



HAL
open science

Cognitive brain-computer interfaces : From feature engineering to neurophenomenological validation

Aldo Mora Sánchez

► **To cite this version:**

Aldo Mora Sánchez. Cognitive brain-computer interfaces : From feature engineering to neurophenomenological validation. *Neurons and Cognition [q-bio.NC]*. Sorbonne Université, 2018. English. NNT : 2018SORUS217 . tel-02489785

HAL Id: tel-02489785

<https://theses.hal.science/tel-02489785>

Submitted on 24 Feb 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Sorbonne Université

École doctorale EDITE

Brain plasticity unit / Brain-computer interfaces team

Cognitive brain-computer interfaces

*From feature engineering to neurophenomenological
validation*

By Aldo Mora Sánchez

Doctoral thesis of Neuroengineering

Supervised by François Vialatte

Presented and defended publicly on November 19, 2018

Members of the committee:

Dr. Maureen Clerc
Dr. Jordi Solé-Casals
Dr. Fabien Lotte
Dr. Patrick Gallinary
Dr. François Vialatte
Dr. Gérard Dreyfus

DEDICATION AND ACKNOWLEDGEMENTS

Many wonderful people helped me through this journey. My parents and siblings always encouraged me to follow my interests without much second thoughts. My Latinos in Paris brought a piece of home with them, helping as much as my French friends with a smooth transition. I especially want to thank Sarah for her constant support through thick and thin, my flatmates Arturo and Martín for being my de facto brothers, and Isui, for all the intellectual and real-life adventures. In general, special thanks to all the friends who, besides making this stage of my life memorable, shared their time and brainwaves during my endless experiments. A piece of each of them takes part of all these numbers and figures.

Academically, I am very grateful to my research group. To François for always keeping my interest alive, and for giving me the freedom to follow my curiosity. To Gérard, for patiently sharing his expertise on domains that were obscure to me. To Antoine, for helping me land on both feet on a field that was entirely new to me. To Takfarinas, for implementing the toolbox and for being at the lab always with a smile. To Aïda, for sharing her points of view about the mind and life and the universe and everything. And last but not least, to Alfredo, for making the online system actually work, and for cleaning our countless amount of data.

Finally, I would like to thank the Consejo Nacional de Ciencia y Tecnología and my research unit for the funding that allowed me to perform this work.

ABSTRACT

This thesis aims at describing in detail the design, implementation and validation of cognitive brain-computer interfaces (BCI). This work comprises three chapters. Chapter 1 introduces the general concepts of cognitive BCI design and brain metastability. In chapter 2, a specific cognitive function (Working Memory) is selected for the construction of a cognitive BCI. In chapter 3, we explore the possibility of using spatio temporal properties of brain dynamics as biomarkers for cognitive BCIs, and we address at the same time scientific questions concerning cognition-driven brain metastability.

The goal of the BCI built in chapter 2 is to continuously monitor Working Memory (WM) load in real-time. The BCI relies on spectral properties of electroencephalography (EEG) as biomarkers. The usefulness of estimating WM in real time ranges from learning applications to security in industrial environments. There are several studies in the literature aiming at estimating the WM load in real time. However, to our knowledge, this represents the first research in which different key elements are included simultaneously in a study concerning WM load estimation. The BCI was successfully tested on a cross-task. A cross-task addresses generalizability, or whether the device could be used in flexible real-world environments, and not only on the task for which it was conceived. Control tests were performed to disentangle possible cognitive or motor confounding factors. These tests address specificity, or whether the device is targeting the desired function and not a correlate of it. Finally, neurophenomenological validation provided an agreement between the objective estimate of WM load and the subjective WM load reported by the user.

In chapter 3, we develop a data-driven framework for studying the spatio temporal structure of brain state switches under cognition, with two specific objectives. First, this framework allows us to perform feature engineering by taking advantage of regularities, or patterns of brain activity, elicited

by cognition. If cognition produces changes in said patterns, then these patterns can be used as descriptors, or features, in a cognitive BCI. The second objective motivating the development of the framework is to answer scientific questions regarding brain organization and brain dynamics. These questions arose during the construction of the WM BCI. Specifically, we aim at investigating how the brain self-organizes allowing different regions to engage and disengage in joint activity in a manner driven by cognition. Assuming brain metastability (in the context of statistical physics), we propose a set of local variables that are expected to be spatially and temporarily affected by cognitive states. We support the latter claim by correlating these variables with cognitive conditions, such as high-WM load, Alzheimer disease, and positive emotional valence. We also analyse whether the switching between states occurs at discrete times, as often proposed in the literature, and find evidence challenging the discrete model.

Each of these two chapters is meant to become a journal article, with the author of this thesis as the main author. In addition, the appendix contains previously published journal and conference papers.

TABLE OF CONTENTS

	Page
1 Introduction	1
1.1 Design of a real time brain–computer interface to continuously monitor Working Memory	2
1.1.1 Models of Working Memory	2
1.1.2 Why and how can the Working Memory load be estimated?	8
1.2 Brain metastability under cognition	15
1.2.1 Transient coordination in the brain	15
1.2.2 Metastability in the brain	17
1.2.3 EEG non-stationarity	20
1.2.4 A data-driven motivation	21
2 Working Memory load estimation in real time	23
2.1 Development of the BCI	23
2.1.1 Initial WM task	23
2.1.2 Online tests: BCI validation and cross-task	28
2.1.3 Online control tests: Disentanglement of potential confounding factors	30
2.1.4 Data Acquisition	34
2.2 Results	36
2.2.1 BCI validation	36
2.2.2 Confounding factors	38
2.2.3 Neurophenomenological validation	40
2.2.4 Temporal behaviour of the WMLE	40

TABLE OF CONTENTS

2.2.5	The EEG biomarkers: Spectral EEG changes due to WM load	41
2.3	Additional material	43
2.3.1	Images used for Task 1	43
2.3.2	Design and training of the BCI	44
2.3.3	Arithmetic operations in Task 2	49
2.3.4	Estimation of statistical significance	50
3	A framework for investigating the spatio temporal brain organization under cognition	53
3.1	Data acquisition	53
3.1.1	WM dataset	54
3.1.2	AD dataset	54
3.1.3	Emotions dataset	54
3.2	The framework	55
3.3	Method 1: Temporal structure of the switches	62
3.3.1	Results specific to Method 1	64
3.3.2	Discussion specific to Method 1	66
3.4	Method 2: Spatial synchrony between states	69
3.5	Method 3: Power law of the power spectra	70
3.5.1	Results specific to Method 3	71
3.5.2	Discussion specific to Method 3	71
3.6	Method 4: The most visited states, and how many of them are available	72
3.6.1	Results specific to Method 4	74
3.6.2	Discussion specific to Method 4	74
3.7	Baseline: power spectral density	77
3.8	Control tests	77
3.8.1	Control test 1: Destroying temporal structure and assessing statistical significance	77
3.8.2	Control test 2: Removing known discrete events	78
3.8.3	Results specific to control tests	79
3.9	Global results	80

TABLE OF CONTENTS

4 Discussion	81
4.1 Working Memory BCI discussion	81
4.2 Brain metastability discussion	92
5 Appendix: published works	99
Bibliography	131

INTRODUCTION

Our aim was to design a system to monitor WM in real-time, from EEG measurements. The developed model was designed using the paradigm of passive BCI. As will be explained in detail later in this chapter, the goal of a passive BCI is to improve human—machine interactions by providing contextual information of the cognitive state of the user, in a manner not consciously driven by the user. The cognitive function that we selected for collecting data and building the online BCI was Working Memory (WM). WM is a general, complex theoretical construct that encompasses the mechanisms of storing, maintenance and processing of information while an individual performs any cognitive task. As a specific example, let us analyse an everyday task -driving- as a WM task. While driving, our sensory information is translated into estimates of the distances and speeds of relevant objects. This information is held and processed in WM to take appropriate actions, such as turning, accelerating or braking. Any activity that involves a cognitive task will then involve the WM system, from keeping a phone number in mind to engaging in speech comprehension. Due to this ubiquity, WM is a key element of cognition. Developing a consistent definition of WM with both explanatory and predictive power is a relevant issue in psychology and in neuroscience, since, by definition, it is a

theoretical construct relevant to all cognitive functions.

1.1 Design of a real time brain–computer interface to continuously monitor Working Memory

1.1.1 Models of Working Memory

The time-limited ability to retain information has been described by different concepts, such as short-term memory (STM), short term store or primary memory. While STM is the cognitive system responsible for holding sensory, motor and cognitive information [73], the WM construct has more intrinsic features [2], as it is thought of a whole interface between perception, long term memory (LTM) and action [10]. In addition, the concept of WM stresses the processing nature of this temporary storage. Miller coined the term *Working Memory* while studying the everyday formation, transformation and execution of plans in the context of behavioural science [123]. Even for simple plans in our everyday life, we need to combine and mentally manipulate information from different sources, such as our habits, current context, and expectations. Later, Atkinson and Shiffrin used the term coined by Miller in their short-term store model [9], influenced by information flow in machines.

Finally, Baddeley and Hitch [13] built on Atkinson and Shiffrin’s model to create their multiple-component model of WM. Although it has been further modified since its first appearance, Baddeley’s model remains the most influential WM model. Several authors, however, have pointed out possible improvements or limits of its applicability. In the next section, Baddeley’s model and other alternative models developed in Cognitive Psychology are outlined.

1.1. DESIGN OF A REAL TIME BRAIN–COMPUTER INTERFACE TO CONTINUOUSLY MONITOR WORKING MEMORY

1.1.1.1 Alan D. Baddeley’s multiple-component model.

Baddeley’s original model included two slave storage subsystems, the phonological loop and the visuospatial sketchpad. The model also included a coordinating system, the central executive.

The phonological loop is in charge of the storage and maintenance of auditory information. It is further subdivided into two components, one responsible for storing memory traces that quickly decay over time and the other responsible for rehearsal, which helps to actively maintain memory traces.

The visuospatial sketchpad comprises two other subsystems. These are the visual cache, holding information concerning shapes and colours, and the inner scribe, responsible for spatial information, such as location, trajectories and speed.

The function of the central executive, an attentional-based control system, is coordinating the slave subsystems, activating memory traces from the LTM, selecting coding strategies and shifting attention. Two of the main criticisms of the idea of a central executive are that it is depicted as a homunculus, an all-powerful person running WM, and that the lack of rigorous evidence makes it impossible to falsify [90]. Parkin argues that the evidence does not suggest a centralized executive function, but rather a pattern of executive tasks associated with different neural substrates. To tackle the homunculous problem, Baddeley [11] proposes explicitly characterising all the executive roles of the homunculus until it becomes redundant.

A new slave system, the episodic buffer, was recently introduced by Baddeley [11]. The episodic buffer is the storage counterpart of the central executive. Multi dimensional information from different sources, bound by the central executive, is stored in time-ordered *episodes*, like the fragment of a story. These episodes are then linked to multi-dimensional represen-

tations in LTM. The episodic buffer was mainly proposed to explain how densely amnesiac patients, with important LTM impairment, can perform normally in immediate recall tasks. Such tasks, which contain more than 20 elements, and hence, are beyond the capacities of the verbal and spatial systems, were previously thought to be mediated by LTM.

As multiple components are concerned, the dual task paradigm has been the main experimental tool for investigating WM under this model. This paradigm explores the possible interference of two tasks to determine whether they compete for resources. Two WM subsystems are said to be independent if they do not compete for resources, like the phonological loop and the visuospatial sketchpad. An example of interference is the articulatory suppression task, where a subject is instructed to speak while trying to remember a collection of words. Memory is impaired when performing this task because speech and rehearsal in the phonological loop are expected to share resources.

1.1.1.2 Nelson Cowan's embedded-process model.

The WM model proposed by Cowan [30] outlines more precisely how the mechanisms underlying attention interact with WM, and it proposes that the slave subsystems could employ more general types of encoding (other than auditory and visual ones). In terms of information flow, Cowan's model can be characterised as in Figure 1.1.

Information enters the brief sensory store and is retained for several hundred milliseconds, whereupon LTM representations (sensory or semantic) become active and remain so for a few seconds. Depending on the salience of the stimuli and/or voluntary attention, the activated memories may enter into the focus of attention or remain outside of it, while still being active. The attentional processes are mediated by the central executive, which can direct attention either outward, to perceived stimuli, or inward,

1.1. DESIGN OF A REAL TIME BRAIN-COMPUTER INTERFACE TO CONTINUOUSLY MONITOR WORKING MEMORY

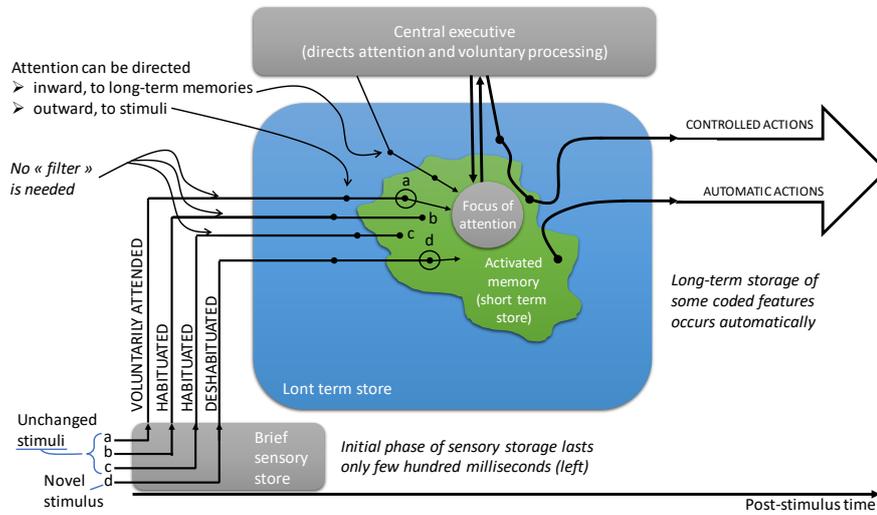


Figure 1.1: Information flow in Cowan's model, modified from Cowan 1988. Time is represented along the horizontal axis, events are discrete approximations to continuous processes that may take place in parallel or in cascades. The focus of attention is represented as a subset of the short-term storage, which is in turn represented as a subset of the long-term storage. The only items that enter the focus of attention are changed stimuli, items (sensory or not) voluntarily selected, and long-term memory items spontaneously activated based on associations (not shown in the figure).

to LTM. The processing of activated traces of LTM may lead to controlled actions, if information passed through the focus of attention, or automatic actions otherwise. LTM storage of some coded features occurs automatically.

Under this model [32], outside the focus of attention, physical encoding of features is more likely than semantic encoding, except for information that is highly relevant for the subject. The latter is evidenced by the attentional switch after hearing one's name in an unattended channel. Experimental evidence [33] suggests that the capacity of the focus of attention is 4 ± 1 items, while the whole number of activated items is 7 ± 2 . Processing can also be performed on active items out of the focus of attention, without awareness, as evidenced by patients with hemispatial neglect able to perform operations on nonconsciously perceived elements. The main mechanism of information degradation is interference, and temporal decay probably plays a role in this.

1.1.1.3 Ericsson and Kintsch's long-term working memory model

The usual models of WM have succeeded in characterizing WM tasks performed in the laboratory. Nevertheless, the large storage required to perform text comprehension and other skilled activity, like playing chess or becoming a digit span expert, can not be explained by these models, which rely only on temporal limited capacity [38]. For instance, performance on a specific memory task can improve tenfold after practice; however, this shows no correlation with the memory span of a different task [37]. Furthermore, some individuals with STM impairments perform normally during skilled activities, such as text comprehension. In general, the number of available items in WM can not explain skilled activity. It is in this context that Ericsson and Knitsch proposed their long-term Working Memory (LTWM) model.

These researchers proposed that skilled activity in everyday life does not heavily rely on any temporal storage. In contrast, while developing skills, semantic structures are built in LTM that allow efficient coding and fast retrieval. Hence, LTM largely mediates expert performance. As these semantic structures are domain specific, skills acquired on a particular type of task involving memory are not necessarily useful for a different memory task. Thus, performance is only correlated if new materials can be meaningfully encoded using the developed semantic structures and retrieval techniques.

1.1.1.4 Barrouillet et al.'s time-based resource-sharing model

The time-based resource-sharing model (TBRS) [17] proposes an interesting way of defining cognitive load. Under the TBRS model, attentional resources are needed not only for the processing of information, but also for its activation and maintenance. They are required for complex tasks and for simple activities, such as reading letters or digits.

1.1. DESIGN OF A REAL TIME BRAIN–COMPUTER INTERFACE TO CONTINUOUSLY MONITOR WORKING MEMORY

Attention, a serial resource, is shared among at least three functions: processing, activation and maintenance of information. Quick pauses are required during processing to maintain the memory traces, which would otherwise decay over time. It is important to note that this process does not necessarily correspond to the rehearsal in the phonological loop proposed by Baddeley [13]. Different mechanisms could occur instead, like the rapid and covert retrieval process through attentional focusing proposed by Cowan [31]. As the activation of an element is about to fade away, the processing is paused, and it can be resumed after maintenance takes place. Attentional switches may occur constantly and at the micro-level, as described in terms of the micro-task-switching process proposed by Towse et al. [113]. This process is serial in nature at the micro-level but rapid enough to seem parallel at the macro-level.

As a consequence of the attentional constraint, it is important to redefine the concept of cognitive load. A high WM load condition may involve not only a high number of active items, but also the potentially available time that can be devoted to attentional switches aiming to refresh memory traces. If the task allows enough time to ensure a proper maintenance of memory traces, it is said to correspond to a low cognitive load; conversely, if high processing demands leave little time for refreshing, the task is said to impose a high cognitive load. Considering the latter, the concept of load can be thought of task dependent.

To show how performance depends on cognitive load, defined in this way, Barrouillet et al. [19] studied the change in memory span as they changed the number-of-digits-to-time ratio, which measures both the number of digits and the time available for refreshing memory traces. There is a clear linear relationship between these two quantities, where the span decreases as the ratio increases. Hence, performance is impaired by both, increasing the number of elements and decreasing the available time.

Furthermore, Barrouillet et al. studied how this ability to perform suc-

successful attentional switches was consolidated through human development. They found that it often appears at the age of 7 years and improves during childhood and adolescence [19].

1.1.2 Why and how can the Working Memory load be estimated?

To answer the question of why and how the WM load can be estimated, we must first address the consequences of a potential WM overload. According to Cognitive Load Theory [89], the design of instructional material should aim at imposing the right amount of WM load to attain optimal learning. In an industrial context, a potential WM overload may lead to accidents. Especially, WM errors may lead to incidents and accidents in fields like air traffic control[105]. WM load estimation might lead to interesting applications in the context of education or industrial environments. The difficulty of instructional materials, and therefore the WM load they impose, might be adjusted automatically to design adaptive, personalized learning strategies. When a human operator is in a situation in which WM overload could compromise the environment's security, early detection might prevent accidents.

Furthermore, one study has shown that General Fluid Intelligence (Gf), that is the ability to solve novel problems (as opposed to acquired skills) may be improved by training WM [60]. The study showed that Gf increases with the amount of training on a multi-modal n-back task. WM monitoring in real time would therefore facilitate the developing of new neuroeducational tools to train Gf for cognitive augmentation.

Finally, brain imaging is necessary for investigating the neural correlates of any cognitive function. Several brain imaging techniques, such as magnetic imaging or electroencephalography (EEG), are used by researchers to infer structural and functional properties of the brain. The technique chosen to develop this work was EEG. Among its advantages are high temporal

1.1. DESIGN OF A REAL TIME BRAIN–COMPUTER INTERFACE TO CONTINUOUSLY MONITOR WORKING MEMORY

resolution (making it ideal for real-time applications), non-invasive nature and low costs. One of its disadvantages is its low spatial resolution.

1.1.2.1 Spectral properties of EEG

EEG recordings consist of a series of measurements of brain electrical activity, measured in microvolts, over time. Recordings are performed through electrodes, which can measure the time evolution of voltage amplitude. Such variations are correlated with changes in the synchronization of the underlying neural ensembles. Local synchrony produces oscillatory and phasic patterns, which can be characterized by the way in which power is distributed over frequencies. A description of power as a function of the frequency, per unit frequency, is called the power spectral density (PSD).

If we consider, for instance, a pure sinusoidal signal of frequency f_0 spanning through an infinite time, all the power will be concentrated at f_0 . Real-life signals however are finite, and in that case the power is concentrated around f_0 , the longer the signal lasts, the narrower the peak of the power becomes. A sum of two finite sinusoids of frequencies f_0 and f_1 , respectively, will generate a PSD with two peaks, at f_0 and f_1 , of heights that depend on each signal's amplitude. We used Welch's method [124] for estimating the PSD, a method especially tailored for finite, noisy data. Examples of PSD estimation can be seen in Figure 1.2.

The EEG PSD is usually divided into different bands, and it is common to compute the power at each band. The power at a specific band is the integral of the PSD in the region containing the frequencies of interest. There is no universal convention for choosing the exact boundaries of the bands; in our work we use the delta (1–4 Hz), theta (4–8 Hz), alpha (8–12 Hz), lower beta (12–20 Hz), upper beta (20–30 Hz) and lower gamma (30–45 Hz) ranges. The band's boundaries are partly historical and partly scientific: It has been observed that a change in the power of a band correlates with motor or

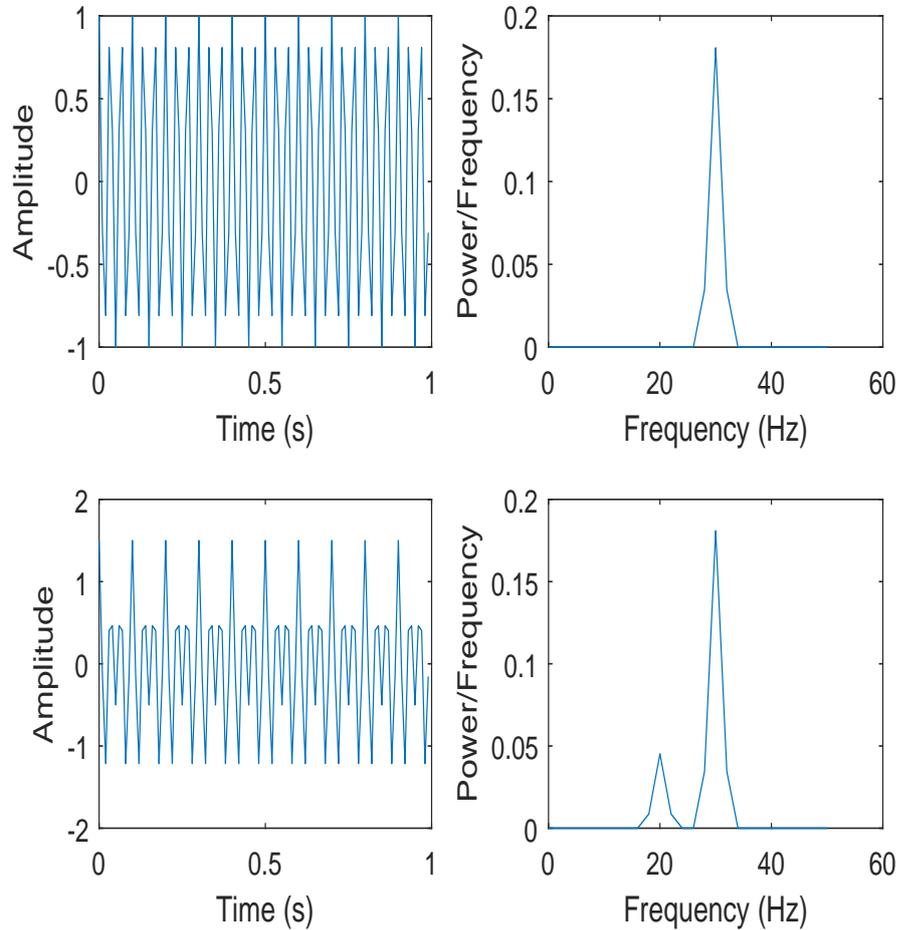


Figure 1.2: Two different power spectral density (PSD) estimations. Top: constant amplitude, finite time sinusoid, bottom: sum of two constant amplitude, finite time sinusoids. The time series are in the left column, PSD estimations are in the right column.

cognitive activity. For instance, alpha waves are observed when an individual is awake with eyes closed, and they are thought to predict mistakes [80]. It is a matter of debate, and outside the scope of this work, whether specific oscillations are necessary for brain functioning or a byproduct of it. However, regarding WM, several types of oscillations have been proposed as neural correlates by other authors. Sauseng et al. [101] report alpha synchronization (leading to an increase of alpha power) in the prefrontal

1.1. DESIGN OF A REAL TIME BRAIN–COMPUTER INTERFACE TO CONTINUOUSLY MONITOR WORKING MEMORY

areas, with alpha de-synchronization (decrease in power) in the occipital areas. In a different study [102], they report increased theta long-range coherence¹ and decreased anterior upper-alpha short-range connectivity for increasing demands on the central executive. Jensen et al. [62] find that theta power at the frontal electrodes increases as a function of the number of items stored in WM. More generally, Antonenko et al. [8] review the use of EEG for measuring cognitive load.

1.1.2.2 Other types of EEG measures

EEG signal processing techniques aim at extracting informative measures, or features, from the raw recordings. While power at different frequency bands remains a popular feature choice, there are several alternatives. Unlike PSD estimations that are timeless, short time Fourier transforms [48] [59] and wavelets [6] [3] allow researchers to resolve the temporal evolution of how power is distributed over frequencies. Complexity measures aim at estimating how regular, or predictable, a signal is. Multiple complexity measures have been applied to EEG signal processing, including the correlation dimension in schizophrenia [74] and creative thinking [83], Lempel–Ziv complexity for Alzheimer disease [1] and depth of anesthesia [130], and multiscale entropy in autism [23]. In autoregressive models, future values of a time series are estimated as a function (weighted sums in general) of past events. Autoregressive models of EEG signals have been proposed, and used, for instance, for classifying motor imagery [92] or mental tasks [87].

In chapter 3, we develop a novel set of techniques to extract meaningful information about cognition from EEG signals. The motivation behind generating such framework is to assess assumptions about brain dynamics.

¹Two regions are said to display coherence if, for a given band, their oscillations entail a constant phase difference.

1.1.2.3 A brain–computer interface approach

We developed this WM monitoring system in the brain–computer interface (BCI) framework. Broadly, a BCI is a device that allows humans to interact with machines by using brain activity directly, bypassing the motor system [127]. These devices take neuroimaging signals as an input and generate a desired output via a translation algorithm. The output is often a command that helps the user in either conveying a message or controlling an object. Examples include choosing a letter in the first case, or moving a wheelchair in the second case. A diagram of a typical BCI is shown in Figure 1.3.

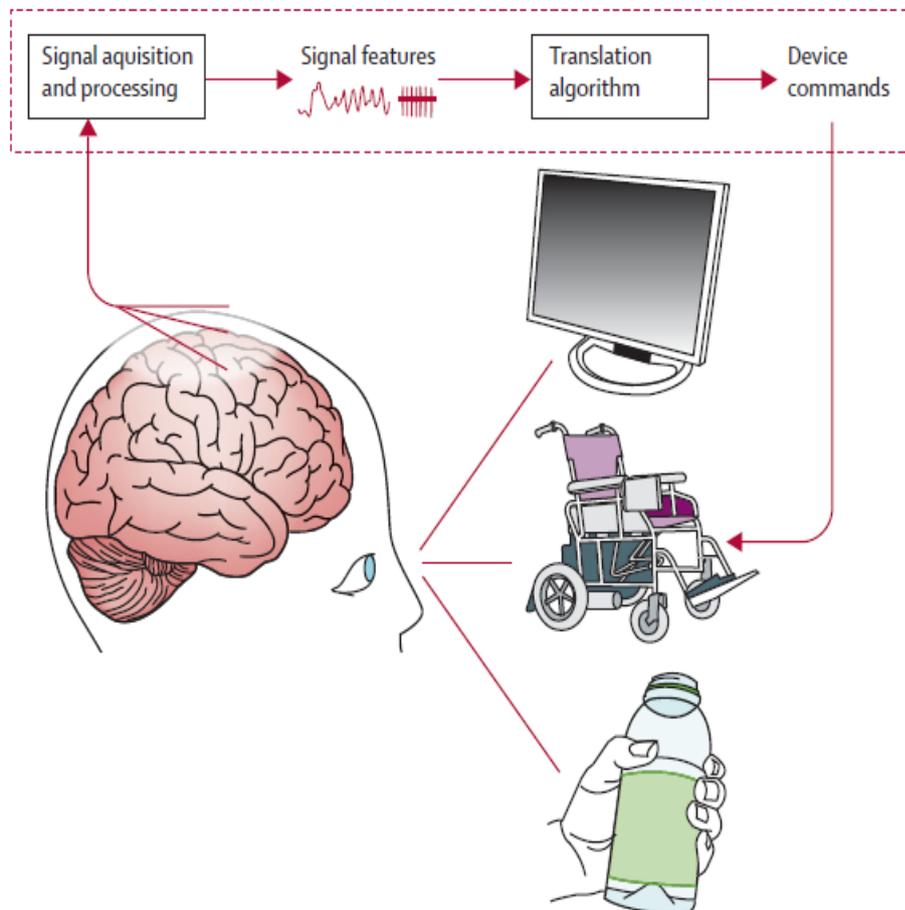


Figure 1.3: Figure taken from Allison 2007. The four BCI components are signal acquisition, signal processing, output applications and operating environment.

1.1. DESIGN OF A REAL TIME BRAIN–COMPUTER INTERFACE TO CONTINUOUSLY MONITOR WORKING MEMORY

BCIs were initially designed for patients whose conditions have impaired their main communication or command channels, such as speech or the motor system. Locked-in patients [57] or paralysed patients [86] are typical examples. In healthy subjects, regular speech is a more efficient way of communication. In addition, the motor system provides a fine-tuned means of control with a high number of degrees of freedom. Both the motor system and speech demand relatively low cognitive effort compared with a BCI. Nevertheless, BCI systems can go beyond communication and control; they can also be useful for cognitive monitoring, as suggested in [85].

Zander and Kothe [128] conceptualised a new way of using BCI. They proposed a new classification of BCI by the type of mechanism used to achieve control of the device, as follows: conscious voluntary control (active BCI), conscious voluntary control aided by external stimulation (reactive BCI), or non-intentional control (passive BCI). The system developed here is a passive interface. The goal of a passive BCI is to improve a system's performance by obtaining relevant information about the user's cognitive state. Such information is provided in a way that is non-voluntarily driven by the user. Having this information, appropriate commands can be triggered, allowing the system to adapt to the user. In the case of healthy users, passive BCI could be of special interest in improving human–machine interactions. Furthermore, feedback-based learning may greatly benefit from neurofeedback protocols based on a precise detection of cognitive states. [45].

1.1.2.4 Key properties of the proposed approach

To our knowledge, this is the first time that different important elements have been integrated into a single WM study, namely, neurophenomenological validation, a cross-task, and different control tests. Neurophenomenological validation refers to subjects confirming the agreement between their subjective experience and an objective measure proposed as a correlate to that experience. Although the main goal of neurophenomenology [117] is

beyond the scope of this work, it has interesting proposals regarding the relationship between subjective experiences and their objective, physical substrates. Lutz [78] provided the first case study in which first-person (subjective experience) and third-person (objective measures) data were related in the context of neurophenomenology.

A cross-task is a task differing from the one used to design the BCI, but nevertheless mobilizing the very same cognitive function. Testing the model with a cross-task is a necessary control before we attempt to generalize our findings to real-world environments. A BCI that truly measures the load of the central executive trained on a given WM task, should be able to transpose the classification of the same load on a different WM task.

Finally, disentangling confounders through control tests is crucial for supporting the claim that our EEG biomarkers correspond specifically to WM and not to the correlated electrical activity of brain or muscle origin. For a discussion of the importance of cross-tasks and disentangling confounders in a WM BCI, the reader can refer to [46]. Our hypothesis in this regard is that WM-load estimations will not be high when subjects are instructed to perform actions that are less demanding of WM but induce the presence of confounders. (More details in section 2.1.3). The confounding factors controlled in this work were attention, attentional filters, internal speech, sub-vocalization, frustration and arousal. Attention and attentional filters are indeed part of WM under some models [32], but they do not encompass the whole WM construct. A relevant approach in this regard is that of Vogel et al. [121], who show evidence suggesting that individual differences in WM capacity may be at least partially explained by individual differences in filtering efficiency. Individuals with a low WM capacity may have less efficient filtering mechanisms, which can lead to deficient encoding strategies and the consequent storage and maintenance of irrelevant information. The tasks developed in this work involved the phonological loop [10]; therefore, we need to disentangle potential correlates of internal speech and subvocalization. In addition, frustration is known to be highly correlated with

mental effort [89]. Eye blinks and other visible electromyographic artifacts were removed from the training set as explain in [100] to prevent biasing of the model. Eye blink removal was performed using the *runica* function of EEGLAB [34].

1.2 Brain metastability under cognition

As mentioned in the review of models of WM, the central executive is in charge of different sub-functions, namely, the maintenance of memory traces and information processing. In the series of experiments performed for the BCI, both sub-functions differed in intensity and pace across conditions (low or high WM load). For instance, refreshing of patterns via the phonological loop was performed more frequently in the high WM condition. In fact, according to the TBRs model discussed above, the WM load is not only determined by the storage requirements, but also by the use of the other sub-functions.

If we consider the working hypothesis that localised brain structures are responsible for performing the different sub-functions of the central executive, then it is reasonable to believe that there may be localised regions where properties of the neural substrates change to support said sub-functions. In chapter 3, a set of these properties that change across different WM (and other cognitive) conditions are postulated and studied. More specifically, the transient coordination in the brain under cognition is investigated borrowing concepts from physics such as metastability, to derive precise formulations of the above mentioned properties.

1.2.1 Transient coordination in the brain

Perception provides an illustrative background for elaborating on the type of cognition-driven brain dynamics that we are interested in capturing.

How do we perceive objects? In the 1960s, Jerome Lettvin coined the term “grandmother cell,” referring to the idea that the brain has neurons where the activity is elicited selectively for familiar faces, with a single neuron for a specific person. Later, this view came to be regarded as simplistic. However, while investigating the degree of selectiveness of neurons, Quian Quiroga et al. [94] found that although we may not have grandmother cells, neurons fire in a specific way for familiar objects. In a patient with implanted electrodes, they found a neuron that fired when pictures of Halley Berry were presented. They did not fire when pictures of anyone else were presented, and they also fired when the patient viewed a drawing of Halley Berry, the name of Halley Berry, and Halley Berry dressed as catwoman. They found similar behaviour for other familiar faces and even for buildings. Later, Chang and Tsao [26] decoded the way in which primates encode faces. They were able to reconstruct any face (not necessarily familiar) with an impressive accuracy, by recording the joint activity of 205 specific cells. Each cell codes for a specific face attribute, so its firing rate corresponds to the degree to which the face can be described by the corresponding attribute.

To summarize, familiar objects have a sparse representation, and faces, perhaps due to evolutionary pressure, are coded in a sparse manner that involves a few hundreds of cells. However, as neurons do not generally fire for single objects, how do we integrate low-level information into high-level categories? The *feature binding problem* describes the integration of this type of information.

Feature binding [114] refers to the process by which the selection and integration of the different properties of objects takes place, in the correct order. Sensory modalities are registered at different brain regions, and the same modality may even be registered at different locations [96]. Considering vision, the ventral pathway is responsible for registering colours and shapes, while the dorsal pathway is responsible for registering motion and space [82]. Our visual field is populated by a collection of objects that usually change over time. Thus, it can be concluded that, as objects or their

properties change in our visual field, different brain regions engage and disengage in transient states of coordination. In addition, visual perception does not only involve low-level signals; rather, the mere concept of an object requires us to identify properties like manipulability or topological connectedness [111]. Beyond the visual modality, our perception of the environment consists of a broadband, high temporal resolution stream of information.

Generally, cognitive functions require the interaction of low- and high-level information, both external and internal. Emotions, external stimuli, intentions and memories all interact in a coordinated fashion in our brain, and the mind can be thought of as the workspace in which these interactions occur. Writing down an idea (in a syllabic writing system), a simple everyday task, can illustrate this intricate set of interactions. To accomplish this, an abstract idea needs to be phrased in words in our mind, and further decomposed into its constituent phonemes. The graphemes corresponding to the phonemes must be retrieved from memory, and visual, motor, and haptic information must be integrated to perform the actual writing.

How does the brain manage to self-organise to create and annihilate these transient coordination involving low- and high-level information? As early as 1974, Katchalsky et al. (cited in [125]) wrote, “waves, oscillations, macrostates emerging out of cooperative processes, sudden transitions, patterning, etc. seem made to order to assist in the understanding of integrative processes of the nervous system”. More recently, the concept of metastability has attracted attention.

1.2.2 Metastability in the brain

Freeman’s [43] work regarding perception led to the discovery of cortical activity that carries perceptual elements [44]. Freeman describes metastability as the recurrence of spatial patterns of phase and amplitude in the neocortex, that occur in a discrete fashion like frames in a movie. An example of these perception frames is displayed in Figure 1.4, adapted from [44].

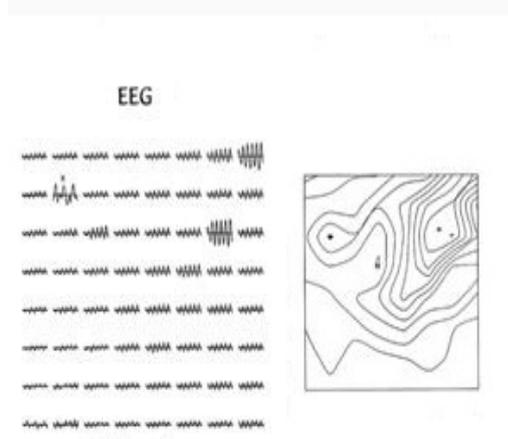


Figure 1.4: A Freeman frame, taken from Freeman’s Scholarpedia article. Left: 64 EEG signals, from a 8x8 array. Right: contour plot, formed with the amplitudes of the array. The pattern that arises from the contour plot is a frame.

Buzsáki [24, chapter 5] suggests that the brain is in a high complexity, critical state, as may be evidenced by the power law in the EEG PSD. He also proposes that the most important property of cortical brain dynamics is the ability to rapidly switch between metastable pink noise and oscillatory behaviour. Under this view, sensory or motor activity represent perturbations (which we will refer to as disturbances) that can temporarily reorganize the effective connectivity to induce transient stability by oscillations. An oscillatory, short-lived regime can hold information required for psychological constructs, whereas the critical state allows for an efficient switching between states. *Coordination dynamics* [65] represents a theoretical framework in which complex systems theory is used to model this transient coordination. In this framework, metastability is a dynamical regime for the relative phase of coupled oscillators in which all stable fixed points have disappeared. Phase trapping, temporarily convergent dynamics, and phase scattering, temporarily divergent dynamics, are the result of competing tendencies: On the one hand, segregation, or modularity, pro-

notes independent behaviour and local coupling; on the other, integration represents a global attempt for cooperation. One of the reasons this author proposes metastability is that, unlike multistability (where stable fixed points still exist), metastability does not require disengage mechanisms (such as stochastic noise or energy flow) for state switching. Tognoli and Kelso [112] point out that, although the concept of phase locking² has gained increasing relevance in the study of neural assembly synchronization, transients have not received adequate attention, perhaps due to the lack of truly dynamical approaches. According to these researchers, metastability has yet to be demonstrated and fully treated from a spatiotemporal perspective. To take a further step in that direction, we propose the framework presented in this study.

We are interested in finding measurable physiological variables that are related to brain states, in characterizing the coordination of these variables, and in correlating this coordination with cognition. We begin by assuming metastability, not in the dynamical systems context, but rather in that of physics. Again, a full description of metastability in the context of physics is out of the scope of this thesis; nevertheless, we can outline it. A metastable system is a system out of equilibrium, with several available states (for instance, liquid, solid and gas for water), and near the boundary between a subset of them. External inputs (energy, noise, matter, etc.) can drive the system into one state or the other. If the system stays in a given state, the system's state variables (distribution of molecular velocities in the case of water) remain stationary. In our case, the state variables will be the output voltage of the neurons, as measured by EEG (space-averaged). According to the definitions of metastability and stationarity, a system that switches between metastable states will have constant statistical parameters over time during the existence of a state. In other words, the entire process generating the state variables will be piecewise stationary.

²Functionally coupled neurons that spike at a constant delay [116].

1.2.3 EEG non-stationarity

EEG is known to be non stationary [16]. However it is considered to be composed of concatenated stationary segments. Kaplan et al. [64] suggest that these non stationarities may arise from the switching of the metastable states of neural assemblies during brain functioning. Thus, the lack of EEG stationarity is a suitable candidate for studying brain metastability. It is surprising that, although EEG non stationarity may result from normal brain functioning, few researchers have directly investigated whether non stationarities can convey relevant information about cognition, or about brain functioning in general. Usually, EEG nonstationarity is either not discussed at all or considered as an issue to overcome, given that many techniques, such as PSD estimations, complexity measures and autoregressive models require stationarity. Common approaches are signal segmentation into stationary epochs [4] [42] or the use of techniques that do not assume stationarity [55] [76]. To list a few exceptions, Kaplan et al. [64] developed a technique estimating synchrony between any two channels (*operational synchrony*) as the degree to which they undergo simultaneous switches. Cao and Slobounov [25] studied the change of the dominant frequency of the EEG signal over time, and they used this measure for detecting residual abnormalities in concussed individuals. In a study regarding depth of anaesthesia, Kreuzer et al. [75] found that during loss of consciousness, stationarity is heavily influenced by the anaesthetic used. Fingelkurts and Fingelkurts [39] further developed operational synchrony to propose the framework of *operational architectonics*, aiming at characterizing the temporal structure of information flow in functionally connected neural networks.

Most of the above-cited research assumes discrete timing. Switches occur in an abrupt manner, and the region remains in the same state until the next switch. Fingelkurts and Fingelkurts [40] discuss the differences between the concepts of *elements of thought* and *stream of consciousness*. They review psychophysical, electrophysiological, neurophysiological and computational support for either discreteness or continuity of timing in

cognition.

1.2.4 A data-driven motivation

For the sake of clarity, the above introduction was presented from a theoretical perspective. In our case, however, the original motivation was data driven. Signal processing techniques involving spectral properties and complexity measures require the signal to be stationary. In contrast, as mentioned in the previous section, EEG is non-stationary. Our initial approach, like the approaches of most researchers who have addressed the issue, was to segment the signal into stationary epochs. As stationarity implies constant statistical properties over time by definition, we decided to investigate the time evolution of statistical properties of the EEG signal. Visual inspection of the WM dataset (see section 3.1) led us to realise that there were some patterns of change of the statistical properties that were more prominent in the high WM condition.

Figure 1.5 shows two valleys, A and B. While visually inspecting the data, these type of valleys seemed to appear more often in the high WM condition as compared with the low WM condition, at least for central EEG channels. We realised then that if the actions performed by the central executive differed across conditions in nature, intensity or pace, the segmentation process itself could be used to derive a biomarker, provided that changes in the statistical properties were adequately represented. In addition, these signal changes, which may reflect underlying changes to support different functions, could shed some light on brain dynamics.

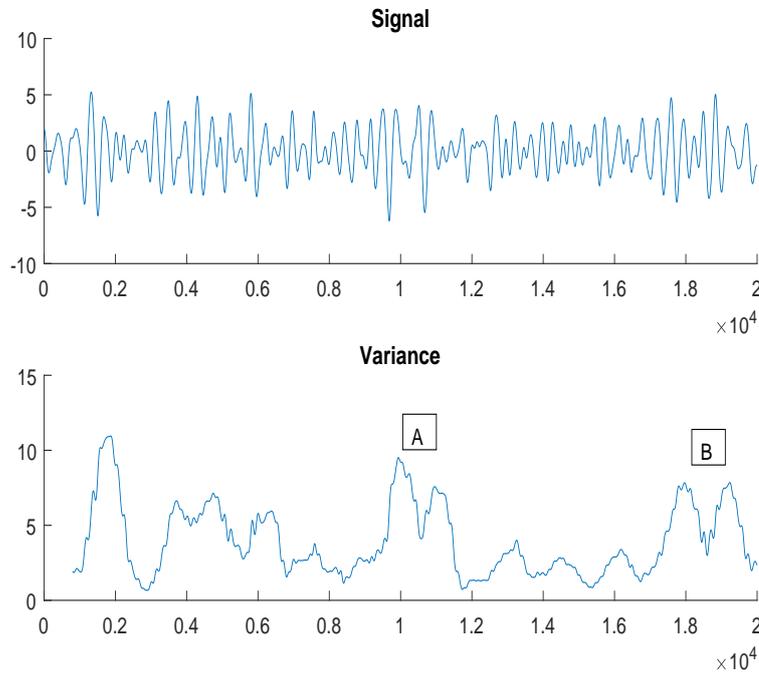


Figure 1.5: Top: typical segment of the EEG signal over time. Bottom: variance computed over a sliding window. For central channels, during the visual inspection of the data, valleys like A and B were found more frequently in the high WM condition than in the low WM condition.

We decided therefore to investigate the spatio temporal structure of the switches, and more importantly, whether they correlated with WM enough to be predictors of WM load at the single trial level. After obtaining successful results, we decided to explore whether the findings were WM specific or a more general aspect of cognition. We applied the technique to a dataset of emotions and an Alzheimer disease dataset with positive results.

WORKING MEMORY LOAD ESTIMATION IN REAL TIME

2.1 Development of the BCI

2.1.1 Initial WM task

The ultimate goal of the present chapter is to reliably and continuously estimate a subject's WM load in real time from EEG recordings taken during a task. More technically, we need to perform EEG-based single trial classification of the WM maintenance and processing load in real time. To achieve this goal, our first step was to collect an EEG database of subjects performing a task with two conditions. Ideally, the conditions should differ only in the amount of WM load induced. (See section 2.1.3 for a feasibility discussion)

An offline analysis of the above database allowed us to design a classifier that, given a newly recorded EEG signal, provides real-time estimation of the posterior probability of the signal belonging to one of two classes, namely high-WM load and low-WM load. The design and implementation of

the task, which will be referred to as *Task 1*, are described as follows:

- Subjects sit in front of a computer screen and are presented with a collection of figures that will be used during the experiment. They are asked to assign a short name to each figure, to become familiar with the set. There are different sets of figures, and each set corresponds to a different semantic field such as animals, vehicles or geometric shapes (see section 2.3.1);
- The *target* to be memorised appears on the screen. The target is a specific sequence of the previously displayed figures. There are two conditions, one in which the target contains two figures (low-WM-load condition) and one in which the target length is five or six figures (high-WM-load condition). The number of figures in the high condition was determined for each subject depending on his or her WM span after performing 21 preliminary trials;
- The target disappears and a sequence of figures, generated from the same semantic field, slides from right to left on the screen. The sliding speed is 222 pixels per second. The subjects must press a button whenever they find the target within the sequence. This is considered one trial. If the subject presses the button before the target appears or misses the target, the trial is over and is not analysed. Trials last 25 seconds on average. An example of a low-load trial is shown in Figure 2.1. Subjective feedback about frustration is collected after each successful trial by means of an analog Likert scale. The question was taken from the NASA Task Load Index questionnaire [54]; and
- A new target (belonging to the other WM condition) is shown, and the whole process repeated. Both conditions are alternated to prevent the BCI from learning slow EEG drifts that are WM independent.

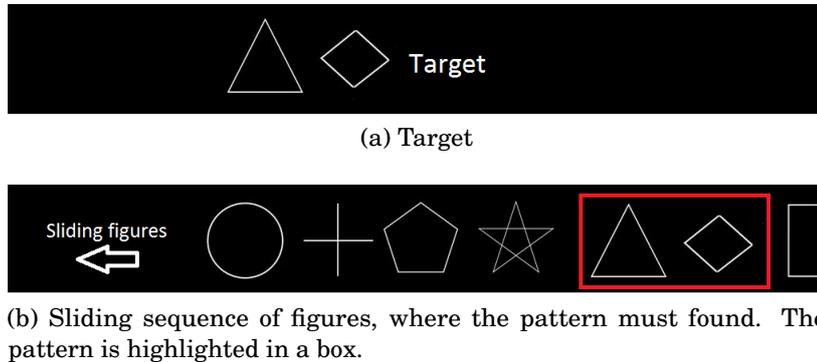


Figure 2.1: Low-WM task

The fact that subjects were asked to assign a short name to each figure induced in them a simple storage-retrieval technique—to internally repeat (using the phonological loop [10]) the names of the elements of the target and to compare them with the observed sliding items. Subjects were indeed instructed to do this to ensure a homogeneous encoding strategy.

While it may appear unnecessary or even trivial to impose a storage-retrieval technique, it is natural for subjects to explore other techniques in order to perform better. According to [63], in a digit span task, the typical initial strategy is simply to rehearse groups of numbers. As subjects become familiar with the task and obtain expertise, the encoding strategy moves towards associating numbers with their own pre-existing knowledge. The authors describe how a subject with a normal memory span develops a (task-specific) span far beyond the limits of WM. The encoding strategy exploited information with which the subject was familiar, namely racing times. (The subject was interested in racing sports, and hence he dealt with racing times often).

It was crucial in our experiment to prevent subjects from developing such strategies for several reasons: first, to have a constant WM load; second, to reduce variability, as the potential encoding strategies are as different as the body of knowledge of every subject; and finally, because we do not want to involve LTM. Mnemonic techniques, for instance, use long-term

WM structures for information retrieval [67].

A specific strategy of this kind, observed in preliminary tests, was to create short stories. Therefore, subjects were specifically instructed not to do so. To further discourage this possibility, all the figures belonged to an evident semantic field. To reduce habituation, the semantic field was changed.

For estimating the WM span of the subjects and the associated length of the target, the subjects performed seven trials for each of three different semantic fields (section 2.3.1). For the actual recordings, subjects performed 10 trials for each of four different semantic fields.

WM is a construct that is responsible not only for information storage, but also for its maintaining and processing [12]. While designing the task, we focused on these three sub-functions of WM. Storage is controlled by the design of the experiment, as both conditions differ in the number of items to be remembered. However, the maintaining and processing loads are harder to impose and monitor without largely complicating the experiment. Therefore, the subjects were instructed to internally maximise the difference in both conditions regarding these two remaining aspects. In the high-WM-load condition, they were asked to internally refresh the items as rapidly as possible (fast and continuous internal speech) and perform the processing (figure comparison) as intensively as possible. An intensive comparison could be, for example, not only deciding if the items are different, but also finding some differences, such as the number of lines. For the low-WM-load condition, they were asked to do the opposite—slow refreshing and low processing. Even though the task covered the three sub-functions of WM, the recordings for both the training and testing sessions were performed after the subjects memorised the target. Therefore, the recordings corresponded to the maintenance and processing of information with a number of items varying across conditions. Before the recordings, the subjects were left to interact with the system to gain familiarity with the instructions,

especially this maximization strategy. As pointed out by Lotte et al. [77], the effectiveness of feedback partly relies on the clarity of instructions and goals.

Distracters, or sequences similar to the target, prevent the use of memorization strategies focusing on subsets of the target. It is unlikely that distracters will appear by chance for targets with five figures or more. Hence, the randomized sequence was generated so that distracters appeared in exactly 50% of the trials. With distracters appearing only half the time, subjects do not learn to expect the target after a distracter. A distracter was defined as a pattern similar to the target, differing only in one item that could be located in any position from the third onwards. The duration of each trial is a random number between 15 and 30 seconds. A random value prevents subjects from implicitly learning the task length instead of performing the task itself, which would severely bias the results.

The area on the screen where the figures slid was 100×300 pixels; the size of the figures was 100×100 pixels. The small window size reduces eye movements, which are known to produce artifacts. Because of the size of the area, size of the figures, and sliding speed, the subjects could only see one complete figure at a time. The experiment was written in Matlab 2015a using the Psychophysics Toolbox extensions [68].

The distance between the subjects and the screen was 60 centimetres; the screen model was ProLite E2208HDD. The lighting and noise conditions were normal office conditions.

A photodiode connected directly to the EEG amplifier auxiliary input allowed synchronization between the EEG recordings and visual stimulation. The BPW-21R photodiode was chosen for its sensitivity to visible light (420–675 nm) and theoretical response time of about $3 \mu\text{s}$, lower than any other time scale in our setup.

2.1.2 Online tests: BCI validation and cross-task

As online tests are the only means of providing solid support to our working hypotheses, the present section is devoted to a description of the online testing of the BCI system. The classifier C, which is at the core of the BCI system, was designed and trained offline, as described in section 2.3.2, prior to testing.

The first online test was the BCI validation, that is whether the classifier correctly predicted the WM load online when subjects were performing Task 1, the task for which the classifier was trained. Classifier C was used to analyse the stream of incoming EEG data. The output of classifier C was the estimated probability that the current EEG epoch corresponded to a high-WM load, which is referred to as the WM-load estimate (WMLE). This value, being a probability, is a continuous number between 0 and 1, which is small for a typical low-WM-load EEG epoch, and large for a typical high-WM-load epoch. With the WMLE values, a receiver operating characteristic (ROC) curve was computed to assess the classifier's performance. An ROC curve is a plot of false-positive versus true-positive rate for different threshold values (for details see for instance [110]). The threshold is the value of the WMLE above which we consider an EEG epoch as corresponding to a high-WM-load state. The area under the curve (AUC) of a ROC curve is a useful indicator of the classifier's performance. AUC values have a lower bound of 0.5 for a random classifier and an upper bound of 1 for a perfect classifier; the larger the value, the better the classifier. Each subject except one (see section 2.1.4 for details about the subjects) performed 20 trials; the other due to his time constraints, performed only 6 trials.

The second online test aimed at obtaining neurophenomenological validation on a cross-task. Performing a cross-task is necessary to control task-related confounders [46]. The classifier C, trained on Task 1, was used to predict the WM load of the subjects in an entirely different WM-based task. A mental arithmetic task was chosen as the cross-task. During this

task, which we will refer to as *Task 2*, subjects were instructed to perform arithmetic computations (details in section 2.3.3), while a visual cue was shown on the screen. After 8.5 seconds, the visual cue disappeared, and subjects stopped the mental arithmetic; the trial lasted for 11.5 extra seconds after the visual cue disappeared. Classifier C analysed the stream of EEG data to provide a WMLE, as in Task 1. Unlike in Task 1, however, the WMLE was used to display real-time (every 150 ms) continuous feedback, in the form of a gauge with a height proportional to the WMLE. The gauge was shown throughout the whole trial (except for the first 2.5 seconds, as the buffer of the classifier C requires 2.5 seconds of data to produce an output). The neurophenomenological validation took place after each mental arithmetic trial, where the subjects were asked to decide whether the feedback provided by the gauge matched the dynamics of their subjectively perceived WM load. The subjects had to complete the sentence, *I believe that the feedback gauge was...* with one of the following: *a) correlated with my WM load, b) not correlated with my WM load, or c) I don't know*. To prevent an optimistic estimation due to a potential obsequiousness bias, half the time, sham feedback was provided. The subjects were aware that the aim was to validate whether the feedback indeed reflected their load, and that we would provide sham feedback half the time. The sham feedback took the form of a reversed estimate, that is, a large bar when the WMLE was low and a small bar when it was high. A reversed gauge has the advantage that its dynamical behaviour cannot be distinguished from the real feedback dynamics. Neither these questions nor the instructions given to the subjects mentioned that the sham feedback was reversed. This information was withheld to prevent subjects from being tempted to devote their cognitive resources to inverting their estimations and evaluating whether they matched the feedback. Ultimately, what we evaluated was the subjects' ability to identify whether the feedback was real or sham, which in turn, assesses the reliability of the BCI, provided confounders were disentangled. Indeed, if the subject is able to recognize the nature of the feedback (real or sham), then it means that the WMLE signal is matching with the subject's own internal evaluation of her or his cognitive load. Note

that in any case, if the subject were able to identify the sham feedback reversion (which did occur with some subjects) it would even more indicate a successful neurophenomenological validation: in order to identify that the signal is reversed, the subject would have to be able to interpret the information to be reversed (and hence understand its meaning as congruent to his internal perceptions of WM load). Each subject performed 20 trials, except for one, who performed only 10 due to personal time limitations. Before the recordings, there was a training session with three trials using the real estimation and three trials using the reversed estimation.

2.1.3 Online control tests: Disentanglement of potential confounding factors

In section 2.1.1, we described the requirement of a task with two conditions, differing only in the amount of WM load imposed on the subject. In practice, as WM is a multimodal complex construct, there may be *confounding factors* involved, that is, factors unspecific to WM, or task-dependent factors, that change across conditions [46].

Unspecific factors can be motor or cognitive confounders, such as frustration, attentional filters, eye blinks, subvocalization or muscle contractions. These confounding factors may or may not be part of the WM construct, but they do not encompass the whole construct, and basing a classifier only on them would be misleading. In contrast, a cross-task is meant to remove task-dependent factors.

Figure 2.2 is a graphical representation of the process of confounder disentanglement. The plane containing the ellipses is an abstract plane representing EEG biomarkers, with no specific order within the plane. The leftmost ellipse represents the set of biomarkers that change across conditions in Task 1. The rightmost ellipse corresponds to biomarkers that change across conditions in Task 2. The upper vertical ellipse depicts biomarkers

that change when a subject experiences high WM load, while the remaining ellipses represent biomarkers that change under the presence of the respective confounders.

If a classifier is trained with data from Task 1, and tested with data from Task 2, the EEG biomarkers that trigger a high response of the classifier (with the test data) are represented by the intersection of the leftmost and rightmost ellipses in Figure 2.2. These biomarkers are ideally task independent, due to the difference in nature between the tasks. However, this set of biomarkers is not free from biomarkers elicited by confounders, and therefore, we need to disentangle them.

In Figure 2.2, area 1 represents the ideal set of biomarkers. Area 2 contains cognitive activity necessary but not sufficient for WM, such as attention, that could potentially be shared by both tasks and change across conditions. Area 3 contains potential motor confounders that could also be shared by both tasks and change across conditions, like sub-vocalization. After all the confounders have been identified, the remaining part of the ellipse, area 4, should be empty. The recordings of a subject systematically producing electromyographic artifacts during the high-WM-load condition of both tasks (see section 2.2.2) belong to this area, and are therefore discarded from the results. According to the embodiment theory [118], the ellipses concerning cognitive activity and motor activity may not be disjoint. However, we are not considering this hypothesis in the present work.

The last set of online tests comprised control tests aiming at disentangling potential confounding factors. The goal of these tasks was to induce the identified confounders in a task that created a low WM demand, to verify whether a high response of classifier C was observed. (We do not call this response the WMLE because we are performing control tests). The confounders analysed were attentional filters, attention, internal speech, sub-vocalization, and frustration. Three control tasks, described in the next paragraph, covered these potential confounders. In addition, arousal was

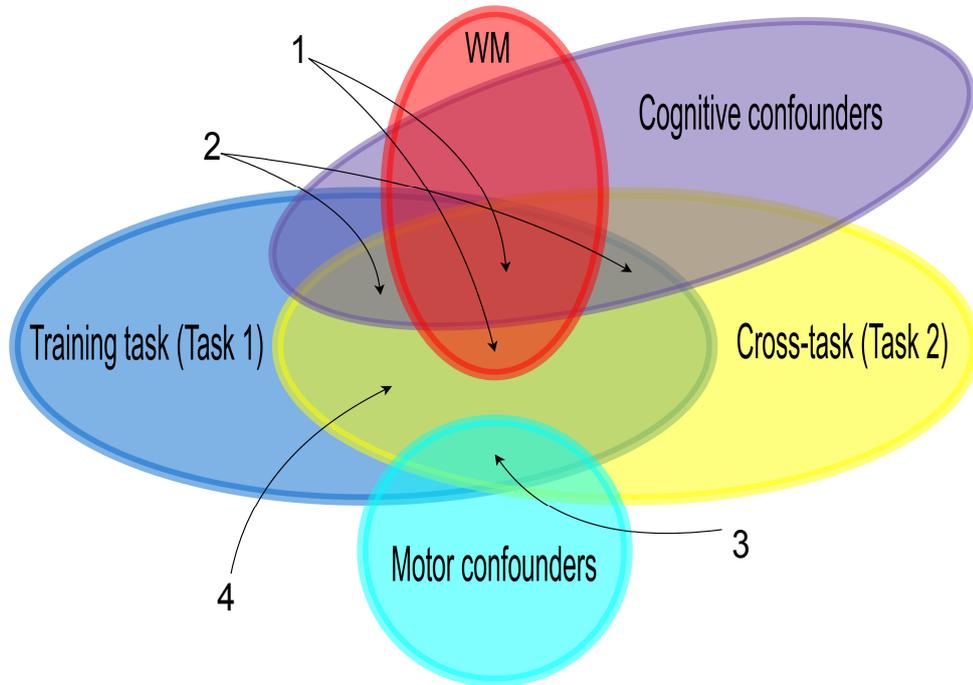


Figure 2.2: Confounders to be disentangled. The plane represents EEG biomarkers. Each ellipse is the set of biomarkers that change across conditions (for Task 1 and Task 2), or that change whenever the associated notion is present (WM, motor confounders and cognitive confounders). Area 1 represents the ideal WM markers. Area 2 represents (cognitive) activity necessary but not sufficient for WM. Area 3 represents potential motor confounders. Area 4, the remaining part of the crosshatched area, should be empty if all the potential confounders were correctly identified

analysed offline, and eye blinks were removed from the learning database; see section 2.3.2.

The first task was identical to the low-WM-load condition of Task 1, with one difference. Above the sliding figures where the target was contained, the picture of a red fly followed a chaotic trajectory for a random duration between 1 and 2 seconds. Subsequently, it would spin around for another random duration between 1 and 2 seconds. The fly alternated between these behaviours. This extra item, spanning the visual field with an unpredictable motion, forced the subjects to make greater use of their attentional filters

to succeed. The aim of this task was to test whether the attentional filters elicited part of the EEG biomarkers found. Twelve trials were performed for this test.

For the second control task, subjects completed a visual reaction time test. The goal of the test was to press a key whenever a visual cue appeared on the screen. As the cue appeared at random times, subjects needed to be attentive to press the key at the correct time. Therefore, the potential confounder of interest here was attention. Ten trials of 10 seconds each were performed.

A third control task involved subjects internally repeating a lengthy word of their choice slowly and continuously. The elements to be disentangled here were internal speech and sub-vocalization. Ten trials of 10 seconds each were performed.

At this point, we had already saved the WMLE values of subjects performing Task 1 in the low-WM-load condition. Our null hypothesis was that, if our EEG biomarkers were specific to WM, then potential confounders would not trigger a high response of classifier C. Therefore, the null hypothesis translates to the following: The response of the classifier C from control tests is not higher than WMLE values from Task 1 in the low-WM-load condition. Failing to reject the null hypothesis after an adequate statistical test would then support the claim that our BCI is WM specific. This test was a paired Student t -test, given that the same set of subjects performed Task 1 and the control tests.

Subjective information about frustration was collected after the Task 1 trials. As mentioned in section 2.1.1, this information was collected via an analogue Likert scale. For each trial, the mean value of the WMLE was compared with the subjective frustration level provided by the subject. A possible correlation between these two values was studied using a conjunctive analysis, including Bonferroni corrections [120]. This method,

previously proven useful for EEG, assesses statistical significance without losing statistical power when performing multiple hypothesis testing, each subject being a test in this case. A lack of correlation would support a lack of effect of frustration on the WMLE.

2.1.3.1 Potential confounding factor analysed offline: arousal

A recording session lasted between 2 and 3 hours. To avoid an adverse effect of fatigue, we estimated the effect of arousal offline. It has been reported in the literature that an increase in central frontal beta activity [51], decrease in central frontal theta activity [109], and increase in global alpha activity [109] are good markers of arousal. To test the effect of arousal, we decorrelated the information contained in these markers from the classifier's output. The latter was done by performing Gram–Schmidt orthogonalisation [27]. We tested this approach by classifying our offline database. If our classifier is based on arousal, after decorrelating these markers of arousal we expect the classification performance to drop to chance levels. If it is not based on arousal, we expect only a slight variation in the classification accuracy.

2.1.4 Data Acquisition

Brain activity was recorded using a 16-channel EEG device (Brain Products V-Amp) at a sampling rate of 500 Hz. The ground electrode was AFz and the reference electrode was FCz. The electrode set-up is shown in Figure 2.3. The collected data can be divided into two groups, namely the data obtained offline and data obtained online. For the offline data, 20 healthy subjects aged 21–31 years were recorded, including 10 males and 10 females. For the online tests, nine subjects were recorded, five males and four females. The online BCI validation was performed by all of the participants. Six of them, three males and three females did the cross task. Confounder disentanglement tests were performed on 4 subjects, 2 males and 2 females. Then, the subjects were asked if they had experienced mental fatigue. If

they responded affirmatively, the experiment was stopped, which was one reason why not all the subjects performed the full test battery. In addition, one subject was lost due to illiteracy after BCI validation (section 2.2.1). One subject did not perform the neurophenomenological validation (cross-task) due to cognitive difficulties manifested while performing the task (see section 4.1 for more details). All the subjects had normal or corrected to normal vision and the absence of any brain disorder or drug consumption. The study followed the principles outlined in the Declaration of Helsinki. All participants were given explanations about the nature of the experiment and signed an informed consent form before the experiment started.

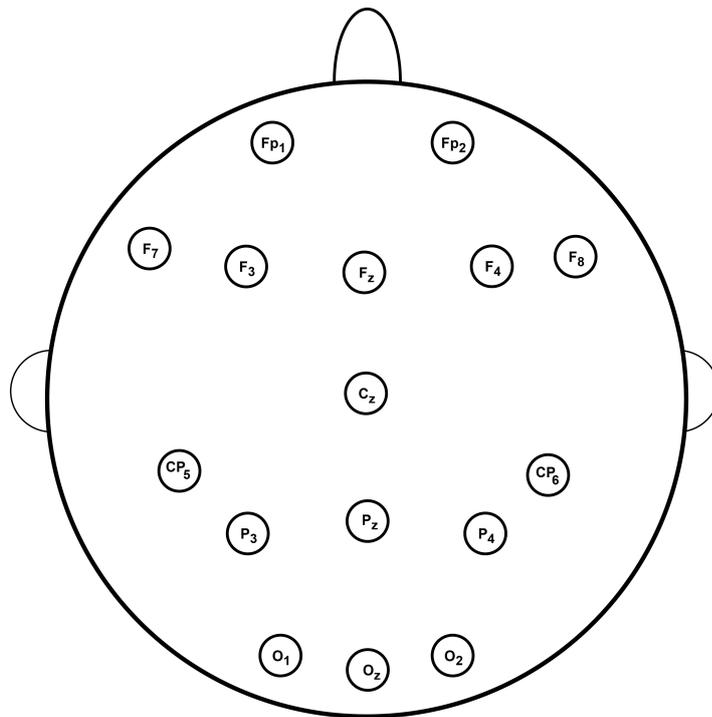


Figure 2.3: Electrode setup.

All the EEG epochs analysed were 2.5 seconds long. For the offline training data, a total of 1744 non-overlapping windows were analysed, 59% corresponding to a low-WM load. Online, for each subject, 90 non-overlapping windows were collected for calibration (see section 2.3.2.3),

representing 50% of each condition. Online tests were performed on the continuous stream of EEG data.

2.2 Results

2.2.1 BCI validation

A two-parameter ROC curve for the nine subjects was generated using the 126 artifact-free, successful trials of Task 1 online. The usual parameter in a ROC curve is the classification threshold; however, an additional parameter was relevant, namely the required sustained activity. In this work, there is a continuous estimate, in other words, a set of WMLE values over time for each trial instead of a global estimate of the WM load. We can, for instance, classify a trial as corresponding to high-WM load only if the activity stays above the threshold for a certain duration. Thus, for every threshold and for every required time (each pair being a possible BCI design) a sensitivity–specificity pair is available. Values are displayed in Figure 2.4 for the whole set of subjects. The curve is thick because of the two parameters.

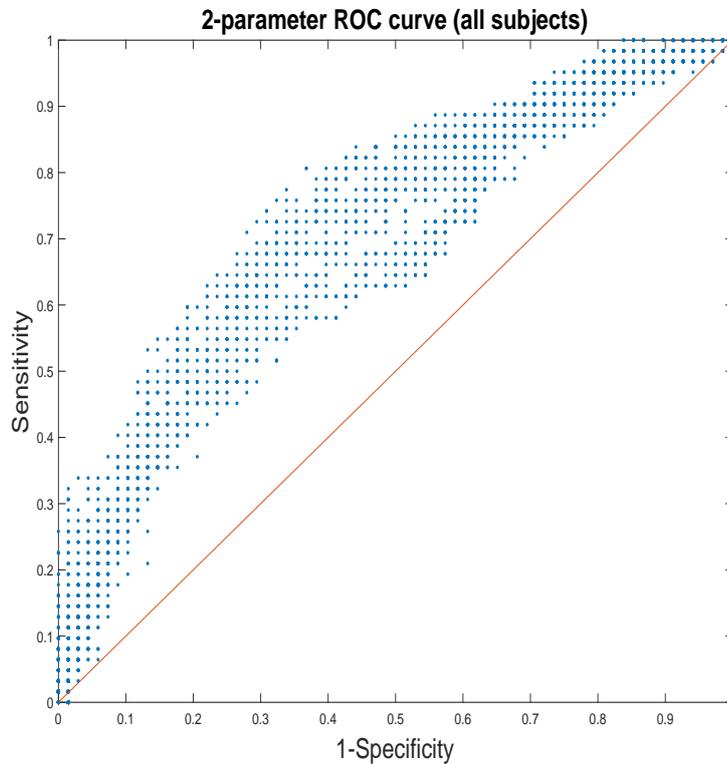


Figure 2.4: Two-parameter ROC curve for Task 1 performed online. The curve has thickness because there are two parameters—the classification threshold and the required time of sustained activity. Each point represents a possible BCI design, and the corresponding specificity–sensitivity pair is the global value when all the subjects are considered.

For a given specificity value, for instance, we can find the optimum threshold and required time so that sensitivity is maximised. Each value of the required time is different, but on average, the best value is 4.84 seconds of sustained activity. The AUC of the online classifier was 0.78 ($p < 0.0001$, see section 2.3.4 for details on how p -values were computed), well above the value of 0.5 of a random uniform classifier (i.e., a classifier that assigns each epoch randomly to one of the classes, with probability 0.5).

One subject was not asked to continue the experiment after the BCI validation due to the classifier’s low performance. The subject may have

been *illiterate*. It has been documented that up to 20% of the users [7] have too-high signal variability for using EEG-based BCI systems; this has not been fully studied for cognitive BCI. Nonetheless, the subject is included in Figure 2.4.

2.2.2 Confounding factors

We compared the distribution of WMLE values from Task 1 in the low-WM-load condition, with WMLE values of control tasks 1, 2 and 3. With a significance level $\alpha = 0.05$, the distribution of the WMLE values of the control tests was not statistically different (paired t-test) from the distribution of WMLE values of Task 1 in the low-WM-load condition. The WMLE of Task 1 in the low and in high conditions were indeed statistically different ($p = 0.037$).

Arousal did not show significant effects. The ROC curve obtained after decorrelating the information contained in markers of arousal is shown in Figure 2.5. Here, the AUC under the corrected curve is only 7% smaller than the AUC under the original curve.

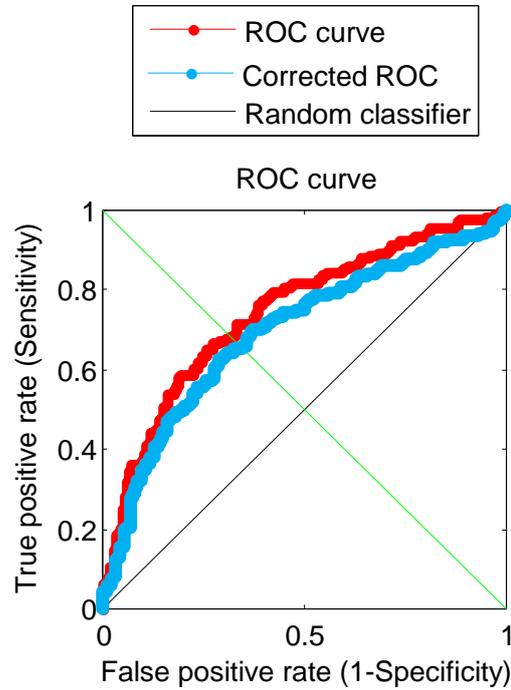


Figure 2.5: Corrected ROC curve, after decorrelating markers of arousal

The conjunctive analysis did not indicate any effect of frustration on the WMLE. The analysis yielded a value of $p \gg 0.1$.

To double check for possible motor confounding factors, EEG data from the online trials were visually inspected at the end of each experiment. For instance, subject 3 initially had an accuracy of 100%; however, visual inspection of the EEG signal allowed us to see that the subject consistently produced electromyographic artifacts in the occipital region in the high-WM-load condition. All the data were discarded, and the subject repeated the experiment on a different day without occipital electrodes. The performance the second time was slightly lower but still well above the chance level (85% correct classification).

2.2.3 Neurophenomenological validation

A total of 92 trials were analysed, with the subjects providing the correct answer 82% of the time. The data for individual subjects are summarized in Table 2.1.

Subject	Trials not answered	Noisy trials removed	Total trials analyzed	Correct answer
1	1	1	18	83%
2	3	0	17	76%
3	0	0	20	85%
4	0	0	10	100%
5	1	0	19	73%
6	7	5	8	75%
TOTAL	12	6	92	82% ($p < 0.0001$)

Table 2.1: Percentage of correct answers, per subject, to the question for assessing whether the feedback was sham or real. Artifacts trials and trials where subjects did not answer were not considered.

2.2.4 Temporal behaviour of the WMLE

Half the subjects had a stable WMLE. Figure 2.6 shows the average over the 20 trials of one of these subjects during Task 2. One can observe that the WMLE begins to decrease systematically after 10 seconds. It is important to remember that, during the first 8.5 seconds, the subjects performed mental arithmetic. Afterward, following a visual cue, the subjects stopped the mental arithmetic. The observed behaviour is consistent with the WM-load switch expected at 8.5 seconds, plus the BCI delay. The length of such a delay is less than 2.5 seconds, as the WMLE at time t_0 considers all the EEG activity that took place between $t_0 - 2.5$ and t_0 . After reaching the lowest value, the WMLE systematically increases again, this time possibly due to the feedback information being processed by the subject. Subjects at this point were still processing information, while performing the comparison between the WMLE and their subjectively estimated WM load. Indeed, the new values are relatively high; however, not as high as they were in the

first part of the task.

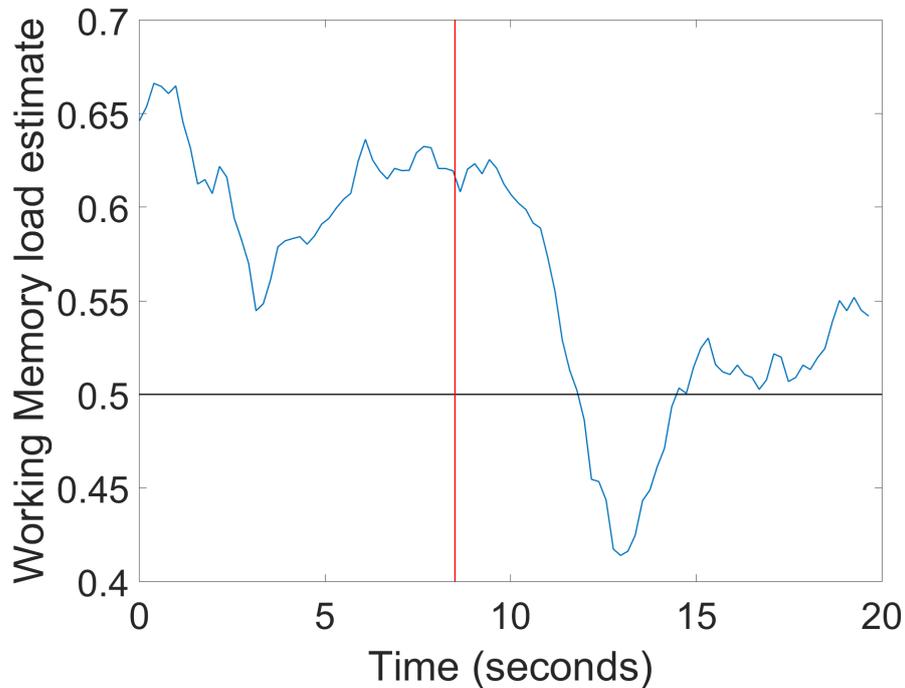


Figure 2.6: Average over trials of the WMLE time evolution of a typical “good” subject

For the other half of the subjects, the behaviour was not as stable across trials, and the averages across the trials were flattened, suggesting no systematic behaviour. Nevertheless, even for these subjects, there was a high rate of correct answers, which means that the WMLE successfully matched their subjective perceptions of WM load.

2.2.5 The EEG biomarkers: Spectral EEG changes due to WM load

After ensuring the reliable single-trial estimation of the WM load, it is useful to go back to the question of what changes are induced in the brain

due to WM activity. For a visual representation of these changes, we can afford windows of 10 seconds instead of 2.5 seconds. Shorter windows are useful for a low-latency system, while larger windows allow a more accurate spectral decomposition of the signal.

The power at a certain band for a given channel is a potential biomarker. We computed the grand average across trials and across subjects for each biomarker. Changes across conditions of this grand average are displayed in Figure 2.7, using all the artifact-free trials. A lighter colour corresponds to biomarkers that had higher values, on average, in the high-WM condition, whereas a darker colour corresponds to biomarkers where the average values were lower in the high-WM-load condition. We only used biomarkers conveying useful information for WM prediction. Due to our multivariable approach, we are not interested in biomarkers that were statistically different across conditions (see section 4.1 for more details on why this may not be informative). Instead, we are interested in biomarkers that, when combined, produce patterns that can be identified as typical low-WM or high-WM activity. To determine how many biomarkers are relevant, we ranked them with the Orthogonal Forward Regression (OFR) feature (in this work, the term “feature” is used interchangeably with “EEG biomarker”) selection technique (see section 2.3.2.1) and added them to the model one by one until performance decreased or did not increase significantly.

The biomarkers that were the best predictors of WM load were the following:

- Relative lower beta power, electrode Fp1;
- Relative lower beta power, electrode Cz;
- Lower gamma power*, electrode Fp1;
- Relative upper beta power, electrode Cz;
- Alpha power, electrode Oz; and

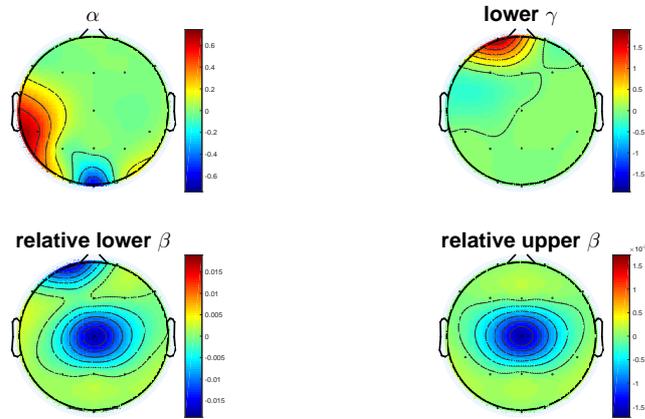


Figure 2.7: Mean difference between the high- and low-WM-load conditions for the relevant biomarkers. Light-coloured values are typically higher in the high-WM-load condition. Dark-coloured values are typically lower in the high-WM-load condition.

- Alpha power*, electrode CP5.

Biomarkers indicated with a star (*) increased with increasing WM load, while the others decreased with increasing WM load.

2.3 Additional material

2.3.1 Images used for Task 1

Table 2.2: Figures used to determine the memory span

Geometric Shapes	Fruits	Landscape
circle	apple	house
pentagon	banana	building
square	orange	church
rhombus	pear	castle
cross	grape	bridge
star	watermelon	tower
triangle	pineapple	tree

Table 2.3: Figures used for the test

Animals	Vehicles	Supplies	Clothes
cat	plane	book	trousers
deer	train	scissors	shirt
dog	skateboard	pen	hat
elephant	truck	ruler	shoes
penguin	car	backpack	socks
snake	ship	compass	belt
turtle	bicycle	set square	tie

2.3.2 Design and training of the BCI

2.3.2.1 Offline study: Designing the BCI

Twenty subjects performed Task 1 offline, which allowed us to collect a WM database and design a classifier. In our case, designing the classifier meant choosing appropriate parameters (P_1, P_2, \dots), as explained below. In general, the classification process takes an EEG epoch as input, extracts meaningful features sensitive to the cognitive function being classified, and gives as output the probability (WMLE) that the EEG epoch analysed belongs to the high-WM class. More specifically, the offline analysis took place as follows:

1. Subjects performed Task 1 while wearing the EEG set. Frequencies below 1 Hz and above 45 Hz were removed from the EEG signal with a third-order Butterworth filter. The EEG data were segmented into epochs of P_1 seconds. Each epoch was visually inspected, and all the epochs contaminated with noise or muscular artifacts were rejected. Especially, epochs with eye blinks or arousal flags were rejected. An arousal flag was placed on an epoch if either a distracter or the target was displayed during its course. During the preliminary tests, the subjects had reported an arousal effect due to the appearance of distracters or targets;

2. For each epoch, and for each channel, spectral features were extracted using the Matlab p-Welch function, with a Hamming window of 0.5 seconds. The spectral features were absolute and relative power in the following bands: delta (1 – 4 Hz), theta (4 – 8 Hz), alpha (8 – 12 Hz), lower beta (12 – 20 Hz), upper beta (20 – 30 Hz) and lower gamma (30 – 45 Hz). The relative power in a band is the fraction of the total power in that band. Normalising the latter has the advantage of reducing inter subject variability. With 16 channels, two features per band, and six bands, we obtained 192 features for each epoch;

3. For each subject, we performed cross-validation using data from the other subjects and P_2 epochs of the current subject as calibration data. We expanded the calibration data by adding noisy copies of the original data with noise parameters P_3 and P_4 (see section 2.3.2.3). At this point we had a $(192, M)$ matrix of features, where M represents the number of epochs and an M -vector of binary labels (low-WM load or high-WM load). The number M depends on the parameters P_1 , P_2 and P_3 ;

4. To select relevant features, OFR [108] was performed on the above matrix, the best P_5 features were kept. OFR is a linear regression technique that can be used as a supervised feature selection approach. In the first step of OFR, features are ranked in order of decreasing correlation to the classifier output; the first selected feature is the top-ranking feature. Spatial filters could have been used at this step instead of OFR, however as we were interested in complex combinations of biomarkers (such as for instance cross-frequency couplings between pairs of channels), we preferred this more exhaustive approach. In the second step, all remaining features, as well as the output, are orthogonalized with respect to the first selected feature, thereby discarding the part of the output that was explained by that feature; the projected features are ranked in order of decreasing corre-

lation to the projected output, and the top-ranking feature is selected. Orthogonalization, ranking and selection are iterated until P_5 features are selected; and

5. Finally, a linear discriminant analysis (LDA [41]) classifier was trained with the selected features and their corresponding labels (high- or low-WM load). The output of the classifier is the WMLE.

Epochs containing eye-blinks or with arousal flags were not included in the training set to obtain clean markers. Nevertheless, they were included in the testing set, both online and offline.

A specific set of parameters (P_1, P_2, P_3, P_4, P_5) represents a potential BCI design. The set of parameters that represented the best trade-off between classification performance and feasibility was chosen to build the BCI. The values are shown in the table below:

Table 2.4: Final set of parameters for building the BCI

Parameter	Description	Value
1	Epoch length	2.5 seconds
2	Calibration epochs	45 per class
3	Subject weight	65%
4	Noise level	1.5
5	Number of features	8

2.3.2.2 Online analysis: From EEG recordings to a WMLE

The online experiments begin with a calibration step. The goal of the calibration is to train a classifier customised to the subject, but still using the information of the previous 20 subjects, to obtain robust and reliable results.

Nine new subjects were recorded for online testing. The EEG data were filtered from 1 to 45 Hz. Subjects were asked to perform Task 1 until we had

collected 45 (P_2) clean epochs for each condition. The criteria for considering an epoch as clean included being free from muscular artifacts, eye blinks, and the arousal flag. Epochs with a visible excess of electric noise were also discarded. The epochs lasted 2.5 seconds (P_1). Features were extracted from the 90 epochs, the feature matrix was expanded (using P_3 and P_4) as described in section 2.3.2.3, and it was added to the existing offline feature database. With the new, subject-customised database, we performed OFR, selected the best eight (P_5) features, and trained an LDA classifier.

Figure 2.8 shows how the information from the previous section (offline study) was integrated with that in this section (calibration) to select a good set of features and design and train the classifier.

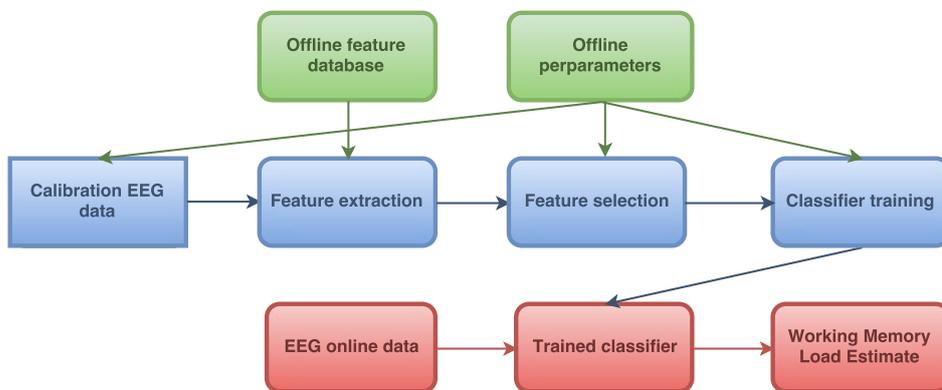


Figure 2.8: Design methodology

At this point, the BCI is ready to provide a WMLE. A continuous stream of data is analysed, and a sliding window including the last 2.5 seconds of EEG is used as input.

2.3.2.3 Cross-validation and calibration

Cross-validation (see [71] for a more detailed description) aims to estimate the performance of a trained model when presented with new data, to select a model that has the appropriate complexity given the available data. One way to achieve this, the leave-one-out method, involves removing one example (one epoch in this case), training the classifier with the remaining examples, and predicting the true class of the removed example. The process is repeated for all the examples, and the expected error can be computed. When we test a single model, an LDA in our case, we simply estimate the average error we would obtain when facing new data. When performing cross-validation using different models (a linear classifier, quadratic classifier, artificial neural network, etc.) we can choose the one that best captures the complexity of the data (thereby minimising the cross-validation error).

Some caution is required here in the context of EEG. Due to subject idiosyncrasies, the features of different epochs of the same subject tend to be highly correlated; however, they are not necessarily correlated to the features of another subject's epochs. Assuming that, during the offline analysis, N epochs of a specific subject are available, performing leave-one-epoch-out would answer the question, "What is the generalization error if the classifier is trained on the data from the rest of the subjects *plus* $N - 1$ epochs of the current subject?" The performance computed in such a way is likely to be an overestimation, as $N - 1$ epochs from a subject may not, in general, be available during a real-time experiment.

Removing the subject entirely would lead to the underestimation of the performance. This is because, in a typical online experiment, we collect data from the individual before we use the BCI to customise it to the subject (calibration). The number of calibration epochs required for an acceptable performance is a parameter itself, referred to as P_2 in this case (see section 2.3.2.1).

To summarise, for a specific left-out epoch, we train the classifier with P_2 epochs from the same subject and all the remaining data from the other subjects. We test our accuracy on predicting that epoch and repeat the process for all the available epochs. The result is an estimate of the classifier’s performance when using P_2 epochs as calibration data.

One problem arises here. The larger our offline database, the less influence our new subject has on the final classifier. The approach taken was to add noisy copies of the subject feature matrix to balance the database. The idea of expanding a dataset with corrupted copies has been developed before [103, 115] in other fields. The subject database was expanded with these copies to a sufficiently large size so that the data associated with the subject (subject weight) represented P_3 percent of the total database.

For each feature, the noise added to the subject matrix was Gaussian noise with zero mean and standard deviation equal to P_4 times the standard deviation of the feature. The noise was added directly to the features.

2.3.3 Arithmetic operations in Task 2

A random sequence of digits, d_1, d_2, \dots, d_n , was presented to the subjects in each trial. Three possible ways of manipulating the digits were suggested to the subjects, who were asked to choose the one that felt more resource demanding for them, as follows:

- Progressive multiplication. Multiply $d_1 d_2 \dots d_i$ until the time is over;
- Pairwise multiplication and successive addition. Multiply d_1 and d_2 and store the result. Add the result to the product of d_3 and d_4 , replace the result. Add the result to the product of d_5 and d_6 , replace the result. Continue until the time is over; and
- Free choice. Subjects comfortable with their arithmetic skills were left to choose the structure of the operations, provided they maintained a

high level of use of their mental resources.

2.3.4 Estimation of statistical significance

We developed a method of analytically estimating the statistical significance of a two-class classifier's performance. The null hypothesis is that the results come from a random classifier, whereas the alternative hypothesis is that the classifier is based on informative features. The first step towards estimating significance, then, is to choose a random classifier and to determine its success rate. Let us assume that our dataset consists of N examples, with N_1 examples of class 1 and $N - N_1$ examples of class 2, with $N_1 \geq N/2$. The best that a random classifier can do is to consider the classes' prior probabilities. Denoting by q the prior probability of class 1, assumed to be larger than 0.5, and estimated by N_1/N , a possible random classification rule is to assign any object to class 1 with probability q . The probability of correct classification of this classifier (which can be estimated by its rate of correct classification) is given by: $c_0 = q^2 + (1 - q)^2$. Let us define a random variable where the realisation z_i , for example, i , is

$$z_i = \begin{cases} 1 & \text{if the random classifier classified example } i \text{ correctly} \\ 0 & \text{otherwise.} \end{cases}$$

The total number of successes, $Z = \sum_{i=1}^N z_i$, follows a binomial distribution $Z \sim B(N, c_0)$, and hence, the probability of obtaining exactly k successes is

$$Pr(Z = k) = \binom{N}{k} c_0^k (1 - c_0)^{N-k}$$

By definition, a p -value is the probability of obtaining results at least as extreme as the observed ones, assuming that the null hypothesis is true. Our goal is comparing a random classifier with a specific classifier that yields c correct answers. In this case, "results as extreme" means observing

at least c correct answers in a random classifier. Therefore, the p -value associated with the null hypothesis defined above can be computed as

$$p = \sum_{k=c}^N Pr(Z = k)$$

In general, the use of any other random classifier would lead to a different c_0 . Especially, the most efficient classification rule under a complete lack of informative predictors is the *zero classifier*, which assigns all the objects to the largest class. In the above notation, the rate of correct classifications of a zero classifier is $c_{0z} = q$. As, by definition, $0.5 < q < 1$, it is easy to show that $c_{0z} > c_0$ for all q . However, although the zero classifier is the best classification rule when no relevant predictors are available, for a zero classifier, $Pr(Z = k) = 0$ if $k \neq N_1$ (by definition, a zero classifier can correctly predict only N_1 objects). Therefore, for any classifier with a number of correct predictions larger than N_1 , p would be zero. It is a good practice to compare classification results with those of a zero classifier when facing imbalanced datasets. However, a zero classifier is not useful for computing p -values.

A FRAMEWORK FOR INVESTIGATING THE SPATIO TEMPORAL BRAIN ORGANIZATION UNDER COGNITION

The present chapter involves the development of a data-driven framework for investigating how cognition affects the spatio temporal properties of brain state switches.

3.1 Data acquisition

As mentioned in chapter 1, three datasets were used to investigate whether the findings were WM specific or a general property of cognition. We used WM, Alzheimer's Disease (AD) and emotions-related datasets. All recordings performed by us followed the principles outlined in the Declaration of Helsinki. All the participants were given explanations about the nature of the experiment and signed an informed consent form before the experiment started.

3.1.1 WM dataset

Twenty subjects performed a WM task, as described in [99]. The dataset consisted of 530 trials lasting 10 seconds each. Of these, 281 correspond to a low WM load and the rest to a high WM load. All the artifacted trials were discarded, and eye blinks were removed with Independent Component Analysis [20]. The sampling rate was 500 Hz, and 16 channels of the international 10–20 system were used, as follows: Fp1, Fp2, F7, F3, Fz, F4, F8, Cz, CP5, CP6, P3, Pz, P4, O1, Oz, and O2. The WM dataset was the same as in Chapter 2.

3.1.2 AD dataset

We used the same dataset as [119]. Recordings from 61 subjects were collected, 23 of which were AD patients, and the rest were healthy, age-matched controls. For each subject, 20 seconds of continuous recordings were available. The data were sampled at 200 Hz, using 21 channels, as follows: Fp1, Fp2, F3, F4, C3, C4, P3, P4, O1, O2, F7, F8, T7, T8, P7, P8, Fz, Cz, Pz, FPz, and Oz. Each 20-seconds recording was divided into eight epochs. In total, 488 epochs were analysed, 184 corresponding to AD patients and the rest to healthy controls.

3.1.3 Emotions dataset

We used the processed version of the DEAP dataset for emotions [70]. Thirty-two participants watched videos and rated them based on valence, arousal and dominance, on an integer scale from 1 to 9. EEG data were recorded while the subjects watched the videos, and our objective was to classify valence in two classes, one corresponding to the first half of the range of the scores and the other class corresponding to the second half. The data were sampled at 128 Hz, using 32 channels, as follows: Fp1, AF3, F3, F7, FC5, FC1, C3, T7, CP5, CP1, P3, P7, PO3, O1, Oz, Pz, Fp2, AF4, Fz, F4, F8, FC6, FC2, Cz, C4, T8, CP6, CP2, P4, P8, PO4, and O2. Electrooculographic

artifacts were removed as indicated in [70]. The original 60-second trials were divided into 10-second epochs, which gave rise to 7680 epochs, 3876 of which corresponded to negative valence.

3.2 The framework

To be more specific about what we mean by the spatio temporal organization of brain dynamics under cognition, we can imagine dividing the brain into N regions, r_1, r_2, \dots, r_N . We consider the state of region r_i at time t to be $s_i(t)$. Whenever $s_i(t_1) \neq s_i(t_2)$, at least one *switch* is said to have occurred for a given t in the interval $[t_1, t_2]$. So far, we do not know what these states are, but we want them to represent a specific operation; for the sake of concreteness, $s_i(t)$ may be, for instance, *registering the colour red in brain region i at time t* . A more specific description of the variables proposed to characterize these states is provided later in this section. Meanwhile, we can have a graphic intuition of the spatio temporal dynamics. Suppose, for simplicity, that we study only three regions, time is discrete and region one has access to states A, B and C ; region two has access to states D, E and F ; and region three has access to states G, H and I . Now imagine two different cognitive conditions (see the Methods section for more details about what we call cognitive conditions), for instance, a low WM load and high WM load. If we follow the dynamics for a few time steps, we could observe something similar to what is depicted in Figure 3.1.

Evidently, in the brain, the number of accessible states is not necessarily finite, and time is not necessarily discrete. Nonetheless, this toy example allows us to observe the behaviour we expect to capture. We can follow any of the regions over time, let us take r_1 . At the boundary between conditions, a dynamical change occurs, the succession of states becomes slower, and more states become available (especially state C). In addition, while r_1 is engaged with r_3 in condition 1 (they have similar dynamics), in condition 2, it disengages from r_3 and engages with r_2 .

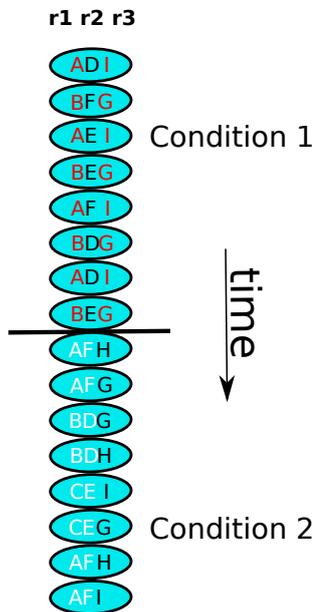


Figure 3.1: Simplified illustration of the spatio temporal organization that we expect to capture. There are three regions, r_1 , r_2 and r_3 . Each region has three available states. If we follow the state of r_1 over time, we will observe that, not only did the temporal behaviour change when condition 2 started, but r_1 also engaged in joint activity with a different region.

We ask the following questions:

1. Do regions experience state switches more often in one cognitive condition than in the other? Do state switches occur in a discrete manner, or continuously?
2. Are pairs of regions more or less engaged depending on the cognitive condition?
3. Is there some underlying criticality affecting state switching dynamics? Do critical parameters depend on cognition?
4. Does the cognitive condition affect the available states and the time spent in each state?

In the context of this work, the first question arose while studying WM, and it led to the others. Refreshing items in WM is a fundamental component of WM [95]. Under some models [18], the central executive is engaged alternatively in refreshing and processing, and its limited capacity has a direct effect on what we call the WM load. If there is a physical, bounded region that performs the activities of what we call the central executive, it may be conjectured that its dynamical regime will change to support both tasks. The current study is based on preliminary results presented in [84]. While conducting WM experiments, we collected subjective evidence that refreshing occurred at a higher pace in the high-WM condition compared with the low-WM condition. As mentioned in the introduction, the motivation behind questions 1, 2 and 3 is to explore whether cognition affects brain state switching—its dynamics, the spatial coordination across brain regions, and possibly any critical parameters. In contrast, question 4 is rooted in the concept of ergodicity, an important notion in dynamical systems. With this question, we are interested in knowing whether the number of states available to the brain and the time spent by brain regions in certain states, change depending on the cognitive conditions.

A way of addressing these questions is constructing spatially localized variables that represent the time evolution of the states we want to study. We then derive measurable properties from the above questions.

To construct the desired variables, let us start by assuming metastability. As mentioned in the introduction, by definition, successive state switches take place in metastability, and each state can be considered stationary. We take EEG as our brain imaging method and consider that, for the region being recorded with a specific EEG channel, a stationary EEG segment will be recorded during the time course of a state. Therefore, non-stationarities can help us track brain state switches.

In this paragraph, we describe how to generate a time series from the EEG signal, that will hopefully represent the region's underlying state. For

a stationary process, all the statistical properties remain constant over time. As any (constant in time) statistical distribution has constant statistical moments, at any time t , we can compute the statistical moments over a short window¹ centred at t . Changes in the underlying distribution will induce changes in its moments. As a graphical example, take a random variable drawn from a normal distribution $\mathcal{N}(-3, 1)$. After a certain time, $t_0 = 0$, the underlying distribution changes to $\mathcal{N}(2, 3)$, as we observe in Figure 3.2a. Using a sliding window, we can compute the estimation of the first two statistical moments in figure 3.2b, where the mean is shown in blue, and the variance in red. The time series of the estimation of the mean and variance can be considered a two-dimensional trajectory, as shown in Figure 3.2c. In this two-dimensional space, a point is a state, and evidently, there is estimation noise arising from the fact that the window has a finite length. If switches occur in a discrete manner, fluctuations in Figure 3.2b at times other than t_0 are artifacts due to this noise only. In section 3.3 we show that as these fluctuations correlate with cognition, they are not only estimation noise, thus, the assumption that switches occur at discrete times must be questioned.

To summarize, we consider the EEG signal to be generated by an underlying process that changes its statistical properties when a brain state switch occurs. To study these changes in the statistical properties, for a given channel, at a given time t , we compute $v(t)$, $s(t)$, and $k(t)$ —the estimations of the variance, skewness and kurtosis² over a short time window centred at t . The length of the window depends on the frequencies to be investigated, as it is explained in the methods section. Given the above considerations, and the fact that we are not using the infinite set of statistical moments, but rather a subset of size three, we hypothesize that the vector $(v(t), s(t), k(t)) \in \mathbb{R}^3$ can be seen as a third-order estimation of a multidimensional feature sensitive to brain state switches. For each chan-

¹An ensemble of identical brains is clearly impossible; therefore, by taking a time window, we assume that the process generating the signal is ergodic.

²The EEG time series have been detrended via high pass filtering, so we do not estimate the mean.

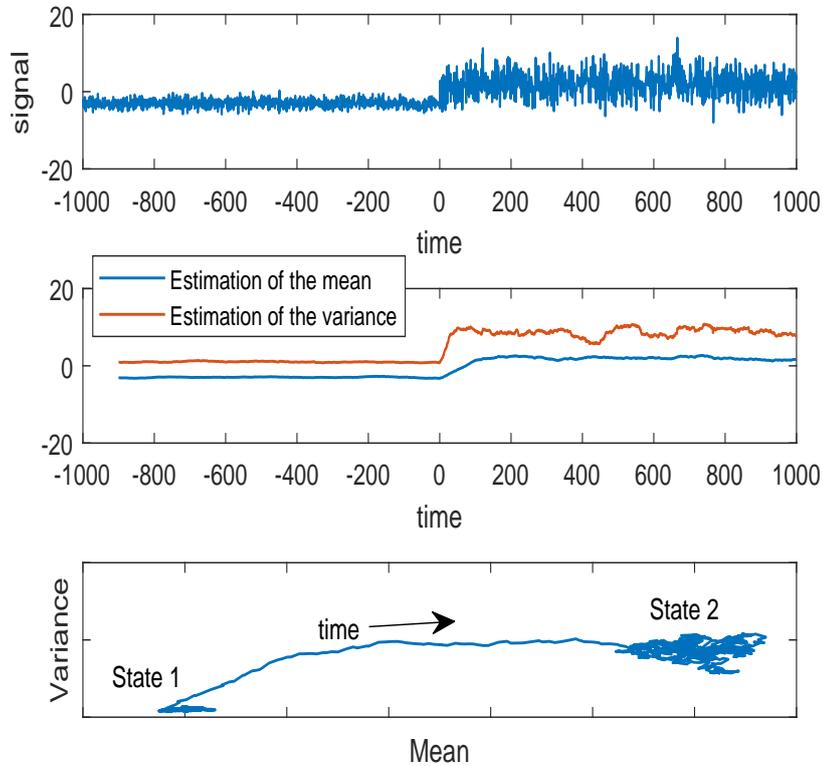


Figure 3.2: Top: Synthetic signal. The underlying statistical distribution changes at time $t = 0$. Middle: Time series of the evolution of the estimation of the two statistical moments over time, when using a window of 50 time points. Bottom: Process viewed as a trajectory in the space of states. Each dimension of this space is the estimation of a statistical moment.

nel, if we compute the third-order estimation over a sliding window, we can create a (vector) time series, that we will call the channel state surrogate time series ($\mathbf{CSS}(t)$, or simply CSS). It is expected to be an indicator of the time evolution of brain states in the region of the cortex whose activity was inferred by the scalp recordings. We suppose, as in [122], that changes in EEG stationarity are already visible in the first statistical moments. By considering n moments, each CSS is a trajectory in an n -dimensional space, and we expect to test whether cognition affects these trajectories' spatio

temporal structure. As the signal is sampled at a constant rate, and the window size remains constant over time, if we define *neighbouring* states as states with similar statistical moments, the norm of the velocity of the CSS (the speed time series), $\|\mathbf{vel}(t)\| = \left\| \frac{d}{dt} \mathbf{CSS}(t) \right\| = \sqrt{v'(t)^2 + s'(t)^2 + k'(t)^2}$, is proportional to the distance between states (or switch size) if switching occurs continuously.

Once a biologically plausible way to numerically characterize states has been developed, we can return to our questions. All the questions, except that concerning discreteness or continuity, can be rephrased as, “*Does cognition affect property X?*” In the next section, we derive a measurement or feature (feature being taken here in its machine learning sense, not to be confused with the feature binding problem discussed above) from each property X . Then, we study the potential of each of these properties to correctly classify cognitive conditions.

Regarding the question of continuity versus discreteness, we assume discreteness and provide inconsistent evidence in the following way: We only keep the switches considered spurious (due to estimation noise) if discreteness holds, and assess whether we are still able to classify cognitive conditions with performance above random classification.

For methods 1, 2 and 4 in the methods section, before computing the CSS, the EEG signal spectrum was segmented into the usual physiological bands: delta (1–4 Hz), theta (4–8 Hz), alpha (8–12 Hz), lower beta (12–20 Hz), upper beta (20–30 Hz) and lower gamma (30–45 Hz). For bandpass-filtered data, we computed the CSS using a sliding window of length $L = 1/f_{min}$ seconds, where f_{min} is the minimum frequency of the corresponding band, so that each window contains at least one full oscillation of the smallest frequency. Method 3 did not involve bandpass filtering, as its purpose was estimating the shape of the CSS power spectrum. For method 3, the window length was 0.1 seconds, considering the length of stationary EEG segments reported in the literature [69].

The goal is to test the hypotheses that certain measurable quantities, or features, convey information about cognition. After feature extraction, relevant features are selected by OFR [108], and then feed to an LDA [41] classifier. The task of the classifier is to discriminate between cognitive conditions, namely high versus low WM load in the first dataset, AD versus control in the second dataset, and positive versus negative valence in the third dataset. If, after cross-validation, the performance of the classifier is better than that of a random classifier, we can conclude that the features carry information about cognition. The results of the process are the classification performance estimated by cross-validation, and the set of most informative features.

Due to the imbalance of the AD dataset, the AUC [52] of an ROC curve was used as the measure of classification performance. An ROC curve is not influenced by the imbalance of the classes, and its AUC value is typically 0.5 for a random classifier, and 1 for a perfect classifier. Values larger than 0.5 indicate performance better than random. The statistical significance of the performance of the classifier was estimated by replacing the features with random numbers and iterating the classification process 300 times. We counted the fraction of times that the performance with the random features was higher than the observed performance. We decided not to perform a permutation test or to estimate significance via the ROC curve because we also wanted to assess the cross-validation procedure. A wrong cross-validation procedure (for instance, performing feature selection before the cross-validation loop) or multiple testing (while optimising hyperparameters) may artificially provide good classification results even if there is no correlation between the data and the output. For cross-validation using the AD dataset, when classifying a subject, the whole subject data was left out of the learning set, given that in a real-life diagnosis task, there is no available information about the subject to be diagnosed. In contrast, only half the subject data was left out for the WM and emotions datasets, simulating the calibration process common in BCIs. Half the subject data

was left in the learning dataset for calibration; however, classification was always done on EEG epochs that had not been “seen” by the classifier. The analysis was performed as in chapter 2.

Each of the following subsections is an attempt to address one of our main questions.

3.3 Method 1: Temporal structure of the switches

The motivation behind Method 1 is the study of the temporal structure of brain switches. Especially, it aims to consider whether these switches occur at discrete or continuous times and to what extent cognition affects such continuous or discrete dynamics.

A way of collecting evidence supporting either the continuous or discrete hypothesis is pinpointing times at which switches occurred. If switches are spaced by intervals larger than the period of the studied brain rhythms, then switching can be considered discrete. Conversely, if the intervals are smaller than the period, switching can be considered effectively continuous. Therefore, estimating the number N of switches per second can shed some light on whether the switching is discrete or continuous.

By definition, the switching points will be those at which the statistical properties of EEG change. The central part of Figure 3.2b shows how the statistical moments estimated over a window behave near a switch. Due to the sliding window estimation, the presence of an instantaneous change of the statistical properties of the signal results in a gradual change of their estimates. The transient’s length should be equal to that of the sliding window. Real-world measurements are noisy, and there should be small discontinuities in the estimation of the statistical moments, but we

expect these noise-related discontinuities to be small compared with the state-switch-related discontinuities.

To estimate the number of switches occurring per second, it is possible to consider the speed time series, in which the above-mentioned transients give rise to non-zero points, which are informative points. Let us denote by N the mean number of switches per second. We assume that we choose a window where the duration L is smaller than $1/N$, representing the mean inter-switch duration (the mean time spent in a particular state); that is, $NL < 1$. Consequently, a window contains at most 1 switch at any given time. If this switch is present in the window, it produces non-zero points during the window's whole duration. As the sampling frequency is constant, the ratio ρ of the number of non-zero (informative) points to the total number of points is equal to the ratio of the window duration to the inter-switch duration; therefore, $\rho = L/(1/N) = LN$. Alternatively, by estimating ρ experimentally, we can infer the value of N , as explained in the next paragraph. The size of the window should be smaller than the inter-switch duration, but large enough to contain at least one full oscillation of the smallest frequency of the band investigated. Therefore, the minimum acceptable value of L is $L = 1/f_{min}$, as stated in the previous section.

For estimating ρ , the speed time series was sorted by amplitude. Only the fraction F of the samples for which the velocity was smallest was used to compute the mean speed, and the mean speeds were used as features for classification. The classifier's performance was studied as a function of F . For small values of F , we would only select noise under the assumptions of discreteness (and hence, piece-wise stationarity), enough temporal resolution ($NL < 1$), and transition size larger than noise. As F increases, if the above assumptions are true, the classification performance should start gradually improving and be better than random classification above a certain threshold value. This threshold value must coincide with $1 - \rho$, the fraction of non-informative points: For values of F larger than $1 - \rho$, there is a non vanishing probability that some points selected to compute the

mean are informative points. If this threshold value does not exist, that is, if all the points are informative regardless of the size of the transition that originated them, we can argue that cognition-related brain-state switches occur at intervals smaller than the studied periods, and therefore, they can be thought of as continuous. As ρ is not defined in the continuous case, neither is N .

For each channel and each band, Method 1 consisted of using the parameter F that maximises the performance of the classifier under cross-validation. In addition, the classification performance as a function of F can be used to estimate, if it exists, ρ , and hence N .

To compare how small transitions behave as compared with large transitions, we also studied the performance of the classifier when F represents the fraction of the samples for which the velocity was largest.

3.3.1 Results specific to Method 1

Figure 3.3 shows the performance of the classifier using Method 1, as a function of the fraction F , for the WM dataset when using only the upper beta and lower gamma bands (combined, for simplicity, to use a single window. These bands were chosen to investigate high frequencies and therefore be able to use small sliding windows). The window size used was $L = 1/20$ seconds, or 25 points. Only high frequencies were considered so that a small sliding window could be used, and only the WM dataset was employed for this figure due to its high sampling rate. The blue dots correspond to the classification performance obtained using only the fraction F of the transitions that had the smallest velocities. For the red dots, we used the fraction F that had the largest velocities. A schematic representation of what would be expected under piecewise stationarity is also displayed. For every N , there is a threshold value $1 - \rho = 1 - NL$ of F ; below this threshold value, the classifier performs as a random classifier ($AUC = 0.5$); beyond

3.3. METHOD 1: TEMPORAL STRUCTURE OF THE SWITCHES

this value, the performance of the classifier increases until it reaches its maximum value. It is clear that the experimental results are not consistent with a discrete model. The latter is discussed in the next section, as well as in the general discussion in chapter 4.

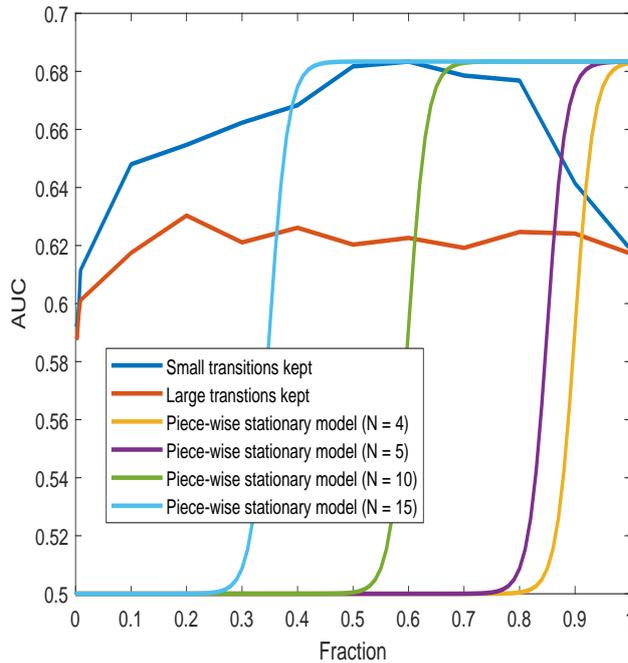


Figure 3.3: WM dataset using only the upper beta and lower gamma ranges. The mean speeds of the CSSs were used as features. To compute the mean, the smallest (blue) and largest (red) F fraction of amplitudes were used. The classification performance is studied as a function of F . The solid lines are diagrams of what should be expected under piecewise stationarity, for different numbers N of switches per second.

Figures 3.4–3.6 show the performance of the classifier, using Method 1, as a function of the fraction F when using all the bands, for the three datasets. For each band and each channel, there was a time series where the amplitudes were sorted by size; hence, the total number of features of the classifier in these cases equals the number of channels multiplied by the number of bands.

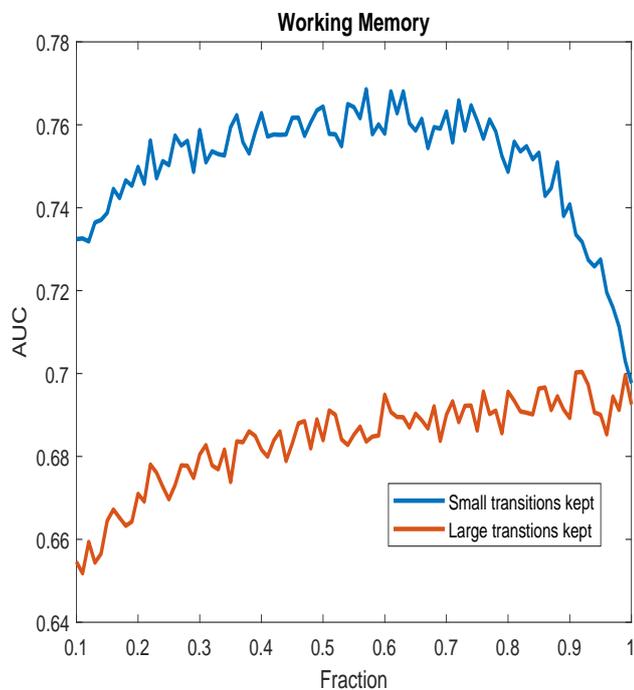


Figure 3.4: WM dataset. Performance of the classifier as a function of the fraction of transitions kept. In blue, keeping small transitions only, in red, keeping large transitions only.

3.3.2 Discussion specific to Method 1

In the literature, one of the motivations for proposing switching at discrete times is the stability required for sustaining oscillations. Thus, we also proposed that assuming discreteness implies assuming that the duration of a state should be larger than one oscillation; consequently, we studied the dynamics with windows as small as one full oscillation. After all, discreteness at intervals smaller than the effective resolution is difficult to falsify.

It seems clear that the spatio temporal structure of the statistical properties of EEG carries information about cognition. This evidence supports the claim that the proposed variables correspond to brain states driven by

3.3. METHOD 1: TEMPORAL STRUCTURE OF THE SWITCHES

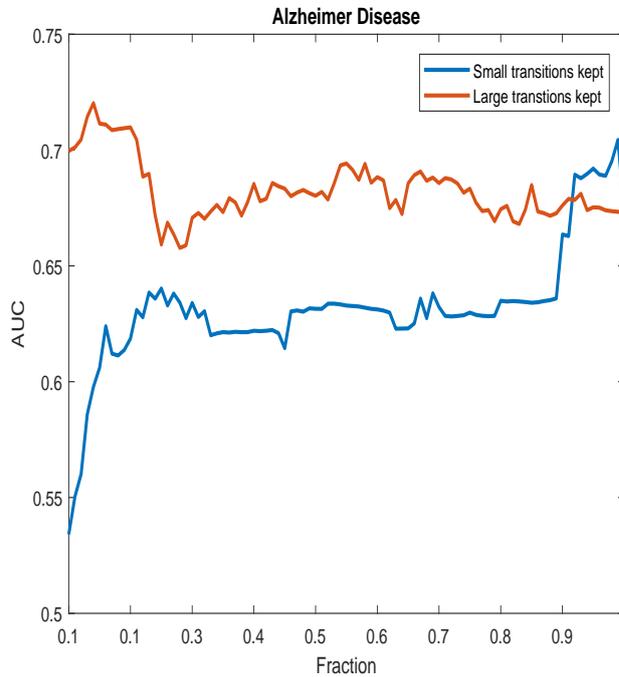


Figure 3.5: AD dataset. Performance of the classifier as a function of the fraction of transitions kept. In blue, only the small transitions were kept, in red, only the large ones.

cognition. In contrast, as mentioned in the introduction and in the above paragraph, theoretical considerations led other researchers to postulate discrete timing. Based on this hypothesis, they found evidence that large transitions between states, as observed by large changes in the EEG properties (rapid transition processes or phase resetting events), correlate with cognition. In this chapter, we provide evidence showing that not only do small transitions convey information, but indeed, they seem to carry more information than large ones under certain circumstances. In Figure 3.3, which concerns only the WM dataset, we use only high frequency bands to be able to use a small sliding window (a single window of 1/20 s, as both the bands were merged into one). The classification results are 11% lower than when using all the bands (see Table 3.3); however, the image reveals an interesting behaviour that remains when all the bands are used. The figure shows how the classification performance varies as a function of the fraction

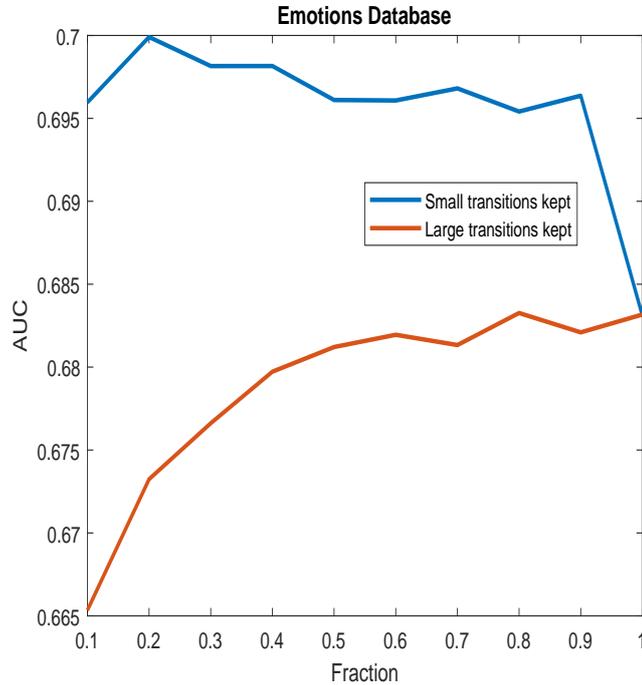


Figure 3.6: Emotions dataset. Performance of the classifier as a function of the fraction of transitions kept. In blue, small transitions; in red, large transitions. Due to the large size of the dataset, the performance of the classifier was computed for a smaller number of values of F as compared with the other datasets, and hence the graph is less smooth.

F . The blue curve reflects the behaviour when the fraction F concerns the smallest transitions, while for the red curve, the fraction F pertains to the largest transitions. We can observe two things. First, in the left part of the blue curve, the classification performance is already above random, and therefore, small transitions are not noise. Second, and more importantly, the blue curve is always above the red curve, which means that given any fraction F , it is always more informative to take the F smallest transitions than it is to take the F largest transitions. As we gradually increase F following the blue curve, we select larger transitions and classification performance increases, up to the point at which performance starts decreasing when we add larger and larger transitions.

In contrast, in a discrete, piecewise stationary model, we should observe something similar to the solid lines. For small values of F , the classification performance should be random (AUC near 0.5) because we are only selecting noise. As F increases, we should find a value at which the classification suddenly increases, as we start considering transitions that correspond to actual switches and not spurious changes. With N switches per second, the expected value of F at which the classification performance starts increasing should be $1 - \rho = 1 - NL = 1 - N\frac{1}{20}$. In the figure, it is clear that it is only as N tends towards 20 that we obtain a discrete model in which all transition sizes are useful. However, it is precisely for $N \geq 20$ that we can not resolve switches with a window of $1/20$ s, and states would last less than a full oscillation of the targeted band. Thus, our results seem to contradict the discrete model. We decided to focus only on high frequencies aiming at contrasting the results with the reported lengths of stationary segments, usually larger than $1/20$ s. Figure 3.4 reveals that this behaviour is even more obvious when considering all the bands. Figure 3.6 shows a similar mechanism for the emotions dataset. The exception was the AD dataset, as shown in Figure 3.5, suggesting that large transitions are more relevant for diagnosing AD. The latter is compatible with findings of a “slowing” of the brain rhythms found on AD patients, for which a decrease of the alpha power and an increase in the delta and theta power (slow rhythms) has been reported, as compared to healthy controls [104]. With dominant slower rhythms, it would be expected that most of the information is carried by slow transitions.

3.4 Method 2: Spatial synchrony between states

If regions engage and disengage in joint activity depending on cognition, we can compute synchrony between pairs of CSSs and test whether synchrony values are different for different cognitive conditions. The CSS is three-dimensional, and therefore, the norm of the CSS was used as a proxy-

CSS (pCSS) to have a one-dimensional representation of the state. For each EEG epoch, we computed the mutual information between pairs of pCSSs, and these values were used as features for building the classifier. Having six series per channel, and 16 (WM dataset), 21 (AD dataset) or 32 (emotions dataset) channels, the potential number of combinations is in the order of several thousands. To prevent overfitting, only the CSSs where the time derivatives provided the best features for Method 1 were considered for measuring the spatial synchrony in terms of their mutual information estimation. The number of combinations considered was set by cross-validation, but it was required to be lower than 10. While Method 1 was meant to capture regularities in the temporal structure of the proposed variables for each condition, Method 2 aimed to capture the spatial structure by using mutual information between selected pCSSs as features.

3.5 Method 3: Power law of the power spectra

Power laws are not sufficient to guarantee criticality [88]; however, scale-free behaviour, like power laws, emerges from self-organized criticality. If the brain is indeed in a critical state that allows effective switching, we would expect a power law in the PSD of the pCSS. If there is some functional meaning of this power law, its properties should be affected by cognition. An important property to look at is the scaling factor of the power law, as it determines its memory properties [81]—the extent to which past events affect the present, and hence the extent to which disturbances (sensory or motor, in this context) propagate.

A power law was fitted to the PSD of the pCSS. The CSS used for the power law was not filtered in any specific band, as we are studying the whole spectrum; therefore, there is only one pCSS per channel. A linear fit in a log-log plot of the PSD of the pCSS was performed, and the slope was used as a feature. The power law hypothesis was tested using the criteria in [28].

For a given sample, the test fits the sample to a power law and generates synthetic samples drawn from the same power-law distribution. Following this, a Kolmogorov–Smirnov test is used to decide whether the real and the synthetic samples belong to the same distribution (null hypothesis). Here, we failed to reject the null hypothesis when the Kolmogorov–Smirnov test yielded a p -value larger than 0.1. It is important to note that whereas statistical hypothesis testing does not allow accepting the null hypothesis, failing to reject it means at least that the null hypothesis is a plausible explanation. The estimation of the coefficient suggested by the same article was not used, given that it provided significant, yet lower classification performance. As mentioned above, this method is meant to test whether there is cognitive-driven criticality in the proposed variables, and hence, the slope of the fit was used to feed the classifier.

3.5.1 Results specific to Method 3

The Kolmogorov–Smirnov test failed to reject the null hypothesis (power law) 98% of the time for the WM dataset, 83% of the time for the AD dataset, and 62 % of the time for the emotions dataset. The fit of a randomly chosen AD trial is shown in Figure 3.7, where x values are the amplitudes of the PSD of the pCSS.

3.5.2 Discussion specific to Method 3

It is important to remember that, for Method 3, we computed the PSD of the CSS, not that of the raw EEG. Therefore, high frequencies mean small transitions. EEG is an especially noisy signal, and for high frequencies, the noise may be larger than the signal. A power-law fit, however, allows us to infer the behaviour at the tail (large frequencies, or small transitions in this case) by studying more accessible regions of the system. In addition, the power-law hypothesis supports the claim of criticality, where information (such as sensory or motor information) is optimally transferred, as discussed

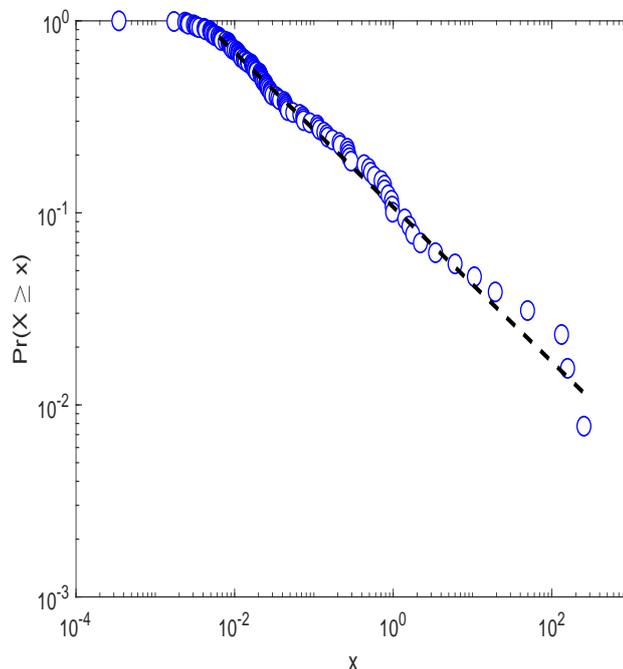


Figure 3.7: Power law fit of a randomly selected AD trial. The x values are the amplitudes of the PSD of the pCSS.

in the introduction.

3.6 Method 4: The most visited states, and how many of them are available

Having variables that represents local brain states, it is interesting to ask whether certain cognitive conditions impose a richer set of states, and whether these states are equally present.

As we are employing a dynamical framework, we can borrow the concept of phase space. For an n -dimensional system, the phase space is a $2n$ -dimensional space able to express all the possible positions and velocities of all of the n components. A point in the phase space is a specific *dynamical*

3.6. METHOD 4: THE MOST VISITED STATES, AND HOW MANY OF THEM ARE AVAILABLE

state—a specific value of all the n positions and velocities that completely captures the instantaneous dynamical properties. We will analyse each CSS separately, especially the one-dimensional pCSS, so that our phase space is two dimensional. To clarify, a dynamical state is not exactly the same as the brain states described above. Here, a dynamical state is the two-dimensional combination of the pCSS and its time derivative. In other words, it is the surrogate of the current brain state plus dynamical information about it.

We discretized the phase space in the following way: The full range of the pCSS was divided into 20 bins, with a range of 0—25. Its time derivative was divided into 20 bins, ranging from -2 to 2 . The ranges were chosen after analysing the intervals in which the pCSS and its time derivative usually fell. The number of bins was not thoroughly optimized, given that the results were robust to different numbers of bins. For each pCSS, this discretization scheme produced a 20x20 grid (see Figure 3.8) spanning the phase space. The element i, j of the grid is a dynamical state $s_{i,j}$.

With the discretised version of the phase space, we used an entropy measure to characterize it:

$$H = - \sum_{i=1}^{20} \sum_{j=1}^{20} p(s_{i,j}) \log(p(s_{i,j}))$$

where $p(s_{i,j})$ is the probability of dynamical state $s_{i,j}$, measured as the fraction of time that the system spent in dynamical state $s_{i,j}$.

The above measure is small when, for a given period of time, the dynamical system is found only in a small set of dynamical states. In contrast, it is large when the set of states is large and the probability of observing each state is similar. The values of H for each pCSS were used as features for this method.

3.6.1 Results specific to Method 4

Figure 3.8 provides visual information concerning how the dynamical states were visited for each cognitive condition. For generating the figure, we selected only the feature ranked first by the OFR algorithm for each dataset, that is, the feature that carries the largest amount of information about the output. In this case, it is the specific channel and band at which the entropy measure related to the phase space changed the most across cognitive conditions. For the WM dataset, the best feature was channel F4 filtered in the lower beta range; for the AD dataset, the best feature was channel Oz filtered in the alpha range; and finally, for the emotions dataset, the best feature was electrode T8 in the lower gamma range. For a given condition we performed the grand average with data from all the subjects, and we plotted how the discretised phase space was visited as a logarithmic heatmap. Red tones represent states that were more visited. For the sake of visual clarity, for generating the images, the ranges were slightly modified as compared with the above description of method 4. The pCSS was divided into 20 bins as before, but for each dataset, the discretisation range was chosen as the interval ranging from the minimum to the maximum value of all the pCSSs (from all the subjects and trials). The discretisation range for the time derivative of the pCSS was chosen in the same manner.

3.6.2 Discussion specific to Method 4

For each dataset, the OFR method selected features consistent with other studies in the literature. Occipital alpha activity has been reported as a marker of AD [58] [93], as has lateralised activity elicited by valence [53] [66]. However, a dynamical approach allows us to go a bit further. Consider occipital (probably visual) activity in AD disease. Healthy controls have access to a richer set of dynamical states. Furthermore, if we ignore the dynamical part (velocity axis), and observe only how the distribution of brain states changes across conditions, the available number of states turns out to be also richer for the healthy controls, suggesting loss of functions

3.6. METHOD 4: THE MOST VISITED STATES, AND HOW MANY OF THEM ARE AVAILABLE

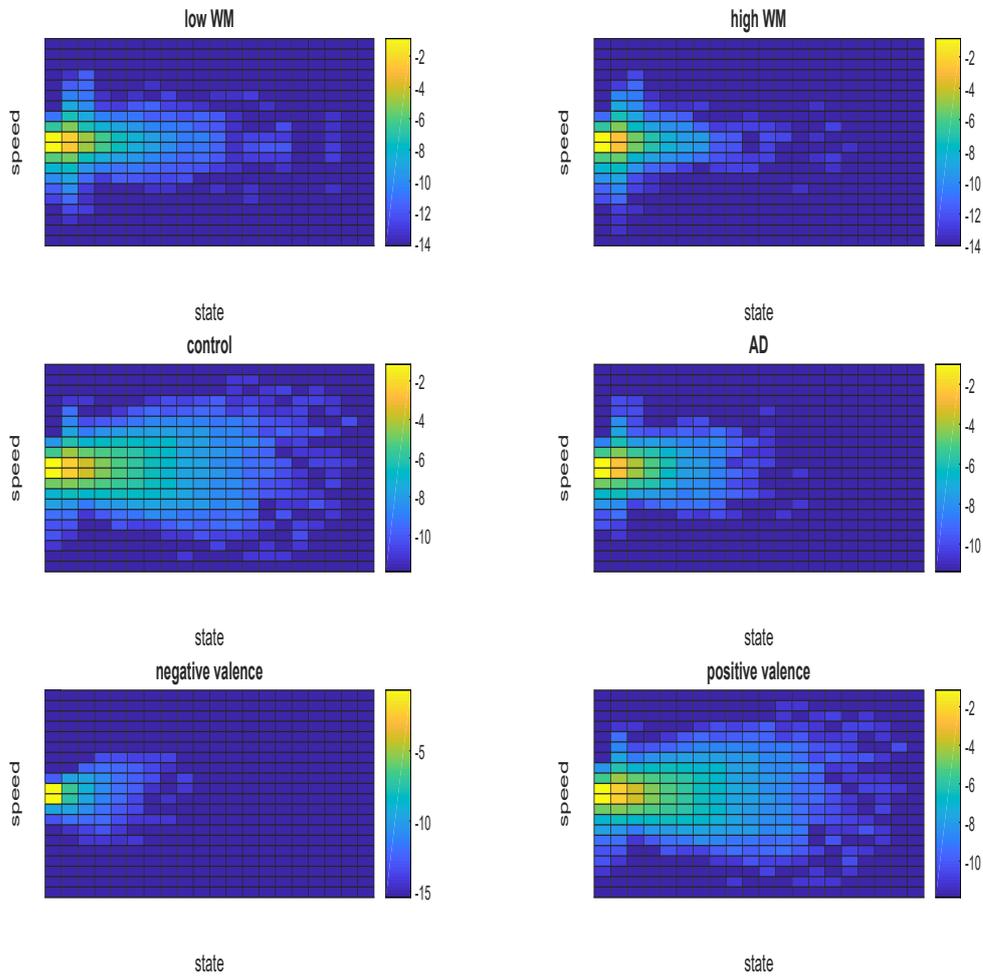


Figure 3.8: Visual representation of the most visited states of the (discretized) phase pace for each condition, for each dataset. The feature (a specific channel at a specific band) ranked first by OFR was selected for generating the image.

(both dynamic and in terms of available brain states) related to AD disease. It is important to consider that subjects recorded for this dataset were not performing any particular task, and therefore, the observed states are due to spontaneous ongoing activity. By contrast, the other two datasets were collected while subjects performed specific tasks, and therefore, the observed states may be task specific. The WM figures show that the low WM condition has a richer set of states. The low WM condition did not require full engagement, and the subjects reported performing various mental activities while

completing the task, from planning their evening activities to be attentive to possible background conversations. The high WM-condition, in contrast, required full engagement, and the small set of states may be specifically related to WM. Regarding valence, let us observe the third row of images, corresponding to the T8 electrode, located in the right hemisphere. The dynamical richness of the right hemisphere is much higher for a positive valence than for negative valence. It has been proposed that negative emotions are processed in the right hemisphere [5]. The right hemisphere could be engaged in a small set of task-specific states while exposed to negative valence material. In contrast, the observed large set of states related to the positive valence condition could be ongoing activity. Evidently, the right hemisphere does not devote all its resources to emotion processing. In the figures, we observe only the most visited states, not the totality of them. The same reasoning applies to the WM dataset. The above findings can be summarized by stating that task-related activity seems to elicit a small, perhaps more specific set of states. The task-free AD dataset, in contrast, suggests that AD decreases the dynamical richness of ongoing activity. Steyn-Ross et al. extended mean-field models³, that consider only chemical synapses, to include diffusive effects via electrical synapses [107]. They found different patterns of self-organisation depending on the time-scale of somatic and dendritic dynamics. If soma voltage remains almost constant during dendritic integration, their model displays patterns consistent with ongoing activity. On the other hand, if both time-scales are comparable, they observe faster dynamics, consistent with cognitive activity. They provide clinical evidence supporting the findings of their model. In their model as well as in our empirical analysis, patterns of self-organisation are different in nature for ongoing and for cognitive activity; in addition, cognition-driven activity exhibits faster dynamics (with small, more frequent transitions being more informative in our analysis).

³Instead of modelling individual neurons, the mean-field approach considers the activity of space averaged cortical patches. These models are expected to reproduce properties observed in space-averaged brain imaging techniques, such as EEG, MRI or MEG.

3.7 Baseline: power spectral density

Although the aim of the study is not to develop a feature extraction method, but rather to address questions about brain dynamics, spectral features were used as a baseline for comparing classification performances. The spectral features included power in the delta (1–4 Hz), theta (4–8 Hz), alpha (8–12 Hz), lower beta (12–20 Hz), upper beta (20–30 Hz) and lower gamma (30–45 Hz) ranges. The frequency bands used were the same as in Methods 1 and 4, which was done to maintain an equal number of features. Table 3.3 shows the comparison between methods, using the power spectral density as baseline.

3.8 Control tests

3.8.1 Control test 1: Destroying temporal structure and assessing statistical significance

A large part of the motivation of this work is studying the temporal structure of brain state switches. To provide more convincing evidence that what we are measuring is indeed a result of the time organization, we used shuffled variables. We computed the CSS, and before computing its time derivative, we shuffled it. We then applied Methods 1, 3 and 4 and tested whether the classification power disappeared. Method 4 is a mixture of static (distribution of states) and dynamic (states with a certain speed) information, and therefore, using shuffled data should not necessarily destroy all the information. Method 2 is about spatial synchrony, and thus, is not concerned: In fact, mutual information is not affected by the temporal structure of the data. We iterated the above procedure 300 times for each dataset for each method, and we observed how often the classification results were better than the ones we observed with no shuffling. In addition to the shuffled data, random features drawn from a uniform distribution were used. Three hundred iterations were performed with random data, and we computed the

fraction of the iterations for which the classification performance was higher than the observed results. While random features helped in assessing the statistical significance of the methods, by shuffling the data, we explored the validity of a specific claim, namely that the observed results arise from the temporal organization of the switches. In the next section, we further refine the control tests by targetting not the general temporal structure of the variables, but rather, events that may be considered as switches in the discrete model.

3.8.2 Control test 2: Removing known discrete events

Models based on discrete switches assume that switches between states occur at precise instants that can be tracked. In Freeman’s work, these points correspond to phase resetting in the original EEG signal [98]. In Kaplan’s work, they are the rapid transition processes [64]. In general, any segmentation technique could be used as well. These transitions can be removed from the time series, to assess their contribution to the classification performance. If classification is not substantially degraded, we could be even more confident in saying that most information comes from the small transitions that occur continuously. We removed points associated with phase resetting and rapid transition processes with severe criteria, in order to reduce the risk of failing to remove the postulated events. Due to the latter, more than 80% of the signal was removed, as shown in Table 3.2. We removed not only the phase resetting points but also their neighbours. As for the rapid transition processes, the segmentation algorithm proposed by Kaplan [64] first finds a large set of pre candidates to be rapid transition processes. The elements of this set are further tested and considered rapid transitions processes only if they fulfil the remaining criteria. We decided to remove the whole set of pre-candidates for far more certainty. This test was performed in the WM dataset because it had the highest temporal resolution.

3.8.3 Results specific to control tests

The results concerning control test 1 are shown in Table 3.1. For each cell, the first value corresponds to the percentage of iterations for which random features outperformed real features. The second number reflects the percentage of the iterations for which shuffled data outperformed real data. As mentioned in section 3.8.1, the shuffled data are not expected to completely destroy all the information provided by Method 4, as it also involves static information about the distribution of states. Table 3.2 shows the results of control test 2.

Dataset	Method 1	Method 2	Method 3	Method 4*
AD	0% / 0%	0% / NA	89% / 34%	0% / 0%
WM	0% / 18%	0% / NA	0% / 0%	0% / 18%
Emotions	0% / 0%	0% / NA	0% / 0%	0% / 0%

Table 3.1: Results of Control test 1. Out of the 300 iterations, the table shows the fraction of times in which random features outperformed real data (first number of the cell) and fraction of times in which shuffled data outperformed real data (second number of the cell). Method 2 was not considered for generating shuffled data because it deals with spatial synchrony, not the temporal structure. *Shuffled data are not expected to destroy all the information provided by Method 4.

Target	Percentage of data removed	Decrease in classification performance
Phase resetting points	81 %	4 %
Rapid transition processes	80 %	14 %

Table 3.2: Control test 2. Results of removing known discrete events. Specific points associated with transition events acknowledged in the literature were removed. The decrease in classification performance is shown in the third column.

3.9 Global results

The summary of the performance of all the methods for all the datasets is presented in this section. Table 3.3 shows classification results reported as the area under the ROC curve. As a reference, the baseline technique (spectral properties of the EEG signal) is displayed for comparison. The performance on the WM and emotions datasets was non-deterministic, as the calibration step involved adding noisy copies of the data. For all the non-deterministic estimations of performance, 20 realizations were executed, and the displayed results correspond to the average. The number of features encompassed by each method is displayed in parentheses.

Dataset	Method 1	Method 2	Method 3	Method 4	Baseline
AD	0.70 (126)	0.71 (126)	0.48 (21)	0.71 (126)	0.54 (126)
WM	0.76 (96)	0.74 (96)	0.68 (16)	0.75 (96)	0.75 (96)
Emotions	0.70 (192)	0.70 (192)	0.64 (32)	0.76 (192)	0.68 (192)

Table 3.3: AUCs for the different methods. Spectral properties of the signal (power at different frequency ranges) were used as a baseline. The number of features is indicated in parentheses.

DISCUSSION

4.1 Working Memory BCI discussion

We developed a cognitive BCI able to perform real-time estimation of WM load. We validated this model and obtained satisfactory online results. In addition we controlled the model for potential cognitive and motor confounders, and we compared the model output with subjective WM-load estimates of the BCI users.

The successful neurophenomenological [117] validation is one of the main features of this work. Experimenting with human subjects provided us with the unique possibility of establishing links between subjective states and objective measures [78]. These links can be meaningfully validated by the subject only under appropriate experimental conditions. It is necessary, first, to design an adequate online protocol, and second, to perform careful control tests. We expect our phenomenological validation to encourage researchers interested in rigorous cognitive monitoring and neurofeedback to pursue such avenues. Narrowing down the gap between the subjective world and objective measures opens the door to new theoretical approaches

and practical implementations.

The statistical analysis of the confounders suggests independence between the EEG biomarkers and the tested potential confounders. This holds true even for cognitive confounders that are necessarily correlated to WM, like the attentional filters or phonological loop. The original WM model from Baddeley [10] considers the phonological loop as a core element of WM, and the embedded-processes model of WM [32] explicitly refers to attentional filters. This independence is a satisfactory result for real-world testing, given that these confounders, being part of WM under certain models, are necessary but not sufficient for an activity to be demanding for WM.

There are several studies on EEG-based WM load estimation; however, to the best of our knowledge, there are none with all the properties required for a real-world, real-time continuous monitoring system.

Studies [62, 101, 102] describing statistical differences of biomarkers across WM conditions aim to make general claims about the neural correlates of WM. Nevertheless, statistically significant differences across conditions are not necessarily sufficient for single-trial classification. Jensen et al. [61], for instance, reported theta activity in frontal areas due to WM activity. However, their further examination highlighted that the theta activity revealed by the grand average was the result of the contribution of a single subject.

Some studies perform single-trial classification, which is a necessary condition for a system to work online. Nevertheless, many of them [47, 129] are offline. Online (real-time) neurofeedback experiments have different advantages, from neurophenomenological validation to overfitting prevention. It is only possible online for subjects to validate in a continuous manner that the feedback is indeed reflecting their instantaneous cognitive state, precisely due to our WM limitations. In general, online approaches allow experimenters to interactively redesign experiments until conclusive hypotheses

are attained [100]. Regarding overfitting, an overfitted model will learn noise and explain the data that was used to train it, but it will be unable to explain new data. The analysis of brain signals may involve complex models with many parameters and variables. For these models, there is a high risk of overfitting. A classifier with good online performance ensures that no positive results come from overfitting, as testing data are acquired on the fly.

Other studies [36, 49, 50, 56] describe implementations that are sufficiently fast to work in real time; however, no actual real-time testing was performed. These are useful feasibility studies, but an online validation would be necessary to assess the prototype reliability.

While studies performed online are indeed an important step towards a practical implementation of BCIs, there is still significant room for improvement, and potential confounders must be controlled for. Wilson et al. [126] train their system with EEG data from subjects performing the NASA Multi-Attribute Task Battery [29]. The task has a motor component (manipulating a joystick and a mouse), and different cognitive load levels are imposed by changing the number of events. Hence, with an imbalance of motor activity across conditions, there is a high risk of motor confounding factors being learned by the system. It is not clear, then, whether they are measuring cognitive load or motor activity. The same issue applies to the study by Berka et al. [22]. The system developed by Kohlmorgen et al. [72] seems to have balanced motor components; however, there is no cross-task. The training session and *application* session involved the same type of tasks. Consequently, it is unclear whether their results are general to WM or task-specific. It has been shown [15] that accuracies can drop to chance levels when trying to classify workload using a testing task different from the training one, even if both address the same cognitive function, the latter meaning that the system had learned particularities of the task instead of generalities of the underlying cognitive system. None of these studies specifically disentangled potential confounding factors, and furthermore, none of them performed neurophenomenological validation.

In addition to the methodological aspects, there are two factors, at the level of design, that could explain the success of our prototype. The training task was defined in such a way that the three functions involved in WM (storing, processing and refreshing) were at full capacity in the high-WM-load condition. This may be the reason for the good generalizability to a different task. In contrast, adding noisy copies of subjects' individual data (calibration data, section 2.3.2.3) allowed us to deal with a necessary compromise when facing large inter-subject variability—a trade-off between performance and the need to strictly measure WM activity, irrespective of subject idiosyncrasies. Training the classifier with only the data from the current subject may result in good performance, but there is a risk that what is ultimately measured will not be WM. By contrast, assigning an equal weight to all subjects would only detect changes that are common to all of them, minimising inter-subject differences. Evidently, not all brains respond in the same way, and we are dealing with this variability in a robust way. As an example, Grimes et al. [49] found that alpha activity increases with memory load for some subjects, while it decreases for others. This is an important remark, given that, as mentioned in the introduction, alpha activity is thought of as a potential WM signature. Addressing this trade-off allows us to build a BCI that is adapted to the user while ensuring that a general underlying cognitive function is measured. By adding noisy copies, we are also making the classifier more robust to noise.

Aiming at generalizability, an additional source of variability was imposed in Task 2. Subjects were able to choose what kind of arithmetic operations to do. In spite of this imposed variability, subjects consistently identified the feedback provided as theirs whenever this was the case.

The results of the neurophenomenological validation, although positive, could be a conservative estimation: Even when using a functional WM BCI, subjects may fail at the neurophenomenological validation. The reason is that introducing the feedback gauge and asking subjects for neurophe-

nomenological validation imposes an additional WM load that cannot be neglected. Beyond the intrinsic WM load imposed by the task, the neurophenomenological validation adds three additional sources of WM load. The first source arises from the fact that subjects are required to estimate their own WM load, which imposes an additional load due to introspection. In addition, subjects need to compare their load estimation with the feedback provided, making a binary judgement on whether it is correct or incorrect feedback. Furthermore, as we analyse EEG epochs of 2.5 seconds, our estimate is delayed. Thus, subjects must compare the current feedback with the WM load they experienced a few instants ago. This comparison is the second source of WM load. All that is described above is repeated at different moments during the trial, and all the partial binary judgements are stored so that the subject can provide a global decision at the end. Storing the binary decisions is the third additional source of WM load. As Lutz et al. [79] point out, generating first-person reports about an experience can modify that experience. Due to the additional cognitive resources required, subjects were given six trials to become familiar with the procedure. After these trials, one of the subjects expressed feeling unable to perform the task and did not continue. The subject explained that the information to be processed was overwhelming. Another subject, subject 6 in table 2.1, expressed difficulties providing an answer for the same reason. This issue is reflected in the relatively high number of unanswered questions on this subject. These testimonies suggest that the results in Table 2.1 could be a conservative estimate of the BCI performance: Subjects required a certain level of skills and training to perform the neurophenomenological validation, and before reaching an adequate level of expertise, their answers are error prone. This task is more demanding than a classical mental calculation task, which only requires an intrinsic task load. However, once the BCI is validated by enough subjects to achieve significance, we can hypothesise that it is usable for all literate subjects (subjects with an adequate performance in the BCI validation), regardless of the individual results of the neurophenomenological validation. In other words, literate subjects who cannot perform the neurophenomenological validation may still be able to

use the BCI in a real-world task, in which they are expected to believe the feedback, not rate it. In an adaptive system, users may not even receive any feedback, as the feedback could be used for the system to trigger an action (lower the task difficulty, start an autopilot, etc.).

The choice of a reversed gauge as sham feedback was made to preserve the dynamic behaviour of the feedback. Had we presented, random feedback, for instance, subjects could have learned that random motion of the bar implies sham feedback, depending on the underlying distribution and dynamics. In addition, the choice of a disappearing cue in Task 2 imposed a time-locked change in the WM load, helping subjects handle BCI delays.

Regarding the WMLE, given that no previous studies had been done on performing neurophenomenological validation, the dynamics of the WM load remained an open question. Of our six online subjects doing the cross-task, two mentioned that it was specifically the dynamics of the feedback gauge that helped them decide whether it was sham or real. In other words, the most informative event for them was whether the gauge increased or decreased at key moments, rather than the absolute value of the gauge. For another two, it was both the absolute value and the dynamics; they mentioned that it was extremely easy to know when it was real or sham. For the remaining two, there was no clear distinction. Generally, they all stated that the (true) feedback was a measure of their WM load, which in turn, was reflected in the high rate of correct answers of the neurophenomenological validation. Furthermore, most of them mentioned spontaneously that it was clear that the sham feedback signal was the reverse of their WM load. This information was not disclosed to them in advance.

Although the global mean of the WMLE remained lower for the low-WM-load condition than for the high-WM-load condition, there were WMLE peaks in both conditions. Further investigation needs to be carried out for determining whether these peaks correspond to refreshing, processing, or simply noise. One subject spontaneously reported that they corresponded to

processing peaks; however, this was not reported by the other subjects. This is not surprising, as subjects are generally not used to monitoring their WM or thinking about it in terms of its subprocesses. Expertise and knowledge of the WM subprocesses would be required to answer that question, and the above mentioned subject had some prior knowledge about WM.

It is of theoretical relevance to investigate which aspects of WM induce more load in the central executive and how these events are temporally distributed depending on the WM load. In addition, if these peaks represent true activity and not noise, then they could be used to improve the performance of the BCI as well.

Some of the training trials of the low condition will happen to contain these peaks, although the subject is engaged in the low-WM-load condition. Peaks in the low-WM-load condition could be present for different reasons. The subject could have been temporarily allocating mental resources to non-task-related activities—attending to external stimuli, mind wandering due to lack of motivation, and so on. In other words, we cannot impose a specific, constant WM load. Moreover, if peaks are refreshing events, they must occur as well in the low-WM-load condition. Perhaps at a different pace and/or with different intensity, but they must occur. That being said, if we have an objective measurement (the WMLE), we could iteratively improve the quality of our offline database. For instance, we could accomplish this by reallocating peaks in the low-WM-load condition to the other class (high-WM load), run the algorithm again, and repeat until stability is reached (the sizes of both classes remaining constant), or alternatively, until another stopping criterion is met, in the absence of stability. In fact, these peaks of activity in the low-WM condition may be the reason why the performance of the neurophenomenological validation seems better than the BCI validation does. Neurophenomenological validation shows the prediction in real time. If low-WM-load trials are contaminated with high-WM-load activity (label noise), the subject may know this and still perform an adequate validation. However, the label (low WM or high WM) remains constant throughout the

whole trial, regardless of the subject's state, and the BCI validation performance considers the label—rather than the internal state of the subject, as truth.

As for the biomarkers, we observe different markers where the joint activity predicts WM load (Figure 2.7). Instead of using single biomarkers to estimate WM load, we derived composite biomarkers from weighted combinations of several biomarkers. Such an estimate is more reliable and more realistic, considering that WM is a complex cognitive function involving the coordination of several brain areas. Thus, we are associating patterns of joint activity with WM conditions. Some of the biomarkers are consistent with the literature, like the decrease in the alpha power in the occipital regions mentioned in the introduction. However, it is important to stress again that considering them as isolated markers of WM activity may be misleading. Conversely, there are also biomarkers in which the change across the WM conditions was statistically significant; however, including them in the analysis decreased the ability of the BCI to correctly estimate the WM load. One potential explanation for this is that these changes are due to subject variability and not to a fundamental aspect of WM. The above mentioned study ([61]), in which the grand average showed significant changes in the theta power because of the contribution of one single subject, is an example of this. Incidentally, we did not find any significant changes in the theta range. Another explanation is multiple hypothesis testing. If we test hundreds of biomarkers, by chance some of them will appear statistically significant, even if there is no real correlation. This illustrates how multivariable classification techniques can be powerful approaches for making inferences. Although some changes across both conditions would appear as statistically significant, their lack of generality renders them useless for prediction. Thus, we have a simple yet powerful criterion: If, by adding information (biomarkers), our ability to classify decreases, then it is not relevant information, although it may seem statistically significant. In addition, combining several biomarkers into a single score (the WMLE) reduces the risks associated with multiple hypothesis testing.

A situation in between may occur. It could be that adding biomarkers will neither increase nor decrease the classification power. One potential explanation for this is that these biomarkers are highly correlated with previously selected biomarkers. Therefore, our list of relevant biomarkers is by no means exhaustive, as mentioned in section 2.2.5. Biomarkers redundant to the selected ones were not chosen, and this could be the reason for the apparent asymmetry in Figure 2.7. In fact, running the code with different parameters would sometimes lead to the selection of the symmetric electrode, for instance, electrode CP6 instead of electrode CP5. Moreover, biomarkers that do not increase the classification power may increase it by using more sophisticated techniques, such as support vector machines or neural networks, that better capture the complexity of the underlying system. The latter is further developed at the end of the discussion.

Let us graphically illustrate how joint activity provides a more reliable estimate. Figure 4.1 shows the first selected biomarker, that is, the one with the highest correlation to the WM load. Each element of the plot is a trial. Crosses correspond to high-WM-load trials, while circles correspond to low-WM-load trials. The variable chosen was the relative lower beta power, at electrode Fp1.

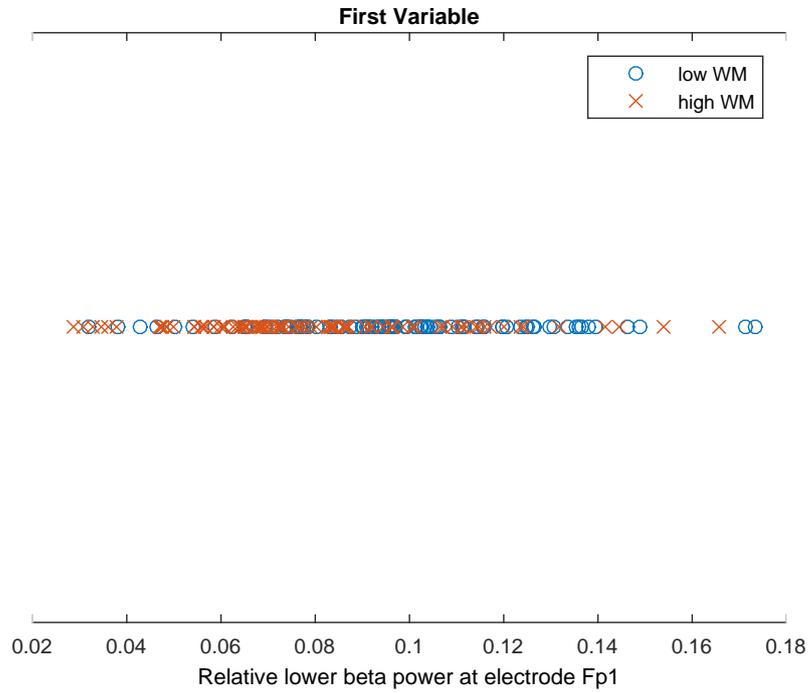


Figure 4.1: Relative lower beta power at electrode Fp1 for the low-WM-load condition (circles) and high-WM-load condition (crosses)

As illustrated in the figure, the distributions of the examples conditioned to the class have different means; however, they completely overlap, and hence, class separation for this variable is completely inefficient. However, if we add a second variable, the relative lower beta power at electrode Cz, we start finding some structure. In Figure 4.2, we already start to distinguish both conditions.

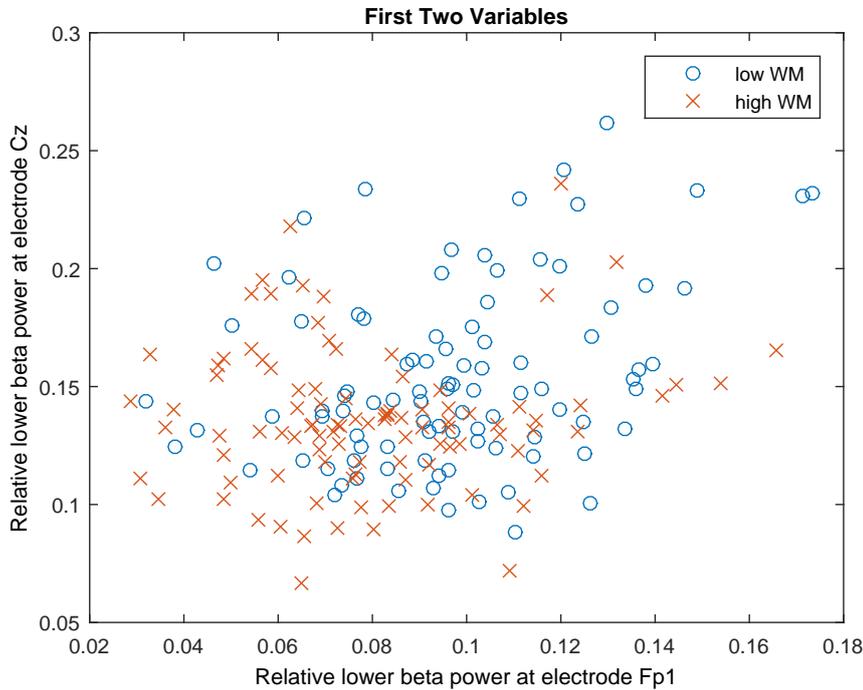


Figure 4.2: Relative lower beta power at electrode Fp1 versus relative lower beta power at electrode Cz. Trials of low-WM-load condition are represented with circles, high-WM-load trials are crosses

As we add more variables, we continue this process in higher dimensions. To perform this visually is impossible for the human eye if there are four or more dimensions; however, this is exactly one of the uses of multivariable statistical analysis or machine learning. This means that claiming that beta activity changes at Fp1 due to WM load may be misleading. It is much more accurate to investigate specific multivariable correlates of WM. For instance, we could hypothesize that our observation of beta-range activity decrease is a correlate of the cognitive flexibility state induced by the intensive rehearsal strategy used by our subjects, rather than a specific correlate of WM maintenance and processing. Indeed, according to [35], beta-range activity could serve the maintenance of the current cognitive state, and would decrease when cognitive flexibility is needed, an event which could occur in any situation where a cognitive load (related or not to WM) occurs. It is only in the perspective of the other identified features (in our report

the alpha and gamma range) that we could indeed target WM. This is also confirmed in recent publications, for instance Bahramisharif et al. [14] have shown that WM maintenance involves couplings of multiple frequencies (in their implanted EEG study, the alpha, theta and gamma ranges were co-organized during WM maintenance, confirming previous publications in scalp EEG, such as [97]).

Our approach is still far from conveying a global view. EEG consists of a few scalp recordings of the activity of an extremely complex underlying system. Acknowledging our relatively small (for a multivariable analysis) database, aiming at robustness and as a first attempt, we chose a linear feature-selection technique and linear classifier. A linear feature-selection technique may not necessarily work when correlations between features and the output are not linear. A linear classifier assumes a certain topology in the feature space: Classes can be separated with a line, plane or hyperplane. A larger dataset would allow the use of feature-selection techniques and classifiers that better capture the underlying complexity. In addition, we are ignoring potential neural mechanisms that could be active when subjects stay at the limit of their WM capacity for a long time. If such mechanisms exist, the distribution of biomarkers in the feature space could be completely different. A real-world application should thus explore the extent to which WM-load-detection protocols may need to be modified when a high WM load is imposed for long periods.

4.2 Brain metastability discussion

Motivated by the way information is registered and processed in parallel by different areas of the brain, which engage in joint activity and disengage on demand, we proposed certain assumptions that we tested. First, like other authors, we assumed metastability driven by cognition. Following the definition of metastability, we proposed local variables that reflect the time evolution of brain states, and developed a framework for studying how

cognition affects different properties of the spatio temporal organization of these variables.

By looking at Table 3.3 and Table 3.1, we can easily conclude that the proposed properties of the spatio temporal structure of the proposed variables are affected by cognition in a way that cannot be explained by chance. An exception is Method 3 in the AD dataset. However, it is worth noting that the number of features computed for Method 3 is six times lower than the number of features in the other methods, and therefore, we cannot directly compare them. Furthermore, the AD dataset is much smaller. Although the number of subjects is larger, the available data for each one and the sampling rates are much smaller. In the emotions dataset, we did not disentangle interactions from other emotions, which would likely have improved the classification results. The experiment from which the data were collected involved axis that were not disentangled, namely valence, like/dislike, dominance and familiarity. The participants rated videos according to these dimensions, but no material was created for investigating one axis while controlling for the others. In addition, several parameters like the type of norm of the velocity or the number of full oscillations considered to fix the window length could be optimised for each scenario (method and dataset), with a significant improvement in performance. However, we decided to keep those parameters fixed to avoid the potential issue of multiple testing [21], as well as because our main goal was to answer a set of scientific questions.

We can observe in Table 3.2 that targeting specific events recognised as candidates for state switching did not substantially degrade the quality of the information. Our conservative approach, which removed as well neighbours and false positives, discarded more than 80% of the data, with only small decreases in the classification performances—3.8% when removing phase-resetting points and 14.4% when removing rapid-transition-process points.

We suggest two possible hypotheses. The first is that, although the spatio temporal structure of the statistical properties of EEG, and especially their dynamics, are relevant to predict cognitive conditions, timing is not discrete. The brain undergoes large transitions at seemingly discrete times, but it keeps fluctuating between neighbouring states in a way that is affected by cognition. In other words, these small fluctuations are not artifacts due to any source of noise, but rather they are induced by cognition. Computations in the brain are analogue, and even under continuous switching, the neighbourhood of a state could provide enough stability to induce the oscillations thought to be required for psychological constructs. Cognition may drive the segregation tendencies considered in the introduction, which push the dynamics away in a way that depends on the condition. In this direction, Werner [125] suggested that metastability can be given an operational meaning: Instead of considering integration–segregation as two poles, a continuous range of tendencies of neural coordination seems more appropriate. According to this author, the continuum seems to be supported at the neuronal dynamics level by the flexibility of coupling coefficients amongst different neuron groups. Assuming continuity may be thought of as rejecting the existence of metastable states, as there are no states of finite duration with constant statistical properties. It is still possible to draw on the less restrictive concept of metastable *regimes* in the dynamical systems point of view [112]: Dynamics takes place in a region where all the attractors have disappeared.

A second hypothesis is that switches are discrete, but due to volume conduction, the recordings reflect the influence of neighbouring regions. Continuity would be then an artifact of the lack of spatial resolution of the measurements. The further the region, the less its changes in statistical properties affect the local recording. EEG source localization could be used to either support or rule out this possibility. The CSS can be computed using the EEG sources instead of EEG raw recordings. If the timing is discrete we should observe piecewise continuity in the statistical properties at the level of the sources.

An argument favouring the first hypothesis is that the power-law fit of the PSD of the CSS demonstrates a fractal temporal structure of the CSS. As mentioned above, the power-law coefficient is a parameter that expresses the extent to which disturbances propagate. We showed that this parameter is affected by cognition, which makes sense if we consider, as other authors have, sensory and motor information as disturbances in this context. The second hypothesis then requires in turn an extra hypothesis. The inter switch duration of a specific brain source should be influenced by the neighbouring sources, as their joint switching dynamics should still be fractal in time. In other words, we need to translate the fractal time structure of a single region (first hypothesis) into the spatial organization of subregions (second hypothesis). Postulating the latter should also involve postulating a mechanism producing this spatial organization. This mechanism should be at least as parsimonious biologically as that in which dynamics evolve for allowing an efficient propagation of disturbances under the first hypothesis.

The above discussion may be theoretically relevant at different levels. We have discussed the biological implications of continuity and discreteness; however, other aspects are relevant as well. Phenomenologically, a fragmented flow of perception or consciousness is essentially different from a continuous flow. Whereas a thought, an action or perceiving an object might seem granular after a quick exercise of introspection, “microcognitive science”, or neurophenomenology at the sub-second level, suggests otherwise. Petitmengin et al. [91] investigate how elicitation techniques provide access to micro-states, at the sub-second level, where boundaries across sensory modalities, and between object and subject, begin to blur. They advocate for finding correlations between these sub-second, first person experiences and third-person, objective measurements. The proposed spatiotemporal analysis of brain-state switches is a possible candidate tool to investigate such correlations. In addition, regarding the mathematical description of natural systems, discrete and continuous mathematical models may have different properties. As a simple case in point, we can consider the logistic

map, one of the simplest discrete dynamical systems able to exhibit chaos. Its continuous version, in contrast, is never chaotic. For a discrete system, there is always a “next” value, whereas this is not the case for a continuous system. Nevertheless, if continuity is an artifact of volume conduction, the presented framework has proven useful still. If the temporal organization turns out to be disguised spatial organization, we have no reason to discard the information obtained from small (far?) transitions. Evidently, the next step would be to identify and study the biological mechanisms generating this spatial organisation.

A view where time is continuous is more compatible with an analogue computer metaphor, and in this regard, our proposals are compatible with the work of Spivey [106]. He suggests that, if we could take the activity of single neurons as variables, cognition would be a continuous trajectory in a high dimensional space, where each coordinate is the activity of a neuron. In his proposal, a specific cognitive task would be a point in this space, and performing such a task would be a trajectory moving toward this point. Perceiving a face, for instance, would be moving toward the point that corresponds to that specific face. Interestingly, he claims that we spend more time near such points than at them. Experimentally, Chang and Tsao [26] could reconstruct human faces with impressive accuracy by reading the activity of 205 neurons in primates. Each neuron codes for a specific facial feature, and the joint activity of the 205 neurons, that is, a point in a space of dimension 205, represents a specific face.

While face recognition is so important that evolution may have given it a sparse representation, in the general case of cognition, we do not have the experimental and computational means for exploring such points and trajectories. Nevertheless, using a low (third) order estimation and a few scalp recordings, we showed that the idea of cognition as a trajectory in an abstract space is worth further investigation.

In general terms, we have developed a framework that produces evidence

that aims to enrich theoretical discussions about brain dynamics. As it is important for us to show that cognition is driving these phenomena, we developed tools for classifying cognitive conditions on a single-trial basis, and practical applications, such as BCIs can benefit too from these methods. In addition, the predictive power of this framework resides in a signal property that is often overlooked, or even considered as a problem to overcome—the lack of stationarity.

CHAPTER



APPENDIX: PUBLISHED WORKS

CHAPTER 5. APPENDIX: PUBLISHED WORKS

Neuroscience and Biobehavioral Reviews 68 (2016) 891–910



Contents lists available at ScienceDirect

Neuroscience and Biobehavioral Reviews

journal homepage: www.elsevier.com/locate/neubiorev



A psychoengineering paradigm for the neurocognitive mechanisms of biofeedback and neurofeedback



A. Gaume^{a,b}, A. Vialatte^{a,c}, A. Mora-Sánchez^{a,b}, C. Ramdani^d, F.B. Vialatte^{b,*}

^a Université Pierre et Marie Curie, 4 Place Jussieu, 75005 Paris, France

^b Laboratoire Plasticité du Cerveau, UMR 8249, ESPCI Paris Tech, PSL Research University, 10 rue Vauquelin 75005 Paris, France

^c Institut Langevin, UMR 7587, ESPCI Paris Tech, PSL Research University, 1 Rue Jussieu, 75005 Paris, France

^d Institut de recherche biomédicale des armées, BP 73, 91223 Brétigny sur Orge Cedex, France

ARTICLE INFO

Article history:

Received 18 February 2016

Received in revised form 3 June 2016

Accepted 14 June 2016

Available online 20 June 2016

Keywords:

Biofeedback

Neurofeedback

Learning

Psychoengineering

Executive function

Brain plasticity

Development

ABSTRACT

We believe that the missing keystone to design effective and efficient biofeedback and neurofeedback protocols is a comprehensive model of the mechanisms of feedback learning. In this manuscript we review the learning models in behavioral, developmental and cognitive psychology, and derive a synthetic model of the psychological perspective on biofeedback. We afterwards review the neural correlates of feedback learning mechanisms, and present a general neuroscience model of biofeedback. We subsequently show how biomedical engineering principles can be applied to design efficient feedback protocols. We finally present an integrative psychoengineering model of the feedback learning processes, and provide new guidelines for the efficient design of biofeedback and neurofeedback protocols. We identify five key properties, (1) perceptibility = can the subject perceive the biosignal?, (2) autonomy = can the subject regulate by himself?, (3) mastery = degree of control over the biosignal, (4) motivation = rewards system of the biofeedback, and (5) learnability = possibility of learning. We conclude with guidelines for the investigation and promotion of these properties in biofeedback protocols.

© 2016 Elsevier Ltd. All rights reserved.

Contents

1. Introduction	892
1.1. Potential of feedback approaches	892
1.2. Towards higher standards	892
1.3. Modeling neurofeedback and biofeedback: how does it work?	893
2. Biomedical perspective	893
2.1. Acquiring skills	893
2.2. Volitional and conscious strategy?	894
2.3. Synthetic biomedical model	894
3. Psychological perspective	894
3.1. Operant conditioning: the reward problem	894
3.2. Developmental psychology and schemata	895
3.3. Skill learning	896
3.4. Executive function and attention	897
3.5. Working memory models	897
3.5.1. Multiple-component model	897
3.5.2. Embedded-process model	897
3.5.3. Long-term working memory	898
3.5.4. Time-based resource-sharing model	898

* Corresponding author.

E-mail address: francois.vialatte@espci.fr (F.B. Vialatte).

<http://dx.doi.org/10.1016/j.neubiorev.2016.06.012>

0149-7634/© 2016 Elsevier Ltd. All rights reserved.

3.6.	Volitional action, agency, and fluency	898
3.7.	Synthetic psychological models	898
4.	Neuroscience perspective	899
4.1.	Neural correlates of schemata formation	899
4.2.	Schemata and working memory	900
4.3.	Executive functions and the prefrontal cortex	900
4.4.	Self-control networks	901
4.5.	Consciousness of errors and error potentials	901
4.6.	Motivation and reward	902
4.7.	General model of feedback learning	902
5.	Engineering perspective	903
5.1.	Process control models of feedback learning	903
5.1.1.	Controllability	903
5.1.2.	Observability	903
5.1.3.	Basic model	903
5.1.4.	Explicit model of neurofeedback and biofeedback	904
5.1.5.	Implicit model of neurofeedback and biofeedback	904
5.2.	Limits of process control models	904
5.2.1.	Linearity	904
5.2.2.	Stability	904
5.2.3.	Temporality: transients and steady-states	904
5.2.4.	Precision	904
5.3.	Serious games	904
6.	Psychoengineering model	905
6.1.	The missing keystone: toward a psychoengineering model	905
6.2.	Bridging the disciplinary gaps	905
7.	Conclusion	905
7.1.	Investigating and promoting perceptibility	905
7.2.	Investigating and promoting autonomy	906
7.3.	Investigating and promoting mastery	906
7.4.	Investigating and promoting motivation	907
7.5.	Investigating and promoting learnability	907
	References	907

1. Introduction

1.1. Potential of feedback approaches

When children go to school to learn how to read and write, they receive guidance and feedback from their teachers. Through hard work and receptivity to instruction, their cognitive skills will adapt and they will eventually acquire reading and writing skills. This adaptation is crucial to human development and central to the acquisition of what makes us human; tutored interaction plays a key role in culture acquisition. Biofeedback provides a subject with a similar type of training, but instead of acquiring knowledge, the subject acquires self-regulation mechanisms in order to control affective, biological, and/or cognitive skills. Such psychophysiological self-regulation could theoretically extend to the functioning of both the autonomic and the central nervous systems (Prinzel et al., 2001). Common modalities of biofeedback include respiratory, cardiovascular, neuromuscular, skin conductance and temperature, and central nervous system (Khazan, 2013).

Biofeedback can be explicit or implicit information (Dekker and Champion, 2007; Kuikkaniemi et al., 2010; Nacke et al., 2011). In the explicit model, feedback is given to the controller so that the controller can act on the system. This is the most typical case of biofeedback or neurofeedback: the user observes a (generally visual or auditory, less frequently tactile) feedback signal, which is a direct correlate of the biosignal to regulate. For example, the user hears a sound with an amplitude directly proportional to his heart rate, providing him/her with an additional perception to help him/her regulate this biosignal. In implicit biofeedback, the signal is not explicitly presented to the subject, but instead changes some detail(s) of the experimental conditions. For example, a person

using a videogame whose content (e.g., changing levels of difficulty or access to bonus items) evolves depending upon his heart rate is receiving implicit feedback; he/she does not know directly that his heart rate has dropped, but he/she experiences indirect effects of this physiological change. The user is not directly aware of his biosignal, but since it changes the behavior of the system he/she is observing, he/she gets implicit access to a correlate of that biosignal. Implicit feedback is used for subtle and indirect interactions (e.g., changing implicitly the game difficulty) rather than to provide information (Dekker and Champion, 2007; Kuikkaniemi et al., 2010). Such indirect biofeedbacks have an effect on motivational variables (Nacke et al., 2010), and are typically used in affective videogames (Gilleade and Dix, 2005). However, note that if the user of an implicit biofeedback starts learning how the system works and thereby gains control over it, implicit biofeedback becomes explicit (Kuikkaniemi et al., 2010).

Biofeedback is also one of the best approaches to the problem of neurophenomenology (Varela et al., 2001). Especially when applied to the brain (neurofeedback), it is a promising new scientific avenue to explore phenomenology and to investigate the self and consciousness (Bagdasaryan and Le Van Quyen, 2013), thereby attempting to solve the so-called hard problem of consciousness (Chalmers, 1995).

Finally, biofeedback holds a prominent position in the transhumanist agenda (Hansell and Grassie, 2011). Transhumanism is an international and intellectual movement that aims to enhance human intellectual, physical, and psychological capacities (Bostrom, 2006). The cybernetics perspective on biofeedback (Anliker, 1977) opens new perspectives about human enhancement, attracting the attention of a growing scientific community.

CHAPTER 5. APPENDIX: PUBLISHED WORKS

1.2. Towards higher standards

In order to clearly evaluate the clinical efficacy of biofeedback interventions, the Association for Applied Psychophysiology and Biofeedback and the Society for Neuronal Regulation developed guidelines with five levels of performance (Moss and Gunkelman, 2002): (1) not empirically supported, (2) possibly efficacious, (3) probably efficacious, (4) efficacious, (5) efficacious and specific. In order to reach level 4 and be considered efficacious, a treatment must be replicated in at least two independent studies, the data analysis must not be flawed, the outcome must be evaluated with precise inclusion criteria, and the experimental setting must involve randomized control trials. Level 5 is reached if the treatment satisfies level 4 conditions, and in addition is statistically superior to credible sham therapy, pill, or alternative bona fide treatment in at least two studies. In a review of 41 treatments, urinary incontinence in females was the only biofeedback treatment found to be efficacious and specific (Yucha and Montgomery, 2008). In the same study, biofeedback was deemed efficacious for ten other conditions: anxiety, attention deficit hyperactivity disorder (ADHD), chronic pain, epilepsy, constipation, headache, hypertension, motion sickness, Raynaud's disease, and temporomandibular disorder. Note that the survey criteria did not require double-blind investigations; consequently, some of the treatments ranked at level 4 may still be biased by placebo effects. In other words, despite several well-conducted studies exist, the effectiveness of biofeedback has not been fully demonstrated yet, due to insufficient evidences. We hope future biofeedback studies will reach higher standards, so they can meet with level 5 condition with double-blind protocols.

1.3. Modeling neurofeedback and biofeedback: how does it work?

Previous studies attempted to describe the cognitive adaptation mechanisms supporting neuro and biofeedback (Sherlin et al., 2011; Bagdasaryan and Le Van Quyen, 2013; Gevensleben et al., 2014; Ros et al., 2014; Micoulaud-Franchi et al., 2015). We believe that the missing keystone to design effective and efficient approaches is a clear and comprehensive model synthesizing the existing medical, neurological, psychological and engineering perspectives. Considering that information processing is impacted by biofeedback, one would expect to see a model—or at least an explanation—of how these processes will adapt. Due to disciplinary barriers, even though these cognitive adaptation processes have been described in the scientific literature, a general model has never been proposed. In the interest of removing those barriers, we will review existing models of biofeedback from biomedical, psychological, brain science, and bioengineering perspectives. We will then synthesize those views and present a general model of the cognitive adaptation mechanisms underlying biofeedback. As was stated by Georges Box, all models are essentially wrong, but some are useful (Box and Draper, 1987). We will prove the usefulness of this model by providing guidelines for proper development of efficient biofeedback and neurofeedback protocols and the means to control key parameters for successful feedback learning.

2. Biomedical perspective

Psychophysiological self-regulation, also commonly termed biofeedback (biological feedback), can be investigated from a biomedical perspective. In this section we will review the existing models of biofeedback mechanisms from the perspective of biomedical interventions, where the aim is to improve biological variables impaired by dysfunctions (e.g., blood pressure, tension, heart rate variability, etc.). The variable of interest is fed back to

the subject as a biosignal that he/she then attempts to regulate. Consequently, investigations in the biomedical field are more concerned with optimizing conditions for the provision of effective and efficient treatments. In other words, most manuscripts in this field focus more on biofeedback efficiency rather than on biofeedback mechanisms. Consequently, we will review in this section the interpretations found in the biomedical literature about the conditions for efficient biofeedback design, considered as a treatment.

Indeed, a medical approach to biofeedback necessarily means an approach centered on treatment of pathologies, for the purposes of improving health and performance (Yucha and Montgomery, 2008). This perspective is to be distinguished from the transhumanist goal of performance enhancement (Maheu et al., 2004), and should not be confused with the entertainment perspective of biofeedback games (Arns et al., 2015). In other words, medical biofeedback seeks to cure, not to enhance or entertain. Some authors defend the thesis that biofeedback would normalize biological functions, thereby treating pathologies. For instance, for Arns the main goal neurofeedback is to normalize deviant brain activity (Arns, 2011). The biofeedback aim would in this case be to train the patient so he can reach normality. However, judging a statistically abnormal feature as pathology is rather a normative judgement than a scientific one. One shall always refer to the individual's own reference when defining pathology¹ (Canguilhem, 1966). According to Canguilhem's perspective, the aim of medicine in general and biofeedback in particular would be to seek improvements in impaired functions, instead of seeking normality. However, as the distinction between the vital norms of the body and the disciplinary norms of society is becoming difficult to maintain in the modern times, this ethical question remains to be solved (Rose, 2009).

2.1. Acquiring skills

For decades, the biomedical literature has emphasized biofeedback's basis in the acquisition of self-regulation and self-control skills that subjects could use to correct their states toward an optimum (Schwartz and Schwartz, 2003; Norris, 1986; Epstein and Blanchard, 1977; Hauri, 1975). The consequence of this acquisition of new self-control skills would be an improved "calibration" of the nervous system (Brenner, 1974). The key to understanding the effect of biofeedback would then be to model how these volitional skills or strategies are acquired during biofeedback sessions. One can identify two specific skills: **discrimination**, which is the aptitude to achieve an inner perception of the biological variable, and **self-maintenance**, which is the ability to affect the biological variable and effectively change it in the intended direction (Epstein and Blanchard, 1977). These skills would in turn allow subjects to regulate their biological constants through a volitional psychosomatic process (Leigh, 1978). This model provides an important guideline for evaluating biofeedback systems—a model that unfortunately has not been taken into account in several studies. Indeed, if discrimination and self-maintenance are acquired, then a proper evaluation of biofeedback should be based on an evaluation of this acquisition. Biofeedback, therefore, should be evaluated pre- and post-training to determine whether the subject has an improved perception of and action on the targeted biological variable (Epstein and Blanchard, 1977). This should be done by comparing the subject's perceptions before and after training (rather than merely evaluating objective performances). There is, however, a surprising lack of reflection in the biomedical literature regarding the nature of those self-control skills and what those strategies could be. Nevertheless, one could easily make the small leap to define

¹ En matière de normes biologiques, c'est toujours à l'individu qu'il faut se référer.

discrimination and self-maintenance skills as *cognitive processes*. We will attempt to provide a proper definition of this new class of cognitive processes in Sections 3 and 4.

2.2. Volitional and conscious strategy?

The existing literature presents contradictory theories about the effects of biofeedback: it is either attributed to *volitional control* over the biological variables (involving executive function) or to *autonomic regulation* of subcognitive systems. The biofeedback literature most often argues that observed effects are due to volitional control of biological variables (e.g., Abukonna et al., 2013), and neurofeedback is known to be more efficient when based on volitional and conscious cognitive strategies demanding the use of attentional processes (Bagdasaryan and Le Van Quyen, 2013). However, one could argue that improved regulation could be achieved without volitional control (in which case the subject would not exert voluntary control over the regulation). In a recent review, for example (Lehrer and Gevirtz, 2014), the effect of heart rate variability biofeedback was attributed to a combination of causes, including homeostasis in the baroreceptors, parasympathetic reflex stimulation, improved gas exchange, mechanical stretching of airways, anti-inflammatory effects, and attentional effects. Nevertheless, as we will see in Section 3.6 it is difficult to defend a hypothesis involving a total absence of volitional control.

Furthermore, one could consider the learning strategy to acquire the biofeedback skills (discrimination and self-maintenance) to be conscious or unconscious. The discord between a cognitive model and an infra-cognitive model is more visible in neurofeedback publications, where two different models can easily be identified. On the one hand, a recent manuscript suggested that neurofeedback relies on a top-down processing mechanism, where higher cognitive functions percolate down from large-scale oscillations to small-scale and single-neuronal activities (Bagdasaryan and Le Van Quyen, 2013). On the other hand, operant conditioning (OC) has historically been the dominant interpretation of neurofeedback mechanisms; the feedback would in that case be modeled as an implicit infra-cognitive reinforcement signal (Lawrence et al., 2014; Caria et al., 2011; Koralek et al., 2012; Sterman and Egner, 2006). These two models lead to opposing perspectives on proper feedback designs: one based on a behavioral paradigm using conditioning strategies, discrete trials, reinforcement approaches, and exercises excluding entertainment (Sterman and Egner, 2006); and another based on a cognitive paradigm linking inner events with the corresponding neural signals (Bagdasaryan and Le Van Quyen, 2013). These two conflicting models have led to a dual-process theory for neurofeedback mechanisms (Wood et al., 2014), a theory that categorizes the cognitive functions supporting neurofeedback into two main types of processing: more automatic and capacity-free processes vs. more controlled and capacity-limited processes. One possible resolution to this contradiction would be to postulate the existence of interactions between these two types of processing. From this perspective, biofeedback could be considered as a self-investigation tool, where the patient improves his volitional control over autonomic mechanisms (Zolten, 1989). As we will see in Section 3.2, it is possible to reconcile these two apparently opposing perspectives, as recent connectionist models in developmental psychology can integrate both perspectives on a continuum. The question of entertainment and biofeedback will be discussed further with the paradigm of serious games in Section 5.3.

2.3. Synthetic biomedical model

From a biomedical perspective, biofeedback paradigms are based either on cognitive training—or subcognitive regulation—of two specific skills acquired using a biosignal (Fig. 1): discrimination

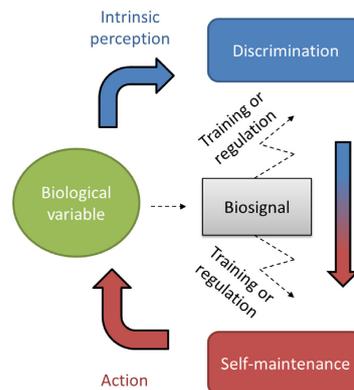


Fig. 1. Biomedical model of biofeedback.

(perception of the target biological variable) and self-maintenance (action over the biological variable). Successful training in either or both of these skills would lead to improved balance in the biological variable for patients suffering from medical conditions involving that variable, and the positive effect should remain when the feedback is turned off (otherwise the patient would be dependent upon the feedback system).

3. Psychological perspective

3.1. Operant conditioning: the reward problem

As we mentioned in Section 2.2, the mechanisms of biofeedback have traditionally been theorized using a behavioral approach inspired by Skinner's theories of OC (Skinner, 1938; Sherlin et al., 2011) and reinforcement learning (RL). The OC paradigm states that when a behavior has consequences (either rewards or punishments), it will be reinforced or repressed. In the case of biofeedback, the behavior is the regulation of an underlying biological variable, and the reinforcement signal is the success or failure of the subject to modulate the feedback signal. Such an approach is supported by animal studies: for example, prefrontal cortex neurons can be controlled by rhesus monkeys through an OC paradigm (Schafer and Moore, 2011). RL has two possible mechanisms (Sutton and Barto, 1998; Dayan and Berridge, 2014): either the subject is in a goal-directed setup and supports his learning from an internal model, in which case learning is termed as model-based RL; or the subject has no model of the outside events and learning arises from simple associations, termed as model-free RL. In the case of biofeedback and neurofeedback based on explicit feedbacks, a model-based RL is triggered: the subject seeks to reach a goal (regulating the feedback signal). In the case of biofeedback and neurofeedback based on implicit feedbacks, learning is more likely to follow a model-free RL mechanism. OC, and more specifically the SORC model (Goldfried and Sprafkin, 1976), has been used for decades to model the functional analysis of behaviors (Bellack and Hersen, 1988). In the SORC model (see Fig. 2), the behavior of an organism is modulated by the environmental feedback that is the consequence of its action. In other words, the action consequence acts as a reward signal.

However, whether for implicit or explicit feedbacks, the OC model for biofeedback has a fundamental limitation. The prob-

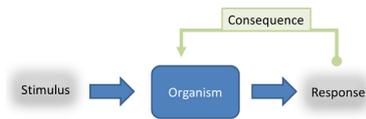


Fig. 2. The SORC model (Goldfried and Sprafkin, 1976), a behavioral model inspired by Skinner's theories on operant conditioning. SORC is an acronym for S-Stimuli, O-Organism variables, R-Responses, and C-Consequences. In this model, an individual's responses are thought to be a joint function of immediate environmental variables (stimuli and consequences) and of organism variables (physiological characteristics and past learning history) that the individual brings to the situation (Nelson-Gray and Farmer, 1999).

lem lies with the definition of the reinforcement signal. In animal experiments, it is standard practice to withhold food from a rat or a monkey and provide it later as a reward when the animal successfully modulates a biosignal, but it would be more difficult (and clearly unethical) to deprive human subjects of food. Furthermore, there is no guarantee that a human subject would interpret the biosignal as a reward: interpretation of the signal would depend upon the motivational state of the subject (see Section 5.3 about motivation).

The challenge is, therefore, to find an appropriate and effective reward to motivate human subjects. Rewards can be extrinsic when they take the form of external motivations (such as money, a pat on the back, or food) or intrinsic when based on self-motivation. Extrinsic rewards are maladaptive for human subjects: even if it were possible to control the rewarding effect of the biosignal, Lepper's studies on the overjustification effect (Lepper et al., 1973) demonstrated that extrinsic rewards have a detrimental effect on long-term motivation in human subjects, as they are perceived by human subjects as constraints rather than motivations. Extrinsic-reward-based strategies can therefore induce short-term stimulation followed by long-term aversive effects.

Another, more plausible option would be to base the reinforcement signal on intrinsic rewards. Intrinsic rewards are triggered when the action a subject takes is congruent with his internal motivation. When a human subject achieves learning toward proficiency in a skill (in the case of biofeedback, the skill would be discrimination and/or self-maintenance), exercising that skill provides an intrinsic reward. This intrinsic reward is the so-called flow state, obtained whenever a good balance is achieved between task difficulty and skill proficiency (Csikszentmihalyi, 1990). From this perspective, the implicit reward value of the biosignal in biofeedback paradigms would be intrinsic and due to achievement of a flow state, which involves voluntary attentional processes and higher cognitive functions. By integrating higher cognitive functions in the OC paradigm, one moves from behavioral learning theories to cognitive and developmental theories (explored in more depth in Section 3.2), which could explain the recent trend in neurofeedback publications toward cognitive strategies for the training of human subjects (Bagdasaryan and Le Van Quyen, 2013).

3.2. Developmental psychology and schemata

A central question in developmental psychology is how best to understand the acquisition of complex behaviors. Rats cannot surf the internet, dance the tango, or even solve the towers of Hanoi problem. These tasks involve the coordination of complex skills whose emergence cannot be attributed to simple reinforcement learning. To model the acquisition of such complex skills, psychologists have had to move away from learning theories grounded in behavioral psychology and notions of conditioning or reward and toward schemata formation and working memory (WM) span. We

will review these concepts here and explain how they can be used to model biofeedback mechanisms.

Piaget was the first to model human development, with a specific interest in childhood development. In 1926 he introduced the concept of schema (plural schemata), a cognitive structure representing organized knowledge of some part of the world that is acquired on the basis of experience (Piaget, 1971). The concept was further developed by Bartlett (Bartlett, 1932) and later by other developmental and cognitive psychologists. When new elements are encountered, a given schema could either be adapted to assimilate the element through an abstraction process, or be revised in order to accommodate the schema to the new element (Lewis and Durrant, 2011). The schemata theory has been successfully extended to development in adults and is still used to model skill acquisition (Weeks et al., 2013; Plant and Stanton, 2013). A neuroscience perspective regarding schemata formation and integration is presented in Section 4.

While it might appear that behavioral learning theories and developmental schemata integration theories are incompatible perspectives, it is possible to reconcile them. From a connectionist perspective, schemata emerge at the moment they are needed from the interaction of large numbers of much simpler elements all working in concert (Rumelhart et al., 1986). Reinforcement learning at a lower level can interact with integration mechanisms to become higher level skills, as has been suggested in recent cognitive neuroscience models of schemata (Lewis and Durrant, 2011), which we will discuss in Section 4.

Early in the development of schemata theories of skill acquisition, questions began to arise about how these skills evolve, since it became apparent that humans acquire skills through successive non-linear "steps." Strikingly, these steps are even evident in the acquisition of complex skills when children have already acquired their subcomponents. A child can learn motor and cognitive skills through apparently abrupt transformations. In order to model what happens between these discontinuous evolutions, the successors of Piaget (the so-called neo-Piagetians) introduced the concept of memory span (Pascual-Leone and Goodman, 1979; Case, 1985). Memory span is a limit on WM during the execution of tasks, the idea being that it is impossible to keep too many items in mind or execute too many cognitive operations simultaneously. The explanatory power of memory span resides in the explanation of developmental "steps" observed in children. Instead of proposing that schemata are created "out of nowhere," neo-Piagetians theorize mechanisms of progressive integration in which schemata with a higher degree of integration have a lower WM cost. Development, therefore, would move in observable steps, since whenever children have finished integrating their schemata they are suddenly able to coordinate more schemata and perform combinations of tasks. (A neuroscience perspective regarding schemata formation and integration is presented in Sections 4 and 4.2.)

A common observation in developmental skill acquisition is the U-shaped learning curve, representing a three-step process: good performance, followed by bad performance, followed by good performance once again (Carlucci and Case, 2013). The adoption of novel processing strategies leads to an increased cognitive load and to temporary losses of processing efficiency (Pauls et al., 2013; Siegler, 2004). Language acquisition models confirm that U-shaped behavior is unavoidable since human learners are limited by cognitive constraints (Carlucci and Case, 2013). If the cognitive load of a task is too high, performance will decline. This effect was observed early for biofeedback, where a transient decrease of galvanic skin response (usually following a U-shaped evolution) can be observed, representative of the increased attentional demand with the biofeedback (Gatchel et al., 1978; Montgomery, 1988; Freedman and Ianni, 1983; Gevensleben et al., 2014).

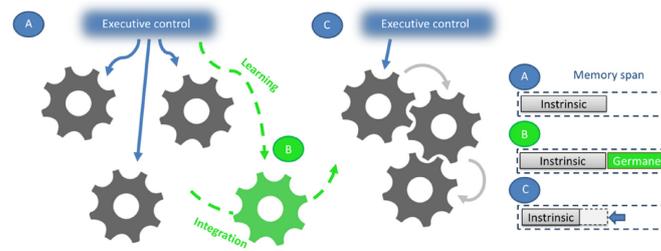


Fig. 3. The neo-Piagetian memory span model, cognitive load theory, and schemata integration. When a subject begins learning a task involving volitional control over a combination of schemata, the memory span is high. When learning begins (A), several storage and processing schemata have to be controlled, inducing a high intrinsic cognitive load. During learning (B), the use of learning schemata (in green) to integrate the processing and storage schemata increases the cognitive load. At this point, performance drops (performance follows a U-shaped curve) as the cognitive load increases (due to a germane cognitive load increase). After learning (C), the schemata are integrated and the cognitive load drops, leading to improvement in performance. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

One explanation for this learning curve can be found in cognitive load theory (CLT) (van Merriënboer and Sweller, 2010). In CLT, WM is considered a resource divided between three different cognitive loads: intrinsic, extraneous, and germane. Extraneous load refers to the complexity of task presentation and is external to the subject. Intrinsic load refers to the amount of WM dedicated to task performance; it is high when element interactivity is high, i.e., when the subject has to process numerous elements (new elements not yet integrated into his own memorized schemata in long term memory). Germane load refers to the process learning load involving induction or “mindful abstraction,” whereby the subject performs abstractions on the schemata associated with the intrinsic cognitive load. We can see how this theory relates to the U-shaped learning curve (as illustrated in Fig. 3): when learning begins, subjects need to devote part of their WM to performing mindful observation of their WM in order to aggregate their schemata into a coherent new process. This increased germane cognitive load will in turn decrease performance. When the new schema is formed, performance improves again (since the intrinsic load is lower due to better schemata integration). It was already defended in 1989 by Zolten that biofeedback indeed follows the CLT predictions: “the better the clients are able to control their autonomic processes, the more efficient will be the organization of those processes when routinization occurs, and the clients will be able to direct their attentional abilities toward other important problem issues” (Zolten, 1989).

The classical perspective of Piaget restricts development of schemata to self-acquired experience (Piaget, 1971). However, both the social cognitive theory of Bandura (Bandura, 1986) and the social learning theories of Vygotski (Valsiner, 2012) placed social interaction at the heart of child development. Children learn more easily when learning is mediated by social interactions with a tutor (Dixon-Krauss, 1996). The tutor provides scaffolding, i.e., elements of a task that are initially beyond the learner’s capacity, thus permitting the child to concentrate upon and complete those elements that are within his range of competence (Wood et al., 1976). What a child is able to do today with instructional scaffolding, he/she will be able to do tomorrow alone (Valsiner, 2012). While the subject has limited abilities, with the help of a supervisor he/she is able to perform more complex tasks. What he/she can do alone is termed the “autonomy zone,” what he/she can do with help is termed the “zone of proximal development,” and what he/she cannot do even with help is termed the “rupture zone.” Numerous experiments have supported this model, demonstrating the direct impact of scaffolding on executive function, WM emergence, and

cognitive self-regulation (Valsiner, 2012; Hammond et al., 2012; Dilworth-Bart et al., 2010; Freund, 1990). Interestingly, cognitive self-regulation corresponds to the definition of self-maintenance in Section 2.1.

We will now illustrate with a simple example why schemata theory accurately models biofeedback effects. As noted in Section 2.2, a recent review (Lehrer and Gevirtz, 2014) attributed the effect of heart rate variability biofeedback to a combination of causes including: homeostasis in the baroreceptors, parasympathetic reflex stimulation, improved gas exchange, mechanical stretching of airways, anti-inflammatory effects, and attentional effects. As explained in Section 2.1, the effect of biofeedback is to improve biosignal control through the acquisition of two skills: discrimination and the self-maintenance. Here, both discrimination and self-maintenance can be seen as complex tasks; though a given subject may know how to sustain his attention, relax, or slow down his breathing, the coordination of these tasks is not necessarily a straightforward process. Similarly, though a subject can monitor his breathing, notice if he/she is relaxed or tense, and observe when his attention drops, the combined monitoring of multiple states can be challenging. In the case of heart rate variability biofeedback, therefore, discrimination and self-maintenance skills could be modeled as schemata. This example is not exceptional, as most biofeedback paradigms include involvement of executive function or attention (see Section 3.4 below). From this perspective, biofeedback provides scaffolding for the subject (Sanders and Welk, 2005), helping him/her to acquire or improve task-related discrimination and self-maintenance schemata.

3.3. Skill learning

Skill learning is a paradigm that describes the mechanisms involved in the acquisition of complex perceptual, cognitive, or motor skills. The effect of feedback is a variable of interest in skill learning—for example, it could be useful for the description of efficient coaching practices for motor skill acquisition. One can identify two significant properties of a motor action (Salmoni et al., 1984): its performance, i.e., the quality of the subject’s own movement (how to do the action); and its result, i.e., the success or failure of the action (what shall be done). The subject can learn about these two properties either by himself or with external help. When the subject has direct access to these two observables, it is termed “intrinsic feedback.” When the information comes from an exter-

nal source (for example, a sports coach or a device), it is termed “external feedback.”

The efficiency of external feedback for skill learning has been the object of several studies and some foundational truths have been demonstrated. First, extrinsic feedback helps to accelerate and facilitate the learning process (Poole, 1991), especially when it is not redundant with internal feedback (Schmidt and Wrisberg, 2007). It has informational functions and motivational properties with important influences on learning (Wulf et al., 2010), but it can also induce dependency (the so-called *guidance effect*): if administration of extrinsic feedback is not appropriate, performance decreases after the feedback is withdrawn (Buchanan and Wang, 2012). Second, the subject must be able to act upon his internal feedback when the external feedback is removed; successful feedback learning, therefore, is an adaptation of internal feedback in a way that incorporates the external feedback (Syzozik et al., 2006). Finally, performance feedback is generally more effective for real-world tasks (Schmidt and Wrisberg, 2007).

The dissociation of performance feedback and result feedback can be observed for instance in skilled typists. Logan and Crump provided skilled typist with fake result feedbacks (Logan and Crump, 2010), either corrected errors that typists made or inserting errors in correct responses. When asked to report errors, typists took credit for corrected errors and accepted blame for inserted errors, claiming authorship for the result feedback. However, their typing rate showed no evidence of these illusions, slowing down after corrected errors but not after inserted errors. This dissociation suggests two error-detection processes: an outer loop sensitive to the appearance of the screen (result feedback) and an inner loop sensitive to keystrokes (intrinsic performance feedback). Another example in motor learning is voice control training or rehabilitation. Visual feedback on voice spectral properties can be used to train singers, and as one would expect, novice and expert singers require training tailored to their individual skill level; while beginners prefer simple and continuous information, experienced singers prefer more complex and discontinuous feedback (Hoppe et al., 2006). Internal result feedback develops with expertise, and therefore simple external result feedback is redundant and ineffective for experts. The results of the Sing & See project (Wilson et al., 2005) are of particular interest, as they illustrate how developmental psychology can explain feedback learning mechanisms: singers' performance dropped during feedback presentation but improved after feedback training (as compared to a control group). This is typical of a U-shaped performance curve (see Section 3.2).

Though motor skill learning theories cannot be directly adapted to explain biofeedback training, their core principles are similar in practice, and assumptions about efficiency of feedback from the motor skill model are likely to hold true for biofeedback. This model can easily be extended to any kind of feedback learning in general, including biofeedback and neurofeedback. The implications of skill learning for neurofeedback has already been debated by Strehl (Strehl, 2014). Skill learning theory models systems with explicit feedbacks, and therefore would relate to model-based RL mechanisms.

3.4. Executive function and attention

Biofeedback could not exist without involvement of executive functions and/or attention. Executive functions comprise the mental processes that enable individuals to take control over otherwise automatic responses of the brain in order to produce goal-oriented behaviors (Lamar and Raz, 2007; Garon et al., 2008; Lezak et al., 2012). They are strongly, but not exclusively, associated with neural networks located in the prefrontal cortex (Miller and Cohen, 2001) (more details in Section 4.3.). These executive functions allow individuals to handle new and/or complex situations where rou-

tine behavior does not exist or would prove suboptimal, and they include processes such as planning, goal setting, decision making, voluntary attention, task switching, set shifting, behavioral and perceptual inhibitions, voluntary emotional regulation, and error correction. In biofeedback paradigms, and especially when training is based on cognitive strategies (see Section 3), several executive functions appear to be essential to setting up an internal reward system (goal setting), integrating feedback information (voluntary attention, set-shifting), and adapting behavior toward self-maintenance (error correction).

Most of the aforementioned cognitive functions interact with attention, a broad concept that can be defined as the set of processes dealing with the allocation of WM to the different neural representations available in the brain (Knudsen, 2007). Many studies point to the common neural mechanisms that support both WM and attention (Ikkai and Curtis, 2011a,b; Gazzaley and Nobre, 2012), reinforcing the idea of an overlap between the two functions. Because high-level cognition relies on WM's limited span (Cowan, 2005; Cowan et al., 2005), attention plays a crucial role in learning tasks where WM is partly occupied by learning schemata (see Section 3.2).

3.5. Working memory models

There are good reasons to hypothesize that WM plays a key role in biofeedback learning. The central role of WM is emphasized in motor skill learning (Seidler et al., 2012), and by definition, this theoretical construct intersects with all cognitive functions (see Section 3.4). While performing any cognitive task, information being processed is stored and maintained in WM. Miller coined the term “working memory” while studying everyday formation, transformation, and execution of plans in the context of behavioral science (Wallace, 1960).

3.5.1. Multiple-component model

Baddeley and Hitch's model (Baddeley and Hitch, 1974) remains the most influential model of WM. The original model included two slave storage subsystems in charge of storage and maintenance of visual and auditory information; and a coordinating system, the central executive. The central executive coordinates the slave subsystems, activates memory traces from long-term memory (LTM), selects coding strategies, and shifts attention. Two main criticisms of the concept of a central executive have been (1) that it is depicted as an homunculus, an all-powerful man running WM, and (2) that the lack of rigorous evidence makes it impossible to falsify (Parkin, 1998). A new slave system, the episodic buffer, was later introduced by Baddeley (Baddeley, 2000). The episodic buffer stores multi-dimensional pieces of information integrated by the central executive into time-ordered episodes, like fragments of a story. These episodes are then linked to multi-dimensional representations in LTM.

3.5.2. Embedded-process model

Cowan's model of WM (Cowan, 1988) outlines more precisely the mechanisms underlying attention and extends the notion of slave subsystems to more general types of encoding. In terms of flow, information enters the brief sensory store and is retained for several hundred milliseconds, whereupon LTM representations (sensory or semantic) become active and remain so for a few seconds. Depending on the salience of the stimuli and/or voluntary attention, the activated memories may enter the focus of attention or remain outside of it (yet still active). The attentional processes are mediated by the central executive, which can direct attention either outward to perceived stimuli or inward to LTM. The processing of activated traces of LTM might lead to controlled actions if information passes through the focus of attention or to automatic

actions otherwise. LTM storage of some coded features occurs automatically. Processing in this model can also be performed on active items outside the focus of attention.

3.5.3. Long-term working memory

Traditional models of WM perform rather well on laboratory tasks. However, the large storage demands of text comprehension and other skilled activity (e.g., good chess players, digit span experts) cannot be explained by models that rely only on temporally limited capacity (Anders Ericsson and Kintsch, 1995). To address this problem, Anders Ericsson and Knitsch proposed their long-term working memory (LTWM) model. Based on experimental findings (Anders Ericsson and Delaney, 1999) that conflicted with other WM models, they proposed the idea that skilled activity in everyday life does not rely heavily on temporal storage. On the contrary, while skills are developing, domain specific semantic structures are built in LTM that allow for efficient coding and fast retrieval, and hence LTM largely mediates expert performance.

3.5.4. Time-based resource-sharing model

A model of WM that proposes an interesting definition of cognitive load is the time-based resource-sharing model (TBRS) (Barrouillet et al., 2004). The main assumption of the TBRS model is that attentional resources, serial in nature, are needed not only for processing information, but are also shared with activation and maintenance processes. This holds true both for complex tasks as well as for simple activities like reading letters or digits.

Within this model, quick pauses are required during processing in order to maintain the memory traces, which would otherwise decay over time. This process does not necessarily correspond to rehearsal in the phonological loop proposed by Baddeley since different mechanisms could occur, such as the rapid and covert retrieval process through attentional focusing proposed by Cowan (Cowan, 1992). This attentional switch might occur constantly and at the micro level, as described in the micro-task-switching process by Towse et al. (Towse et al., 2007). This process is serial in nature at the micro level, yet rapid enough to seem parallel at the macro level.

Due to this attentional constraint, it is important to redefine the notion of cognitive load. A high load condition should involve not only the number of active items, but also the available time that can be devoted to attentional switches to refresh memory traces. If the task allows enough time to ensure proper maintenance of memory traces, it is said to correspond to low cognitive load, and conversely,

if high processing demands leave little time for refreshing, the task is said to involve high cognitive load. In this sense, the concept of load becomes task dependent.

3.6. Volitional action, agency, and fluency

In Section 2.2, we mentioned that biomedical models of biofeedback disagree over the need for volitional control of the regulated biological variables. Volitional action is associated with authorship of the action, a sense of agency or self-agency—the sense that “I am the one who is causing or generating an action” (Gallagher, 2000). In other words, a sense of agency refers to the feeling of controlling an external event through one’s own actions. Agency is at the center of neurocognitive models of schizophrenia as an explanation for volitional delusions (Lafargue and Franck, 2009). Interestingly, agency seems to be linked with both internal and external feedback about self-control (Syznozik et al., 2006) and therefore has a direct relationship to fluency. Fluency is the subjective experience of ease or difficulty associated with completing a mental task (Oppenheimer, 2008) and therefore relates to the perception of self-control or self-regulation. Monitoring of physical efforts by a subject, for example, can lead to a retrospective sense of fluency, which can in turn contribute to a sense of agency (Demanet et al., 2013). This is not a new observation; Maine de Biran proposed in 1805 that the sensation of effort might provide an internal cue for distinguishing self-caused changes from other changes in the environment (Maine de Biran, 1805). Recent reports have shown that a sense of agency would be derived from both a prospective (action selection) and retrospective (action outcome) fluency (Chambon et al., 2014).

This relationship between agency and self-regulation is critically important for biofeedback training. First of all, successful volitional biofeedback induces improved fluency in regulation of the biological variable and consequently involves a sense of agency. Furthermore, self-regulation can be seen as one aspect of executive function, whose depletion has negative effects on task performance—the so-called ego depletion effect (Vinney and Turkstra, 2013). Again, this ego depletion effect predicts a drop in performance during effective biofeedback, in line with developmental psychology models (Section 3.2).

3.7. Synthetic psychological models

Biofeedback is concerned with a specific subtype of skill learning: biological variable regulation. Biofeedback setups provide the

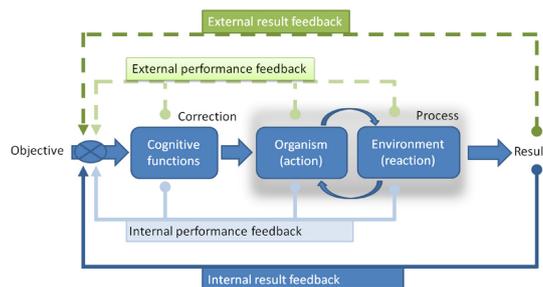


Fig. 4. Four-component biofeedback flow chart. The subject has access to two internal feedbacks (bottom of flow chart). The internal performance feedback corresponds to the discrimination skill. Succeeding or failing to regulate the biological variable is the result feedback. The biofeedback provides either external result or external performance evaluations to the subject (top of the flow chart). The self-maintenance skill integrates both internal and external feedback and regulates the biological variable based on these inputs.

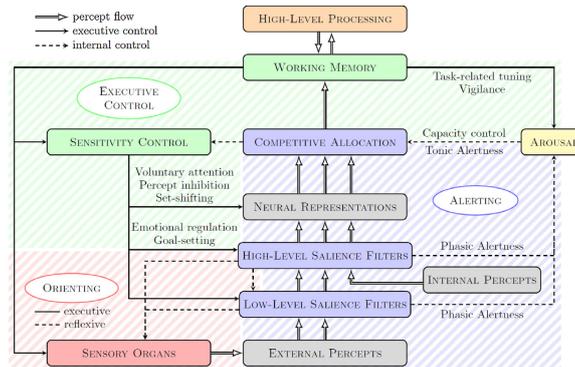


Fig. 5. Integrative model of attention and executive control. The role of attention is assumed to be the allocation of working memory between available percepts. Blue nodes represent the main components of attention, responsible for evaluation and selection of relevant information as well as executive feedback over the selection process. Red nodes are brain structures strongly linked with attention. Long-term memory could be represented as a red node as well. Green nodes are abstract representations of information flow and processing. Executive functions linked to the control of attention are represented in purple. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

user with external feedback, while the discrimination skill is an internal performance feedback. The self-maintenance skill integrates both internal and external feedback and acts on the biological variable. As with motor learning, biological variable regulation seeks an effect involving the organism and its environment and has a directed functional goal; succeeding or failing to reach this goal is the result feedback. The main difference from motor learning resides in the type of action involved: while motor action learning involves sensorimotor processes, biofeedback is more general and can include any kind of biological variable. We can summarize these elements in a general biofeedback flow chart with four types of feedbacks: external result feedback, external performance feedback, internal result feedback, and internal performance feedback (Fig. 4).

A framework for the different executive functions involved in biofeedback, largely inspired by the work of Knudsen (Knudsen, 2007), is shown in Fig. 5. This model includes several levels of salience filters that attribute weights to both external and internal percepts based on their physical, temporal, motivational, and emotional properties (Menon and Uddin, 2010). The resulting neural representations then go through a competitive selection process to determine which information enters WM. This filtering layer is referred to as bottom-up attention and will, for example, allow a loud, unexpected sound to enter almost anyone's WM (in addition to triggering subcortical responses). Top-down signals can alter this selection process by modifying the behavior of salience filters (e.g., emotional regulation) or by enhancing or inhibiting a neural representation that has already entered WM and has gained or lost salience through high-level processing (voluntary attention and percept inhibition, respectively). Feedback signals can also modify the behavior of sensory organs at several levels of this weighting/selection process—for example, by orienting the eyes toward a stimulus to enhance its relative importance in the visual cortex. Other executive functions deal with the temporal allocation of WM and can therefore be considered components of attention. Set-shifting, for example, refers to the ability to switch between different high-level neural representations of a percept on the basis of feedback and is therefore an important mechanism in learning and biofeedback paradigms (Kehagia et al., 2010). Sustained atten-

tion is another key component of attention and refers to the ability to maintain neural representations in WM over time (Gazzaley and Nobre, 2012). This cognitive function is also strongly involved in the learning process described in Section 3.2, as both feedback information and learning schemata should be maintained in WM during the integration process.

As explained in Section 2.1, the interest of biofeedback is to help train two cognitive functions related to a target biological variable: discrimination and self-maintenance. Acquisition of the discrimination skill requires the subject to find an internal or autogenous percept that matches the fluctuation of the external feedback. This process requires the subject to scan the different percepts available to him/her at a given time (selective attention) and to manipulate their different neural representations (set-shifting) in order to find out if a correlation can be established with the feedback. Training of the discrimination skill is greatly facilitated by joint or prior development of the self-maintenance function, i.e., the ability to affect the biological variable voluntarily. Intended modification of the biological variable allows the subject to more easily confirm or contradict a possible correlation between an internal or autogenous percept and the external feedback than would mere observation of natural fluctuations in the biological variable. Development of the self-maintenance skill also requires the subject to try several approaches to infer whether or not a behavior has an influence on the feedback. Both functions are therefore acquired using typical learning strategies that involve reinforcement in the case of a positive correlation and error detection/correction otherwise.

4. Neuroscience perspective

4.1. Neural correlates of schemata formation

Straightforward links can be established between schemata theory and functional neuroanatomy (Johnson and Grafton, 2003; Cannon et al., 2008). Schemata correspond closely to biological networks of neurons usually termed "neural assemblies." A neural assembly is a small set of interconnected neurons that can persist without external stimulus, connected by learning and supported by synchronous firing behavior (Huyck and Pasmore, 2013).

The “information overlap to abstract” (iOta) model of Lewis and Durant (Lewis and Durrant, 2011) theorizes that schemata are created through reinforcement of synaptic connections of overlapping memories: when a group of neural assemblies are activated simultaneously, their common overlapping networks are reinforced. Through progressive abstraction due to synaptic homeostasis, a new assembly of neurons could be gathered into abstract schemata combining elements of these memories. Despite the likelihood that other biological mechanisms may also underlie the formation of schemata, such mechanisms have not yet been described (Huyck and Passmore, 2013), so the iOta model is the most complete available.

The formation of neural assemblies occurs in two steps (Frankland and Bontempi, 2005). First, a transient neuronal assembly is formed to deal with a task, leading to short-term memory organization. The hippocampus probably plays a key role at this stage, especially for episodic memories (Shirvalkar, 2009). Reactivation of the assembly leads to its consolidation and the formation of a long-term memory through reinforcement learning (RL), stored in cortical networks. Classical models assume that memories are consolidated during sleep, but experimental evidence shows that this process can also occur during waking states (Axmacher et al., 2009). The ventromedial prefrontal cortex and the hippocampus may interact at this stage for schema formation and possibly in the representation of partially consolidated schemata (van Kesteren et al., 2010). Furthermore, schema acts as memory containers facilitating encoding: when a schema exists, the assimilation of new mnemonic traces into the schema can occur extremely quickly, and become rapidly hippocampal-independent (Tse et al., 2007).

A functional model of the neural correlates of schemata could be found in the notion of actor strategies in the prefrontal cortex (Koechlin, 2014; Collins and Koechlin, 2012). Actors are task sets driving ongoing behavior, stored in long-term memory. Koechlin's theory (Koechlin, 2014, 2016) provides a model integrating schemata learning and self-control networks. While existing actors are used and reinforced through model-free RL (see Section 3), they are evaluated by the prefrontal cortex (PFC), and monitored by the anterior cingulate cortex (ACC). Cascade of interactions can be observed between the dorsolateral PFC and the anterior cingulate cortex, involved in response evaluation upon action performance (Banich, 2009). When the ACC detects suboptimal strategies, a model-based RL mechanism would be triggered in order to create a new actor. Once a new efficient actor is learned, model-free RL progressively dominates with time. Model-free RL and model-based RL form two cooperative systems with model-free RL driving online behavior and model-based RL working offline in the background to continuously adjust model-free RL (Sutton and Barto, 1998; Gershman et al., 2014; Koechlin, 2016). It can easily be seen that this theory articulates model-free RL mechanisms for schema assimilation and model-based RL mechanisms for schema accommodation (new actor creation), which bridges the gap between the CLT in psychology (Section 3.2) and RL mechanisms in neuroscience. The germane load could find a potential neural correlate in the frontopolar cortex, involved in the cognitively costly evaluation of new strategies in model-based RL by (Koechlin, 2014). A recent study illustrates this effect and indicate a neural correlate of the germane load: subjects exposed to a slow cortical potential followed a U-shaped evolution of neuronal resource allocations, measurable using the contingent negative variation (CNV) at the Cz electrode (Gevensleben et al., 2014), which was not observed in the sham group. The role of the PFC, the cognitive control network and the ACC are discussed with more details thereafter in Sections 4.3, 4.4 and 4.5.

4.2. Schemata and working memory

WM, or the processing of short-term memory, is fundamental to the functioning of schemata. As neural activity persists in subregions of the PFC and posterior parietal cortex (PPC) during maintenance of WM representations (Ikkai and Curtis, 2011a,b), one could consider these two brain regions together as the location of WM neural substrates.

Two subfunctions of WM have been identified: information storage and executive processing of stored data. Neuroimaging evidence links the short-term memory storage function with the ventrolateral PCF (Smith and Jonides, 1999; Stokes, 2015; Ester et al., 2015) and the PPC (Ester et al., 2015; Ikkai and Curtis, 2011a,b). The executive component, on the other hand, appears to be mediated by the dorsolateral PFC (Smith and Jonides, 1999), whose causal role is supported by transcranial magnetic stimulation studies (Mottaghy, 2006).

There are three hypotheses regarding the neural basis of WM storage. First, information could be stored in the PFC and PPC themselves; in fact, brain activity in these areas can be used to reconstruct orientation bars stored in visual WM (Ester et al., 2015). A second hypothesis is that WM is not stored in persistent neural activity, but instead in the combined interaction of ongoing activity and the hidden state (activity-silent states) in the brain's structural connectivity (Stokes, 2015). This hypothesis is supported by the fact that dynamic states of neural networks are combinations of their ongoing activity, underlying connections, and short-term synaptic plasticity (Buonomano and Maass, 2009). The final hypothesis proposes a mediating role for the lateral PFC. Recent studies combining TMS and neural measures have shown that the lateral PFC modulates sensory activity during WM tasks and enhances selectivity of representations in the sensory cortex (Sreenivasan et al., 2014). According to these results, and in line with Cowan's WM model (Section 3.5), the WM would not be stored in the lateral PFC, but instead stored in the sensory cortex and mediated by the lateral PFC (whose activity would therefore be a correlate of sensory cortex activity). There is no consensus yet on these three models. However, a recent study demonstrated that noise learning is accompanied by rapid formation of sharp neural selectivity to arbitrary and complex acoustic patterns within sensory regions (Andrillon et al., 2015). This is the first experimental confirmation that schemata bridge the gap between sensory and memory processes, and a validation of Cowan's hypotheses.

The iOta model is compatible with all three theories of WM storage, fitting best with the second (activity-silent states) theory. The model describes schemata as neural assemblies involving structural networks of neurons, a description that is consistent with activity-silent states. The TBRs model (see Section 3.5.4) is more compatible with the third theory (lateral PFC mediation of WM), as it separates the storage function from the storage location.

4.3. Executive functions and the prefrontal cortex

As mentioned in Section 3.4, executive functions play a key role in the integration of feedback in skill learning. Two frontal brain regions are central to several executive functions (Logue and Gould, 2014): the medial PFC, involved in general attention and set-shifting tasks; and the orbitofrontal cortex, involved in reversal learning and response inhibition tasks.

Koechlin's hierarchical model of cognitive control (Koechlin et al., 2003; Koechlin and Summerfield, 2007) is a multistage architecture along the anterior–posterior axis of the lateral PFC where each stage maintains active representations that are controlled by

higher stages and that exert control on representations in lower stages. Control signals owing to events which occurred in the more and more distant past would arise from successively more anterior cortical regions. In this model, the apex of the prefrontal executive system is implemented in the most anterior prefrontal regions and corresponds to control processes underlying multitasking and the temporary maintenance of pending behavioral episodes. Logan and Crump's hierarchy of loops (Logan and Crump, 2010) involved in result (outer loop) and performance (inner loop) error-detection processes is compatible with this hierarchical model. Together, these models can explain the differences observed between result and performance feedbacks in skill learning (see Section 3.3).

There is general consensus about the nature of the PFC's mediation of executive functions (Smith and Jonides, 1999). PFC areas modulate the activity in sensory cortices, thereby allowing for voluntary control of brain functions. Similarly, emotional regulation involves a network of areas in the PFC, hippocampus, and parahippocampus (Phillips et al., 2008). The PFC most likely plays a central role in executive control of the brain: several reports indicate that top-down signals originating in the LPFC (representing current task goals) implement cognitive control by biasing information flow across multiple large-scale functional networks (Miller and Cohen, 2001; Cole et al., 2013). This specific role in cognitive control will be addressed in the next section.

4.4. Self-control networks

As noted in Section 2.1, the voluntary control of biosignals attempted in biofeedback paradigms depends on two functions: discrimination and self-maintenance. Here we will report recent evidence about the neural correlates of cognitive control, which could stand as potential candidates for the neural basis of self-maintenance. Recently it has been hypothesized that neurofeedback might tune brain oscillations toward a homeostatic point through a top-down regulation mechanism (Ros et al., 2014). If this theory is true, then top-down control of brain functions would play a key role in neurofeedback, even in autonomous (non-volitional) regulation neurofeedback models (see Section 2.2 for a discussion of volitional and autonomous regulation strategies).

The cognitive control network (CCN) is a brain network thought to underlie cognitive control capacity (Dosenbach et al., 2006; Cole and Schneider, 2007); to correlate with fluid intelligence (Cole et al., 2012a,b); and to support executive functions in general (Niendam et al., 2012). Regions within the CCN include the ACC and pre-supplementary motor area (pSMA), inferior frontal junction (IFJ), anterior insular cortex (AIC), dorsal premotor cortex (dPMC), and a subnetwork termed the frontoparietal network (FPN) that includes portions of the lateral prefrontal cortex (LPFC) and the posterior parietal cortex (PPC) (Cole and Schneider, 2007; Cole et al., 2013). The FPN acts as a hub that coordinates cognitive control (Cole et al., 2013); it centralizes functional connections with multiple brain networks and is involved in a wide variety of tasks. Furthermore, these connections form an organized framework, with systematic relationships between the types of tasks and the corresponding connectivity patterns. Consequently, the FPN can coordinate brain networks according to the requirements of the task, thereby enabling the transfer of abilities across tasks. The CCN is considered the neural seat of cognitive control, and therefore is a good candidate for the neural basis of self-maintenance; in a recent fMRI study, sham neurofeedback was indeed associated with activation in three areas of the CCN: the LPFC, ACC, and AIC (Ninaus, et al., 2013).

The CCN is likely not the only neural network supporting the self-maintenance function. In situations of wakeful rest such as day-dreaming, activity in a network of brain areas termed the default mode network (DMN) can be observed (Buckner et al.,

2008). Recent investigations have observed that cognitive control may actually be the outcome of dynamic functional couplings between the FPN system, the cingulo-opercular network, and the DMN (Cocchi et al., 2013). By applying network control theory on human diffusion tensor imaging, Gu et al. recently confirmed that: (i) DMN areas may be important in low cognitive effort tasks, (ii) the FPN and cingulo-opercular areas may be important in high cognitive effort tasks, and (iii) attention areas may be important in manipulating information across different cognitive processes (Gu et al., 2015). Furthermore, the FPN is anatomically positioned to integrate information from the attention system and the DMN (Vincent et al., 2008). From this perspective, the CCN, attentional networks, and DMN would share access to cognitive processes depending on the type of task. This observation confirms recent evidence pointing to correlations between dynamic interactions of the CCN and DMN on the one hand, and cognitive control performance of adolescent subjects on the other (Dwyer et al., 2014). This is also consistent with the dual-process theory mentioned in Section 2.2 (Wood et al., 2014), with the DMN corresponding to low-level processing and the CCN to high-level processing.

Finally, as explained in Section 3.6, a sense of agency would be directly related to the perception of self-maintenance. According to one meta-analysis, self-agency appears to involve the insula and the experience of a "global emotional moment" representative of the sequential integration of perceptive and motivational information (Sperduti et al., 2011). The angular gyrus (AG) may also play a key role in monitoring signals relating to action selection in the dorsolateral prefrontal cortex in order to prospectively inform subjective judgments of control over action outcomes. The online monitoring of these signals by the AG might provide a subject with subjective markers of volition prior to the action itself (Chambon et al., 2013), and therefore the AG might be a neural substrate of the sense of agency (Chambon et al., 2014). The main electrophysiological markers of a sense of agency in EEG signals are the alpha-band relative power in the central, parietal, and right temporal areas, as well as alpha phase coherence in frontal areas (Kang et al., 2013). The correlates of fluidity in EEG are the error potentials reported in the next section.

4.5. Consciousness of errors and error potentials

Action monitoring and error processing are two critical stages of executive control in humans, allowing for efficient behavioral adjustment and optimization of performance. These functions therefore play a central role in skill learning and are good candidates for neuronal markers of the discrimination function defined in Section 2.1.

Correct overt responses are frequently preceded by an early sub-threshold electromyographic burst recorded from the hand that is associated with the incorrect response (Burl and Bonnet, 1999). These bursts that occur in about 20% of correct response trials represent partial errors (Hasbroucq et al., 2009). If the correct response is provided by the subject, this means that the partial error has been identified and corrected, preventing an overt error. Rochet et al. studied whether partial errors are consciously detected by subjects (Rochet et al., 2014), and they showed that less than one-third of partial errors were reported. Even if partial errors are not consciously detected, however, they are being corrected for before producing an overt error.

One might ask: is it helpful to be aware of our errors if two-thirds are not reported but still corrected? Biofeedback could be used to explore brain mechanisms implicated in error monitoring and whether being aware of our errors has consequences on error processing and skill learning. Indeed, errors can be corrected without awareness before they reach the threshold of response. However, in situations where partial errors have been consciously detected,

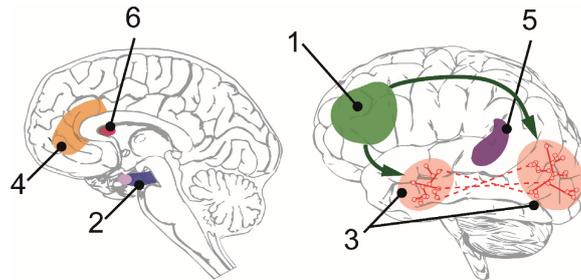


Fig. 6. Feedback learning from a neuroscience perspective. The user is focusing his executive functions on the task, involving the DLPFC (1) and the CCN (not represented in the illustration). Working memory is coordinated by these networks, involving both his hippocampus (2) and the neural assemblies supporting task performance (3) under the supervision of the DLPFC (1). Error monitoring in the ACC (4) allows the user to perceive fluidity, which is then converted into agency by the AG (5). If agency is perceived, and the user is training through a trial-and-error process, then the ventral striatum (6) activates. This leads to the formation of a schemata, progressively integrated and abstracted from the areas involved in the task (3) and consolidated into long-term memory as a skill.

it would be of interest to investigate whether they are corrected through the same processing mechanism or if other adjustments occur (such as a change in strategy). Evoked related potential (ERP) can be useful in exploring error monitoring and might be employed in biofeedback to investigate error monitoring mechanisms.

Errors in reaction-time tasks induce a response-locked ERP that peaks within 50–100 ms after the erroneous response. This ERP is a fronto-central negative deflection, and because it was originally reported as being absent following correct responses, it has been called error negativity or Ne (Falkenstein et al., 1991), or error-related negativity or ERN (Gehring et al., 1993). The Ne is strong evidence for the existence of an action monitoring system able to quickly separate errors from correct responses at the very moment of response (Vidal et al., 2015). In 2000, Vidal et al., by applying the Laplacian transformation, observed a smaller Ne-like potential following correct responses. Actually, Laplacian-transformed data argues in favor of a single-generator hypothesis (Vidal et al., 2015) for the Ne, and the Ne is sensitive to the correctness of the ongoing response. Previous studies suggest that Ne may remain present even when subjects are unaware of having made a partially erroneous eye-movement. It seems that Ne is generated independently of the conscious detection of errors (Nieuwenhuis et al., 2001; Endrass et al., 2007; O'Connell et al., 2007). More generally, the midline frontal theta power—the position and frequencies where Ne is observed—might be the best EEG marker for cognitive control (Cavanagh and Frank, 2014). The error positivity (Pe) is a positive deflection with more parietally distributed ERP than the Ne. It occurs 200–400 ms after a consciously erroneous response (Falkenstein et al., 1991; Nieuwenhuis et al., 2001; Overbeek et al., 2005). The amplitude of Pe is sensitive to the degree of awareness of an error (Dockree and Roberston, 2011) and is larger for conscious than unconscious errors (O'Connell et al., 2007; Charles et al., 2013; Logan et al., 2015). Ne and Pe are therefore neural correlates of volitional self-monitoring of errors.

4.6. Motivation and reward

As explained in Section 3, motivation and reward are central components of biofeedback mechanisms. Motivation involves dopaminergic circuits in the reward system, where the striatum plays a key role (Yager et al., 2015). Monitoring the neural correlates of motivation and reinforcement learning would provide direct insights into biofeedback learning mechanisms.

Volitional self-monitoring of errors is associated with Ne and Pe. When feedback is presented, a specific Ne can be recorded: feedback-related negativity (FRN), which follows the display of negative feedback (Miltner et al., 1997; Walsh and Anderson, 2012). FRN may be the best neural correlate of the reinforcement learning process (Walsh and Anderson, 2012) for several reasons: (1) FRN represents a quantitative prediction error; (2) it is evoked by rewards and by reward-predicting stimuli; (3) FRN and behavior change with experience; and (4) the system that produces FRN is maximally engaged by volitional actions. According to a recent joint EEG-fMRI investigation by Hauser and colleagues (Hauser et al., 2014), FRN could be a neural correlate of surprise signals involving top-down cognitive control in the ACC and may therefore be a good neural marker of fluency in feedback learning (see Section 3.6).

One well-studied ERP component that seems to play a role in reward processing is the P3 (or P300), a positive wave usually peaking between 300 and 600 ms post-stimulus with its largest amplitude at centroparietal scalp sites. When comparing P300 and FRN, reward magnitude (how much reward is received) is reflected by the P300 ERP but not by feedback negativity, while reward valence (positive or negative reward) was reflected by feedback negativity only (Yeung and Sanfey, 2004).

4.7. General model of feedback learning

Feedback learning is the generalization of skill learning to cognitive functioning. The principal brain areas involved in this learning process are illustrated in Fig. 6. The user is performing a learning task that involves both the executive functions and the self-control networks. During learning, working memories and neural assemblies are activated under the monitoring of the central executive (involving the CCN and the DLPFC). Error detection is related to fluency and agency, involving the ACC and the AG: AG playing a key role in the sense of agency (Chambon et al., 2014), while ACC is involved with error detection (Bush et al., 2000). If the protocol leads to motivating conditions (mainly intrinsic motivation in volitional biofeedback), then the reward system activates. Finally, feedback learning leads to the formation of coordinated and integrated neural assemblies through reinforcement of synaptic connections among overlapping memories. The resulting schemata have lower intrinsic cognitive load because the CCN will not need to coordinate the underlying neural assemblies anymore (the task is automated now; see Fig. 3).

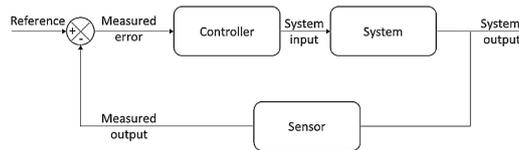


Fig. 7. Basic control theory model. Block Diagrams are graphical representations of processes. This diagram represents a closed-loop model, where a feedback is a correlate of the output returned back to the input to form part of the systems excitation.

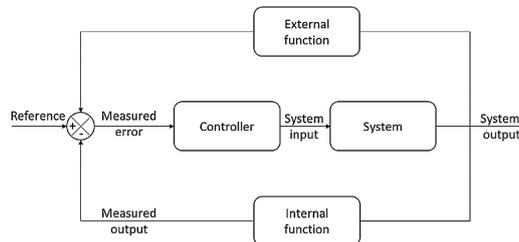


Fig. 8. Explicit model of biofeedback and neurofeedback. The feedback is subdivided into internal and external feedback, where the external feedback comes from an external device and the internal one is within the central nervous system of the subject.

Note that whereas conscious error monitoring involves the ACC, subliminal error monitoring does not (Dehaene et al., 2003). Implicit feedback strategies may indeed not involve the same mechanisms: they are more likely to be based on model-free RL mechanisms (Dayan and Berridge, 2014). Explicit feedbacks would foster schema accommodation with a model-based RL mechanism; whereas implicit feedbacks would foster schema assimilation with a model-free RL mechanism.

5. Engineering perspective

5.1. Process control models of feedback learning

In engineering, process control is a discipline that aims to maintain the output of a process in a certain desired state (Murrill, 2000; Bennett, 1993; Levine, 2010). For example, a thermostat on a heater can turn the heater on or off by comparing the temperature measured by a sensor to a reference temperature. Once the target temperature is reached, the difference between the room temperature and the target temperature is zero, so the thermostat stops the heater. Process control can work in an open loop or by using feedback (Wilts, 1960). It can be continuous or discrete – causing a sequence of events (Levine, 2010). Its application to biomedical engineering models was first suggested by Norbert Wiener in 1948 who introduced cybernetics to model self-regulating mechanisms (Wiener, 1948 2nd revised ed. 1961; Mindell, 2002; Ross Ashby, 1956), and was soon identified as a framework to model biofeedback (Anliker, 1977). It is now commonly used to model systems biology (Cosentino and Bates, 2011), and was recently applied for instance to model biological motor systems (Scott, 2004) and their cognitive control (Frith et al., 2000), or speech acquisition (Tourville and Guenther, 2011; Vinney and Turkstra, 2013), and generally the behavior of biological organisms (Cowan et al., 2014). Biofeedback and neurofeedback are also often modeled using control theory, such as neurofeedback training of implanted brain-computer interface (Guenther et al., 2009), for biofeedback training of postural

control (Ersal and Sienko, 2013), biofeedback techniques in renal replacement therapy (Paolini and Bosetto, 1999), or electrodermal biofeedback of arousal (Parnandi et al., 2013). It was also suggested as a general model for neurofeedback (Ros et al., 2014).

Feedback can be positive or negative (Ross Ashby, 1956; Black, 1934)—terms that can refer either to the way we widen or narrow the gap between reference and measurement of a parameter, or to the valence of the action on the gap, which can have positive or negative emotional connotations.

5.1.1. Controllability

A deterministic system can be fully described by the set of values of all its state variables at a given time. These state variables are characterized by dynamic equations, and prior knowledge is not necessary to predict future states given the current state and current and future values of control variables. Controllability describes the ability to control the internal state of a system from an initial state to a final state in a finite time interval (Kalman, 1960). Controlling a system means being able to move it in all its configuration space using some determined displacements.

5.1.2. Observability

Observability is a measure of a system’s predictability according to knowledge of its external outputs. A system is observable if the current state can be determined in a finite time using only its outputs, for any possible sequence of states and control (Kalman, 1960). If a system is not observable, it means that the current value of some of its state cannot be determined using the output sensors; they are unknown to the controller, but can be estimated under certain conditions.

5.1.3. Basic model

The basic model of feedback in process control theory can be illustrated as in Fig. 7 (Murrill, 2000). A sensor is used to measure the output of a system. This output is then compared to a reference value so that the error between measured output and reference can

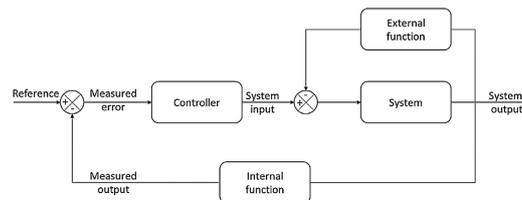


Fig. 9. Implicit model of biofeedback and neurofeedback. The feedback signal is not provided to the subject (controller input), but instead used to change the system conditions.

be reduced. In this model, the comparison only involves the output of the system.

5.1.4. Explicit model of neurofeedback and biofeedback

Biofeedback can be explicit or implicit (Dekker and Champion, 2007; Kuikkaniemi et al., 2010; Nacke et al., 2011); we will first provide a model for explicit biofeedback (see Fig. 8). In this model, both internal and external functions are used to control the current state of the system as compared to the target state.

5.1.5. Implicit model of neurofeedback and biofeedback

The process control model of implicit biofeedback is shown in Fig. 9. In this model, there are two kinds of comparisons: the system can use both internal and external functions to reach a target state. Internal functions refer to inner sensors of the system, while external functions are not directly accessible. External feedback is provided as an input to the system (it has an impact on the system input but not on the error measurement).

5.2. Limits of process control models

When using modelling biofeedback, one should bear in mind the typical limitations of process control methods, reflected in the following five “good practice” precautions.

5.2.1. Linearity

Linear process control models are generally only applicable to linear systems, and when applied to non-linear systems, definitions are only valid for small movements in the neighborhood of a functioning point (Trentelman et al., 2001). Physiological regulation is typically non-linear, and consequently, biofeedback systems need to be individually calibrated, with each user having his own functioning point depending upon both his physiology and his proficiency at regulating the biosignal of interest. Furthermore, the regulation task should also target small enough variations in performances to prevent non-linearity effects. In addition, neurophysiological regulation is allostatic (Sterling, 2004): the brain performs predictive regulations and retune its parameters according to changes. Therefore, experimental protocols should take into account the fact that individuals’ reference point may vary over time depending on task demands and learning.

5.2.2. Stability

Process control systems can be stable or unstable (Ross Ashby, 1956; Routh and Fuller, 1975; Lopez-Caamal et al., 2014). Unstable systems can be heavily perturbed by the slightest change in the input command, whereas stable systems can regulate even the most discontinuous perturbations (e.g., Dirac pulses, which have finite impulse responses). Biofeedback systems are in most cases unstable, and consequently, tolerance to variations around the functioning point is fairly limited.

5.2.3. Temporality: transients and steady-states

When the command changes in complex systems, transient variations typically occur before the system reaches its steady-state (Wilts, 1960). Controlling the amplitude of these variations usually leads to a tradeoff between convergence speed and transient variation amplitude. In other words, fast systems tend to have lots of fluctuations before they reach their goals, whereas slow systems tend to be more precise. This means that the temporality of biofeedback can be a crucial issue: while transient variations correspond to task performances, steady-state error relates to the task result (see Section 3.3 about performance and results). Consequently, continuous feedback about transient states is usually more efficient than discrete feedback about steady-state errors, unless steady-state error perception is not readily available to subjects. For example, in sports training, result feedback can provide useful information to beginners but is of less interest to trained subjects. In biofeedback, the type of feedback presented (transient or steady-state) has to match the subject’s level of fluency in the task.

5.2.4. Precision

In process control, a system’s precision (or accuracy) is defined by its ability to reach a zero steady-state error (Levine, 2010). This precision, or static error, is one of the key estimates of the system’s performance. Therefore, the precision of the biofeedback system (i.e., the precision of biosignal monitoring, the subject’s performance with or without feedback) should always be evaluated.

5.3. Serious games

“Serious games” are games with teaching, training, and informational purposes that utilize play as motivational leverage (Abt, 1970; Prensky, 2001; Michael and Chen, 2006; McGonigal, 2011). Such games have been designed and engineered to stimulate motivation in subjects learning new tasks. Video gaming has several effects on cognitive functions, and in particular may be efficient training for learning how to learn (Bavelier et al., 2012): action video game players have been shown to learn how to extract regular patterns in their environment, thereby improving their ability to learn new tasks. Furthermore, video gaming may lead to lasting changes in reward processing mechanisms (Lorenz et al., 2015). For example, it has been shown that cancer patients playing a serious game to encourage treatment-related behavior markedly activated neural circuits implicated in reward (caudate, putamen, and nucleus accumbens) as compared to patients observing the same audio-visual stimuli without playing (Cole et al., 2012a,b). Biofeedback can be considered a type of serious game: the user “plays” with his biological variable through an interface. Understanding the effective design of serious games is therefore critical to knowing how to design efficient biofeedback systems.

Games are interesting learning strategies because they stimulate motivation and therefore the reward system. Humans have

Table 1
Integrative perspective on biofeedback models, from biomedicine to neuroscience. WM = working memory, LPFC = lateral prefrontal cortex, CCN = cognitive control network.

Biomedicine	Engineering	Psychology	Neuroscience	Psychoengineering
Discrimination	Observability	Cognitive load, WM Agency	LPFC, sensory cortex Insula, Angular gyrus CCN, Error potentials	Perceptibility Autonomy
Self maintenance	Volitional	Explicit paradigms	Reward system, Ventral striatum Neural assembly formation	Mastery Motivation
Biosignal regulation	Autonomous	Implicit paradigms Controllability		Learnability

genuinely high motivation to play video games because they stimulate intrinsic motivation factors, i.e., psychological needs of mastery, autonomy, and relatedness (Przybylski et al., 2010; Lorenz et al., 2015). Several studies have been published on video games and flow (Olson, 2010; Swanson and Whittinghill, 2015), a state of being pleasantly and completely absorbed in a goal-driven activity with hyper-focused attention (Csikszentmihalyi and LeFevre, 1989). The flow state occurs when information processing matches the user's aptitudes and the task becomes a realizable challenge (neither too frustrating nor too boring). According to Csikszentmihalyi, the amount of information a human subject can process amounts to a bit rate of 126 bits/s (Csikszentmihalyi and Csikszentmihalyi, 1992), placing a higher bound on manageable cognitive load (which is modulated by the person's skills). This intrinsic motivation is mainly reported as "fun" by the video game player (Olson, 2010), associated with biological rewards with dopamine release in the ventral striatum (Lorenz et al., 2015).

Certain errors must be avoided to take full advantage of the "fun factor" in biofeedback treatments. Simply because a process is required during game play does not guarantee changes in that process (Bavelier et al., 2012). Unfortunately in some clinical studies the goal has been to "entertain" children with "EEG-driven games," rather than really applying a learning procedure the children could benefit from for a longer period (Arns et al., 2015). The game should be designed to induce training, and this is done by controlling the game's validity—in particular, its predictive validity, proving that performance in the game leads to better outcomes in reality (Graafland et al., 2014). In the field of biofeedback, the problem of transfer is as important as it is for serious games; the skill must be transferable to real life or the user will not benefit from treatment.

6. Psychoengineering model

6.1. The missing keystone: toward a psychoengineering model

In the previous sections, we have explored the various existing models of biofeedback: biomedical, psychological, neuroscience and bioengineering perspectives. We could argue in favor of any of these four perspectives, as each one answers a set of critical questions. However, we believe that a blended model would best describe the mechanisms of biofeedback and produce useful experimental paradigms. This model should represent the perspective of biofeedback itself and bridge the gaps among the aforementioned four disciplines. From a biofeedback perspective, the brain is regulating its own control over biosignals, thereby building itself anew. We have therefore coined the term "psychoengineering" to define our perspective and will attempt to develop such a model in this section.

6.2. Bridging the disciplinary gaps

First, we recapitulate the key points of the above-mentioned four models in the table below (Table 1). As we can see, there

is no direct mapping between the applied models (biomedicine and engineering perspectives) and the theoretical models (psychology and neuroscience perspectives) of self-maintenance. We can identify five key properties of an efficient biofeedback system: perceptibility, autonomy, mastery, motivation, and learnability. Controlling these five variables is necessary for evaluation of a biofeedback prototype.

Perceptibility refers to the potential for the subject to access the perception of the biosignal he/she has to regulate. Autonomy refers to the potential for the subject to regulate the biosignal by himself, without the help of biofeedback, once the training protocol is over. Mastery refers to the degree of control the subject can exercise over the biosignal. Motivation refers to the reward system of the biofeedback device—the reinforcement signal that will induce learning. Learnability refers both to the conditions for achieving long-term memory formation (e.g., sufficient amount of time and repetitions) and to the possibility of learning itself.

7. Conclusion

The learning mechanisms involved in biofeedback should be thoroughly investigated, as the existing literature is largely insufficient to understand biofeedback and explain how it works. We conclude thereafter with five directions that ought to be pursued to better investigate these mechanisms and to improve biofeedback and neurofeedback protocols. These guidelines are representative of the existing literature and should not be seen as established laws but rather as future research directions. They can be used to design a good practice guide for biofeedback and neurofeedback—a tool that is of critical importance to the clinical evaluation of these interventions (Micoulaud-Franchi et al., 2015). Thanks to these guidelines, we hope future biofeedback studies will reach higher standards.

Note that different standardized psychological scales are mentioned for each property, complicating the investigation of feedback protocols. Performing all these evaluations during online feedback procedures might lead to an increased cognitive load for the subject, which could lead to negative interactions with the feedback protocol. Therefore, developing a general feedback learning experience scale involving the main items of all five properties may provide us with a new and useful direction in biofeedback research. Furthermore, one should distinguish here the research purposes (when one evaluates a feedback procedure) and the clinical purposes (when one uses a feedback procedure for treatment). The evaluation of psychological scales during feedback trials is certainly useful for research purposes; however for the final clinical applications such evaluations could be pointless in many cases.

7.1. Investigating and promoting perceptibility

An efficient biofeedback system has to ensure that both the external and internal signals of interest can be perceived with sufficient precision and be effectively organized so that their bit rate will not exceed the user's perception capabilities. One method that

might improve perceptibility in explicit models would be to provide multimodal feedbacks (Lotte et al., 2013): each sensory modality would correspond to a different slave subsystem of working memory. Consequently, a sensory modality not involved in the task should be preferred in the feedback design (see Section 3.5 about working memory models).

From a psychological perspective, perceptibility is related to cognitive load, which can be measured using scales such as the NASA TLX (Hart and Staveland, 1988) or equivalent standardized measures while the subject performs the feedback learning task. The cognitive load is expected to be anti-correlated with the U-shaped evolution of performance (Pauls et al., 2013; Siegler, 2004) and should not reach too high a level, or the subject will experience a cognitive overload and a subsequent loss of motivation. Unfortunately, the issue of cognitive load is often overlooked or ignored in biofeedback studies. For instance, in Angelakis et al. (2007), the same neurofeedback is presented both in auditory and visual modalities (Angelakis et al., 2007) – without any discussion about the impact on cognitive load of this strategy. In Keizer et al. (2010), an auditory neurofeedback rate is bounded to a maximum of 1 feedback per second (Keizer et al., 2010). The amount of information a given subject can process during 1 s is limited, this limit has an interaction with the extraneous load of this task. In Kober et al. (2015), an SMR neurofeedback uses 3 bars, the subject having to modulate both SMR, alpha range and theta range (Kober et al., 2015). As the theta and beta range bars were used to prevent muscle contraction and eye blinks, they could have been replaced by an auditory feedback, and thereby the cognitive load could have been decreased. In Penzlin et al. (2015), a visual feedback is presented to indicate heart rate variability, while another visual cue is presented to indicate breath-pacing (Penzlin et al., 2015). The breath pacing cue could have been auditory, likely reducing the cognitive load.

Furthermore, one has to ensure that the informative external and internal feedback can be perceived as well. This concerns the validity of the external feedback signal: one must demonstrate that the signal is indeed correlated with biosignal regulation. It also concerns the precision of the feedback signal—a classical modelling problem, this precision is a test error that should be evaluated on an independent test set (not on the database used to develop the feedback model). For instance, in neurofeedback, the appropriate approach is to evaluate precision using the same methods as in brain-computer interface paradigms.

Finally, evaluating internal feedback is an observability issue: without the presence of internal feedback there is nothing to be learned. This evaluation can be achieved using psychophysiological scales measuring perception of the internal biosignal: subjects with a complete infirmity in the trained biosignal regulation would not be good candidates for a neurofeedback procedure (since they will never be able to develop autonomy, as explained in the next section). Instead, they would be limited to using the biofeedback system as a palliative measure—something akin to a wheelchair, which cannot be used to rehabilitate movement in hemiplegic individuals (though still useful to them). Inter-individual differences in the ability to monitor interoceptive signals, to concentrate on one's own internal representations and inhibit external, task-irrelevant, stimulation, should be tracked (Corbetta and Shulman, 2002; Burgess et al., 2007). Measuring these individual profiles in biofeedback subjects could be of use to adapt the protocol to individual needs. For instance, Lazarov et al. reported that individuals with obsessive compulsive disorders may suffer from interoceptive deficits, with deficits in internal signal perception, when exposed to biofeedbacks (Lazarov et al., 2010). One could use scales such as Rotter's Locus of control scale to evaluate whether subjects are internally or externally oriented (Rotter, 1966). Internal state per-

ceptibility might also be promoted by combining biofeedback with mindfulness interventions and strategies (Khazan, 2013).

7.2. Investigating and promoting autonomy

From a psychological perspective, autonomy could be promoted following the "guidance hypothesis" (Winstein and Schmidt, 1990; Strehl, 2014). Biofeedback aims to be a scaffolding system rather than palliation for a missing internal signal; otherwise, learning cannot occur. The biofeedback signal should help the subject identify his own internal signals and become progressively more independent of the external feedback, promoting the user's sense of agency. The biofeedback protocol should be as close to reality as possible (high predictive validity, with feedback progressively withheld to promote memorization and intrinsic motivation; see Sections 3 and 3.5.3). For instance, in O'Connell et al. (2007), a protocol promoting autonomy in an explicit biofeedback setting is presented (O'Connell et al., 2007): volitional control is promoted by allowing the subject to progressively initiate the biofeedback task, instead of externally cued; furthermore the subject has to progressively gain autonomy by learning to rely on his internal feedback, the final training step being performed with a withheld external feedback. Such procedures will induce generalization, a process whereby the learner control is progressively experienced without feedback (Sherlin et al., 2011; Strack, 2011).

Predictive validity is necessary to allow transfer from the task-training protocol to real-life positive outcomes. It can be investigated using task performance properties: the so-called "game metrics" used in serious-game designs. These game metrics must be reliable, valid, and cause-specific (Graafland et al., 2012). For example, in neurofeedback, predictive validity requires specificity of the feedback signal: is it targeting only the function to be regulated or a confused signal involving the target function together with additional brain systems? The feedback setup is also of interest: is the training related to real-life conditions or to an abstract conditioning protocol that has no meaning for the subject? Virtual reality setups, for example, seek to improve predictive validity by immersing the subject in a realistic task environment.

If the sense of agency is too low, the biofeedback protocol will not trigger intrinsic motivation and could have a negative impact on learning. Sense of agency can be measured using scales such as SOARS (Polito et al., 2013) or equivalent standardized measures. Other implicit, preverbal, measures such as action-outcome temporal compression or sensory attenuation following voluntary action could also be used to estimate agency (Brown et al., 2013; Dewey and Knoblich, 2014).

From a neuroscience perspective, monitoring the neural correlates of agency could be attempted by measuring the alpha-band relative power and phase coherence during feedback performance.

7.3. Investigating and promoting mastery

Biofeedback systems should provide the user with the possibility to experiment with a progressive experience of control over the regulatory task, promoting the user's sense of fluency. Mastery can be promoted by maintaining a reasonable challenge level, which can be achieved by breaking the treatment down into several sessions of progressive difficulty. In order to respect conditions of linearity and stability, a typical solution is to estimate a psychophysiological curve of subject performance during a calibration phase. This curve estimates both the optimal functioning point and the tolerance to variations around this point. The curve could then be optimized online while the subject is training with the biofeedback system. Task difficulty could either be regularly recalibrated

CHAPTER 5. APPENDIX: PUBLISHED WORKS

at the beginning of each session, or controlled in real-time using an adaptive calibration strategy.

A sense of fluency can be measured using scales such as SCS (Dong et al., 2015) or equivalent standardized measures. Task performance (such as biosignal modulation ability) or cognitive load physiological markers (such as galvanic skin response, or CNV in Cz) are objective but indirect indicator of mastery, and should follow a U-shaped evolution (Pauls et al., 2013; Siegler, 2004; Gevensleben et al., 2014) in explicit feedback protocols. Despite U-shape cognitive demand evolution was first reported in 1978 in biofeedback systems (Gatchel et al., 1978), and is still reported in recent investigations (Gevensleben et al., 2014), most studies investigate performances before and after feedback administration, instead of during – and are therefore ignoring this issue. Note however that this U-shaped evolution might not be observed in implicit reward feedback systems, as the subject is not focusing his attention directly onto the feedback. Optimally, the type of feedback (steady-state or transient) should be consistent with the subject's fluency without feedback. Fluent subjects will not be interested in steady-state discrete feedback but rather in transient continuous feedback, whereas subjects with low fluency may find discrete steady-state feedback useful.

From a neuroscience perspective, FRN may be a good neural marker of fluency in feedback learning and can be measured during tasks comparing naive, trained, and control subjects while they receive feedback (sham feedback for control subjects).

7.4. Investigating and promoting motivation

Within the five properties of efficient biofeedback systems, motivation is probably the most important research avenue. Most existing biofeedback systems are actually extremely boring: the subject sits in a chair and observes a biosignal correlate over a long period of time. A biofeedback system should be motivating (targeting extrinsic or intrinsic motivation) to best promote learning.

From a psychological perspective, though it is well-known that human interactions are catalysts of intrinsic motivation (Ryff and Keyes, 1995), biofeedback and neurofeedback paradigms are too often based on solitary human-computer interactions, and the “human variable” is seldom mentioned or investigated. Much biofeedback research seems to assume a treatment model, as if biofeedback is a procedure “done to” an individual (Yucha and Montgomery, 2008). As was previously stated by Strehl, neurofeedback and biofeedback will always take place within a patient-therapist interaction (Strehl, 2014). Furthermore, it should be noted that this human factor can have an effect both on feedback groups and on control groups in controlled studies (possibly biasing outcomes). Interactions with instructors are key motivational variables (Middaugh et al., 2001; Khazan, 2013) that should be taken into account and evaluated rigorously, for instance using principles taken from instructional design (Lotte et al., 2013).

Finally, from the perspective of OC, the reward percentage (positive feedback), the reward delay and the strategy of reward presentation can also play a key role (Sherlin et al., 2011). In any case, the subjective experience of motivation should be controlled, for instance using items from standardized flow-state evaluation scales such as the FSSOT (Yoshida et al., 2013) or equivalent standardized measures.

From a neuroscience perspective, monitoring the neural correlates of motivation and reinforcement learning would be of great interest. For instance, EEG signatures such as Ne, Pe, FRN, P300, or midline frontal theta power would provide direct insights into biofeedback learning mechanisms.

7.5. Investigating and promoting learnability

Learnability introduces a controllability issue: is the subject able to regulate his biosignal—at least slightly—before the biofeedback or neurofeedback protocol starts? Otherwise, the subject will never be able to learn anything: whatever the precision of the biofeedback, it cannot be used to train nonexistent internal mechanisms. This can be evaluated by determining the subject's fluency without feedback before training begins, which can be measured using scales such as the SCS (Dong et al., 2015).

From a neuroscience perspective, it could also be of great interest to measure a subject's aptitude in brain wave modulation as an indicator of his ability to be trained by neuro or biofeedback. For instance, performance in neurofeedback is usually defined as the ability to up-regulate the targeted neuromarker during feedback training sessions (Escolano et al., 2012; Witte et al., 2013; Zoefel et al., 2011; Reichert et al., 2015; Escolano et al., 2011). The investigation of biomarkers predicting learnability is of great interest for the design and evaluation of efficient bio and neurofeedback, and should be generalized. For instance, Reichert et al. reported a relationship between the controllability of the biosignal (ability to modulate the SMR) and the measurement of an EEG marker (rest signal pre-training value) in SMR neurofeedback (Reichert et al., 2015).

For other types of biofeedback, learnability could be measured by evaluating the modulation performance of the subject during the first training sessions: a low initial performance in explicit biofeedback (i.e. an absence of improvement, or an absence of aptitude to modulate the biosignal) would indicate poor learnability.

Finally, learning is constrained by mechanisms of long-term memory formation. Learning follows a succession of steps: memories are abstracted into functionally efficient schemata (see Section 4) and progressively consolidated. This process takes time, and it calls for a succession of sessions separated by nights of recuperation (sleep being a necessary ingredient for memory consolidation). The number of sessions, session duration, and time intervals between sessions are therefore all crucial parameters of biofeedback and neurofeedback protocols, and the long-term effects of feedback training should be evaluated to determine training stability.

References

- Abt, C., 1970. *Serious Games*. The Viking Press, New York.
- Abukonna, A., Xiaolin, Y., Zhang, C., Zhang, J., 2013. Volitional control of the heart rate. *Int. J. Psychophysiol.* 90, 143–148.
- Anders Ericsson, K., Delaney, P.F., 1999. Long-term working memory as an alternative to capacity models of working memory in everyday skilled performance. *Models of working memory: mechanisms of active maintenance and executive control*, 257–297.
- Anders Ericsson, K., Kintsch, W., 1995. Long-term working memory. *Psychol. Rev.* 102 (2), 211–245.
- Andrillon, T., Kouider, S., Agus, T., Pressnitzer, D., 2015. Perceptual learning of acoustic noise generates memory-evoked potentials. *Curr. Biol.* 25 (21), 2823–2829.
- Angelakis, E., Stathopoulou, S., Frymiare, J.L., Green, D.L., Lubar, J.F., Kounios, J., 2007. EEG neurofeedback: a brief overview and an example of peak alpha frequency training for cognitive enhancement in the elderly. *Clin. Neurophysiol.* 21 (1), 110–129.
- Anliker, J., 1977. Biofeedback from the Perspectives of Cybernetics and Systems Science. In: Beatty, J., Legewie, H. (Eds.), *Biofeedback and Behavior*, NATO Conference Series, 2. Plenum Press, New York, USA, pp. 21–45.
- Arns, M., 2011. Personalized medicine in ADHD and depression: a quest for EEG treatment predictors. *Ipskamp Drukkers*, Amsterdam.
- Arns, M., Heinrich, H., Ros, T., Rothenberger, A., Strehl, U., 2015. Editorial: Neurofeedback in ADHD. *Frontiers in Human Neuroscience* 9 (602), 1–3.
- Axmacher, N., Draguhn, A., Elger, C.E., Fell, J., 2009. Memory processes during sleep: beyond the standard. *Cell. Mol. Life Sci.* 66, 2285–2297.
- Baddeley, A.D., Hitch, G., 1974. Working memory. In: Bower, G.H. (Ed.), *The Psychology of Learning and Motivation: Advances in Research and Theory*, vol. 8. Academic Press, New York, pp. 47–89.
- Baddeley, A., 2000. The episodic buffer: a new component of working memory? *Trends Cogn. Sci.* 4 (11), 417–423.

- Bagdasaryan, J., Le Van Quyen, M., 2013. Experiencing your brain: neurofeedback as a new bridge between neuroscience and phenomenology. *Front. Hum. Neurosci.* 7 (680), 1–19.
- Bandura, A., 1986. *Social Foundations of Thought and Action: A Social Cognitive Theory*. Prentice-Hall, Englewood Cliffs, NJ.
- Banich, M.T., 2009. Executive function: the search for an integrated account. *Curr. Direct. Psychol. Sci.* 18, 89–94.
- Barrouillet, P., Bernardin, S., Camos, V., 2004. Time constraints and resource sharing in adults' working memory spans. *J. Exp. Psychol. Gen.* 133 (1), 83–100.
- Bartlett, F.C., 1932. *Remembering: A Study in Experimental and Social Psychology*. Cambridge University Press, Cambridge, England.
- Bavelier, D., Green, C.S., Pouget, A., Schrater, P., 2012. Brain plasticity through the life span: learning to learn and action video games. *Annu. Rev. Neurosci.* 35, 391–416.
- Bellack, A.S., Hersen, M., 1988. *Behavioral Assessment: A Practical Handbook*, 3rd ed. Pergamon Press, Elmsford, NY (USA).
- Black, H.S., 1934. Stabilized feed-back amplifiers. *Electr. Eng.* 53, 114–120.
- Bostrom, N., 2006. A history of transhumanist thought. *Anal. Metaphys.* 5 (1–2).
- Box, G.E., Draper, N.R., 1987. *Empirical Model Building and Response Surfaces*. John Wiley & Sons, New York, NY.
- Brenner, J., 1974. A general model of voluntary control applied to the phenomenon of learned cardiovascular change. In: Obrist, P.A., Black, A.H., DiCara, L.V. (Eds.), *Cardiovascular Psychophysiology*. Aldine, Chicago, pp. 365–391.
- Buchanan, J.J., Wang, C., 2012. Overcoming the guidance effect in motor skill learning: feedback all the time can be beneficial. *Exp. Brain Res.* 219 (2), 305–320.
- Buckner, R.L., Andrews-Hanna, J.R., Schacter, D.L., 2008. The brain's default network: anatomy, function, and relevance to disease. *Ann. N.Y. Acad. Sci.* 1124 (1), 1–38.
- Buonomano, D.V., Maass, W., 2009. State-dependent computations: spatiotemporal processing in cortical networks. *Nat. Rev. Neurosci.* 10, 113–125.
- Burle, B., Bonnet, M., 1999. What's an internal clock for? From temporal information processing to temporal processing of information. *Behav. Process.* 45, 59–72.
- Bush, G., Luu, P., Posner, M., 2000. Cognitive and emotional influences in anterior cingulate cortex. *Trends Cogn. Sci.* 4 (6), 215–222.
- Cannon, R., Lubar, J., Baldwin, D., 2008. Self-perception and experiential schemata in the addicted brain. *Appl. Psychophysiol. Biofeedback* 33 (4), 223–238.
- Canguilhem, G., 1966. *Le normal et le pathologique, réédité en 1979*. PUF, collection Galien, Paris.
- Caria, A., Sitaram, R., Birbaumer, N., 2011. Real-time fMRI: a tool for local brain regulation. *Neuroscientist* 18, 487–501.
- Carlucci, L., Case, J., 2013. On the necessity of U-shaped learning. *Top. Cogn. Sci.* 5 (1), 56–88.
- Case, R., 1985. *Intellectual Development: Birth to Adulthood*. Academic Press, Orlando.
- Cavanagh, J., Frank, M., 2014. Frontal theta as a mechanism for cognitive control. *Trends Cogn. Sci.* 18 (8), 414–421.
- Chalmers, D.J., 1995. Facing up to the problem of consciousness. *J. Conscious. Stud.* 2, 200–219.
- Chambon, V., Wenke, D., Fleming, S.M., Prinz, W., Haggard, P., 2013. An online neural substrate for sense of agency. *Cereb. Cortex* 23 (5), 1031–1037.
- Chambon, V., Sidarus, N., Haggard, P., 2014. From action intentions to action effects: how does the sense of agency come about? *Front. Hum. Neurosci.* 8 (220), 1–9.
- Charles, L., Van Opstal, F., Marti, S., Dehaene, S., 2013. Distinct brain mechanisms for conscious versus subliminal error detection. *Neuroimage* 73, 80–94.
- Cocchi, L., Zalesky, A., Fornito, A., Mattingley, J.B., 2013. Dynamic cooperation and competition between brain systems during cognitive control. *Trends Cogn. Sci.* 17, 493–501.
- Cole, M.W., Schneider, W., 2007. The cognitive control network: integrated cortical regions with dissociable functions. *Neuroimage* 37, 343–360.
- Cole, M.W., Yarkoni, T., Repovš, G., Anticevic, A., Braver, T.S., 2012a. Global connectivity of prefrontal cortex predicts cognitive control and intelligence. *J. Neurosci.* 32 (26), 8988–8999.
- Cole, S.W., Yoo, D.J., Knutson, B., 2012b. Interactivity and reward-related neural activation during a serious videogame. *PLoS One* 7 (3), 1–9.
- Cole, M.W., Reynolds, J.R., Power, J.D., Repovš, G., Anticevic, A., Braver, T.S., 2013. Multi-task connectivity reveals flexible hubs for adaptive task control. *Nat. Neurosci.* 16, 1348–1355.
- Cowan, N., Elliott, E.M., Saults, J.S., Morey, C.C., Mattox, S., Hismjatullina, A., et al., 2005. On the capacity of attention: its estimation and its role in working memory and cognitive aptitudes. *Cognit. Psychol.* 51 (1), 42–100.
- Cowan, N.J., Ankaran, M.M., Dyhr, J.P., Madhav, M.S., Roth, E., Sefati, S., Sponberg, S., Stamper, S.A., Fortune, E.S., Daniel, T.L., 2014. Feedback control as a framework for understanding tradeoffs in biology. *Integr. Comp. Biol.* 54 (2), 223–237.
- Cowan, N., 1988. Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychol. Bull.* 104 (2), 163–191.
- Cowan, N., 1992. Verbal memory span and the timing of spoken recall. *J. Mem. Lang.* 31 (5), 668–684.
- Cowan, N., 2005. *Working Memory Capacity*. Psychology Press, Hove, East Sussex, UK.
- Csikszentmihalyi, M., LeFevre, J., 1989. Optimal experience in work and leisure. *J. Pers. Soc. Psychol.* 56, 815–822.
- Csikszentmihalyi, M., 1990. *Flow: The Psychology of Optimal Experience*. Harper & Row, New York.
- Csikszentmihalyi, M., Csikszentmihalyi, L.S., 1992. *Optimal Experience: Psychological Studies of Flow in Consciousness*. Cambridge University Press, New York.
- Dehaene, S., Artiges, E., Naccache, L., Martelli, C., Viard, A., Schifxpsl#rhoff, F., Recasens, C., Martinot, M.L., Leboyer, M., Martinot, J.L., 2003. Conscious and subliminal conflicts in normal subjects and patients with schizophrenia: the role of the anterior cingulate. *Proc. Natl. Acad. Sci. U.S.A.* 100 (23), 13722–13727.
- Demanet, J., Muhle-Karbe, P.S., Lynn, M.T., Blotenberg, I., Brass, M., 2013. Power to the will: how exerting physical effort boosts the sense of agency. *Cognition* 129 (3), 574–578.
- Dilworth-Bart, J., Poehlmann, J., Hilgendorf, A.E., Miller, K., Lambert, H., 2010. Maternal scaffolding and preterm toddlers' visual-spatial processing and emerging working memory. *J. Pediatr. Psychol.* 35 (2), 209–220.
- Dixon-Krauss, L., 1996. *Vygotsky in the Classroom, Mediated Literacy Instruction and Assessment*. Longman Publishers, White Plains, NY.
- Dockree, P.M., Robertson, L.H., 2011. Electrophysiological markers of cognitive deficits in traumatic brain injury: a review. *Int. J. Psychophysiol.* 82, 53–60.
- Dong, M.Y., Sandberg, K., Bibby, B.M., Pedersen, M.N., Overgaard, M., 2015. The development of a sense of control scale. *Front. Psychol.* 6 (1733), 1–14.
- Dosenbach, N., Visscher, K., Palmer, E., Miezin, F., Wenger, K., Kang, H., et al., 2006. A core system for the implementation of task sets. *Neuron* 50 (5), 799–812.
- Dwyer, D.B., Harrison, B.J., Yücel, M., Whittle, S., Zalesky, A., Pantelis, C., et al., 2014. Large-scale brain network dynamics supporting adolescent cognitive control. *J. Neurosci.* 34 (42), 14096–14107.
- Endrass, T., Reuter, B., Kathmann, N., 2007. ERP correlates of conscious error recognition: aware and unaware errors in an antisaccade task. *Eur. J. Neurosci.* 26, 1714–1720.
- Epstein, L.H., Blanchard, E.B., 1977. Biofeedback, self-control, and self-management. *Biofeedback Self-Regul.* 2 (2), 201–211.
- Escolano, C., Aguilar, M., Minguez, J., 2011. EEG-based Upper Alpha Neurofeedback Training Improves Working Memory Performance. 33rd Annual International Conference of the IEEE EMBS. Institute of Electrical and Electronics Engineers, Boston, MA, USA, pp. 2327–2330.
- Escolano, C., Oliván, B., Lopez-del-Hoyo, Y., Garcia-Campayo, J., Minguez, J., 2012. Double-blind single-session neurofeedback training in upper-alpha for cognitive enhancement of healthy subjects. 34th conference of the IEEE EMBC. IEEE Engineering in Medicine and Biology Society, San Diego, CA, USA, pp. 4643–4647.
- Ester, E.F., Sprague, T.C., Serences, J.T., 2015. Parietal and frontal cortex encode stimulus-specific mnemonic representations during visual working memory. *Neuron* 87 (4), 893–905.
- Falkenstein, M., Hohnsbein, J., Hoormann, J., 1991. Effects of crossmodal divided attention on late ERP components. II. Error processing in choice reaction time tasks. *Electroencephalogr. Clin. Neurophysiol.* 78, 447–455.
- Frankland, P.W., Bontempi, B., 2005. The organization of recent and remote memories. *Nat. Rev. Neurosci.* 6 (2), 119–130.
- Freund, L.S., 1990. Maternal regulation of children's problem-solving behavior and its impact on children's performance. *Child Dev.* 61, 113–126.
- Frith, C.D., Blakemore, S., Wolpert, D., 2000. Abnormalities in the awareness and control of action. *Trans. R. Soc. Lond.* 355, 1771–1788.
- Gallagher, S., 2000. Philosophical conceptions of the self: implications for cognitive science. *Trends Cogn. Sci.* 4 (1), 14–21.
- Garon, N., Bryson, S.E., Smith, I.M., 2008. Executive function in preschoolers: a review using an integrative framework. *Psychol. Bull.* 31, 31–60.
- Gratchel, R.J., Korman, M., Weis, C.B., Smith, D., Clarke, L., 1978. A multiple-response evaluation of EMG biofeedback performance during training and stress-induction conditions. *Psychophysiology* 15, 253–258.
- Gazzaley, A., Nobre, A.C., 2012. Top-down modulation: bridging selective attention and working memory. *Trends Cogn. Sci.* 16 (2), 129–135.
- Gehring, W.J., Goss, B., Coles, M.G., Meyer, D.E., Donchin, E., 1993. A neural system for error detection and compensation. *Psychol. Sci.* 4, 385–390.
- Gershman, S., Markman, A., Otto, A., 2014. Retrospective reevaluation in sequential decision making: a tale of two systems. *J. Exp. Psychol. Gen.* 143 (1), 182–194.
- Gevensleben, H., Albrecht, B., Lütcke, H., Auer, T., Dewiputri, W.L., Schweizer, R., et al., 2014. Neurofeedback of slow cortical potentials: neural mechanisms and feasibility of a placebo-controlled design in healthy adults. *Front. Hum. Neurosci.* 8 (990), 1–13.
- Goldfried, M.R., Sprafkin, J.N., 1976. Behavioral personality assessment. In: Spence, J.T., Carson, R.C., Thibaut, J.W. (Eds.), *Behavioral Approaches to Therapy*. General Learning Press, Morristown (USA): NJ, pp. 295–321.
- Graafland, M., Schraagen, J., Schijven, M.P., 2012. Systematic review of serious games for medical education and surgical skills training. *Br. J. Surg.* 99 (10), 1322–1330.
- Graafland, M., Dankbaar, M., Mert, A., Lagro, J., De Wit-Zuurendonk, L., Schuit, S., et al., 2014. How to systematically assess serious games applied to health care. *JMIR Serious Games* 2 (2), e11.
- Gu, S., Pasqualetti, F., Cieslak, M., Telesford, Q.K., Yu, A.B., Kahn, A.E., et al., 2015. Controllability of structural brain networks. *Nat. Commun.* 6 (8414).
- Guenther, F.H., Brumberg, J.S., Wright, E.J., Nieto-Castanon, A., Tourville, J.A., Pankov, M., 2009. A Wireless Brain-Machine Interface for Real-Time Speech Synthesis. *PLoS One* 4 (12), e8218.
- Hammond, S.I., Müller, U., Carpendale, J.I., Bibok, M.B., Liebermann-Finestone, D.P., 2012. The effects of parental scaffolding on preschoolers' executive function. *Dev. Psychol.* 48 (1), 271–281.

CHAPTER 5. APPENDIX: PUBLISHED WORKS

A. Gaume et al. / Neuroscience and Biobehavioral Reviews 68 (2016) 891–910

909

- Hart, S.G., Staveland, L.E., 1988. Development of NASA-TLX (task load index): results of empirical and theoretical research. *Adv. Psychol.* 52, 139–183.
- Hasbroucq, T., Burle, B., Vidal, F., Possamai, C.-A., 2009. Stimulus-hand correspondence and direct response activation: an electromyographic analysis. *Psychophysiology*, 1160–1169.
- Hauri, P.P., 1975. Biofeedback and self-control of physiological functions: clinical applications. *Int. J. Psychiatry Med.* 6 (1–2), 255–265.
- Häusser, T.U., Iannaccone, R., Siampli, P., Drechsler, R., Brandeis, D., Walitza, S., et al., 2014. The feedback-related negativity (FRN) revisited: new insights into the localization: meaning and network organization. *Neuroimage* 84, 159–168.
- Hoppe, D., Sadakata, M., Desain, P., 2006. Development of a real-time visual feedback assistance in singing training: a review. *J. Comp. Assist. Learn.* 22, 308–316.
- Huyck, C.R., Passmore, P.J., 2013. A review of cell assemblies. *Biol. Cybern.* 107 (3), 263–288.
- Ikka, A., Curtis, C.E., 2011a. Common neural mechanisms supporting spatial working. *Neuropsychologia* 49 (6), 1428–1434.
- Ikka, A., Curtis, C.E., 2011b. Common neural mechanisms supporting spatial working memory, attention and motor intention. *Neuropsychologia* 49 (6), 1428–1434.
- Johnson, S.H., Grafton, S.T., 2003. From 'acting on' to 'acting with': the functional anatomy of object-oriented action schemata. In: Fraibanc, C., Pellissier, D., Rossetti, Y. (Eds.), *Neural Control of Space Coding and Action Production*. Progress in Brain Research, vol. 142. Elsevier Science B.V., pp. 127–139.
- Kalman, R.E., 1960. On the General Theory of Control Systems. Proc. First IFAC Congress Automatic Control, 1. Butterworths, London, Moscow, pp. 481–492.
- Kang, S.Y., Im, C.-H., Shim, M., Nahab, F.B., Park, J., Kim, D.-W., et al., 2013. Brain networks responsible for sense of agency: an EEG study. *PLoS One* 10 (8).
- Kehagia, A.A., Murray, G.K., Robbins, T.W., 2010. Learning and cognitive flexibility: frontostriatal function and monoaminergic modulation. *Curr. Opin. Neurobiol.* 20 (2), 199–204.
- Keizer, A.W., Verschoor, M., Verment, R.S., Hommel, B., 2010. The effect of gamma enhancing neurofeedback on the control of feature bindings and intelligence measures. *Int. J. Psychophysiol.* 75 (1), 25–32.
- Khazan, I.Z., 2013. The Clinical Handbook of Biofeedback: A Step-by-Step Guide for Training and Practice with Mindfulness. John Wiley & Sons, Ltd., Chichester, UK.
- Knudsen, E.I., 2007. Fundamental components of attention. *Annu. Rev. Neurosci.* 30, 57–78.
- Kober, S.E., Witte, M., Stangl, M., Väjjäme, A., Neuper, C., Wood, G., 2015. Shutting down sensorimotor interference unblocks the networks for stimulus processing: an fMRI neurofeedback training study. *Clin. Neurophysiol.* 126 (1), 82–95.
- Koechlin, E., 2014. An evolutionary computational theory of prefrontal executive function in decision-making. *Phil. Trans. R. Soc. B* 369, 20130474.
- Koechlin, E., 2016. Prefrontal executive function and adaptive behavior in complex environments. *Curr. Opin. Neurobiol.* 37, 1–6.
- Koechlin, E., Ody, C., Kouneither, F., 2003. The Architecture of Cognitive Control in the Human. *Science* 302 (5648), 1181–1185.
- Koralek, A.C., Jin, X., Long, J.D., Costa, R.M., Carmenta, J., 2012. Corticostriatal plasticity necessary for learning intentional neuroprosthetic skills. *Nature* 483, 331–335.
- Kuikkaniemi, K., Laitinen, T., Turpeinen, M., Saari, T., Kosunen, I., Ravaja, N., 2010. The influence of implicit and explicit biofeedback in first-person shooter games. CHI '10 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Brains and Brawn. ACM, Atlanta, GA, USA, pp. 859–868.
- Lafargue, G., Franck, N., 2009. Effort awareness and sense of volition in schizophrenia. *Conscious. Cogn.* 18 (1), 277–289.
- Lamar, M., Raz, A., et al., 2007. Neuropsychological assessment of attention and executive functioning. In: Ayers, S., Baum, A., McManus, C., Newman, S., Wallston, K., Weinman, J. (Eds.), *Cambridge Handbook of Psychology, Health, and Medicine*, second edition. Cambridge University Press, New York, pp. 290–294.
- Lazarov, A.D., Liberman, N., Oded, Y., 2010. Obsessive/compulsive tendencies may be associated with attenuated access to internal states: Evidence from a biofeedback-aided muscle tensing task. *Conscious. Cognit.* 21 (3), 1401–1409.
- Lawrence, E.J., Sua, L., Barker, C.J., Medford, N., Dalton, J., Williams, S.C., et al., 2014. Self-regulation of the anterior insula: reinforcement learning using real-time fMRI neurofeedback. *Neuroimage* 88, 113–124.
- Leher, P.M., Gewirtz, R., 2014. Heart rate variability biofeedback: how and why does it work. *Front. Psychol.* 5 (756).
- Leigh, H., 1978. Self-control, biofeedback, and change in 'psychosomatic' approach. *Psychother. Psychosom.* 30 (2), 130–136.
- Lepper, M.P., Greene, D., Nisbett, R.E., 1973. Undermining children's intrinsic interest with extrinsic reward: a test of the overjustification hypothesis. *J. Pers. Soc. Psychol.* 28 (1), 129–137.
- Levine, W.S., 2010. *The Control Handbook: Control System Fundamentals*, Second Edition. Taylor & Francis, Boca Raton, USA.
- Lewis, P.A., Durrant, S., 2011. Overlapping memory replay during sleep builds cognitive schemata. *Trends Cogn. Sci.* 15 (8), 343–351.
- Lezak, M.D., Howieson, D.B., Bigler, E.D., Tranel, D., 2012. *Neuropsychological Assessment*, fifth edition. Oxford University Press, New York.
- Logan, D.M., Hill, K.R., Larson, M.J., 2015. Cognitive control of conscious error awareness: error awareness and error positivity (Pe) amplitude in moderate-to-severe traumatic brain injury (TBI). *Front. Hum. Neurosci.* 9 (397), 1–12.
- Logue, S.F., Gould, T.J., 2014. The neural and genetic basis of executive function: attention, cognitive flexibility, and response inhibition. *Pharmacol. Biochem. Behav.* 123, 45–54.
- Lopez-Gamal, F., Middleton, R.H., Huber, H., 2014. Equilibria and stability of a class of positive feedback loops. *J. Math. Biol.* 68, 609–645.
- Lorenz, R.C., Gleich, T., Gallinat, J., Kühn, S., 2015. Video game training and the reward system. *Front. Hum. Neurosci.* 9 (40), 1–9.
- Lotte, F., Larrue, F., Mühl, C., 2013. Flaws in current human training protocols for spontaneous Brain-Computer Interfaces: lessons learned from instructional design. *Front. Hum. Neurosci.* 7 (568), 1–11.
- Maheu, M.M., Pulier, M.L., Wilhelm, F.H., McMenamin, J.P., Brown-Connelly, N.E., 2004. *The Mental Health Professional and the New Technologies: A Handbook for Practice Today*. Lawrence Erlbaum Associates, Inc., Mahwah, New Jersey, USA.
- Maine de Biran, F.P., 1805. *Mémoire sur la décomposition de la pensée (Tome III)*. Vrin (1963), Paris.
- McGonigal, J., 2011. *Reality Is Broken: Why Games Make Us Better and How They Can Change the World*. Penguin press, New York.
- Menon, V., Uddin, L.Q., 2010. Saliency, switching, attention and control: a network model of insula function. *Brain Struct. Funct.* 214 (5–6), 655–667.
- Michael, D., Chen, S., 2006. *Serious Games: Games That Educate, Train and Inform*. Thomson, Boston, MA.
- Micoulaud-Franchi, J.A., McGonigal, A., Lopez, R., Daudet, C., Kotwas, I., Bartolomei, F., 2015. Electroencephalographic neurofeedback: level of evidence in mental and brain disorders and suggestions for good clinical practice. *Neurophysiol. Clin.* 45 (6), 423–433.
- Middaugh, S.J., Haythornthwaite, J.A., Thompson, B., Hill, R., Brown, K.M., Freedman, R.R., et al., 2001. The Raynaud's Treatment Study: biofeedback protocols and acquisition of temperature biofeedback skills. *Appl. Psychophysiol. Biofeedback* 26 (4), 251–278.
- Miller, E., Cohen, J., 2001. An integrative theory of prefrontal cortex function. *Annu. Rev. Neurosci.* 24, 167–202.
- Miltner, W.H., Braun, C.H., Coles, M.G., 1997. Event-related brain potentials following incorrect feedback in a time-estimation task: evidence for a generic neural system for error detection. *J. Cogn. Neurosci.* 9, 788–798.
- Mindell, D.A., 2002. *Between Human and Machine: Feedback, Control, and Computing Before Cybernetics*. JHU Press, Baltimore, Maryland, USA.
- Montgomery, G.T., 1988. Variations in Digital Temperature During Frontal EMG Biofeedback Training in Normal Subjects. *Biofeedback and Self-Regulation* 13 (2), 151–168.
- Moss, D., Gunzelman, J., 2002. Task force report on methodology and empirically supported treatments: introduction and summary. *Biofeedback* 30 (2), 19–20.
- Mottaghy, F.M., 2006. Interfering with working memory in humans. *Neuroscience* 139, 85–90.
- Murrill, P.W., 2000. *Fundamentals of Process Control Theory*, 3rd Edition. ISA: The Instrumentation, Systems, and Automation Society, Research triangle park, North Carolina, USA.
- Nacke, L.E., Grimshaw, M.N., Lindley, C.A., 2010. More than a feeling: Measurement of sonic user experience and psychophysiology in a first-person shooter game. *Interact. Comput.* 22 (5), 336–343.
- Nacke, L.E., Kalyn, M., Lough, C., Mandryk, R.L., 2011. Biofeedback Game Design: Using Direct and Indirect Physiological Control to Enhance Game Interaction. Proceedings of the CHI 2011 Conference on Human Factors in Computing Systems. ACM, Vancouver, BC, Canada, pp. 103–112.
- Nelson-Gray, R.O., Farmer, R.F., 1999. Behavioral assessment of personality disorders. *Behav. Res. Ther.* 37, 347–368.
- Niendam, T.A., Laird, A.R., Ray, K.L., Monica Dean, Y., Glahn, D.C., Carter, C.S., 2012. Meta-analytic evidence for a superordinate cognitive control network subserving diverse executive functions. *Cogn. Affect. Behav. Neurosci.* 12, 241–268.
- Nieuwenhuis, S., Ridderinkhof, K.R., Blom, J., Band, G.P., Kok, A., 2001. Error-related brain potentials are differentially related to awareness of response errors: evidence from an antisaccade task. *Psychophysiology* 38, 752–760.
- Ninaus, M., Kober, S.E., Witte, M., Koschnig, K., Stangl, M., Neuper, C., et al., 2013. Neural substrates of cognitive control under the belief of getting neurofeedback training. *Front. Hum. Neurosci.* 7 (914), 1–10.
- Norris, P., 1986. Biofeedback, voluntary control, and human potential. *Biofeedback Self Regul.* 11 (1), 1–20.
- O'Connell, R.G., Dockree, P.M., Bellgrove, M.A., Kelly, S.P., Hester, R., Garavan, H., et al., 2007. The role of cingulate cortex in the detection of errors with and without awareness: a high-density electrical mapping study. *Eur. J. Neurosci.* 25 (8), 2571–2579.
- Olson, C.K., 2010. Children's motivations for video game play in the context of normal development. *Rev. Gen. Psychol.* 14 (2), 180–187.
- Oppenheimer, D.M., 2008. The secret life of fluency. *Trends Cogn. Sci.* 12 (6), 237–241.
- Overbeek, T.J., Nieuwenhuis, S., Ridderinkhof, K.R., 2005. Dissociable components of error processing: on the functional significance of the Pe vis-à-vis the ERN/Ne. *J. Psychophysiol.* 19, 319–329.
- Parkin, A.J., 1998. The central executive does not exist. *J. Int. Neuropsychol. Soc.* 4 (5), 518–522.
- Parnandi, A., Son, Y., Gutierrez-Osuna, R., 2013. A Control-Theoretic Approach to Adaptive Physiological Games. Proceedings of the 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction. Association for the Advancement of Affective Computing, Geneva, Switzerland, pp. 7–12.

- Pascual-Leone, J., Goodman, D., 1979. Intelligence and experience: a neopiagetian approach. *Instr. Sci.* 8, 301–367.
- Pauls, F., Macha, T., Petermann, F., 2013. U-shaped development: an old but unsolved problem. *Front. Psychol.* 4 (301), 1–6.
- Penzlin, A.I., Siepmann, T., Illigens, B.M., Weidner, K., Siepmann, M., 2015. Heart rate variability biofeedback in patients with alcohol dependence: a randomized controlled study. *Neuropsychiatr. Dis. Treat.* 11, 2619–2627.
- Phillips, M.L., Ladouceur, C.D., Drevets, W.C., 2008. A neural model of voluntary and automatic emotion regulation: implications for understanding the pathophysiology and neurodevelopment of bipolar disorder. *Mol. Psychiatry* 13 (9), 829–857.
- Piaget, J., 1971. *Biology and Knowledge*. University of Chicago Press, Chicago.
- Plant, K.L., Stanton, N.A., 2013. The explanatory power of Schema Theory: theoretical foundations and future applications in Ergonomics. *Ergonomics* 56 (1), 1–15.
- Polito, V., Barnier, A.J., Woody, E.Z., 2013. Developing the Sense of Agency Rating Scale (SOARS): an empirical measure of agency disruption in hypnosis. *Conscious. Cogn.* 22 (3), 684–696.
- Poole, J.L., 1991. Application of motor learning principles in occupational therapy. *Am. J. Occup. Ther.* 45, 531–537.
- Premsky, M., 2001. *Digital Game-Based Learning*. McGraw-Hill, New York.
- Prinzl, L., Pope, A., Freeman, F., 2001. Application of physiological Self-regulation and Adaptive Task Allocation Techniques for Controlling Operator Hazardous States of Awareness. NASA/TM-2001-211015, L-18075, NAS 1.15:211015. NASA Langley Research Center, Hampton, VA.
- Przybylski, A.K., Rigby, C.S., Ryan, R.M., 2010. A motivational model of video game engagement. *Rev. Gen. Psychol.* 14 (2), 154–166.
- Reichert, J.L., Kober, S.E., Neuper, C., Wood, G., 2015. Resting-state sensorimotor rhythm (SMR) power predicts the ability to up-regulate SMR in an EEG-instrumental conditioning paradigm. *Clin. Neurophysiol.* 126 (11), 2068–2077.
- Rochet, N., Spieser, L., Casini, L., Hasbroucq, T., Burle, B., 2014. Detecting and correcting partial errors: evidence for efficient control without conscious access. *Cogn. Affect. Behav. Neurosci.* 14 (3), 970–982.
- Rose, N., 2009. Normality and pathology in a biomedical age. *Sociolog. Rev.* 57 (s2), 66–83.
- Ross Ashby, W., 1956. *An Introduction to Cybernetics*. Chapman & Hall, London, UK.
- Ros, T., Baars, B.J., Lanus, R.A., Vuilleumier, P., 2014. Tuning pathological brain oscillations with neurofeedback: a systems neuroscience framework. *Front. Hum. Neurosci.* 8 (1008), 1–22.
- Rutter, J.B., 1966. Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs: General and Applied* 80, 1–28.
- Rumelhart, D.E., Smolensky, P., McClelland, J.L., Hinton, G.E., 1986. Schematic and sequential thought processes in PDP models. In: McClelland, J.L., Rumelhart, D.E. (Eds.), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. Volume 2: Psychological and Biological Models. MIT Press, Cambridge, MA, pp. 7–57.
- Routh, E.J., Fuller, A.T., 1975. *Stability of Motion*. Taylor & Francis, London, UK.
- Ryff, C.D., Keyes, C.L., 1995. The structure of psychological well-being revisited. *J. Pers. Soc. Psychol.* 69 (4), 719–727.
- Salmoni, A.W., Schmidt, R.A., Walter, C.B., 1984. Knowledge of results and motor learning: a review and critical reappraisal. *Psychol. Bull.* 95 (3), 355–386.
- Sanders, D., Welk, D.S., 2005. Strategies to scaffold student learning: applying Vygotsky's Zone of Proximal Development. *Nurse Educ.* 30 (5), 203–207.
- Schafer, R.J., Moore, T., 2011. Selective attention from voluntary control of neurons in prefrontal cortex. *Science* 332, 1568–1571.
- Schmidt, R.A., Wrisberg, C.A., 2007. *Motor Learning and Performance*, 4th edition. Human Kinetics, Champaign, IL, USA.
- Schwartz, N.M., Schwartz, M.S., 2003. Definitions of biofeedback and applied psychophysiology. In: Schwartz, M.S., Andrasik, F. (Eds.), *Biofeedback*. The Guilford Press, New York, pp. 27–42.
- Scott, S.H., 2004. Optimal feedback control and the neural basis of volitional motor control. *Nat. Rev. Neurosci.* 5, 532–546.
- Seidler, R.D., Bo, J., Anguera, J.A., 2012. Neurocognitive contributions to motor skill learning: the role of working memory. *J. Mot. Behav.* 44 (6), 445–453.
- Shirvalkar, P.R., 2009. Hippocampal neural assemblies and conscious remembering. *J. Neurophysiol.* 101 (5), 2197–2200.
- Sherlin, L.H., Arns, M., Lubar, J., Heinrich, H., Kerson, C., Strehl, U., 2011. Neurofeedback and basic learning theory: implications for research and practice. *J. Neurother.* 15, 292–304.
- Siegler, R.S., 2004. U-shaped interest in U-shaped development and what it means. *J. Cogn. Dev.* 5 (1), 1–10.
- Skinner, B.F., 1938. *The Behavior of Organisms: An Experimental Analysis*. Appleton-Century-Crofts, New York.
- Smith, E.E., Jonides, J., 1999. Storage and executive processes in the frontal lobes. *Science* 283, 1647–1661.
- Sperduti, M., Delaveau, P., Fossati, P., Nadel, J., 2011. Different brain structures related to self- and external-agency attribution: a brief review and meta-analysis. *Brain Struct. Funct.* 216, 151–157.
- Sreenivasan, K.K., Curtis, C.E., D'Esposito, M., 2014. Revisiting the role of persistent neural activity during working memory. *Trends Cogn. Sci.* 18 (2), 82–89.
- Sterman, M.B., Egner, T., 2006. Foundation and practice of neurofeedback for the treatment of epilepsy. *Appl. Psychophysiol. Biofeedback* 31 (1), 21–35.
- Stokes, M.G., 2015. 'Activity-silent' working memory in prefrontal cortex: a dynamic coding framework. *Trends Cogn. Sci.* 19 (7), 394–405.
- Strack, B.W., 2011. *Biofeedback & Neurofeedback Applications in Sport Psychology*. Association for Applied Psychophysiology and Biofeedback, Wheat Ridge, CO, USA.
- Strehl, U., 2014. What learning theories can teach us in designing neurofeedback treatments. *Front. Hum. Neurosci.* 8 (894), 1–8.
- Sterling, P., 2004. Principles of allostasis: Optimal design, predictive regulation, pathophysiology, and rational therapeutics. In: Schulkin, J. (Ed.), *Allostasis, homeostasis, and the costs of physiological adaptation*. Cambridge University Press, New York, NY, pp. 17–64.
- Swanson, L.R., Whittinghill, D.M., 2015. Intrinsic or extrinsic? Using videogames to motivate stroke survivors: a systematic review. *Games Health J.* 4 (3), 253–258.
- Sweller, J., 2010. Element interactivity and intrinsic, extraneous: and germane cognitive load. *Educ. Psychol. Rev.* 22, 123–138.
- Szynofzik, M., Thier, P., Lindner, A., 2006. Internalizing agency of self-action: perception of one's own hand movements depends on an adaptable prediction about the sensory action outcome. *J. Neurophysiol.* 96 (3), 1592–1601.
- Towse, J.N., Hitch, G.J., Horton, N., 2007. Working memory as the interface between processing and retention: a developmental perspective. *Adv. Child Dev. Behav.* 219–251.
- Trentelman, H., Stoorvogel, A.A., Hautus, M., 2001. *Control Theory for Linear Systems*. Springer-Verlag, London, UK.
- Tse, D., Langston, R., Kakeyama, M., Bethus, I., Spooner, P., Wood, E., 2007. Schemas and memory consolidation. *Science* 316 (5821), 76–82.
- Valsiner, J., 2012. *The Oxford Handbook of Culture and Psychology*. Oxford University Press, New York, NY.
- Varela, F., Lachaux, J.P., Rodriguez, E., Martinerie, J., 2001. The brainweb: phase synchronization and large-scale integration. *Nat. Rev. Neurosci.* 2, 229–239.
- Vidal, F., Meckler, C., Hasbroucq, T., 2015. Basics for sensorimotor information processing: some implications for learning. *Front. Psychol.* 6 (33).
- Vincent, J., Kahn, I., Snyder, A., Raichle, M., Buckner, R., 2008. Evidence for a frontoparietal control system revealed by intrinsic functional connectivity. *J. Neurophysiol.* 100 (6), 3328–3342.
- Vinney, L.A., Turkstra, L.S., 2013. The role of self-regulation in voice therapy. *J. Voice* 27 (3), 390, e1–e11.
- Wallace, A.F., 1960. Plans and the structure of behavior. George A. Miller, Eugene Galanter, and Karl H. Pribram. Book review. *Am. Anthropol.* 62 (6), 1065–1067.
- Walsh, M.M., Anderson, J.R., 2012. Learning from experience: event-related potential correlates of reward processing, neural adaptation, and behavioral choice. *Neurosci. Biobehav. Rev.* 36, 1870–1884.
- Weeks, K.W., Higginson, R., Clochesy, J.M., Coben, D., 2013. Safety in numbers 7: Veni, vidi, duci: a grounded theory evaluation of nursing students' medication dosage calculation problem-solving schemata construction. *Nurse Educ. Pract.* 13 (2), e78–87.
- Wiener, N., 1948. *Cybernetics: Or Control and Communication in the Animal and the Machine*. The MIT Press, Cambridge, Massachusetts, 2nd revised ed. 1961.
- Wilson, P.H., Thorpe, C.W., Callaghan, J., 2005. August 4–6. Looking at singing: does real-time visual feedback improve the way we learn to sing. Second APSCOM Conference: Asia-Pacific Society for the Cognitive Sciences of Music, South Korea, Seoul.
- Wilts, C.H., 1960. *Principles of Feedback Control*. Addison-Wesley Educational Publishers Inc., Reading, Massachusetts, USA.
- Winstein, C.J., Schmidt, R.A., 1990. Reduced frequency of knowledge of results enhances motor skill learning. *J. Exp. Psychol. Learn. Mem. Cogn.* 16, 677–691.
- Witte, M., Kober, S., Ninaus, M., Neuper, C., Wood, G., 2013. Control beliefs can predict the ability to up-regulate sensorimotor rhythm during neurofeedback training. *Front. Hum. Neurosci.* 7, 1–8.
- Wood, D., Bruner, J., Ross, G., 1976. The role of tutoring in problem solving. *J. Child Psychol. Child Psychiatry* 17, 89–100.
- Wood, G., Kober, S.E., Witte, M., Neuper, C., 2014. On the need to better specify the concept of control in brain-computer-interfaces/neurofeedback research. *Front. Syst. Neurosci.* 8 (171), 1–4.
- Wulf, G., Shea, C., Lewthwaite, R., 2010. Motor skill learning and performance: a review of influential factors. *Med. Educ.* 44 (1), 75–84.
- Yager, L.M., Garcia, A.F., Wunsch, A.M., Ferguson, S.M., 2015. The ins and outs of the striatum: role in drug addiction. *Neuroscience* 301, 529–541.
- Yeung, N., Sanfey, A.G., 2004. Independent coding of reward magnitude and valence in the human brain. *J. Neurosci.* 24 (28), 6258–6264.
- Yoshida, K., Asakawa, K., Yamauchi, T., Sakuraba, S., Sawamura, D., Murakami, Y., et al., 2013. The flow state scale for occupational tasks: development, reliability, and validity. *Hong Kong J. Occup. Ther.* 23 (2), 54–61.
- Yucha, C.B., Montgomery, D., 2008. *Evidence-based Practice in Biofeedback and Neurofeedback*. Association for Applied Psychophysiology and Biofeedback, Wheat Ridge, CO.
- van Kesteren, M.T., Fernández, G., Norris, D.G., Hermans, E.J., 2010. Persistent schema-dependent hippocampal-neocortical connectivity during memory encoding and postencoding rest in humans. *Proc. Natl. Acad. Sci. U. S. A.* 107, 7550–7555.
- van Merriënboer, J.J., Sweller, J., 2010. Cognitive load theory in health professional education: design principles and strategies. *Med. Educ.* 44, 85–93.
- Zoefel, B., Huster, R., Herrmann, C., 2011. Neurofeedback training of the upper alpha. *NeuroImage* 2, 1427–1431.
- Zolten, A.J., 1989. *Constructive Integration of Learning Theory and Phenomenological Approaches to Biofeedback Training*. *Biofeedback and Self-Regulation* 14 (2), 89–99.

2015 IEEE INTERNATIONAL WORKSHOP ON MACHINE LEARNING FOR SIGNAL PROCESSING, SEPT. 17–20, 2015, BOSTON, USA

A COGNITIVE BRAIN-COMPUTER INTERFACE PROTOTYPE FOR THE CONTINUOUS MONITORING OF VISUAL WORKING MEMORY LOAD.

Aldo Mora Sánchez^{1,2}, Antoine Gaume^{1,2}, Gérard Dreyfus², François-Benoît Vialatte^{1,2}

1 - Brain plasticity Unit. CNRS, UMR8249, Paris, 75005, France
2 - ESPCI-ParisTech, PSL Research University, Paris, 75005, France

ABSTRACT

We investigate the neural correlates of visual working memory using electroencephalography (EEG). Our objective is to develop a cognitive Brain-Computer Interface (BCI) able to monitor visual working memory load in real-time. A system with these properties would eventually have different applications, such as training, rehabilitation, or safety while operating dangerous machinery. The BCI performances were evaluated using cross-validation. With an appropriately chosen classification threshold, it is possible to detect high working memory load with a sensitivity of 68% and a specificity of 72%. However, it is well known that some subjects are BCI *illiterate*, meaning that up to 30% of the users have too high signal variability to use EEG-based BCI systems. If we analyse each subject individually, it is possible to detect high working memory load with a sensitivity of 78% and a specificity of 81% (accuracy = 81%) for a typical good subject. Changes due to visual working memory load were observed in frontal, parietal, and occipital regions.

Index Terms— Brain-computer interfaces, cognitive information processing, pattern recognition, classification.

1. INTRODUCTION

Humans can interact directly with machines using their brain activity. This is, in general terms, a Brain-Computer Interface (BCI) [1]. Such devices take brain activity as an input, and transform it into an output via a translation algorithm, bypassing the motor system. The outcome is often a command, for instance, choosing a letter or a certain movement. Electroencephalography (EEG) is the most frequently used technique due to its non-invasive nature and low acquisition and operational costs. Furthermore, EEG is also a signal of interest because electrical brain signals provide access to neural dynamics with a very high temporal resolution.

In healthy subjects, regular speech is an efficient channel of communication. Besides, the motor system provides a fine-tuned means of control with several degrees of freedom. Both the motor system and speech demand relatively

low cognitive effort as compared to BCI. Therefore, BCIs were initially thought for patients whose conditions prevent them from adequately using those systems. For instance, locked-in patients [2]. However, BCI systems can go beyond communication, for instance, they can be useful for cognitive monitoring as suggested by [3].

Zander and Kothe [4] propose a re-thinking of BCI usage. They suggest a new classification according to the BCI functional mechanisms, and outline usability according to the target (healthy or disabled users). BCI are classified into active, reactive and passive. This paper deals with passive interfaces.

1.1. Passive BCI

The current state of BCI performance can hardly compete with the above mentioned mechanisms available to healthy users. Therefore, it is proposed to combine BCI technology with cognitive monitoring in a new approach: passive BCI. Cognitive monitoring refers to the analysis of brain signals in order to infer information about the cognitive state of the user. The idea of a passive BCI is to feed this information to a system to improve its performance, in a way which is non-voluntarily driven by the user. Contextual information about the cognitive state could be critical in safety-related tasks such as driving or in industrial environments. Cognitive load or attention, for instance, have a high impact on the performance of these activities. By including this information, certain commands can be triggered depending on the cognitive state, allowing the system to adapt to the users, without them intentionally triggering the actions.

1.2. Working Memory Load

Working memory is regarded as a system that keeps information (with storage and time limitations) while it is being manipulated. It works as an interface between perception, long-term memory and action [5].

There are precedents for the assessing of working memory load using EEG spectral features, particularly, in prefrontal and parietal regions of the brain. Furthermore, after training working memory it is possible to observe and measure certain changes [6].

2. MATERIALS AND METHODS

2.1. The task

According to Cowan's model [7], working memory acts as an activation buffer for items previously stored in the long term memory. This accounts for the remarkable consistency of the number of items that can be stored at a given time, regardless the size of the items themselves. The number of items is then limited, however, the human brain is extraordinarily good at finding associations between items in working memory, in order to reduce the number of occupied items (mnemonics, for instance). This was a key point while designing the task, that can be described as follows:

- Subjects sit in front of a computer screen and are presented with a collection of figures that will be used during the experiment. They are asked to give a short name for them. There are different sets of figures, and each set corresponds to a semantic field: animals, transport, etc.
- The *target*, a specific sequence of figures, containing either a small number (2) or a large number (determined by the calibration, typically 5) of elements is presented to the subject who is asked to keep it in memory. An example of target corresponding to low visual working memory load would be, for instance, a train followed by a bicycle.
- A random sequence of figures generated from the total set slides from right to left. The subjects simply have to press a button whenever they find the target. This is considered one trial. Trials last on average 25 seconds.
- Subjective feedback is collected on every successful trial: reported cognitive state (bored, stressed or challenged) and reported number of figures actually remembered.

The distance between the subjects and the screen was 60 centimeters, the screen model was ProLite E2208HDD. Lighting conditions were normal room conditions. The size of the figures was 100x100 pixels.

The fact that subjects were asked to verbalize the name of the figures induces in them a homogeneous storage-retrieval technique: simply to repeat the names of the sequence of

stored items and to compare it with the observed sliding items. Subjects were asked, after the experiment, which technique they used. They reported to have used precisely this method most of the time. This has several advantages. Firstly, all the items are kept as separate items. Secondly, visual comparisons guarantee that we are measuring visual working memory. Furthermore, this comparison is made in a sufficiently continuous manner. Finally, the fact that a given group of objects belongs to an evident semantic field, reduces the likelihood of item merging. In fact, the set of figures is changed several times during the experiment to prevent subjects from developing compression strategies. Or, at least, to reduce them.

As it is unlikely that patterns similar to the target, or distractors, appear by chance, parameters are tuned so that exactly fifty percent of the time a distractor appears. The (low) probability of a distractor appearing by chance was also taken into account. Distractors are important to prevent subjects from storing only a smaller amount of items (for instance, first and last, or first few). Distractors appear half of the time so that subjects do not learn that it is more likely to find, for instance, first the distractor and then the pattern, or vice-versa.

The subjective data was not included in the analysis, however, it was used to verify that the task was indeed inducing a change in the perceived cognitive state.

Finally, the small window size prevents eye movements, that are known to produce eye-related artifacts. The experiment was written in Matlab® 2015a using Psychophysics Toolbox extensions [6].

2.2. Data acquisition and pre-processing

Brain activity was recorded using a 16 channel EEG device (Brain Products V-Amp) at a sampling rate of 2000 Hz. The electrode set-up is shown in figure 1. 12 healthy subjects of age between 21 and 31 were recorded, 6 males and 6 females, with normal or corrected to normal vision, and absence of any brain disorder or drug consumption. The study followed the principles outlined in the Declaration of Helsinki. All participants were given explanations about the nature of the experiment and signed an informed consent form before the experiment started.

Data was filtered in the range of 1 to 90 Hz with a 3rd order butterworth filter. A notch filter at 50 Hz was used to remove the line noise as well.

In order to obtain a clean marker that can be further used even in noisy, real-life conditions, special attention was paid

that those features were actually cross terms, i.e., the product of two features. All the following estimations were obtained by setting the number of features to 6. Figure 3 shows that the mean error rate obtained using cross-validation was 32%. However, due to the problem of BCI literacy, it is more interesting to analyse this error on a subject basis rather than globally. Nine subjects out of the twelve, had both an error rate lower than that of a random classifier, and a ROC curve with area bigger than 0.5. Those two values are evidently correlated, as we can see in figure 4. Subjects plotted in the top left square were considered as *good subjects*. Figure 5 shows the ROC curve of the classifier, together with the ROC curve of a typical good subject. For the whole set of subjects, we can expect to detect high working memory load 68% of the time with 72% of accuracy. In particular, for a single good subject, we can observe that we can detect high working memory load 77% of the time with 81% of specificity.

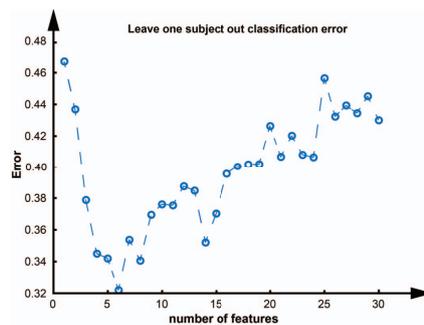


Fig. 3. Error rate as a function of the number of spectral features

Feature selection was performed on a subject basis, because GSO algorithms are fast and it can be implemented on-line. However, for a descriptive characterization of the underlying process, we performed *bagging* with the selected features of each individual, in order to choose the most representative ones. For each subject a set of features obtained by GSO was formed. The size of the set being the number of features that minimized the classification error for that individual. All these sets were merged into one, and features were ranked according to the number of times they appeared. As a result, most of the chosen features were present in most of the subjects. The final features were the following:

- relative theta power, electrode Fz
- relative upper gamma power, electrode Pz

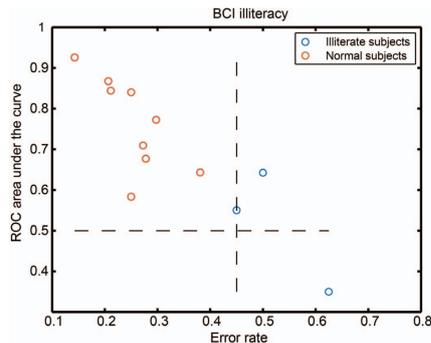


Fig. 4. Usability of the BCI prototype. Three subjects over twelve (25%) were considered illiterate.

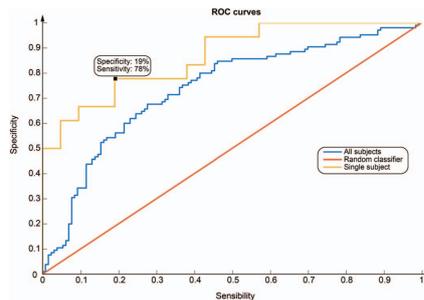


Fig. 5. ROC curves for the whole set of subjects, and a typical good subject

- upper beta power, electrode Fz
- relative lower beta power, electrode F4
- relative delta power, electrode Fp1
- relative alpha power, electrode O1
- lower gamma power, electrode Fz
- relative lower gamma power, electrode Fz
- relative upper beta power, electrode Pz

There are 9 features instead of 12 because 3 of the single features were in more than one cross term. It is important to

note that although both, linear features and cross terms were included in the GSO procedure, all the selected features for all the subjects were cross terms. Single features then did not convey as much information as cross-terms. Figure 6 shows that for most of the selected features, classification power was rather poor when they were used alone.

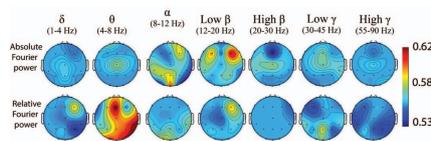


Fig. 6. Accuracy of classification for each single feature

4. DISCUSSION

By looking at the ROC curves, we can observe that the classifier consistently performs much better than a random classifier across the whole range of possible thresholds, and that this improves even more when we consider the classification on a subject basis. Furthermore, the threshold value can be adjusted depending on the expected applications. For instance, in a learning environment, we can afford skipping some true positives (high memory load), as long as we can be certain about the detected positives.

In [8], authors conclude that the last 30 years of BCI research have shown that between 20% and 30% of the users cannot use BCI systems with enough accuracy to achieve control. However, that paper addresses BCI for communication, not for monitoring as it is the case of this work. Hence, their claims, although useful for establishing a comparison, should not be blindly generalized. In our case, even though we kept illiterate subjects in our database, for a typical non-illiterate subject, we can observe that, if we are willing to accept only a 19% false positive rate, we can still accurately detect high memory load approximately 78% of the time.

The main features found in this study are consistent with literature reports of working memory. In [9], it is discussed that changes due to working memory load are often observed in the theta power of midline electrodes, and alpha power in occipital electrodes.

5. REFERENCES

- [1] Jonathan R Wolpaw, Niels Birbaumer, William J Heetderks, Dennis J McFarland, P Hunter Peckham, Gerwin Schalk, Emanuel Donchin, Louis A Quatrano, Charles J Robinson, Theresa M Vaughan, et al., "Brain-computer interface technology: a review of the first international

meeting," *IEEE transactions on rehabilitation engineering*, vol. 8, no. 2, pp. 164–173, 2000.

- [2] Thilo Hinterberger, Andrea Kübler, Jochen Kaiser, Nicola Neumann, and Niels Birbaumer, "A brain-computer interface (bci) for the locked-in: comparison of different eeg classifications for the thought translation device," *Clinical Neurophysiology*, vol. 114, no. 3, pp. 416–425, 2003.
- [3] Klaus-Robert Müller, Michael Tangermann, Guido Dornhege, Matthias Krauledat, Gabriel Curio, and Benjamin Blankertz, "Machine learning for real-time single-trial eeg-analysis: from brain-computer interfacing to mental state monitoring," *Journal of neuroscience methods*, vol. 167, no. 1, pp. 82–90, 2008.
- [4] Thorsten O Zander and Christian Kothe, "Towards passive brain-computer interfaces: applying brain-computer interface technology to human-machine systems in general," *Journal of Neural Engineering*, vol. 8, no. 2, pp. 025005, 2011.
- [5] Alan Baddeley, "Working memory," *Science*, vol. 255, no. 5044, pp. 556–559, 1992.
- [6] Pernille J Olesen, Helena Westerberg, and Torkel Klingberg, "Increased prefrontal and parietal activity after training of working memory," *Nature neuroscience*, vol. 7, no. 1, pp. 75–79, 2004.
- [7] N. Cowan, *Attention and memory: an integrated framework*. Oxford University Press, Oxford, Oxfordshire, 1995.
- [8] Benjamin Blankertz, Claudia Sanelli, Sebastian Halder, Eva-Maria Hammer, Andrea Kübler, Klaus-Robert Müller, Gabriel Curio, and Thorsten Dickhaus, "Predicting bci performance to study bci illiteracy," 2009.
- [9] Lars Michels, Kerstin Bucher, Rafael Lüchinger, Peter Klaver, Ernst Martin, Daniel Jeanmonod, and Daniel Brandeis, "Simultaneous eeg-fmri during a working memory task: modulations in low and high frequency bands," *PLoS One*, vol. 5, no. 4, pp. e10298, 2010.

Is the lack of EEG stationarity useful? The dynamics of metastable brain states under cognition

Aldo Mora-Sánchez^{1,2} and François-Benoît Vialatte^{1,2}

Abstract—We developed a technique showing that non stationarities in EEG signal carry information about cognition. This technique was successfully tested in two different databases: a working memory database, and an Alzheimer disease database. We also provide evidence suggesting that EEG might not be even piecewise stationary. Therefore, as changes between different stationary regimes are linked to transitions between metastable states in the brain, transitions between those states might not occur in a discrete manner, after a short period of metastability, but rather in a continuous way. Transitions between neighbouring states would occur more often, whereas large transitions occur as well. Large transitions suggesting discreteness had been detected by other techniques, but small fluctuations are not noise, as they can be successfully used to infer aspects of cognition.

I. INTRODUCTION

Electroencephalographic (EEG) recordings are known to be non stationary [1], however, it is considered that they are composed of concatenated stationary segments of an average length of 0.25 seconds. Non stationarity poses technical issues given that many popular techniques require stationarity: spectral methods, complexity measures and autoregressive models being examples. Moreover, the analysis of non stationarity of EEG signals has deeper theoretical relevance as well, when it is considered in the context of metastability.

A system in a metastable state will have stationary parameters, and according to Kaplan et al. [2], EEG non stationarities might arise from the switching of the metastable states of neural assemblies during brain functioning. Buzsáki [3, Chapter 5] suggests that the brain is in a high complexity, critical state, characterised by a power law in the EEG power spectral density, in the shape of pink noise. He proposes as well that the most important property of cortical brain dynamics is the ability to rapidly switch between metastable pink noise and oscillatory behaviour. Under this view, sensory or motor activity are perturbations that can temporarily reorganize the effective connectivity to induce transient stability by oscillations. An oscillatory, short lived regime can hold information required for psychological constructs, whereas the critical state allows for an efficient transition between states.

In the context of perception, Freeman [4] describes metastability in the neocortex with the recurrence of spatial patterns of phase and amplitude, or frames. These frames [5] carry the meaning of sensory information in spatial patterns of cortical activity that resemble discrete movie frames.

Despite the evidence suggesting that transitions between states are part of the brain functioning, few researchers investigate directly whether stationarity can convey relevant information about cognition. In general, the lack of stationarity is either not discussed at all, or regarded as an issue to overcome. The most popular approaches being signal segmentation in stationary epochs [6] [7], or the use of techniques that do not require the stationarity assumption [8] [9]. There are a few exceptions, for instance, Kaplan et al. [2] developed a technique in which synchrony between two channels is estimated as the degree to which they undergo simultaneous transitions. Cao and Slobounov [10] developed a measure of non stationarity based on the shift of the dominant frequency of the EEG signal over time. They use this measure to detect residual abnormalities in concussed individuals. While studying depth of anaesthesia, Kreuzer et al. [11] found that during loss of consciousness, stationarity is heavily influenced by the anaesthetic used.

In our study we make some hypotheses and explore whether they are consistent with experimental results obtained from the analysis of a Working Memory (WM) database and an Alzheimer Disease (AD) database. Our first hypothesis is about the relationship between stationarity and brain metastability. For a stationary process, all the statistical moments remain constant, and as our EEG time series have been detrended via high pass filtering, the first non vanishing moments are the variance, skewness and kurtosis. For a given channel, at a given time t_0 , let $v(t_0)$, $s(t_0)$, and $k(t_0)$ be the values of the variance, skewness and kurtosis over a short time window centred at t_0 . We hypothesize that the vector $(v(t_0), s(t_0), k(t_0)) \in \mathbb{R}^3$ can be used as a third order approximation to the current metastable state of the brain. For each channel, if we compute the third order approximation over a sliding window, we can create a time series, that we will call the channel state proxy time series (CSP(t), or simply CSP). It is expected to be an indicator of the time evolution of brain states in the region of the cortex where the recording was made. We suppose, just as [12], that changes in stationarity are already visible in the first statistical moments. We suggest that if cognition affects metastable state transitions, dynamical properties of the CSP can be linked with cognitive properties (high WM, AD disease, ...). By considering n moments, each CSP is a trajectory in an n -dimensional space, and dynamical aspects (velocity, acceleration) of this trajectory would be affected by cognition if the hypothesis holds true. If we define *neighbouring* states as states with similar statistical moments, the norm of the velocity of the CSP, $v_{el}(t) =$

¹ Brain plasticity Unit, CNRS, UMR8249, Paris, 75005, France
² ESPCI Paris, PSL Research University, Paris, 75005, France

$\frac{d}{dt}CSP(t) = \sqrt{v'(t)^2 + s'(t)^2 + k'(t)^2}$, expresses some distance between states, or transition size. If cognition affects either the occurrence or size of transitions, $v_{el}(t)$ would be sensitive to it. Our second hypothesis is about criticality. Power laws are not sufficient to guarantee criticality [13], however, scale free behaviour like power laws emerges from self-organized criticality. If indeed the brain is in a critical state that allows effective switching, we would expect a power law in the Power Spectral Density (PSD) of the CSP. If there is some functional meaning of this power law, its properties should be affected by cognition. An important property to look at is the scaling factor of the power law, as it determines its memory properties [14]: the extent to which past events affect the present, and hence the extent to which (sensory or motor) perturbations propagate.

Finally, the dynamical behaviour of the CSP can shed some light on the occurrence of state transitions. In particular, if metastability and hence piecewise stationarity hold, the CSP should remain constant except when transitions occur. Therefore, the velocity, being the derivative, should contain only informative data about cognition around times of transitions. The rest, being zero plus external and statistical noise (under or over estimation of the statistical moments in the sliding window). Due to the sliding window, the discontinuity in the CSP would not be completely discrete, but rather each transition would spread its influence to an interval of time equal to the length of the sliding window (from the moment the window encounters the changing point, to the moment it leaves it). The density of informative data can be easily computed: if every second N transitions occur and the moving window consists of M points, every second we have NM useful points, considering that the transition effect spreads through the entire window. If there are S points in a second, the density of useful points becomes $\rho = NM/S = NL$, where L is the time length of the moving window in seconds. The previous holds if $NL < 1$, which means that the window size is smaller than the average time spent in a particular state. Alternatively, by estimating ρ experimentally we can infer the value of N .

II. DATA ACQUISITION

Two datasets were used, a WM dataset and an AD dataset. Both studies followed the principles outlined in the Declaration of Helsinki. All participants were given explanations about the nature of the experiment and signed an informed consent form before the experiment started.

A. WM dataset

20 subjects performed a WM task as described in [15]. The dataset consists of 530 trials lasting 10 seconds each. 53 percent of them correspond to low WM load and 47 percent to high WM load. The dataset contains no obviously artifacted trials, and eye blinks were removed with Independent Component Analysis. The sampling rate is 500 Hz, and 16 channels of the international 10-20 system were used: Fp1, Fp2, F7, F3, Fz, F4, F8, Cz, CP5, CP6, P3, Pz, P4, O1, Oz, and O2.

B. AD dataset

Recordings from 60 subjects were collected, 22 of which were Mild Cognitive Impairment (MCI) patients, and the rest healthy, age-matched controls. AD patients were not included, as it is more interesting to classify between MCI and controls. For each subject 20 seconds of continuous recordings were available. Data were sampled at 200 Hz, using 21 channels: Fp1, Fp2, F3, F4, C3, C4, P3, P4, O1, O2, F7, F8, T7, T8, P7, P8, Fz, Cz, Pz, FPz, and Oz.

III. METHODS

All the hypotheses made are that a certain measurable quantity, or feature, conveys information about cognition. To test the hypotheses, features were fed to a Linear Discriminant Analysis (LDA) classifier. The task of the classifier was to predict whether the features corresponded to high WM load in the first dataset, or to MCI in the second dataset. If, after cross validation, the performance of the classifier was better than that of a random classifier, we can conclude that the features carry information about cognition. The design of the classifier was as in [15]. After feature extraction, relevant features were selected by orthogonal forward regression [16], and an LDA was trained with these features. The output of the classifier was the performance estimated with cross validation, and the set of most informative features. Due to the imbalance of the AD dataset, the Area Under the Curve (AUC) [17] of a Receiver Operating Characteristic (ROC) curve was used as performance measure. A ROC curve is not influenced by the imbalance of the classes, and its AUC value is typically 0.5 for a random classifier, and 1 for a perfect classifier, values larger than 0.5 indicate performance better than random. The statistical significance of the performance of the classifier was estimated by a permutation test.

Spectral features (Fourier analysis) were used as benchmark for comparison. They included absolute and relative power in the delta (1-4 Hz), theta (4-8 Hz), alpha (8-12 Hz), lower beta (12-20 Hz), upper beta (20-30 Hz) and lower gamma (30-45 Hz) ranges. Relative power being the fraction of the total power corresponding to a particular band.

A. Method 1: Dynamics of the CSP

For each channel, EEG data were filtered in the same bands as the spectral features, which gave rise to 6 time series per channel. For each channel and for each band, method 1 consisted in computing the CSP using sliding windows of length $L = 1/f_{min}$ seconds, where f_{min} is the minimum frequency of the corresponding band, so that each window contained at least one full oscillation of the smallest frequency. For every epoch, the mean value of the velocity and acceleration of the CSPs were computed as features.

In order to estimate the density of transitions N , each velocity time series was sorted by size. Only the smallest F fraction of the points was used to compute the mean velocity and mean acceleration. The performance of the classifier was studied as a function of F . For small values of F , we would pick only noise under the assumptions of piecewise stationarity, enough temporal resolution ($NL < 1$), and

transition size larger than noise. As F increases, at a certain threshold value $1 - \rho$ we would start obtaining useful points, and classification performance should begin to continuously increase and differ from random classification (see fig 1). N can be estimated as $N = \rho/L$.

Method 1 consisted of computing velocity and acceleration with the parameter F that maximises the performance of the classifier under cross validation. For attempting to estimate N , on the other hand, only the upper beta and lower gamma ranges were used, in order to have an adequate resolution. Also, only the WM database was used, as it was sampled at a higher rate.

B. Method 2: Power law of the power spectra

The CSP of order 3 is 3-dimensional. The norm of the CSP was used as an auxiliary-CSP (aCSP) to have a 1-dimensional representation of the state. A power law was fitted to the tail of the PSD of the aCSP. The CSP used for the power law was not filtered in any particular band, as we are studying the whole spectrum, therefore there is only one aCSP per channel. The power law hypothesis was successfully tested using the criteria of [18]. A linear fit in a log-log plot was performed and both of the parameters were used as features. The estimation of the slope suggested by the same paper [18] was not used, given that it provided significant, yet lower classification performance.

C. Method 3: Measuring synchrony. Spatial correlations between states

An alternative way to test the relevance of the CSP as a proxy of brain states is through potential synchrony. If it is indeed a proxy of brain states, we would expect brain synchrony to produce correlations between CSPs at certain channels and bands. The linear correlation coefficient between pairs of CSP was computed for each trial. Having six series per channel, and 16 (WM database) or 21 (AD database) channels, the potential number of combinations are in the order of several thousands. In order to prevent overfitting, only the CSP that provided the best features for Method 1 were taken into account for measuring the correlations. The number of combinations considered was obtained by cross validation, but was kept lower than 10.

IV. RESULTS

We begin by showing in figure 1 the performance of the classifier as a function of the fraction F , for the WM database, when using only features in the upper beta and lower gamma ranges. A diagram of what would be expected under piecewise stationarity is displayed as well. The expected value of ρ (the value at which classification should differ from random) was computed as $\rho = NL$. The window size used was $L = 1/20$ seconds, or 25 points.

When using all the bands, for the WM database the value of F that maximises the performance was the same, 0.6. For the AD database it was lower: 0.19.

The classification results of the different features are shown in figure 2. As a reference, the benchmark technique

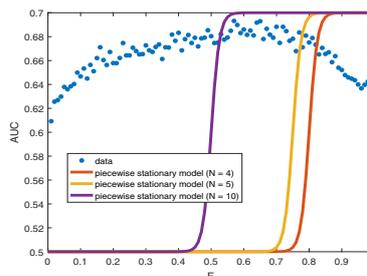


Fig. 1. WM database. In blue, the mean velocity of the CSP (using only upper beta and lower gamma ranges) was used as feature. To compute this mean, only the smallest F fraction of values was used. The classification performance is studied as a function of F . The solid lines are diagrams of what would be expected under piecewise stationarity, for different numbers N of transitions per second.

(spectral features of the EEG signal) is displayed for comparison. The p-values associated to the AUC of synchrony and power law fit for the AD database were $p < 0.01$, the other 8 AUCs yielded a p-value $p < 0.001$.

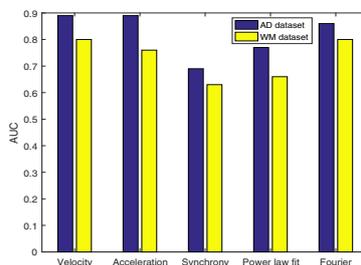


Fig. 2. Classification performance measured as the AUC for both databases.

V. DISCUSSION

Besides the hypothesis, we made some assumptions. First, that the size of the transitions, measured as the norm of the derivative of the CSP was larger than the noise in the system. It seems to be plausible as otherwise it would be hard to obtain classification results. Second, that the discrete, piece-wise stationary model implies that transitions occur at intervals larger than the smallest window length that can be used, the length being a function of the frequencies studied. Otherwise, discreteness is not a falsifiable statement, and the model becomes effectively continuous if transitions occur at intervals smaller than the minimum resolution. In addition,

we are not claiming that only cognition affects stationarity, as it was mentioned above anaesthetics do as well, and probably sleep, we nevertheless restrict our study to cognition.

Several parameters could have been optimised for each scenario (method and database), and performance would significantly improve. We decided however to keep those parameters fixed for all the methods and both of the datasets, to avoid the potential issue of multiple testing [19].

The first unexpected result is the shape of the plot in figure 1. N was impossible to estimate as there was no point at which suddenly classification increased from random. The behaviour was the opposite of what was expected. Indeed, taking the smallest 20% is better than taking all of them, thus removing the largest transitions indeed improved classification. It seems that small transitions are not just not noise, but they might carry more information than larger ones. This behaviour was observed also when considering all the bands, and for both of the datasets. In addition, the tail of the PSD of the CSP carries information about the high frequencies, or small transitions, and it was useful for classification as well. It is important to remember here that it is the PSD of the CSP, not of the raw EEG, therefore high frequencies mean small transitions, or transitions between neighbouring states. EEG is a particularly noisy signal, and for high frequencies the noise might be larger than the signal. A power law fit however allows us to infer the behaviour at the end of the tail (large frequencies, small transitions in this case) by observing more accessible regions of the spectrum.

All this evidence is more compatible with a scenario in which the brain undergoes large transitions at seemingly discrete times, but it keeps fluctuating between neighbouring states in a way that is affected by cognition. These findings are also compatible with the ideas about the continuity of mind exposed by Spivey [20]. He suggests that if we could possibly take the activity of each neuron as a variable, we could represent cognition as a (continuous) trajectory in a high dimensional space, where each coordinate represents the activity of a neuron. It might be not possible to explore the hypothetical trajectory in such a high dimensional state space. Nevertheless, if by using a low order proxy and a few scalp recordings it is possible to obtain useful dynamical quantities that are affected by cognition, the idea seems to be promising enough to further investigate it. It is not absurd to consider cognition as a trajectory in an abstract space.

At the introduction, several experimental findings considering discrete transitions were presented. Metastability implies piecewise stationarity, lack of stationarity does not necessarily imply lack of state stability, because there could be external sources of non-stationarities (slow drifts from electrodes drying, etc), but if we can correlate small transitions with cognition, we might benefit from modifying our view of stability. A continuous view in which small transitions occur much more often than large ones is compatible with evidence supporting discreteness. Small transitions would then seem like noise, whereas large ones would seem to occur at discrete intervals. Furthermore, if most of the fluctuations are small, the brain, although not properly piecewise

stationary, could still be considered so in practical terms. If we are cautious, some techniques requiring stationarity still work. However, a discrete view would at the very least ignore useful information contained in the small transitions, not to mention potential insights about brain functioning.

REFERENCES

- [1] Włodzimirz Klonowski, "Everything you wanted to ask about eeg but were afraid to get the right answer," *Nonlinear Biomedical Physics*, vol. 3, no. 1, pp. 2, 2009.
- [2] Alexander Ya Kaplan, Andrew A Fingelkurts, Alexander A Fingelkurts, Sergei V Borisov, and Boris S Darkhovsky, "Nonstationary nature of the brain activity as revealed by eeg/meg: methodological, practical and conceptual challenges," *Signal processing*, vol. 85, no. 11, pp. 2190–2212, 2005.
- [3] György Buzsáki, *Rhythms of the Brain*, Oxford University Press, 2006.
- [4] Walter J Freeman and Mark D Holmes, "Metastability, instability, and state transition in neocortex," *Neural Networks*, vol. 18, no. 5, pp. 497–504, 2005.
- [5] Walter J Freeman and Robert Kozma, "Freeman's mass action," *Scholarpedia*, 5(1):8040, 2010.
- [6] Rajeev Agarwal and Jean Gotman, "Adaptive segmentation of electroencephalographic data using a nonlinear energy operator," in *Circuits and Systems, 1999. ISCAS'99. Proceedings of the 1999 IEEE International Symposium on*. IEEE, 1999, vol. 4, pp. 199–202.
- [7] Gernot Florian and Gert Pfurtscheller, "Dynamic spectral analysis of event-related eeg data," *Electroencephalography and clinical neurophysiology*, vol. 95, no. 5, pp. 393–396, 1995.
- [8] Neep Hazarika, Jean Zhu Chen, Ah Chung Tsoi, and Alex Sergejew, "Classification of eeg signals using the wavelet transform," in *Digital Signal Processing Proceedings, 1997. DSP 97., 1997 13th International Conference on*. IEEE, 1997, vol. 1, pp. 89–92.
- [9] Andrew D Krystal, Raquel Prado, and Mike West, "New methods of time series analysis of non-stationary eeg data: eigenstructure decompositions of time varying autoregressions," *Clinical Neurophysiology*, vol. 110, no. 12, pp. 2197–2206, 1999.
- [10] Cheng Cao and Semyon Slobounov, "Application of a novel measure of eeg non-stationarity as 'shannon-entropy of the peak frequency shifting' for detecting residual abnormalities in concussed individuals," *Clinical Neurophysiology*, vol. 122, no. 7, pp. 1314–1321, 2011.
- [11] Matthias Kreuzer, Eberhard F Kochs, Gerhard Schneider, and Denis Jordan, "Non-stationarity of eeg during wakefulness and anaesthesia: advantages of eeg permutation entropy monitoring," *Journal of clinical monitoring and computing*, vol. 28, no. 6, pp. 573–580, 2014.
- [12] Paul Von Büna, Frank C Meinecke, Franz C Király, and Klaus-Robert Müller, "Finding stationary subspaces in multivariate time series," *Physical review letters*, vol. 103, no. 21, pp. 214101, 2009.
- [13] Mark EJ Newman, "Power laws, pareto distributions and zipf's law," *Contemporary physics*, vol. 46, no. 5, pp. 323–351, 2005.
- [14] Octavio Miramontes and Pejman Rohani, "Estimating $1/f$ α scaling exponents from short time-series," *Physica D: Nonlinear Phenomena*, vol. 166, no. 3, pp. 147–154, 2002.
- [15] Aldo Mora Sánchez, Antoine Gaume, Gerard Dreyfus, and François-Benoît Vialatte, "A cognitive brain-computer interface prototype for the continuous monitoring of visual working memory load," in *2015 IEEE 25th International Workshop on Machine Learning for Signal Processing (MLSP)*. IEEE, 2015, pp. 1–5.
- [16] Hervé Stoepiglia, Gérard Dreyfus, Rémi Dubois, and Yacine Oussar, "Ranking a random feature for variable and feature selection," *The Journal of Machine Learning Research*, vol. 3, pp. 1399–1414, 2003.
- [17] James A Hanley and Barbara J McNeil, "The meaning and use of the area under a receiver operating characteristic (roc) curve," *Radiology*, vol. 143, no. 1, pp. 29–36, 1982.
- [18] Aaron Clauset, Cosma Rohilla Shalizi, and Mark EJ Newman, "Power-law distributions in empirical data," *SIAM review*, vol. 51, no. 4, pp. 661–703, 2009.
- [19] Yoav Benjamini and Yoel Hochberg, "Controlling the false discovery rate: a practical and powerful approach to multiple testing," *Journal of the royal statistical society. Series B (Methodological)*, pp. 289–300, 1995.
- [20] Michael Spivey, *The continuity of mind*. Oxford University Press, 2008.

Session Name: PI-9

Room: R1

Session: Poster Session

Time slot: 17:30 – 20:00

Day: 1

**SIGMABOX: TOWARDS A SIMPLE AND EFFICIENT MATLAB
TOOLBOX FOR EEG SIGNAL PROCESSING AND
CLASSIFICATION**

Takfarinas Medani^{1,2}, Aldo Mora-Sánchez^{1,2}, Aurore Jaumard-Hakoun^{1,2}, Gerard Dreyfus^{1,2},
Francois Vialatte^{1,2}

Brain Plasticity Unit. Cnrs, Umr8249, Paris, 75005, France¹; Espci-Paris, Psl, Research
University, Paris, 75005, France²

E-mail address: {takfarinas.medani, aldo.mora-sanchez, aurore.hakoun, gerard.dreyfus,
francois.vialatte}@espci.fr

ABSTRACT: During the past few years, significant progress was made in devices and software for recording and analyzing bio-physiological signals. This short paper presents a general description of our current project: a new open-source Matlab-based toolbox, designed in order to help with biosignal data processing. SIGMAbox (SIGnal processing and MACHine Learning toolbox) gathers several pre-configured methods and algorithms for signal processing, statistics and classification. This toolbox is based on a graphical user interface (GUI) designed for end-users without expert skills in programming, and should be useable with very limited intervention from the user.

INTRODUCTION

SIGMABOX encapsulates a collection of existing Matlab functions and scripts. Those methods are pre-initialized and configured; however, their hyper parameters may be chosen by the user. The parameters for the implemented functions are initialized according to the best ones found in the literature and validated on our data-base, and can be optimized by the user if necessary. Various visualization options for the data and the results will be included on the GUI.

MATERIALS AND METHODS

The present version of the toolbox allows the design of two-class classifiers for EEG data [1]. It can be used on offline analysis for pre-recorded data-bases. Also it can be adapted for online system such as a brain-computer interface. The data analysis on SIGMABOX is divided into two phases: a training and validation phase, and a test phase. The user selects the database, and chooses the suitable method(s) for preprocessing, artifact detection and rejection, feature extraction and selection, and classification. The implemented features extraction methods include spectral and statistical analysis, complexity and synchrony measures. The classifications algorithms use the built-in Matlab toolboxes [2] for linear discriminant analysis (LDA), quadratic discriminant analysis (QDA) and support vector machines (SVM). The users can visualize their data, compare the performance of the different classifiers, and display the

CHAPTER 5. APPENDIX: PUBLISHED WORKS

sensitivity, specificity, error rate and ROC curves. In the case of EEG signals, options for visualizing the electrical topography of the detected brain activity and for signal source localization are available using the Brainstorm packages [3].

DISCUSSION

Contrary to the other available tools [4][5], advanced skills in programming are not needed to use SIGMAbox, most of the option can be reached from the main GUI. The toolbox offers options that users can select and run to get the desired results, together with illustrations helping in their interpretations. This allows for instance supervisors to verify the proper use of the toolbox by non-experts (*e.g.* master's degree students) involved in a research project. Other options are under investigation, and will be added in future versions of the toolbox for specific types of signals such as electrodermal responses or breathing signals. The first version is expected to be available at the end of this year.

REFERENCES

- [1] Lotte, F., Congedo, M., Lécuyer, A., Lamarche, F., & Arnaldi, B. (2007). A review of classification algorithms for EEG-based brain-computer interfaces. *Journal of neural engineering*, 4(2), R1.
- [2] Paluszek, M., & Thomas, S. (2016). *MATLAB Machine Learning*. Apress.
- [3] Tadel F, Baillet S, Moshier JC, Pantazis D, Leahy RM, "Brainstorm: A User-Friendly Application for MEG/EEG Analysis," *Computational Intelligence and Neuroscience*, vol. 2011, Article ID 879716, 13 pages, 2011. doi:10.1155/2011/879716
- [4] Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of neuroscience methods*, 134(1), 9-21.
- [5] Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational intelligence and neuroscience*, 2011, 1.

BIBLIOGRAPHY

- [1] D. ABÁSULO, R. HORNERO, C. GÓMEZ, M. GARCÍA, AND M. LÓPEZ, *Analysis of eeg background activity in alzheimer's disease patients with lempel–ziv complexity and central tendency measure*, *Medical engineering & physics*, 28 (2006), pp. 315–322.
- [2] B. ABEN, S. STAPERT, AND A. BLOKLAND, *About the distinction between working memory and short-term memory*, *Frontiers in psychology*, 3 (2012), p. 301.
- [3] H. ADELI, Z. ZHOU, AND N. DADMEHR, *Analysis of eeg records in an epileptic patient using wavelet transform*, *Journal of neuroscience methods*, 123 (2003), pp. 69–87.
- [4] R. AGARWAL AND J. GOTMAN, *Adaptive segmentation of electroencephalographic data using a nonlinear energy operator*, in *Circuits and Systems, 1999. ISCAS'99. Proceedings of the 1999 IEEE International Symposium on*, vol. 4, IEEE, 1999, pp. 199–202.
- [5] G. L. AHERN AND G. E. SCHWARTZ, *Differential lateralization for positive versus negative emotion*, *Neuropsychologia*, 17 (1979), pp. 693–698.
- [6] M. AKIN, *Comparison of wavelet transform and fft methods in the analysis of eeg signals*, *Journal of medical systems*, 26 (2002), pp. 241–247.
- [7] B. Z. ALLISON AND C. NEUPER, *Could anyone use a bci?*, in *Brain-computer interfaces*, Springer, 2010, pp. 35–54.

BIBLIOGRAPHY

- [8] P. ANTONENKO, F. PAAS, R. GRABNER, AND T. VAN GOG, *Using electroencephalography to measure cognitive load*, *Educational Psychology Review*, 22 (2010), pp. 425–438.
- [9] R. C. ATKINSON AND R. M. SHIFFRIN, *Human memory: A proposed system and its control processes*, *The psychology of learning and motivation*, 2 (1968), pp. 89–195.
- [10] A. BADDELEY, *Working memory*, *Science*, 255 (1992), pp. 556–559.
- [11] —, *The episodic buffer: a new component of working memory?*, *Trends in cognitive sciences*, 4 (2000), pp. 417–423.
- [12] —, *Working memory: looking back and looking forward*, *Nature reviews neuroscience*, 4 (2003), pp. 829–839.
- [13] A. D. BADDELEY AND G. HITCH, *Working memory*. In G.H. Bower (Ed.). *The psychology of learning and motivation: Advances in research and theory*, vol. 8, New York: Academic Press, 1974.
- [14] A. BAHRAMISHARIF, O. JENSEN, J. JACOBS, AND J. LISMAN, *Serial representation of items during working memory maintenance at letter-selective cortical sites*, *PLoS biology*, 16 (2018), p. e2003805.
- [15] C. L. BALDWIN AND B. PENARANDA, *Adaptive training using an artificial neural network and eeg metrics for within-and cross-task workload classification*, *NeuroImage*, 59 (2012), pp. 48–56.
- [16] J. S. BARLOW, *Methods of analysis of nonstationary eegs, with emphasis on segmentation techniques: a comparative review.*, *Journal of clinical neurophysiology: official publication of the American Electroencephalographic Society*, 2 (1985), pp. 267–304.
- [17] P. BARROUILLET, S. BERNARDIN, AND V. CAMOS, *Time constraints and resource sharing in adults' working memory spans.*, *Journal of Experimental Psychology: General*, 133 (2004), p. 83.

-
- [18] P. BARROUILLET AND V. CAMOS, *The time-based resource-sharing model of working memory*, *The cognitive neuroscience of working memory*, 455 (2007), pp. 59–80.
- [19] P. BARROUILLET, N. GAVENS, E. VERGAUWE, V. GAILLARD, AND V. CAMOS, *Working memory span development: a time-based resource-sharing model account.*, *Developmental psychology*, 45 (2009), p. 477.
- [20] A. J. BELL AND T. J. SEJNOWSKI, *An information-maximization approach to blind separation and blind deconvolution*, *Neural computation*, 7 (1995), pp. 1129–1159.
- [21] Y. BENJAMINI AND Y. HOCHBERG, *Controlling the false discovery rate: a practical and powerful approach to multiple testing*, *Journal of the royal statistical society. Series B (Methodological)*, (1995), pp. 289–300.
- [22] C. BERKA, D. J. LEVENDOWSKI, M. M. CVETINOVIC, M. M. PETROVIC, G. DAVIS, M. N. LUMICAO, V. T. ZIVKOVIC, M. V. POPOVIC, AND R. OLMSTEAD, *Real-time analysis of eeg indexes of alertness, cognition, and memory acquired with a wireless eeg headset*, *International Journal of Human-Computer Interaction*, 17 (2004), pp. 151–170.
- [23] W. BOSL, A. TIERNEY, H. TAGER-FLUSBERG, AND C. NELSON, *Eeg complexity as a biomarker for autism spectrum disorder risk*, *BMC medicine*, 9 (2011), p. 18.
- [24] G. BUZSAKI, *Rhythms of the Brain*, Oxford University Press, 2006.
- [25] C. CAO AND S. SLOBOUNOV, *Application of a novel measure of eeg non-stationarity as ‘shannon-entropy of the peak frequency shifting’ for detecting residual abnormalities in concussed individuals*, *Clinical Neurophysiology*, 122 (2011), pp. 1314–1321.

BIBLIOGRAPHY

- [26] L. CHANG AND D. Y. TSAO, *The code for facial identity in the primate brain*, *Cell*, 169 (2017), pp. 1013–1028.
- [27] W. CHENEY AND D. KINCAID, *Linear algebra: Theory and applications*, The Australian Mathematical Society, 110 (2009).
- [28] A. CLAUSET, C. R. SHALIZI, AND M. E. NEWMAN, *Power-law distributions in empirical data*, *SIAM review*, 51 (2009), pp. 661–703.
- [29] J. R. COMSTOCK JR AND R. J. ARNEGARD, *The multi-attribute task battery for human operator workload and strategic behavior research*, (1992).
- [30] N. COWAN, *Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system.*, *Psychological bulletin*, 104 (1988), p. 163.
- [31] —, *Verbal memory span and the timing of spoken recall*, *Journal of Memory and Language*, 31 (1992), pp. 668–684.
- [32] —, *An embedded-processes model of working memory.*, (1999).
- [33] —, *The magical mystery four: How is working memory capacity limited, and why?*, *Current directions in psychological science*, 19 (2010), pp. 51–57.
- [34] A. DELORME AND S. MAKEIG, *Eeglab: an open source toolbox for analysis of single-trial eeg dynamics including independent component analysis*, *Journal of neuroscience methods*, 134 (2004), pp. 9–21.
- [35] A. K. ENGEL AND P. FRIES, *Beta-band oscillations—signalling the status quo?*, *Current opinion in neurobiology*, 20 (2010), pp. 156–165.
- [36] D. ERDOGMUS, A. ADAMI, M. PAVEL, T. LAN, S. MATHAN, S. WHITLOW, AND M. DORNEICH, *Cognitive state estimation based on eeg*

-
- for augmented cognition*, in Conference Proceedings. 2nd International IEEE EMBS Conference on Neural Engineering, 2005., IEEE, 2005, pp. 566–569.
- [37] K. A. ERICSSON AND P. F. DELANEY, *Long-term working memory as an alternative to capacity models of working memory in everyday skilled performance*, *Models of working memory: Mechanisms of active maintenance and executive control*, (1999), pp. 257–297.
- [38] K. A. ERICSSON AND W. KINTSCH, *Long-term working memory.*, *Psychological review*, 102 (1995), p. 211.
- [39] A. A. FINGELKURTS AND A. A. FINGELKURTS, *Operational architectonics of the human brain biopotential field: towards solving the mind-brain problem*, *Brain and Mind*, 2 (2001), pp. 261–296.
- [40] ———, *Timing in cognition and eeg brain dynamics: discreteness versus continuity*, *Cognitive Processing*, 7 (2006), pp. 135–162.
- [41] R. A. FISHER, *The use of multiple measurements in taxonomic problems*, *Annals of eugenics*, 7 (1936), pp. 179–188.
- [42] G. FLORIAN AND G. PFURTSCHELLER, *Dynamic spectral analysis of event-related eeg data*, *Electroencephalography and clinical neurophysiology*, 95 (1995), pp. 393–396.
- [43] W. J. FREEMAN AND M. D. HOLMES, *Metastability, instability, and state transition in neocortex*, *Neural Networks*, 18 (2005), pp. 497–504.
- [44] W. J. FREEMAN AND R. KOZMA, *Freeman’s mass action*, *Scholarpedia*, 5(1):8040, (2010).
- [45] A. GAUME, A. VIALATTE, A. MORA-SÁNCHEZ, C. RAMDANI, AND F. VIALATTE, *A psychoengineering paradigm for the neurocognitive mechanisms of biofeedback and neurofeedback*, *Neuroscience & Biobehavioral Reviews*, 68 (2016), pp. 891–910.

BIBLIOGRAPHY

- [46] P. GERJETS, C. WALTER, W. ROSENSTIEL, M. BOGDAN, AND T. O. ZANDER, *Cognitive state monitoring and the design of adaptive instruction in digital environments: lessons learned from cognitive workload assessment using a passive brain-computer interface approach*, Using Neurophysiological Signals that Reflect Cognitive or Affective State, (2015), p. 20.
- [47] A. GEVINS, M. E. SMITH, H. LEONG, L. MCEVOY, S. WHITFIELD, R. DU, AND G. RUSH, *Monitoring working memory load during computer-based tasks with eeg pattern recognition methods*, Human Factors: The Journal of the Human Factors and Ergonomics Society, 40 (1998), pp. 79–91.
- [48] D. GORUR, U. HALICI, H. AYDIN, G. ONGUN, F. OZGEN, AND K. LEBLEBICIOGLU, *Sleep spindles detection using short time fourier transform and neural networks*, in Neural Networks, 2002. IJCNN'02. Proceedings of the 2002 International Joint Conference on, vol. 2, IEEE, 2002, pp. 1631–1636.
- [49] D. GRIMES, D. S. TAN, S. E. HUDSON, P. SHENOY, AND R. P. RAO, *Feasibility and pragmatics of classifying working memory load with an electroencephalograph*, in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM, 2008, pp. 835–844.
- [50] E. HAAPALAINEN, S. KIM, J. F. FORLIZZI, AND A. K. DEY, *Psychophysiological measures for assessing cognitive load*, in Proceedings of the 12th ACM international conference on Ubiquitous computing, ACM, 2010, pp. 301–310.
- [51] C. HAENSCHER, T. BALDEWEG, R. J. CROFT, M. WHITTINGTON, AND J. GRUZELIER, *Gamma and beta frequency oscillations in response to novel auditory stimuli: a comparison of human electroencephalogram (eeg) data with in vitro models*, Proceedings of the National Academy of Sciences, 97 (2000), pp. 7645–7650.

- [52] J. A. HANLEY AND B. J. MCNEIL, *The meaning and use of the area under a receiver operating characteristic (roc) curve.*, *Radiology*, 143 (1982), pp. 29–36.
- [53] E. HARMON-JONES AND J. J. ALLEN, *Anger and frontal brain activity: Eeg asymmetry consistent with approach motivation despite negative affective valence.*, *Journal of personality and social psychology*, 74 (1998), p. 1310.
- [54] S. G. HART AND L. E. STAVELAND, *Development of nasa-tlx (task load index): Results of empirical and theoretical research*, *Advances in psychology*, 52 (1988), pp. 139–183.
- [55] N. HAZARIKA, J. Z. CHEN, A. C. TSOI, AND A. SERGEJEV, *Classification of eeg signals using the wavelet transform*, in *Digital Signal Processing Proceedings, 1997. DSP 97., 1997 13th International Conference on*, vol. 1, IEEE, 1997, pp. 89–92.
- [56] D. HEGER, F. PUTZE, AND T. SCHULTZ, *Online workload recognition from eeg data during cognitive tests and human-machine interaction*, in *Annual Conference on Artificial Intelligence*, Springer, 2010, pp. 410–417.
- [57] T. HINTERBERGER, A. KÜBLER, J. KAISER, N. NEUMANN, AND N. BIRBAUMER, *A brain-computer interface (bci) for the locked-in: comparison of different eeg classifications for the thought translation device*, *Clinical Neurophysiology*, 114 (2003), pp. 416–425.
- [58] C. HUANG, L.-O. WAHLUND, T. DIERKS, P. JULIN, B. WINBLAD, AND V. JELIC, *Discrimination of alzheimer’s disease and mild cognitive impairment by equivalent eeg sources: a cross-sectional and longitudinal study*, *Clinical Neurophysiology*, 111 (2000), pp. 1961–1967.
- [59] A. HYVÄRINEN, P. RAMKUMAR, L. PARKKONEN, AND R. HARI, *Independent component analysis of short-time fourier transforms for*

BIBLIOGRAPHY

- spontaneous eeg / meg analysis*, *NeuroImage*, 49 (2010), pp. 257–271.
- [60] S. M. JAEGGI, M. BUSCHKUEHL, J. JONIDES, AND W. J. PERRIG, *Improving fluid intelligence with training on working memory*, *Proceedings of the National Academy of Sciences*, 105 (2008), pp. 6829–6833.
- [61] O. JENSEN, J. GELFAND, J. KOUNIOS, AND J. E. LISMAN, *Oscillations in the alpha band (9–12 hz) increase with memory load during retention in a short-term memory task*, *Cerebral cortex*, 12 (2002), pp. 877–882.
- [62] O. JENSEN AND C. D. TESCHE, *Frontal theta activity in humans increases with memory load in a working memory task*, *European journal of Neuroscience*, 15 (2002), pp. 1395–1399.
- [63] E. K ANDERS AND P. F DELANEY, *Working memory in everyday skilled performance*, *Models of working memory: Mechanisms of active maintenance and executive control*, 20 (1999), p. 274.
- [64] A. Y. KAPLAN, A. A. FINGELKURTS, A. A. FINGELKURTS, S. V. BORISOV, AND B. S. DARKHOVSKY, *Nonstationary nature of the brain activity as revealed by eeg / meg: methodological, practical and conceptual challenges*, *Signal processing*, 85 (2005), pp. 2190–2212.
- [65] J. S. KELSO, *Multistability and metastability: understanding dynamic coordination in the brain*, *Phil. Trans. R. Soc. B*, 367 (2012), pp. 906–918.
- [66] S. KHALFA, D. SCHON, J.-L. ANTON, AND C. LIÉGEOIS-CHAUVEL, *Brain regions involved in the recognition of happiness and sadness in music*, *Neuroreport*, 16 (2005), pp. 1981–1984.

-
- [67] W. KINTSCH, V. L. PATEL, AND K. A. ERICSSON, *The role of long-term working memory in text comprehension*, *Psychologia*, 42 (1999), pp. 186–198.
- [68] M. KLEINER, D. BRAINARD, D. PELLI, A. INGLING, R. MURRAY, AND C. BROUSSARD, *What's new in psychtoolbox-3*, *Perception*, 36 (2007), p. 1.
- [69] W. KLONOWSKI, *Everything you wanted to ask about eeg but were afraid to get the right answer*, *Nonlinear Biomedical Physics*, 3 (2009), p. 2.
- [70] S. KOELSTRA, C. MUHL, M. SOLEYMANI, J.-S. LEE, A. YAZDANI, T. EBRAHIMI, T. PUN, A. NIJHOLT, AND I. PATRAS, *Deap: A database for emotion analysis; using physiological signals*, *IEEE Transactions on Affective Computing*, 3 (2012), pp. 18–31.
- [71] R. KOHAVI ET AL., *A study of cross-validation and bootstrap for accuracy estimation and model selection*, in *Ijcai*, vol. 14, 1995, pp. 1137–1145.
- [72] J. KOHLMORGEN, G. DORNHEGE, M. BRAUN, B. BLANKERTZ, K.-R. MÜLLER, G. CURIO, K. HAGEMANN, A. BRUNS, M. SCHRAUF, W. KINCSES, ET AL., *Improving human performance in a real operating environment through real-time mental workload detection*, *Toward Brain-Computer Interfacing*, (2007), pp. 409–422.
- [73] B. KOLB AND I. WHISHAW, *Fundamentals of human neuropsychology. 2009*, New York: Worth, 5.
- [74] M. KOUKKOU, D. LEHMANN, J. WACKERMANN, I. DVORAK, AND B. HENGGELER, *Dimensional complexity of eeg brain mechanisms in untreated schizophrenia*, *Biological psychiatry*, 33 (1993), pp. 397–407.
- [75] M. KREUZER, E. F. KOCHS, G. SCHNEIDER, AND D. JORDAN, *Non-stationarity of eeg during wakefulness and anaesthesia: advan-*

BIBLIOGRAPHY

- tages of eeg permutation entropy monitoring*, *Journal of clinical monitoring and computing*, 28 (2014), pp. 573–580.
- [76] A. D. KRYSTAL, R. PRADO, AND M. WEST, *New methods of time series analysis of non-stationary eeg data: eigenstructure decompositions of time varying autoregressions*, *Clinical Neurophysiology*, 110 (1999), pp. 2197–2206.
- [77] F. LOTTE, F. LARRUE, AND C. MÜHL, *Flaws in current human training protocols for spontaneous brain-computer interfaces: lessons learned from instructional design*, *Frontiers in Human Neuroscience*, 7 (2013), p. 568.
- [78] A. LUTZ, *Toward a neurophenomenology as an account of generative passages: A first empirical case study*, *Phenomenology and the Cognitive Sciences*, 1 (2002), pp. 133–167.
- [79] A. LUTZ AND E. THOMPSON, *Neurophenomenology integrating subjective experience and brain dynamics in the neuroscience of consciousness*, *Journal of consciousness studies*, 10 (2003), pp. 31–52.
- [80] A. MAZAHERI, I. L. NIEUWENHUIS, H. VAN DIJK, AND O. JENSEN, *Prestimulus alpha and mu activity predicts failure to inhibit motor responses*, *Human brain mapping*, 30 (2009), pp. 1791–1800.
- [81] O. MIRAMONTES AND P. ROHANI, *Estimating $1/f$ α scaling exponents from short time-series*, *Physica D: Nonlinear Phenomena*, 166 (2002), pp. 147–154.
- [82] M. MISHKIN, L. G. UNGERLEIDER, AND K. A. MACKO, *Object vision and spatial vision: two cortical pathways*, *Trends in neurosciences*, 6 (1983), pp. 414–417.
- [83] M. MÖLLE, L. MARSHALL, W. LUTZENBERGER, R. PIETROWSKY, H. L. FEHM, AND J. BORN, *Enhanced dynamic complexity in the human eeg during creative thinking*, *Neuroscience Letters*, 208 (1996), pp. 61–64.

-
- [84] A. MORA-SÁNCHEZ AND F.-B. VIALATTE, *Is the lack of eeg stationarity useful? the dynamics of metastable brain states under cognition*, in 2017 Eight International IEEE EMBS Conference on Neural Engineering (NER'17), IEEE, 2017.
- [85] K.-R. MÜLLER, M. TANGERMANN, G. DORNHEGE, M. KRAULEDAT, G. CURIO, AND B. BLANKERTZ, *Machine learning for real-time single-trial eeg-analysis: from brain-computer interfacing to mental state monitoring*, Journal of neuroscience methods, 167 (2008), pp. 82–90.
- [86] G. R. MULLER-PUTZ AND G. PFURTSCHELLER, *Control of an electrical prosthesis with an ssvep-based bci*, IEEE Transactions on Biomedical Engineering, 55 (2008), pp. 361–364.
- [87] H. NAI-JEN AND R. PALANIAPPAN, *Classification of mental tasks using fixed and adaptive autoregressive models of eeg signals*, in Engineering in Medicine and Biology Society, 2004. IEMBS'04. 26th Annual International Conference of the IEEE, vol. 1, IEEE, 2004, pp. 507–510.
- [88] M. E. NEWMAN, *Power laws, pareto distributions and zipf's law*, Contemporary physics, 46 (2005), pp. 323–351.
- [89] F. PAAS, A. RENKL, AND J. SWELLER, *Cognitive load theory and instructional design: Recent developments*, Educational psychologist, 38 (2003), pp. 1–4.
- [90] A. J. PARKIN, *The central executive does not exist*, Journal of the International Neuropsychological Society, 4 (1998), pp. 518–522.
- [91] C. PETITMENGIN AND J.-P. LACHAUX, *Microcognitive science: bridging experiential and neuronal microdynamics*, Frontiers in human neuroscience, 7 (2013), p. 617.
- [92] G. PFURTSCHELLER, C. NEUPER, A. SCHLOGL, AND K. LUGGER, *Separability of eeg signals recorded during right and left motor*

BIBLIOGRAPHY

- imagery using adaptive autoregressive parameters*, IEEE transactions on Rehabilitation Engineering, 6 (1998), pp. 316–325.
- [93] P. N. PRINZ AND M. V. VITIELL, *Dominant occipital (alpha) rhythm frequency in early stage alzheimer’s disease and depression*, Electroencephalography and clinical neurophysiology, 73 (1989), pp. 427–432.
- [94] R. Q. QUIROGA, L. REDDY, G. KREIMAN, C. KOCH, AND I. FRIED, *Invariant visual representation by single neurons in the human brain*, Nature, 435 (2005), pp. 1102–1107.
- [95] C. L. RAYE, M. K. JOHNSON, K. J. MITCHELL, E. J. GREENE, AND M. R. JOHNSON, *Refreshing: A minimal executive function*, Cortex, 43 (2007), pp. 135–145.
- [96] L. C. ROBERTSON, *Binding, spatial attention and perceptual awareness*, Nature Reviews Neuroscience, 4 (2003), p. 93.
- [97] F. ROUX AND P. J. UHLHAAS, *Working memory and neural oscillations: alpha–gamma versus theta–gamma codes for distinct wm information?*, Trends in cognitive sciences, 18 (2014), pp. 16–25.
- [98] Y. RUIZ, S. POCKETT, W. J. FREEMAN, E. GONZALEZ, AND G. LI, *A method to study global spatial patterns related to sensory perception in scalp eeg*, Journal of neuroscience methods, 191 (2010), pp. 110–118.
- [99] A. M. SÁNCHEZ, A. GAUME, G. DREYFUS, AND F.-B. VIALATTE, *A cognitive brain-computer interface prototype for the continuous monitoring of visual working memory load*, in 2015 IEEE 25th International Workshop on Machine Learning for Signal Processing (MLSP), IEEE, 2015, pp. 1–5.
- [100] G. SANCHEZ, J. DAUNIZEAU, E. MABY, O. BERTRAND, A. BOMPAS, AND J. MATTOU, *Toward a new application of real-time*

electrophysiology: online optimization of cognitive neuroscience hypothesis testing, Brain sciences, 4 (2014), pp. 49–72.

- [101] P. SAUSENG, W. KLIMESCH, M. DOPPELMAYR, T. PECHERSTORFER, R. FREUNBERGER, AND S. HANSLMAYR, *Eeg alpha synchronization and functional coupling during top-down processing in a working memory task*, Human brain mapping, 26 (2005), pp. 148–155.
- [102] P. SAUSENG, W. KLIMESCH, M. SCHABUS, AND M. DOPPELMAYR, *Fronto-parietal eeg coherence in theta and upper alpha reflect central executive functions of working memory*, International Journal of Psychophysiology, 57 (2005), pp. 97–103.
- [103] B. SCHÖLKOPF, P. SIMARD, V. VAPNIK, AND A. SMOLA, *Improving the accuracy and speed of support vector machines*, Advances in neural information processing systems, 9 (1997), pp. 375–381.
- [104] U. SCHREITER-GASSER, T. GASSER, AND P. ZIEGLER, *Quantitative eeg analysis in early onset alzheimer’s disease: a controlled study*, Electroencephalography and clinical neurophysiology, 86 (1993), pp. 15–22.
- [105] S. T. SHORROCK, *Errors of memory in air traffic control*, Safety Science, 43 (2005), pp. 571–588.
- [106] M. SPIVEY, *The continuity of mind*, Oxford University Press, 2008.
- [107] M. L. STEYN-ROSS, D. A. STEYN-ROSS, M. T. WILSON, AND J. W. SLEIGH, *Modeling brain activation patterns for the default and cognitive states*, NeuroImage, 45 (2009), pp. 298–311.
- [108] H. STOPPIGLIA, G. DREYFUS, R. DUBOIS, AND Y. OUSSAR, *Ranking a random feature for variable and feature selection*, The Journal of Machine Learning Research, 3 (2003), pp. 1399–1414.
- [109] A. M. STRIJKSTRA, D. G. BEERSMA, B. DRAYER, N. HALBESMA, AND S. DAAN, *Subjective sleepiness correlates negatively with global*

BIBLIOGRAPHY

- alpha (8–12 hz) and positively with central frontal theta (4–8 hz) frequencies in the human resting awake electroencephalogram*, Neuroscience letters, 340 (2003), pp. 17–20.
- [110] J. A. SWETS, *Signal detection theory and ROC analysis in psychology and diagnostics: Collected papers*, Psychology Press, 2014.
- [111] D. TARABORELLI, *Feature binding and object perception. does object awareness require feature conjunction?*, in European Society for Philosophy and Psychology 2002, 2002.
- [112] E. TOGNOLI AND J. S. KELSO, *The metastable brain*, Neuron, 81 (2014), pp. 35–48.
- [113] J. N. TOWSE, G. J. HITCH, AND N. HORTON, *Working memory as the interface between processing and retention: a developmental perspective.*, Advances in child development and behavior, 35 (2007), p. 219.
- [114] A. TREISMAN, *Feature binding, attention and object perception*, Philosophical Transactions of the Royal Society of London B: Biological Sciences, 353 (1998), pp. 1295–1306.
- [115] L. VAN DER MAATEN, M. CHEN, S. TYREE, AND K. Q. WEINBERGER, *Learning with marginalized corrupted features.*, in ICML (1), 2013, pp. 410–418.
- [116] F. VARELA, J.-P. LACHAUX, E. RODRIGUEZ, AND J. MARTINERIE, *The brainweb: phase synchronization and large-scale integration*, Nature reviews neuroscience, 2 (2001), p. 229.
- [117] F. J. VARELA, *Neurophenomenology: A methodological remedy for the hard problem*, Journal of consciousness studies, 3 (1996), pp. 330–349.
- [118] F. J. VARELA, E. THOMPSON, AND E. ROSCH, *The embodied mind: Cognitive science and human experience*, MIT press, 2017.

- [119] F. VIALATTE, A. CICHOCKI, G. DREYFUS, T. MUSHA, T. M. RUTKOWSKI, AND R. GERVAIS, *Blind source separation and sparse bump modelling of time frequency representation of eeg signals: New tools for early detection of alzheimer's disease*, in Machine Learning for Signal Processing, 2005 IEEE Workshop on, IEEE, 2005, pp. 27–32.
- [120] F.-B. VIALATTE AND A. CICHOCKI, *Split-test bonferroni correction for qeeg statistical maps*, Biological Cybernetics, 98 (2008), pp. 295–303.
- [121] E. K. VOGEL, A. W. MCCOLLOUGH, AND M. G. MACHIZAWA, *Neural measures reveal individual differences in controlling access to working memory*, Nature, 438 (2005), pp. 500–503.
- [122] P. VON BÜNAU, F. C. MEINECKE, F. C. KIRÁLY, AND K.-R. MÜLLER, *Finding stationary subspaces in multivariate time series*, Physical review letters, 103 (2009), p. 214101.
- [123] A. F. WALLACE, *General: Plans and the structure of behavior. george a. miller, eugene galanter, and karl h. pribram*, American Anthropologist, 62 (1960), pp. 1065–1067.
- [124] P. WELCH, *The use of fast fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms*, IEEE Transactions on audio and electroacoustics, 15 (1967), pp. 70–73.
- [125] G. WERNER, *Metastability, criticality and phase transitions in brain and its models*, Biosystems, 90 (2007), pp. 496–508.
- [126] G. F. WILSON AND C. A. RUSSELL, *Real-time assessment of mental workload using psychophysiological measures and artificial neural networks*, Human Factors: The Journal of the Human Factors and Ergonomics Society, 45 (2003), pp. 635–644.

BIBLIOGRAPHY

- [127] J. R. WOLPAW, N. BIRBAUMER, W. J. HEETDERKS, D. J. MCFARLAND, P. H. PECKHAM, G. SCHALK, E. DONCHIN, L. A. QUATRANO, C. J. ROBINSON, T. M. VAUGHAN, ET AL., *Brain-computer interface technology: a review of the first international meeting*, IEEE transactions on rehabilitation engineering, 8 (2000), pp. 164–173.
- [128] T. O. ZANDER AND C. KOTHE, *Towards passive brain–computer interfaces: applying brain–computer interface technology to human–machine systems in general*, Journal of Neural Engineering, 8 (2011), p. 025005.
- [129] P. ZARJAM, J. EPPS, AND F. CHEN, *Characterizing working memory load using eeg delta activity*, in Signal Processing Conference, 2011 19th European, IEEE, 2011, pp. 1554–1558.
- [130] X.-S. ZHANG, R. J. ROY, AND E. W. JENSEN, *Eeg complexity as a measure of depth of anesthesia for patients*, IEEE Transactions on Biomedical Engineering, 48 (2001), pp. 1424–1433.