some problems hard? Evidence from Tower of Hanoi. Cognit Psychol 17: 248–294.

Kozlowski S. W. J., Gully S. M., Brown K. G., Salas E., Smith E. M., and Nason E. R (2001). «Effects of Training Goals and Goal Orientation Traits on Multidimensional Training Outcomes and Performance Adaptability», *Organ. Behav. Hum. Decis. Process.*, vol. 85, n. 1, pagg. 1–31, may. 2001.

Kramer AF, Larish JL, Weber TA, Bardell L. (1999). Training for executive control: task coordination strategies and aging. In: Gopher, Koriat, editors. Attention and performance XVII: cognitive regulation of performance: interaction of theory and application. Cambridge, MA: The MIT Press. p 617-652.

Kramer AF, Larish JF, Strayer DL. (1995). Training for attentional control in dual task settings: a comparison of young and old adults. J Exp Psychol Appl 1(1):50--76.

Kropotov, J. (2008). Quantitative EEG, event-related potentials and neurotherapy. San Diego: Academic Press.

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

Kubota, Y., Sato, W., Toichi, M., Murai, T., Okada, T., Hayashi, A., and Sengoku, A. (2001). Frontal midline theta rhythm is correlated with cardiac autonomic activities during the performance of an attention demanding meditation procedure. *Cognitive Brain Research*, *11*(2), 281-287.

Laberge D and Buchsbaum MS (1990). Positron emission tomographic measurements of pulvinar activity during an attention task. Neuroscience 10: 613–619, 1990.

Laberge D. (2001). Attention, consciousness, and electrical wave activity within the cortical column. Int J Psychophysiol 43: 5–24, 2001.

Laberge, D. (1995). Attentional Processing: The Brain's Art of Mindfulness. Cambridge, MA: Harvard University Press.

Lal SKL, Craig A (2002). Driver fatigue: Electroencephalograpy and psychological assessment. Psychophysiology 39: 313-321.

Lawton, R., Ward, N.J., (2005). A systems analysis of the Ladbroke Grove rail crash. Accid. Anal. Prev. 37, 235–244. doi:10.1016/j.aap.2004.08.001.

Leape L. L. and Fromson J. A. (2006). «Problem doctors: is there a system-level solution?», *Ann. Intern. Med.*, vol. 144, n. 2, pagg. 107–115, jan. 2006.

Lee D. H. and Park K. S. (1990). «Multivariate analysis of mental and physical load components in sinus arrhythmia scores», *Ergonomics*, vol. 33, n. 1, pagg. 35–47, gen. 1990.

Leonard M., Graham S., and Bonacum D. (2004). «The human factor: the critical importance of effective teamwork and communication in providing safe care», *Qual. Saf. Health Care*, vol. 13, n. Suppl 1, pagg. i85–i90, oct. 2004.

Leonard M. and Tarrant C. (2001). «Culture, systems, and human factors—Two tales of patient safety: The KP Colorado region's experience.», *The Permanente Journal*, vol. 5, pag. 46 −49, 2001.

Liao, L.-D., Lin, C.-T., McDowell, K., Wickenden, A.E., Gramann, K., Jung, T.-P., Ko, L.-W., Chang, J.-Y., (2012). Biosensor Technologies for Augmented Brain #x2013;Computer Interfaces in the Next Decades. Proc. IEEE 100, 1553–1566. doi:10.1109/JPROC.2012.2184829.

Logan GD (1988). Toward an instance theory of automatization. Psychol Rev 95: 492–527.

Lopes da Silva F. H (1996): The generation of electric and magnetic signals of the brain by local networks. In: R. Greger and U. Windhorst, eds., Comprehensive Human Physiology. Heidelberg, Germany: Springer-Verlag.

Lopes da Silva F. (1991). Neural mechanisms underlying brain waves: from neural membranes to networks. Electroencephalogr Clin Neurophysiol 79: 81–93.

Luft, A. R., & Buitrago, M. M. (2005). Stages of motor skill learning. Molecular Neurobiology, 32(3), 205–216.

Luu P, Tucker DM (2001). Regulating action: alternating activation of midline frontal and motor cortical networks. Clin. Neurophysiol. 112: 1295–1306.

Mackie MA, Van Dam NT, Fan J (2013): Cognitive control and attentional functions. Brain Cogn. 82: 301–312.

Maquet, P. (2008). Both the hippocampus and striatum are involved in consolidation of motor sequence memory. Neuron, 58(2), 261–272.

Marcel, S., Millan, J.D.R., 2007. Person Authentication Using Brainwaves (EEG) and Maximum A Posteriori Model Adaptation. IEEE Trans. Pattern Anal. Mach. Intell. 29, 743–752. doi:10.1109/TPAMI.2007.1012.

Marcel, A.J. (1980). Conscious and preconscious recognition of polysemous words: Locating the selective effects of prior verbal context. Attention and Performance, 8.

Martindale, C. (1999). Biological bases of creativity. In R. Sternberg (Ed.), Handbook of creativity (pp. 137–152). Cambridge, UK: Cambridge University Press.

Martindale, C., & Hasenfus, N. (1978). EEG differences as a function of creativity, stage of the creative process, and effort to be original. Biological Psychology, 6, 157–167.

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

Miller, (2001). «LITERATURE REVIEW Workload Measures Document N01-006 National Advanced Driving Simulator».

Mizuseki, K., Sirota, A., Pastalkova, E., & Buzsáki, G. (2009). Theta oscillations provide temporal windows for local circuit computation in the entorhinal-hippocampal loop. Neuron, 64(2), 267–280.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., ... & Petersen, S. (2015). Human-level control through deep reinforcement learning. *Nature*, *518*(7540), 529-533.

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

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

Murphy K. (2012). Machine Learning: A Probabilistic Perspective (MIT Press, Cambridge, MA, 2012).

Murkin, J.M., Arango, M., (2009). Near-infrared spectroscopy as an index of brain and tissue oxygenation. Br. J. Anaesth. 103 Suppl 1, i3–13. doi:10.1093/bja/aep299.

NASA, (2003). NASA Task Load Index (TLX): Computerized Version (Version 2.0) [Computer Software]. Moffett Field, CA: NASA-Ames Research Center, Aerospace Human Factors Research Division.

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 Astronaut. 43, 211–222.

Neumann, O. (1984). Automatic processing: A Review of recent findings and a plea for an old theory. In W. Prinz & A. F. Sanders (Eds.), Cognition and automatic processing (pp. 255-293). Berlin: Springer-Verlag.

Neuper, C., & Pfurtscheller, G. (2001). Event-related dynamics of cortical rhythms: frequency-specific features and functional correlates. Int J Psychophysiol. 2001 Dec;43(1):41-58.

Newell A, Rosenbloom PS (1993): The Soar Papers: Research on integrated intelligence. In: Rosenbloom PS, Laird JE, Newell A (Eds.), MIT Press, Cambridge, MA, USA. 1438 p.

Niedermayer E. and Lopes da Silva F. H. (1993): Electroencephalography. Basic Principles, Clinical Applications, and RelatedFields, 3rd ed. Baltimore, MD: Williams & Wilkins.

Nieuwenhuis, I. L., & Takashima, A. (2011). The role of the ventromedial prefrontal cortex in memory consolidation. Behavioural Brain Research, 218(2), 325–334.

Noback, C. R., Strominger, N. L., Demarest, R. J., & Ruggiero, D. A. (2005). *The human nervous system: structure and function* (No. 744). Springer Science & Business Media.

Norman, D.A., Berkrot, P., (2011). The Design of Everyday Things, MP3 - Unabridged CD edition. ed. Tantor Audio.

Norman, D. A. and Shallice, T. (1986). Attention to action: Willed and automatic control of behaviour. In Davidson, R. J., Schwartz, G. E., and Shapiro, D., editors, Consciousness and Self-Regulation: Advances in Research and Theory. Plenum Press.

Noyes, J. M., Bruneau, D. P. J. (2007). A self-analysis of the NASA-TLX workload measure. Ergonomics, 50(4), 514-519.

Nunez PL. (1995). Neocortical dynamics and human EEG rhythms. Oxford University Press, New York. 708 p.

O'Connor, D.H., Fukui, M.M., Pinsk, M.A., and Kastner, S. (2002). Attention modulates responses in the human lateral geniculate nucleus. Nat Neurosci 5: 1203–1209.

O'Donnell, R.D., Eggemeier, 1986. Workload assessment methodology. Handbook of Perception and Human Performance. John Wiley and Sons, Inc.

O'Reilly R.C., Braver T.S., Cohen J.D. (1999): A biologically based computational model of working memory, in: A. Miyake, P. Shah (Eds.), Models of Working Memory: Mechanisms of Active Maintenance and Executive Control, Cambridge Univ. Press, New York, NY, USA, 1999, pp. 375– 411.

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

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

Palomba D, Angrilli A, Mini A (1997). Visual evoked potentials, heart rate responses and memory to emotional pictorial stimuli. Int J Psychophysiol Off J Int Organ Psychophysiol 27: 55–67.

Parasuraman, R., and McKinley, R. A. (2014). Using noninvasive brain stimulation to accelerate learning and enhance human performance. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 0018720814538815.

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

Parasuraman, R., (2003). Neuroergonomics: Research and practice. Theor. Issues Ergon. Sci. 4, 5–20. doi:10.1080/1463922021019975 3.

Parasuraman, R., Hancock, P., (2001). Adaptive control of mental workload, in: Stress, Workload, and Fatigue. pp. 305–320.

Parasuraman, R.; Sheridan T.B. and Wickens, C. (2000): A Model for Types and Levels of Human Interaction with Automation. IEE Transactions on

Systems, Man, and Cybernetics-Part A: Systems and Humans, Vol 30, No. 3, May 2000.

Parasuraman, R., Mouloua, M., Molloy, R., (1996). Effects of adaptive task allocation on monitoring of automated systems. Hum. Factors 38, 665–679.

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

Patent II: Borghini G., Aricò P., Di Flumeri G. and Babiloni F., (2015). Metodo per stimare un livello di addestramento, e relativo apparato (A Method for the training assessment and related device). Patent pending. Number P1107IT00.

Pawlak W.S., Vicente K.J. (1996). "Inducing effective operator control through ecological interface design", International Journal of Human-Computer Studies, Elsevier, 1996.

Petersen SE, van Mier H, Fiez JA, Raichle ME (1998). The effects of practice on the functional anatomy of task performance. Proc Natl Acad Sci USA 95:853--860.

Petersen SE, Robinson, D.L., and Morris, J.D. (1987). Contributions of the pulvinar to visual spatial attention. Neuropsychologia 25: 97–105.

Pfurtscheller, G., & Lopes da Silva, F. H. (2005). EEG event-related desynchronization (ERD) and event-related synchronization (ERS). In E. Niedermeyer, & F. Lopes da Silva (Eds.), Electroencephalography: Basic principles, clinical applications, and related fields (5th ed., pp. 1003–1016). Philadelphia: Lippincott Williams & Wilkins.

Pfurtscheller, G. (1999). Quantification of ERD and ERS in the time domain. In G. Pfurtscheller, & F. H. Lopes da Silva (Eds.), Event-related desynchronization. Handbook of electroencephalography and clinical neurophysiology (Revised ed., Vol. 6, pp. 89–105). Amsterdam: Elsevier.

Pfurtscheller, G., Stancak, A., Jr., & Neuper, C. (1996). Event-related synchronization (ERS) in the alpha band—An electrophysiological correlate of cortical idling: A review. International Journal of Psychophysiology, 24, 39–46.

Pfurtscheller G. and Klimesch W. (1991). «Event-related desynchronization during motor behavior and visual information processing», *Electroencephalogr Clin Neurophysiol Suppl*, vol. 42, pagg. 58–65, 1991.

Pivik R. T., Broughton R. J., Coppola R., Davidson R. J., Fox N., and Nuwer M. R. (1993). Committee Report: Guidelines for the recording and quantitative analysis of electroencephalographic activity in research context. Psychophysiology 1993; 30:547–558.

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

Polanyi M. (1958). *Personal Knowledge.* London: Routledge & Kegan Paul, 1958.

Poldrack RA, Gabrieli JD (2001). Characterizing the neural mechanisms of skill learning and repetition priming: evidence from mirror reading. Brain 124:67--82.

Poldrack RA (2000). Imaging brain plasticity: conceptual and methodological issues--a theoretical review. Neuroimage 12:1-13.

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

Posner MI, Rothbart MK (2007). Research on attention networks as a model for the integration of psychological science. Annu Rev Psychol. 2007; 58:1–23.

Posner M.I., Petersen S.E. (1990). The attention system of the human brain, Annu. Rev. Neurosci. 13 (1990) 25–42.

Prinzel LJ, Freeman FG, Scerbo MW, Mikulka PJ, Pope AT (2003). Effects of a Psychophysiological System for Adaptive Automation on Performance, Workload, and the Event-Related Potential P300 Component. Hum Factors J Hum Factors Ergon Soc 45: 601–614.

Raichle ME, Fiez JA, Videen TO, MacLeod AM, Pardo JV, Fox PT, Petersen SE (1994). Practice-related changes in human brain functional anatomy during nonmotor learning. Cereb Cortex 4: 8-26.

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

Rankin, W., Hibit, R., Allen, J., Sargent, R., (2000). Development and evaluation of the Maintenance Error Decision Aid (MEDA) process. Int. J. Ind. Ergon. 26, 261–276. doi:10.1016/S0169-8141(99)00070-0.

Rasmussen, J., & Vicente, K. J. (1989). Coping with human errors through system design: implications for ecological interface design. *International Journal of Man-Machine Studies*, *31*(5), 517-534.

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

Rasmussen, J. (1986). {Information Processing and Human-Machine Interaction. An Approach to Cognitive Engineering}.

Rasmussen, J. (1983). Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models. *Systems, Man and Cybernetics, IEEE Transactions on*, (3), 257-266.

Rasmussen, J. (1982). Human errors. A taxonomy for describing human malfunction in industrial installations. J. Occup. Accid. 4, 311–333. doi:10.1016/0376-6349(82)90041-4.

Raz A. and Buhle J. (2006). Typologies of attentional networks. Nature Reviews Neuroscience 7, 367-379.

Razumnikova, O. M. (2000). Functional organization of different brain areas during convergent and divergent thinking: An EEG investigation. Cognitive Brain Research, 10, 11–18.

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

Rechtschaffen A, Kales A. (1968). A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects. Rechtschaffen and Kales, editors. Bethesda, Md: U. S. National Institute of Neurological Diseases and Blindness, Neurological Information Network.

Reid, G.B., Nygren, T.E., (1988). The Subjective Workload Assessment Technique: A Scaling Procedure for Measuring Mental Workload, in:

Meshkati, P.A.H. and N. (Ed.), Advances in Psychology, Human Mental Workload. North-Holland, pp. 185–218.

Reiner, M., Rozengurt, R., & Barnea, A. (2014). Better than sleep: Theta neurofeedback training accelerates memory consolidation. Biological psychology, 95, 45-53.

Rihs, T. A., Michel, C. M., & Thut, G. (2007). Mechanisms of selective inhibition in visual spatial attention are indexed by α-band EEG synchronization. European Journal of Neuroscience, 25, 603–610.

Ritter F, Schooler L (2001). The Learning Curve. In: International Encyclopedia of the Social & Behavioral Sciences. Elsevier Science Ltd. 1508 p.

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

Roscoe A. H. (1993). «Heart rate as a psychophysiological measure for in-flight workload assessment», *Ergonomics*, vol. 36, n. 9, pagg. 1055–1062, set. 1993.

Roscoe A. H. (1992). «Assessing pilot workload. Why measure heart rate, HRV and respiration?», *Biol Psychol*, vol. 34, n. 2–3, pagg. 259–287, nov. 1992.

Rosenbluth A., Wiener N., and Bigelow J. (1943). "Behaviour, purpose and teleology," *Phil. Sci.*, vol. 10, pp. 18-24, 1943.

Rudner M. and Lunner T. (2014). «Cognitive spare capacity and speech communication: a narrative overview», *BioMed Res. Int.*, vol. 2014, pag. 869726, 2014.

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

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

Sadaghiani S, Scheeringa R, Lehongre K, Morillon B, Giraud AL, Kleinschmidt A (2010). Intrinsic connectivity networks, alpha oscillations,

and tonic alertness: a simultaneous electroencephalography/functional magnetic resonance imaging study. J Neurosci Off J Soc Neurosci 30: 10243–10250.

Sakai, K., Hikosaka, O., Miyauchi, S., Sasaki, Y., Fujimaki, N., and Pu ¨ tz, B. (1999). Presupplementary motor area activation during sequence learning reflects visuo-motor association. J. Neurosci. 19.

Salas E., Burke C. S., Bowers C. A., and Wilson K. A. (2001). «Team training in the skies: does crew resource management (CRM) training work?», *Hum. Factors*, vol. 43, n. 4, pagg. 641–674, 2001.

Salmon, P., Regan, M., Johnston, I., (2005). Human Error and Road Transport: Phase One - Literature Review. Monash University Accident Research Centre.

Sammer, G., Blecker, C., Gebhardt, H., Bischoff, M., Stark, R., Morgen, K., and 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.

Sandkühler, S., & Bhattacharya, J. (2008). Deconstructing insight: EEG correlates of insightful problem solving. PloS ONE, 3(1), e1459.

Satterfield J. M. and Hughes E. (2007). «Emotion skills training for medical students: a systematic review», *Med. Educ.*, vol. 41, n. 10, pagg. 935–941, oct. 2007.

Sauseng, P., Klimesch, W., Schabus, M., & Doppelmayr, M. (2005). Fronto-parietal EEG coherence in theta and upper alpha reflect central executive functions of working memory. International Journal of Psychophysiology, 57(2), 97–103.

Schmidt, R. (2001). "Attention." In P. Robinson (Ed.), Cognition and second language instruction (pp. 3-32). Cambridge University Press.

Schmidt, R. A., Lange, C., & Young, D. E. (1990). Optimizing summary knowledge of results for skill learning. *Human Movement Science, 9*(3), 325-348.

Schmorrow, D.D., (2002). DARPA's Augmented Cognition Program-tomorrow's human computer interaction from vision to reality: building cognitively aware computational systems, in: Proceedings of the 2002 IEEE

7th Conference on Human Factors and Power Plants, 2002. Presented at the Proceedings of the 2002 IEEE 7th Conference on Human Factors and Power Plants, 2002, pp. 7–1–7–4. doi:10.1109/HFPP.2002.1042859.

Schneider W, Shiffrin RM (1977). Controlled and automatic human information processing: I. Detection, search, and attention. Psychol Rev 84: 1–66.

Schubert, R., Tangermann, M., (2008). "Parieto-occipital alpha power indexes distraction during simulated car driving," abstracts of the 14th World Congress of Psychophysiology. Int. J. Psychophysiol. - INT J PSYCHOPHYSIOL 69. doi:10.1016/j.ijpsycho.2008.05.033.

Schumacher EH, Lauber EJ, Glass JM, Zurbriggen EL, Gmeindl L, Kieras DE, Meyer DE. (1999). Concurrent response-selection processes in dual-task performance: evidence for adaptive executive control of task-scheduling. J Exp Psychol Hum Percept Perform 25:791-814.

Seidler, R.D. (2010). Neural correlates of motor learning, transfer of learning, and learning to learn. Exerc. Sport Sci. Rev. 38, 3–9.

Seidler RD, Purushotham A, Kim SG, Ugurbil K, Willingham D, Ashe J (2002) Cerebellum activation associated with performance change but not motor learning. Science 296:2043--2046.

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

Shadmehr R, Holcomb HH (1997). Neural correlates of motor memory consolidation. Science 277:821--5.

Shalev-Shwartz S., Shamir O., Tromer E. (2012): Using more data to speed up training time, Proceedings of the Fifteenth Conference on Artificial Intelligence and Statistics, Canary Islands, Spain, 21 to 23 April, 2012.

Shallice T. (1982). Specific impairments of planning, Philos. Trans. R. Soc. London Ser. B 298 (1982) 199– 209.

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

Shiffrin RM, Schneider W (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic.

Shorrock, S.T., Kirwan, B., (2002). Development and application of a human error identification tool for air traffic control. Appl. Ergon. 33, 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. J. Neurosci. Methods 209, 22–34. doi:10.1016/j.jneumeth.2012.05.022.

Shultz S, Klin A, Jones W (2011). Inhibition of eye blinking reveals subjective perceptions of stimulus salience. Proc Natl Acad Sci USA 108: 21270–21275.

Sirevaag E. J., Kramer A. F., Wickens C. D., Reisweber M., Strayer D. L., and Grenell J. F. (1993). «Assessment of pilot performance and mental workload in rotary wing aircraft», *Ergonomics*, vol. 36, n. 9, pagg. 1121–1140, set. 1993.

Sirota, A., Montgomery, S., Fujisawa, S., Isomura, Y., Zugaro, M., & Buzsáki, G. (2008). Entrainment of neocortical neurons and gamma oscillations by the hippocampal theta rhythm. Neuron, 60(4), 683–697.

Smit, A.S., Eling, P. a. T.M., Hopman, M.T.E., Coenen, A.M.L., (2005). Mental and physical effort affect vigilance differently. 218.

Smith ME, Gevins A, Brown H, Karnik A, Du R (2001). Monitoring Task Loading with Multivariate EEG Measures during Complex Forms of Human-Computer Interaction. Hum Factors J Hum Factors Ergon Soc 43: 366–380.

Spearman C. (1928). «The Origin of Error», J. Gen. Psychol., vol. 1, n. 1, pagg. 29–53, gen. 1928.

Stanton, N.A., Harris, D., Salmon, P.M., Demagalski, J.M., Marshall, A., Young, M.S., Dekker, S.W.A., Waldmann, T., (2006). Predicting design induced pilot error using HET (human error template) - A new formal human error identification method for flight decks. [WWW Document]. URL https://dspace.lib.cranfield.ac.uk/handle/1826/1158 (accessed 11.25.15).

Steel, R. G. D., and Torrie, J. H. (1960). *Principles and procedures of statistics: with special reference to the biological sciences*. McGraw-Hill.

Steinvorth, S., Wang, C., Ulbert, I., Schomer, D., & Halgren, E. (2010). Human entorhinal gamma and theta oscillations selective for remote autobiographical memory. Hippocampus, 20(1), 166–173.

Stern e Skelly (1984). «The eye blink and workload considerations», in *Proceedings of the Human Factors Society - 28th Annual Meeting*, San Antonio, Texas, 1984.

Sterman M. B., Mann C. A., Kaiser D. A., and Suyenobu B. Y.(1994). «Multiband topographic EEG analysis of a simulated visuomotor aviation task», *Int J Psychophysiol*, vol. 16, n. 1, pagg. 49–56, feb. 1994.

Strimmer K (2008). A unified approach to false discovery rate estimation. BMC Bioinformatics 9: 303.

Sutton R. S., Barto A. G. (1998): Reinforcement Learning: An Introduction (MIT Press, Cambridge, MA, 1998).

Sweller J, van Merrienboer JG, Paas FGWC (1998). Cognitive Architecture and Instructional Design. Educ Psychol Rev 10: 251–296.

Sweller J. (1999). Instructional Design in Technical Areas. Australian Council for Educational. 168 p.

Taggart W. R. (1994). «Crew Resource Management: Achieving enhanced flight operations», Aviation Psychology in Practice, pagg. 309–339, 1994.

Tannenbaum S. and Yukl G. (1992). «Training and Development in Work Organizations», *Annu. Rev. Psychol.*, vol. 43, n. 1, pagg. 399–441, 1992.

Trejo LJ, Shensa MJ (1999). Feature Extraction of Event-Related Potentials Using Wavelets: An Application to Human Performance Monitoring. Brain Lang 66: 89–107.

Uebbing-Rumke M., Gürlük H., Schulze-Kissing D. (2012). "Adaptive Automation Support for Time-Based Operations in ATC", Third International Air Transport & Operations Symposium ATOS, Delft, Nederland, 2012.

Van Orden, Makeig, Jung, e Limbert (1999). «Eye activity correlates of workload during a visuospatial memory task». Defense Technical Information Center, 1999.

Vanduffel, W., Tootell, R.B., and Orban, G.A. (2000). Attention-dependent suppression of metabolic activity in the early stages of the macaque visual system. Cereb Cortex 10: 109–126.

Vapnik, V.N., 2000. The Nature of Statistical Learning Theory. Springer New York, New York, NY.

Varela, F., Lachaux, J. P., Rodriguez, E., & Martinerie, J. (2001). The brainweb: Phase synchronization and large-scale integration. Nature Reviews Neuroscience, 2(4), 229–239.

Veltman J. A. and Gaillard A. W. (1996). «Physiological indices of workload in a simulated flight task», *Biol Psychol*, vol. 42, n. 3, pagg. 323–342, feb. 1996.

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

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

Vicente K. J. (2002). "Ecological Interface Design: Progress and Challenges", Human Factors: The Journal of the Human Factors and Ergonomics Society, Spring, vol. 44, no. 1, pp. 62-78, 2002.

Vicente K. J., Rasmussen J. (1992). "Ecological Interface Design: Theoretical Foundations", IEEE Transactions on Systems Man and Cybernetics, Vol. 22, no. 4, July/August 1992.

Vink, M., Pas, P., Bijleveld, E., Custers, R., & Gladwin, T. E. (2013). Ventral striatum is related to within-subject learning performance. Neuroscience, 250, 408–416.

Von Luxburg, U., Schoelkopf, B., (2008). Statistical Learning Theory: Models, Concepts, and Results. ArXiv08104752 Math Stat.

Von Stein, A., & Sarnthein, J. (2000). Different frequencies for different scales of cortical integration: From local gamma to long range alpha/theta synchronization. International Journal of Psychophysiology, 38, 301–313.

Ward LM. (2003). Synchronous neural oscillations and cognitive processes. Trends Cogn Sci (Regul Ed); 7:553-9.

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

Welke, S., Jurgensohn, T., Roetting, M., (2009). Single-Trial Detection of Cognitive Processes for Increasing Traffic Safety. Proc. 21st Esv Int. Tech. Conf. Enhanc. Saf. Veh. Held June 2009 Stuttg. Ger.

Wexley K. N. (1984). «Personnel training», *Annu. Rev. Psychol.*, vol. 35, pagg. 519–551, 1984.

Wickens, C.D., (1984). Processing resources in attention. In R. Parasuraman and D.R. Davies (Eds.). Varieties of attention. (pp. 63-102). London: Academic Press.

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

Wierwille, W.W., Eggemeier, F.T., (1993). Recommendations for Mental Workload Measurement in a Test and Evaluation Environment. Hum. Factors J. Hum. Factors Ergon. Soc. 35, 263–281. doi:10.1177/001872089303500205.

Wilson GF, Russell CA (2003). Real-Time Assessment of Mental Workload Using Psychophysiological Measures and Artificial Neural Networks. Hum Factors J Hum Factors Ergon Soc 45: 635–644.

Wilson G.F., Lambet J.D., Russell C.A. (2000). "Performance enhancement with Real-Time Physiologically controlled Adaptive Aiding", Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Vol. 44, No. 13, pp. 61-64, 2000.

Wilson G. F., Swain C. R., and Ullsperger P. (1999). «EEG power changes during a multiple level memory retention task», *Int J Psychophysiol*, vol. 32, n. 2, pagg. 107–118, mag. 1999.

Wilson G. F., Fullenkamp P., and Davis I. (1994). «Evoked potential, cardiac, blink, and respiration measures of pilot workload in air-to-ground missions», *Aviat Space Environ Med*, vol. 65, n. 2, pagg. 100–105, feb. 1994.

Wilson G. F. (1993). «Air-to-ground training missions: a psychophysiological workload analysis», *Ergonomics*, vol. 36, n. 9, pagg. 1071–1087, set. 1993.

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

Wolpaw, J.R., Birbaumer, N., McFarland, D.J., Pfurtscheller, G., Vaughan, T.M., (2002). Brain-computer interfaces for communication and control. Clin. Neurophysiol. Off. J. Int. Fed. Clin. Neurophysiol. 113, 767–91.

Womelsdorf, T., Schoffelen, J. M., Oostenveld, R., Singer, W., Desimone, R., Engel, A. K., et al. (2007). Modulation of neuronal interactions through neuronal synchronization. Science, 316(5831), 1609–1612.

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

Yamada F. (1998). «Frontal midline theta rhythm and eyeblinking activity during a VDT task and a video game: useful tools for psychophysiology in ergonomics», *Ergonomics*, vol. 41, n. 5, pagg. 678–688, mag. 1998.

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

Zander, T.O., Kothe, C., Welke, S., Roetting, M., (2009). Utilizing Secondary Input from Passive Brain-Computer Interfaces for Enhancing Human-Machine Interaction, in: Schmorrow, D.D., Estabrooke, I.V., Grootjen, M. (Eds.), Foundations of Augmented Cognition. Neuroergonomics and Operational Neuroscience. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 759–771.

Zar JH (1999). Biostatistical analysis. Prentice Hall PTR. 663 p.

Zhang, J.H., Chung, T.D., Oldenburg, K., (1999). A Simple Statistical Parameter for Use in Evaluation and Validation of High Throughput Screening Assays. J. Biomol. Screen. 4, 67–73.

Scientific Writings

Patents

1. G. Borghini, P. Aricò, G. Di Flumeri and Fabio Babiloni. “Metodo per stimare un livello di addestramento, e relativo apparato”. Patent submitted, REF: P1107IT00.
2. P. Aricò, G. Borghini, G. Di Flumeri and F. Babiloni. “Metodo per stimare uno stato mentale, in particolare un carico di lavoro, e relativo apparato”. Patent submitted, REF: P1108IT00.

Journal Papers

1. G. Borghini, P. Aricò, I. Graziani, Y. Sun, F. Taya, A. Bezerianos, N.V. Thakor and F. Babiloni, “Quantitative assessment of the training improvement in a motor-cognitive task by using EEG, ECG and EOG signals”. Published on Brain Topography Journal. January, 2015 (DOI: 0.1007/s10548-015-0425-7).
2. G. Di Flumeri, G. Borghini, P. Aricò, S. Pozzi, S. Bonelli, A. Golfetti, A. Colosimo and F. Babiloni. “Mental workload evaluation of ATCOs during ecological ATM scenarios. IJASM. In press. (Prize), 2015.
3. G. Borghini, J. Toppi, L. Astolfi, M. Petti, E. He, V. De Giusti, B. He, F. Babiloni. “Investigating Cooperative Behavior in Ecological Settings: An EEG Hyperscanning Study". Submitted to NeuroImage, September 2015.
4. W. Kong, W. Lin, F. Babiloni, S. Hu and G. Borghini. “Investigating Drivers Fatigue versus Alert by Granger Causality Network". Physical Sensors - Sensors in New Road Vehicles, 2015.
5. G. Borghini, G. Vecchiato, P. Aricò, I. Graziani, A.G. Maglione, P. Cherubino and F. Babiloni. “Investigation on the effect of EEG-BCI on the simultaneous execution of flight simulation and attentional tasks”. Accepted for publication on Journal of Neuroscience Methods, November 2015.
6. P. Aricò, G. Borghini, I. Graziani, J.P. Imbert, G. Granger, R. Benhacene, S. Pozzi, L. Napoletano and F. Babiloni, “ATCo: neurophysiological analysis of the training and of the workload”.

Journal of Aerospace Medicine. May 2014. In press. (Winner of the prize for the best research in the Italian ATM field).

7. P. Aricò, G. Borghini, I. Graziani, F. Bianchini, F. Cincotti, F. Babiloni, "A brain computer interface system for the online evaluation of ATCs' workload". Italian Journal Of Aerospace Medicine. In press. Submitted in June, 2013.

Conference Papers

1. G. Borghini, P. Aricò, G. Di Flumeri, S. Salinari, A. Colosimo, S. Bonelli, L. Napoletano, A. Ferreira, and F. Babiloni. "Avionic Technology Testing By Using a Cognitive Neurometric Index: a Study with Professional Helicopter Pilots". 37th Annual International Conference of the IEEE EMBC, 2015.
2. G. Di Flumeri, G. Borghini, P. Aricò, A. Colosimo, S. Pozzi, S. Bonelli, A. Golfetti, W. Kong and F. Babiloni. "On the use of cognitive neurometric indexes in aeronautic and air traffic management environments", SYMBIOTIC 2015.
3. G. Borghini, P. Aricò, G. Di Flumeri, I. Graziani, A. Colosimo, S. Salinari, F. Babiloni, J.P. Imbert, G. Granger, R. Benhacene, A. Golfetti, S. Bonelli and S. Pozzi. "Skill Rule and Knowledge based behaviors detection during realistic ATM simulations by means of ATCOs brain activity". Fifth SESAR Innovation Days, October 2015.
4. F. Taya, Y. Sun, G. Borghini, P. Aricò, F. Babiloni, A. Bezerianos, N. Thakor. "Training-induced changes in information transfer efficiency of the brain network: A functional connectome approach". 7th International IEEE EMBS Conference on Neural Engineering of the IEEE Engineering in Medicine and Biology Society, 2015.
5. P. Aricò, G. Borghini, G. Di Flumeri, A. Colosimo, I. Graziani, J.P. Imbert, G. Granger, R. Benhacene, M. Terenzi, S. Pozzi and F. Babiloni. "Reliability over time of EEG-based mental workload evaluation during Air Traffic Management (ATM) tasks". 37th Annual International Conference of the IEEE EMBC, 2015.
6. Bezerianos, Y. Sun, Y. Chen, K.F. Wong, F. Taya, P. Aricò, G. Borghini, F. Babiloni and N. Thakor. "Cooperation driven coherence: brains working hard together". 37th Annual International Conference of the IEEE EMBC, 2015.G. Cartocci, A.G. Maglione, A. Scorpecci,

P. Marsella, P. Malerba, F. Babiloni, G. Di Flumeri, G. Borghini, P. Aricò. “Mental workload estimations in unilateral deafened children". 37th Annual International Conference of the IEEE EMBC, 2015.
8. G. Borghini, P. Aricò, F. Ferri, I. Graziani, S. Pozzi, L. Napoletano, J. P. Imbert, G. Granger, R. Benhacene, and F. Babiloni. A neurophysiological training evaluation metric for Air Traffic Management. 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, August 2014.
9. G. Borghini, P. Aricò, I. Graziani and Fabio Babiloni. "Training improvement assessment by EEG, ECG and EOG analysis ". IV Congresso Nazionale di Bioingegneria. Pavia, 25th -27th June, 2014.
10. G. Borghini, F. Babiloni, W. Kong, G. Vecchiato and A. G. Maglione. Evaluation of the cerebral workload and drowsiness during car driving by using high resolution EEG activity and neurophysiologic indices. 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, August 2014.
11. P. Aricò, G. Borghini, I. Graziani, F. Taya, Y. Sun, A. Bezerianos, N. Thakor, F. Cincotti and F. Babiloni. Towards a multimodal bioelectrical framework for the online mental workload evaluation. 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, August 2014.
12. P. Aricò, G. Borghini, F. Taya, I. Graziani, Y. Sun, A. Bezerianos, F. Cincotti, N.V. Thakor and F. Babiloni. Towards a multimodal bioelectrical framework for the online mental workload evaluation. IV Congresso Nazionale di Bioingegneria. Pavia, 25th -27th June, 2014.
13. F. Babiloni, G. Vecchiato, A. G. Maglione, G. Borghini and P. Aricò. The great beauty: a neuroaesthetic study by neuroelectric imaging during the observation of the real Michelangelo’s Moses sculpture. 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, August 2014.
14. G. Borghini, P. Aricò, L. Astolfi, J. Toppi, F. Cincotti, D. Mattia, G. Vecchiato, A. G. Maglione and Fabio Babiloni. "Frontal EEG theta changes assess the training improvements of novices in flight simulation tasks". 35th Annual International IEEE EMBS Conference of the IEEE Engineering Medicine and Biology Society in Osaka, Japan, August, 2013.

15. F. Babiloni, G. Borghini, D. Mattia, D. Picconi, I. Graziani, C. Patrizia, A. Trettel, F. Infarinato, G. Vecchiato, A. G. Maglione, "Neuroelectric Brain Imaging During a Real Visit of a Fine Arts Gallery: A Neuroaesthetic Study of XVII Century Dutch Painters". 35th Annual International IEEE EMBS Conference of the IEEE Engineering Medicine and Biology Society in Osaka, Japan, August, 2013.
16. G. Borghini, P. Aricò, F. Babiloni, G. Granger, J-P. Imbert, R. Benhacene, L. Napoletano, S. Pozzi. NINA: Neurometrics Indacators for ATM. Third SESAR Innovation Days, Stockholm, Sweden, November, 2013.

List of Abbreviations

AA	Adaptive Automation
ACC	Area Control Center
ACC	Anterior Cingulate Cortex
AI	Artificial Intelligence
AIS	Aeronautical Information Services
ANOVA	Analysis of Variance
ANS	Autonomic Nervous System
ATC	Air Traffic Controller
ATCO	Air Traffic Controller Officer
ATFM	Air Traffic Flow Management
ATM	Air Traffic Management
ATP	Adenosine Triphosphate
AUC	Area Under Curve
AUGCOG	Augmented Cognition
asSWLDA	Automatic Stop StepWise Linear Discriminant Analysis
BA	Brodmann Area
BCI	Brain Computer Interface
BOLD	Blood Oxygen Level Dependent
CC	Cortico Cerebellar
CI	Confidence Interval
Class-ndx	Classification index
Class-THR	Classification threshold
CMA	Cingulate Motor Area
CNS	Central Nervous System
COMM	auditory monitoring
CS	Cortico Striatal
CWP	Controller Working Position
DARPA	Defense Advanced Research Projects Agency
DLPFC	Dorsolateral Prefrontal Cortex
DM	Decision Making
DLPFC	Dorsolateral Prefrontal Cortex
EBR	Eye Blinks Rate
ECG	Electrocardiography
ECoG	Electrocorticography
EEG	Electroencephalogram
ENAC	École Nationale de l'Aviation Civile

ENAV	Ente Nazionale di Assistenza al Volo
EOG	Electrooculogram
EPSP	Excitatory or depolarizing postsynaptic potential
FDR	False Discovery Rate
FEF	Frontal Eye-Field
FL	Flight Level
FFT	Fast Fourier Transform
fMRI	Functional Magnetic Resonance Imaging
fNIR	Functional Near InfraRed
FOS	Fast Optical Signal
GABA	Gamma Amino Butyric Acid
GSR	Galvanic Skin Response
HF	Human Factor
HMI	Human Machine Interaction
HR	Heart Rate
HUD	Head Up Display
IAF	Individual Alpha Frequency
ICA	Independent Component Analysis
IPS	Intraparietal Sulcus
IPSP	Inhibitory or hyperpolarizing postsynaptic potential
ISA	Instantaneous Self Assessment
LDA	Linear Discriminant Analysis
LT	Laplace transform
MA-Class	Moving Average Classification index
MATB	Multi Attribute Task Battery
MEG	Magneticencephalogram
mPFC	Medial Prefrontal Cortex
MWL	Mental Work Load
NASA-TLX	NASA Task Load Index
NINA	Neurometric Indicators for ATM
NN	Neural Networks
NUS	National University of Singapore
OFC	Orbitofrontal Cortex
p-BCI	Passive Brain Computer Interface
PET	Positron Emission Tomography
PFC	Prefrontal Cortex
PNS	Peripheral Nervous System
PPC	Posterior Parietal Cortex

PSD	Power Spectral Density
RMAN	Resource Management task
ROC	Receiver Operating Characteristic
ROI	Region of Interest
RT	Reaction Times
S-R-K	Skill, Rule and Knowledge
SRT	Serial Reaction Time
STCA	Short Term Conflict Alert
SVM	Support Vector Machines
SYSM	Response to Event onsets task
SWLDA	Stepwise Linear Discriminant Analysis
tfICA	Time-Frequency Independent Component Analysis
TMS	Transcranial Magnetic Stimulation
TP	Task Performance
TPJ	Temporo-Parietal Junction
TRCK	Tracking task
TTB	Take-the-Best
VMPFC	Ventromedial Prefrontal Cortex
VLPFC	Ventrolateral Prefrontal Cortex
WEEG	EEG-based Mental Workload Index
WM	Working Memory
WMRT	Weighted Mean Reaction Time
WL	Work Load