ALMA MATER STUDIORUM

UNIVERSITÀ DEGLI STUDI DI BOLOGNA

DOTTORATO DI RICERCA IN

BIOINGEGNERIA

CICLO XXVI

SETTORE CONCORSUALE: 09/G2

SETTORE SCIENTIFICO DISCIPLINARE DI AFFERENZA: ING-INF/06

MENTAL STATES MONITORING THROUGH PASSIVE BRAIN-COMPUTER INTERFACE SYSTEMS

Pietro Aricò

RELATORE:

Prof. ssa **Serenella Salinari**

CO-RELATORE:

Prof. **Febo Cincotti**

Ing. **Fabio Aloise**

REVISORI:

Prof. **Mauro Ursino**

Dr. **Ricardo Chavarriaga**

COORDINATORE DOTTORATO:

Prof. ssa **Elisa Magosso**

ESAME FINALE ANNO 2014

ABSTRACT

The monitoring of cognitive functions aims at gaining information about the current cognitive state of the user by decoding brain signals. In recent years, this approach allowed to acquire valuable information about the cognitive aspects regarding the interaction of humans with external world. From this consideration, researchers started to consider passive application of brain–computer interface (BCI) in order to provide a novel input modality for technical systems solely based on brain activity. The objective of this thesis is to demonstrate how the passive Brain Computer Interfaces (BCIs) applications can be used to assess the mental states of the users, in order to improve the human machine interaction. Two main studies has been proposed. The first one allows to investigate whatever the Event Related Potentials (ERPs) morphological variations can be used to predict the users' mental states (e.g. attentional resources, mental workload) during different reactive BCI tasks (e.g. P300-based BCIs), and if these information can predict the subjects' performance in performing the tasks. In the second study, a passive BCI system able to online estimate the mental workload of the user by relying on the combination of the EEG and the ECG biosignals has been proposed. The latter study has been performed by simulating an operative scenario, in which the occurrence of errors or lack of performance could have significant consequences. The results showed that the proposed system is able to estimate online the mental workload of the subjects discriminating three different difficulty level of the tasks ensuring a high reliability.

1 INTRODUCTION

Simultaneous control of multiple devices, while maintaining high attentional levels, represents an important feature in several operating environment. For example, pilot a plane or drive a car represents the classic situations where the operator has to manage simultaneously the available devices, while maintaining a high level of attention. There are also situations in which the required cognitive load can become very high, for example in safety-critical applications. These considerations point out the usefulness of a system that continuously monitors the user's mental states and that at the same time can act on the system itself using the subjective collected information.

A BCI is typically defined as a communication system, which relies on brain activity to control an external device bypassing muscular and nerves pathway (e.g., using electroencephalogram (EEG) technique, Wolpaw et al. 2002). BCI research was originally driven by the goal to provide an alternative/additional channel to restore communication and interaction with the external world in people with severe motor disabilities. More recently, Wolpaw and Wolpaw (2012) defined a Brain-Computer Interface as *"a system that measures Central Nervous System (CNS) activity and converts it into artificial output that replaces, restores, enhances, supplements, or improves natural CNS output and thereby changes the ongoing interactions between the CNS and its external or internal environment"*. Thus, researchers suggested new application fields for BCI systems, developing applications that also involve subjects in operational environments, as military and commercial pilots and car drivers (Zander et al., 2009; Mueller et al., 2008; Blankertz et al. 2010). In fact, the meaning of the

term "BCI" (which originally only included the translation of the users' intentions through the classification of their voluntarily modulated brain activity) was broadened to comprise monitoring of cognitive states (e.g. mental workload, attention levels) identified through the users' spontaneous brain activity.

The objective of this PhD thesis is to design and validate a passive Brain Computer Interface (BCI) system able to estimate the user's mental state through the analysis of neurophysiological signals.

This thesis is organized in five main sections:

- ✓ In the first part basic concepts about the nervous system, the EEG signal and the ECG signal will be provided. In addition, a review of the state of the art concerning the covert and the overt attention modalities, the mental workload and the Brain Computer Interface (BCI) systems will be reported.

- ✓ In the second section, the studies regarding the event-related potentials and changes in their morphology during the use of two BCI interfaces used in overt and covert attention modalities will be reported and discussed.

- ✓ The third section will deal the design and the development of a monitoring system of the user's mental workload in operational environments using EEG rhythms and the ECG signal.

- ✓ In the fourth section, the general conclusions about the carried out research will be discussed.

2 PRELIMINARY CONCEPTS

2.1 The nervous system

Before discussing physiological measures, it is important to have at least a brief understanding of the extremely complex human nervous system (NS). The NS is a complex network of nerves and cells that carry messages to and from the brain and spinal cord to various parts of the body (Figure 2.1). The nervous system includes both the Central nervous system and Peripheral nervous system. The Central nervous system is made up of the brain and spinal cord and The Peripheral nervous system is made up of the Somatic and the Autonomic nervous systems.

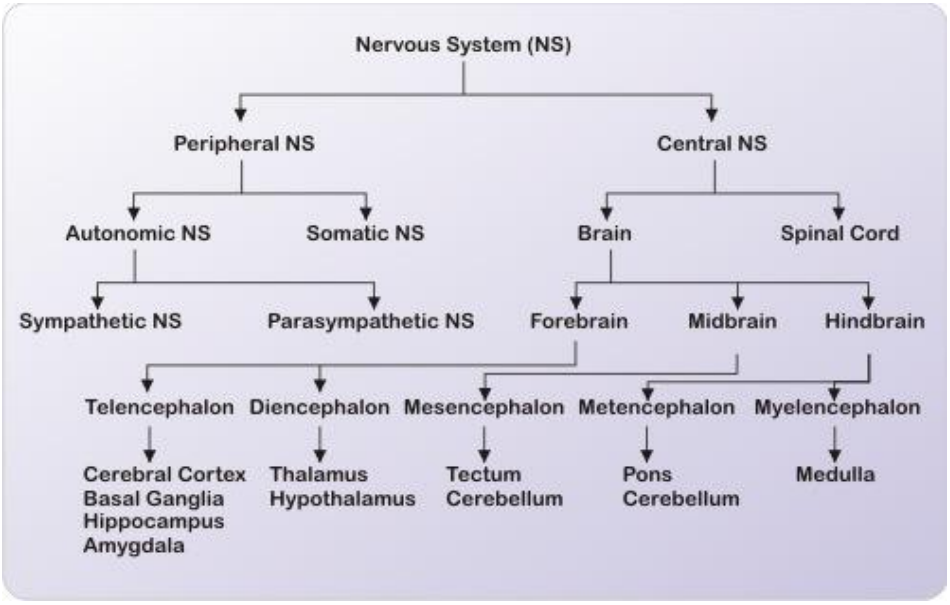


Figure 2.1: Schematic overview of the nervous system

2.1.1 The Central Nervous System

The Central Nervous System (CNS) gathers information about the environment through sensations, controls thought and motor control. Central to this effort and to the understanding of our existence is the brain. First, it is important to discuss the basic functions of the brain as related to its anatomy. The brain is made up of several components, which work in concert to perform the myriad of functions, which we use to survive (Figure 2.2).

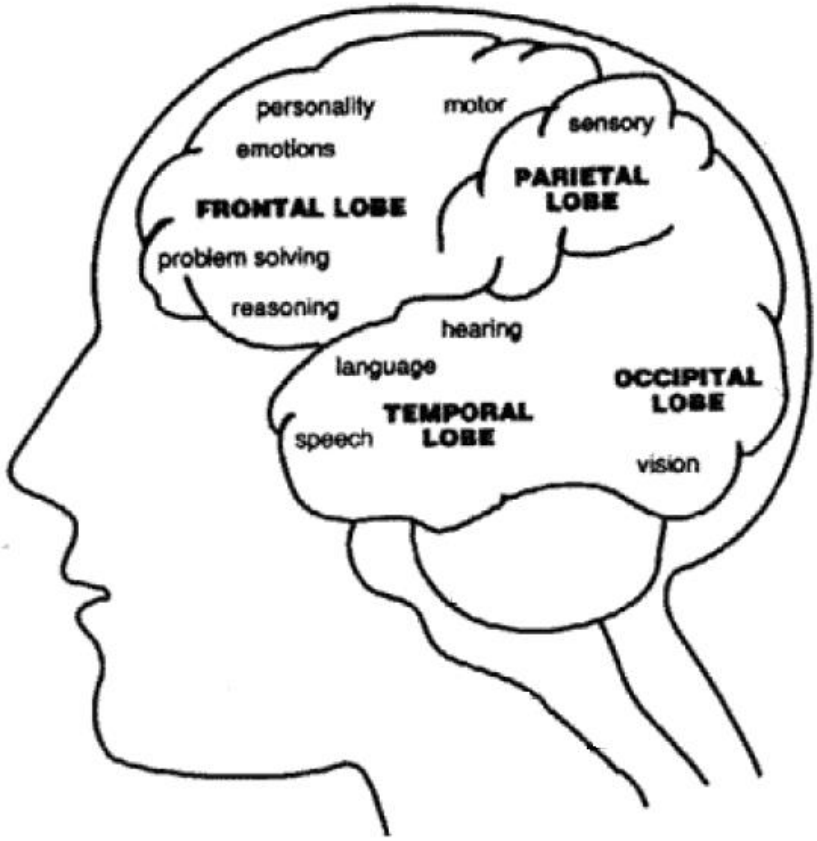


Figure 2.2: Structures of the brain

2.1.2 Temporal Lobes

The temporal lobes are highly associated with memory skills and are involved in the primary organization of sensory input (Read, 1981). This area of the brain is involved with emotional response, memory, and speech recognition. The responsibility of these lobes also includes language functions such as naming and verbal comprehension. Evidence suggests that the temporal lobes are involved in high-level visual processing of complex stimuli and scenes as well as object perception and recognition. This part of the brain handles the transfer of memory from short to long term and control spatial memory.

2.1.3 Occipital Lobe

The ability to process visual images is located in the occipital lobe. This part of the brain handles the perception of motion, color discrimination and visual/spatial processing.

2.1.4 Parietal Lobe

There are several functions carried out by this part of the brain. First, the cognitive functions of sensation and perception. This sensory input is then integrated to form a corresponding spatial coordinate system to the environment. The parietal lobe has been associated with various visuo-spatial abilities and analogical mental rotations (Dehaene et al., 1999).

2.1.5 Frontal Lobe

The frontal lobe area of the brain involved several important activities including motor function, problem solving, memory, language, judgment, impulse control, and social behavior. The left and right frontal lobes are involved in different behaviors, for example, the left controls language related movement (e.g. muscle activation necessary for speech) and the right lobe is involved with non-verbal abilities.

2.2 The Neuron

There are about 10^{10} neurons in CNS organized in a multilevel hierarchical system (Shepherd, 1998). The nervous system provides a lot of diversity of neuron type, connectivity, functionality, etc. Therefore, pretty much all of what is said refers to the most common behaviour despite the whole variability present (Figure 2.3).

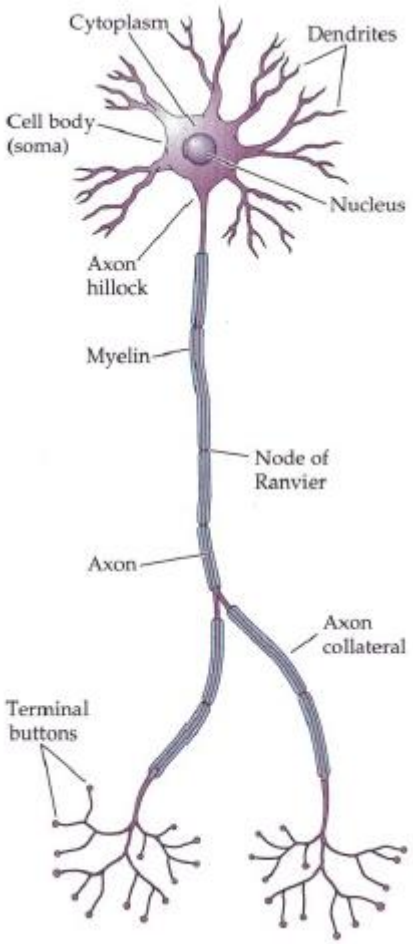


Figure 2.3: Schematic representation of a neuron.

The three main parts of a neuron are the dendrites, the soma (cell body) and the axon. Most of the incoming current to a neuron comes from the dendrites. Probably the great distinctive features of neurons is the presence of large dendritic trees. They are responsible for most of the variety in neuron size, shape and types. The dendritic tree contains many post synaptic terminals of chemical synapses. Several functions (Stuart et al., 1999) have also been claimed to be performed by dendritic arbors such as biological gates and coincidence detectors, learning signaling by dendritic spikes, to increase the learning capacity of the neuron (Poirazi and Mel, 2001) or to increase the

ability to differ incoming stimulus intensity into a neuron (or enhance the dynamic, Gollo et al., 2012). However such dendritic computation properties are still far away from been clearly understood. The cell body (soma) contains the nucleus and most of the cytoplasmic organelles. It is mainly where the metabolic process occur. The axon goes very far away from the soma. It might have different size (from 0.1 to 2.000 mm) depending on its functionality (Kandel et al., 2000). It starts at the axon hillock where the action potential is generated and present ramifications at the extremities. From those terminal buttons come out most of the pre synaptic terminals. It might be involved by myelin to protect and control some properties as the propagation velocity.

2.2.1 The action potential

The neurons are nonlinear excitable elements, i.e., they generate a spike when its membrane potential goes above a defined threshold (about 20-30 mV above the rest potential, Gerstner and Kistler, 2002). This excitation is also called action potential (Figure 2.4). When the membrane potential of a given neuron is perturbed, for instance via the incoming activity from a neighbour, it relaxes back to its rest potential in a time scale determined by the membrane time (τm) if it does not exceed the threshold. The spike is generated in a particular region called axon hillock located in between the soma and the axon. The pulse propagates (Bishop and Davis, 1960) mainly throw the axon (forward propagation) but may also propagate in the other direction (backpropagating spike, Falkenburger et al., 2001). The spike occurs in a very narrow time window followed by a fall of the membrane potential bellow the rest state. At that point, the neuron is hyperpolarized and its potential difference is greater with respect to the exterior region (arbitrarily defined as 0 mV). This stage is called refractory period and the neuron is typically not allowed to reach the threshold and consequently

to spike. Typically, the membrane potential relaxes to the rest potential before another cycle happens.

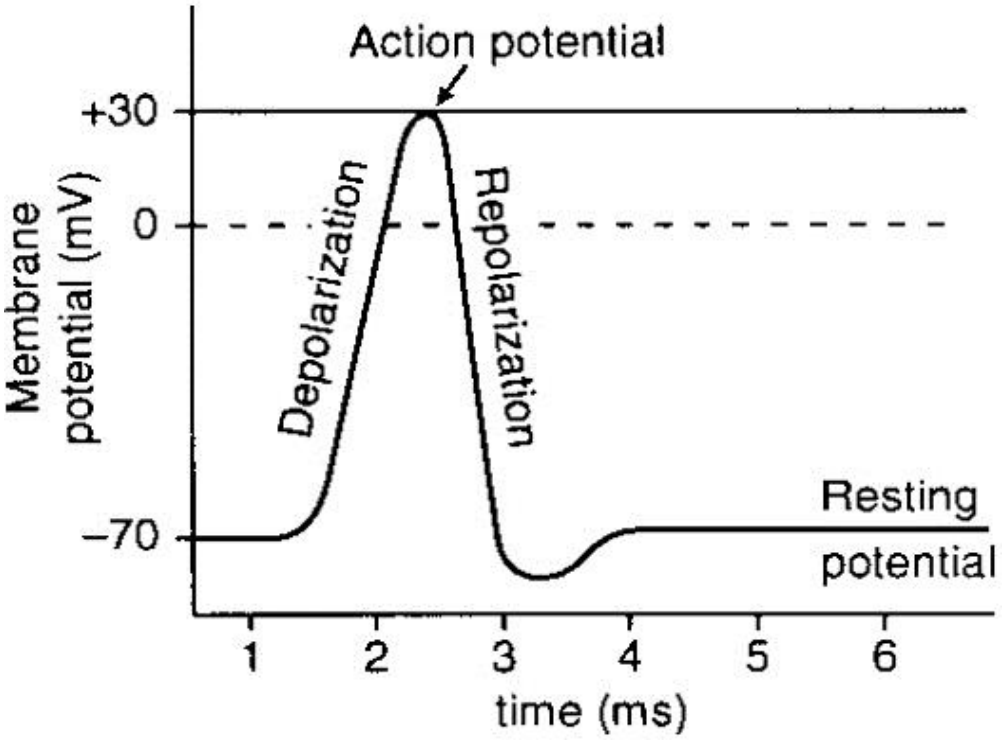


Figure 2.4: The action potential

2.3 The electroencephalography

The EEG is a recording of the brain's electrical activity, in most cases, made from electrodes over the surface of the scalp or from needle electrodes inserted into the brain. One of the first ever reports about EEG was by Richard Caton (1875), who recorded the EEG oscillations from monkeys and rabbits. In 1929, Hans Berger reported the first reliable recording of the EEG from a human scalp and a first categorization of EEG oscillation into alpha (8-13 Hz) and beta waves (14-30 Hz). Here, we refer EEG only to that measured from the head surface. Generally, the EEG recordings could be categorized into two types: the spontaneous activity and the

evoked potentials. Spontaneous activity is often referred to the unprovoked occurrence of brain activity, in terms of the absence of an identifiable stimulus, with or without behaviour manifestation. The bandwidth of this signal is from under 1 Hz to over 100 Hz. The evoked potentials are time-locked components in the EEG that arise in response to a stimulus, which may be electric, visual, auditory, tactile, etc. Such signals are often evaluated by averaging a number of trials to improve the signal-to-noise ratio. EEG is measured using scalp electrodes, which record the difference in the electric potential between an electrode with an active neural signal and an electrode placed over a supposedly inactive region that serves as a reference. These recordings are the resultant field potentials containing many active neurons. However, the action potential in axons is revealed to contribute little to the scalp surface records, as they are asynchronous while the axons run in many different directions. Surface records are thought to be the net effect of local postsynaptic potentials of the cortical cells. Mostly, the EEG measures the currents that flow during synaptic excitations of the dendrites of many pyramidal neurons, a type of neuron found in areas of the brain including the cerebral cortex (Teplan, 2002). Although there are various EEG recording systems in the market, such systems conventionally include four parts: electrodes with conductive media, amplifiers with filters, A/D converter, and recording device. Electrodes are used to read the signal from the scalp; amplifiers increase the magnitude of the microvolt signals into a range which can be digitalized accurately; the converter changes the signals from analog to digital form; and the recorder system (normally personal computer) stores and displays the obtained data (Teplan, 2002). Additionally, a 10-20 system (Figure 2.5) EEG measurement has been adopted by the International Federation in Electroencephalography and Clinical

Neurophysiology (Jasper, 1958). Such a system provides the standardized physical placement of electrodes on the scalp. The electrodes are labelled according to adjacent brain areas: F (frontal), C (central), T (temporal), P (posterior), and O (occipital), with odd numbers on the left side and even numbers on the right side.

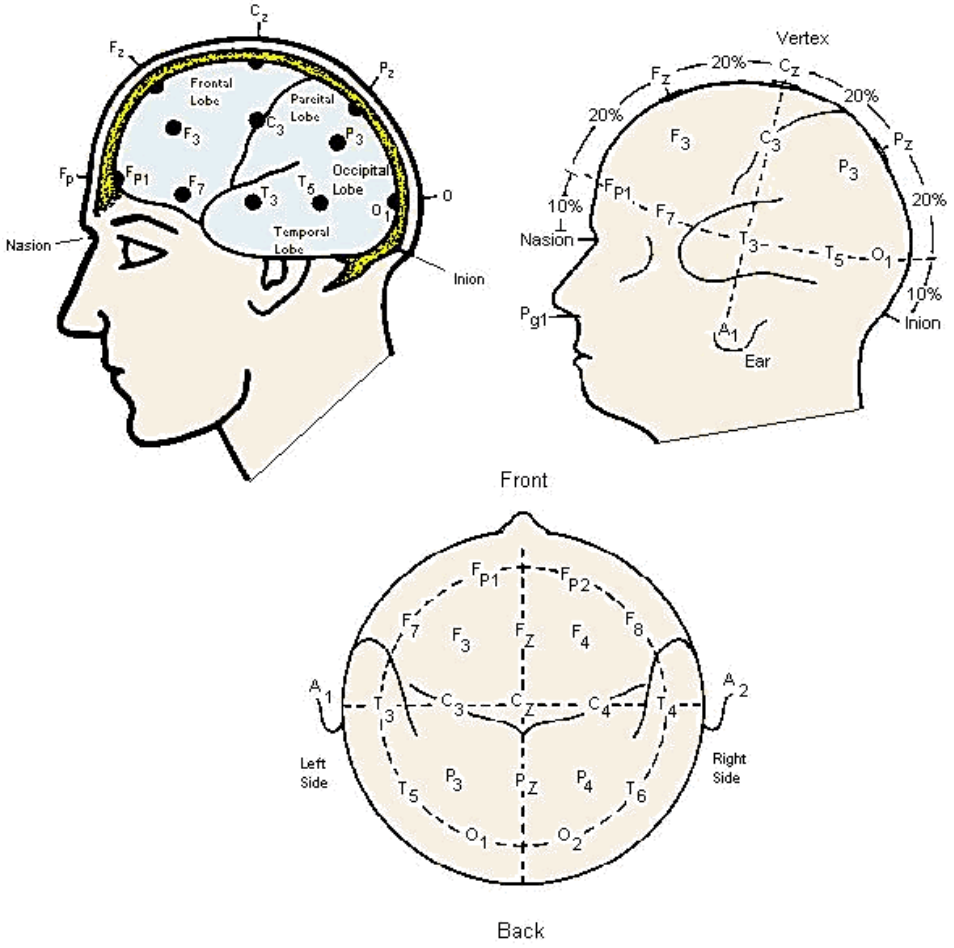


Figure 2.5: 10-20 system for the standardized electrode placement.

Two basic approaches are commonly used for the EEG analysis: (i) the analysis of evoked potentials (EPs); and (ii) the power spectrum analysis (EEG rhythms). These two methods have been applied in various experimental or field researches into human cognitive activities.

2.3.1 The evoked potentials

The evoked potential (EP) is a response induced by the presentation of an external stimulus that can be isolated from the electroencephalographic spontaneous activity. This means that, for any external stimulation, the brain reacts with a specific wave, characterized by a particular latency, an amplitude and a polarity. A given evoked potential appears at a time interval approximately constant from the presentation of the stimulus. Because the amplitude of each EP is smaller than the fluctuations in the amplitude of the spontaneous EEG, normally the EP is extracted from the EEG as the average of a series of single responses (synchronized averaging) in order to remove the random fluctuations of the EEG. In this way, the EEG variations which are not synchronized with the stimulus are deleted, while the EPs become more visible. From a morphological point of view, the EP is named according to the polarity of the peak that can be positive or negative (P or N) and to the latency with respect to the onset of the external stimuli. From the physiological point of view, the evoked potentials are defined as the electrical changes that occur in the central nervous system in response to an external stimulus: in this way, their latency and amplitude depend on the physical characteristics of the stimulus applied (e.g. tone and intensity for the auditory system; contrast, luminance, and spatial frequency for the visual system; intensity and stimulation mode for the somatosensory system). The evoked potentials are categorized into two basic types: the evoked potentials stimulus related (e.g. visual EPs, VEPs), which morphology depends from the physical characteristics of the stimulus, and the event related potentials (ERPs), which generation is independent from the physical characteristics of the stimulus but reflects the attentional resources of the subject.

2.1.1.3 *Visual Evoked Potentials (VEPs)*

The terms visually evoked potential (VEP), visually evoked response (VER) and visually evoked cortical potential (VECP) are equivalent. They refer to electrical potentials, initiated by brief visual stimuli, which are recorded from the scalp overlying visual cortex, VEP waveforms are extracted from the electroencephalogram (EEG) by means of a signal averaging synchronized to the onset of the stimuli. VEPs are used primarily to measure the functional integrity of the visual pathways from retina via the optic nerves to the visual cortex of the brain. Visually evoked potentials elicited by flash stimuli can be recorded from many scalp locations in humans. Visual stimuli stimulate both primary visual cortices and secondary areas. Clinical VEPs are usually recorded from occipital scalp overlying the calcarine fissure. This is the closest location to primary visual cortex. The time period analyzed is usually between 50 and 300 milliseconds following the onset of each visual stimulus. The most common stimulus used is a checkerboard pattern, which reverses every half-second. Pattern reversal is a preferred stimulus because there is more inter-subject VEP reliability than with flash or pattern onset stimuli.

In the morphology of a VEP it is possible to differentiate few components (Figure 2.4): there is a prominent negative component at peak latency of about 70 ms (N1), a larger amplitude positive component at about 100 ms (P1) and a more variable negative component at about 140 ms (N2). The major component of the VEP is the large positive wave peaking at about 100 milliseconds. This "P100″ or P1 in the jargon of evoked potentials, is very reliable between individuals and stable from about age 5 years to 60 years. The mean peak latency of the "P100″ only slows about one millisecond per decade from 5 years old until 60 years old.

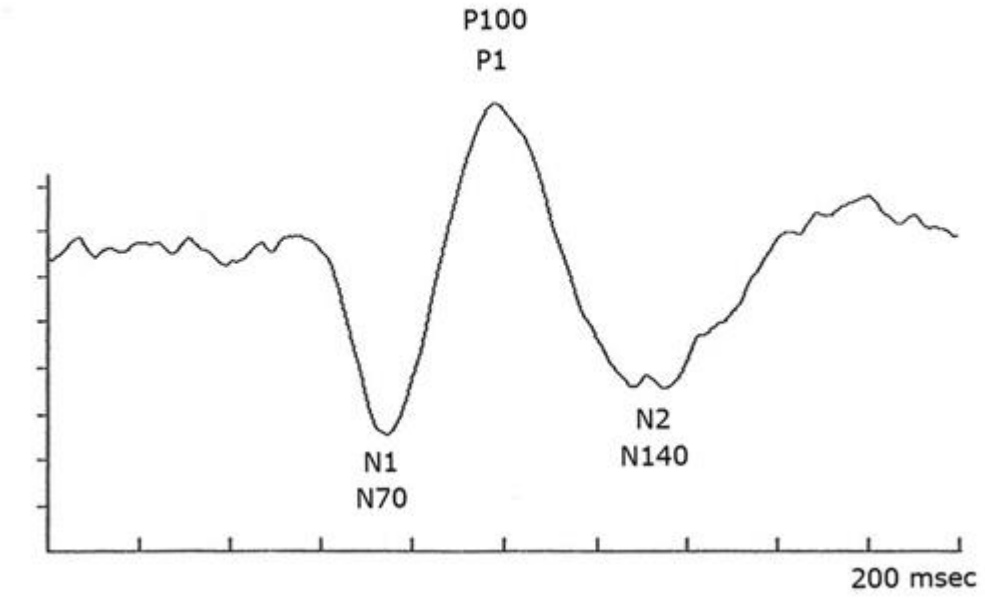


Figure 2.6: Representative normal pattern reversal VEP recorded from mid-occipital scalp using 50′ checkerboard pattern stimuli.

2.1.2.3 *Event Related Potentials (ERPs)*

Event-related potentials (ERPs) represent the voltage fluctuations that are associated in time with some physical or mental occurrence (Picton et al., 2000). ERP is a complex potential consisting of both time-locked fast and slow components, which could both precede an event or follow it (Kotchoubey, 2006). Since the late 1950s, the ERP analysis has been established as a psychophysiological approach to provide information about the cognitive processing of an event or a stimulus in the brain. ERP components are supposed to allow obtaining information about how the intact human brain processes signals and prepares actions.

Specifically, the event related potential (ERP) P300 is a positive deflection of the EEG signal elicited in the process of decision-making (Fabiani et al., 1987). The P300 (P3) wave is an event related potential (ERP) component elicited in the process of decision-making. It is assumed an endogenous potential, as its occurrence links not to the physical attributes of a stimulus, but to a person's reaction to it. More specifically, the P300 is thought to reflect processes involved in stimulus evaluation or categorization. It is usually elicited using the oddball paradigm, in which low-probability target items are mixed with high-probability non-target (or "standard") items. The P300 component is measured by assessing its amplitude and latency. Amplitude is defined as the difference between the mean pre-stimulus baseline voltage and the largest positive-going peak of the ERP waveform within a time window (e.g., 250–500 ms, although the range can vary depending on stimulus modality, task conditions, subject age, etc.). Latency (ms) is defined as the time from stimulus onset to the point of maximum positive amplitude within a time window. P300 scalp distribution is defined as the amplitude change over the midline electrodes (Fz, Cz, Pz), which typically increases in magnitude from the frontal to parietal electrode sites (Johnson, 1996). The P300 potential can be evoked through different paradigms: The single-stimulus task presents an infrequent target (T) in the absence of any other stimuli. The oddball task presents two different stimuli in a random sequence, with one occurring less frequently than the other does (target = T, standard = S). In each task, the subject has to respond only to the target and otherwise to refrain from responding.

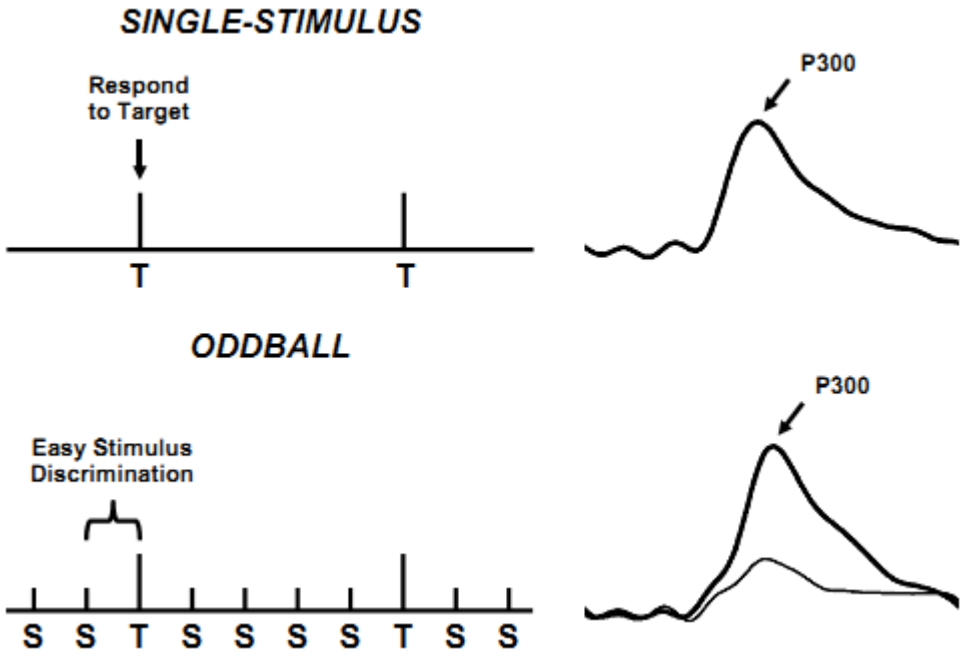


Figure 2.7: Schematic illustration of the single-stimulus (top) and oddball (bottom), with the elicited ERPs from the stimuli of each task at the right.

Latency and amplitude of the P300 potential can be influenced by several internal and external factors. Some of the determinants affecting P300 amplitude and latency include exercise and fatigue (Yagi et al., 1999), commonly used drugs, age, IQ, handedness, and gender, as well as some personality variables (Polich and Kok, 1995). Subjects who have eaten recently show a higher amplitude and a shorter latency than those who have not. Recent nicotine consumption affects both behavioral and P300 measures in some tasks (Houlihan et al., 1996). Caffeine, alcohol, and other substances have also been shown to influence the P300 morphology (Sommer et al., 1999). The amplitude and the latency of the P300 potential is reported to change with the age: latency increases of 1.8 ms/year and the amplitude decreases of 0.2μV/year. (Goodin et al., 1978). Rare no-target stimuli elicit different ERP components in

children and adults, while equally rare, target stimuli elicit similar components in children and adults (Courchesne, 1977). Also, the P300 increases in amplitude and decreases in latency with age (Polich et al., 1990). Auditory P300 has a centro-parietal distribution that increases in amplitude and decreases in latency (Martin et al., 1988) steadily from age 5 to age 19. Apart the physiological aspect, the P300 morphology is also dependent from the stimulation timing: the time between stimuli affects P300 amplitude, in particular the P300 potentials elicited with shorter timing within the stimuli have smaller amplitudes and longer latencies than those obtained with longer timing (Picton et al., 2000). The P300 potential is a measure of the attentional resources of the subject. In particular, the amplitude of the P300 is proportional to the amount of attentional resources engaged in processing a given stimulus (Johnson, 1986) and it is not influenced by factors related to response selection or execution (Crites et al., 1995). Gray et al., (2003) reported that the P300 amplitude therefore served as our covert measure of attention that arises independently of behavioral responding. Further, P300 latency is thought to reflect stimulus classification speed, such that it serves as a temporal measure of neural activity underlying attention allocation and immediate memory operations (Duncan and Johnson, 1981; Magliero et al., 1984; Polich, 1986). Finally, a large number of studies using ERPs to evaluate the mental user's load have been conducted which proved that the amplitude and latency of P300 provide effective tools for the assessment of mental workload (Johnson, 1986, for further details refer to the 2.6.1 section).

The ERPs serve as important adjuncts to studies of human information processing, a fundamental problem with this method is the signal-noise ratio. The magnitude of the ERP signal is around 5-10 μV, which is far smaller than the amplitude of the

background EEG (0-100 μV ; Hagemann, 2008). Therefore, the classic approach for ERP extraction is to average the signal over a number of trials in order to obtain a stable response with a sufficient signal-to-noise ratio. In this regard, in the section 3.3 a method to enhance the signal to noise ratio (SNR) and to extract the single epoch P300 potential was reported (Aricò et al., 2014, [J 1], [C 3]).

2.1.3.3 *Oddball paradigm*

The oddball paradigm is a method used in evoked potential research in which trains of stimuli (usually auditory or visual) are used to assess the neural reactions to unpredictable but recognizable events (Figure 2.6). It has been found that the P300 event related potential across the parieto-central area of the skull is larger after the target stimulus (Polich et al., 2007). In the oddball paradigm, two stimuli are presented in a random series such that one of them occurs relatively infrequently (target stimuli). The subject is required to distinguish between the stimuli by noting the occurrence of every target mentally counting, button press and by not responding to the standard stimulus (Polich and Margala, 1997). For example, in a visual oddball task, there might be a 95% chance for a square to be presented and a 5% chance for a circle. When the targets (e.g. circles) appear, the subject must make a response, such as pressing a button or updating a mental count. This task has provided much of the fundamental data for the theoretical interpretation of P300 in terms of memory updating (Johnson, 1986), as well as in studies that suggest P300 amplitude is proportional to the amount of attentional resources required for a given task (Wickens et al., 1983; Kramer and Strayer, 1988). The oddball paradigm was widely used in several works in the brain computer interface (BCI) field (see section 2.7 for further details).

2.3.2 EEG rhythms analyses

The oscillatory activity of the spontaneous EEG is typically categorized into five different frequency bands: delta (0-4 Hz), theta (4-7), alpha (8-12), beta (12-30) and gamma (30-100 Hz), as shown in Figure 2.8. These frequency bands are suggested to be a result of different cognitive functions.

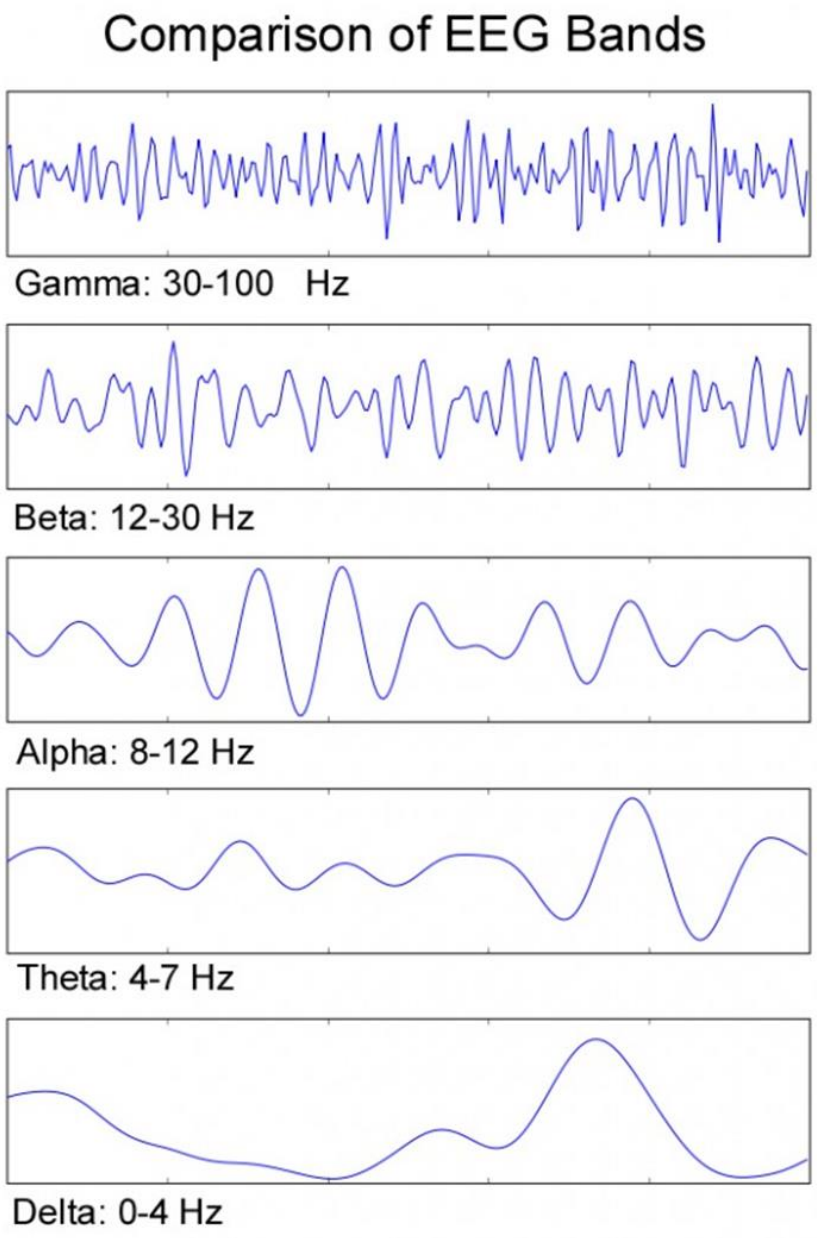


Figure 2.8: Comparison of EEG bands over one second of activity. Gamma (30-100Hz), Beta (12-30Hz), Alpha (8-12Hz), Theta (4-7Hz), and Delta (0-4Hz).

- Delta (0 -4 Hz): The delta activity is characterized as high amplitude and low frequency. It is usually associated with the slow-wave sleep in the sleep research. It is suggested that delta waves represent the onset of deep sleep phases in healthy adults (Rechtschaffen and Kales, 1968). In addition, contamination of the eye activity is mostly represented in the delta frequency band.

- Theta (4-7Hz): The generation of theta power is associated with the hippocampus (Buzsáki, 2002) as well as neocortex (Cantero et al., 2003). The theta band is thought to be associated with deep relaxation or meditation (e.g. Hebert and Lehmann, 1977; Kubota et al., 2001) and it has been observed at the transition stage 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 et al., 2001; Smith et al., 2001) and is often shown to increase with a higher cognitive task demand (e.g. Gundel and Wilson, 1992; Gevins et al., 1998).

- Alpha (8-12Hz): The alpha band activity is found at the visual cortex (occipital lobe) during periods of relaxation or idling (eyes closed but awake). It is characterized by high amplitude and regular oscillations with a maximum over parietal and occipital electrodes in the continuous EEG. The modulation of alpha activity is thought to be a result of resonation or oscillation of the neuron groups (Lopes da Silva et al., 1980; Smith et al., 2001). High alpha power has been assumed to reflect a state of relaxation or cortical idling. However, when the operator devotes 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 suggested that alpha is involved in auditory attention processes and the inhibition of task irrelevant areas to enhance signal-to-noise ratio (Cooper et al., 2006; Klimesch et al., 2007; Hagemann, 2008). Additionally, some researchers divide the alpha activity further into sub-bands to achieve a finer grained description of its functionality (e.g. Klimesch et al., 1999). For instance, the "mu" band (10-12 Hz) occurs with actual motor movement and intent to move with an associated diminished activation of the motor cortex (Dooley 2009).

- Beta (13-30Hz): The beta wave is predominant when the human is awake. Spatially, it predominates in the frontal and central areas of the brain. It has been described that the high power in the beta band is associated with the increased arousal and activity. Dooley (2009) pointed out that the beta wave represents cognitive consciousness and an active, busy, or anxious thinking. Furthermore, it has been revealed to reflect visual concentration and the

orienting of attention (Birbaumer and Schmidt, 1996). The beta band can be further divided into several sub-bands: low beta wave (12.5-15 Hz); middle beta wave (15-18 Hz); high beta wave (> 18 Hz). These three sub-bands are associated with separate physiological processes. For instance, the high beta waves are suggested to be linked with the dopaminergic system (Gruzelier et al., 1990; Hagemann, 2008), while the low beta activities are thought to reflect the inhibition of phasic movements during sleep (Hagemann, 2008).

- Gamma (30-100Hz): The gamma band is the fastest activity in EEG and is thought to be infrequent during waking states of consciousness (Dooley, 2009). It is reported that gamma waves are associated with perceptual blinding problem (Gray et al., 1989). More specifically, Tallon-Baudry et al. (2005) revealed that areas of lateral occipital cortex play an important role in visual stimulus encoding and show large gamma oscillations differently affected by attentional modulation. Recent studies reveal that gamma is linked with many other cognitive functions such as attention, learning, memory (Jensen, et al., 2007), and language perception (Eulitz et al., 1996).

2.4 The electrocardiography

The electrocardiogram (ECG) interprets the electrical activity caused by depolarization and polarization of the heart muscle. It reflects the electrical impulses produced by heart contraction. The ECG can be analyzed by means of three approaches: (a) time domain measures; (b) amplitude measures; and (c) frequency domain measures.

A typical time-domain ECG tracing of the cardiac cycle (heartbeat) consists of a P wave, a QRS complex, a T-wave, and a U-wave (Figure 2.9). The QRS complex is often used to detect peaks while the time between peaks i.e. namely, Inter-Beat-Interval (IBI), can be extracted. Typically, heart rate (HR) and heart rate variability (HRV) are widely used for the representation of the mental workload. HR is determined by the number of heart beats within a fixed period of time (usually per minute) and is non-linearly related to IBI. Compared with IBI, HR is less normally distributed in samples (Jennings et al., 1974). Additionally, the amplitude of T-wave (TWA) is another variable in the ECG signal reflecting sympathetic nervous system (SNS) activity (Furedy, 1987). Müller et al. (1992) reported that the amplitude of TWA decreased with increases in SNS activity

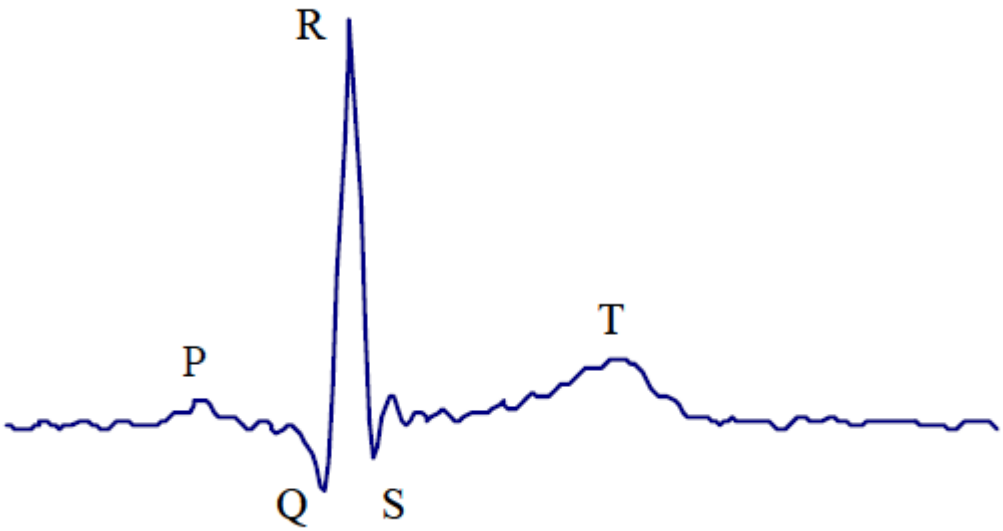


Figure 2.9: The typical time-domain ECG tracing of the cardiac cycle

Compared with HR and TWA, the analysis of HRV is more complex. HRV is usually defined as the changes in the interval between heart beats in either time or frequency domain. It reflects the irregularities in heart rate caused by a continuous feedback

between the CNS and peripheral autonomic receptors. Three frequency components have been defined: a very low frequency range (VLF; 0.02-0.06 Hz), a low frequency range (LF; 0.06-0.15 Hz; also called '0.1 Hz' component), and a high frequency range (HF; 0.15-0.4 Hz). The VLF is believed to be linked to the regulation of the body temperature; LF is assumed to be involved in the regulation of short-term blood pressure; HF is shown to be related to respiratory fluctuations reflecting parasympathetic influences that are dependent on respiration frequency (Kramer, 1990; Grossman, 1992).

2.5 The attention: Overt vs Covert

Each time we open our eyes we are confronted with an overwhelming amount of information. Despite this, we experience a seemingly effortless understanding of our visual world. This requires selecting relevant information out of irrelevant noise. Attention is the key to this process; it is the mechanism that turns looking into seeing. In perception, ignoring irrelevant information is what makes it possible for us to attend to and interpret the important parts of what we see. Attention allows us to selectively process the vast amount of information with which we are confronted, prioritizing some aspects of information while ignoring others by focusing on a certain location or aspect of the visual scene. In general, three typologies of attention modalities can be identified:

- Selective attention: The ability to process or focus on one message in the presence of distracting information.

- Divided attention: The ability to process more than one message at a time.

- Visual attention: The mechanism determining what information is or is not extracted from our visual field.

The appeal of visual attention seems to be related to an observation that is likely to disconcert a traditional vision scientist: changing an observer's attentional state while keeping the retinal image constant can affect perceptual performance and the activity of 'sensory' neurons throughout visual cortex. For over a century, the study of visual attention has attracted some of the greatest thinkers in psychology, neurophysiology and perceptual sciences, including Hermann von Helmholtz, Wilhelm Wundt and William James. More recently (1960–1980s), many psychologists, including Michael Posner, Anne Treisman, Donald Broadbent and Ulric Neisser, have provided distinct theories and developed experimental paradigms to investigate what attention does and what perceptual processes it affects. Initially, there was a great deal of interest in categorizing mechanisms of vision as pre-attentive or attentive. The interest in that distinction has waned as many studies have shown that attention actually affects tasks that were once considered pre-attentive, such as contrast discrimination, texture segmentation and acuity. The influence of attention increases along the hierarchy of the cortical visual areas, resulting in a neural representation of the visual world affected by behavioral relevance of the information, at the expense of an accurate and complete description of it (e.g., Treue, 2001). Attention can affect perception by altering performance – how well we perform on a given task –and/or by altering the subjective appearance of a stimulus or object. These aspects will be discussed in the sections 3.1 and 3.2, in which the effects of the overt and covert attention modalities in the performance of two P300 BCI systems will be reported (Aloise et al., 2012a, [J 1] [J 6] [J 7] [J 10] [C 7] [C 28]).

There are three main types of visual attention: i) spatial attention, which can be either overt, when an observer moves his/her eyes to a relevant location and the focus of attention coincides with the movement of the eyes, or covert, when attention is deployed to relevant locations without accompanying eye movements; ii) feature-based attention (FBA), which can be deployed covertly to specific aspects (e.g., color, orientation or motion direction) of objects in the environment, regardless of their location; and iii) object-based attention in which attention is influenced or guided by object structure (Olson, 2001; Scholl, 2001).

2.5.1 Spatial (c)overt attention

Attention can be allocated by moving one's eyes toward a location (overt attention) or by attending to an area in the periphery without actually directing one's gaze toward it (covert attention). The deployment of covert attention aids us in monitoring the environment and can inform subsequent eye movements. Hermann von Helmholtz was the first scientist to provide an experimental demonstration of covert attention (Suzuki and Cavanagh, 1997). Looking into a wooden box through two pinholes, Helmholtz would attend to a particular region of his visual field (without moving his eyes in that direction). When a spark was lit to briefly illuminate the box, he found an impression of only the objects in the region he had been attending to, thus showing that attention could be deployed independently of eye position and accommodation. In general, to investigate covert attention, it is necessary to ensure that observers' eyes remain fixated at one location, and to keep both the task and stimuli constant across conditions while manipulating attention. Spatial resolution, our ability to discriminate fine patterns, is not uniform across locations in the visual field. It decreases with

eccentricity. Correspondingly, signals from the central parts of the visual field are processed with greater accuracy and faster reaction times (e.g., Cannon, 1985; Carrasco, Evert et al., 1995; Rijsdijk et al., 1980). In many tasks, these performance differences are eliminated when stimulus size is enlarged according to the cortical magnification factor, which equates the size of the cortical representation for stimuli presented at different eccentricities (e.g., Rovamo and Virsu, 1979). There are several factors contributing to differences in spatial resolution across eccentricities. A greater proportion of the cortex is devoted to processing input from the central part of the visual field than from the periphery (cortical magnification) in many cortical visual areas (Sutter, 1992).

2.6 The mental workload

The mental workload is a measure of the resources required to process information during a specific task (O'Donnell and Eggemeier, 1986). Workload concept can be divided into five dimensions: instantaneous workload, peak workload, accumulated workload, average workload, and overall workload. The instantaneous workload measures dynamic changes in the workload values during task performance. The typical examples for such measures are the physiological markers. The peak workload is referred to as the maximal value of instantaneous workload. Accumulated workload is the total amount of instantaneous workload. The average workload is defined as the average of the instantaneous workload. Finally, the overall workload is the individual's experienced mental workload which maps instantaneous workload (or accumulated and averaged workload) in the operator's brain (Xie and Salvendy, 2000). In general, the mental workload is thought of as a mental construct, a latent

variable, or perhaps an "intervening variable" (Gopher and Donchin 1986), reflecting the interaction of mental demands imposed on operators by tasks they attend to. The capabilities and the effort of the operators in the context of specific situations all moderate the workload experienced by the operator. Workload is thought to be multidimensional and multifaceted. Workload results from the aggregation of many different demands and so is difficult to define uniquely. Casali and Wierwille (1984) note that as workload cannot be directly observed, it must be inferred from observation of overt behavior or measurement of psychological and physiological processes. Gopher and Donchin (1986) feel that no single, representative measure of workload exists or is likely to be of general use, although they do not provide guidance on how many workload measures they feel are necessary or sufficient (Cain, 2007). Mental workload can be influenced by numerous factors that make a definitive measurement difficult. Jex (1988) implies that mental workload derives from the operator's meta-controller activities: the cognitive "device" that directs attention, copes with interacting goals, selects strategies, adjusts to task complexity, sets performance tolerances, etc. This supports the intuitive notion that workload can be represented as a function, and the utility of univariate workload measures as globally sensitive estimates of workload, while acknowledging that tasks of differing characteristics interfere differently. Alternatively, Wierwille (1988) suggests that an operator faced with a task is fully engaged until the task is done, then is idle or engages in another task.

2.6.1 Workload measurement techniques

The principal reason for measuring workload is to quantify the mental cost of performing tasks in order to predict operator and system performance. As such, it is an interim measure and one that should provide insight into where increased task demands may lead to unacceptable performance. In the comparison of system designs, procedures, or manning requirements, workload measurement can be used to assess the desirability of a system if performance measures fail to differentiate among the choices. Implicit in this approach is the belief that as task difficulty (workload) increases: performance usually decreases; response times and errors increase; control variability increases; fewer tasks are completed per unit time; task performance strategies change (Huey and Wickens 1993); and, there is less residual capacity to deal with other issues.The mental workload can be evaluated using mainly three approaches: i) subjective or self-assessment evaluation, ii) performance evaluation and iii) psychophysiological variables assessment.

2.6.2 Subjective evaluation

Subjective measures have been used extensively to assess operator workload in many studies (Tsang and Johnson, 1989; Zaklad, and Christ, 1989; Eggemeier and Stadler, 1984). The reasons for the frequent use of subjective procedures include their practical advantages (ease of implementation, non-intrusiveness) and current data which support their capability to provide sensitive measures of operator load. Many subjective procedures exist to measure mental workload. The most outstanding among them are the Cooper-Harper Scale (Cooper and Harper, 1969), the Bedford Scale

(Roscoe and Ellis, 1990), the SWAT (Subjective Assessment Technique) (Reid and Nygren, 1988) and the NASA-TLX (Task Load Index) (Hart and Staveland, 1988). Between these procedures, we will take into account only the NASA-TLX questionnaire that we used in many reported works (See section 2.2.1.6). Self-assessments involve rating demands on numerical or graphical scales, typically anchored either at one or two extremes per scale. Some subjective techniques use scales that are categorical, with definitions at every level, such as the Modified Cooper-Harper scale. Other techniques use an open-ended rating with a "standard" reference task as an anchor and subjects rate other tasks relative to the reference task. Despite this kind of measure is quite direct because the subject him/herself assesses the perceived workload, the repeatability and validity of such quantitative subjective techniques are sometimes uncertain and data manipulations are often questioned as being inappropriate.

2.2.1.6 *NASA-Task Load Index (TLX)*

The NASA Task Load Index (Hart and Staveland, 1988) uses six dimensions to assess mental workload:

1. Mental demand: How much mental and perceptual activity was required? Was the task easy or demanding, simple or complex?

2. Physical demand: How much physical activity was required? Was the task easy or demanding, slack or strenuous?

3. Temporal demand: How much time pressure did you feel due to the pace at which the tasks or task elements occurred? Was the pace slow or rapid?

4. Performance: How successful were you in performing the task? How satisfied were you with your performance?

5. Effort: How hard did you have to work (mentally and physically) to accomplish your level of performance?

6. Frustration: How irritated, stresses, and annoyed versus content, relaxed, and complacent did you feel during the task?

Twenty-step bipolar scales are used to obtain ratings for these dimensions. A score from 0 to 100 (assigned to the nearest point 5) is obtained on each scale. A weighting procedure is used to combine the six individual scale ratings into a global score; this procedure requires a paired comparison task to be per-formed prior to the workload assessments. Paired comparisons require the operator to choose which dimension is more relevant to workload across all pairs of the six dimensions. The number of times a dimension is chosen as more relevant is the weighting of that dimension scale for a given task for that operator. A workload score from 0 to 100 is obtained for each rated task by multiplying the weight by the individual dimension scale score, summing across scales, and dividing by 15 (the total number of paired comparisons).

2.6.3 Performance evaluation

The performance evaluation provides a direct correlation between the performance achieved by the subject during the task and the required mental workload. It can be classified into two major types: primary task measures and secondary task measures. In most investigations, performance of the primary task will always be of interest as its generalization to in-service performance is central to the study. Primary task measures

attempt to assess the operator's performance on the task of interest directly, and this is useful where the demands exceed the operator's capacity such that performance degrades from baseline or ideal levels. Speed, accuracy, reaction or response times, and error rates are often used to assess primary task performance (e.g. Multi Attribute Task Battery, MATB, Comstock, 1994).

In secondary task methods, performance of the secondary task itself may have no practical importance and serves only to load or measure the load of the operator. Secondary task measures provide an index of the remaining operator capacity while performing primary tasks, and are more diagnostic than primary task measures alone. The characteristics of the secondary task are used to infer the interaction between the primary and secondary tasks and this approach is frequently used when the operator can adapt to demand manipulations such that primary-task performance is apparently unaffected (Colle and Reid,1999).

2.3.1.6 *Multi Attribute Task Battery (MATB)*

The Multi-Attribute Task Battery (MATB, Figure 2.10) provides a benchmark set of tasks for use in a wide range of laboratory studies about operator performance and workload (Comstock, 1994). The MATB simulates the activities inside an aircraft's cockpit and provides a high degree of experimental tasks control in terms of complexity and difficulty. Furthermore, task features include an auditory communications task (to simulate Air-Traffic-Control communications), a fuel resources management task of maintaining target performance (e.g. to keep the fuel level around 2500 lbs), an emergency lights control and a task of cursor tracking, that

is, it simulates the control of the aircraft flight level (this can be switched from manual to automatic mode).

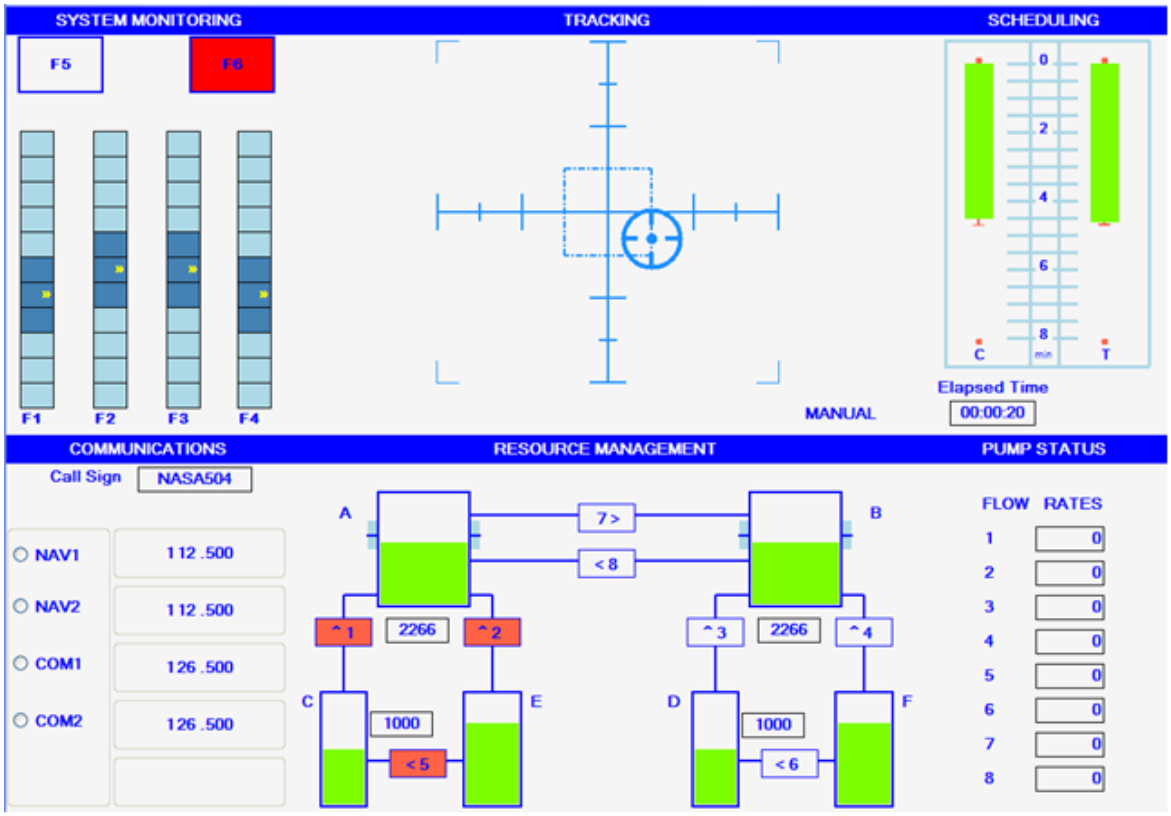


Figure 2.10: Screenshot of the Multi Attribute Task Battery (MATB) interface. On the top left corner, there is the emergency lights task; on the top, in the center, there is the task of cursor tracking; on the left bottom corner, there is the radio communication task and, finally, in the center on the bottom, there is the fuel levels managing.

2.6.4 Psychophysiological variables assessment

Finally, the psycho - physiological measure, consists in the evaluation of the variability (and of the correlation) of one or more neurophysiological signals

(electroencephalogram (EEG), electrocardiogram (ECG), galvanic skin response (GSR), etc.) with respect to the mental workload required to the subject during the task (Kramer, 1990; Hancock and Desmond, 2001). This class of measures is based on the concept that increasing workload, for example by means of the increment of mental demand, leads to an activation in physical response from the body. Normally, a requirement of most psychophysiological measures is for reference data that establishes the operator's unstressed background state. Such background states are subject to many factors and may change markedly over time so an operational baseline state is often used. For the purpose of this work, we will take into account only the EEG and the ECG signals as a measure of the user's mental workload.

2.4.1.6 *Electroencephalography*

Characteristic changes in the EEG reflecting levels of mental workload have been identified in different works. In general, two kind of EEG features can be took into account for the representation of the human operator mental workload: ERPs and EEG rhythms modulation.

Event Related Potentials measurement

In the last decades, a large number of studies using ERPs to evaluate the mental workload have been conducted which proved that the amplitude and latency of P300 provide effective tools for the assessment of mental workload (Johnson, 1986; Schultheis and Jameson, 2004). For the workload assessment, three features from the P300, the latency, the latency jitter and the amplitude, are used.

P300 latency: Different studies demonstrated that the P300 latency provides a chronometric index for assess the duration of perceptual processing (Leuthold and Sommer, 1998). Kramer and Parasuraman (2007) also pointed out that the latency of P300 reflects the timing of stimulus identification and categorization processes. Previous research indicated that increasing the mental workload may lead to an extension of the P300 latency. Kutas et al. (1977) stated that increasing the difficulty of identifying the target stimulus also increased the latency of the P300 wave. Such conclusion was confirmed also by Fowler (1994). However, an increasing in the difficulty of response selection do not affect P300 latency (Magliero, 1984). This led to a discussion on whether the latency of the P300 provides a relatively pure measurement of perceptual processing and categorization time, independent of response selection and execution stages (Kutas et al., 1977; McCarthy and Donchin, 1981).

P300 latency jitter: This phenomenon happens when the lag between the onset of the stimulus of interest and the evoked P300 potential peak is not constant over the different stimuli. In different studies, the authors demonstrated that large latency variations were observed when the attention was divided between two tasks (Polich, 2007, Kutas et al., 1977). Aricò et al., (2014) demonstrated that a ERP-BCI used in covert attention modality, increase the workload perceived with respect to the overt attention, and at the same time the P300 latency jitter over the target stimuli significantly increases (for further details, please refer to the section 3.3).

P300 amplitude: It has been assumed that the amplitude of P300 is proportional to the amount of attentional resource allocation for the task performance (Johnson, 1986).

This assumption is in line with the findings in the oddball paradigm that the amplitude of P300 is sensitive to the probability of the presentation of stimulus. Gopher and Donchin (1986) suggested that the P300 amplitude could index the perceptual/central processing load, until the moment performance declines, in which case the amplitude remains unaffected. It is assumed that the amplitude of P300 may show different changes in the single and dual task performance. In a primary-task-only-condition, it was suggested that the P300 amplitude increases with task complexity. In a dual-task paradigm, the diversion of processing resources away from target discrimination leads to a reduction in P300 amplitude (Kramer and Parasuraman, 2007).

EEG rhythms modulation

An extensive body of literatures exists concerning the EEG spectra modulation according to the variation of the cognitive workload and the allocation of mental effort (Gundel and Wilson, 1992; Berka et al., 2007; Lei et al., 2009; Lei and Roetting, 2011) and applied settings (Wilson, 2002; Kohlmorgen et al., 2007; Aricò et al., 2013). Several studies described the correlation of spectral power of the electroencephalogram (EEG) with the complexity of the task that the subject is performing. In fact, an increase of the theta band spectral power (4 - 7 (Hz)) especially on the frontal cortex and a decrease in alpha band (8-12 (Hz)) over the parietal and occipital cortexes have been observed when the required mental workload increases (Lei and Roetting, 2011; Borghini et al., 2012). Specifically, at the Fz site the theta power was increased during high-load task relative to low-load task, whereas alpha power tended to be attenuated in the high-load task compared to low-load tasks. Consistent results have been found not only in similar working memory (WM) task

(Gundel and Wilson, 1992; Gevins et al., 1998), but also in more complex cognitive tasks (Smith et al., 2001; Wilson, 2002; Wilson and Russell, 2003). Smith et al. (2001) recorded continuous EEG while 16 participants performed versions of the compute based flight simulation task, the Multiple-Attribute Task Battery (MATB; Comstock, 1994), in low, moderate and high difficulty. As task difficulty increased, frontal midline theta EEG activity increased while parietal midline alpha decreased. In field research, Wilson (2002) reported a study involving ten pilots who flew an approximately 90-minute scenario containing both visual and instrument flight conditions. Multiple variables including EEG parameters were analyzed. Wilson (1992) found that parietal alpha band showed significant reduction in high workload condition, but an increasing in the theta power spectrum could only be observed at a few scattered electrode sites. However, disputing voices on theta power can be also heard. Decreases in theta activity were found with transitions from single to dual-tasks. Pigeau et al. (1987) revealed that theta power initially increases with increments in the task difficulty of an additional task and then decreases at high levels of difficulty. Alpha oscillation was found to systematically decrease in power as the task load increases. This inverse proportion has been found in numerous earlier studies (Sterman et al., 1988; Gevins et al., 1998) and is consistent with current understanding of the underlying neural mechanisms in the generation of the alpha rhythms.

Many studies attempted to combine the EEG parameters for a reliable index of neural activity, for example, using the ratio of the different band powers (Brookhuis and De Waard, 1993; Pope el al., 1995; Prinzel et al., 2000). Pope et al. (1995), who reported the first brain-based adaptive system, established a system to index the task engagement based upon ratios of EEG power bands (theta, alpha, beta, etc.). While

these changes are reproducible across subjects, and stable over the time, their estimations are relatively slow (more than five minutes in order to highlight differences between different mental workload levels).

Another approach towards real-time assessment of mental workload, instead of the EEG spectral components, is to use Brain-Computer-Interaction (BCI) technology, e.g. linear discriminate analysis (LDA), support vector machine (SVM), artificial neural network (ANN), etc. These studies classified workload into several levels (e.g. low, moderate and high) using the various EEG parameters in either a simple, single-task (Wilson and Fisher, 1995; Gevins et al., 1998; Nikolaev et al., 1998; Gevins and Smith, 1997) or complex tasks with skilled operators (Noel, et al., 2005; Russell and Wilson, 1998; Wilson and Russell, 2003; Grimes et al., 2008; Heger et al., 2010; Putze et al., 2010; Aricò et al., 2013). The use of the machine learning techniques allows to assess the subject's mental workload in a short time (few seconds) reaching a high accuracy (>90%) (Aricò et al., 2013, Kohlmogoren et al., 2007).

2.4.2.6 *Electrocardiography*

Heart Rate (HR): Since the hearth rate (HR) measure is easy to obtain and less sensitive to artefacts (Kramer, 1990), it is one of the most popular physiological parameters for mental workload assessment within various environments (Backs and Seljos, 1994; Wilson, 2002; Brookhuis and De Waard, 1993; Mehler et al., 2009). It is assumed that an increased mental workload leads to an increased cardiovascular activity, a heightened cortical energy transformation, and corresponding enhanced metabolic demands (Backs and Seljos, 1994). Although this generalization is widely accepted, not all studies agree with the findings. Some articles are critical of the use of

heart rate to measure workload because of the various psychological, environmental, and emotional factors that can influence the response (Jorna, 1993; Lee and Park, 1990; Roscoe, 1992). For example, "feelings of uncertainty and anxiety can significantly raise heart rate" (Jorna, 1993). Other research has determined that HR "does not appear to be of value as a sole measure of pilot workload but it can be strongly recommended as a technique to augment a good subjective rating scale" (Roscoe, 1992). HR is also sensitive to mental effort. Numerous studies have found correlations between cognitive demands and HR (Roscoe, 1992; Veltman and Gaillard, 1996, 1998; Caldwell et al., 1994). HR is sensitive to variations in task demand, but is also influenced by the contamination from physical effort, emotions and stress (Kramer, 1990). In a study on multitasking performance, Fairclough et al. (2005) explored the interaction between learning and task demand on psychophysiological reactivity. These authors used EEG activity, cardiac activity and respiration rate 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. In another study, Wilson (2002) evaluated in a flight experimental scenario, ten pilots who were required to fly a 90-minute to test the reliability of psychophysiological measures of workload. Each pilot performed the same scenario twice to assess the test-retest reliability of the measures. Cardiac, electrodermal and electrical brain activity measures were highly correlated and exhibited changes in response to the demands of the flights. Wilson found that HR was more sensitive to the workload level than heart rate variability. Therefore, the majority of previous

researches consistently demonstrate that the increased workload leads to increased HR.

Heart Rate Variability (HRV): Several studies investigated the relationship between HRV and mental workload. It was demonstrated that HRV is sensitively decreased with increased mental demands in a binary choice task (Backs and Seljos, 1994; Lee and Park, 1990; Mehler et al., 2009). For instance, Lee and Park showed that both increased physical load and mental load could lead to decreased HRV. Brookhuis and De Waard (2001) stated that HRV shows sensitivity to computational effort but not to compensatory effort, while HR has generally been sensitive to both. As reported by Miller (2001), in laboratory studies, HRV has consistently responded to changes from rest to task conditions and to a range of between-task manipulations (Aasman et al., 1987; Sirevaag et al., 1987). In the experimental contexts, especially in flight-related studies, HRV increases as an indicator of the extent of task engagement in information processing requiring significant mental effort (Kramer, 1990; Sirevaag et al., 1987; Wilson and Eggemeier, 1991). HRV has been reported to respond rapidly to changes in operator workload and strategies, usually within seconds (Aasman et al., 1987; Coles and Sirevaag, 1987). Thus, HRV has been able to detect rapid transient shifts in mental workload (Kramer, 1990).

2.7 Brain Computer Interfaces (BCIs)

A 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 (Figure 2.11). For example, in an EEG-based BCI the

messages are encoded in EEG activity. A BCI provides its user with an alternative method for acting on the world (Wolpaw et al., 2002). More recently, Wolpaw and Wolpaw (2012) defined a Brain-Computer Interface as *"a system that measures 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 particular, the electroencephalography (EEG) is the most commonly used technique to realize a BCI system, because the high temporal resolution and the portability, compared to other neuroimaging techniques (fMRI, MEG, etc.). A first categorization of the BCI systems can be made according to the invasiveness of the system. The invasive BCI systems are based on the use of electrodes implanted in the cerebral cortex of the user, which allow to obtain a high signal (control feature) to noise (basic EEG) ratio. On the contrary, the non-invasive systems use the surface electrodes.

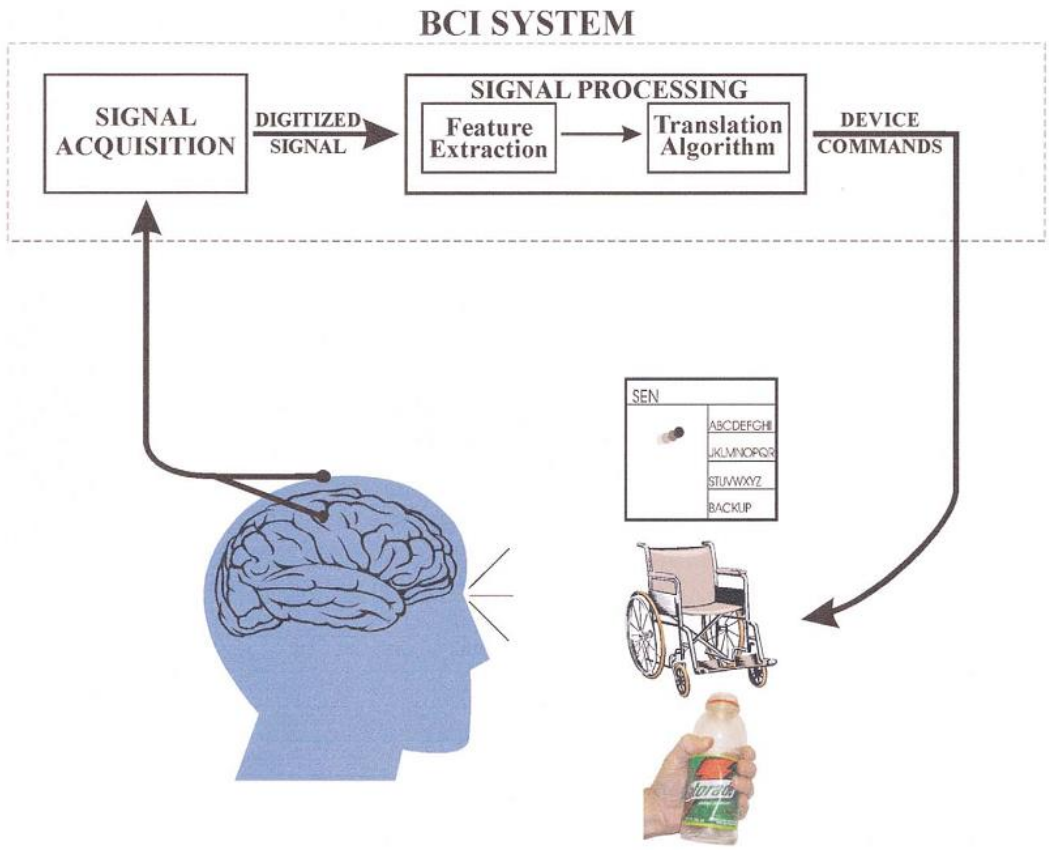


Figure 2.11: Basic design and operation of any BCI system. Signals from the brain are acquired by electrodes on the scalp or in the head and processed to extract specific signal features (e.g. amplitudes of evoked potentials or sensorimotor cortex rhythms, firing rates of cortical neurons) that reflect the user's intent. These features are translated into commands that operate a device (e.g. a simple word processing program, a wheelchair, or a neuroprosthesis).

Regarding the non-invasive BCI, a second categorization is referred to the EEG feature used for act the control. Present-day BCIs fall into 4 groups: slow cortical potentials, SSVEP, P300 evoked potentials and EEG rhythms based BCIs.

Slow Cortical Potentials based BCIs: Among the lowest frequency features of the scalp-recorded EEG are slow voltage changes generated in cortex. These potential shifts occur over 0.5–10.0 s and are called slow cortical potentials (SCPs). Negative SCPs are typically associated with movement and other functions involving cortical activation, while positive SCPs are usually associated with reduced cortical activation (Rockstroh et al., 1993; Birbaumer, 1997). In studies over more than 30 years, Birbaumer and his colleagues have shown that people can learn to control SCPs and thereby control movement of an object on a computer screen (Elbert et al., 1980, Birbaumer et al., 1999, 2000). This demonstration is the basis for a BCI referred to as a 'thought translation device' (TTD). The principal emphasis has been on developing clinical application of this BCI system. It has been tested extensively in people with late-stage ALS and has proved able to supply basic communication capability (Kübler et al., 2001).

SSVEP based BCIs: It has long been established that any stimulus in the visual field that flickers at a specific frequency can cause neurons in visual areas to fire at the same frequency. These neural oscillations are called SSVEPs, also known as Steady State Visual Evoked Responses or SSVERs (Regan 1966). This effect is enhanced by attending to the flickering stimulus (Galloway 1990; Müller and Hillyard 2000). This suggests that users can indicate their interest in specific stimuli by choosing to attend or ignore it, thus providing the basis for a BCI (Ding et al. 2006).

P300 ERPs based BCIs: Infrequent or particularly significant auditory, visual, or somatosensory stimuli, when interspersed with frequent or routine stimuli, typically evoke in the EEG over parietal cortex a positive peak at about 300 ms (Walter et al.,

1964; Sutton et al., 1965). Donchin and his colleagues have used this 'P300', or 'oddball' response to act BCI (Farwell and Donchin, 1988; Donchin et al., 2000). The most widespread approach of P300-based BCIs relies on the 'P300 speller' (P300 Speller) paradigm, proposed by Farwell and Donchin (1988). The subject can choose among 36 alphanumeric characters, arranged in a 6 by 6 matrix. The stimulation entails the random intensification of the rows and columns on a computer screen. During this stimulation, the subject is required to focus his attention on the character (target) that he intends to select (for instance, mentally counting the occurrences of the target stimulus). The intensification of the target elicits a P300 potential, which is not detected when other characters (non-targets) are presented (Figure 2.12).

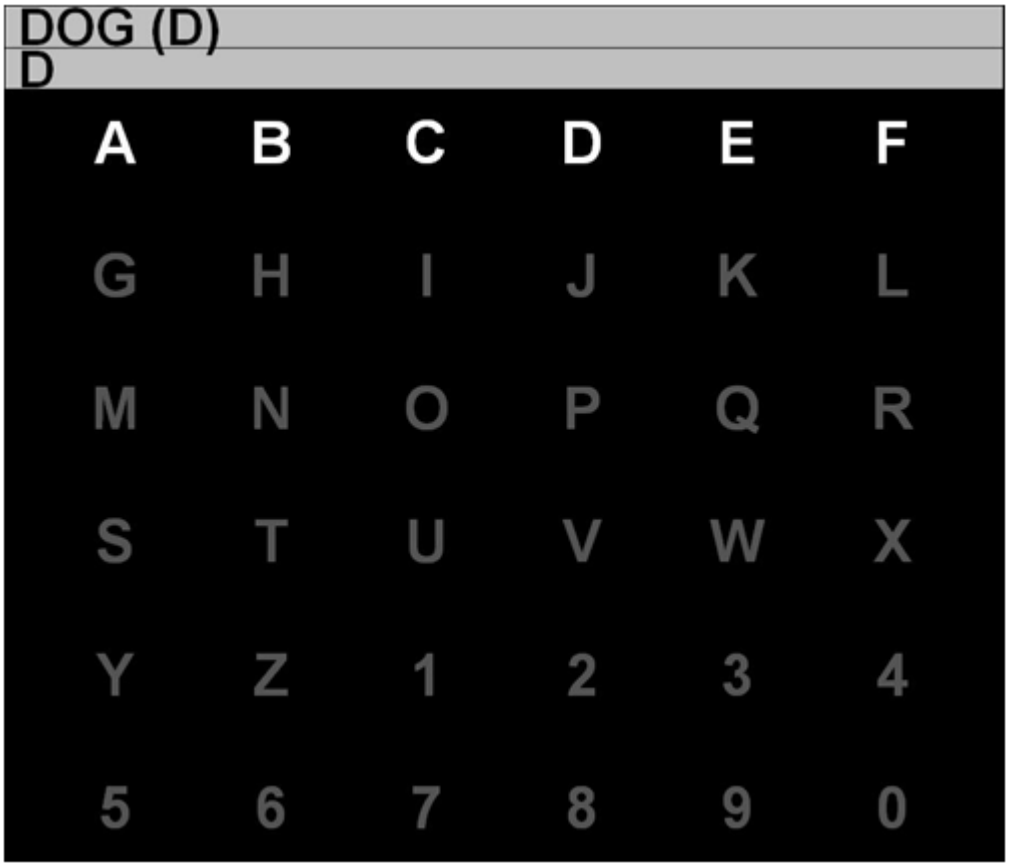


Figure 2.12: P300 Speller matrix interface

EEG rhythms based BCIs: In awake people, primary sensory or motor cortical areas often display 8–12 Hz EEG activity when they are not engaged in processing sensory input or producing motor output (Kozelka and Pedley, 1990). This idling activity, called mu rhythm when focused over somatosensory or motor cortex and visual alpha rhythm when focused over visual cortex, is thought to be produced by thalamocortical circuits (Lopes da Silva, 1991). Several factors suggest that mu rhythms could be good signal features for EEG-based communication. They are associated with those cortical areas most directly connected to the brain's normal motor output channels. Movement or preparation for movement is typically accompanied by a decrease in mu and beta rhythms, particularly contralateral to the movement. This decrease has been labeled 'event-related desynchronization' or ERD (Pfurtscheller and Lopes da Silva, 1999; Pfurtscheller, 1999). Its opposite, rhythm increase, or 'event-related synchronization' (ERS) occurs after movement and with relaxation (Pfurtscheller, 1999). Furthermore, and most relevant for BCI use, ERD and ERS do not require actual movement, they occur also with motor imagery (i.e. imagined movement) (Holmes, 2002; McFarland et al., 2000). This kind of BCI was also widely used for the neurorehabilitation of post-stroke patients (Pichiorri et al., 2011).

At the state of the art, the BCI systems still suffer of a very low bit-rate in comparison with other types of communication systems (e.g., eye-tracker, keyboards, voice recognition). For this reason, potentially they constitute a functional support for people with severe motor disabilities (e.g., stroke, amyotrophic lateral sclerosis). Recently in the BCI community, has emerged the possibility of using the BCI systems in different

contexts from communication and control, developing applications that also involve healthy subjects (Zander et al., 2009; Mueller et al., 2008 Blankertz et al. 2011). In particular, the classic meaning of BCI, in which the user is to modulate voluntarily its brain activity to communicate its intention to the system was changed, but rather, it is the system itself to recognize the spontaneous activity of the user (not modulated by voluntary control) related to the current mental state (e.g., emotional state, workload, attentional levels), and to monitor and use these information to improve the human-machine interaction. In this regard, Zander and colleagues (2011) proposed a categorization of BCI systems, dividing applications based on BCI technology into active, reactive and passive BCI systems.

- Active BCI. An active BCI is one that derives its outputs from brain activity which is directly and consciously controlled by the user, independent of external events, for controlling an application.

- Reactive BCI. A reactive BCI is one that derives its outputs from brain activity arising in reaction to external stimulation, which is indirectly modulated by the user for controlling an application.

- Passive BCI. A passive BCI is one that derives its outputs from arbitrary brain activity arising without the purpose of voluntary control, for enriching a human–machine interaction with implicit information on the actual user state.

2.7.1 Passive Brain Computer Interfaces

A recent direction within the research field of BCI attempts to broaden the general BCI approach by substituting the user's command with passively conveyed implicit

information (Zander et al., 2010; Zander and Kothe, 2011). Such a passive BCI derives its outputs from arbitrary brain activity arising without the purpose of voluntary control, for enriching a human-machine interaction with implicit information on the user state. Systems based on passive BCI can provide information about covert aspects of the user state, i.e. task-induced states which can only be detected with weak reliability using conventional methods such as behavioral measures (Zander and Jatzev, 2012). The signals extracted by these BCI techniques are then employed to exploit this novel information for improved man–machine interaction. This allows to optimize and to enhance human performance, and to achieve potentially novel types of skills. Mental state monitoring is of particular interest in safety-critical applications where human performance is often the least controllable factor. In this regard, there are many examples in which a passive BCI could be useful. For example, BCI technology can reveal valuable information about the user state in safety-critical applications, such as driving (Welke et al., 2009; Borghini et al., 2012), industrial environments or security surveillance. With respect to driving assistance applications, recent studies have explored the use of BCI systems in a driving simulation for assessing driving performance and inattentiveness (Schubert et al., 2008), as well as for robustly detecting emergency brakes before braking onset (Welke et al., 2009). In addition, BCI systems can potentially be used for cognitive monitoring in real time the mental workload of operators (Kohlmorgen et al., 2007, Aricò et al., 2013, [J 5]). In a different context, initial steps have been taken towards assistive technologies that use the current mental state of a user for avoiding accidents in industrial environments (Venthur et al., 2010).

3 OBJECTIVES

The main objective of this thesis work is to design and develop a passive Brain Computer Interface (BCI) system able to estimate the user's mental state through the analysis of neurophysiological signals. Different methodologies to realize a passive BCI usable in operative environments have been defined and validated.

In the following chapters, two main streams regarding the passive BCI systems have been taken into account, depending on the EEG characteristics and on the biosignals used for assess the mental states of the users.

- In the first part (section 4), it was analyzed the variations in the morphology of the event-related potentials (ERPs) detected using two reactive BCI interfaces based on two different attention modalities. Possible correlations between variations in ERPs morphology and the current mental state of the user and/or BCI performance were investigates. The long-term purpose of this study is to use these physiological indexes in a closed loop, in order to automatically adapt the reactive BCI interfaces to the current users' mental states [J 1; J 6; J 7; J 10; C 3; C 7; C 24; C 28].

- In the second part (section 5), a passive BCI system able to estimate online the mental workload of the user, by using the combination of several biosignals (EEG rhythms and the ECG signals) has been proposed. The long-term

purpose of this study is the estimation and the control of the workload level in operative environments and under high pressure and stress conditions [J 5].

4 MORPHOLOGICAL VARIATIONS IN THE ERPS (C)OVERT ATTENTION MODALITIES

4.1 A Covert Attention P300-based Brain-Computer Interface: GeoSpell

4.1.1 Introduction

People who suffer from neurodegenerative diseases, such as amyotrophic lateral sclerosis (ALS), experience a progressive loss of motor abilities. In their advanced stages, these pathologies can even affect the control of eye movement (complete locked-in syndrome). The application of brain-computer interfaces (BCIs) as communication aids for these patients has prompted the recent growing interest in new and more effective paradigms of gaze-independent stimulation.

A BCI is a communication system in which messages and commands that a user wishes to send to the environment are not conveyed through the normal output channels of the central nervous system, such as peripheral nerves and muscles (Wolpaw et al. 2002); instead, the user's intention is detected directly, based on the (electrical) activity of the brain, and translated into messages and actions. One of the most commonly used brain signals that are used to operate non invasive electroencephalogram (EEG)-based BCIs is the P300 event-related potential (ERP,

Donchin and Smith 1970). P300 is a positive deflection of a subject's EEG potentials, occurring 250–400 ms after delivery of a rare or task-relevant stimulus (target), within a train of frequent or non relevant stimuli (nontarget) (Fabiani et al. 1987, Polich and Kok 1995).

The largest hurdle that is impeding the practical application of BCIs in assistive solutions for persons with disabilities is the need to improve this technology from laboratory prototypes to devices that can be used in the user's environment. As opined by Riccio et al. (2011), the need for this translation has necessitated an evaluation of the system's usability, among other metrics.

4.1.1.1 *P300 Speller interface*

The most widespread approach of P300-based BCIs relies on the "P300 speller" (P3Speller) paradigm, proposed by Farwell and Donchin (1988). The subject can choose among 36 alphanumeric characters, arranged in a 6 by 6 matrix. The stimulation entails the random intensification of the rows and columns on a computer screen.

During this stimulation, the subject is required to focus his attention on the character (target) that he intends to select (for instance, mentally counting the occurrences of the target stimulus). The intensification of the target elicits a P300 potential, which is not detected when other characters (nontargets) are intensified. Further, VEPs are relevant features for the P3Speller in the classification process (Krusienski et al. 2008, Sellers et al. 2006). In the P3Speller interface, stimuli have different spatial positions, allowing the subjects to gaze at the target letter and wait for its intensification, keeping

nontarget letters at the periphery of the visual field. Higher VEPs are thus elicited by stimuli versus nontargets.

In studies on selective attention, 2 conditions are defined, wherein the subject can focus his attention on a specific target of the visual field overt and covert attention (de Haan et al. 2008). The former relates to the condition in which the subject turns his gaze toward the target, whereas in the latter condition, he focuses his attention on the target without gazing at it directly.

Recently, Brunner et al. (2010) evaluated the performance of 15 healthy subjects using the P300 speller interface in overt and covert states that distinguished the 'letter' and 'center' conditions, respectively. In the former, the subjects gazed at the intended letter (overt condition), and in the latter, the subjects gazed at a fixation cross in the center of a screen, paying attention to the target item (covert condition). Due to the consistent decrease in accuracy under the covert attention conditions, the authors concluded that the performance of the P300 speller depends on gazing. This conclusion has paramount relevance when P300-based BCIs are proposed as communication aids for completely locked-in people.

This issue was addressed by Treder and Blankertz (2010), who developed the ERP-based Hex-o-Spell, a 2-level speller that comprises 6 discs that are arranged on the vertices of an invisible hexagon, allowing subjects to focus their attention on the stimulation without moving their eyes. The authors compared this new speller with the classical matrix approach, using the interfaces under overt and covert conditions, evaluating their performance and the elicited potential waveforms of 13 healthy subjects. They noted that the Hex-o-Spell increased accuracy compared with the P300

speller under covert attention conditions, but insufficiently high (approximately 60%) to consider the interface an effective communication channel (Kübler and Birbaumer 2008).

In a subsequent study, the same authors improved the interface by introducing 3 alternative gaze-independent spellers, wherein each group of letters was associated with a different color. Using this color code, the recognition accuracy exceeded 90% on average (Treder et al. 2011). However, to be effective, the proposed approach requires the subject to remember the color coding. Although this paradigm is effective for a speller interface in which the number and positions of characters are fixed a priori, it might fail to have sufficient flexibility in other contexts. This approach cannot be extended to paradigms in which the number of items on the interface changes dynamically, such as in Aloise et al. (2009).

Further, Liu et al. (2010) proposed 2 gaze-independent brain-computer speller approaches, using the covert visual search task. With their system, subjects achieved an accuracy that was comparable with the classical Farwell and Donchin speller (95% on average). However, these results were obtained using a stimulus onset asynchrony (SOA) of 400 ms, which is significantly longer than the conventional time that is used for other P300-based BCIs (eg, 160–250 ms in Allison and Pineda 2006 and Treder et al. 2010) negatively affecting the written symbol rate (WSR, Furdea et al. 2009, Liu et al. 2010).

The usability of BCIs has seldom been evaluated. Two studies (Riccio et al. 2012 and Zickler et al. 2011) compared the performance of the 2 P300-based BCI systems with regard to:

i) effectiveness, defined as the accuracy and wholeness with which users accomplished the tasks;

ii) efficiency, the measure of the amount of human, economic, and temporal resources that are expended in attaining the required level of product effectiveness; and

iii) satisfaction, a measure of the immediate and the long-term comfort and acceptability of the overall system.

The efficiency was tested in terms of accuracy and WSR; the efficiency was assessed in terms of workload (Hart and Staveland 1988), using the NASA Task Load Index (TLX) workload assessment; and overall device satisfaction was scored on a visual analogue scale (VAS), ranging from 0 (not at all satisfied) to 10 (absolutely satisfied).

The purpose of this study was to introduce and evaluate a novel P300-based speller interface, GeoSpell (Geometric Speller), which was designed for operation under covert attention conditions, even in protocols that contemplate a dynamically variable number of stimulus classes. We compared GeoSpell with the classical P300 speller (P3Speller) under overt attention conditions in terms of performance (accuracy and WSR) and usability (ISO 9241-210).

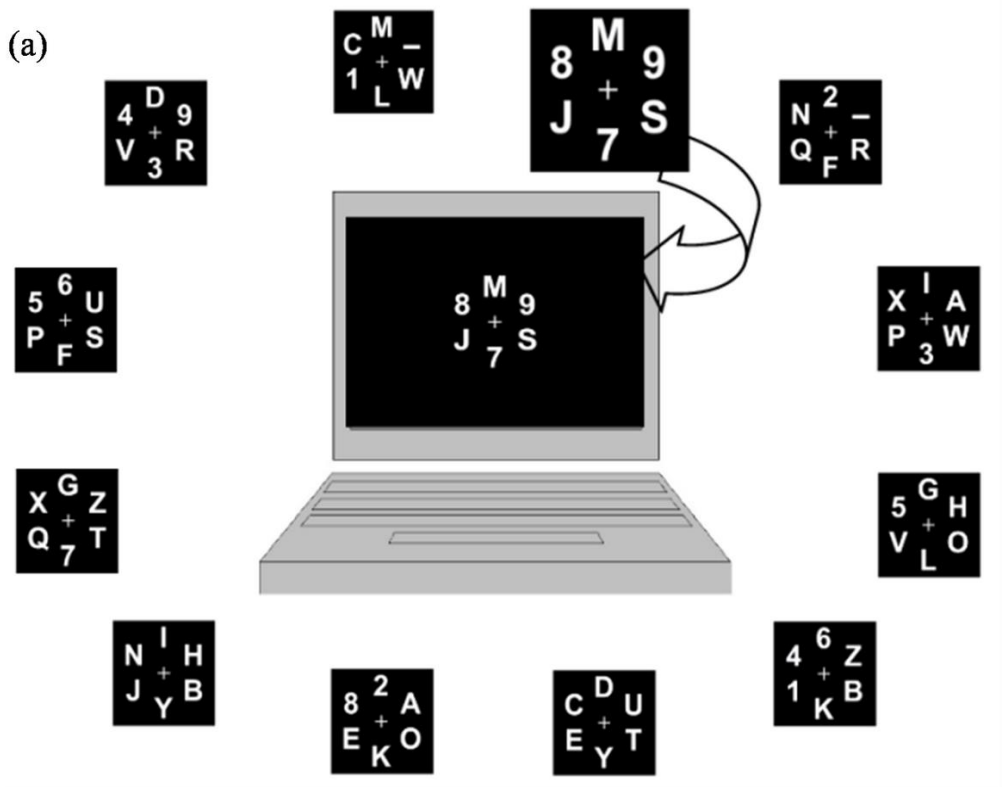

(a)
(b)
M _ W L 1 C
9 R 3 V 4 D
S F P 5 6 U
7 Q X G Z T
J N I H B Y
8 2 A O K E

Figure 4.1: (a) The proposed GeoSpell (Geometric Speller) BCI. Each group contains 6 alphanumeric items that are presented in a random sequence in the center of a screen. (b) Group organization. Each group contains the characters of one row or one column of a matrix; thus, the new interface maintains a similar approach as the row-by-column P300 Speller for the stimulation, but it can be used under covert attention conditions.

4.1.2 Methods and Materials

4.2.1.1 *GeoSpell interface*

In the GeoSpell interface, characters are organized per the same logic as an N by N matrix of a P3Speller: a total of N^2 characters are grouped into 2N sets of N characters (analogous to rows and columns of a P3Speller). In this arrangement, each character belongs to exactly 2 sets. In the visual layout of each set, characters are displayed at the vertices of a regular geometric figure.

During the presentation, each set of characters is displayed transiently on the screen. Notably, each character is displayed at the same position for each of the 2 sets to which it belongs. All 2N sets are displayed in a pseudorandom sequence that is repeated several times in a trial (Figure 4.1). A fixation point is placed in its center to help the subject avoid eye movements. Classification of the attended character can be performed at the end of each sequence. As in the P3Speller, the selection of a character is given by the intersection of the 2 most likely selected sets.

The angular distance between the fixation cross and each character in the group was fixed; the subject sat 1 meter from a 17" LCD monitor, and the distance between the cross and letters was 2.64 cm, yielding a 0.90° angle. The visual angle that was subtended by the subject's eyes did not exceed 1°, allowing stimuli to fall within the subject's fovea (Sutter 1992). While we assembled the sets of characters, we ensure that the numbers of white pixels in each layout were comparable (mean 3274.33 pixels; SD = 2.93%) to minimize the differences between the visual evoked potentials (VEPs) that were elicited by each set, preventing any influence on the system's accuracy. This approach was conducted in order to avoid an unbalanced contribution of the VEP elicited by target and nontarget stimuli.

4.2.2.1 *Experimental protocol*

Ten healthy subjects (6 males, 4 females, mean age = 26.82, SD = 4.21) with previous experience with P300-based BCIs were recruited. Scalp EEG potentials were measured using 16 Ag/AgCl electrodes that covered the left, right, and central scalp (Fz, FCz, Cz, CPz, Pz, Oz, F3, F4, C3, C4, CP3, CP4, P3, P4, PO7, PO8) per the 10-10 standard (Jurcak et al. 2007), arranged on an elastic cap (Electro-Cap International, Inc.). Each electrode was referenced to the linked earlobes and grounded to the right mastoid. The EEG was acquired using a g.USBamp amplifier (g.Tec, Austria), digitized at 256 Hz, high pass- and low pass-filtered with cutoff frequencies of 0.1 Hz and 20 Hz, respectively. The electrode impedance did not exceed 10 kΩ. Visual stimulation, acquisition, and online classification were performed with BCI2000 (Schalk et al. 2004) using a stimulus presentation that was modified for this study.

During the recording sessions, eye movements were monitored on an eye tracker system with 0.5° spatial resolution. The system comprised an infrared light camera (iSlim 320, Genius corp., Taipei, Taiwan) that was managed by the open-source program ITU GazeTracker (San Agustin et al. 2010). Eye gaze coordinates (in pixels) were sent via UDP protocol to the BCI2000 program, which stored them, keeping the temporal correspondence with the EEG data and stimulation markers. This step allowed us to quantify ocular movements and eye blinks and correlate them with the stimuli during an offline analysis. The eye tracker system was mounted on a chinrest, on which the subject placed his head during the recording session to avoid head movements.

The experimental protocol consisted of 5 recording sessions, during which we compared the P300 speller and GeoSpell interfaces with regard to reaction times, lost targets, ERP components, and usability (effectiveness, efficiency and satisfaction).

Before describing the experimental protocol in detail, we introduce terms that will be used below.

- Stimulation sequence: a series of presentations of all 2N stimuli (character sets), each stimulus being presented once;
- Trial: a series of contiguous stimulation sequences during which the target is unchanged;
- Run: a series of trials that entail the continuous acquisition of data;
- Session: a series of one or more runs, acquired without removing the electrode cap (different sessions typically take place on different days).In the GeoSpell

Training sessions

As discussed, each participant in the protocol had previous experience with the P3Speller interface. To avoid any bias due to greater familiarity with the classical interface, each subject participated in 4-6 training sessions to become accustomed to the GeoSpell interface, before the actual data were collected. Each training session consisted of 9 runs of 6 trials. The system prompted the target character at the beginning of each trial. Eight stimulation sequences were presented per trial; thus, each item was presented 16 times. No EEG data were acquired during these sessions, but subjects were instructed to attend the stimulation and push a button when they recognized the target character. New training sessions were scheduled for each subject until the number of missed targets stabilized.

Offline sessions

The offline sessions were categorized as Reaction Time and Copy Mode.

- Reaction Time sessions (RT): These sessions were used to evaluate the response times, and no EEG data were acquired; the subject was required to attend the simulation and push a button each time a target stimulus appeared. The data acquired in these sessions were used to compare the subjects' reaction times and relative missed targets using the P300 Speller in overt attention and the GeoSpell in covert attention.

- Copy Mode sessions (CM): During these 2 sessions, subjects were required to pay attention to target stimuli. The EEG signal was acquired, but no feedback regarding the classification results was provided to the subjects. The data from

these sessions were used to evaluate the subjects' ERP components (amplitude and latency of the P100, N100, P200, N200, and P300 potentials) and the offline accuracy of the classifier.

Offline sessions were composed of 6 runs of 6 trials. The targets of each run formed random 6-character words ("AX6L1O", "TVM3CH", "2EWY_8", "BJZN7G", "DR5K9Q", and "FU4SPI"). The characters in a word were chosen to encompass all possible positions through the sets of characters. At the end of a session, each character of the interface was prompted as the target exactly once. The Reaction Time and Copy Mode sessions were alternated (RT-CM-RT-CM); each pair of RT and CM sessions shared the same target word.

During each session, the subject performed 3 runs with each of the stimulation interfaces. At the beginning of each trial, before the stimulation began, the system prompted the subject with the character that he had to attend. The target prompt appeared during a 2-second pretrial interval. The target appeared in the same position as in the following stimulation to allow the subject to focus his spatial attention before the trial started.

A trial consisted of 8 stimulation sequences and, thus, 16 intensifications of the target character. Each stimulus was intensified for 125 ms, with an inter stimulus interval (ISI) of 125 ms, yielding a 250 ms lag between the appearance of 2 stimuli (SOA). To avoid the attentional blink phenomenon, which occurs when the target-to-target interval (TTI) is shorter than 500ms (Raymond et al. 1992), pseudorandom stimulation sequences were assembled so that each target intensification would not occur within

500 ms after the previous one. The same parameters were set for the GeoSpell and P3Speller.

In the offline analysis, the EEG signal was segmented into overlapping epochs that lasted 800 ms, starting at the onset of each stimulus. The classification was performed by 3-fold cross validation, exploring all permutations of the training (2 runs) and testing (1 run) datasets for each interface. A series of 8 classification scores were obtained per cross validation fold, including only data that belonged to the first i sequences of each trial in the datasets, thus simulating various trial durations. Differences in the amplitudes of ERPs that were elicited by the stimulus types (target vs nontarget) were quantified using the coefficient of determination (R^2). R^2 values range from 0 to 1, wherein higher values correspond to larger explained variances (and thus separability of classes). A signed R^2 index was derived by multiplying R^2 by the sign of the slope of the corresponding linear model, which was positive when the amplitudes of the ERPs that were elicited by the target stimuli were higher than by nontargets, and vice versa.

Online session

The last session (online) compared the online performance of the GeoSpell and P3Speller. Data from the previous copy mode sessions of each subject constituted a training set that was to calibrate the classifier in the online session. Stepwise linear discriminant analysis (SWLDA) was used to select the most relevant features and estimate the weights of the linear classifier that was used to discriminate target and nontarget stimuli from the EEG data (Krusienski et al. 2006).

During the pretrial interval (6 s), the subject was prompted with a target character, which appeared in the same position on the screen as in the subsequent stimulation. The target appeared as a static intensification in the P3Speller interface, whereas in GeoSpell (in which only 6 characters can be shown simultaneously), 2 sequences of stimulation (ie, 4 target intensifications) were used. We chose 2 Italian words to be spelled in the run, which required subjects to select characters in different positions, for both interfaces (same as targets in the offline sessions): "ENFASI" ("emphasis") and "NAPOLI" ("Naples"). The following 8 stimulation sequences were used to acquire EEG data for the online classification. At the end of each trial, the classification results were feedback to the subjects.

During the online session, subjects performed 4 runs with each stimulation interface.

4.2.3.1 *Written Symbol Rate (WSR)*

To compare performance of the GeoSpell and P3Speller, we used the written symbol rate index (WSR, symbols/min, Furdea et al. 2009). Compared with the bit rate and information transfer rate index (McFarland and Wolpaw 2003), WSR accounts for corrections in erroneously selected letters; thus, it estimates the number of symbols that a subject spells correctly in a unit of time more accurately.

For the WSR evaluation, the target prediction accuracy was assessed by leave-one-word-out (LOWO) cross validation, considering the entire dataset for the offline and online sessions (Liu et al. 2010). Thus, the target stimuli of one run were tested by the classifier that was trained on the targets that were related to the remaining runs,

exploring all possible combinations, and the LOWO accuracy was computed as the average value at each stimulation sequence.

4.2.4.1 *NASA-Task Load Index (TLX) and Visual Analogue Scale (VAS)*

Workload has a direct effect on the usability of a software interface. If fewer mental resources are requested, the efficiency is higher, and the effectiveness and satisfaction that are associated with the interface also increase. Users' subjective workload for both interfaces was assessed with the NASA-TLX index (Hart and Staveland 1988). NASA-TLX measures the workload by considering 6 factors: mental, physical, and temporal demands; frustration; effort; and performance.

During each session, after runs with a specific interface, subjects were asked to complete the NASA-TLX. Participants were asked to rate subjective workload for each dimension on bipolar scales, scored from 0 to 100. The 6 subscales were then combined into 14 pairs, and for each pair of scales, the subjects were asked to indicate identify the factor that contributed more to their workload. A weighted average technique was used to compute an overall measure of workload (between 0 and 100) and the relative contribution of each subscale.

We evaluated user satisfaction with each interface (GeoSpell and P3Speller). At the end of the GeoSpell- or P3Speller-related runs, subjects were asked to provide a satisfaction score by visual analogue scale (VAS), ranging from 0 (not at all satisfied) to 10 (absolutely satisfied).

At the end of each session, users were asked to express their preference between the interfaces, marking their choice on a continuous line, ranging from -5 to 5. A score of

0 denoted no preference between the 2 interfaces, whereas -5 and 5 corresponded to a strong preference for the P3Speller and Geospell interface, respectively. Two labels that indicated the 2 interfaces were placed at the extremities of the VAS.

4.1.3 Results

4.3.1.1 *Reaction times and Missed targets*

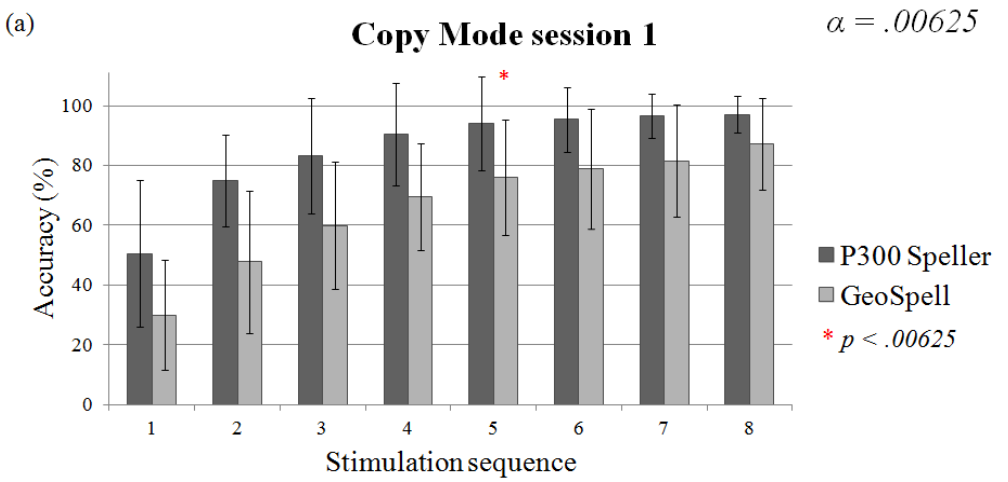
We compared the subjects' reaction times and the relative missed targets that were detected during the 2 reaction time sessions using the P300 Speller and GeoSpell. To analyze the differences between the 2 interfaces, we used two-way repeated measures ANOVA (confidential interval = .95) with interfaces and sessions as factors.

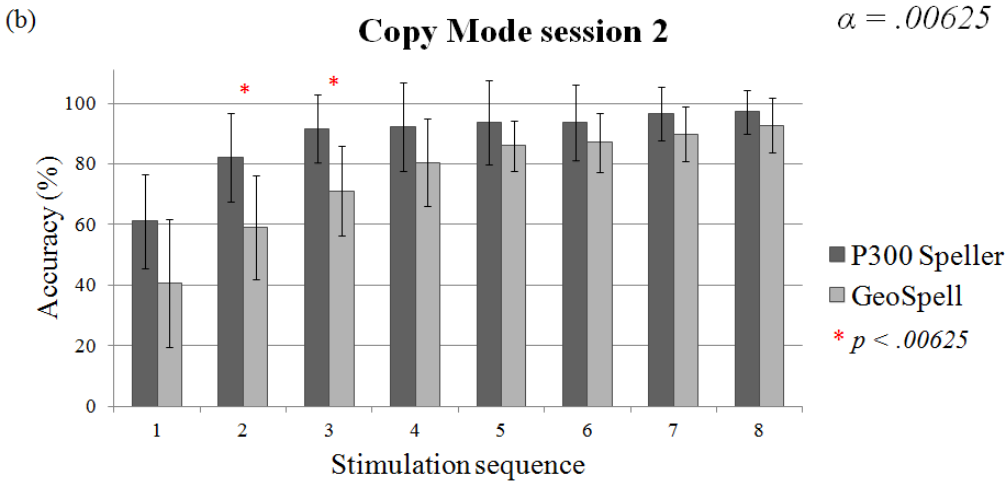
Mean reaction times in the GeoSpell test (Session 1: 454.63 ± 42.80 ms; Session 2: 452.74 ± 30.51 ms) differed (Session 1: $F = 21.848$, $p = .00019$; Session 2: $F = 48.408$, $p = .000002$) from those using the P3Speller (Session 1: 372.14 ± 35.81 ms; Session 2: 361.39 ± 28.17 ms). The number of missed targets increased (Session 1: $F = 4.599$, $p = .004589$; Session 2: $F = 13.702$, $p = .00163$) with the GeoSpell (Session 1: 3.09 ± 3.34%; Session 2: 3.89 ± 2.31%) versus P3Speller (Session 1: 0.76 ± 0.78%; Session 2: 1 ± 0.84%).

4.3.2.1 *Offline BCI accuracy*

Whitney-Mann-Wilcoxon test ($\alpha = .05/nr$, where nr = 8 stimulation sequences - Bonferroni correction) was used to compare the accuracy of the 2 interfaces for each number of the stimulation sequences. We observed statistically significant differences in the fifth stimulation sequence in *copy mode* session 1 ($z = 2.97$, $p = 0.003$) and in

the second (z = 2.77, p = 0.006) and third (z = 2.82, p = 0.005) stimulation sequences in c*opy mode* session 2. Figure 4.2a and 4.2b show the offline accuracy of both interfaces in the 2 *copy mode* sessions. We also noted differences in the fifth stimulation sequence in session 1 and in the second and third sequences in session 2.





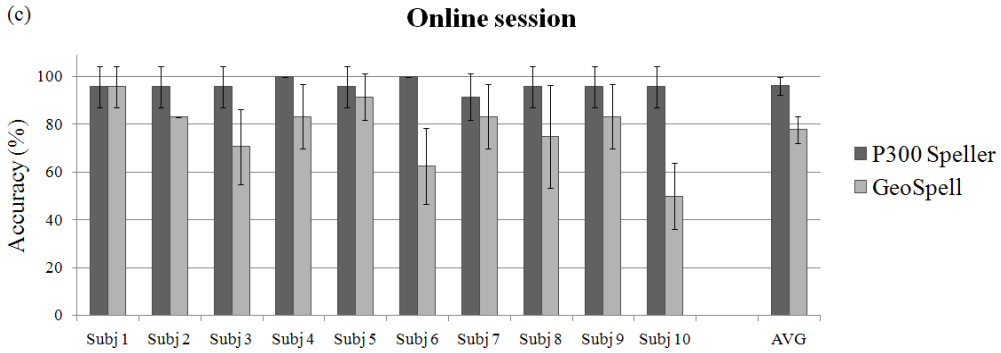


Figure 4.2: (a,b) Mean accuracy and standard deviation of subjects' performance for each interface and for each copy mode session. Asterisks denote that the 2 distributions are statistically different ($p < .05/8$, Bonferroni-corrected). (c) Online classification accuracy for each subject and interface. The subjects' SD values denote the interrun variability, and the SD of the average value is related to the intersubject variability.

4.3.3.1 *Target/Nontarget stimulus-related potentials*

Data from the copy mode sessions were used to assess differences in amplitudes and latencies of the potentials between the 2 interfaces. In addition to the P300 and N200 components, we considered the contributions from P100, N100, and P200. The grand averages of the waveforms on the best channel set for discriminating target versus nontarget evoked ERPs for all subjects (Fz, Cz, Pz, Oz, P3, P4, PO7, PO8, Krusienski et al. 2006) are shown in Figure 4.3.

For the P300 and N200 components, the peak amplitudes and latencies (related to the target stimuli) were determined for each subject by selecting the largest positive or negative peak on channels Fz, Cz and Pz (Krusienski et al. 2008). The intervals were

selected using the grand average of the EEG signals of all subjects for the 2 interfaces. Three-way repeated measures ANOVA (confidential interval = .95) was used to analyze the differences, with interfaces, channels and amplitude/latency as factors.

Overall, there were significant differences in N200 and P300 amplitude and latency for between interfaces (Amplitude: [Fz-N200] Interface, F(1.16) =.00676 p = .936; [Fz-P300] Interface, F(1.16) = 2.7346 p = .118 ; [Cz-N200] Interface, F(1.16) =.196 p = .664; [Cz-P300] Interface, F(1.16) = 4.757 p = .044; [Pz-N200] Interface, F(1.16) =.002 p = .966; [Pz-P300] Interface, F(1.16) =5.77 p = .0287); (Latency: [Fz-N200] Interface, F(1.16) = 25.698 p = .00011 ; [Fz-P300] Interface, F(1.16) = 33.18 p = .00003; [Cz-N200] Interface, F(1.16) = 25.853 p = .00011 ; [Cz-P300] Interface, F(1.16) = 41.773 p = .00001; [Pz-N200] Interface, F(1.16) = 33.279 p = .00003 ; [Pz-P300] Interface, F(1.16) = 45.131 p = .00001) (see Figure 4.3).

The respective contributions to classification of the VEPs using the 2 interfaces were determined by two way repeated measures ANOVA (confidential interval = .95) using interface (GeoSpell and P3Speller) and channels (Oz, P3, P4, PO7 and PO8) as factors and the signed-R^2 of the 2 distributions (target and nontarget) of potentials in the first 200 ms of the epoch as dependent variable.

The contribution of the VEPs to the classification stage using the GeoSpell interface was significantly lower ($p < .05$) compared with that of the P3Speller one ([Oz-VEP] Interface, F(1.104) = 44.254, $p \sim 0$; ([P3-VEP] Interface, F(1.104) = 89.922, $p \sim 0$; ([P4-VEP] Interface, F(1.104) = 74.109, $p \sim 0$; ([PO7-VEP] Interface, F(1.104) = 25.083, $p \sim 0$; ([PO8-VEP] Interface, F(1.104) = 30.102, $p \sim 0$).

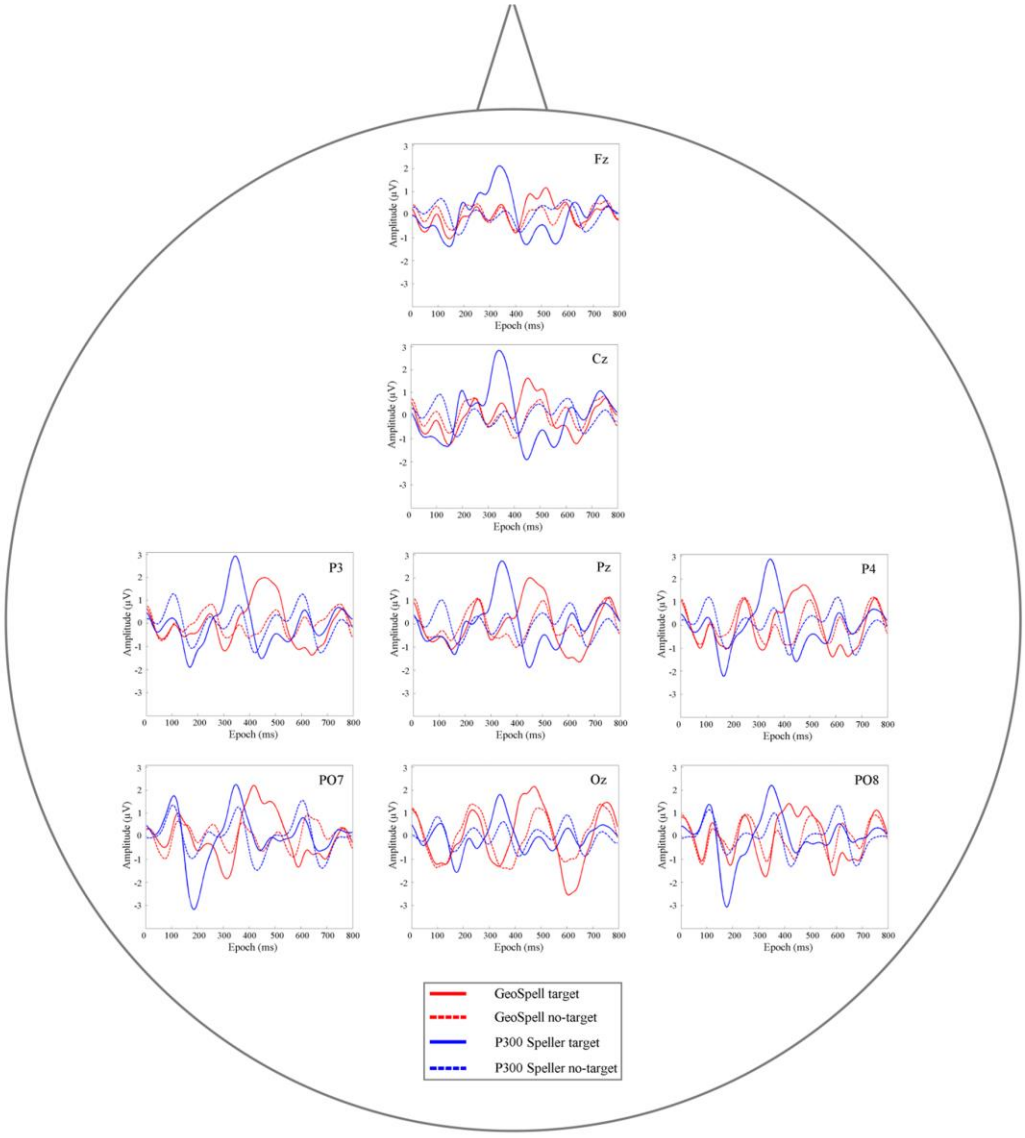


Figure 4.3: Grand average of all subjects' waveforms on channels Fz, Cz, Pz, Oz, P3, P4, PO7, and PO8. The EEG signal was reorganized in overlapping epochs lasting 800 ms and following the onset of each stimulus; figure shows the potentials related to the target and nontarget stimuli for each interface in the 2 Copy Mode sessions.

Two-sample t-test ($\alpha = .05/nr$, where nr = 3 intervals x 5 electrodes = 15, Bonferroni corrected) of the R^2 values that were evaluated with regard to the difference between the 2 target and nontarget classes, in terms of amplitude of elicited waveforms for each

subject, was used to analyze differences between the contribution of the N100, P100, and P200 components using GeoSpell and P3Speller. We considered the occipital and the parietal-occipital sites as test channels (ie, Oz, P3, P4, PO7 and PO8) (Krusienski et al. 2008) (Table 4.1)

Table 4.1: t and p values (in red: $p < \alpha$) of the two-sample t-test ($\alpha = .05/15$, Bonferroni-corrected) for each subject and channel (Oz, P3, P4, PO7, PO8), performed on the signed R^2 values of the amplitudes of the elicited waveforms, using the stimulus type (target vs. nontarget) as the independent variable, between the two interfaces (GeoSpell and P3Speller). Analysis was performed on data acquired in the 2 copy mode sessions.

t-test ($\alpha = .05/15$)	*Oz*		*P3*		*P4*		*PO7*		*PO8*	
	p	***t***	***p***	***t***	***p***	***t***	***p***	***t***	***p***	***t***
Subj 1	2.1×10^{-5}	-4.48	2.1×10^{-5}	-4.48	0.13	-1.54	9.5×10^{-4}	-3.45	6.0×10^{-4}	-3.56
Subj 2	5.7×10^{-6}	-4.88	1.58×10^{-14}	-8.98	6.2×10^{-7}	-5.47	1.27×10^{-9}	-6.74	1.3×10^{-17}	-10.45
Subj 3	4.8×10^{-3}	-2.89	0.0077	-2.73	0.377	-0.89	0.031	-2.20	7.3×10^{-7}	-5.27
Subj 4	1.00×10^{-3}	-3.40	1.61×10^{-11}	-7.58	1.61×10^{-9}	-6.65	5.2×10^{-8}	-5.94	2.8×10^{-8}	-6.02
Subj 5	4.0×10^{-6}	-4.88	0.32	-1	0.0017	-3.22	0.07	-1.82	5.1×10^{-6}	-4.84
Subj 6	3.9×10^{-5}	-4.32	1.21×10^{-14}	-9	1.09×10^{-14}	-9.17	5.2×10^{-5}	-4.22	7.7×10^{-10}	-6.78
Subj 7	1.90×10^{-9}	-6.66	1.93×10^{-8}	-6.13	8.4×10^{-10}	-6.76	4.8×10^{-10}	-6.88	1.6×10^{-6}	-5.12
Subj 8	0.052	-1.96	0.0078	-2.71	1.7×10^{-5}	-4.52	9.3×10^{-14}	-8.59	3.0×10^{-16}	-9.89
Subj 9	4.4×10^{-4}	-3.63	4.2×10^{-5}	-4.28	3.5×10^{-7}	-5.52	0.19	-1.32	0.06	-1.87
Subj 10	1.09×10^{-5}	-4.69	5.7×10^{-9}	-6.39	6.0×10^{-5}	-4.19	2.9×10^{-9}	-6.50	3.1×10^{-7}	-5.56

4.3.4.1 *Online BCI accuracy*

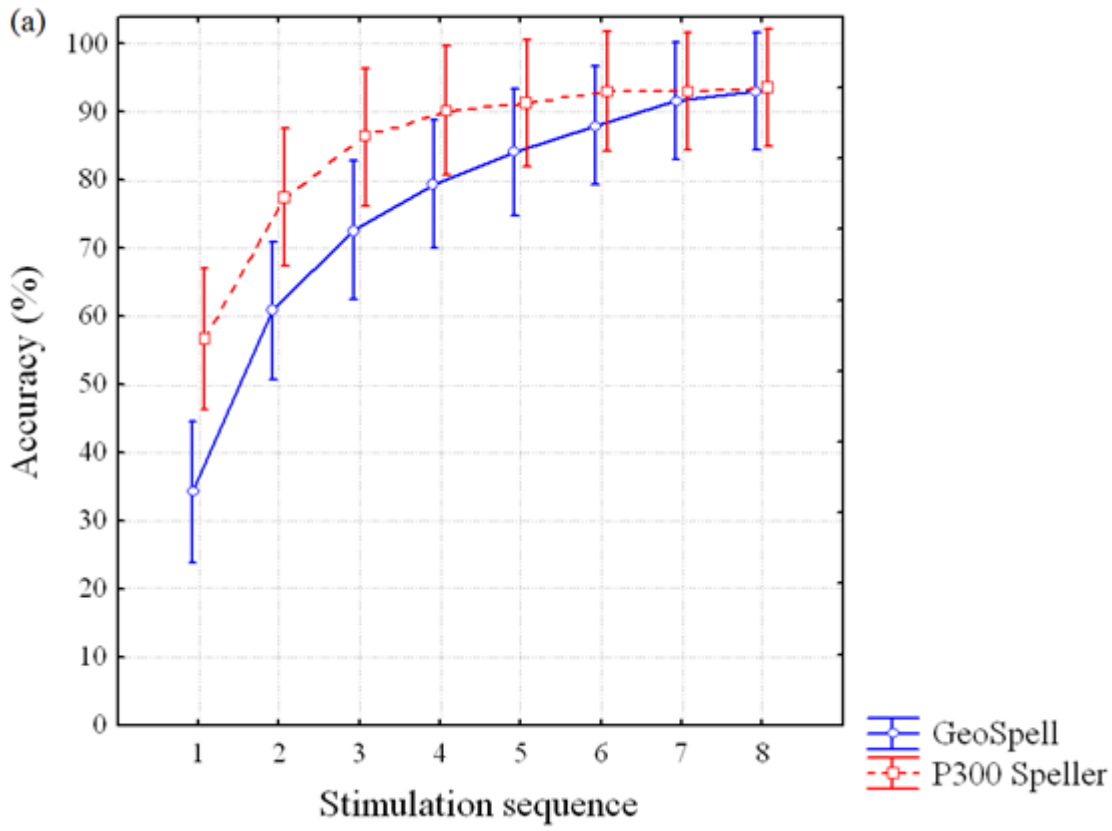
The accuracy in the online session for each subject and the mean accuracy are shown in Figure 4.2c.

By two-way repeated measures ANOVA, with interfaces and runs as factors (CI = .95), we observed a significant difference in accuracy between interfaces ($F = 17.388$, $p = .00058$).

The results of the online session confirmed those of the copy mode sessions: the P300 Speller effected better performance (Mean = 96.17%, SD = 3.68) than GeoSpell (Mean = 77.82%, SD = 5.63); the SD value of the accuracy with GeoSpell showed greater intersubject variability in performance compared with the P300 Speller.

4.3.5.1 *WSR analysis*

Figure 4.4 shows the target classification accuracies (Figure 4.4a) and the corresponding WSRs (Figure 4.4b) for the GeoSpell and P3Speller with regard to the LOWO cross validation (error bars – CI = .95). GeoSpell WSR values were lower overall compared with the P300 Speller, differing significantly ($p < .05$) from the second to sixth stimulation sequence. This result was confirmed by the LOWO target classification accuracies. With GeoSpell, the performance on the first 3 stimulation sequences was significantly lower ($p < .05$) compared with the P300 Speller. WSR peaks of 1.86 symbols/min and 3.76 symbols/min in the seventh and the third stimulation sequences were achieved with the GeoSpell and P300 Speller, respectively. Mean time to select a character was 21 seconds and 9 seconds, respectively. Mean LOWO target classification accuracy was 91.6% and 86.3%, respectively.


(a)
Accuracy (%)
100
90
80
70
60
50
40
30
20
10
0
1
2
3
4
5
6
7
8
Stimulation sequence
GeoSpell
P300 Speller

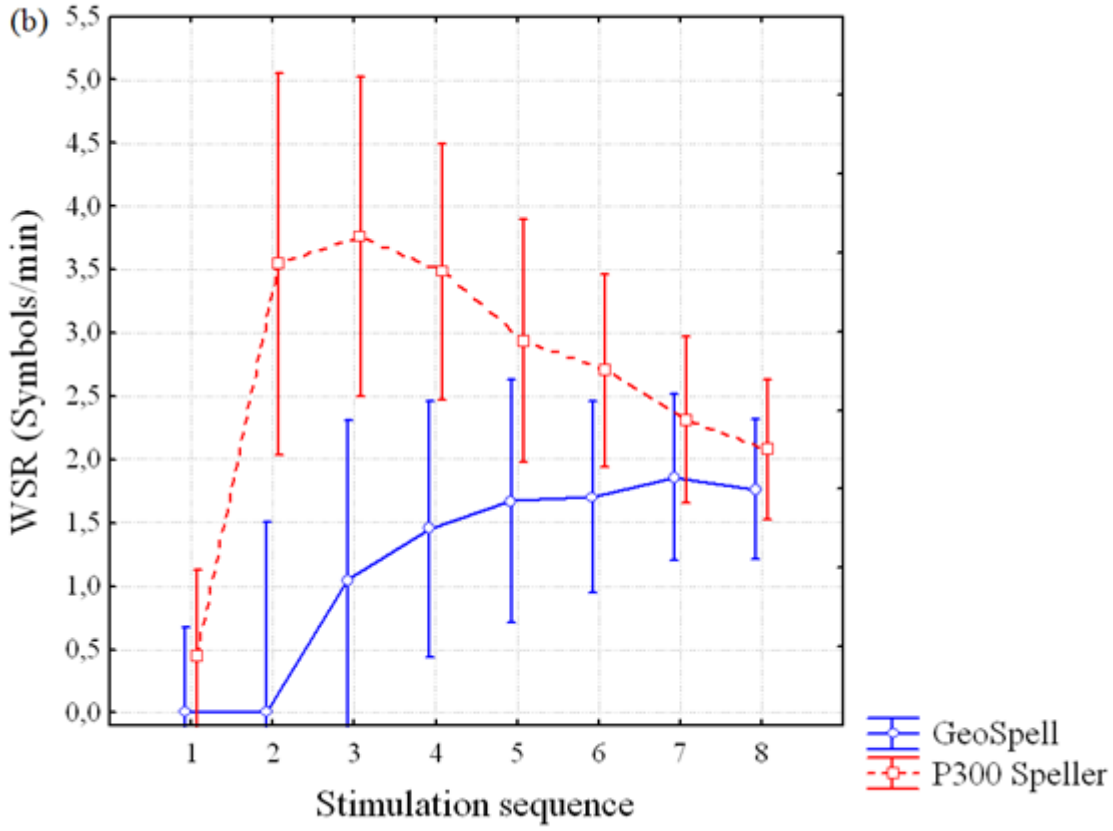


Figure 4.4: (a) Mean and confidence intervals (α = 0.05) of the LOWO target classification accuracies and (b) the corresponding WSRs for the GeoSpell and the P300 Speller interfaces, for each stimulation sequence. Labels on the plots indicate the peak WSR values and the related system accuracies for the GeoSpell and P300 Speller interface. The peak WSR values were, respectively, 1.86 Symbols/min (91.62% of accuracy) for the seventh stimulation sequence and 3.76 Symbols/min (86.35% of accuracy) for the third stimulation sequence.

4.3.6.1 *Workload and overall device satisfaction analysis*

Two analyses by repeated measures ANOVA (CI = .095) were performed separately for the workload scores that were assessed using NASA-TLX and for overall device

satisfaction by VAS scale of the online and copy mode sessions, with GeoSpell interface and P300 speller interface as independent factors.

Although the workload scores of the GeoSpell interface (Copy Mode sessions: Mean=45.584 SD=16.447; Online session: Mean=45.801 SD=19.009) were higher versus the P300 speller interface (Copy Mode sessions: Mean=32.400 SD=21.592; Online session: Mean=30.699 SD=21.066), there was no significant difference between them in the copy mode (p=.142) or online sessions (p=.109).

The mean VAS scores with the P300 Speller were higher compared with GeoSpell for the copy mode and online sessions (Copy Mode sessions: GeoSpell_VAS=7.2±2.05; P300 Speller_VAS=7.94±1.55; Online session: GeoSpell_VAS=7.04±2.17; P300 Speller_VAS=7.71±1.40), but this difference was not significant (Copy Mode sessions: p=.296; Online session: p=.398).

These results were confirmed by the observation that overall, users did not have a preference of interface in the copy mode or online sessions (Copy Mode_Preference=0.041; Online Preference=-0.04).

4.1.4 Discussion

In this study, a novel P300-based BCI text writer that required no eye gaze was developed and validated with regard to effectiveness, efficiency, and satisfaction, comparing the P3Speller interface in the overt attention condition and GeoSpell in the covert attention condition. We decided to compare the new interface with the P3Speller, because the latter is used widely in studies that include end users (Nijboer et al. 2010, Aloise et al. 2011). Our analysis of offline accuracy demonstrated that

despite the lower accuracy with the GeoSpell versus the P3Speller for the first stimulation sequences, the performance of the 2 interfaces was comparable when the number of stimulation sequences increased.

In addition, we demonstrated that the stimulation modality of the GeoSpell, in which the luminance of all stimuli was matched, allowed us to avoid the contributions of the early components of VEPs in the classification process. In contrast, the P3Speller, used under overt attention conditions, relied on these components, which depended on the subject gazing at the target. This result is consistent with Krusienski et al. (2008), who showed that these potentials improve the classification by the P3Speller. Based on our data and previous findings, the term "P300-based interface" is an inaccurate description of this interface (Treder and Blankertz, 2010).

To compare the speed of selection of the GeoSpell with the system that was described by Liu et al. (2010), we performed a WSR analysis, evaluating the given target prediction accuracy by LOWO cross validation using data from both offline sessions. Our interface showed a higher peak WSR and related accuracy (WSR = 1.86 symbols/min; Accuracy = 91.6%) with respect to one of their approaches (WSR = 1.38 symbols/min; Random Position (RP) Accuracy = 87.8%; Fixed Position (FP) Accuracy = 84.1%). In the online session, subjects spelled with an average accuracy of 77.8%, lower than our study's offline accuracy and Liu's online accuracy. These differences might be attributed to the choice of the calibration data, which, in our online session, were obtained from the previous session rather than from data that were acquired on the same day.

Due to the GeoSpell's need for a higher number of stimulation sequences versus the P3Speller, the former had lower WSR values, but the performance of the 2 interfaces was comparable when the number of stimulation sequences increased. The offline and online performance with the GeoSpell interface exceeded 70% the threshold above which an interface is defined as efficient with regard to communication (Kübler and Birbaumer 2008).

We observed a significant increase ($p < .05$) in reaction time and lost targets using the GeoSpell versus the P3Speller. Further, by ERP analysis, we noted lower amplitudes for the P300 component and longer latency values of the N200 and P300 waveforms that were elicited by the GeoSpell stimulation compared with the P3Speller. Allison and Pineda (2003) demonstrated that changes in ERP component latency between groups and conditions reflect changes in the processing of the stimulus a high P300 latency often correlates with task difficulty; in particular, P300 latency is directly proportional to the task difficulty.

Our most significant result regards the workload scores that were assessed by NASA-TLX using the 2 interfaces, which were statistically comparable, demonstrating that although the GeoSpell interface requires a higher level of concentration than the P3Speller, the user's workload is not impacted. This finding is an important aspect, as it relates to the effective usability of the interface with actual end users (Riccio et al. 2011). This result was qualitatively confirmed by the users' preferences, which did not differ significantly between the interfaces.

The above mentioned approach, which highlights the importance of user feedback in the evaluation of the usability of a device, spurs us toward a user-centered approach.

The needs and feedbacks of end users should be taken into account during the development of the system. Considering that potential end users could encounter problems, such as fatigue and fluctuations in attention, the usability of the system should be improved through an asynchronous approach (Aloise et al. 2011). The potential advantages of a new interface should be tested online with potential end users. As discussed by Aloise et al. (2011) with patients, an approach that confers minor advantages to healthy users could have a robust impact on the end user acceptability of the device.

Eye movements toward the target stimuli that were detected during the EEG recording sessions were considered irrelevant for the purposes of this study due to their negligible number (~1% of presented target stimuli). Moreover, based on their timing, they might be interpreted as involuntary and nontarget-related movements. For this reason, we chose not to eliminate trials with eye movements from our analysis. Quantitative assessment of the absence of eye movements confirmed the hypothesis that users are able to operate GeoSpell under covert attention conditions.

People with severe motor disabilities, such as those who are locked in by amyotrophic lateral sclerosis (ALS), use their remaining resources to communicate with the outside world; in general, their control over their eye muscles is maintained, even in the advanced stages of the disease, and until it is compromised, they can use eye tracker systems, which have several advantages over the classic P300-based BCI systems (eg, P3Speller). Eye movements are detected quicker, more easily, and more accurately than ERPs; also, the bit rate of eye tracker systems is higher compared with BCIs eg,

an eye tracker-based text writing system has a spelling rate of 10 words per minute with unimpaired eye movements (Majaranta et al. 2006).

Conversely, a BCI system that is operable during covert attention may be the sole method of communication for ALS subjects who have lost the ability to control their eye movements.

Thus, the GeoSpell approach is a valid solution of restoring communication for such patients; this interface can also be used with impaired eye movement, performing above the 70% threshold and handling a workload that is comparable with that of the classical Speller matrix.

The 2 interfaces have been used under disparate conditions of attention; under covert attention conditions, the P3Speller causes a decrease in performance (Brunner et al. 2010), rendering it unsuitable as a "communicative mean" (Kübler and Birbaumer 2008, Furdea et al. 2009).

4.2 Influence of P300 latency jitter over ERPs based BCIs performance

4.2.1 Introduction

The Farwell and Donchin's P300 Speller (Farwell and Donchin, 1988) is among the most widely validated Brain Computer Interface (BCI) paradigms for communication applications. Brunner and colleagues (Brunner et al., 2010) have recently shown that the P300 Speller recognition accuracy was significantly decreased if the subject was not allowed to gaze at the target stimulus. Several user interfaces designed to be used in covert attention modality, (i.e. in the absence of eyes movements) have been implemented and tested (Fabio Aloise et al., 2012; Liu et al., 2010; Treder and Blankertz, 2010) with the overall result of a lower system performance in covert with respect to overt attention usage. The observed superiority in the system performances under overt usage modality was mainly ascribed to the contribution of visual evoked potential (VEP) components recorded at occipital and parieto-occipital sites (Fabio Aloise et al., 2012; Treder and Blankertz, 2010). In this regard, it has been clearly demonstrated that short latency VEPs represent relevant features for successful control of the P300 Speller interface (Krusienski et al., 2008; Sellers et al., 2006). In fact, in the P300 Speller interface the stimuli are arranged in a way that the users can gaze the target letter and wait for its intensification while the non-target letters are spatially distributed at the periphery of the visual field. Higher amplitudes of these VEP

components are elicited by target as compared with non-target stimuli since only the former stimuli fall in the foveal part of the retina. Reducing the visual crowding, that is similarly as in the covert modality, would greatly affect the VEP component amplitude while leaving the P300 component amplitude almost unaffected due to its independence from whether the target is foveated or not (Brunner et al., 2010). More specifically, in the covert attention-based interfaces there is no spatial difference between target and non-target stimuli, thus there is no difference between VEP amplitude elicited by target and non-target stimuli.

Other factors might be also relevant in influencing the classification accuracy of P300-based BCI paradigms, such as the trial-by-trial stability of latencies of the potentials elicited by the visual stimulation (Thompson et al., 2013). Specifically, the P300 is a positive deflection of the EEG signal elicited in the process of decision-making (Fabiani et al., 1987). The P300 latency and amplitude can be influenced by several internal and external factors such as exercise, fatigue (Yagi et al., 1999), age and gender (Polich and Kok, 1995). Greater latency variations were also observed when the attention is divided between two tasks (Polich, 2007). This phenomenon, known as latency jitter, occurs when the lag between each target stimulus onset and the related potential peak is not constant for the different stimulus repetitions. Kutas and colleagues (Kutas et al., 1977) showed that for a P300 potential elicited by means of an odd-ball paradigm, measures of the peak amplitude performed on the averaged potential are biased because of the inter-trial variability (i.e. the jitter) of the peak latency. In fact, the latency jitter would induce a decrease in P300 amplitude (peak height) and a lengthening of the P300 latency window (peak width) (Fjell et al., 2009). The inter-trial variability was ascribed to the stimulus evaluation time defined as the

amount of time to perceive and categorize the relevant stimulus. A probabilistic method to estimate the P300 latency across trials and to realign the P300 potentials in order to obtain an unbiased peak amplitude was also proposed in (Kutas et al., 1977).

In the context of ERP-based BCI paradigms, each stimulus is presented to the subject several times (e.g. ten times) and a signal average is performed (e.g. by means of the output scores of the classifier) before a classification decision is generated. Thompson and colleagues (Thompson et al., 2013) demonstrated that the accuracy achieved with the P300 Speller was strongly correlated with the jitter in the P300 latency.

In this study we addressed the issue of whether the accuracy of BCIs used in covert attention modality i) is fully explained by the lack of VEP contribution to the classification accuracy and/or ii) is correlated with a lower stability of the P300 potential elicited in the covert attention with respect to the overt attention modality. We hypothesize that i) the jitter would be significantly greater when a specific BCI is utilized relying on covert rather than overt visual attention; ii) a negative correlation would exist between BCI performance and latency jitter in a wide combination of visual interfaces and attention modalities; iii) compensating for the P300 latency jitter through an analysis of single trials would significantly improve the performance of a BCI classifier.

To test our hypotheses, we first evaluated the effect of presenting stimuli through a given visual interface (i.e. the GeoSpell) in either covert or overt modality. Secondly, we evaluated the influence of the P300 latency jitter on the performance of a BCI classifier, in a set of 3 different BCI visual interfaces, and tested whether the expected

differences could be reduced by pre-processing single trials to compensate for the P300 latency jitter.

4.2.2 Materials and Methods

4.2.1.2 *Data collection*

Subjects were requested to complete a spelling task using a BCI. For this purpose, visual stimuli containing 36 alphanumeric characters for the GeoSpell and the P300 Speller interface, and 2 characters for the Visual Oddball interface, were delivered in different arrangements, through three alternative visual interfaces. EEG potentials were acquired for offline analysis. The study protocols were approved by the local Ethics Committee and all subjects gave their informed consent.

4.2.2.2 *Stimulation interfaces*

P300 Speller. In the first interface ((Farwell and Donchin, 1988), Figure 4.5a, cues are organized in a 6 by 6 matrix and each character is always visible on the screen and spatially separated from the others. By design, no fixation cue is provided, as the subject is expected to gaze the target character. Stimulation consists in the intensification of whole lines (rows or columns) of 6 characters.

GeoSpell. In the second interface (Fabio Aloise et al., 2012, Figure 4.5b only six characters at a time are presented at the vertices of a hexagon, at the same angular distance (0.9°) from a central foveation point, marked by a fixation cross. Thus, in its intended operation, stimuli must be attended by the subject using covert attention only. New sets of 6 characters are presented in a sequence, until all 36 have been delivered

after 6 intensifications; sequences are designed so that a given character is only presented at a specific vertex, which the subject had previously learned by practicing.

Oddball. A simple Visual Oddball paradigm interface (Figure 4.5c) was also tested being a conventional paradigm to elicit P300 potentials. Only two characters ('O' and 'X') were successively presented at the same spatial location (corresponding to the foveation point), the former being the target 'rare' stimulus.

For all interfaces, the frequency of target stimuli was 16.7% (i.e. 1/6).

4.2.3.2 *BCI settings*

Scalp EEG signals were recorded (g.USBamp, gTec, Austria) from 8 Ag/AgCl electrodes (Fz, Cz, Pz, Oz, P3, P4, PO7 and PO8, referenced to the right earlobe and grounded to the left mastoid; electrode impedance not exceeding 10 kΩ) according to the 10-10 standard (Jurcak et al., 2007) at 256 samples/second. Visual stimulation and acquisition were operated by means of the BCI2000 software (Schalk et al., 2004). At the beginning of each trial the system suggested to the subject the character to be written before the stimulation started. No feedback regarding the classification results was provided to the subjects.

4.2.4.2 *Experimental task*

Recordings took place in four sessions (on separate days). In the first two sessions, the experimental task was carried out using the GeoSpell interface (see section 4.2.2.2) in the overt attention modality, i.e. the fixation cross was removed and subjects were allowed to gaze the specific spatial location where the target character was designed to

appear during the stimulation sequence, as described in Experiment I. In the third and the fourth sessions, the experiment was carried out using the GeoSpell interface in the covert attention modality. In these two sessions, further measurements were performed, as described in Experiment II.

Each session consisted of 3 runs for each interface and 6 trials (i.e. characters) per run. Subjects were required to spell 6 words (3 words per session) chosen so that the spatial position of the target characters covered as much as possible all the positions on the screen, using either the GeoSpell and the P300 Speller interfaces; subjects were required to spell the sequence "OOOOOO" (all 'rare' stimuli) using the Visual Oddball interface. This latter sequence was repeated for six runs. Each trial consisted of 8 stimulation sequences and corresponded to the selection of a single character displayed on the interface. With the term stimulation sequence we refer to a single intensification of all the available items. In summary, for each subject and interface we collected a total of 576 target stimuli (2 sessions x 3 runs x 6 trials (i.e. characters) x 8 stimulation sequences x 2 target stimuli (e.g. in the P300 Speller 1 row and 1 column) and 2880 non-target stimuli (2 sessions x 3 runs x 6 trials x 8 stimulation sequences x 10 non-target stimuli (e.g. in the P300 Speller 5 rows and 5 columns)). Each character was intensified for 125ms (Stimulus duration), with an Inter Stimulus Interval (ISI) of 125ms, yielding a 250ms Stimulus Onset Asynchrony (SOA).

4.2.5.2 *Experiment I*

In Experiment I, we preliminarily tested the effect of using the GeoSpell interface in either overt or covert attention modalities on the P300 latency jitter. The aim was to describe the effects of the attention modality on latency and jitter of P300 regardless of

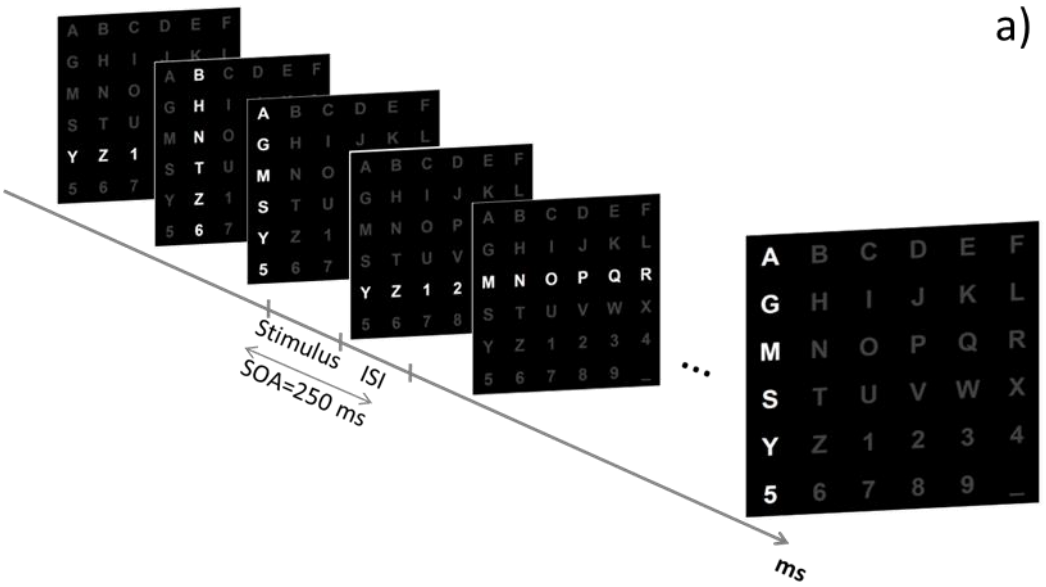
the stimulation interface, and provide the rationale for experiment II. Furthermore, a comparison with the P300 Speller and the Visual Oddball interfaces was performed. Six healthy subjects (3 females and 3 males, mean age 31±5 years) participated in the experiment.

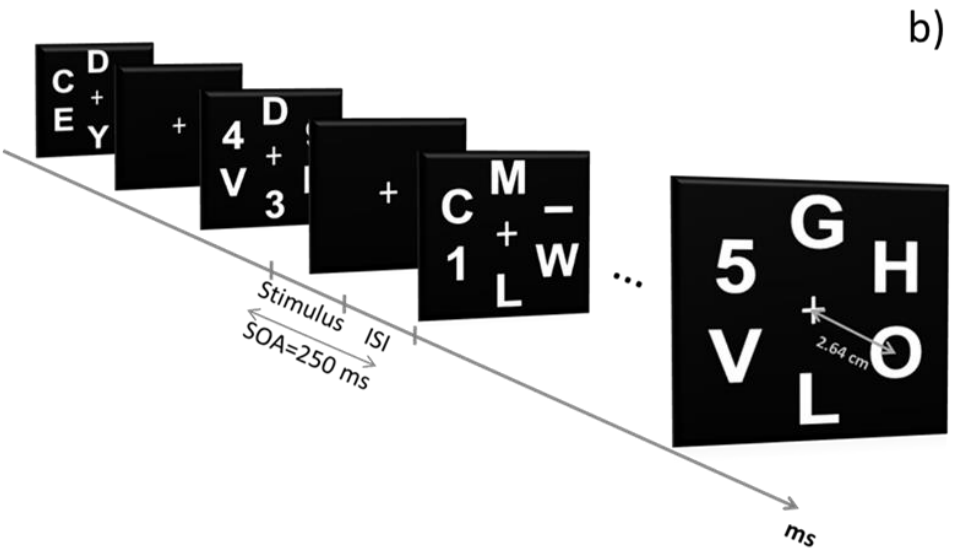
4.2.6.2 *Experiment II*

Experiment II aimed to investigate the influence of the P300 latency jitter on the BCI spelling accuracy when each of the visual interfaces described in Section 4.2.2.2 were used. According to their original design, the P300 Speller and the Oddball interfaces were used in overt attention modality whereas the GeoSpell was tested under the covert attention modality.

Twenty healthy volunteers (14 females and 6 males, mean age 28±5 years) were involved in the study including those who participated in Experiment I. All subjects had normal or corrected to normal vision. Each of them had previous experience with P300-based BCIs and with the interfaces used in this study.

In the following we will refer to the GeoSpell interface used in covert and overt attention modality as Covert GeoSpell and Overt GeoSpell, respectively.


a)
Stimulus
ISI
SOA=250 ms
ms


b)
Stimulus
ISI
SOA=250 ms
2.64 cm
ms

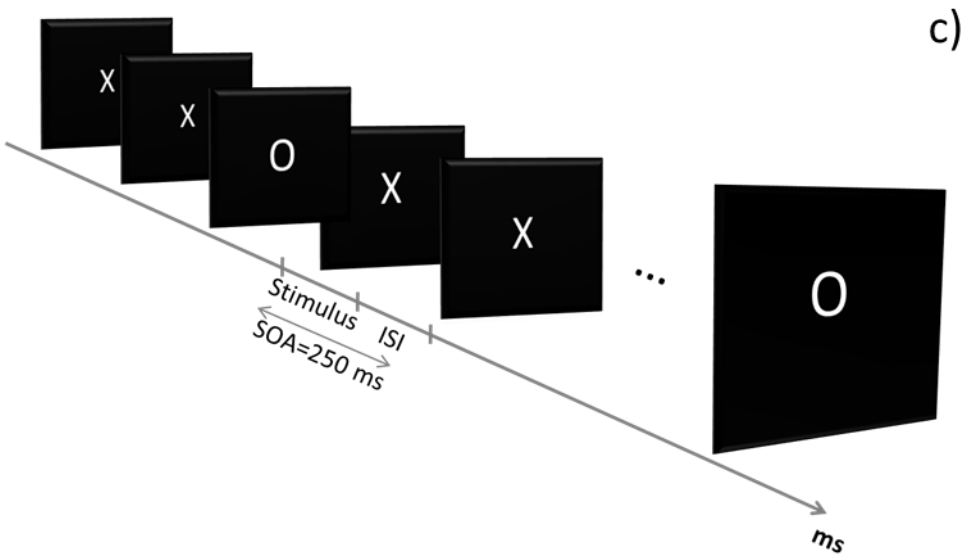


Figure 4.5: The three visual interfaces: a) P300 Speller; b) GeoSpell; c) Visual Oddball

4.2.7.2 *EEG pre-processing*

The EEG signals were segmented into 800 ms overlapping epochs following the onset of each stimulus.

Two runs of each recording session were considered as training set while the remaining run provided the data for the testing set, exploring all possible permutations. This procedure was applied in both the waveform and the performance analyses, which were based on an offline cross-validation (see below).

4.2.8.2 *Waveform analysis*

To evaluate the influence of the P300 latency jitter on the classification accuracy, it was necessary to reconstruct the P300 potential waveform for each single epoch. To this aim, we applied a method described in (Hu et al., 2010), based on the use of a wavelet transform to increase the signal to noise ratio (SNR) of the P300 potentials recorded during the experimental tasks. Figure 4.6 shows a schematic overview of the signal processing procedure applied to estimate the P300 latency jitter. We decomposed each single target epoch into its time-frequency representation by evaluating the continuous wavelet transform (CWT) for each channel, both for the training and testing runs. In the CWT we used a complex Morlet wavelet, with frequency content ranging from 1 to 20 Hz with a frequency resolution of 0.5 Hz and a time window of 800 ms. We computed the power spectrum (PWT) for each transformed single epoch of the training runs, defined as the squared magnitude of the CWT. Finally, we computed the average PWT over all epochs, to identify the wavelet coefficients with the highest power. Coefficients below a specified power threshold were filtered out, according to the following procedure: the empirical cumulative distribution function (CDF) of the power spectrum was calculated through the Kaplan–Meier estimation (Lawless, 1982); the filtering model consisted of a matrix (PMask) whose time–frequency elements were set to 1 when the CDF of the corresponding wavelet coefficient was greater than the threshold, and set to 0 otherwise. We computed the best threshold value referring to the original method used in (Hu et al., 2010), aiming to eliminate as much noise as possible while preserving the shape of the P300 potential. A filtered version of the target single epochs (training and testing sets) was finally obtained by evaluating the inverse CWT (ICWT) of the

coefficient of each single epoch, multiplied for the PMask. When employed in a cross-validation, PMask was estimated from data belonging on the training set.

We estimated the latency of the reconstructed single-trial P300 potential as the latency of the highest peak of the signal falling within a predefined interval (e.g. between 300ms and 600ms). The latter had been manually selected from the averaged waveforms, to embrace the whole P300 shape.

Once the epoch-by-epoch latency of the P300 potential had been estimated, the wavelet-filtered signals were discarded; all amplitude analyses were performed on the original signal (band-pass filtered between 0.1 Hz and 20 Hz, eighth-order Butterworth filter).

For each visual interface, we compared the P300 responses evoked during the different BCI interfaces in terms of amplitude, latency and latency jitter. The P300 peak amplitude was measured both on the original average waveform (Non-Realigned amplitude) and on the waveform obtained by averaging the realigned single epochs (Realigned amplitude), whose time course was shifted according to the estimated P300 latency values. We quantified the jitter of the P300 latency as the difference betweenthe 3rd and the 1st quartile of each distribution for each testing run. We performed the waveform analyses only considering the Cz electrode as representative channel. It should be stressed that the realignment process requires information on the labels of epochs (i.e. target vs. non-target). While it could be a useful analysis method to interpret the timing of single trial ERPs, in its present formulation it cannot be used online.

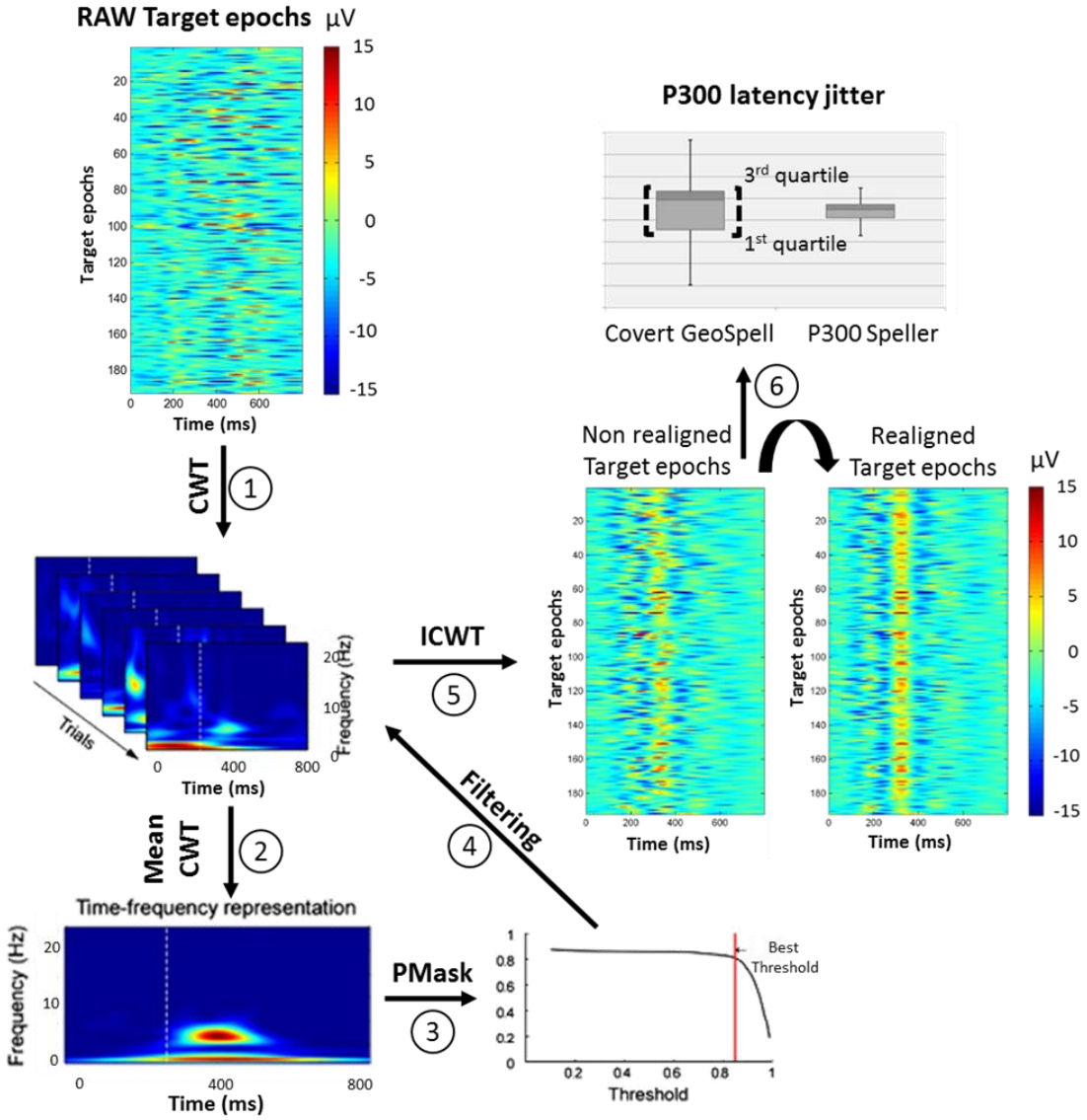


Figure 4.6: Overview of the waveform processing computation. Steps 1: computation of a time-frequency representation of the single target related epoch, based on the continuous wavelet transform (CWT) power spectra; step 2: average of the time-frequency power spectra; step 3: PMask computation starting from the cumulative distribution function (CDF) of the power spectra; step 4: application of the PMask and calculation of the inverse CWT; step 5: evaluation of the P300 latency jitter as the difference between the 3rd and the 1st quartile of the P300 latency

distribution. In the reported example, the Cz electrode for a representative subject was computed.

4.2.9.2 *Performance analysis*

For each participant, we assessed the BCI accuracies offline, as a function of the number of stimulation sequences averaged during each trial. We used a Stepwise Linear Discriminant Analysis (SWLDA, Krusienski et al., 2006) to select the most relevant features that allowed to discriminate between target and non-target stimuli. We performed a three-fold crossvalidations exploring all possible combinations of training (2 runs) and testing (1 run) data set for each session and interface.

We evaluated the performance of the subjects for each interface in the following conditions:

- Whole epoch: the entire time length of the epoch (0-800 ms) is considered. This is the baseline condition against which we compared all others;

- Whole epoch decimated: same epoch length as above, reducing by a factor of 12 the number of time samples (each new sample is the average of 12 original samples). Downsampling is a commonly used procedure to prevent overfitting of the classifier by reducing the number of features (F Aloise et al., 2012; Krusienski et al., 2006), and we considered this condition when referring to state-of-the art performance of a classifier;

- P300 epoch non-realigned: only the epoch segment containing the P300 potential is considered, thus disregarding those VEPs components influenced

by gazing at the target stimuli. The interval extent is subject- and interface-specific;

- P300 epoch realigned: same epoch length as above, using potentials obtained after realignment of the single epochs. In this condition, the effect of latency jitter is compensated.

4.2.10.2 *Correlation between P300 latency jitter and performance*

The information transfer rate (ITR, bit/min) was calculated at each fold of cross-validation as a function of the number of sequences in the trial. The formula described in Pierce (1980) was used to compute the number of bits transmitted per trial. The number of bits transmitted for each stimulation sequence is expressed as:

$$B_i = log_2 N + P_i log_2 P_i + (1 - P_i) log_2 \frac{(1 - P_i)}{(N - 1)} \qquad i = 1 \dots 8 \tag{4.1}$$

where N is the number of possible characters (in our case N = 36), i is the specific stimulation sequence, Pi is the probability that the target is accurately classified at the end of sequence i. From the equation (4.1) the ITR at each stimulation sequence is determined as:

$$ITR_i = \frac{60}{Time_i} B_i \qquad Time_i = SOA \cdot i \cdot M \qquad i = 1 \dots 8 \tag{4.2}$$

Where $Time_i$, represents the time expressed in seconds for the i^{th} stimulation sequence and M is the number of total stimuli (M = 12, e.g. 6 rows and 6 columns for the P300

Speller interface). From equation (4.2) we calculated the mean value of the ITR along the 8 stimulation sequences, in order to have a synthetic measure of the system's performance (4.3):

$$ITR_{Mean} = \frac{\sum_{i=1}^{K} ITR_i}{K} \qquad K = 8 \tag{4.3}$$

To assess the correlation between the ITR_{Mean} and the P300 latency jitter, we estimated the non-parametric Spearman's rank correlation coefficient between these variables. For each subject and for each interface, we considered (i) the latency jitter and (ii) the ITR_{Mean} values calculated at each fold of cross-validation (2 sessions times 3 testing runs).

4.2.3 Results

4.3.1.2 *Experiment I*

Waveform analysis

We performed two one-way repeated measures ANOVA (Confidential Interval = .95) considering the *interfaces* (Overt GeoSpell, Covert GeoSpell, P300 Speller and Visual Oddball) as factors and P300 latency and latency jitter as dependent variables. A significantly influence of *interfaces* factor was found on both the P300 latencies and the P300 latency jitters (P300 Latency: $F(3, 140)=56.18$; $p=1.2x10^{-5}$, P300 Latency jitter: $F(3, 140)=9.3$; $p=10^{-5}$). A post-hoc analysis (Duncan test) revealed that the P300 latency mean values elicited by the Overt GeoSpell (470±16ms), Covert GeoSpell

(476±23ms) and Visual Oddball (451±65ms) interfaces were significantly longer ($p<1.1x10^{-5}$) than the values obtained with the P300 Speller (360±50ms). The same analysis returned a significantly ($p<5x10^{-4}$) larger P300 latency jitter in the Covert GeoSpell (136±33ms) as compared with those observed with the Overt GeoSpell (111±34ms), the P300 Speller (98±18ms) and the Visual Oddball (110±36ms) interfaces. No significant differences ($p>.05$) were found between the Overt GeoSpell, the P300 Speller and the Visual Oddball interfaces in terms of latency jitter.

4.3.2.2 *Experiment II*

Waveform analysis

Figure 4.7 shows, for a representative subject, the average of the waveforms extracted from the testing runs and generated with and without realignment of the single-epoch P300 potentials elicited by the *target* stimuli delivered by each visual interface.

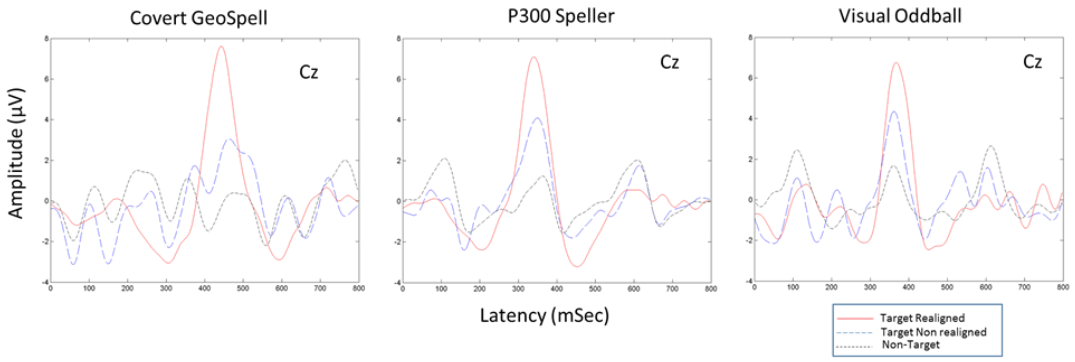


Figure 4.7: Averaged P300 potential waveforms at the Cz electrode position obtained from a representative subject using the 3 visual interfaces.

Significant differences of latency and amplitude of the P300 potential elicited by the 3 interfaces were explored by means of 4 one-way repeated measures ANOVAs

(Confidential Interval = .95) where *interface* was considered as factor and *Realigned P300 amplitude/Not-realigned P300 amplitude/P300 latency/P300 latency jitter* were the dependent variables. Also, a two-way repeated measures ANOVA (Confidential Interval = .95) was performed, where *interface* and *P300 Realignment (P300 Realigned or not)* were considered as factors, and the *P300 amplitude* was the dependent variable.

The analysis revealed a significant difference across the interfaces for latencies ($F(2, 357)=73.56$; $p=1.1x10^{-5}$), the non-realigned amplitudes ($F(2, 357)=6.9$; $p=1.1x10^{-3}$); and latency jitters ($F(2, 357)=52.58$; $p=9x10^{-6}$).

Post-hoc analysis (Duncan test) showed that the P300 Speller elicited P300 waves with lower mean latency than the Covert GeoSpell and the Visual Oddball (353±90 ms, 434±100 ms, and 426±113 ms, respectively; $p<10^{-4}$).

The GeoSpell produced a latency jitter significantly larger than the P300 Speller and the Visual Oddball (mean values: 108±24ms , 76±24ms, and 74±38ms, respectively; $p<10^{-4}$).

The GeoSpell elicited P300 waves with lower amplitudes than the P300 Speller and the Visual Oddball (mean values: 4.7 ±2.0 µV, 6.1 ±3.6 µV, and 5.4 ±3.0 µV, respectively; $p<0.05$).

No significant influence was found on the P300 Realigned amplitude (P300 Speller (9.52 ±3.9µV), GeoSpell (9.58±2.4µV) and Visual Oddball (9.12±3.8µV)) variable ($F(2, 357)=.13$; $p=.88$).

Furthermore, the P300 amplitudes estimated after the realignment exhibited significantly higher values than those evaluated without the realignment, for all the interfaces ($F(1, 714)=380,93$; $p=10^{-5}$).

Figure 4.8 illustrates for an exemplary subject data set, the *target* epochs relative to each interface with and without realignment.

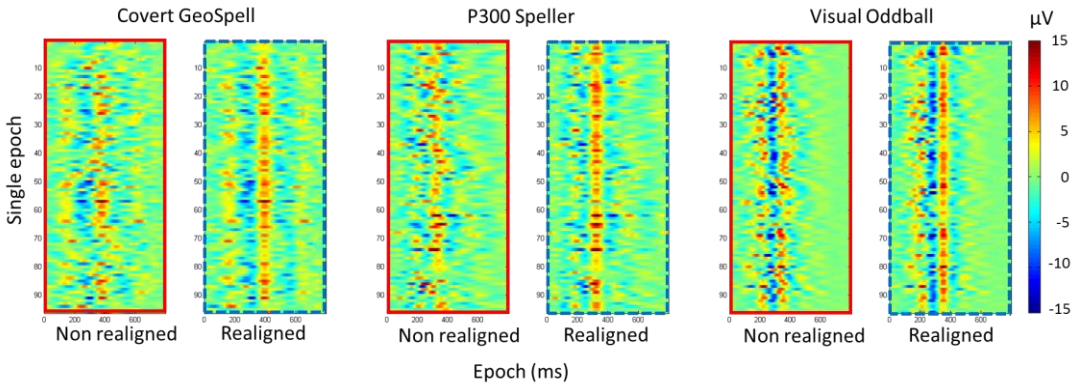
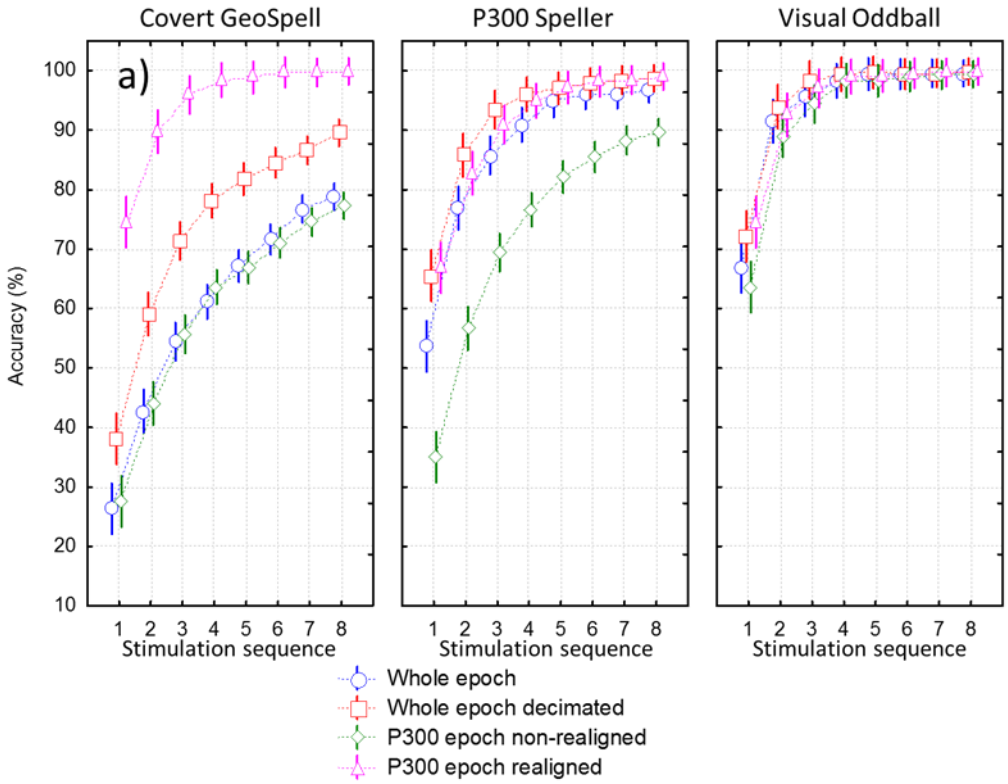


Figure 4.8: Target epochs relative to each interface with (dashed blue boxes) and without (red boxes) realignment. Only one exemplary subject data set over Cz electrode is shown.

Performance accuracy analysis

Differences in the classification accuracy achieved with each of the 3 visual interfaces and each of the 4 conditions introduced in Section 2.2.2 (epochs). Figure 4.9 shows the accuracy for (a) each stimulation sequence and (b) averaged over all the stimulation sequences.


Covert GeoSpell
P300 Speller
Visual Oddball
a)
Accuracy (%)
100
90
80
70
60
50
40
30
20
10
1 2 3 4 5 6 7 8
Stimulation sequence
Whole epoch
Whole epoch decimated
P300 epoch non-realigned
P300 epoch realigned

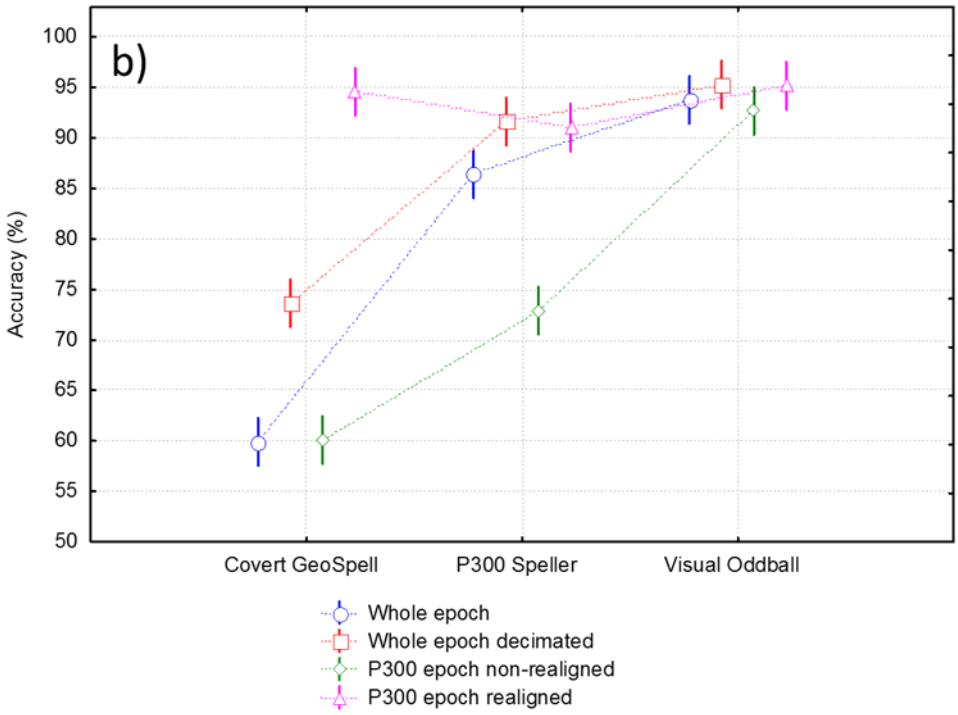

b)
Accuracy (%)
100
95
90
85
80
75
70
65
60
55
50
Covert GeoSpell
P300 Speller
Visual Oddball
Whole epoch
Whole epoch decimated
P300 epoch non-realigned
P300 epoch realigned

Figure 4.9: Mean and confidence intervals (CI = 0.95) of the cross-validation target classification accuracies achieved with the Covert GeoSpell, P300 Speller and Visual Oddball interface, relative to each epoch choice; (a) as a function of the number of stimulations; (b) averaged over all stimulations.

A two-way repeated measures ANOVA (Confidential Interval = .95) was performed with *interfaces* and *conditions* as factors and the *accuracy per stimulation sequences* as dependent variables.

The analysis revealed a significant interaction between the factors ($F(6, 1428)=42.57$; $p=10^{-9}$). The Duncan's multiple range test was used for post hoc comparison. The differences in the epoch choices and the interfaces are summarized in Figure 4.10 and described in detail in the remainder of this section.

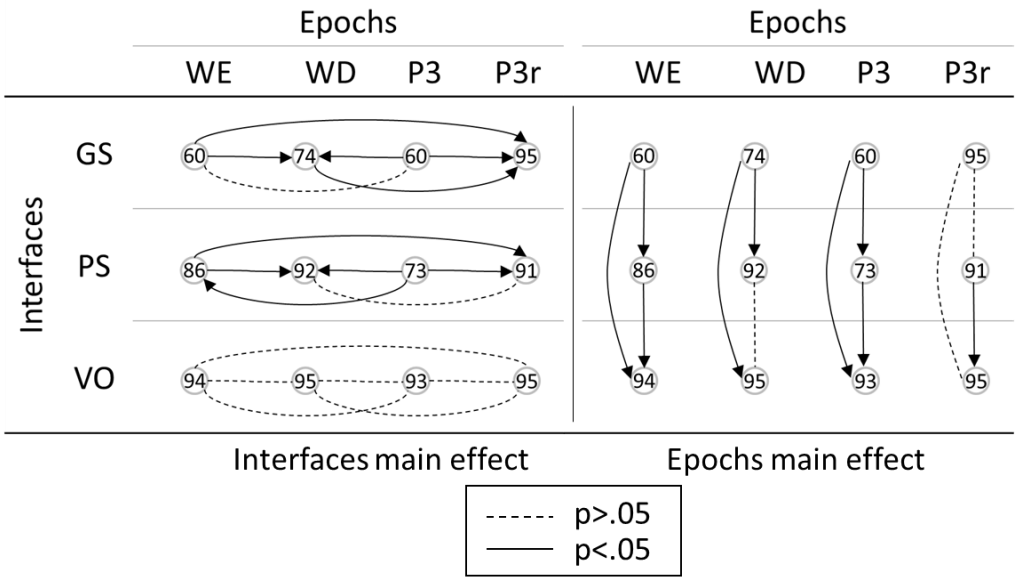


Figure 4.10: Graphical representation of the differences between the epochs (WE: Whole epoch; WD: Whole epoch decimated; P3: P300 epoch non-realigned;

P3r: P300 epoch realigned) and the interfaces (GS: Covert GeoSpell; PS: P300 Speller; VO: Visual Oddball) in terms of accuracy, highlighted by the post hoc test. Each solid line indicates a significant difference ($p<.05$) between the considered epochs or interfaces. Each arrow points to the factor with higher mean value. Dashed lines indicate non significant ($p>.05$) differences between the epochs or interfaces. Numbers in the circles indicate the percent mean accuracy value.

In the *Whole epoch* and *Whole epoch decimated* conditions, the accuracy of the GeoSpell differed significantly from each of the other two interfaces ($p<10^{-5}$). Instead, the Visual Oddball interface exhibited significantly higher accuracy than the P300 Speller only in the *Whole epoch* condition ($p<10^{-3}$).

In the *P300 Epoch Non-realigned* condition, accuracy was higher for the P300 Speller and the Visual Oddball than the GeoSpell interface ($p<10^{-6}$). In addition, the accuracy of the Visual Oddball interface was significantly higher than the P300 Speller ($p<10^{-5}$).

In the *P300 Realigned* condition, only the Visual Oddball interface differed significantly from the P300 Speller ($p<.05$).

Both for the GeoSpell and the P300 Speller interfaces, realignment of the P300 potentials (*P300 Realigned*), yielded a significant increase ($p<10^{-2}$) of the accuracy with respect to the *Whole epoch* condition. Moreover, the decimation of samples (*Whole epoch decimated*) yields a significantly higher accuracy than using the original ($p<10^{-5}$) samples (*Whole epoch* condition) and the *P300 Epoch Non-realigned* condition ($p<10^{-4}$).

Only using the P300 Speller, accuracy in the *Whole epoch* condition is significantly higher ($p<10^{-4}$) than in the *P300 Epoch Non-realigned* ($p<10^{-4}$).

Only for the Covert Geospell interface, accuracy in the *Whole epoch decimated* is significantly lower than in the *P300 Epoch Realigned* condition ($p<10^{-5}$).

Only when the Covert GeoSpell and the P300 Speller interfaces were used, was a significantly higher accuracy obtained in the *P300 Epoch Realigned* with respect to the *P300 Epoch Non-Realigned* condition ($p<10^{-6}$).

Correlation between P300 latency jitter and classification accuracy

The non-parametric Spearman's rank correlation coefficient was used to evaluate the correlation between the classification accuracy as expressed by the ITR_{Mean} values and the P300 latency jitter obtained for each interface. We found a significant negative correlation between the latency jitter and the accuracy achieved by the subjects with all 3 interfaces (GeoSpell: r=.17 p=.04; P300 Speller: r=.35 $p=10^{-4}$; Visual Oddball: r=.18 p=.03).

Figure 4.11 shows the scatter plot and the related regression lines of the P300 latency jitter values and the ITR_{Mean} values for the Covert GeoSpell, the P300 Speller and the Visual Oddball interfaces.

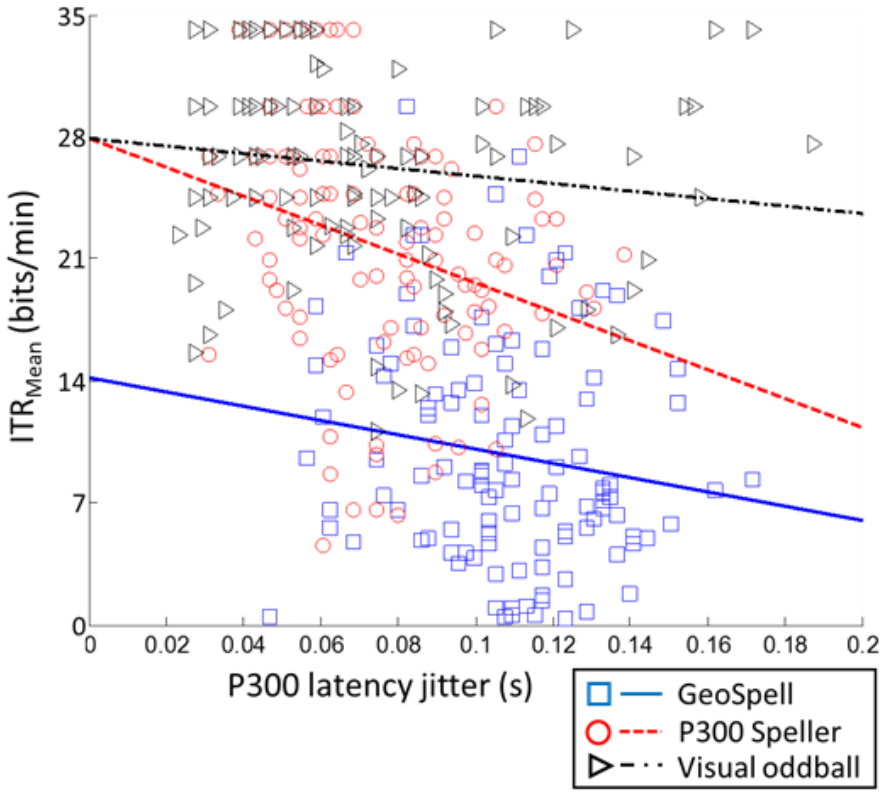


Figure 4.11: Scatter plot and regression lines of the P300 latency jitter and the ITR_{Mean} values relative to Covert GeoSpell, P300 Speller and Visual Oddball interface.

4.2.4 Discussion

The overall aim of this study was to investigate whether and to what extent the decrease of BCI accuracy using the covert attention based GeoSpell interface can be explained by the two following phenomena: (i) the lack of contribution of short latency VEPs (whose amplitude is mainly determined by foveation of the stimuli) in the tasks performed in covert attention modality; (ii) the lower temporal stability of the single-trial P300 potential when compared to corresponding potentials generated by

interfaces based on overt attentive tasks, such as the P300 Speller and a Visual Oddball interface.

In line with previous studies (Brunner et al., 2010; Treder and Blankertz, 2010), our findings on the first phenomenon clearly indicate the significant contribution of the early VEP components to the classification accuracy only for the overt (i.e. P300 Speller) interface. Also, removing the VEP contribution from ERPs elicited using the P300 Speller and the GeoSpell interface, the latter still performed significantly worse than the former, suggesting that the lack of VEPs is not the only reason for the performance decrement in the tasks performed in covert attention modality.

To test the relevance of the second phenomenon, we preliminarily contrasted covert vs. overt attentional tasks using a given visual interface (i.e. the GeoSpell). This first experiment proved that when the user operates a BCI using covert attention, the latency jitter is greater than using overt attention.

Capitalizing on this preliminary result, we evaluated the influence of the P300 latency jitter on the performance of a BCI classifier in a set of 3 different BCI visual interfaces, and tested whether the expected differences are reduced by pre-processing single trials to compensate for the P300 latency jitter.

As the main finding of this experiment, we found that for two out of three interfaces the reduced stability of the P300 potential evoked during the task is a significant contributor to the reduced accuracy.

4.4.1.2 *ERPs and (c)overt attention*

According to the waveform analysis, we found that the **latency** of the P300 evoked by the GeoSpell visual interface was significantly longer with respect to that elicited by

the P300 Speller interface, regardless of the required attention modality (i.e. covert vs. overt). In addition, an increase in the P300 latency also occurred when using the Visual Oddball interface as compared to the P300 Speller.

The finding of an influence of the stimulation interface on the P300 latency was somehow expected if one considers that in the case of the Overt and Covert GeoSpell and the Visual Oddball interfaces, the target and the non-target stimuli appear at the same spatial location of the screen. This implies that the subject cannot use the position of the stimulus as a feature to discriminate target types. Rather, discrimination must happen on the basis of the stimulus' shape only. On the other hand, in the P300 Speller the target and non-target stimuli are arranged in distinct positions in the matrix and the subject is allowed to foveate the target stimulus; in this case, discrimination is performed on the basis of a change of luminance occurring in the foveal region. Differences in latency can thus be plausibly ascribed to the timing of the categorization process, which would introduce longer delays with the GeoSpell and the Visual Oddball interfaces with respect to the P300 Speller.

The waveform analysis also revealed that the P300 latency **jitter** was significantly greater when using the Covert GeoSpell interface than using the Overt GeoSpell, the P300 Speller and the Visual Oddball interfaces. This result indicates that the *attention modality* does influence the magnitude of the jitter in the P300 latency. The latter may be partially ascribed to the dual task nature intrinsic to the covert attention modality (Peterson et al., 2004), which would make the task highly demanding. In fact using the Covert GeoSpell interface the users had to maintain gaze on the center of the screen (fixation cross) and simultaneously she/he had to pay attention to the surrounding stimuli. This interpretation is in line with previous evidence (Polich, 2007) of a larger

deviation in the P300 latency values that occur when attention is divided between two tasks. Additionally, we note that the specific succession of the presented shapes (characters) can facilitate or delay recognition of the target, which plausibly makes categorization timing less deterministic.

In agreement with *Experiment I*, the waveform analysis in *Experiment II* confirmed with an enlarged group of subjects that the P300 elicited by the Covert GeoSpell and the Visual Oddball (categorization of shapes) display longer latencies with respect to those evoked by the P300 Speller (categorization of luminance). The latency jitter was significantly higher for the Covert Geospell (covert attention) than the other two interfaces (overt attention).

As for the **amplitude** of the P300, the Covert GeoSpell interface elicited P300 responses of significantly lower amplitude with respect to the P300 Speller and the Visual Oddball interfaces. After the introduction of the single trial *realignment* procedure, the amplitude values of the P300 did not differ significantly ($p>.05$) between covert and overt interfaces. Also, the P300 amplitude estimated after *realignment* displayed significantly higher values than those calculated without *realignment*, regardless the type of interface. In fact, as expected, a natural consequence of the jitter in the temporal onset of the P300 is a 'smearing out' of the grand average ERPs, resulting in a decrease in P300 amplitude and an increase in the width of the P300 (Chennu et al., 2009; Handy, 2005; Kutas et al., 1977).

4.4.2.2 *BCI performances and latency jitter*

The main hypothesis of this study predicts that the usage of a covert attention-based BCI visual interface such as the GeoSpell would lead to a greater jitter of the P300 latency, which in turn would negatively affect the classification accuracy.

In fact, we confirmed that short latency VEPs, which are modulated by gazing at a flashing target, are a relevant feature when classifying ERPs acquired during an overt attention task: the accuracy of the P300 Speller deteriorates significantly (-19%) when only the P300 component is fed into the classifier. On the other hand, the accuracy attained by the P300 Speller is still significantly higher than the GeoSpell (+13%). Thus, we conclude that the modulation of early VEPs does not entirely account for the lower performance of BCI controlled in covert attention,

On the other hand, we showed that: (i) the attention modality significantly influences the amount of jitter in otherwise fixed experimental conditions (Experiment I); (ii) accuracy negatively correlated with the P300 latency jitter for all interfaces (Experiment II). In other words, covert attention increases the P300 latency jitter, and the greater the jitter the lower the accuracy of the classifier.

To further quantify to what extent the greater jitter accounts for the BCI loss of accuracy (as compared to other possible causes), we introduced an offline single trial analysis which realigns the P300 peaks following each stimulus, thus compensating the latency jitter. Comparing the classifier's performance with such post-processing, we observed a significant increase of the averaged P300 amplitude, and a substantial increase of performance of the BCI classification for both the Covert GeoSpell and P300 Speller interfaces. More importantly, the average accuracy of the Covert GeoSpell using realigned epochs is almost identical to the best performance of the

P300 Speller (94% vs. 92%). Taken together, these results lead to confirm our working hypothesis – the larger latency jitter associated to the tasks performed in covert attention modality largely explains the reduced performance of BCIs designed to be operated in absence of eye movements.

The improvement in performance produced by the realignment procedure may be simply explained by the consequent increase of the P300 peak's amplitude, even if for BCI classification purposes the averaging procedure is only carried out on a small number (5-20) of epochs, i.e. those acquired while a single character is selected. More effectively, the higher discriminability of P300 response may be directly accounted by the higher epoch-by-epoch stability of the feature vectors fed into the classifier; in fact this vector contains the values of the potential at a given latency, and the lower jitter implies more reproducible (less dispersed) features for the classifier.

It is worth noting that the realignment process requires information on the labels of epochs (i.e. target vs. non-target). While it is a useful analysis method to interpret the timing of single trial ERPs, in its present formulation it cannot be employed to improve performances of online BCIs.

5 EVALUATION OF THE OPERATORS' MENTAL WORKLOAD USING EEG RHYTHMS AND THE HEART RATE SIGNAL

5.1 Towards an EEG and HR based framework for realtime monitoring of mental workload

5.1.1 Introduction

A Brain-Computer Interface (BCI) is a communication system, which relies on brain activity to control an external device bypassing muscular and nerves pathway (e.g., using electroencephalogram (EEG) technique, Wolpaw et al., 2002). BCI research was originally driven by the goal to provide an alternative/additional channel to restore communication and interaction with the external world in people with severe motor disabilities. Recently, researchers suggested new application fields for BCI systems, developing applications that also involve subjects in operational environments, as military and commercial pilots and car drivers (Kohlmorgen et al., 2007; Borghini et al., 2012; Müller et al., 2008). Originally, the scope of the term BCI only included the translation of the users' intentions through the classification of their voluntarily modulated brain activity. In the new acceptation, the BCI meaning was broadened to comprise monitoring of cognitive states (e.g. mental workload, attention levels)

identified through the users' spontaneous brain activity. This kind of BCI was recently defined "passive" BCI (Zander and Kothe, 2011).

Mental workload monitoring is of particular interest in safety-critical applications where human performance is often the last controllable factor. In general as cognitive workload increases, maintaining task performance within an acceptable range becomes more difficult. Increased cognitive workload may demand more cognitive resources than that available in the human brain, resulting in performance degradation and errors (Norman and Bobrow, 1976). Objective measures of mental workload based on biomarkers could be used to evaluate alternative system designs, to appropriately allocate imposed workload to minimize errors due to overloads, or to intervene in real-time before operators become overloaded while performing safety-critical tasks (Byrne and Parasuraman, 1996). For example, some studies investigated neurophysiological indexes about the user states in safety-critical applications, such as driving (Welke et al., 2008), industrial environments or security surveillance (Venthur et al., 2010). With respect to driving assistance applications, recent studies have explored the use of psychophysiological measures in a driving simulation for assessing driving performance and inattentiveness (Schubert et al., 2008), as well as for robust detection of emergency brakes before braking onset (Welke et al., 2009). Another example of operative environment where a lack of performance or overloads of the work could be fatal is airplane-flying contexts. Mental workload of pilots could be too high due to complexity of the flying tasks to be performed simultaneously. In fact, besides control of the airplane, the pilot has to navigate, communicate and monitor the system. In these situations, a real-time system to estimate the mental workload of an operator can be useful in a future to avoid possible fatal consequences, by warning the operators or the

system that the task demands are going to be too much for the pilot. In addition, such a future system can be useful for facilitating training of operators by controlling a degree of cognitive efforts to be spent for accomplishing the required tasks.

The mental workload is a measure of the cognitive resources required to process information during a specific task (Nordwall, 1998). Several approaches have been proposed to evaluate the mental workload: i) subjective evaluation, ii) performance evaluation, and iii) psychophysiological variables assessment (Borghini et al., 2012). Firstly, the subjective evaluation is a measure assessed by subjective introspections, providing a rate for the perceived workload during the performed task (e.g. NASA-TLX; Hart and Staveland, 1988). Secondly, the performance evaluation provides a direct relationship between the performance achieved by the subject and the required mental workload (e.g. reaction times evaluation, number of lost targets; Colle and Reid, 1999). Finally, the psychophysiological measure consists in the evaluation of the variability (and of the correlation) of one or more neurophysiological signals (EEG, ECG, EOG, etc.) with respect to the mental workload required to the subject during the task. The assumption here is that modulations in psychophysiological features reflect changes in the operator mental states.

In this work, the EEG rhythms and the Heart Rate (HR) signals are taken into account, thus, only these features are briefly reviewed in the following.

- Several studies have associated the correlation of spectral power of the electroencephalogram (EEG) with the complexity of the task. For example, an increase of the theta band spectral power (4-7 Hz) especially on the frontal sites and a decrease of spectral power in alpha band (8-12 Hz) over the parietal sites have been observed when required mental workload increased (Mogford et al.,

1994; Pawlak et al., 1996). Although these changes in spectral power of the EEG signal are reproducible across subjects and stable over the time, their estimations are often slow (more than five minutes in order to highlight differences between different mental workload levels). In order to address this computational issue, different approaches have been adopted in other studies, e.g., the use of machine learning techniques employing linear and nonlinear classifiers allowed the system to assess subjects' mental workload in a short time (few seconds), reaching a high accuracy (>90%). To the best of our knowledge, only few studies have proposed on-line systems for assessments of the mental workload using the EEG signal (Kohlmorgen et al., 2007; Wilson and Russell, 2002). These systems are capable of predicting only two different workload levels (low and high workload).

- Since the hearth rate (HR) measure is easy to obtain and less sensitive to artefacts (Kramer, 1991), it is one of the most popular physiological parameters for mental workload assessments within various environments (Backs & Seljos, 1994; Wilson, 2002; Brookhuis & De Waard, 1993, 2001, 2010; Mehler et al., 2009). Also, cardiac measures can be used in real-world environments because they are unobtrusive and continuously available (Wilson, 1992). Here, it is assumed that an increased mental workload leads to an increased cardiovascular activity, a heightened cortical energy transformation, and corresponding enhanced metabolic demands (Backs & Seljos, 1994). Although this generalization is widely accepted, not all studies agree with these findings. It is known that HR is also sensitive to mental effort that is defined as the cost of the cognitive processing (Mulder, 1986). Numerous studies have found correlations

between cognitive demands and HR (Roscoe, 1992; Veltman & Gaillard, 1996, 1998; Caldwell et al., 1994). HR is also influenced by the contamination from physical effort, emotions and stress (Kramer, 1990). In a study on multitasking performance, Fairclough et al. (2005) explored the interaction between learning and task demand on psychophysiological reactivity. Authors found that a sustained learning effect was observed during the high demand condition only. In another study, Wilson (2002) evaluated cardiac, electrodermal and electrical brain activities of ten pilots during a 90-minute simulated flight in an experimental flight scenario. To test the reliability of psychophysiological measures of workload, each pilot performed the same scenario. It was shown that cardiac, electrodermal and electrical brain activity measures were highly correlated and exhibited changes in response to the demands of the flights. Recently, Raphaëlle et al. (2013) used HR for assessment of the mental workload of the users performing a modified Sternberg task (Sternberg, 1966), reaching a 57% of accuracy in discriminating two levels of workload. Taken together, the majority of previous researches have consistently demonstrated that an increase of workload led to an increase of HR (Borghini et al., 2012; Subhani et al., 2012; Larue et al., 2010).

Although several works have tried to use the psychophysiological measures to assess the mental workload, on-line systems are used in only few cases. In addition, the most of them were able to classify mental workload at maximum only two levels using EEG or ECG activity (e.g. low and high, Kohlmorgen et al., 2007; Wilson, 2002; Raphaëlle et al., 2013). Only a few studies succeeded in classification of three levels of mental workload using the EEG activity (Hope et al., 2011; Aricò et al., 2013). The purpose

of the present work is to examine whether the integration of the information derived from different biosignals (e.g. EEG, HR) can be a more reliable measure of the mental workload with respect of using just one physiological measure (e.g. EEG or ECG alone). For the purpose, an online passive BCI system to quantify the mental workload of subjects involved in multiple parallel tasks using the combination of the EEG and the HR biosignals was designed, implemented and evaluated. The framework was tested while subjects were performing a multitasking task at different difficulty levels with a clear relevance for the flight control. In addition to the capability to detect multiple levels of mental workload, stability of the system is of a great importance for a practical use of such device in real working contexts. In fact, the need to recalibrate the performance level of the subjects' limits through preliminary recordings each day made such kind of system unusable. Recent attempts were made to use the classification parameters estimated from EEG subjects in the previous day for the on line classification of the cerebral performance (Christensen et al., 2012). However, it was found that the performance of three different classifiers was significantly negatively impacted across days, raising the to classify over extended times. Here, using the combination of EEG and HR signals for the generation of the proper classification parameters, the stability of the estimated workload indexes over the time has been investigated. Results, showing a high reliability of the system for up to one week without recalibration, are reported.

5.1.2 Methods

5.2.1.1 *Subjects*

Ten healthy male subjects (mean age = 25±3) have been involved in this study. All subjects were students and/or staff members of the National University of Singapore (NUS). The study protocol was approved by the local Ethics Committee and all subjects gave their written informed consent. In addition, all the subjects have been paid to take part at the experimental protocol.

5.2.2.1 *Experimental protocol*

Scalp EEG has been recorded by the Waveguard© amplifier (ANT-neuro, Netherlands) with a sample frequency of 256 Hz from 16 EEG electrodes (FPz, F3, Fz, F4, AF3, AF4, C3, Cz, C4, P3, Pz, P4, POz, O1, Oz, O2) referenced to the earlobes and grounded to the AFz electrode. Also, the ECG and the vertical EOG activity were recorded at the same time of the EEG. The task performed by the subjects was the Multi-Attribute Task Battery (MATB, Comstock, 1994, see the section below for further details), which provides a benchmark set of tasks about operator performance and workload. In this study, we introduced three conditions characterized by different task difficulty levels (described successively in the text) to induce different mental workload levels in the subjects. Before the beginning of the protocol, the subjects have been trained to use the MATB software for five days and all the subjects reached a performance level above 90% on average over all the MATB subtasks in a single day, as stated in Borghini et al., (2013). The online evaluation protocol was composed of 6 recording sessions. The first four sessions were performed in two consecutive days

named hereafter as Day 1 and Day 2. Sessions were one in the morning and the other one in the afternoon for each day. The last two sessions were performed after about one week from the fourth session (Day 9). Each session consisted of 7 runs with two baseline conditions. During the first 3 and the last 3 runs (*offline runs*), the subjects performed the three MATB difficulty levels (easy, medium and hard subtasks). The fourth run (*online run*) consisted in a sequence of random combination of the three subtasks (easy, medium, hard). Each subtask with different difficulty levels has been presented twice in the sequence. This online run has been used for testing the online workload evaluation system. In order to avoid habituation effect, some task parameters have been randomly changed across the experimental sessions (e.g. tasks order presentation, radio frequencies, active emergency lights, etc). Each subtask lasted 2.5 minutes. Thus, the online sequence lasted 15 minutes in total. At the end of each run, the subjects were required to fill the NASA-TLX (Task Load Index, Hart and Staveland, 1988) in order to evaluate the perceived workload during the different tasks. Figure 5.1 shows the scheme of the experimental protocol.

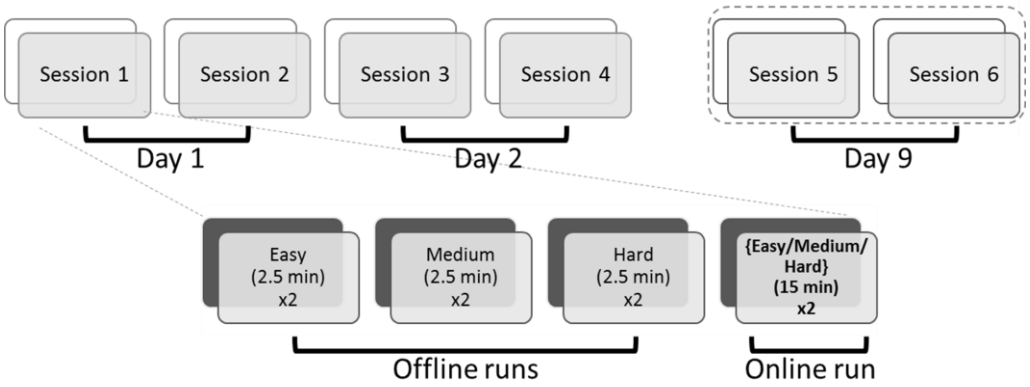


Figure 5.1: Experimental protocol scheme: each subject performed 6 recording sessions in three separate days, two sessions per day. The first four sessions

were performed within two consecutive days, whilst the remaining two sessions were performed after about one week from the fourth session in order to test the stability of the system over time. Each session consisted of 7 runs. During the first 3 and the last 3 runs (offline runs), the subjects performed the three MATB difficulty levels (easy, medium and hard subtasks). The fourth run (online run) consisted in a sequence of random combination of the three subtasks (easy, medium, hard). Each subtask has been presented twice in the sequence, so that the total duration of the online run was 15 minutes (2.5 min each subtask).

5.2.3.1 *Multi Attribute Task Battery*

The Multi-Attribute Task Battery (MATB, version 2.0, Figure 5.2) provides a benchmark set of tasks for use in a wide range of laboratory studies about operator performance and workload (Comstock, 1994; Wilson and Russell, 2003). The MATB simulates the activities inside an aircraft's cockpit and provides a high degree of experimental tasks control in terms of complexity and difficulty. The task features include an auditory communications task (to simulate Air-Traffic-Control communications), a fuel resources management task of maintaining target performance (e.g. to keep the fuel level around 2500 lbs), an emergency lights control and a task of cursor tracking that simulates the control of the aircraft flight level (this can be switched from manual to automatic mode). In this study, we introduce three conditions characterized by different task difficulty levels to induce different mental workload levels in the subject. The chosen tasks simulated three classic showcases in a flight scenario. In easy condition, subjects simply watched the MATB interface and its stimuli as the cruise flight phase. In medium condition, subjects had to maintain the

cursor in the center of the screen by manipulating the joystick to maintain the flight level. Finally, in hard condition, subjects had to perform all the MATB sub-tasks at the same time to simulate a few emergencies. As described in Borghini et al., (2013), four indices have been defined for each sub-task to evaluate performance of the MATB. In particular, the TRCK index is defined by considering the the ratio between the cursor's distance and the maximum of this distance (fixed) from the center of the screen. The indexes of the COMM and SYSM tasks are defined as a linear combination of accuracies in terms of correct answers (e.g., correct frequency selected) and the complement of the ratio between the subject's reaction time and the maximum time for answering.. Finally, the index for the RMAN task is defined as the mean value of the fuel's levels in the tank A and B. The results have been multiplied by "100" in order to obtain a percentage. In order to get a global index for the hard condition, the average of the previous indexes is calculated as single index as a percentage. Instead, the TRCK performance index has been considered for scoring the medium difficulty level condition. No performance index was evaluated for the easy condition because the subject was not to make an active control on the interface.

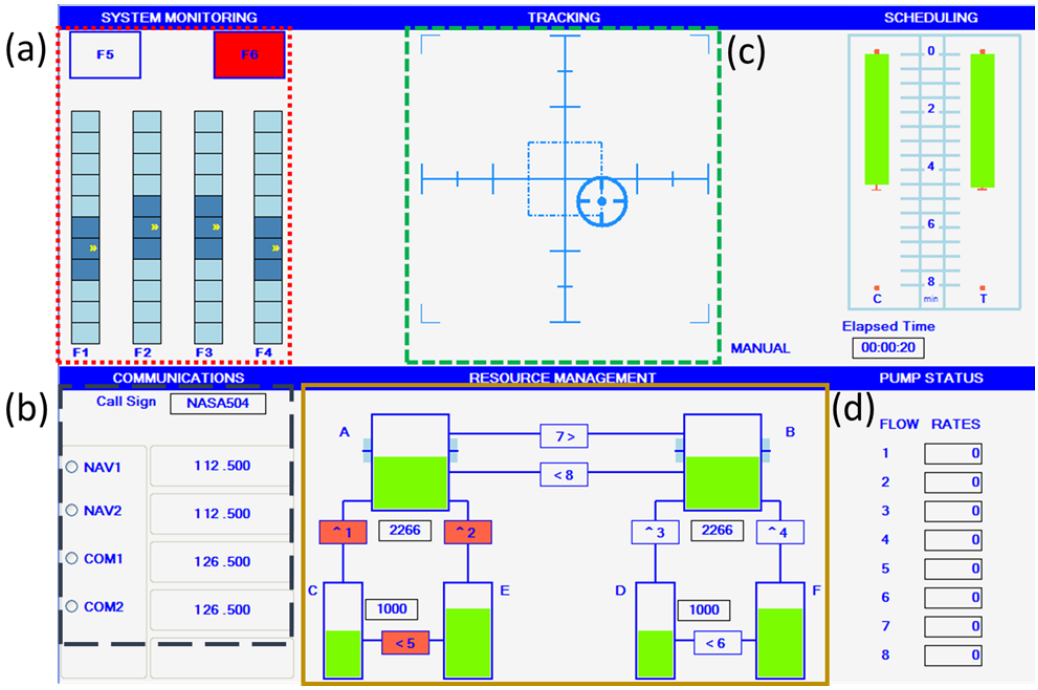


Figure 5.2: Screenshot of the Multi Attribute Task Battery (MATB) interface. On the top left corner (a, little dashed red box), there is the emergency lights task; on the top, in the center (b, medium dashed green box), there is the task of cursor tracking; on the left bottom corner (c, big dashed silver box), there is the radio communication task and, finally, in the center on the bottom (d, solid yellow box), there is the fuel levels managing.

5.2.4.1 *System development*

The developed system is capable of online estimation of the mental workload of the user based on the relevant features of the EEG and the HR signals highlighted by means of the SWLDA. The system is implemented under Matlab® using the TOBI interfaces (Breitwieser et al., 2012), which standardizes the information exchange procedures between different processing modules of the system. Particularly, the Tobi interface A (TiA) is a standardized interface to transmit raw biosignals (e.g. EEG,

EOG, ECG signals), Tobi interface C (TiC) is used to exchange messages between modules (e.g. output of classifier modules) and Tobi interface D (TiD) is used to exchange standardized high-level events between modules (e.g. time line of the experiment, start/stop events, markers, etc.). Data acquisition is driven by the TOBI Signal Server, which sends the acquired data to the following processing blocks (biosignal-signal- processing) in data streams compliant to the TOBI Interface A (TiA) format. The signal processing and classification modules deal with filtering, feature extraction and classification of the input signals. Classification results (further details about the classification process are provided in the *EEG classifier for mental workload evaluation* section and following) are transmitted to the Fusion application in TOBI Interface C (TiC) format. The Fusion module receives classification outputs from both the EEG and the other biosignal classifiers, by transforming them into "fusion classes" and then by transferring the information to the visualization module. The latter uses these signals (classifier output and biosignals) to provide a feedback to the operator and/or to the user. Finally, the controller module provides the clock to the whole system (TiD messages), according to the parameters previously set by the operator (initialization and finalization of each module, the time line of the experiment, markers, synchronization events, etc.). The communication between the modules is managed using the network protocol TCP / IP. A schematic overview of the developed system is provided in the Figure 5.3.

In this study, this online system was tested offline by simulating bio-signals using acquired data during online runs. For the purpose, all the offline runs (i.e., the first 3 runs and the last 3 runs of each session) were used to estimate parameters of classifiers, and the derived classifiers were used to evaluate mental workload during

online runs. To assess efficacy of the different biosignals, we compared workload indices derived from EEG, HR and a combination of EEG and HR as described in details below.

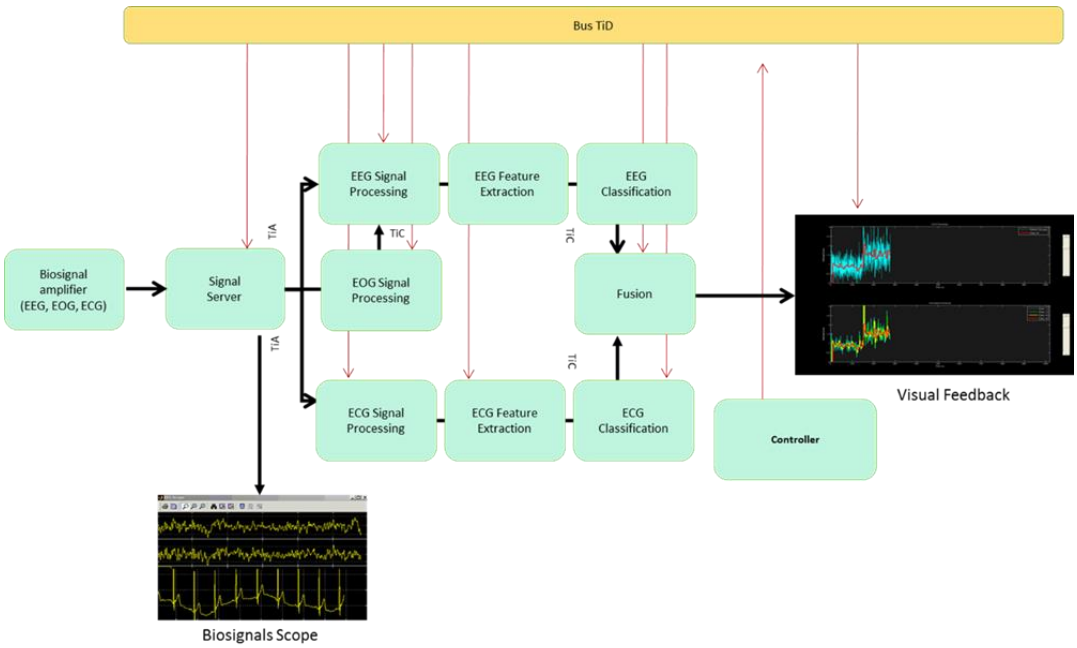


Figure 5.3: Workload measurement system architecture. The system has been entirely implemented in Matlab®, using the TOBI interfaces, that allow exchanging information between all the modules in a standardized way. Biosignals coming from the amplifier (EEG, EOG and ECG) are transmitted to the TOBI Signal Server, which sends the acquired data to the following processing blocks (biosignal-signal-processing). The signal processing modules deal with filtering, feature extraction and classification of the input signals. Classification results are transmitted to the Fusion application. The Fusion module receives classification outputs from both the EEG and the other biosignals classifiers, transforms them into "fusion classes" and then by transfers the information to the visualization module. Finally, the controller module

provides the clock to the whole system in accordance with the parameters previously set by the operator.

5.2.5.1 *EEG classifier for mental workload evaluation*

To train EEG classifiers to be used in the online mental workload evaluation, EEG-based mental workload index (W_{EEG}) was derived offline as follows. The EEG signal was band-pass filtered (0.1-40 Hz) and then segmented into epochs of 2 seconds, overlapping by 0.125 seconds. The EOG signal has been used to remove eyes-blink contribution from each epoch of the EEG signal, by using the Gratton and Coles (1983) algorithm available in the EEGLab toolbox (Delorme & Makeig, 2004). After that, for each epoch, the power spectral density (PSD) was calculated using a periodogram with Hanning window (2 seconds length), and a spectral features matrix for all the EEG channels was obtained within the frequency bands involved in the mental workload estimation (i.e., theta and alpha bands). A Stepwise Linear Discriminant Analysis (SWLDA, see the appendix A for further details) was used to select the most relevant spectral features to discriminate the mental workload levels from the training set (the first and the last 3 runs of the experimental session). Several moving average samples (N_{MA}) were applied to the output of the classifier (EEG based mental workload index, W_{EEG}: $N_{MA}(1) = 0.125$ (s), $N_{MA}(8) = 1$ (s), $N_{MA}(16) = 2$ (s), $N_{MA}(32) = 4$ (s), $N_{MA}(64) = 8$ (s)) to evaluate the stability and the accuracy of the index with the drawback of introducing delays in the workload estimation, inducing a decreasing of the workload refresh time (Figure 5.4a).

5.2.6.1 *HR classifier for mental workload evaluation*

As well for the EEG, HR-based mental workload index (W_{HR}) was derived as below. The ECG signal for each subtask (easy, medium and hard) was first band-pass filtered (0.1-40 Hz) to remove low frequency contributions, and then segmented into epochs of 8 seconds, with 0.125 seconds overlapped. The epoch length of 8 seconds was chosen to have enough R-peaks to calculate the HR. For each epoch, only the R-peaks have been extracted from the ECG signal, by using the method explained in Bhoi et al., (2012). The peak amplitudes of all the conditions were normalized by dividing them with the mean values of the peaks recorded during the easy condition. At this point, for each epoch, the PSD was evaluated using a periodogram with Hanning window (8 seconds length), considering only the frequencies bins closed to the HR (Figure 3b). As for the EEG analysis, using data from the training set (the first and the last 3 runs of the experimental session), a Stepwise Linear Discriminant Analysis (SWLDA) was used to select the most relevant spectral features to discriminate different levels. The same moving average samples (N_{MA}) showed in the EEG analysis were applied to the output of the classifier (HR based mental workload index, W_{HR}, Figure 5.4b).

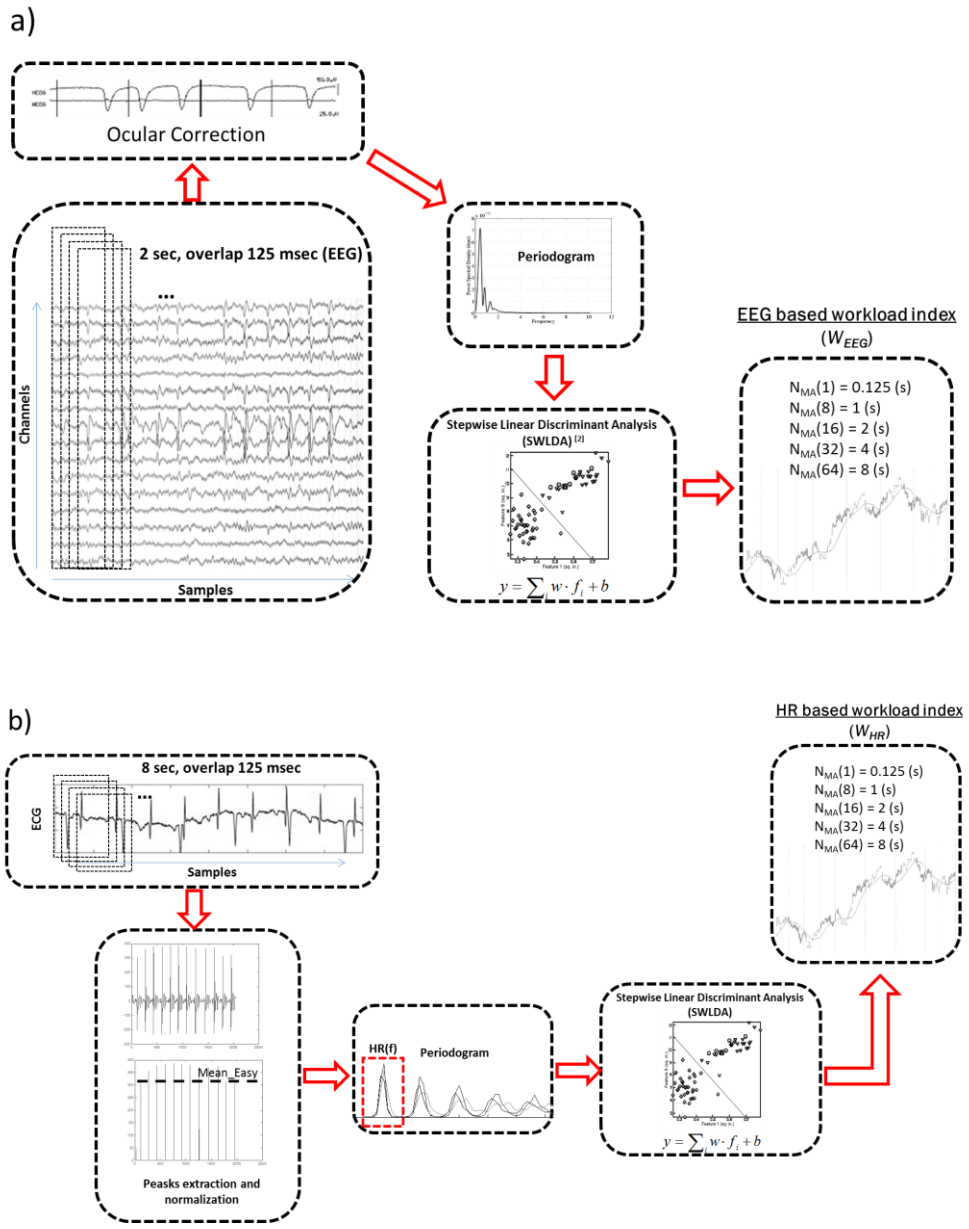


Figure 5.4: (a) EEG based workload index assessment (WEEG). The figure explains the algorithm of the evaluation of the EEG-based workload index. The band-pass filtered (0.1-40 Hz) EEG signal was segmented into epochs of 2 seconds, with 0.125 seconds overlapped. The EOG signal was used to remove the eyes-artefact contribution from the EEG signal. Then, the power spectral density (PSD) was

evaluated for each EEG channel, taking only the frequency bands involved in the mental workload estimation (i.e., theta and alpha bands) into account. After that, a Stepwise Linear Discriminant Analysis (SWLDA) was used to select the most relevant spectral features to discriminate the mental workload levels. Several moving average samples (NMA) were tested to the output of the classifier in order to evaluate the stability and the accuracy of the index. (b) HR based workload index assessment (WHR). The ECG signal for each subtask was first band-pass filtered (0.1-40 Hz), and then segmented into epochs of 8 seconds, with 0.125 seconds overlapped. The R-peaks were extracted from each epoch. The peak amplitudes of all the conditions were normalized by dividing them with the mean values of the peaks recorded during the easy condition. At this point, for each epoch, the power spectral density (PSD) was calculated using a periodogram with Hanning window (8 seconds length), considering only the frequencies bins closed to the heart rate. The SWLDA classifier was used to select the most relevant spectral features to discriminate different mental workload levels. The same moving average samples (NMA) showed in the EEG were applied to the output of the classifier.

5.2.7.1 *Fusion of the classifiers for mental workload evaluation*

A fusion-based workload index (W_{Fusion}) was computed as a combination of the EEG and the HR based workload indices. The two classifiers outputs were first synchronized with each other to eliminate delays, and then a new score (Fusion based workload index, equation 1, W_{Fusion}) was computed as a linear combination of the W_{EEG} and the W_{HR} scores (Figure 5.5).

$$W_{Fusion} = aW_{EEG} + bW_{HR} \quad (5.1)$$

Here, the coefficients *a* and *b* of the equation (1) were estimated for each subject, depending on individual contributions of the W_{EEG} and the W_{HR} scores to classifications. For each subject, these coefficients were calculated by means of a simple linear discriminant analysis (LDA), considering the EEG (W_{EEG}) and the HR score (W_{HR}) distributions over the offline cross validations for the three different difficulty levels. In particular, for each subject, the classifier was trained using the different difficulty levels, and the weights in output from the LDA were those who maximized the separation between the three difficulty levels.

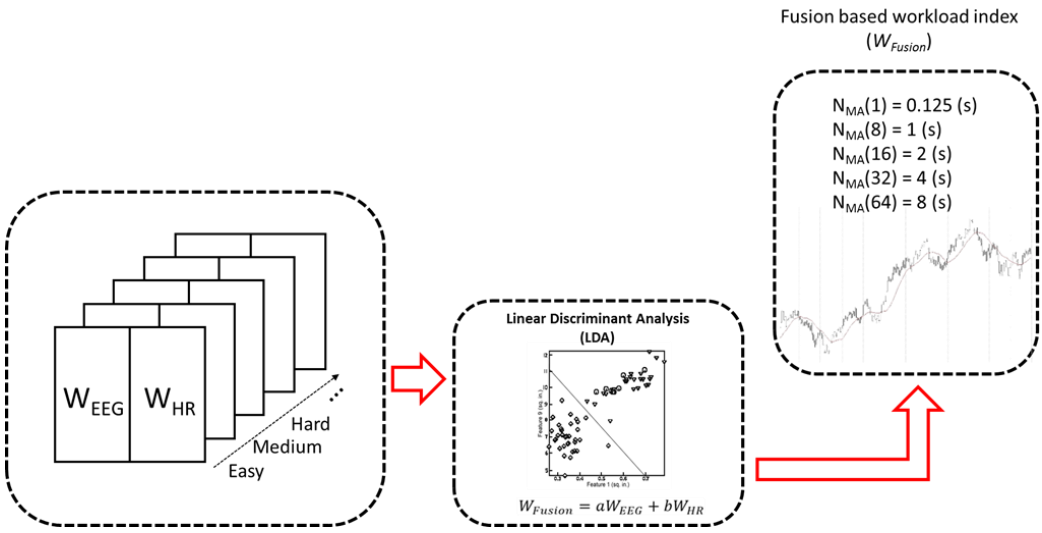


Figure 5.5: Fusion based workload index assessment (W_{Fusion}). The Fusion workload index (W_{Fusion}) has been calculated as a linear combination of the EEG and the HR based workload indices. The two classifiers outputs were synchronized before the computation of the fusion-based index.

5.2.8.1 *Performed analyses*

The analyses described below were organized in two categories; *Offline runs analyses* and *Online runs analyses*. The Offline runs analyses refer to the analyses on data collected during offline runs (the first three and the last three runs for each session) and were performed to see behavioral and electrophysiological difference between different mental workloads, while the Online runs analyses refer to those on data collected during online runs (the fourth run for each session) were conducted to evaluate the online mental workload evaluation system we developed. The Offline runs analyses consisted of i) NASA-TLX assessment of the workload, ii) power spectra analysis and iii) performance analysis. NASA-TLX analyses were used to assess the subjective perceived workload to be sure that the psychophysiological behavior was consistent with the perceived one; power spectra analyses highlighted the EEG and HR patterns that were modulated by the mental workload changes. Finally, performance analyses assessed how the system was able to discriminate different workload levels. In the Online runs analyses, classification parameters derived from the EEG, the HR, or a combination of them were estimated using data during off-line runs (the first three and the last three runs for each session), and then applied to data collected during the on-line run (4^{th}) to estimate the fusion based mental workload index (W_{Fusion}). Visualization of trends of the workload index is possible in real-time on the visual interface (Figure 5.6). Furthermore, different workload indices ($W_{EEG,}$ W_{HR}) were computed, and analyzed their distribution provided by the system. Finally, the performances of the MATB task were evaluated to highlight any differences between experimental sessions and difficulty levels conditions.

5.2.9.1 *Offline runs analyses*

i) NASA-Task Load Index (TLX): Subjective perceived workload evaluation was obtained by filling the standard NASA-TLX questionnaire for each subtask (easy, medium and hard). The given subjective scores were used to estimate the perceived workload by considering six different factors: Mental Demand, Physical Demand, Temporal Demand, Frustration, Effort and Performance. The workload scores ranged from 0 to 100 were obtained for each factor at the end of the questionnaire. The subjective scores of the perceived workload were compared with the workload indices estimated using the system. A one-way ANOVA (CI=.95) was performed on the NASA-TLX scores with the subtask (Easy, Medium, Hard) as an independent variable. In addition, Duncan post-hoc tests were performed to test the differences between all the levels.

ii) Power Spectrum analyses:

a. *EEG:* The differences in the power spectra were evaluated between each couple of conditions; LOW vs. HIGH (i.e., easy vs. hard, easy vs. medium and medium vs. hard). For each couple, the signed Coefficient of Determination (R^2 value; see appendix B for further details) was quantified. The R^2 values range from 0 to 1, and higher values correspond to larger explained variance: an higher discriminability of classification among conditions. The signed R^2 indices were derived by multiplying the R^2 by the sign of the slope of the corresponding linear model; positive sign is obtained when the PSD values of the signals considered during the HIGH (e.g. hard) subtasks are higher than that

related to the LOW (e.g. easy) subtasks, and *vice versa* for a negative sign. The signed R^2 values calculated for the theta (3-7 Hz) and alpha (8-12 Hz) frequency bands at all the scalp positions were subjected to one-way repeated measures ANOVAs (CI = .95) with the three couple of conditions (easy vs. hard, easy vs. medium and medium vs. hard) as an independent variable. In addition, a Duncan post-hoc tests were performed to highlight differences between all the levels.

b. *HR:* In order to highlight the differences in the HR signal between the different subtasks (easy, medium and hard), a dimensionless index taking into account the contribution of both the power spectrum of the HR signal correspondent to the heart beat and the related frequency was calculated. For comparison of these two values, the values were first normalized, dividing them by the maximum value in the easiest condition. After that, they were averaged to have a synthetic index of the HR signal (HR_{index}). In order to analyze the differences between the HR signals recorded during the different subtasks, the derived HR_{index} values were subjected to a one-way repeated-measures ANOVA (CI = .95) with the subtasks (easy, medium and hard) as an independent variable. In addition, a Duncan post-hoc test was performed in order to highlight the difference between all the factors.

iii) Performance analyses: In order to evaluate the performance of the system, the dataset has been re-organized into 12 triplets of runs (easy, medium and hard subtasks; 2 triplets per session). All the possible cross-validations were considered, training a classifier with one of the triplets and testing the

extracted features over the remaining triplets. To evaluate the accuracy of the system, values of the Area Under Curve (AUC) of the Receiver Operating Characteristic (ROC, Bamber, 1975) were calculated from the outputs of the classifiers (for each different refresh rate). The AUC values represent how well the classifier separated two different subtasks, and so how well the classifier could predict the difficulty of the task directly related to the subject's workload level. These kinds of analyses were performed on the W_{EEG}, the W_{HR} and the W_{Fusion} workload indices. A three-way repeated measures ANOVA (CI = .95) was performed on the AUC values using the types of bio-signals used for the classifiers (EEG, HR and Fusion based), the couple of subtasks (easy vs. hard, easy vs. medium, and medium vs. hard), and the moving average lengths ($N_{MA}(x)$, x={1, 8, 16, 32, 64}) as dependent variables. In addition, Duncan post-hoc tests were performed to test the difference between all the levels.

5.2.10.1 *Online runs analyses*

i) *Workload score distributions:* As described before, in the *Online runs analyses*, the score distributions of the single subtasks were simulated offline within the 4th run (online run). First, the classifiers were trained with every combination of the triplet (easy, medium and hard subtasks) during the offline runs (1st-3rd; 5th-7th) within each session. Thus, three classifiers (Day 1, Day 2 and Day 9) were derived for each subject. Then, the extracted features were tested for the online runs (4th). To investigate short- (INTRA) and medium-term (INTER) stability of the classifiers for the workload

evaluation, we performed two types of cross-validations. For the short-term stability test (INTRA), the classifiers trained with the offline data on Day 1, Day 2 or Day 9 were tested with the online data on the same day. For the medium-term stability test (INTER), the classifiers trained with the offline data collected on Day 1 and Day 2 were tested with the online data on Day 9 while the classifier trained with the offline data on Day 9 was tested with the online data on Day 1 and Day 2. Figure 5.7 depicts a schematic overview of the INTRA and INTER type cross-validations. These analyses were performed separately for the W_{EEG}, the W_{HR} and the W_{Fusion} workload indices. The online system was tested during the simulated MATB-sequence based on the data collected from the subjects by visualizing in real-time the output of the classifier onto the visual interface. Moreover, the discriminability between the three estimated workload distributions (easy, medium and hard) was evaluated for the three mental workload indices (W_{EEG}, W_{HR} and W_{Fusion}). Also, the short- (INTRA) and the medium-term (INTER) changes of the workload indices were tested. Three two-way repeated measures ANOVAs (CI = .95) were performed on the workload index distributions (W_{EEG}, W_{HR} and W_{Fusion}) with subtask (easy, medium and hard) and cross-validation type (INTRA and INTER) as independent variables.

ii) *MATB performances:* The MATB performances achieved online by the subjects within conditions (medium and hard) and sessions were evaluated, following the procedure described in Borghini et al., (2013). A two-way repeated-measures ANOVA (CI = .95) was performed on the MATB

performances with the session number of the offline runs (from 1 to 6) and the subtasks (medium and hard) as independent variables.

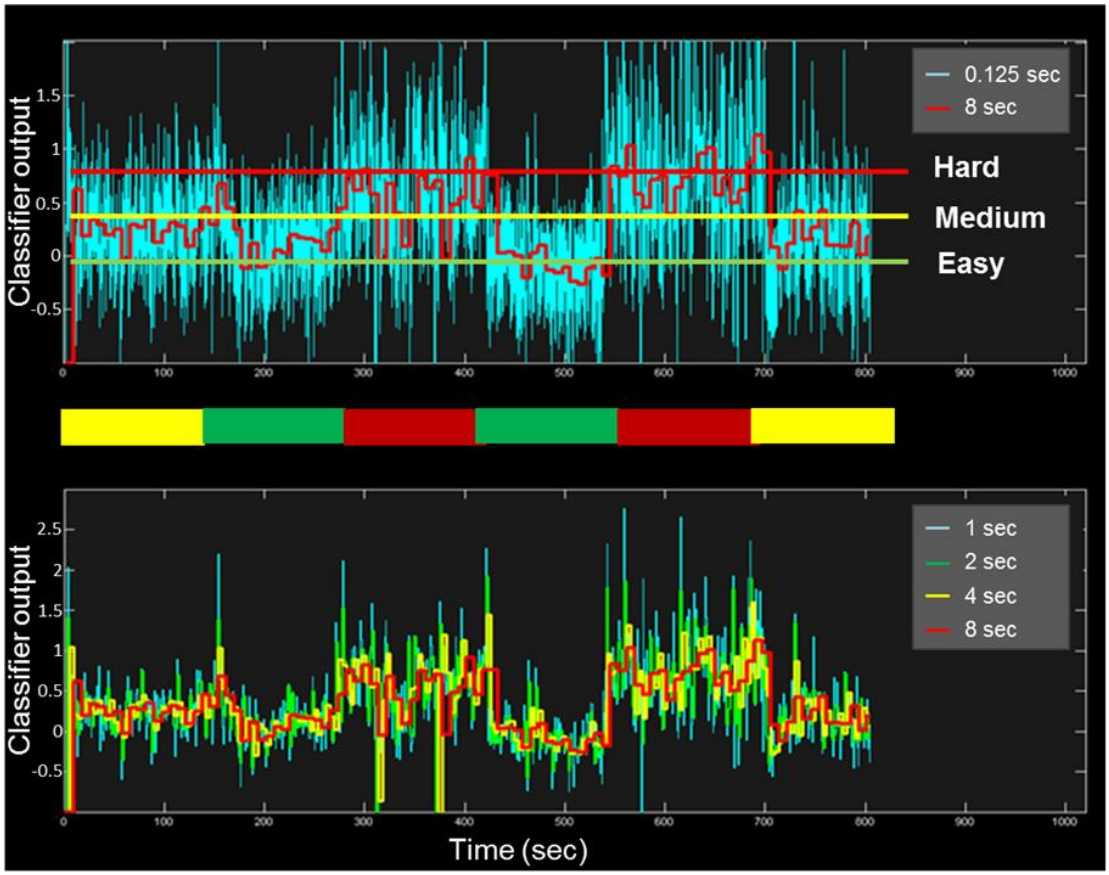


Figure 5.6: Screenshot of the visual interface provided to the operator that allow visualizing the fusion based workload index (W_{Fusion}) over time. In the upper side of the screen the workload index for the low and the high refresh rates are visualized. In the bottom part the NMA(x), x={8, 16, 32, 64} are visualized in real time. It is possible to note the variation of the index level related to the occurrence of the task difficulties.

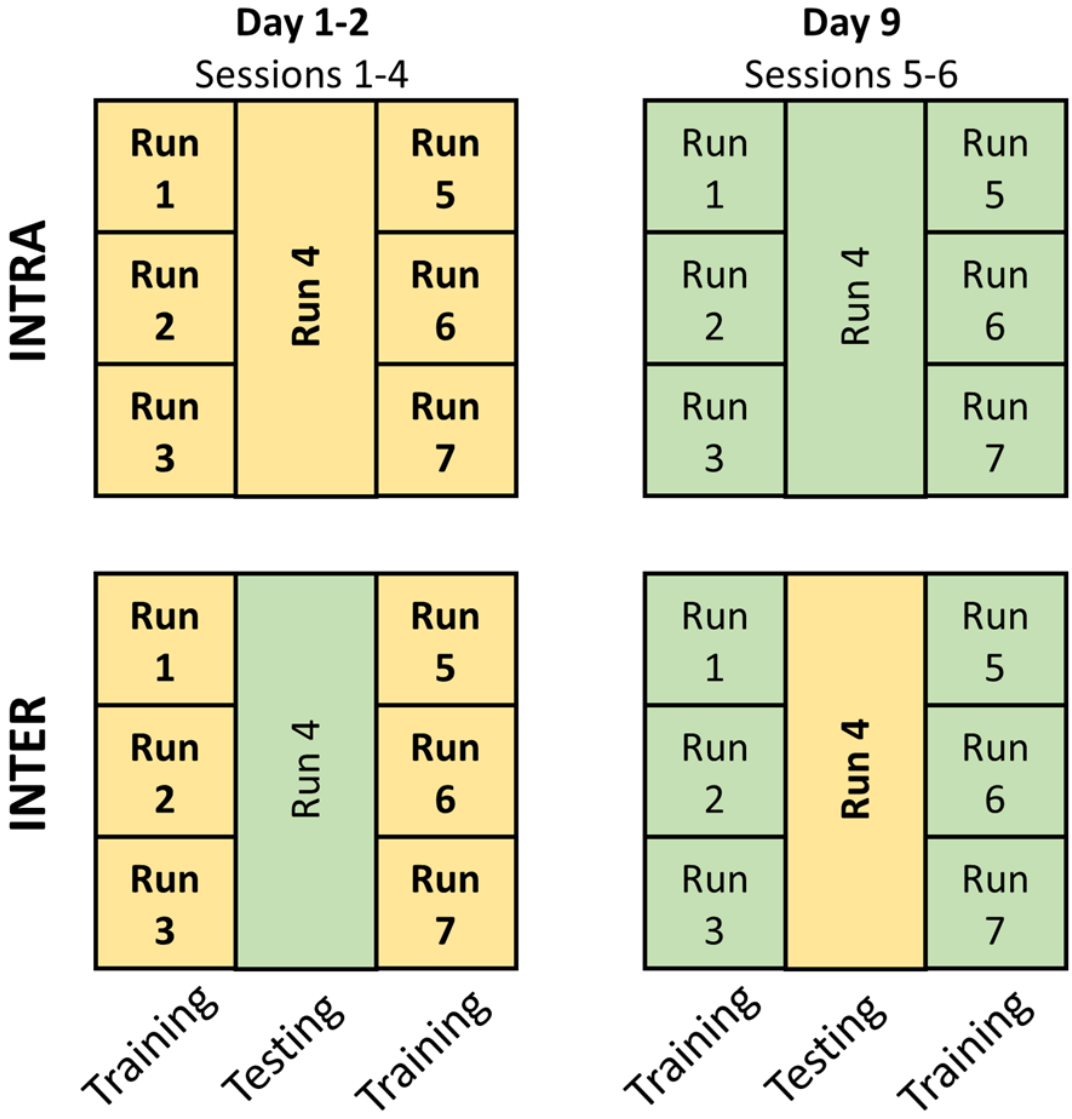


Figure 5.7: Schematic overview of the INTRA and INTER type cross-validations. The INTRA type refers to the cross-validations performed considering as training sessions those related to Day 1 and Day 2, reported in the yellow bold boxes (Day 9, reported in the green boxes) and as testing sessions those performed in the same days, Day 1 and Day 2, reported in the yellow bold box (Day 9, reported in the

green box). Contrariwise, the INTER type refers to the cross-validations performed considering as training sessions those related to Day 1 and Day 2, reported in the yellow bold boxes (Day 9, reported in the green boxes) and as testing sessions those performed in the Day 9, reported in the yellow bold box (Day 1 and Day 2, reported in the green box) and vice versa.

5.1.3 Results

5.3.1.1 *Offline analyses*

NASA-Task Load Index (TLX)

Figure 5.8 shows the changes in the perceived workload estimated by the NASA-TLX scores for the different subtasks. Roughly speaking, the perceived workload increased as the difficulty of the task increased as can be seen in the figure. The repeated-measures ANOVA revealed a main effect of the difficulty levels ($F(2,18)=27.68$, $p=10^{-6}$). The post-hoc test showed that the hard subtask showed a significantly higher workload than the other two subtasks (all $p<10^{-3}$). Although the perceived workload for the medium task was higher than the easy task on average, there was no significant difference between the two subtasks ($p=.56$).

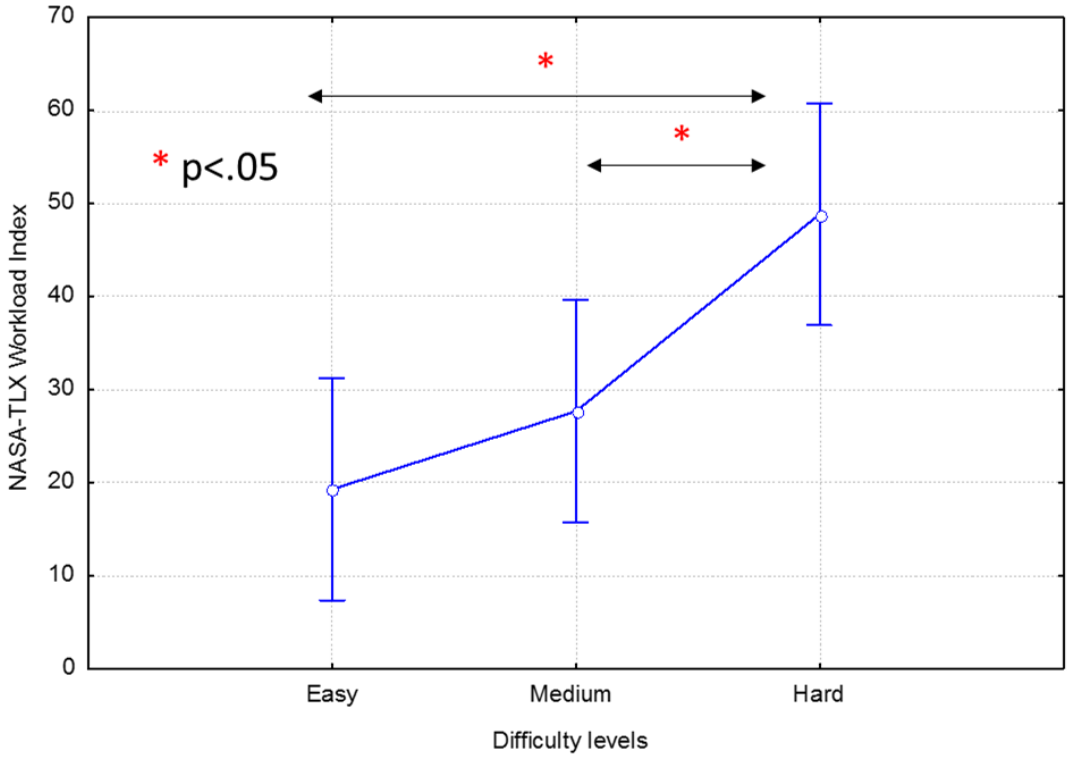


Figure 5.8: Mean and standard error of the perceived workload estimated by the NASA-TLX workload scores over the different subtasks.

EEG Power Spectrum analyses

Differences of the EEG power spectral density between LOW and HIGH subtasks (easy vs. hard, easy vs. medium, and medium vs. hard) was evaluated using the signed Coefficient of Determination (R^2), for each channel and each frequency bin. The Figure 5.9 represents the grand average of the signed R^2 indices of the EEG power spectral density evaluated between the three pairs of LOW and HIGH conditions over all the experimental sessions. The results showed an increment of R^2 in the theta bands over all the scalp positions, especially in the easy vs. medium and the medium vs. hard conditions, and a decrement of R^2 in the alpha band especially over the centro-parietal

areas in all the three conditions. However, the differences between the power spectrum of the frequency bands are smaller for the easy vs medium pairs of conditions, for each considered band (Theta: $F(2,18)=20.54$, $p=2x10^{-4}$; Alpha: $F(2,18)=8.85$, $p=.002$). The Duncan post-hoc test showed that both the signed R^2 values related to the theta and alpha bands are significantly different between the easy and hard and medium and hard conditions (all $p<.05$), but not between the easy and medium conditions.

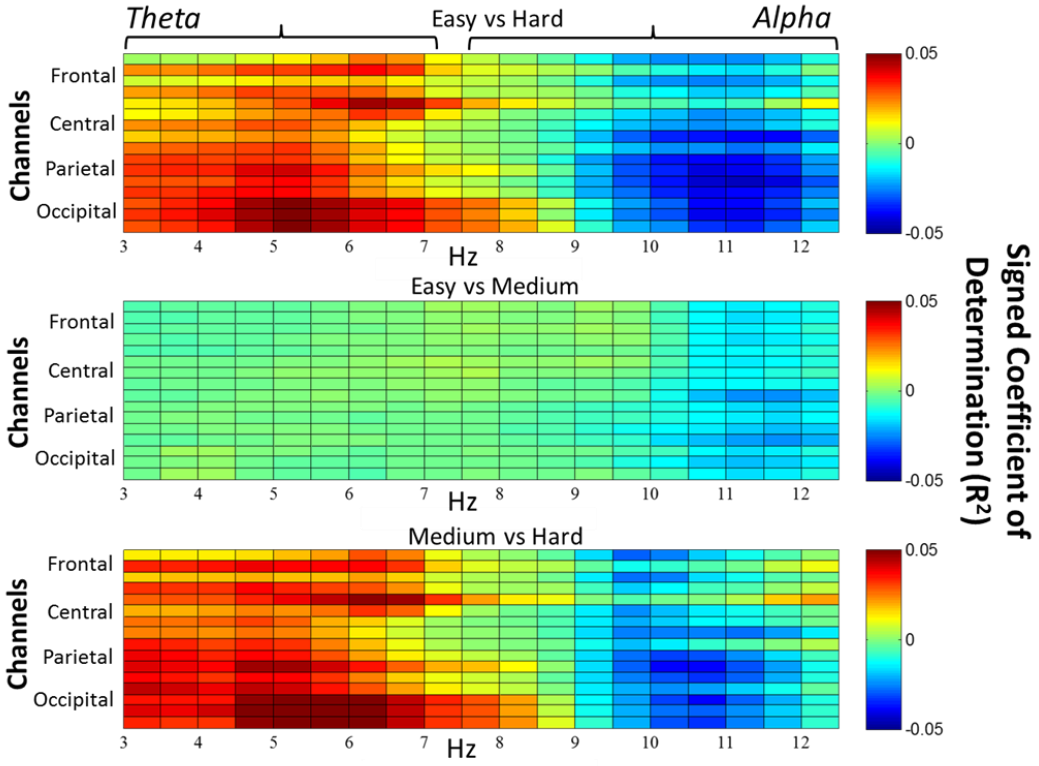


Figure 5.9: Grand average over all the subjects of the signed R^2 indexes of the EEG PSD evaluated between the three pairs of LOW and HIGH conditions (easy vs hard, easy vs medium, medium vs hard) over all the experimental sessions. Abscissa is the frequency (Hz), while on the ordinate represent the scalp electrode locations. The first panel from the top showed the statistical variation of the signed R^2 index

relative to the Easy versus Hard task condition. The central panel is instead related to the variation of the index in the Easy versus Medium condition and the lower panel for the Medium versus Hard condition. Red (blue) colors in the first panel represents a particular conditions between frequency band and scalp electrode location in which the estimated R^2 index is higher (lower) in the Hard condition than in the Easy one. In general, the R^2 analysis shows an increasing value of the EEG PSD in theta band related to the HIGH conditions with respect to the LOW ones, and a decreasing value of EEG PSD in the alpha band.

5.3.2.1 *HR Power Spectra analyses*

The ANOVA shows a main effect of the subtasks on the combined HR index (HR_{index}) values ($F(2,18)=5.95$, $p=.01$). The Duncan post-hoc test showed that the HR_{index} values during the hard task were significantly higher than the easy and medium ones (all $p<.05$). No significant difference was found between the easy and the medium conditions ($p=.25$), though an increasing in the HR_{index} values of the medium task with respect to the easy one (Figure 5.10).

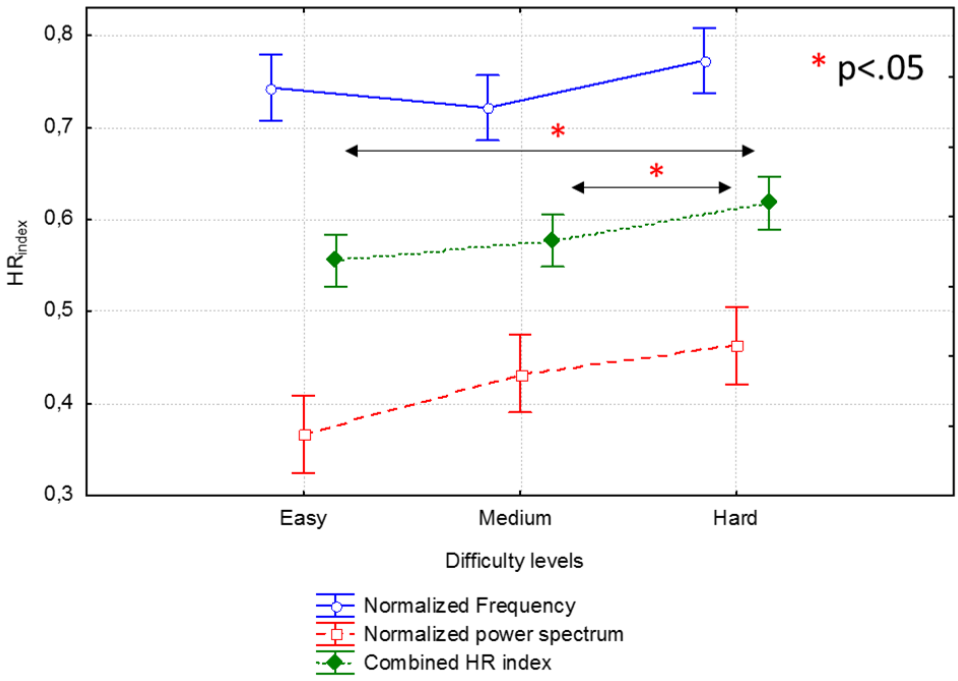


Figure 5.10: Mean values and related standard errors (CI = .95) of the HR index calculated over all the subjects, and the experimental sessions for the three subtasks (easy, medium and hard) for each component of the HR index (the frequency, the power spectrum and the combined).

5.3.3.1 *System performance analyses*

Figure 5.11 represents the accuracy of the system revealed by AUC values calculated using the different moving average lengths for the EEG, the HR and the Fusion based workload indexes. The ANOVA analyses revealed no main effect of the classifiers ($F(2, 18)=.27$, $p=.76$), a main effect of conditions ($F(2, 18)=28.76$, $p=10^{-5}$) and a main effect of refresh time ($F(4, 36)=256.21$, $p=10^{-6}$). The post-hoc test showed that AUC values calculated using the EEG based classifier in the "easy vs medium" couple were significantly lower (all $p<10^{-6}$) than the other two ones. Also, increasing the refresh

rate, the AUCs of the system significantly increase (all p<.05). The same behaviors have been obtained using the Fusion based classifier. For the HR based classifier, the AUC values for all the refresh time values and couples of tasks are not significantly different (all p>.05). Finally, the analysis revealed that the HR and the Fusion based classifiers performed significantly better than the EEG classifier (all p<.05) for the fast refresh rates. Instead, the EEG and the Fusion based classifiers performed better than the HR based classifier (all p<.05) for the high refresh (Figure 5.12).

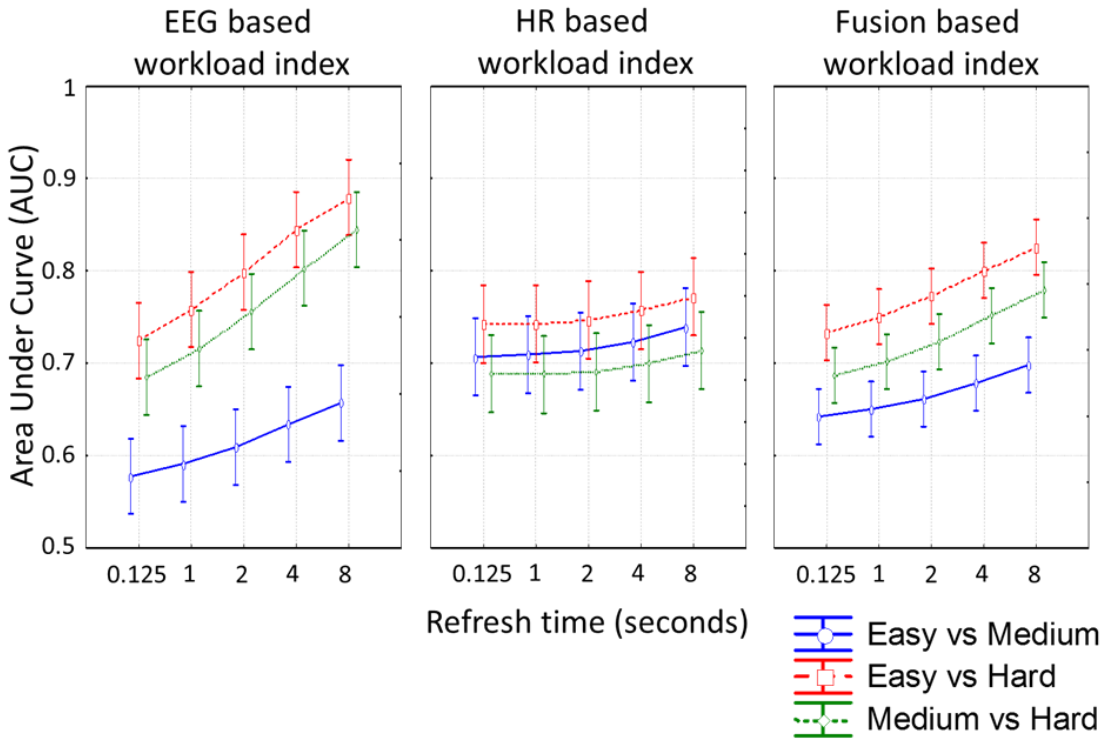


Figure 5.11: Mean values and related standard errors (CI = .95) of the AUC values achieved using the different classifiers (EEG, HR and Fusion-based) for each refresh time value.

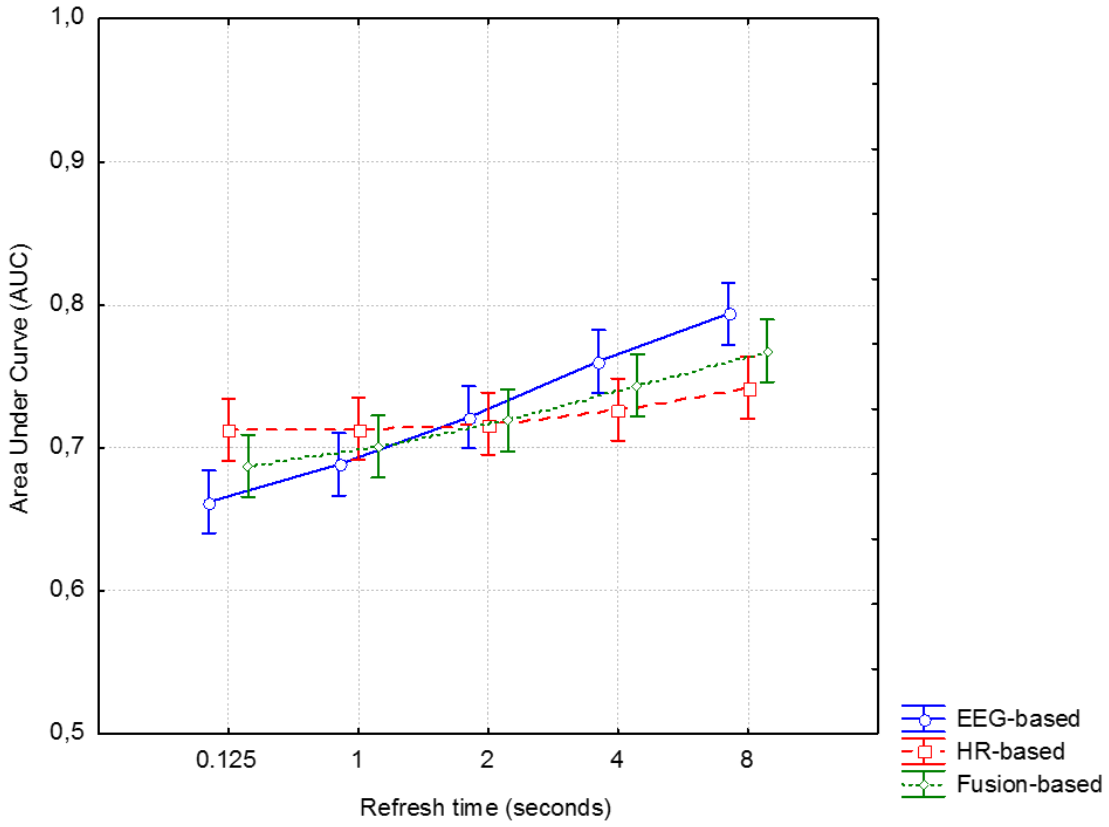


Figure 5.12: Mean values and related standard errors (CI = .95) of the AUC values of the three classifiers (EEG, HR and Fusion based) over the different refresh time values.

5.3.4.1 *Online analyses*

Workload score distributions

The ANOVA analysis revealed that the score distributions related to the different subtasks (Easy, Medium and Hard) for all the three classifiers were significantly separated (EEG-based: $F(2,18)=37.84$, $p=10^{-6}$; HR-based: $F(2,18)=13.69$, $p=2.4\text{x}10^{-3}$, Fusion-based: $F(2,18)=36.52$, $p=10^{-7}$). Furthermore, no significant differences were

found between the workload scores related to the INTER and the INTRA cross-validations, for each classifier (EEG-based: $F(1,9)=.20$, $p=.67$; HR-based: $F(1,9)=.85$, $p=.38$, Fusion-based: $F(1,9)=10^{-4}$, $p=.99$). Figure 5.13 shows the error bars related to the distributions of the workload indexes (W_{EEG}, W_{HR} and W_{Fusion}) evaluated by means of the three classifier (EEG, HR and Fusion-based).

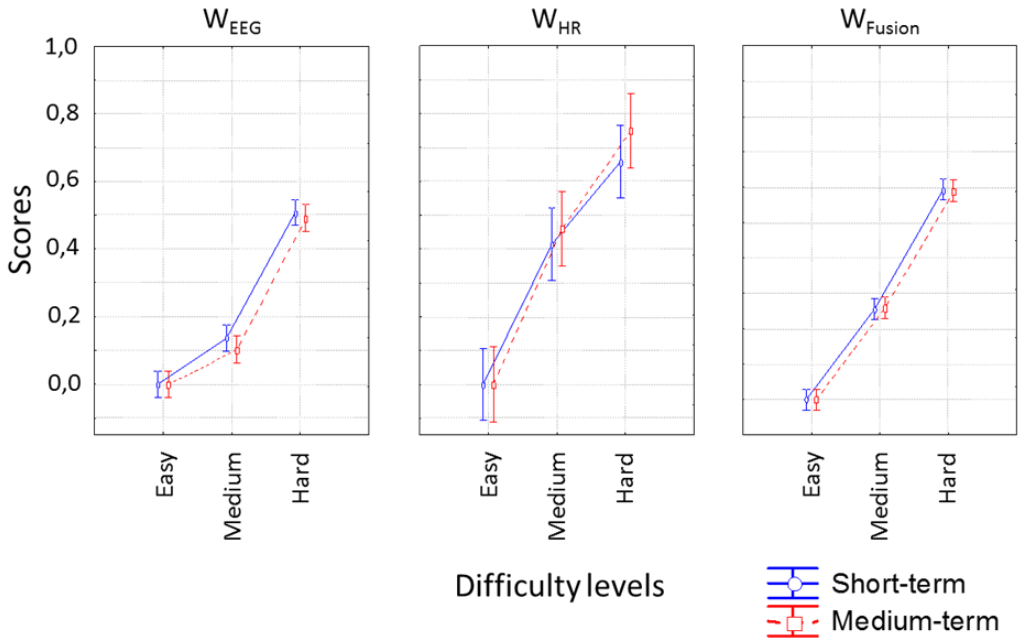


Figure 5.13: Mean values and related standard errors (CI = .95) of the distributions of the workload indices (W_{EEG}, W_{HR} and W_{Fusion}) evaluated by the three classifier (EEG, HR and Fusion based).

MATB performance

The repeated-measure ANOVA revealed no main effects of task difficulty and days ($F(5, 45)=.52$, $p=.76$), suggesting that the subjects were trained well with the MATB task by Day 1. The ANOVA revealed no significant differences between the MATB performances (Mean values: 92.7 ± 2.4) achieved by the subjects within the different

experimental sessions ($F(1,9)=.51$, $p=.49$) and medium (92.9 ± 1.5) and hard (92.4 ± 3.1) conditions.

5.1.4 Discussion

In this work, an online passive BCI system to classify subject's mental workload online has been demonstrated using the brain and heart activities. The system has been tested with ten healthy subjects performing the MATB task which simulates the cockpit of an airplane. In particular, the employed tasks run over three different difficulty levels (Easy, Medium and Hard) resembling different flight conditions (cruise flight phase, flight level maintaining, and emergencies). Three different classifiers have been simulated and tested offline, by using the EEG (EEG based classifier) the HR (HR based classifier) signals alone and the combination of them (Fusion based classifier). Results demonstrate that the EEG spectra show an overall increase in the theta band and a decrease in the alpha band as the difficulty level of the task increased. Furthermore, the HR activity increased in the same way as the difficulty level of the task increased. In particular, the spectral analysis showed less discriminability between the autopilot (easy) and the tracking (medium) conditions associated with the emergencies one. As the mental resources required to perform the autopilot and the tracking tasks are less demanding than in an emergency condition, these neuroelectrical results were expected, as confirmed by the NASA-TLX questionnaire analyses. In addition, it was already demonstrated that the increase of the mental workload induced an increase of the EEG spectral power on the frontal areas as well as a decrease of the EEG spectral power in parietal areas (Mogford et al., 1994; Pawlak et al., 1996; Borghini et al., 2012).

The performance analyses, as well the workload distribution analyses for all the classifiers showed a significant discriminability ($p<.05$) between the different difficulty levels when considering all the classifiers. Furthermore, the statistical analyses of the stability of the computed workload score in the short and medium terms did not show any significantly difference ($p>.05$), demonstrating that the features extracted by the classifiers are stable over the time, and that even after a week may not be necessary to recalibrate the system with new data. Also the MATB performances showed that after a week the performances remained stable and in general within all the experimental sessions. These results demonstrate that the classification features chosen by the classifier do not change significantly after a week and that the system is able to differentiate significantly among the three imposed difficulty levels. These aspects related to stability and accuracy are highly important for the usability point of view of the system. In fact, to use such system in real environments it could be enough to calibrate the system with the specific parameters of the operator once and then just use it without further adjustments maintaining a high reliability over at least a one week period.

Since the refresh time of the system decrease until reaching an AUC of around 0.9 related to the slower refresh rate, the EEG-based classifier finally showed a statistical increase in the performance. The HR based classifier showed no significant improvements, also decreasing the refresh rate of the system, and allowed AUC to reach higher than 0.7 for all the conditions. The fusion-based classifier reached an AUC higher than the EEG-based classifier at fast refresh time values, the same as the EEG-based classifier at fast refresh times and higher than the HR-based classifier at the fast refresh times. These results demonstrate that by combining information

coming from different biosignals (e.g. EEG and HR), it is possible to have more reliable and faster information about the mental states of the user. This multi-modality approach can be used in real operating environments for improving the human machine interaction, not only for pilots, but also for other users, such as air traffic controllers, car drivers or more in general for all the contexts, in which the high stress conditions can cause a critical drop in performance. It is worth of noting that the overt behavior of the subjects did not differ in terms of MATB scores between the Medium and Hard conditions (e.g. they perform the tasks in a statistically similar manner) while their cerebral and cardiac activity across such conditions changes significantly. This implies that the overt behavior measurements of the subject's performance is not a reliable indicator of the mental workload perceived during the task while the link between the mental workload perceived and the increase changes in EEG PSD and HR activity are more robust and stable. In conclusion, a human mental state classification system using the neurophysiological information has been demonstrated with a fairly realistic scenario aircraft pilots may encounter. Our system able to do online estimation of the mental workload by using the combination of EEG rhythms and HR signals has been proposed. We have demonstrated that i) the system is able to significantly differentiate three workload levels related to three difficulty level tasks with a high reliability; ii) the subjective features used for the evaluation of the mental workload remain stable over one week and iii) an online implementation of mental workload assessment if feasible using our approach. Mental states monitoring is of particular interest especially in safety-critical applications where human performance is often the least controllable factor. In this way, the proposed system could be useful during the operator's training to measure its cognitive workload and spare capacity

while facing specific operative and emergency conditions. The innovation with respect to the present literature is the possibility to predict online the mental workload of the user over three difficulty levels, using the combination of multiple biosignals (EEG and HR), that improve the reliability of the estimated mental states as compared with a single measure. Another relevant aspect of innovation of the presented results is that the classification features chosen by the system are stable after a week. This aspect can be very important when using such system in a real work environment scenario. Further experiments will be performed to even further test and extend the long term use of the system, and whether some unsupervised recalibration can be carried out when any decrement in the performance is observed.

6 CONCLUSION

The main purpose of this PhD thesis was to demonstrate how the passive Brain Computer Interfaces (BCIs) concepts could be used to assess the mental states of the users, in order to use them for improving the human machine interaction (HMI, Zander, 2011). For this purpose, different methodologies have been proposed and validated. Two main studies have been reported.

In the first proposed study (section 4), it has been estimated the morphological variations in the Event Related Potentials (ERPs), such as latency, latency jitter and amplitude using two reactive BCI systems in two different attention modalities (overt e covert attention). It was demonstrated as these variations can be used as an objective index to assess the attentional resources and the mental workload perceived by the user during the BCI tasks. Furthermore, they can also be used as a predictor of how well the subjects are performing the BCI task itself. In the perspective of the passive BCI systems, these physiological indexes could be used in closed loop for improve the ergonomics of the reactive BCI interfaces, or also for automatically stop the BCI system control when the mental workload became too high, or more in general to improve the human machine interaction. The innovation respect to the present literature is the concept to use the covert mental states of the user (e.g. mental workload) to act directly to the system, and improve its usability.

In the second study (section 5), it has been proposed a passive BCI system able to estimate online the mental workload of the user by using the EEG rhythms and the ECG signals. It was demonstrated that the system is able to significantly differentiate three workload levels related to three difficulty level tasks with a high reliability. In addition, another relevant aspect is that the classification features chosen by the system are stable after a week. This aspect is strictly required in the perspective of using such system in a real environment scenario. Mental states monitoring is of particular interest especially in safety-critical applications where human performance is often the least controllable factor. In this way, the proposed system could be useful during the operator's training to measure his cognitive workload and spare capacity while facing specific operative and emergency conditions. The innovation respect to the present literature is the possibility to predict online the mental workload of the user, using the combination of several biosignals (EEG and ECG), that allows to improve the reliability of the estimated mental states with respect to using just one information.

In conclusion, an accurate analysis of the human mental states using the neurophysiological information can be employed to optimize the mental states dependent man–machine interaction, and this thesis allowed demonstrating the powerful of using the passive BCI applications in two different scenarios (reactive BCI for communication and control and mental states evaluation in the operative environments). Future improvements have to be performed for making these systems usable in real contexts, such as ease of use, minimal calibration of the system, general usability, wearability and reliability.

7 REFERENCES

Aasman, J., Mulder, G., Mulder, L.J., 1987. Operator effort and the measurement of heart-rate variability. Hum Factors 29, 161–170.

Allison, B.Z., Pineda, J.A., 2006. Effects of SOA and flash pattern manipulations on ERPs, performance, and preference: Implications for a BCI system. International Journal of Psychophysiology 59, 127–140.

Aloise, F., Aricò, P., Schettini, F., Riccio, A., Salinari, S., Mattia, D., Babiloni, F., Cincotti, F., 2012. A covert attention P300-based brain–computer interface: Geospell. Ergonomics 1–14.

Aloise, F., Ferriero, D., Ruiu, A., Santucci, G., Catarci, T., Mattia,, D., Babiloni, F., Cincotti, F., 2009. Controlling domotic appliances via a "dynamical" P300-based Brain Computer Interface. Presented at the AAATE '09, Florence, Italy.

Aloise, F., Schettini, F., Aricò, P., Leotta, F., Salinari, S., Mattia, D., Babiloni, F., Cincotti, F., 2011. P300-based Brain Computer Interface for environmental control: an asynchronous approach. J Neural Eng 8, 016001.

Aloise, F., Schettini, F., Aricò, P., Salinari, S., Babiloni, F., Cincotti, F., 2012. A comparison of classification techniques for a gaze-independent P300-based brain-computer interface. Journal of neural engineering 9, 045012.

Aloise, F., Schettini, F., Aricò, P., Salinari, S., Guger, C., Rinsma, J., Aiello, M., Mattia, D., Cincotti, F., 2011. Asynchronous P300-based BCI to control a virtual apartment: initial tests on patients. Clinical Electroencephalography and Neuroscience.

Aricò, P., Aloise, F., Schettini, F., Salinari, S., Mattia, D., Cincotti, F., 2014. Evaluation of the Latency Jitter of P300 Evoked Potentials during C(o)vert Attention BCI. J Neural Eng In press.

Aricò, P., Borghini, G., Graziani, I., Bianchini, F., Cincotti, F., Babiloni, F., 2013. A brain computer interface system for the online evaluation of ATCs' workload. Italian Journal Of Aerospace Medicine.

Backs, R.W., Seljos, K.A., 1994. Metabolic and cardiorespiratory measures of mental effort: the effects of level of difficulty in a working memory task. Int J Psychophysiol 16, 57–68.

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.

Birbaumer, N., 1997. Slow cortical potentials: their origin, meaning, and clinical use. Brain and Behavior Past, Present, and Future 25–39.

Birbaumer, N., 1999. Neurobiology: Rain Man's revelations. Nature 399, 211–212.

Birbaumer, N., Kubler, A., Ghanayim, N., Hinterberger, T., Perelmouter, J., Kaiser, J., Iversen, I., Kotchoubey, B., Neumann, N., Flor, H., 2000. The thought translation device (TTD) for completely paralyzed patients. IEEE Transactions on Rehabilitation Engineering 8, 190–193.

Birbaumer, N., Schmidt, R.., 1996. Biologische Psychologie(3rd ed.), in: Heidelberg: Springer-Verlag.

Bishop, P.O., Davis, R., 1960. Synaptic potentials, after-potentials, and slow rhythms of lateral geniculate neurones. J Physiol 154, 514–546.

Blankertz, B., Tangermann, M., Vidaurre, C., Fazli, S., Sannelli, C., Haufe, S., Maeder, C., Ramsey, L., Sturm, I., Curio, G., Müller, K.-R., 2010. The Berlin Brain-Computer Interface: Non-Medical Uses of BCI Technology. Front Neurosci 4, 198.

Borghini, G., Astolfi, L., Vecchiato, G., Mattia, D., Babiloni, F., 2012. Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. Neurosci Biobehav Rev.

Breitwieser, C., Daly, I., Neuper, C., Müller-Putz, G.R., 2012. Proposing a standardized protocol for raw biosignal transmission. IEEE Trans Biomed Eng 59, 852–859.

Brookhuis, K.A., de Waard, D., 1993. The use of psychophysiology to assess driver status. Ergonomics 36, 1099–1110.

Brunner, P., Joshi, S., Briskin, S., Wolpaw, J.R., Bischof, H., Schalk, G., 2010. Does the "P300" speller depend on eye gaze? J Neural Eng 7, 056013.

Buzsáki, G., 2002. Theta oscillations in the hippocampus. Neuron 33, 325–340.

Cain, B., 2007. A Review of the Mental Workload Literature. Presented at the NATO Science and Technology Organization.

Cannon, M.W., Jr, 1985. Perceived contrast in the fovea and periphery. J Opt Soc Am A 2, 1760–1768.

Cantero, J.L., Atienza, M., Stickgold, R., Kahana, M.J., Madsen, J.R., Kocsis, B., 2003. Sleep-dependent theta oscillations in the human hippocampus and neocortex. J. Neurosci. 23, 10897–10903.

Carrasco, M., Evert, D.L., Chang, I., Katz, S.M., 1995. The eccentricity effect: target eccentricity affects performance on conjunction searches. Percept Psychophys 57, 1241–1261.

Casali, J.G., Wierwille, W.W., 1984. On the measurement of pilot perceptual workload: a comparison of assessment techniques addressing sensitivity and intrusion issues. Ergonomics 27, 1033–1050.

Chennu, S., Craston, P., Wyble, B., Bowman, H., 2009. Attention Increases the Temporal Precision of Conscious Perception: Verifying the Neural-ST2 Model. PLoS Comput Biol 5, e1000576.

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.

Colle, H.A., Reid, G.B., 1999. Double trade-off curves with different cognitive processing combinations: testing the cancellation axiom of mental workload measurement theory. Hum Factors 41, 35–50.

Comstock, J.R., 1994. MATB - Multi-Attribute Task Battery for human operator workload and strategic behavior research.

Cooper, N.R., Burgess, A.P., Croft, R.J., Gruzelier, J.H., 2006. Investigating evoked and induced electroencephalogram activity in task-related alpha power increases during an internally directed attention task. Neuroreport 17, 205–208.

Courchesne, E., 1977. Event-related brain potentials: comparison between children and adults. Science 197, 589–592.

Crites, S.L., Jr, Cacioppo, J.T., Gardner, W.L., Berntson, G.G., 1995. Bioelectrical echoes from evaluative categorization: II. A late positive brain potential that varies as a function of attitude registration rather than attitude report. J Pers Soc Psychol 68, 997–1013.

De Haan, B., Morgan, P.S., Rorden, C., 2008. Covert orienting of attention and overt eye movements activate identical brain regions. Brain Res 1204, 102–111.

Dehaene, S., Spelke, E., Pinel, P., Stanescu, R., Tsivkin, S., 1999. Sources of mathematical thinking: behavioral and brain-imaging evidence. Science 284, 970–974.

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.

Ding, J., Sperling, G., Srinivasan, R., 2006. Attentional modulation of SSVEP power depends on the network tagged by the flicker frequency. Cereb Cortex 16, 1016–1029.

Donchin, E., Smith, D.B., 1970. The contingent negative variation and the late positive wave of the average evoked potential. Electroencephalogr Clin Neurophysiol 29, 201–203.

Dooley, C., 2009. The Impact of Meditative Practices on Physiology and Neurology: A Review of the Literature, in: Scientia Discipulorum. pp. 35–59.

Duncan-Johnson, C.C., Donchin, E., 1977. On quantifying surprise: the variation of event-related potentials with subjective probability. Psychophysiology 14, 456–467.

Eggemeier, F.T., Stadler, M.A., 1984. Subjective Workload Assessment in a Spatial Memory Task. Proceedings of the Human Factors and Ergonomics Society Annual Meeting 28, 680–684.

Elbert, T., Rockstroh, B., Lutzenberger, W., Birbaumer, N., 1980. Biofeedback of slow cortical potentials. I. Electroencephalography and Clinical Neurophysiology 48, 293–301.

Eulitz, C., Maess, B., Pantev, C., Friederici, A.D., Feige, B., Elbert, T., 1996. Oscillatory neuromagnetic activity induced by language and non-language stimuli. Brain Res Cogn Brain Res 4, 121–132.

Fabiani, M., Gratton, G., Karis, D., Donchin, E., 1987. Definition, identification and reliability of measurement of the P300 component of the event-related brain potential. Advances in Psychophysiology 2, 1–78.

Fairclough, S.H., Venables, L., Tattersall, A., 2005. The influence of task demand and learning on the psychophysiological response. Int J Psychophysiol 56, 171–184.

Falkenburger, B.H., Barstow, K.L., Mintz, I.M., 2001. Dendrodendritic Inhibition Through Reversal of Dopamine Transport. Science 293, 2465–2470.

Farwell, L.A., Donchin, E., 1988. Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. Electroencephalogr Clin Neurophysiol 70, 510–523.

Fjell, A.M., Rosquist, H., Walhovd, K.B., 2009. Instability in the latency of P3a/P3b brain potentials and cognitive function in aging. Neurobiol. Aging 30, 2065–2079.

Fowler, B., 1994. P300 as a measure of workload during a simulated aircraft landing task. Hum Factors 36, 670–683.

Furdea, A., Halder, S., Krusienski, D.J., Bross, D., Nijboer, F., Birbaumer, N., Kübler, A., 2009. An auditory oddball (P300) spelling system for brain-computer interfaces. Psychophysiology 46, 617–625.

Furedy, J.J., 1987. Beyond heart rate in the cardiac psychophysiological assessment of mental effort: the T-wave amplitude component of the electrocardiogram. Hum Factors 29, 183–194.

G. E. Cooper, H., 1969. The use of pilot rating in the evaluation of aircraft handling qualities.

Galloway, N.R., 1990. Human Brain Electrophysiology: Evoked Potentials and Evoked Magnetic Fields in Science and Medicine. Br J Ophthalmol 74, 255.

Gerstner, W., Kistler, W.M., 2002. Spiking neuron models single neurons, populations, plasticity. Cambridge University Press, Cambridge, U.K.; New York.

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 40, 79–91.

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

Gollo, L.L., Kinouchi, O., Copelli, M., 2012. Statistical physics approach to dendritic computation: The excitable-wave mean-field approximation. Phys. Rev. E 85, 011911.

Goodin, D.S., Squires, K.C., Henderson, B.H., Starr, A., 1978. Age-related variations in evoked potentials to auditory stimuli in normal human subjects. Electroencephalogr Clin Neurophysiol 44, 447–458.

Gopher, D., Donchin, E., 1986. Workload – An examination of the concept. Handbook of Perception and Human Performance, in: Cognitive Processes and Performance. K.R. Boff, L. Kaufman and J.P. Thomas, John Wiley and Sons, Inc: 41-1:41-49.

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

Gray, C.M., König, P., Engel, A.K., Singer, W., 1989. Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. Nature 338, 334–337.

Gray, H.M., Ambady, N., Lowenthal, W.T., Deldin, P., 2004. P300 as an index of attention to self-relevant stimuli. Journal of Experimental Social Psychology 40, 216–224.

Grimes, D., Tan, D.S., Hudson, S.E., Shenoy, P., Rao, R.P.N., 2008. Feasibility and Pragmatics of Classifying Working Memory Load with an Electroencephalograph, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '08. ACM, New York, NY, USA, pp. 835–844.

Grossman, P., 1992. Respiratory and cardiac rhythms as windows to central and autonomic biobehavioral regulation: selection of window frames, keeping the panes clean and viewing the neural topography. Biol Psychol 34, 131–161.

Gruzelier, J., Liddiard, D., Davis, L., Wilson, L., 1990. Topographical EEG differences between schizophrenic patients and controls during neuropsychological functional activation. Int J Psychophysiol 8, 275–282.

Gundel, A., Wilson, G.F., 1992. Topographical changes in the ongoing EEG related to the difficulty of mental tasks. Brain Topography 5, 17–25.

Hagemann, K., 2008. The alpha band as an electrophysiological indicator for internalized attention and high mental workload in real traffic driving. [WWW Document]. URL http://docserv.uni-duesseldorf.de/servlets/DocumentServlet?id=8318 (accessed 1.8.14).

Hancock, P.A., Desmond, P.A., 2001. Engineering Psychophysiology.

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.

Hebert, R., Lehmann, D., 1977. Theta bursts: an EEG pattern in normal subjects practising the transcendental meditation technique. Electroencephalogr Clin Neurophysiol 42, 397–405.

Heger, D., Putze, F., Schultz, T., 2010. Online Workload Recognition from EEG Data during Cognitive Tests and Human-Machine Interaction, in: Dillmann, R., Beyerer, J., Hanebeck, U.D., Schultz, T. (Eds.), KI 2010: Advances in Artificial Intelligence. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 410–417.

Holmes, G.L., 2002. Event-Related Desynchronization. Handbook of Electroencephalography and Clinical Neurophysiology, Revised Series, Volume 6. Epilepsy Research 49, 178.

Houlihan, M.E., Pritchard, W.S., Robinson, J.H., 1996. Faster P300 latency after smoking in visual but not auditory oddball tasks. Psychopharmacology (Berl.) 123, 231–238.

Hu, L., Mouraux, A., Hu, Y., Iannetti, G.D., 2010. A novel approach for enhancing the signal-to-noise ratio and detecting automatically event-related potentials (ERPs) in single trials. Neuroimage 50, 99–111.

Huey, B.M., Wickens, C.D., National Research Council (U.S.), Panel on Workload Transition, 1993. Workload transition implications for individual and team performance. National Academy Press, Washington, DC.

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

Jennings, J.R., Stringfellow, J.C., Graham, M., 1974. A Comparison of the Statistical Distributions of Beat-By-Beat Heart Rate and Heart Period. Psychophysiology 11, 207–210.

Jensen, O., Kaiser, J., Lachaux, J.-P., 2007. Human gamma-frequency oscillations associated with attention and memory. Trends Neurosci. 30, 317–324.

Jex, H.R., 1988. Measuring Mental Workload: Problems, Progress, and Promises, in: Peter A. Hancock and Najmedin Meshkati (Ed.), Advances in Psychology, Human Mental Workload. North-Holland, pp. 5–39.

Johnson, R., Jr, 1986. A triarchic model of P300 amplitude. Psychophysiology 23, 367–384.

Jorna, P.G., 1993. Heart rate and workload variations in actual and simulated flight. Ergonomics 36, 1043–1054.

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.

Kandel, E.R., Schwartz, J.H., Jessell, T.M., 2000. Principles of neural science. McGraw-Hill, Health Professions Division, New York.

Klimesch, W., Doppelmayr, M., Schwaiger, J., Auinger, P., Winkler, T., 1999. `Paradoxical' alpha synchronization in a memory task. Cognitive Brain Research 7, 493–501.

Klimesch, W., Sauseng, P., Hanslmayr, S., 2007. EEG alpha oscillations: the inhibition-timing hypothesis. Brain Res Rev 53, 63–88.

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.

Kotchoubey, B., 2006. Event-related potentials, cognition, and behavior: a biological approach. Neurosci Biobehav Rev 30, 42–65.

Kozelka, J.W., Pedley, T.A., 1990. Beta and mu rhythms. J Clin Neurophysiol 7, 191–207.

Kramer, A.E., 1990. Physiological Metrics of Mental Workload: A Review of Recent Progress.

Kramer, A.F., Parasuraman, R., 2007. Neuroergonomics: Applications of Neuroscience to Human Factors, in: Handbook of Psychophysiology. Cambridge University Press.

Kramer, A.F., Strayer, D.L., 1988. Assessing the development of automatic processing: an application of dual-task and event-related brain potential methodologies. Biol Psychol 26, 231–267.

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.

Krusienski, D.J., Sellers, E.W., McFarland, D.J., Vaughan, T.M., Wolpaw, J.R., 2008. Toward enhanced P300 speller performance. J. Neurosci. Methods 167, 15–21.

Kübler, A., Birbaumer, N., 2008. Brain-computer interfaces and communication in paralysis: extinction of goal directed thinking in completely paralysed patients? Clin Neurophysiol 119, 2658–2666.

Kübler, A., Neumann, N., Kaiser, J., Kotchoubey, B., Hinterberger, T., Birbaumer, N.P., 2001. Brain-computer communication: Self-regulation of slow cortical potentials for verbal communication. Archives of Physical Medicine and Rehabilitation 82, 1533–1539.

Kubota, Y., Sato, W., Toichi, M., Murai, T., Okada, T., Hayashi, A., Sengoku, A., 2001. Frontal midline theta rhythm is correlated with cardiac autonomic activities during the performance of an attention demanding meditation procedure. Brain Res Cogn Brain Res 11, 281–287.

Kutas, M., McCarthy, G., Donchin, E., 1977. Augmenting mental chronometry: the P300 as a measure of stimulus evaluation time. Science 197, 792–795.

Lawless, J.F., 1982. Statistical models and methods for lifetime data. Wiley.

Lee, D.H., Park, K.S., 1990. Multivariate analysis of mental and physical load components in sinus arrhythmia scores. Ergonomics 33, 35–47.

Lei, S., Roetting, M., 2011. Influence of Task Combination on EEG Spectrum Modulation for Driver Workload Estimation. Human Factors: The Journal of the Human Factors and Ergonomics Society 53, 168–179.

Lei, S., Welke, S., Roetting, M., 2009. Driver's Mental Workload Assessment Using EEG Data in a Dual Task Paradigm. Proceedings of the 21st (Esv) International Technical Conference On The Enhanced Safety Of Vehicles, Held June 2009, Stuttgart, Germany.

Leuthold, H., 1998. Postperceptual effects and P300 latency. Psychophysiology 35, 34–46.

Liu, Y., Zhou, Z., Hu, D., 2010. Gaze independent brain-computer speller with covert visual search tasks. Clin Neurophysiol.

Lopes da Silva, F., 1991. Neural mechanisms underlying brain waves: from neural membranes to networks. Electroencephalography and Clinical Neurophysiology 79, 81–93.

Lopes da Silva, F.H., Vos, J.E., Mooibroek, J., Van Rotterdam, A., 1980. Relative contributions of intracortical and thalamo-cortical processes in the generation of alpha rhythms, revealed by partial coherence analysis. Electroencephalogr Clin Neurophysiol 50, 449–456.

Magliero, A., Bashore, T.R., Coles, M.G., Donchin, E., 1984. On the dependence of P300 latency on stimulus evaluation processes. Psychophysiology 21, 171–186.

Majaranta, P., MacKenzie, S., Aula, A., Räihä, K.-J., 2006. Effects of feedback and dwell time on eye typing speed and accuracy. Univers. Access Inf. Soc. 5, 199–208.

Martin, L., Barajas, J.J., Fernandez, R., Torres, E., 1988. Auditory event-related potentials in well-characterized groups of children. Electroencephalogr Clin Neurophysiol 71, 375–381.

McCarthy, G., Donchin, E., 1981. A metric for thought: a comparison of P300 latency and reaction time. Science 211, 77–80.

McFarland, D.J., Miner, L.A., Vaughan, T.M., Wolpaw, J.R., 2000. Mu and beta rhythm topographies during motor imagery and actual movements. Brain Topogr 12, 177–186.

McFarland, D.J., Wolpaw, J.R., 2003. EEG-based communication and control: speed-accuracy relationships. Appl Psychophysiol Biofeedback 28, 217–231.

Mehler, B., Reimer, B., Coughlin, J.F., Dusek, J.A., 2009. Impact of Incremental Increases in Cognitive Workload on Physiological Arousal and Performance in Young Adult Drivers. Transportation Research Record: Journal of the Transportation Research Board 2138, 6–12.

Mogford, R.H., n.d. Application of Research Techniques for Documenting Cognitive Processes in Air Traffic Control: Sector Complexity and Decision Making. PN.

Müller, A., Schandry, R., Montoya, P., Gsellhofer, B., 1992. Differential effects of two stressors on heart rate, respiratory sinus arrhythmia, and T-wave amplitude. Journal of Psychophysiology 6, 252–259.

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.

Müller, M.M., Hillyard, S., 2000. Concurrent recording of steady-state and transient event-related potentials as indices of visual-spatial selective attention. Clin Neurophysiol 111, 1544–1552.

Nijboer, F., Birbaumer, N., Kübler, A., 2010. The influence of psychological state and motivation on brain-computer interface performance in patients with amyotrophic lateral sclerosis - a longitudinal study. Front Neurosci 4.

Nikolaev, A.R., Ivanitskiĭ, G.A., Ivanitskiĭ, A.M., 1998. [Reproducible EEG alpha-rhythm patterns in solving psychological tasks]. Fiziol Cheloveka 24, 5–12.

Noel, J.B., Bauer,Jr., K.W., Lanning, J.W., 2005. Improving Pilot Mental Workload Classification Through Feature Exploitation and Combination: A Feasibility Study. Comput. Oper. Res. 32, 2713–2730.

Nordwall, B.D., n.d. FAA programs advance, with caution flags flying.

O'Donnell F.T., R.D., Eggemeier, 1986. Workload assessment methodology. Handbook of Perception and Human Performance. John Wiley and Sons, Inc.

Olson, C.R., 2001. Object-based vision and attention in primates. Curr. Opin. Neurobiol. 11, 171–179.

Pawlak, W.S., Brinton, C.R., Crouch, K., Lancaster, K.M., 1996. A framework for the evaluation of air traffic control complexity. Presented at the AIAA national Conference.

Peterson, M.S., Kramer, A.F., Irwin, D.E., 2004. Covert shifts of attention precede involuntary eye movements. Percept Psychophys 66, 398–405.

Pichiorri, F., De Vico Fallani, F., Cincotti, F., Babiloni, F., Molinari, M., Kleih, S.C., Neuper, C., Kübler, A., Mattia, D., 2011. Sensorimotor rhythm-based brain-computer interface training: the impact on motor cortical responsiveness. J Neural Eng 8, 025020.

Picton, T.W., Bentin, S., Berg, P., Donchin, E., Hillyard, S.A., Johnson, R., Miller, G.A., Ritter, W., Ruchkin, D.S., Rugg, M.D., Taylor, M.J., 2000. Guidelines for using human event-related potentials to study cognition: recording standards and publication criteria. Psychophysiology 37, 127–152.

Pierce, J.R., 1980. An introduction to information theory. Dover Publications, New York.

Pigeau, R., Hoffmann, R., Purcell, S., Moffitt, A., 1988. The Effect of Endogenous Alpha on Hemispheric Asymmetries and the Relationship of Frontal Theta to Sustained Attention,.

Poirazi, P., Mel, B.W., 2001. Impact of active dendrites and structural plasticity on the memory capacity of neural tissue. Neuron 29, 779–796.

Polich, J., 1986. Attention, probability, and task demands as determinants of P300 latency from auditory stimuli. Electroencephalogr Clin Neurophysiol 63, 251–259.

Polich, J., 2007. Updating P300: an integrative theory of P3a and P3b. Clin Neurophysiol 118, 2128–2148.

Polich, J., Kok, A., 1995. Cognitive and biological determinants of P300: an integrative review. Biol Psychol 41, 103–146.

Polich, J., Ladish, C., Burns, T., 1990. Normal variation of P300 in children: age, memory span, and head size. Int J Psychophysiol 9, 237–248.

Polich, J., Margala, C., 1997. P300 and probability: comparison of oddball and single-stimulus paradigms. Int J Psychophysiol 25, 169–176.

Pope, A.T., Bogart, E.H., Bartolome, D.S., 1995. Biocybernetic system evaluates indices of operator engagement in automated task. Biological Psychology 40, 187–195.

Putze, F., Jarvis, J.-P., Schultz, T., 2010. Multimodal Recognition of Cognitive Workload for Multitasking in the Car, in: 2010 20th International Conference on Pattern Recognition (ICPR). Presented at the 2010 20th International Conference on Pattern Recognition (ICPR), pp. 3748–3751.

Raymond, J.E., Shapiro, K.L., Arnell, K.M., 1992. Temporary suppression of visual processing in an RSVP task: an attentional blink? J Exp Psychol Hum Percept Perform 18, 849–860.

Read, D.E., 1981. Solving deductive-reasoning problems after unilateral temporal lobectomy. Brain and Language 12, 116–127.

Regan, D., 1966. Some characteristics of average steady-state and transient responses evoked by modulated light. Electroencephalography and Clinical Neurophysiology 20, 238–248.

Reid, G.B., Nygren, T.E., 1988. The Subjective Workload Assessment Technique: A Scaling Procedure for Measuring Mental Workload, in: Peter A. Hancock and Najmedin Meshkati (Ed.), Advances in Psychology, Human Mental Workload. North-Holland, pp. 185–218.

Riccio, A., Leotta, F., Bianchi, L., Aloise, F., Zickler, C., Hoogerwerf, E.-J., Kübler, A., Mattia, D., Cincotti, F., 2011. Workload measurement in a communication application operated through a P300-based brain-computer interface. J Neural Eng 8, 025028.

Riccio, A., Mattia, D., Simione, L., Olivetti, M., Cincotti, F., 2012. Eye-gaze independent EEG-based brain-computer interfaces for communication. Journal of neural engineering 9, 045001.

Rijsdijk, J.P., Kroon, J.N., van der Wildt, G.J., 1980. Contrast sensitivity as a function of position on the retina. Vision Research 20, 235–241.

Rockstroh, B., Elbert, T., Birbaumer, N., Wolf, P., Düchting-Röth, A., Reker, M., Daum, I., Lutzenberger, W., Dichgans, J., 1993. Cortical self-regulation in patients with epilepsies. Epilepsy Res 14, 63–72.

Roscoe, A.H., 1992. Assessing pilot workload. Why measure heart rate, HRV and respiration? Biol Psychol 34, 259–287.

Roscoe, A.H., Ellis, G.A., 1990. A Subjective Rating Scale for Assessing Pilot Workload in Flight: A decade of Practical Use.

Rovamo, J., Virsu, V., 1979. An estimation and application of the human cortical magnification factor. Experimental Brain Research 37.

Sammer, G., Blecker, C., Gebhardt, H., Bischoff, M., Stark, R., Morgen, K., Vaitl, D., 2007. Relationship between regional hemodynamic activity and simultaneously recorded EEG-theta associated with mental arithmetic-induced workload. Hum Brain Mapp 28, 793–803.

San Agustin, J., Skovsgaard, H., Mollenbach, E., Barret, M., Tall, M., Hansen, D.W., Hansen, J.P., 2010. Evaluation of a low-cost open-source gaze tracker, in: Proceedings of the 2010

Symposium on Eye-Tracking Research & Applications, ETRA '10. ACM, New York, NY, USA, pp. 77–80.

Schalk, G., McFarland, D.J., Hinterberger, T., Birbaumer, N., Wolpaw, J.R., 2004. BCI2000: a general-purpose brain-computer interface (BCI) system. IEEE Trans Biomed Eng 51, 1034–43.

Scholl, B.J., 2001. Objects and attention: the state of the art. Cognition 80, 1–46.

Schubert, R., 2008. 14th World Congress of Psychophysiology The Olympics of the Brain 2008. Abstracts.

Schultheis, H., Jameson, A., 2004. Assessing Cognitive Load in Adaptive Hypermedia Systems: Physiological and Behavioral Methods, in: Bra, P.M.E., Nejdl, W. (Eds.), Adaptive Hypermedia and Adaptive Web-Based Systems. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 225–234.

Sellers, E.W., Kübler, A., Donchin, E., 2006. Brain-computer interface research at the University of South Florida Cognitive Psychophysiology Laboratory: the P300 Speller. IEEE Trans Neural Syst Rehabil Eng 14, 221–224.

Shepherd, G.M., 2003. The Synaptic organization of the brain. Oxford University Press, New York.

Sirevaag, E.J., Kramer, A.F., Coles, M.G., Donchin, E., 1989. Resource reciprocity: an event-related brain potentials analysis. Acta Psychol (Amst) 70, 77–97.

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

Sommer, W., Leuthold, H., Soetens, E., 1999. Covert signs of expectancy in serial reaction time tasks revealed by event-related potentials. Percept Psychophys 61, 342–353.

Sterman, M.B., Schummer, G.J., Dushenko, T.W., Smith, J.C., 1988. Electroencephalographic Correlates of Pilot Performance: Simulation and In-Flight Studies.

Stuart, G., Spruston, N., Häusser, M., 2008. Dendrites. Oxford University Press, Oxford.

Sutter, E., 1992. The brain response interface: communication through visually-induced electrical brain responses. J. Microcomput. Appl. 15, 31–45.

Sutton, S., Braren, M., Zubin, J., John, E.R., 1965. Evoked-potential correlates of stimulus uncertainty. Science 150, 1187–1188.

Suzuki, S., Cavanagh, P., 1997. Focused attention distorts visual space: An attentional repulsion effect. Journal of Experimental Psychology: Human Perception and Performance 23, 443–463.

Tallon-Baudry, C., Bertrand, O., Hénaff, M.-A., Isnard, J., Fischer, C., 2005. Attention modulates gamma-band oscillations differently in the human lateral occipital cortex and fusiform gyrus. Cereb. Cortex 15, 654–662.

Teplan, M., 2002. Fundamentals of EEG measurement [WWW Document]. Scribd. URL http://www.scribd.com/doc/47884335/FUNDAMENTALS-OF-EEG-MEASUREMENT (accessed 1.8.14).

Thompson, D.E., Warschausky, S., Huggins, J.E., 2013. Classifier-based latency estimation: a novel way to estimate and predict BCI accuracy. J Neural Eng 10, 016006.

Treder, M.S., Blankertz, B., 2010. (C)overt attention and visual speller design in an ERP-based brain-computer interface. Behav Brain Funct 6, 28.

Treder, M.S., Schmidt, N.M., Blankertz, B., 2011. Gaze-independent brain-computer interfaces based on covert attention and feature attention. Journal of Neural Engineering 8, 066003.

Treder, M., Sebastian, Schmidt, N., Blankertz, B., 2010. Gaze-independent visual brain-computer interfaces. Presented at the Translational Issues in BCI Development: User Needs, Ethics, and Technology Transfer, Roma, Italy.

Treue, S., 2001. Neural correlates of attention in primate visual cortex. Trends Neurosci. 24, 295–300.

Tsang, P.S., Johnson, W.W., 1989. Cognitive demands in automation. Aviat Space Environ Med 60, 130–135.

Veltman, J.A., Gaillard, A.W., 1998. Physiological workload reactions to increasing levels of task difficulty. Ergonomics 41, 656–669.

Veltman, J.A., Gaillard, A.W.K., 1996. Physiological indices of workload in a simulated flight task. Biological Psychology 42, 323–342.

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). Presented at the 2010 IEEE International Conference on Systems Man and Cybernetics (SMC), pp. 417–421.

Walter, W.G., Cooper, R., Aldridge, V.J., Mccallum, W.C., Winter, A.L., 1964. Contingent Negative Variation: An Electric Sign Of Sensorimotor Association And Expectancy In The Human Brain. Nature 203, 380–384.

Welke, S., Jurgensohn, T., Roetting, M., 2009. Single-Trial Detection Of Cognitive Processes For Increasing Traffic Safety. Proceedings Of The 21st (Esv) International Technical Conference On The Enhanced Safety Of Vehicles, Held June 2009, Stuttgart, Germany.

Wickens, C., Kramer, A., Vanasse, L., Donchin, E., 1983. Performance of concurrent tasks: a psychophysiological analysis of the reciprocity of information-processing resources. Science 221, 1080–1082.

Wierwille, W.W., 1988. Important Remaining Issues in Mental Workload Estimation, in: Peter A. Hancock and Najmedin Meshkati (Ed.), Advances in Psychology, Human Mental Workload. North-Holland, pp. 315–327.

Wilson, G.F., 1992. Applied use of cardiac and respiration measures: practical considerations and precautions. Biol Psychol 34, 163–178.

Wilson, G.F., 2002. An analysis of mental workload in pilots during flight using multiple psychophysiological measures. The International Journal of Aviation Psychology 12, 3–18.

Wilson, G.F., Fisher, F., 1995. Cognitive task classification based upon topographic EEG data. Biological Psychology 40, 239–250.

Wilson, G.F., Russell, C.A., 2003a. Operator functional state classification using multiple psychophysiological features in an air traffic control task. Hum Factors 45, 381–389.

Wilson, G.F., Russell, C.A., 2003b. Real-time assessment of mental workload using psychophysiological measures and artificial neural networks. Hum Factors 45, 635–643.

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

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

Xie, B., Salvendy, G., 2000. Prediction of Mental Workload in Single and Multiple Tasks Environments. International Journal of Cognitive Ergonomics 4, 213–242.

Yagi, Y., Coburn, K.L., Estes, K.M., Arruda, J.E., 1999. Effects of aerobic exercise and gender on visual and auditory P300, reaction time, and accuracy. Eur J Appl Physiol Occup Physiol 80, 402–408.

Zander, T.O., Jatzev, S., 2012. Context-aware brain–computer interfaces: exploring the information space of user, technical system and environment. Journal of Neural Engineering 9, 016003.

Zander, T.O., Kothe, C., 2011. Towards passive brain-computer interfaces: applying brain-computer interface technology to human-machine systems in general. J Neural Eng 8, 025005.

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.

Zickler, C., Riccio, A., Leotta, F., Hillian-Tress, S., Halder, S., Holz, E., Staiger-Sälzer, P., Hoogerwerf, E.-J., Desideri, L., Mattia, D., Kübler, A., 2011. A brain-computer interface as input channel for a standard assistive technology software. Clin EEG Neurosci 42, 236–244.

8 SCIENTIFIC WRITING

8.1 Full Papers

[J 1] **P.Aricò**, F.Aloise, F.Schettini, S.Salinari, D.Mattia, F.Cincotti. "Influence of P300 latency jitter over ERPs based BCIs performance". Accepted for publication on *J Neural Eng*.

[J 2] A. Riccio, E. Holz, **P. Aricò**, F. Leotta, F. Aloise, L. Desideri, M. Rimondini, A. Kübler, D. Mattia, F. Cincotti. "A hybrid control of a P300-based Brain Computer Interface to improve usability for people with severe motor disability" Accepted for publication on *Archives of Physical Medicine and Rehabilitation*.

[J 3] F. Schettini, F. Aloise, **P.Aricò**, S. Salinari, D. Mattia, F. Cincotti. "Self-calibration algorithm in an asynchronous P300-based". Accepted for publication on *J Neural Eng*.

[J 4] F. Aloise, **P. Aricò**, F. Schettini, S. Salinari, D. Mattia, F. Cincotti. "Asynchronous gaze-independent event-related potential-based brain–computer interface". *Artif. Intell. Med*, vol. 59, no. 2, pp. 61–69, Oct. 2013.

[J 5] **P. Aricò**, G. Borghini, I. Graziani, F. Bianchini, F. Cincotti, F. Babiloni. "A brain computer interface system for the online evaluation of ATCs' workload". Accepted for publication on *Italian Journal Of Aerospace Medicine*.

[J 6] F. Aloise, F. Schettini, **P. Aricò**, S. Salinari, F. Babiloni and F. Cincotti. "A comparison of classification techniques for a gaze-independent P300-based brain-computer interface". *J Neural Eng*, vol. 9, n. 4, pag. 045012 (9pp), Aug. 2012.

[J 7] F. Aloise, **P. Aricò**, F. Schettini, A. Riccio, S. Salinari, D. Mattia, F. Babiloni, F. Cincotti. "A Covert Attention P300-based Brain Computer Interface: GeoSpell". *Ergonomics*, vol. 55, n°. 5, pagg. 538–551, May 2012.

[J 8] F. Aloise, F. Schettini, **P. Aricò**, S. Salinari, C. Guger, J. Rinsma, M. Aiello, D. Mattia, F. Cincotti. "Asynchronous P300-based BCI to control a virtual environment: initial tests on end users". *Clin EEG Neurosci*, vol. 42, n°. 4, pagg. 219–224, Oct 2011.

[J 9] F.Aloise, F.Schettini, **P.Aricò**, F.Leotta, S.Salinari, D.Mattia, F.Babiloni, F.Cincotti. "P300-based brain-computer interface for environmental control: an asynchronous approach". *J Neural Eng*. 8(2):025025 (10pp), Apr 2011.

[J 10] **P. Aricò**, F. Aloise, F. Schettini, A. Riccio, S.Salinari, F. Babiloni, D. Mattia, F. Cincotti. "GeoSpell: an alternative P300-based speller interface towards no eye gaze required". *International Journal of Bioelectromagnetism*, Vol. 13, No. 3, pp. 152 – 153, 2011.

[J 11] F. Aloise, **P. Aricò**, F. Schettini, E. Lucano, S.Salinari, F.Babiloni, D. Mattia, F. Cincotti. "Can the P300-based BCI training affect the ERPs?". I*nternational Journal of Bioelectromagnetism*, Vol. 13, No. 3, pp. 148 – 149, 2011.

[J 12] F. Schettini, F. Aloise, **P. Aricò**, F. Leotta, S. Salinari, F. Babiloni, D. Mattia, F. Cincotti. "Improving Asynchronous Control for P300-based BCI: towards a completely autoadaptative system". *International Journal of Bioelectromagnetism*, Vol. 13, No. 3, pp. 150 – 151, 2011.

8.2 Conference proceedings

[C 1] G. Borghini, **P. Aricò**, F. Babiloni, G. Granger, J-P., Imbert, R. Benhacene, L. Napoletano, S. Pozzi. "NINA: Neurometrics Indicators for ATM". The Third SESAR Innovation Days, 26th-28th November 2013, Stockholm, Sweden.

[C 2] G. Borghini, **P. Aricò**, L. Astolfi, J. Toppi, F. Cincotti, D. Mattia, G. Vecchiato, A. G. Maglione, I. Graziani, F. Babiloni. "Frontal EEG theta changes assess the training improvements of novices in a flight simulation tasks" Conf Proc IEEE Eng Med Biol Soc. 2013 Jul; 2013, Osaka, Japan.

[C 3] **P. Aricò**, F. Aloise, F. Schettini, S. Salinari, D. Mattia, F. Cincotti. "Assessment of the P300 evoked potentials latency stability during c(o)vert attention BCI". BCI Meeting 2013 Fifth International Meeting Asilomar, California June 2-3, 2013.

[C 4] F. Schettini, F. Aloise, **P. Aricò**, S. Salinari, D. Mattia, F. Cincotti. "Self-Calibration in an asynchronous P300-based BCI". BCI Meeting 2013 Fifth International Meeting Asilomar, California June 2-3, 2013.

[C 5] A. Riccio, E. Holz, **P. Aricò**, F. Leotta, F. Aloise, L. Desideri, M. Rimondini, A. Kübler, D. Mattia, F. Cincotti. "A hybrid control of a P300-based BCI: a solution to improve system usability?". BCI Meeting 2013 Fifth International Meeting Asilomar, California June 2-3, 2013.

[C 6] I. Daly, F. Aloise, **P. Aricò**, J. Belda, M. Billinger, E. Bolinger, F. Cincotti, D. Hettich, M. Iosa, J. Laparra, R. Scherer, G. Müller-Putz. "Rapid prototyping for hBCI users with Cerebral palsy". BCI Meeting 2013 Fifth International Meeting Asilomar, California June 2-3, 2013.

[C 7] **P. Aricò**, F. Aloise, F. Schettini, S. Salinari, Donatella Mattia, Febo Cincotti. "Evaluation of the Latency Jitter of P300 Evoked Potentials during C(o)vert Attention BCI". 4th Workshop of the TOBI Project: Practical Brain-Computer Interfaces for End-Users: Progress and Challenges. Sion, Switzerland, January 23-25, 2013.

[C 8] **P. Aricò**, F. Aloise, F. Pichiorri, G. Morone, F. Tamburella, S. Salinari, M. Molinari, D. Mattia, F. Cincotti. "Automated Assessment of Pathologic EMG Synergies for BCI-based Neuro-rehabilitation after Stroke". 4th Workshop of the TOBI Project: Practical Brain-Computer Interfaces for End-Users: Progress and Challenges. Sion, Switzerland, January 23-25, 2013.

[C 9] A. Riccio, E. Holtz, **P. Aricò**, F. Leotta, F. Aloise, L. Desideri, E-J. Hoogerwerf, A. Kubler, D.Mattia, F. Cincotti. "Towards a Hybrid Control of a P300-based BCI for Communication in Severely Disabled End-Users". 4th Workshop of the TOBI Project: Practical Brain-Computer Interfaces for End-Users: Progress and Challenges. Sion, Switzerland, January 23-25, 2013.

[C 10] E.M. Holz, A.Riccio, J. Reichert, F. Leotta, **P. Aricò**, F. Cincotti, D. Mattia, A. Kübler. "Hybrid-P300 BCI: Usability Testing by Severely Motor-restricted End-Users". 4th Workshop of the TOBI Project: Practical Brain-Computer Interfaces for End-Users: Progress and Challenges. Sion, Switzerland, January 23-25, 2013.

[C 11] F. Cincotti, F. Pichiorri, **P. Aricò**, F. Aloise, F. Leotta, F. De Vico Fallani, J. del R. Millán, M. Molinari, D. Mattia. "EEG-based Brain-Computer Interface to support post-stroke motor rehabilitation of the upper limb". EMBC 2012, 28th August 1st September, 2012, San Diego, USA.

[C 12] F. Schettini, F. Aloise, **P. Aricò**, S. Salinari, D. Mattia and F. Cincotti. "Control or No-Control? Reducing the gap between Brain-Computer Interface and classical input devices". EMBC 2012, 28th August 1st September, 2012, San Diego, USA.

[C 13] **P. Aricò**, F. Aloise, C. Giovannella. "ERP approach: what could we learn from?".Advanced Learning Technologies (ICALT), 2012 IEEE 12th International Conference, 4-6 July 2012.

[C 14] **P. Aricò**, F. Aloise, F. Pichiorri, F. Leotta, S. Serenella, D. Mattia, F. Cincotti. "FES controlled by a hybrid BCI system for neurorehabilitation – driven after stroke". GNB2012, June 26th-29th 2012, Rome, Italy. ISBN: 978 88 555 3182-5.

[C 15] F. Aloise, **P. Aricò**, F. Schettini, M. Iosa, M. Scarnicchia, S. Salinari, D. Morelli, D. Mattia, F. Cincotti. "The Brain Computer Interface as augmentative and alternative communication aid: the ABC project". GNB2012, June 26th-29th 2012, Rome, Italy. ISBN: 978 88 555 3182-5.

[C 16] **P. Aricò**, F. Aloise, F. Schettini, S. Salinari, D. Mattia, F. Cincotti. "On the correlation between Brain Computer Interface performance and chronotype". GNB2012, June 26th-29th 2012, Rome, Italy. ISBN: 978 88.

[C 17] F. Schettini, F. Aloise, **P. Aricò**, S. Salinari, D. Mattia, F. Cincotti. "Improving Communication Efficiency for gaze independent P300 based Brain Computer Interface". GNB2012, June 26th-29th 2012, Rome, Italy. ISBN: 978 88.

[C 18] M. Iosa, F. Aloise, F. Schettini, **P. Aricò**, S. Paolucci, D. Morelli, D. Mattia, F. Cincotti. "Uso dei sistemi di Brain-Neural-Computer Interface nella comunicazione aumentata: il progetto ABC". XII SIRN 2012 National Congress, 3-5 May, Milan, Italy.

[C 19] **P. Aricò**, F. Aloise, F. Schettini, V. Soragnese, S. Salinari, D. Mattia, F. Cincotti. "Variability of ERPs – based Brain Computer Interface performance across repeated sessions in a day". 3rd Workshop of the TOBI Project: Bringing BCIs to End-Users: Facing the Challenge, Evaluation, User Perspectives, User Needs, and Ethical Questions. Würzburg, Germany Mar 20th-22th, 2012.

[C 20] F. Pichiorri, **P. Aricò**, F. Leotta, F. Aloise, F. Cincotti, M. Secci, M. Petti, D. Mattia. "Neurorehabilitation-driven design of hybrid BCI-controlled FES for motoe recovery after stroke". 3rd Workshop of the TOBI Project: Bringing BCIs to End-Users: Facing the Challenge, Evaluation, User Perspectives, User Needs, and Ethical Questions. Würzburg, Germany Mar 20th-22th, 2012.

[C 21] F. Schettini, F. Aloise, **P. Aricò**, S. Salinari, S. Petrichella, D. Mattia and F. Cincotti. "Comparing efficiency for Synchronous and Asynchronous P300-based BCIs". *5th International BCI Conference*, Graz, Austria September 23 – 24, 2011.

[C 22] F. Aloise, F. Schettini, **P. Aricò**, S. Salinari, C. Guger, J. Rinsma, M. Aiello, D. Mattia and F. Cincotti. "Validation of an asynchronous P300-based BCI with potential end users to control a virtual environment". *5th International BCI Conference*, Graz, Austria September 23 – 24, 2011.

[C 23] **P. Aricò**, F. Aloise, F. Schettini, S. Salinari, S. Santostasi, D. Mattia and F. Cincotti. "On the effect of ERPs-based BCI practice on user's performance". *5th International BCI Conference*, Graz, Austria September 23 – 24, 2011.

[C 24] F.Aloise, **P. Aricò**, F. Schettini, A. Riccio, M. Risetti, S. Salinari, D. Mattia, F. Babiloni, F. Cincotti."A new P300 No Eye-gaze based interface: GeoSpell". International Conference on Bio-Inspired Systems and Signal Processing: BioSignal 2011. January 26-29, 2011.

[C 25] F. Aloise, F. Schettini, **P. Aricò**, F. Leotta, S. Salinari, D. Mattia, F. Babiloni, F. Cincotti. "Towards Domotic appliances control through a self-paced P300-based BCI". International Conference on Bio-Inspired Systems and Signal Processing: BioSignal 2011. January 26-29, 2011.

[C 26] F. Aloise, F. Schettini, **P. Aricò**, L. Bianchi, A. Riccio, M. Mecella, F. Babiloni, D. Mattia, F. Cincotti. "Advanced Brain computer interface for communication and control" International Working Conference on advanced visual interfaces: AVI 2010, Roma 25-29 May, 2010.

[C 27] F. Schettini, F. Aloise, **P. Aricò**, F. Leotta, S. Salinari, F. Babiloni, D. Mattia, F. Cincotti. "Improving Asynchronous Control for P300-based BCI: towards a completely autoadaptative system". 2nd Workshop of the TOBI Project: Translational issues in BCI development: user needs, ethics, and technology transfer. December 2nd-3rd; 2010 Rome, Italy.

[C 28] **P. Aricò**, F. Aloise, F. Schettini, A. Riccio, S. Salinari, F. Babiloni, D. Mattia, F. Cincotti. "GeoSpell: an alternative P300-based speller interface towards no eye gaze required". 2nd Workshop of the TOBI Project: Translational issues in BCI development: user needs, ethics, and technology transfer. December 2nd-3rd; 2010 Rome, Italy.

[C 29] F. Aloise, **P. Aricò**, F. Schettini, E. Lucano, S. Salinari, F. Babiloni, D. Mattia, F. Cincotti. "Can the P300-based BCI training affect the ERPs?". 2nd Workshop of the TOBI Project Translational issues in BCI development: user needs, ethics, and technology transfer. December 2nd-3rd; 2010 Rome, Italy.