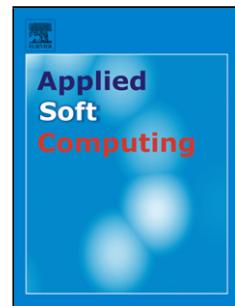


Accepted Manuscript



Title: EEG classification for the detection of mental states

Author: Laurent Vézard Pierrick Legrand Marie Chavent
Frédérique Faïta-Aïnseba Leonardo Trujillo

PII: S1568-4946(15)00186-6

DOI: <http://dx.doi.org/doi:10.1016/j.asoc.2015.03.028>

Reference: ASOC 2857

To appear in: *Applied Soft Computing*

Received date: 27-6-2014

Revised date: 6-2-2015

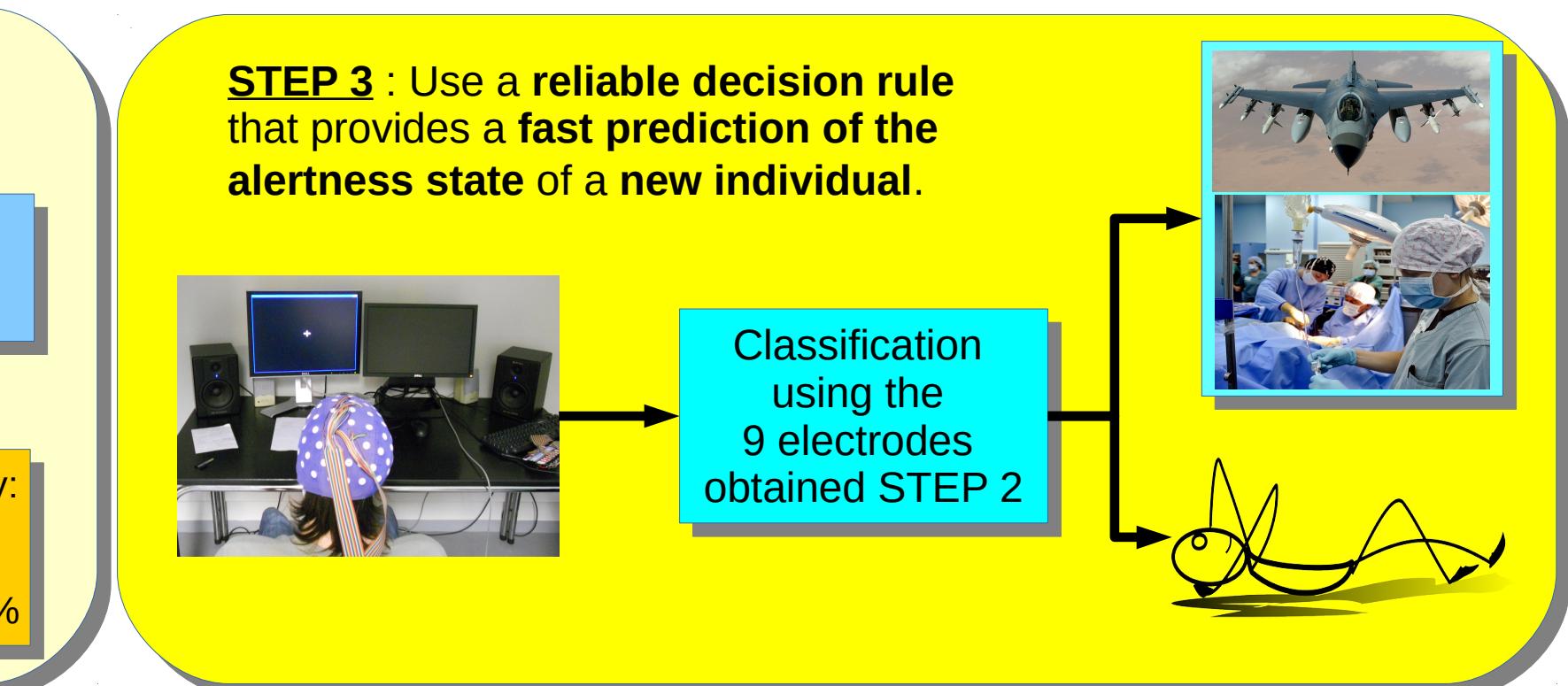
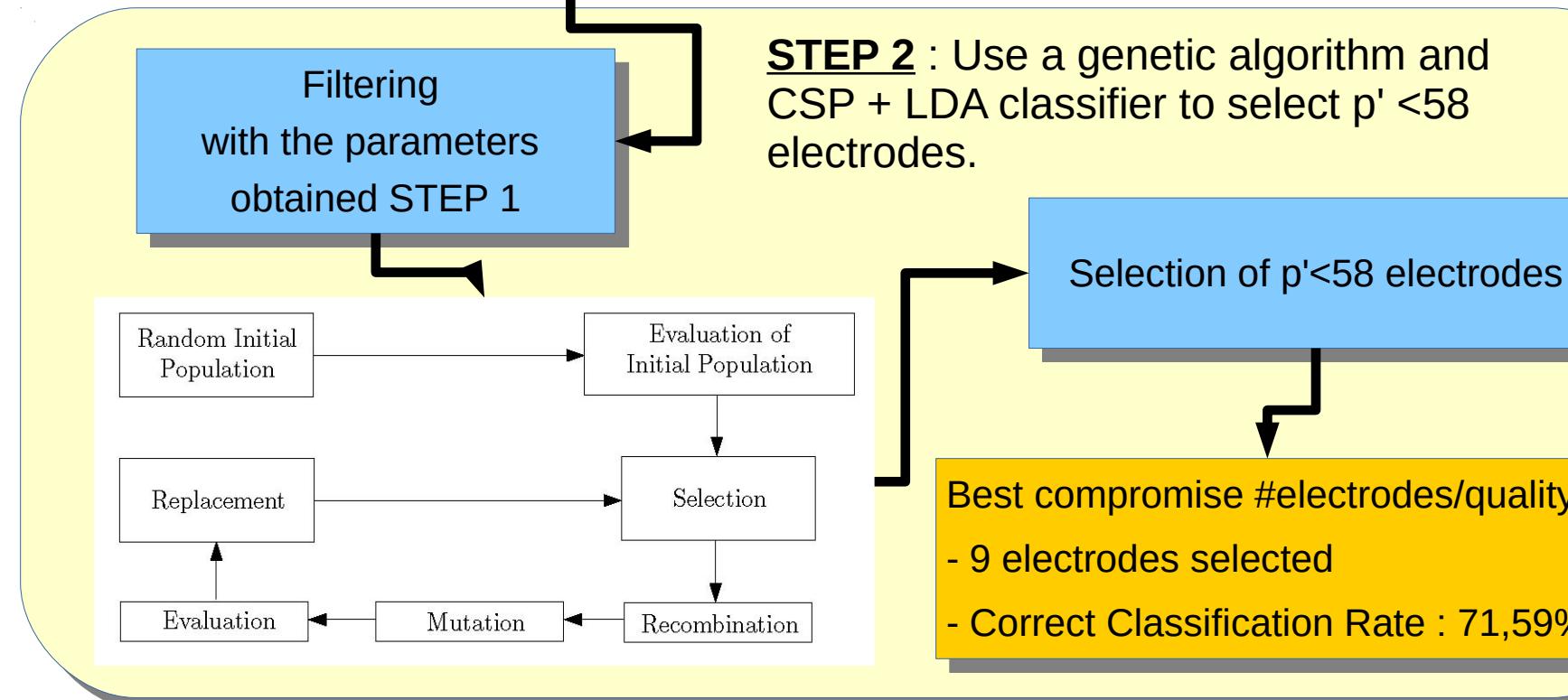
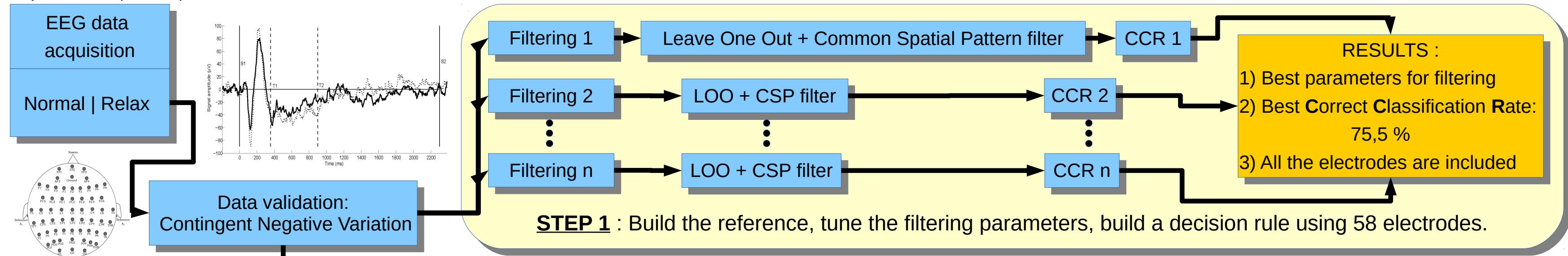
Accepted date: 12-3-2015

Please cite this article as: Laurent Vézard, Pierrick Legrand, Marie Chavent, Frédérique Faïta-Aïnseba, Leonardo Trujillo, EEG classification for the detection of mental states, *Applied Soft Computing Journal* (2015), <http://dx.doi.org/10.1016/j.asoc.2015.03.028>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

- A method able to **automatically determine mental states of alertness** is developed.
- A **complete protocol of EEG data acquisition** is detailed.
- Analysis of the Contingent Negative Variation is used **to validate the EEG data**.
- Best compromise ***number of electrodes / quality of the solution*** : 9 electrodes.
- Provides an **accurate and fast prediction of the alertness** of an **unseen** individual.

Accepted Manuscript



EEG classification for the detection of mental states

Laurent Vézard^{a,b,c}, Pierrick Legrand^{a,b,c,*}, Marie Chavent^{a,b,c},
Frédérique Faïta-Aïnseba^c, Leonardo Trujillo^d

^a*IMB, UMR CNRS 5251, 351, cours de la Libération, 33405 Talence, France*

^b*INRIA Bordeaux Sud-Ouest, 200 Rue Vieille Tour, 33405 Talence, France*

^c*Université de Bordeaux, 3ter place de la Victoire, 33076 Bordeaux, France*

^d*Tree-Lab, Posgrado en Ciencias de la Ingeniería, Instituto Tecnológico de Tijuana,
Calzada Del Tecnológico S/N Fraccionamiento Tomas Aquino, 22414 Tijuana, Baja
California, México*

Abstract

The objective of the present work is to develop a method that is able to automatically determine mental states of vigilance; i.e., a person's state of alertness. Such a task is relevant to diverse domains, where a person is expected or required to be in a particular state of mind. For instance, pilots and medical staff are expected to be in a highly alert state and the proposed method could help to detect possible deviations from this expected state. This work poses a binary classification problem where the goal is to distinguish between a "relaxed" state and a baseline state ("normal") from the study of electroencephalographic signals (EEG) collected with a small number of electrodes. The EEG of 58 subjects in the two alertness states (116 records) were collected via a cap with 58 electrodes. After a data validation step, 19 subjects were retained for further analysis. A genetic algorithm was used to select a subset of electrodes. Common spatial pattern (CSP) coupled to linear discriminant analysis (LDA) was used to build a decision rule and thus predict the alertness of the subjects. Different subset sizes were investigated and the best compromise between the number of selected electrodes and the quality of the solution was obtained by considering 9 electrodes. Even

*Corresponding author

Email addresses: laurentvezard@gmail.com (Laurent Vézard),
pierrick.legrand@u-bordeaux.fr (Pierrick Legrand), marie.chavent@inria.fr
(Marie Chavent), [frédérique faïta@u-bordeaux2.fr](mailto:frederique.faita@u-bordeaux2.fr) (Frédérique Faïta-Aïnseba),
leonardo.trujillo@tectijuana.edu.mx (Leonardo Trujillo)

if the present approach is costly in computation time (GA search), it allows to construct a decision rule that provides an accurate and fast prediction of the alertness state of an unseen individual.

Keywords: Electroencephalographic signals, Alertness, Genetic algorithm, Common spatial pattern, Variable selection

1. Introduction

Over the last decade, Human-Computer Interaction (HCI) has grown and matured as a field [1, 2]. Gone are the days when only a mouse and keyboard could be used to interact with a computer. The most ambitious of

- 5 such interfaces are Brain-Computer Interaction (BCI) systems. The goal in BCI is to allow a person to interact with an artificial system using only his brain activity. The most common approach towards BCI is to analyse, categorize and interpret Electroencephalographic signals (EEG), in such a way that they alter the state of a computer.

10

In particular, the objective of the present work is to study the development of computer systems for the automatic analysis and classification of mental states of vigilance; i.e., a person's state of alertness. Such a task is relevant to diverse domains, where a person is expected or required to be in a
15 particular state. For instance, pilots, security personnel or medical personnel are expected to be in a highly alert state, and a BCI could help to confirm this or detect possible problems. However, this task is by no means a trivial one. In fact, EEG signals are known to be non-stationary, highly noisy, irregular and tend to vary significantly from person to person, making the
20 development of general techniques a very challenging scientific endeavour [3]. Therefore, it is important to develop robust methods, adaptable to different persons, that can give a rapid and accurate prediction of the alertness state.

EEG signals are almost always pre-processed before any further analysis
25 is performed. The main goal of pre-processing methods is to perform feature extraction and pose a classification problem. To obtain these features, signal processing tools such as the Fourier transform or discrete wavelet decomposition (DWT) can be applied. For example, a DWT can be used to select the wavelet sub-band frequencies δ (1-3.5 Hz), θ (4-8 Hz), α (8-12 Hz) and
30 β (19-26 Hz) that are given as an input to a neural network classifier [3].

Another approach that uses coefficients of a discrete wavelet decomposition as features to describe the EEG signal can be found in [4]. An artificial neural network is then used to determine if the EEG record comes from a schizophrenia patient, an obsessive-compulsive disorder patient or a healthy subject. In [5], the EEG signal is decomposed into 23 bands of 1 Hz (from 1 to 23Hz) and a Short Term Fast Fourier transformation (STFFT) is used to calculate the percentage of the power spectrum of each band. In [6], a Fourier transform is used between hidden layers of a convolutional neural networks to switch from the time domain to the frequency domain analysis in the network.

Another common method to obtain features from EEG signals is the common spatial pattern (CSP). This method was introduced by Funkunaga and Koontz [7]. The CSP method was used to analyse EEG signals and to extract features for different classification tasks; for example to differentiate between normal and abnormal EEG signals [8] or for movement classification [9], [10], [11].

Once features are extracted, a classification method is applied to achieve the classification task. When a Fourier transform or DWT is used as pre-processing, the most common classification method is neural network (see for example [3] or [12]). However, the disadvantage of this approach is that it requires having a large set of test subjects relative to the number of predictive variables. To avoid this problem, it is possible to split the EEG signal into several segments of a few seconds, called “epochs” (see [3] and [12]). Other approaches use different statistical methods. For example, [13] uses Support Vector Machine (SVM), [14] Autoregressive Models (AR) and [15] hidden Markov chains.

When the CSP is used to extract features, the most common classification method is Linear Discriminant Analysis (LDA) as described in [16] or [11], even if other classification methods can be used, such as neural networks [17].

Main contributions. The aim of this work is to construct a model that is able to predict the alertness state of a human using the smallest possible number of EEG electrodes. That is why, the two main objectives are:

- 65 ● Reduce the time needed to install the EEG cap on a subject using a variable selection method to determine the best compromise between

the number of variables (electrodes) kept and the quality of the solution (based on classification rate). In fact, in real world applications, it is necessary to reduce the number of electrodes needed because the cap installation process has to be short. Moreover, a long installation of the EEG cap can cause a disturbance of the mental state of the person that we want to study.

- 70 • Obtain a model (decision rule) which is able to give a reliable prediction of the alertness state of a new subject.

75 In this paper, the CSP method coupled to LDA will be used to construct the decision rule and thus predict the alertness of the subjects. Moreover, a genetic algorithm selects a subset of variables on which the CSP method will be executed. In this work, the CSP method is used to extract features from EEG data signals of different individuals. It represents one of the most important contributions of this work because usually the CSP is used on a single subject to obtain subject specific discriminative spatial filters.

80 85 The remainder of this paper proceeds as follows. Section 2 presents the data acquisition procedure that was applied to record EEG signals. Section 3 deals with the CSP method and the genetic algorithm that is used to perform variable selection is detailed in Section 4. Finally, the results of this work are discussed in Section 5.

90 **2. Data acquisition**

This section describes the data acquisition process that was used to record the EEG signals and details how the data is validated.

2.1. Procedure

95 EEG signals are recorded using a cap with 58 electrodes placed according to the international system 10/10 defined in [18] and shown in Figure 1.

100 During the data acquisition procedure, the subject is alone in a sound-proof room, comfortably seated in front of a computer screen. It takes approximately two hours and a half to place the EEG cap on the subject, perform the data acquisition procedure and perform a final explanatory interview with the subject. In particular, the interview takes place at the

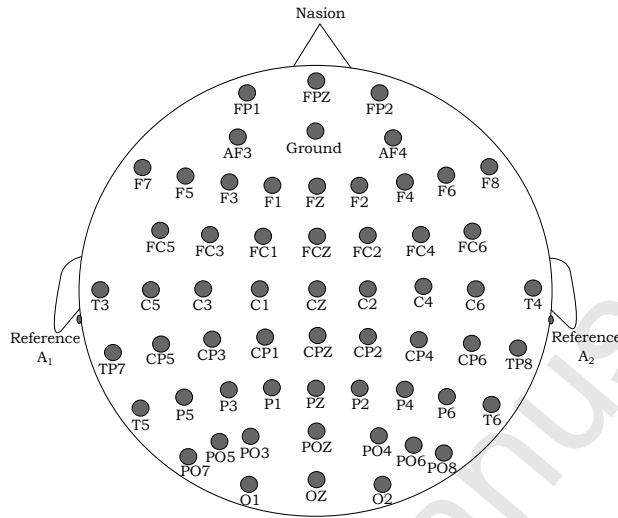


Figure 1: Representation of the distribution of electrodes in the international system 10/10.

end of the data acquisition procedure to not introduce any bias in the EEG records. Data collection was controlled by the acquisition system Coherence 3NT (Deltamed, <http://www.natus.com/>). In the present work, the data acquisition procedure is composed of five main steps which are boxed in Figure 2.

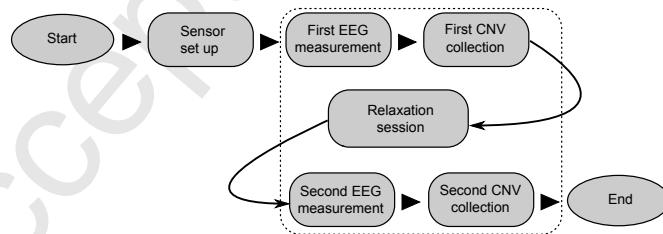


Figure 2: Diagram of the acquisition procedure. The five main steps of the procedure are boxed.

1. First EEG measurement: the subject has to look at a cross (fixation point) at the center of a computer screen to reduce eye movements. This first recording corresponds to the reference state, considered as the normal vigilance state of the subject. A photograph of a member

110

of our team, taken to illustrate the conditions during an EEG recording, is shown in Figure 3.



Figure 3: Photograph that illustrates the conditions during an EEG recording.

115

120

125

130

2. First Contingent Negative Variation (CNV) collection: the subject was instructed to press the space bar as quickly as possible after each time the cross was replaced by a square on the screen. For each appearance of the square, a warning sound (beep) was given 2.5 seconds before it appeared, allowing the subject to prepare his response. The experimental session included 50 pairs of stimuli (S1: beep, S2: square), with a random amount of time elapsing between each pair. The purpose of this attentional task is to collect the CNV of the subject. The CNV is detailed in Subsection 2.3.
3. Relaxation session: the subject was guided by a soundtrack broadcast through loudspeakers placed in the room. The soundtrack suggested the subject to perform three successive exercises of self-relaxation, based on muscular exercises and mental visualization. The first exercise is autogenic training [19] where the subject has to mentally repeat some sentences favouring the self-hypnosis. The second exercise is a progressive relaxation [20], that consists in flexing and unflexing some muscles. The final exercise is mental visualization, where the subject imagines that he is moving in a familiar and lovely place.
The purpose of this relaxation session is to try to bring the subject to a lower level of vigilance, qualified as the “relaxed” state.
4. Second EEG measurement: in this step, 3 minutes of EEG were recorded with the same protocol as in the step 1. This second recording should

reflect the low alertness state of the subject's brain ("relaxed" state) if
 135 it was reached in the prior step.

5. Second CNV collection: CNV is collected using exactly the same protocol as in step 2.

2.2. Subjects

This work uses 58 subjects (35 women and 23 men), aged from 18 to 35
 140 years old. They are right-handed, to avoid variations in the characteristics of the EEG due to age or handedness linked to a functional inter-hemispheric asymmetry. The EEG of 44 subjects were recorded during a first acquisition campaign in 2011 and the EEG of the 14 other subjects were recorded during a second acquisition campaign in 2012.

145 2.3. Contingent negative variation extraction

An Event-Related Potential (ERP) is a brief change of the brain electrical activity which occurs in response to an external stimulus (visual, olfactory, acoustic, ...) or an internal stimulus (preparation of a cognitive or motor action). Evoked potentials are synchronized with the triggering stimulus.
 150 The CNV is a particular ERP that appears when the paradigm detailed in Section 2.1-2 is used [21]. Because of their low-amplitude, ERPs are not easy to detect on an EEG record. That is why, several trials (EEG records after the triggering stimulus) are necessary. For a given electrode, the collected EEG signals are superimposed by taking the triggering stimulus as starting time ($t = 0$). Then, ERP extraction is performed by averaging all these
 155 electrical activities [22]. This allows event-related brain activity components, reflecting stimulus processing, to emerge from the overall cortical electrical activity, unrelated to the task performed. The CNV is a negative deflection of the averaged waveform (on 50 experiments in our case). This attentional component has the property of decreasing in amplitude when the subject
 160 is less alert, either because he is distracted [23], is deprived of sleep [24] or is falling asleep [25]. This fundamental result is shown in Figure 4. In this figure, the CNV is plotted as a dotted line for a subject with "normal" alertness and as a solid line for a subject with "low" alertness ("relaxed").
 165 The amplitude of the CNV is proportional to the alertness of the subject. Thus, for a given subject, the CNV analysis will allow us to determine if the relaxation step was effective.

That is why, although the instruction given to the subject during CNV acquisition was to press the space bar as quickly as possible after the square

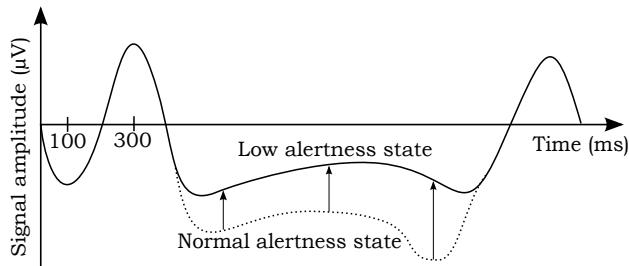


Figure 4: Representation of the amplitude variation of the CNV with respect to the alertness of a subject.

¹⁷⁰ appearance, the reaction time is not investigated in this study. Instead, the way that the subject prepares to perform the task is analysed.

The comparison of the amplitude of the CNV between the tasks performed in steps 2 and 5 is used to determine if the alertness of a subject has changed. ¹⁷⁵ This comparison allows us to know if he is actually relaxed. Only the positive cases, for which the amplitude of the CNV has significantly declined, were selected for comparative analysis of their raw EEG's (stages 1 and 4). Their EEG were then tagged respectively as belonging to the "normal" or "relaxed" state.

¹⁸⁰ An example of a subject kept after studying his CNV is shown in Figure 5 and an example of a rejected subject is given in Figure 6.

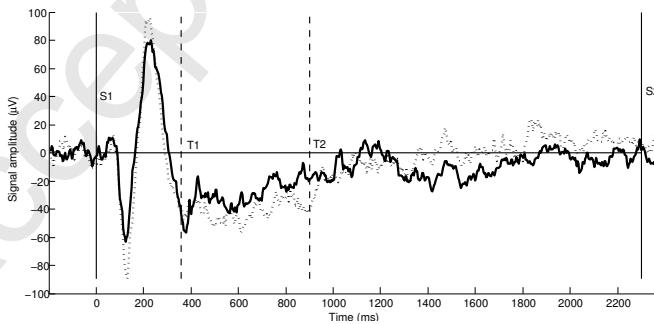


Figure 5: An example of a CNV recorded from subject 4 during steps 2 (dotted curve) and 5 (solid curve) of our methodology. The solid vertical lines correspond to warning signals (S1: beep, S2: square). This subject is kept because the solid curve is mostly above the dotted curve between T1 and T2 (framed by the dotted vertical lines).

In these figures, the dotted curve represents the CNV recorded during

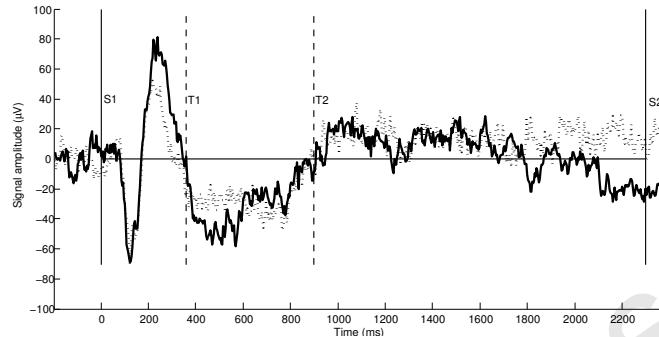
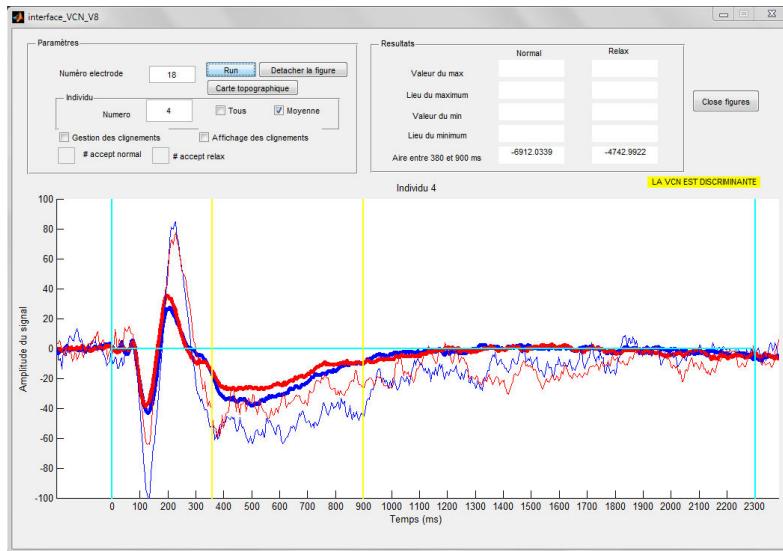


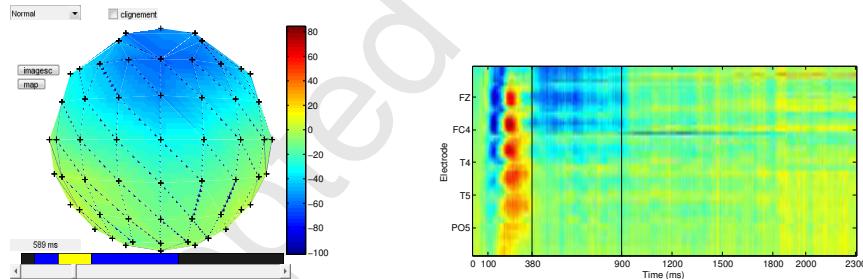
Figure 6: An exemple of a CNV recorded from subject 9 during steps 2 (dotted curve) and 5 (solid curve) of our methodology. The solid vertical lines correspond to warning signals (S1: beep, S2: square). This subject is rejected because the solid curve is mostly below the dotted curve between T1 and T2 (framed by the dotted vertical lines).

step 2 and the solid curve represents the CNV recorded in step 5. The solid vertical lines correspond to warning signals (S1: beep, S2: square). The area
185 between the curve and the x-axis is calculated between T1 and T2 (section framed by the dotted vertical lines). A subject is kept if the area calculated with the CNV recorded in step 5 is lower than the area calculated with the CNV recorded in step 2. To facilitate this validation step an allow a visual inspection of the curves, a graphical user interface was created. This interface,
190 shown in Figure 7, provides three kinds of output given in Figures 7(a), 7(b) and 7(c). Figure 7(a) allows to visualize the CNV curves of a chosen subject for both alertness states. The bold curves represents the mean of these curves over all the subject for the two alertness states. Figure 7(b) is a topographic map of the chosen subject brain in a given alertness state. This map depicts
195 the electrical activity recorded at a given time. Here, the chosen time point is 589 ms after S1 and corresponds to a time instant during the CNV. Thus, this map shows the appearance of the CNV on the scalp and allows us to visualize which brain regions are involved in the CNV appearance. Figure
200 7(c) represents the whole signal of the chosen subject for each electrodes (ordinate) in function of the time. This figure gives a rapid overview of the signal and can be useful to detect possible problems (for example, electrodes that record an abnormally high electrical activity).

The study of CNV was performed on the 58 subjects and 19 subjects were
205 kept for further analysis (13 from the first acquisition campaign and 6 from



(a) Visualization the CNV of a given subject in the two alertness state. The bold curves represents the mean of these curves over all the subject for a given alertness state.



(b) Topographic map that allows to visualize the electrical activity recorded at a given time during the CNV record on a given subject.

(c) Representation of the whole signal for each electrode (ordinate) as a function of time. Vertical lines frame the part of the signal on which the CNV area is calculated.

Figure 7: Graphical user interface for the CNV display.

the second). Thus, several subjects were not used for EEG classification. Their EEG recordings were not useful, probably as a result of the following factors. The stress due to the experiment and the duration of the installation of the cap may be factors that deteriorated the efficiency of the relaxation

210 session. Moreover, to limit the amount of time that a subject wears the EEG cap, the relaxation session was maintained relatively short. Thus, it is possible that the relaxation session (20 minutes) is too short to fully relax the subjects. Therefore, the subjects selected are those that were relaxed in a relatively short period of time and in conditions that can be stressful.

215 *2.4. Data*

A representation of the data matrix is given in Figure 8. The data consist of 38 records of 3 minutes of raw EEG signals from the 19 selected subjects (one “normal” EEG and one “relaxed” EEG for each subject). Each record contains variations of electric potential obtained with a sampling frequency 220 of 256 Hz (Deltamed acquisition system) with 58 active electrodes placed on a cap (ElectroCap). Using this sampling frequency, each signal recorded by an electrode for a given subject in a given alertness state contains 46000 data points.

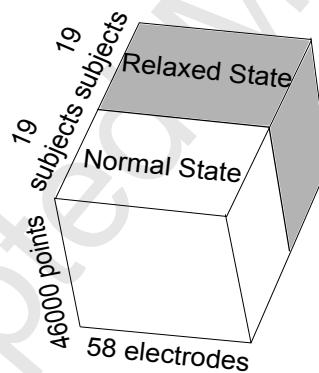


Figure 8: Representation of the data matrix. There are three dimensions: one for the subjects (2 records by subjects), one for the time (46000 points corresponding to the number of points in each 3 minutes EEG signals recorded using a sampling frequency of 256 Hz) and one for the electrodes.

3. EEG signals classification

225 Let X be an original EEG data matrix of dimensions $T \times p$ where p is the number of channels (recording electrodes) and T is the number of sampled time points. A spatial filter $\mathbf{w} \in \mathbb{R}^p$ is used to construct a small number of

synthetic signals defined as a linear combinations (projections) of the columns of X . The linear projection of X using filter \mathbf{w} is given by $\mathbf{f} = X\mathbf{w}$.

²³⁰ Principal Component Analysis (PCA) is a very well-known method to compute \mathbf{w} (loadings) and \mathbf{f} (principal components) in the context of unsupervised classification problems. Here, for EEG signals classification, the Common Spatial Filter (CSP) is used to obtain \mathbf{w} (spatial filters) and \mathbf{f} that are called CSP components.

²³⁵ *3.1. Common spatial pattern (CSP)*

The CSP is widely used to find spatial filters for binary supervised classification of EEG signals [16] [26]. The aim of CSP is to find the linear projections that maximize the variance ($\mathbf{w}'X'X\mathbf{w}$ or $\mathbf{f}'\mathbf{f}$) of signals of one condition and at the same time minimize the variance of another condition.

²⁴⁰ Afterwards, the variances of these synthetic signals are used to train a binary classifier.

²⁴⁵ Let $\{1, \dots, i, \dots, n\}$ be a set of n trials with n_1 in condition 1 (class 1) and n_2 in condition 2 (class 2). Let $X_i, i = 1, \dots, n$ be the EEG data matrix of these trials of dimension $T \times p$. We assume zero mean for EEG signals. The EEG data matrices of trials in class 1 are row concatenated in the matrix \mathbf{X}_1 of dimension $Tn_1 \times p$. In the same way, matrix \mathbf{X}_2 of dimension $Tn_2 \times p$ is obtained by concatenating only trials in class 2.

We can now define \mathbf{C}_1 and \mathbf{C}_2 , the estimates of the covariances matrices of the EEG signals in the two conditions:

$$\mathbf{C}_c = \frac{1}{Tn_c} \mathbf{X}'_c \mathbf{X}_c = \frac{1}{n_c} \sum_{i \in \mathcal{I}_c}^n \frac{1}{T} X'_i X_i = \frac{1}{n_c} \sum_{i \in \mathcal{I}_c}^n C_i$$

²⁵⁵ where $c = 1, 2$, \mathcal{I}_c is the set of indices corresponding to trials belonging in class c and X'_i is the transpose of the matrix X_i . The covariance matrix \mathbf{C}_c is the average of the covariance matrices C_i for trials i in class c . In this work, we will assume that \mathbf{X}_1 and \mathbf{X}_2 are invertible. By this way, \mathbf{C}_1 and \mathbf{C}_2 are definite-positive and invertible.

3.1.1. Search of the first set of spatial filters

In the first step, we search for the filters $\mathbf{w} \in \mathbb{R}^p$ that maximize $\text{var}(\mathbf{X}_1\mathbf{w})$ while $\text{var}(\mathbf{X}_2\mathbf{w})$ is minimal. We have $\text{var}(\mathbf{X}_c\mathbf{w}) = \frac{1}{Tn_c}\mathbf{w}'\mathbf{X}'_c\mathbf{X}_c\mathbf{w} = \mathbf{w}'\mathbf{C}_c\mathbf{w}$

²⁶⁰ for $c = 1, 2$. Thus, the optimisation problem can be written:

$$\max_{\mathbf{w} \in \mathbb{R}^p} \frac{\mathbf{w}' \mathbf{C}_1 \mathbf{w}}{\mathbf{w}' \mathbf{C}_2 \mathbf{w}} \quad (1)$$

According to Proposition 7.1, the solution \mathbf{w}_1 of (1) is the first eigenvector of $\mathbf{C}_2^{-1} \mathbf{C}_1$, i.e. the eigenvector which corresponds to the largest eigenvalue $\lambda_1 = \frac{\mathbf{w}_1' \mathbf{C}_1 \mathbf{w}_1}{\mathbf{w}_1' \mathbf{C}_2 \mathbf{w}_1}$.

²⁶⁵ By considering the k first eigenvectors of $\mathbf{C}_2^{-1} \mathbf{C}_1$ (i.e. which corresponds to the k largest eigenvalues), we obtain a set of k spatial filters $\mathbf{w}_1, \dots, \mathbf{w}_k$ maximizing (1) under the constraints that $\mathbf{w}_j \mathbf{C}_2 \mathbf{w}_l = 0, \forall l \neq j$ (filters are \mathbf{C}_2 orthogonal).

²⁷⁰ 3.1.2. Search of the second set of spatial filters

In the second step, we search for the filters $\mathbf{v} \in \mathbb{R}^p$ that maximize $\text{var}(\mathbf{X}_2 \mathbf{v})$ while $\text{var}(\mathbf{X}_1 \mathbf{v})$ is minimal. The optimisation problem is

$$\max_{\mathbf{v} \in \mathbb{R}^p} \frac{\mathbf{v}' \mathbf{C}_2 \mathbf{v}}{\mathbf{v}' \mathbf{C}_1 \mathbf{v}} \quad (2)$$

Here, \mathbf{v}_1 is the first eigenvector of $\mathbf{C}_1^{-1} \mathbf{C}_2$ which correspond to the largest eigenvalue $\mu_1 = \frac{\mathbf{v}_1' \mathbf{C}_2 \mathbf{v}_1}{\mathbf{v}_1' \mathbf{C}_1 \mathbf{v}_1} = \frac{1}{\lambda_1}$. Symmetrically than for filters \mathbf{w} , the k first eigenvectors of $\mathbf{C}_1^{-1} \mathbf{C}_2$ give a second set of k spatial filters $\mathbf{v}_1, \dots, \mathbf{v}_k$ maximizing (2) under the constraints that $\mathbf{v}_j \mathbf{C}_1 \mathbf{v}_l = 0, \forall l \neq j$.

3.1.3. A single eigenvalue decomposition

We have seen that the two sets of spatial filters $\mathbf{w}_1, \dots, \mathbf{w}_k$ and $\mathbf{v}_1, \dots, \mathbf{v}_k$ are obtained by computing the eigenvalue decomposition of $\mathbf{C}_2^{-1} \mathbf{C}_1$ and $\mathbf{C}_1^{-1} \mathbf{C}_2$. However, the first k eigenvectors of $\mathbf{C}_1^{-1} \mathbf{C}_2$ are also the k last eigenvectors of $\mathbf{C}_2^{-1} \mathbf{C}_1$ (see Proposition 7.2 applied on the product $\mathbf{C}_2^{-1} \mathbf{C}_1$ and Proposition 7.3):

$$\begin{cases} \mathbf{w}_1 = \mathbf{v}_p \\ \vdots \quad \vdots \\ \mathbf{w}_k = \mathbf{v}_{p-k} \end{cases} \quad (3)$$

CSP reduces then to a single eigenvalue decomposition and the filters
 285 are constructed by keeping the k first and the k last eigenvectors of $\mathbf{C}_2^{-1}\mathbf{C}_1$. They form the filter matrix $W = [\mathbf{w}_1, \dots, \mathbf{w}_k, \mathbf{v}_1, \dots, \mathbf{v}_k]$ in which filters are considered by pairs $(\mathbf{w}_j, \mathbf{v}_j)$ for $j = 1, \dots, k$.

Note that the CSP is sometimes presented as the eigenvalue decomposition
 290 of $(\mathbf{C}_1 + \mathbf{C}_2)^{-1} \mathbf{C}_1$. It corresponds to the optimisation of

$$\frac{\mathbf{w}' \mathbf{C}_1 \mathbf{w}}{\mathbf{w}' (\mathbf{C}_1 + \mathbf{C}_2) \mathbf{w}} = \frac{\text{var}(\mathbf{X}_1 \mathbf{w})}{\text{var}(\mathbf{X}_1 \mathbf{w}) + \text{var}(\mathbf{X}_2 \mathbf{w})} \quad (4)$$

According to Proposition 7.4, matrices $\mathbf{C}_2^{-1}\mathbf{C}_1$ and $(\mathbf{C}_1 + \mathbf{C}_2)^{-1} \mathbf{C}_1$ have the same eigenvectors but different eigenvalues. Eigenvalues of $(\mathbf{C}_1 + \mathbf{C}_2)^{-1} \mathbf{C}_1$ are $\beta_j = \frac{\text{var}(\mathbf{X}_1 \mathbf{w}_j)}{\text{var}(\mathbf{X}_1 \mathbf{w}_j) + \text{var}(\mathbf{X}_2 \mathbf{w}_j)}$. This ratio is convenient to interpret.

3.2. Trials classification

295 Once the filter matrix W is obtained by CSP, the EEG data matrices $X_i, i = 1, \dots, n$ of the n trials are linearly projected using W . We obtain n matrices $F_i = X_i W$ of dimension $T \times 2k$. Columns of F_i are the $2k < p$ synthetic signals of the trial i . Each synthetic signal $X_i \mathbf{w}_j$ is summarized by its variance and a logarithm transformation is used to approximate normal
 300 distribution of the data [26]. The n trials are then described on $2k$ dimensions in a matrix $Z = (z_{ij})_{n \times 2k}$ with $z_{ij} = \log(\text{var}(X_i \mathbf{w}_j))$. Finally, a classifier is constructed with Z to predict the class of each trial.

3.3. Data pre-processing

3.3.1. Data band-pass filtering

305 The electrical activity of the brain is divided into different oscillatory rhythms characterized by their frequency bands. The main rhythms in ascending order of frequency are δ (1-3.5 Hz), θ (4-8 Hz), α (8-12 Hz) and β (19-26 Hz). Each of these rhythms is linked to psycho-physiological states. The α waves are characteristic of a diffuse awake state for healthy subjects
 310 and can be used to discern the normal awake and relaxed states, which is the topic of this experimental study. The oscillatory α rhythm appears as visually observable puffs on the electroencephalogram (EEG) (see the EEG signal framed in Figure 9), especially over the occipital brain areas at the back of the skull, but also under certain conditions in more frontal recording sites.
 315

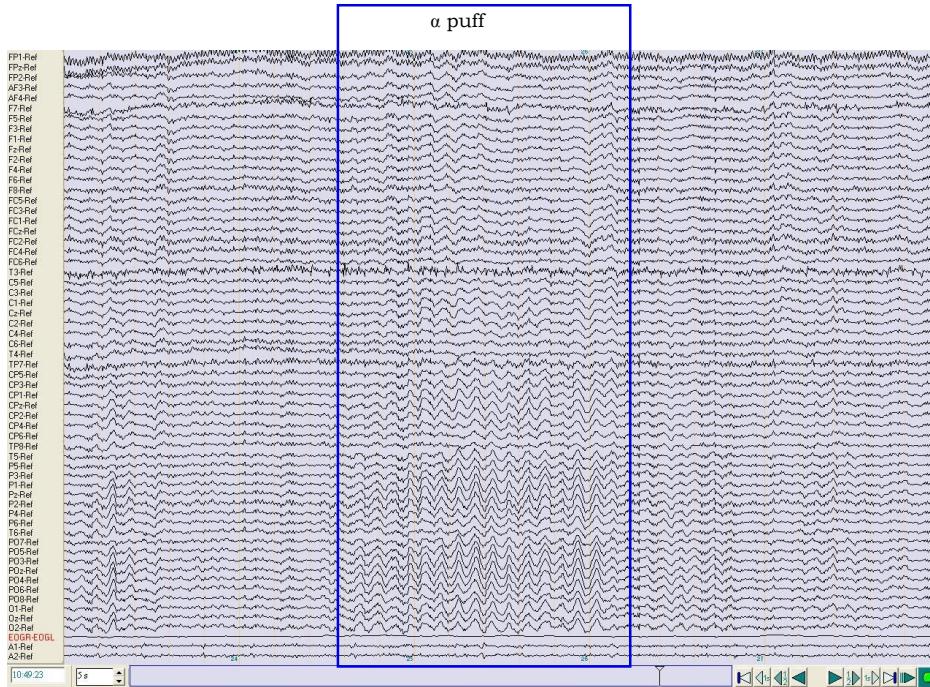


Figure 9: Representation of the EEG of a subject in a low alertness state, acquired by 58 electrodes positioned according to the international 10/10 system. The EEG signal framed represents an α puff recorded.

Before using the CSP method, EEG signals are always band-pass filtered to extract only the most informative frequencies. These frequencies depend on the classification task considered. Even if band-pass filtered signals are already almost centered as mentioned in [11], each X_i matrices must be centered as detailed in Section 3.1.

In [10], a CSP is used to extract synthetic signals and then classify trials recorded using 56 electrodes during left index finger, right index finger or toe movements. Six different filtering strategies are studied using knowledge of the main EEG rhythms. Data is band-pass filtered considering only the α rhythm (8-12 Hz), the lower or upper α (8-10 Hz or 10-12 Hz), the β rhythm (19-26 Hz), the γ rhythm (38-42 Hz) and a broad band 8-30 Hz. On those EEG records, the filtering strategy that gives the best correct classification rate (CCR), computed using a 50-fold cross validation, is the 8-30 Hz band-

330 pass filter. These results can be explained by the fact that these frequencies cover both α and β rhythms, which are the most important rhythms for movement classification [9].

335 Using these results, the 8-30 Hz frequency band is often used in other works without more explorations even if the data is different [26], [16], [11].
 In other works, other frequency ranges are used without justifications (1.5-35 Hz in [8], or 8-35 Hz in [27]).

340 The frequency range proposed in [10] concerns classification of movements. Since we assume that the frequency range depends on the classification task and the data, an approach to tune the frequency range for the band-pass filter is used in this work. In particular, a 5th order Butterworth filter is used to band-pass filtered each X_i trial.

3.3.2. Data normalization

345 In what follows, we will assume that each EEG data matrix X_i is band-pass filtered and centered. In [8], Koles et al. suggest to use for a trial i the normalized covariance matrix defined by

$$\tilde{C}_i = \frac{X'_i X_i}{\text{trace}(X'_i X_i)} = \frac{C_i}{\text{trace}(C_i)} \quad (5)$$

350 where $\text{trace}(C_i)$ is the inertia of X_i . This normalization is used to eliminate magnitude variations in EEGs between trials. This scaling strategy is widely used in other CSP applications [10], [26], [27].

Note that \tilde{C}_i can also be obtained by normalizing the EEG data matrix X_i rather than its covariance matrix C_i . The normalized data matrix \tilde{X}_i is:

$$\tilde{X}_i = \frac{X_i}{\sqrt{\text{trace}(C_i)}}.$$

In that way, \tilde{C}_i is defined as the covariance matrix of the normalized data and

$$\tilde{C}_i = \frac{1}{T} \tilde{X}'_i \tilde{X}_i. \quad (6)$$

355 This distinction between (5) and (6) is important in practice because in CSP, covariance matrices have to be calculated with the data that will be

projected (see Section 3.3.3).

In this work, we propose another normalization strategy which is classical in PCA. It consists in scaling each column of X_i by its standard deviation:

$$\tilde{X}_i = X_i D, \text{ where } D = \begin{pmatrix} \frac{1}{\sigma_1} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{\sigma_p} \end{pmatrix},$$

360 where σ_j , $j = 1, \dots, p$, is the standard deviation of the j^{th} column of X_i . By this way, $\tilde{C}_i = \frac{1}{T} \tilde{X}'_i \tilde{X}_i$ is a correlation matrix instead of a covariance matrix.

In this work, the following three strategies have been compared on our data set:

- 365
- $\tilde{X}_i = X_i$ i.e. the data is not scaled.
 - $\tilde{X}_i = \frac{X_i}{\text{trace}(C_i)}$ i.e. the data is scaled using the inertia of X_i .
 - $\tilde{X}_i = X_i D$ i.e. the data is scaled using the standard deviations.

3.3.3. Data projection

370 In CSP, normalization of covariance matrices C_i or normalization of EEG data matrices X_i , lead to the same filter matrix W . However, even if it is not explicitly specified, it is the normalized data that is projected in this CSP step. For that reason, it is important in the classification step to project the normalized data and not the original data. In other words, in this step 375 we use $F_i = \tilde{X}_i W$ as first proposed in [8] and not $F_i = X_i W$ as used in many works (in [26] for instance). This is precisely the reason why, we made the distinction between normalizing the data and normalizing the covariance matrices in Section 3.3.2.

380 *3.3.4. Number of pairs of spatial filter*

The main parameter of the CSP method is k , which represents the number of pairs of spatial filters used to build W . In [10], the value of this parameter is tested between 1 and 5. In that work, the best CCRs are obtained using

$k = 2$ and $k = 3$. In this work, the k value will be tuned on our data to find
385 the best value based on the CCRs.

3.4. Parameter tuning

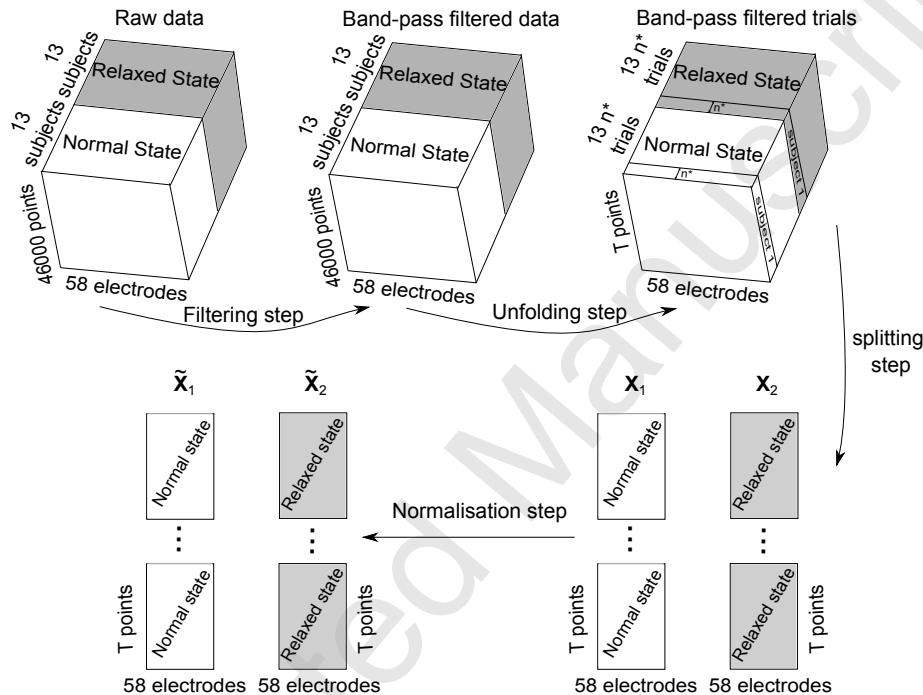


Figure 10: Diagram representing the data pre-processing step. The raw data is filtered between a and b Hz and split into trials of length T . A normalization of the data is applied to obtain $\tilde{\mathbf{X}}_1$ and $\tilde{\mathbf{X}}_2$.

To perform the CSP, the data will be pre-processed as detailed in Figure 10. The data is band-pass filtered between a and b Hz where $a < b$. Then, the data is split into trials of length T . Finally, a normalization of the data is applied on X_i to obtain normalized EEG data matrices $\tilde{\mathbf{X}}_i$. The normalized EEG data matrices of trials in class 1 (respectively class 2) are row concatenated in the matrix $\tilde{\mathbf{X}}_1$ of dimension $Tn_1 \times p$ (resp. $\tilde{\mathbf{X}}_2$). Thus, the same configuration than in 3.1 is obtained. Therefore, n^* data EEG matrices are available for each subject in each state of alertness, with $n^* = \lfloor 46000/T \rfloor$ and where $\lfloor . \rfloor$ is the integer part.
390
395

In what follows, the method to estimate the CCR will be detailed and parameters a , b , T , k and the normalization strategy will be tuned on our

data.

400 *3.4.1. Leave One Out (LOO) estimation of the CCR*

To estimate the correct classification rate obtained using the CSP, a LOO strategy is used. This LOO is adapted to our case. In fact, this LOO is performed on the subject level instead of the trials level. In other words, at each iteration, all the trials of a subject (trials in class 1 and 2) are removed 405 from the data set. The data set without these trials forms the learning set and the removed trials form the test set. The m^{th} iteration of the LOO estimation is represented in Figure 11.

In our work, the LOO can not be performed at the trials level because when a trial is removed from the learning set to form the test set, other 410 trials from the same subject remain in the learning set. This could lead to distorted results.

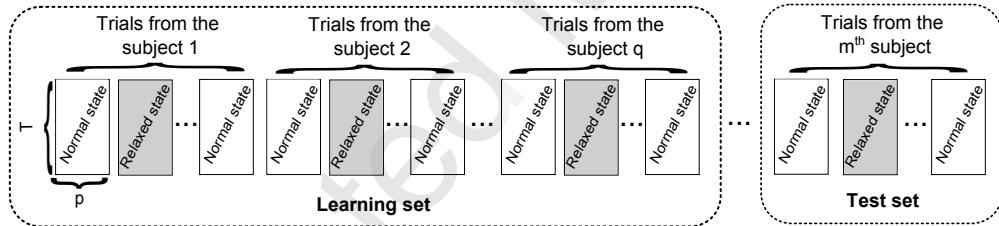


Figure 11: Learning and test set creation in the m^{th} iteration of the LOO estimation of the CCR (with $m = 1, \dots, q$).

Thus, the LOO validation has q iterations (where q is the number of subjects). At each iteration, a CSP is performed on trials from the learning 415 set to create the k pairs of CSP filters. Then, learning trials are projected using these filters and a log var transformation is applied to obtain the Z matrix. This matrix Z is given as an input to a linear discriminant classifier (LDA) to construct a decision rule. The LDA method [28] is widely used to solve multiclass supervised classification problems. It consists in constructing 420 hyperplanes to separate the data that belong to different classes into different regions of \mathbb{R}^p . A classification of a new individual will be obtained by determining to which region of \mathbb{R}^p it belongs.

Once the decision rule is obtained, the filters are used to project trials from the test set. After the log var transformation, the class of each trial
425 is predicted using the decision rule constructed on trials from the learning set.

When the q iterations are done, we obtain a vector of length n that contains the predicted class for each trial. This vector is compared to true labels to obtain the CCR.
430

3.4.2. Tuning results

In this part, the a , b , k and T values and the normalization strategy will be tuned using a LOO over subjects of the first campaign.

An exhaustive search of the best combination of these parameters seems
435 difficult. That is why only some values of these parameters will be tested. In particular, a and b values will be tuned one after the other. Indeed, even if a joint optimisation of these parameter gives better results, this approach is too costly in computation time. Thus, we will choose the best combination of parameter values over all the combination tested.
440

To begin, b will be set to 30 Hz. The a value will vary between 1 and 25 Hz. Three values will be tested for the T value: 1024, 2048 and 4096 points (corresponding respectively to trials of 4, 8 and 16 seconds length). The k value will vary between 1, 2 and 3. Data will be band-pass filtered between
445 a and 30 Hz. Then, the data will be split into trials of length T . A CCR will be obtain as explained in Section 3.4.1. The best CCR obtained will allow us to choose a , k and T and the best normalization strategy. Thus, these parameters will be fixed. Then, the b value will be tuned between $a+3$ and 40 Hz. Results are presented in Figures 12, 14, 16 and 18. For a better readability, Figures 12, 14 and 16 are also splitted by k value respectively in Figures 13, 15 and 17.
450

To begin, a search for the best normalization strategy is done by comparing Figures 12 (data not scaled), 14 (data scaled using the inertia of X_i) and
455 16 (data scaled using the standard deviations). In these figures, The data is split into trials of length 4 (circles), 8 (squares) or 16 (triangles) seconds (corresponding respectively to $T = 1024$, $T = 2048$ and $T = 4096$ points). The number of pairs of CSP filters, denoted by k , varies between 1 (dotted line), 2 (solid line) and 3 (dash-dotted line). Data is band-pass filtered be-

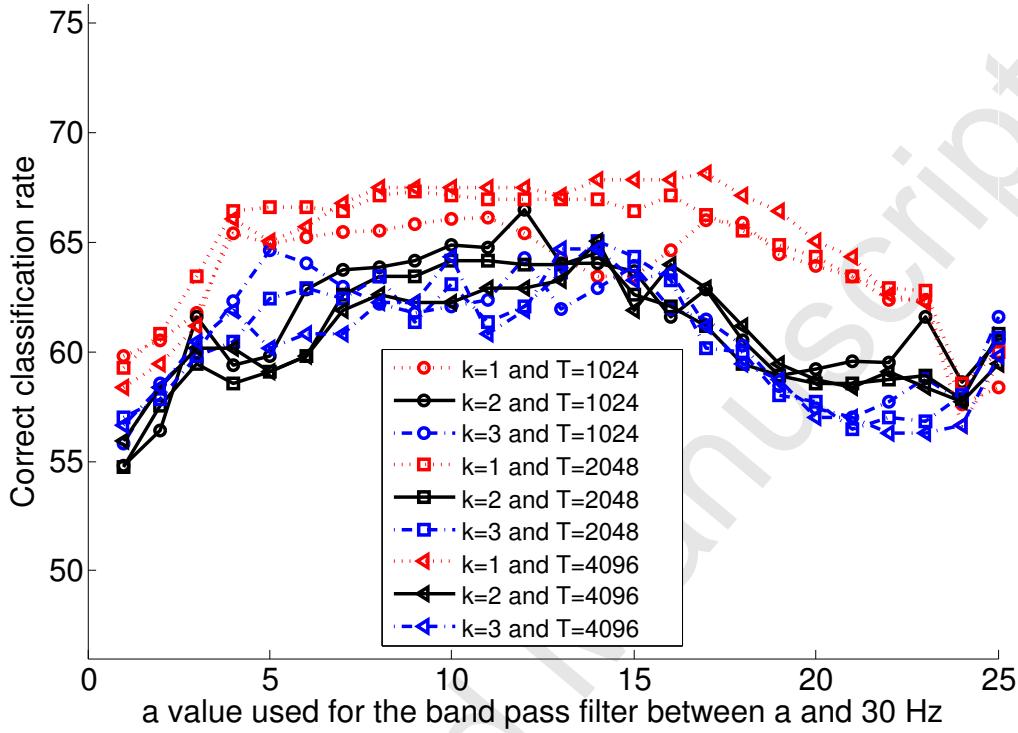


Figure 12: Correct classification rates (CCR) estimated by LOO on data from the first acquisition campaign for different filtering strategies (data band-pass filtered between a and 30 Hz with $a \in [1, 2, \dots, 25]$). **The data is not scaled** and split into trials of 4 (circles), 8 (squares) or 16 (triangles) seconds (corresponding respectively to $T = 1024$, $T = 2048$ and $T = 4096$ points). A CSP is performed using $k = 1$ (dotted line), $k = 2$ (solid line) or $k = 3$ (dash-dotted line) pairs of filters.

- 460 tween a (abscissa) and 30 Hz. For each strategy, a CCR is obtained. The parameter values which exhibit the highest CCR are desired.
 By comparing Figures 12, 14 and 16, it seems that the CCR obtained by scaling the data using the standard deviations (Figure 16) outperform the CCR obtained with the other normalization strategies (Figures 12 and 14).
 465 Thus, we will definitively use this normalization strategy on our data and we will focus on the results shown in Figure 16. It appears that best results are obtained by using 3 pairs of CSP filters and by considering trials of 8 seconds length ($T = 2048$). Moreover, the best result is obtained by using $a = 7$.

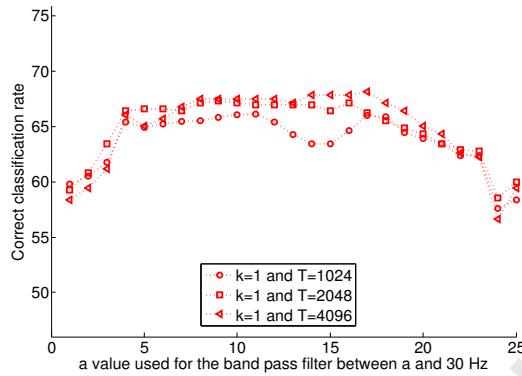
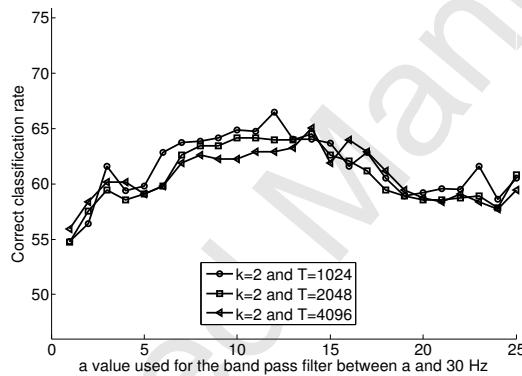
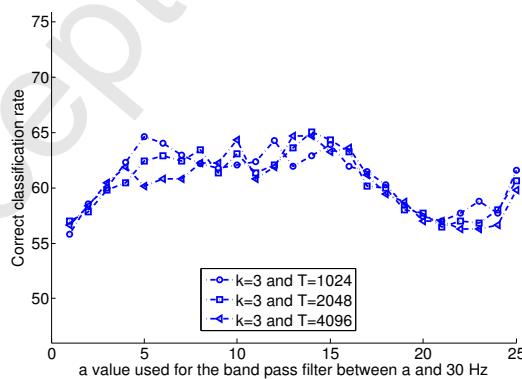
(a) $k = 1$ (b) $k = 2$ (c) $k = 3$

Figure 13: These plots correspond with Figure 12, splitted by the number of pairs of CSP filters (k value). Plots for $k = 1$, $k = 2$ and $k = 3$ are represented in (a), (b) and (c).

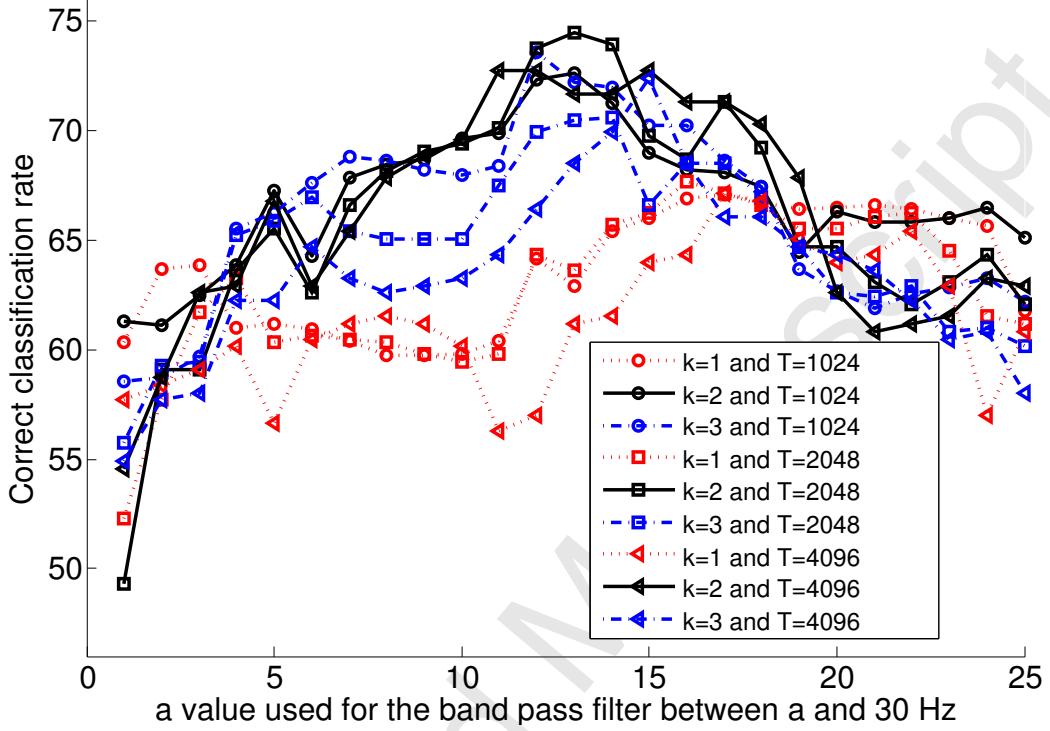


Figure 14: Correct classification rates estimated by LOO on data from the first acquisition campaign for different filtering strategy (data band-pass filtered between a and 30 Hz with a integer between 1 and 25). **The data is scaled using the inertia of X_i** and split into trials of 4 (circles), 8 (squares) or 16 (triangles) seconds length (corresponding respectively to $T = 1024$, $T = 2048$ and $T = 4096$ points). A CSP is performed using $k = 1$ (dotted line), $k = 2$ (solid line) or $k = 3$ (dash-dotted line) pairs of CSP filters.

Afterwards, the b value is tuned between 10 and 40 Hz. The value of b which allows us to obtain the highest CCR is desired. Thus, the data is band-pass filtered between 7 and b Hz and split into trials of 8 seconds length. Then, data is normalized using the standard deviation and a CSP with 3 pairs of filters is used. Figure 18 represents the CCR obtained as a function of b . By analysing Figure 18, it appears that the highest CCR (equal to 75.35%) is obtained by setting $b = 30$ Hz.

To sum up, the best combination of parameter values found during these experiments is:

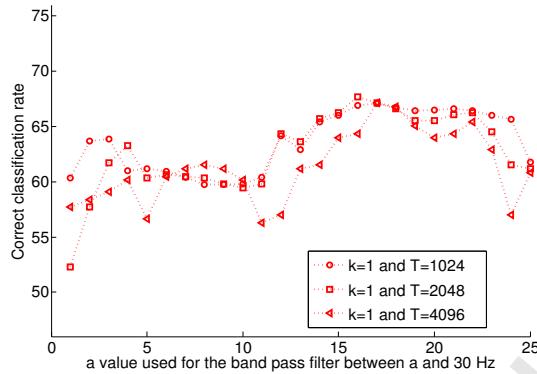
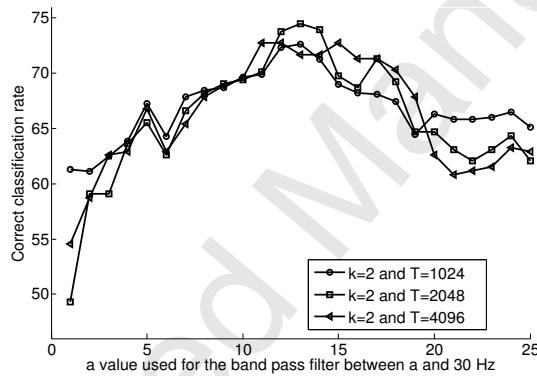
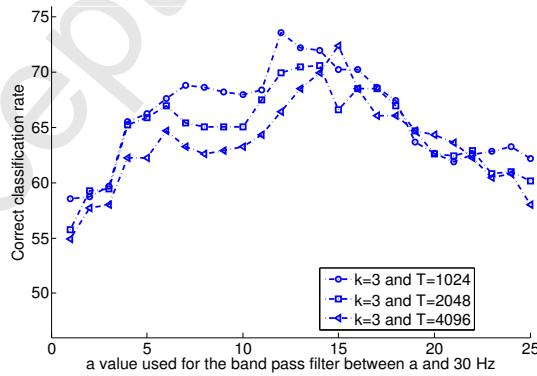
(a) $k = 1$ (b) $k = 2$ (c) $k = 3$

Figure 15: These plots correspond with Figure 14, splitted by the number of pairs of CSP filters (k value). Plots for $k = 1$, $k = 2$ and $k = 3$ are represented in (a), (b) and (c).

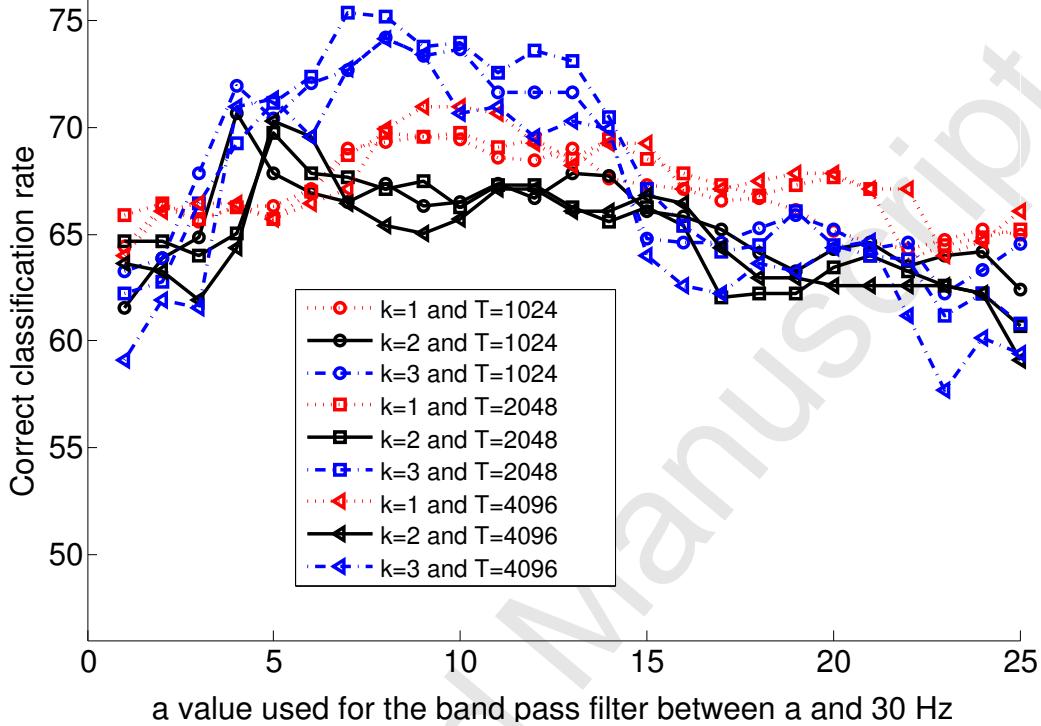


Figure 16: Correct classification rates (CCR) estimated by LOO on data from the first acquisition campaign for different filtering strategies (data band-pass filtered between a and 30 Hz with $a \in [1, 2, \dots, 25]$). **The data is scaled using the standard deviations** and split into trials of 4 (circles), 8 (squares) or 16 (triangles) seconds (corresponding respectively to $T = 1024$, $T = 2048$ and $T = 4096$ points). A CSP is performed using $k = 1$ (dotted line), $k = 2$ (solid line) or $k = 3$ (dash-dotted line) pairs of filters.

- 480
- Band-pass filter: $[7, 30]$ Hz.
 - Trial length: 8 seconds ($T = 2048$ points). Thus, each signal will be split into $n^* = 22$ trials ($\lfloor \frac{46000}{2048} \rfloor = 22$).
 - Normalization: $\tilde{X}_i = X_i D$.
 - Number of CSP filters pairs: 3.

485 Values of a , b and k seem coherent with previous work (see[10]). We have seen that the CSP method coupled to a LDA classifier allows us to

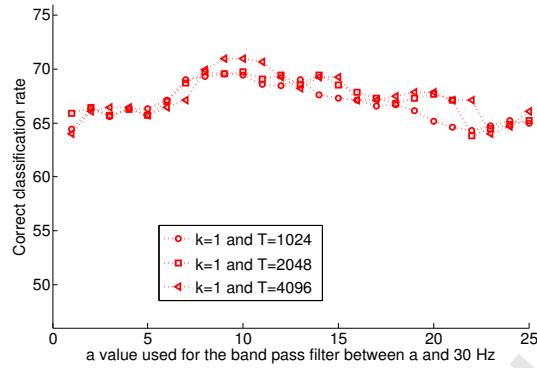
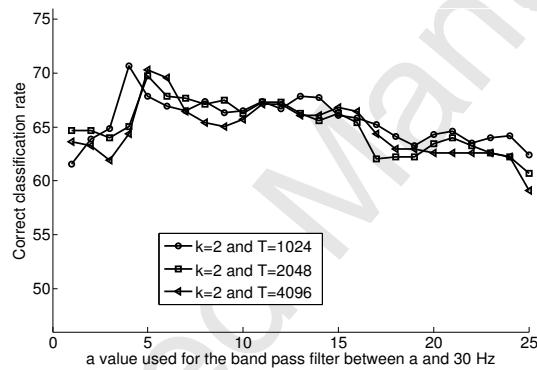
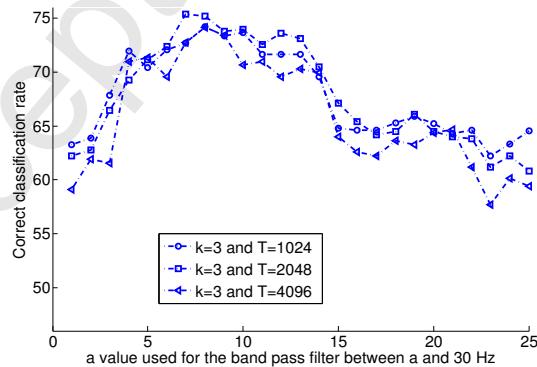
(a) $k = 1$ (b) $k = 2$ (c) $k = 3$

Figure 17: These plots correspond with Figure 16, splitted by the number of pairs of CSP filters (k value). Plots for $k = 1$, $k = 2$ and $k = 3$ are represented in (a), (b) and (c).

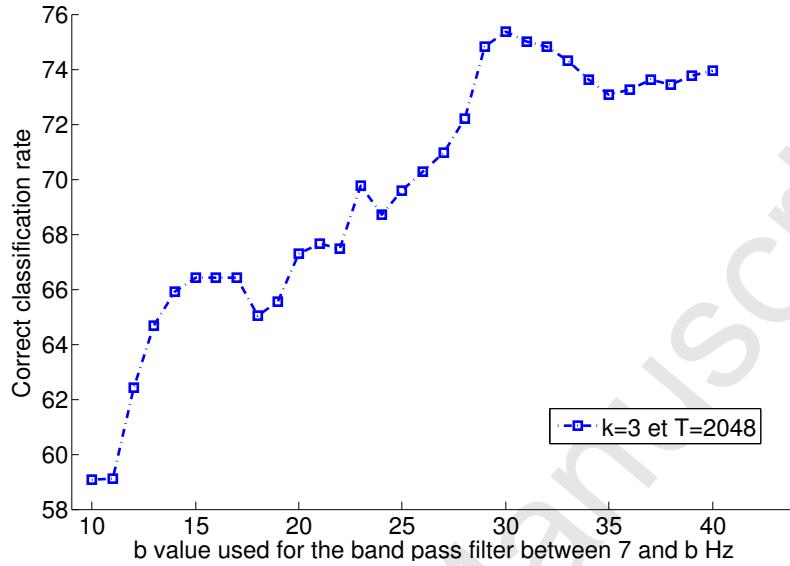


Figure 18: Correct classification rates (CCR) estimated by LOO on data from the first acquisition campaign for different filtering strategies (data band-pass filtered between 7 and b Hz with $b \in [10, 11, \dots, 40]$). **The data is scaled using the standard deviations** and split into trials of 8 seconds (corresponding to $T = 2048$ points). A CSP is performed using 3 pairs of filters.

obtain a good CCR on our data by using all of the variables (75.35% on subjects from the first data acquisition campaign). We will now describe an approach to select a subset of variables to predict the alertness state of an
490 unseen subject.

4. Electrode selection using a genetic algorithm

The goal of our approach is to find a subset of p' electrodes (with $p' < p$) which allows us to obtain an accurate prediction for the alertness state of a new subject. We consider Ω the set of all the combinations of p' electrodes
495 taken among p . The cardinality of Ω is equal to $|\Omega| = C_p^{p'}$. In this work, $p = 58$, thus $|\Omega|$ grows very fast with respect to p' , making an exhaustive search infeasible. For this reason, the proposal in this work is to use a genetic algorithm (GA) to explore the search space defined by Ω .

4.1. General principle of a genetic algorithm

500 These optimization algorithms [29] [30] are based on a simplified abstraction
 505 of the Neo-Darwinian evolution theory of evolution. The general idea is
 that a population of potential solutions will improve its characteristics over
 time, through a series of basic genetic operations called selection, mutation
 and genetic recombination or crossing. The general principle is depicted in
 Figure 19.

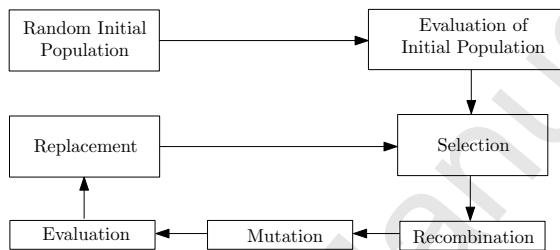


Figure 19: Evolutionary loop of a basic Genetic Algorithm.

The purpose of these algorithms is to optimize a function (fitness) within a given search space of candidate solutions. Solutions (called individuals) correspond to points within the search space, a random set of which are generated, seeding the algorithm with an initial population (set of individuals).
 510 Solutions are represented by their genomes (binary codes or real numbers, with a fixed or variable size). All individuals are evaluated using a problem specific objective function called fitness. Individuals are selected based on their fitness (using a series of tournaments), these selected individuals are called parents. These parents are used to generate new individuals using two
 515 basic genetic (search) operations, recombination (random recombination of two or more individuals) and mutation (random modification of a single individual). These newly generated individuals are called offspring, since they share (genetic) similarities with the parents that were used to generate them. Finally, the best individuals (amongst parents and offspring) are selected and
 520 replace the initial population. The algorithm is iterated until a stop criterion is reached; for instance, when all individuals are identical (convergence of the algorithm) or after a pre-specified number of iterations.

4.2. Genome

In this work, the genome is composed of $p = 58$ binary variables which
 525 correspond to the inclusion (component of the genome equal to 1) or not

(component equal to 0) of each electrode to compute the CSP. Then, each genome defines the electrodes on which the CSP will be performed. We chose to set the number of selected electrodes p' (input parameter of the genetic algorithm) for several reasons. The genetic algorithm is used with an aim 530 of reducing the dimension. A compromise between the number of selected electrodes in the solution and the quality of the solution is required. Thus, we want to see the evolution of the quality of prediction as a function of p' . In addition, we want to obtain feedback from neurobiologists to contribute by bringing their expertise after seeing the results for different values of p' 535 (quality of obtained prediction, selected electrodes). It is also possible that the objective changes and a smaller number of electrodes is desired. Thus we prefer to determine the value of p' after analysing the prediction quality obtained for different values of p' . That is why the fitness function of the genetic algorithm in our work does not depend on p' .

540 *4.3. Genetic operators*

To create an offspring, 2 parents are randomly selected. A tournament is performed to keep only the best individual (the one with the highest rating based on fitness). The selection pressure is not high (tournament of size 2) to maintain a high diversity in the population. The selection and the 545 tournament is repeated twice in order to select two parents (tournament “winners”). Both parents are crossed to create an offspring. The crossing is done using a logical operator given in Table 1. This crossover is used to balance the production of 1s and 0s. For a given electrode, when the two parents share the same value (1 for a selected electrode or 0 for a non-selected 550 electrode), the child inherits this value. When the two parents do not share the same value, then a Bernoulli distribution is used to build the component of the children. Once the offspring is established, a mutation is applied. A pair of components is selected and binary values of these components are 555 switched (the 0 becomes 1 and vice versa). In this way, the number of selected electrodes, p' , is kept constant in the algorithm. Thus, each component of the genome of the offspring mutates with probability $2/p$.

4.4. Evaluation function

Each genome will be evaluated using a CSP method coupled to a LDA classifier. To perform this CSP, data will be pre-processed as detailed in 3.4. 560 Thus, a LOO estimation of the CCR is obtained for each genome as described

Table 1: Logical operator used for the crossing step.

Parent 1	Parent 2	Offspring
0	0	0
1	0	$Bern(\frac{1}{2})$
0	1	$Bern(\frac{1}{2})$
1	1	1

in 3.4.1. An iteration of the LOO estimation of the CCR for a given genome is represented in Figure 20.

The CCR obtained can be related to the quality of the genome. It seems natural that the fitness function corresponds to the CCR obtained for each genome. Thus, the fitness function for each genome \mathbf{u} is given by

$$g(\mathbf{u}) = \frac{\# \text{ of well classified trials}}{\# \text{ of trials}}.$$

The genetic algorithm searches for the genome which maximizes g .

4.5. Stop criterion

The algorithm stops if one of the following four conditions is satisfied:

- The number of iterations exceeds 1000.
- Parents are the same for 10 generations.
- The number of differences among the parents is less than 3.
- The best evaluation is the same for 50 generations.

5. Experimental application

The GA and the predictive classifier were written in MATLAB running on a Linux with 8 * 2.8 GHz Intel Xeon processors X5560 and 32 GB 1333 MHz DDR3 ECC RAM.

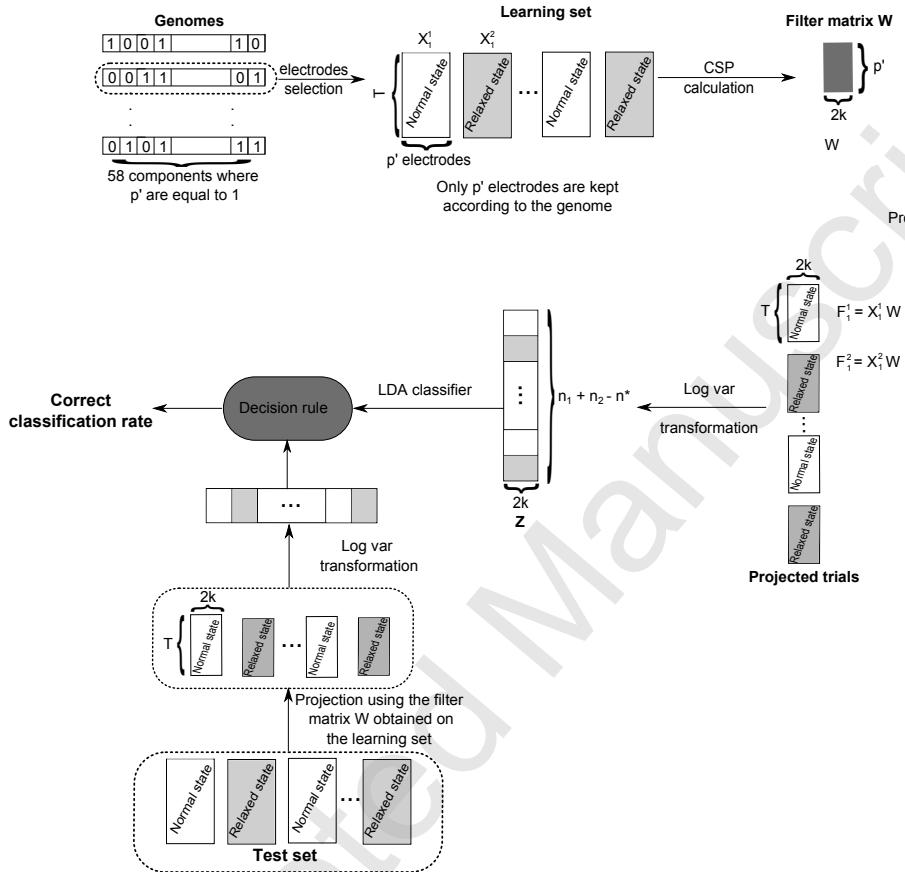


Figure 20: Diagram representing the evaluation step in the genetic algorithm. For a given genome, a subset of electrodes is selected and a CSP is apply to construct the filter matrix W . Then, the trials are projected using W and a log var transformation is applied to obtain Z . Finally, a decision rule is learned using a LDA classifier and a CCR is obtained on the test set.

5.1. Experimental strategy

The genetic algorithm is run for $p' = 1, 2, \dots, 15, 20, 30, 40$ and 50 with 150 parents and 50 offspring. Each run of the genetic algorithm takes between 9 and 15 hours depending on the p' value. Therefore, we run the GA only 15 times for each p' .

When the GA converges, the best genome found by the algorithm is kept.

Then, the 15 best genomes and their associated CCR will be obtained for each value of p' . Given these CCR results, the best compromise between the size of the subset of variables (p' value) and the quality of prediction obtained (CCR) will be chosen. The best genome found over all the 15 runs for the chosen p' value will be kept which gives the subset of chosen electrodes (components with 1 in the genome). A CSP will be run using this subset of electrodes on the trials from the 13 subjects in order to construct 3 pairs of CSP filters ($k = 3$) and a decision rule (LDA classifier) will be constructed. Then, these LDA classifier and CSP filters learned on the training set are used on the test set (formed by the trials of the 6 unseen subjects from the second data acquisition campaign) to calculate an external CCR.

595 5.2. Results

Performances of the GA, the choice of the p' value and selected electrodes are discussed in the followings paragraphs. Finally, a description of how our approach will be used in future applications is given.

600 5.2.1. Performances of the method

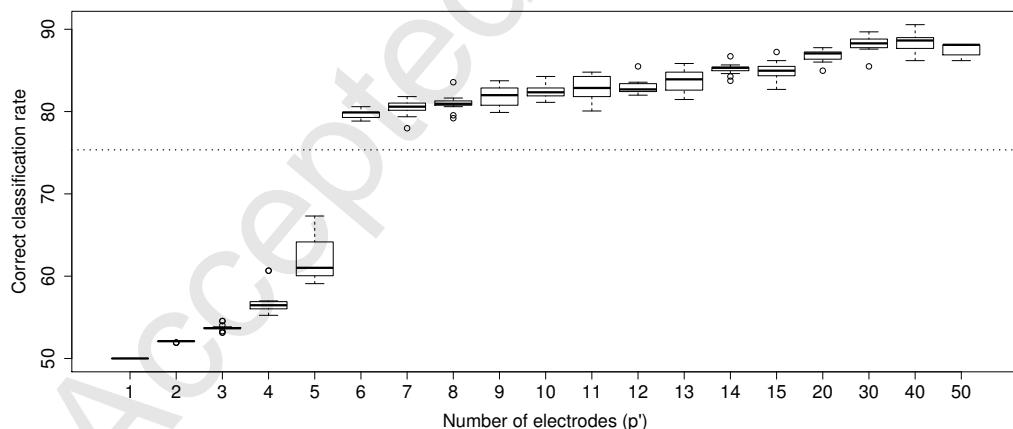


Figure 21; Correct classification rates obtained on the training set with the genetic algorithm for different values of p' . The dotted line represents the CCR of 75.35% obtained using the 58 initial variables for the CSP calculation (no variable selection).

Figure 21 represents, for each value of p' , a boxplot of the 15 CCR (one per run) associated to best genomes found by the GA. The dotted line represents the CCR obtained using the 58 initial variables for the CSP calculation

(no variable selection). Because the first aim of this work is to select a small
 605 subset of variables (electrodes), we will focus on results obtained for p' between 1 and 15. Results for $p' = 20, 30, 40$ and 50 are just given for assessing the general behaviour of our approach.

In Figure 21, it appears that the median CCR globally increases between
 610 $p' = 1$ and $p' = 50$ (except for $p' = 15$ and 50). An important increase occurs between $p' = 5$ and $p' = 6$ where the CCR reaches the 80% of accuracy. Thus, from $p' = 6$ to $p' = 15, 20, 30, 40$ and 50, the CCR obtained by this approach outperforms the CCR obtained by using the CSP on the 58 initial electrodes (dotted line). Results start to decrease between $p' = 40$ and $p' = 50$ to reach the dotted line in $p' = 58$. Thus, an important decrease will occur between
 615 $p' = 50$ and $p' = 58$. It means that, among the eight variables not included, some variables are very noisy and deteriorate the prediction quality. From $p' = 6$ to 15, the CCR obtained by the GA seems quite satisfactory. It allows us to consider the use of this approach in future applications.

In these applications, the aim is to predict the alertness of an unseen
 620 individual. That is why, it is important to assess how the method performs at the subject level. In this work, the signal of a subject in a given alertness state is split into 22 trials. In the GA, for a given genome, a prediction is obtained for each trials. Thus, to predict the alertness state of each subject in a given state, a majority affectation is applied on his trials prediction. As
 625 we have an even number of trials, we remove the first trial of each subject to perform the majority affectation.

Results obtained are synthesized in Table 2. For the different values of p' , Table 2 contains the mean of the number of iterations before the GA convergence. It also contains the mean and the standard deviation of CCR obtained in the GA. Results between parenthesis are obtained by considering the subject level (majority affectation) instead of considering trials level. It appears that the mean of the number of iterations does not exceed 170. The algorithm seems to converge rapidly and the number of iterations never reaches 1000 (the first stop criterion presented in Subsection 4.5 is never
 630 used). In most cases, the GA stops because the parents are the same for 10 generations or because the best evaluation is the same for 50 generations. An important diversity is kept in genomes so the number of differences among the parents is rarely less than 3. The means of the CCR obtained by considering trials are quite close to medians observed in Figures 21. During the
 635 GA, the average of the CCR, obtained considering the trials, increase with p' until $p' = 14$. When $p' > 5$, CCR mean is about 80% with small standard
 640 deviation.

Table 2: Results obtained by LOO in the genetic algorithm for different values of p' . Results between parenthesis are obtained by considering subjects instead of considering trials. The last line corresponds to reference CCR value obtain by keeping all the electrodes.

Inside the GA: LOO estimation on campaign 1			
p'	Average of the number of iterations	Average of CCR of best genomes	Standard deviation of CCR of best genomes
1	12	50 (50)	0 (0)
2	55.93	52.06 (53.85)	0.07 (0)
3	71.07	53.73 (53.85)	0.41 (0)
4	78.6	56.85 (58.72)	1.66 (0.023)
5	87.07	62.05 (65.13)	2.68 (0.057)
6	111.6	79.72 (86.41)	0.59 (0.038)
7	108.2	80.43 (86.67)	0.96 (0.025)
8	130.53	81 (86.41)	0.97 (0.035)
9	121.2	81.89 (90.26)	1.28 (0.029)
10	141.47	82.47 (89.74)	0.93 (0.028)
11	137.87	82.84 (91.03)	1.59 (0.035)
12	141.67	82.96 (90.77)	0.87 (0.032)
13	146	83.8 (90.51)	1.36 (0.032)
14	169.07	85.19 (92.05)	0.68 (0.027)
15	163.67	84.93 (92.56)	1.19 (0.031)
20	165.73	86.75 (94.1)	0.72 (0.025)
30	168.47	88.22 (95.13)	0.99 (0.027)
40	168	88.41 (96.67)	1.29 (0.029)
50	145.8	87.49 (92.56)	0.81 (0.01)
58	-	75.35 (73.08)	-

deviation values.

Results obtained for the different values of p' must be compared to the
645 result obtained without variable selection (last line in Table 2). In fact, when no variable selection is considered, the CCR obtained by LOO over the 13 subjects of the first campaign acquisition data is equal to 75.35%. Then, results obtained performing a dimension reduction (using the GA) exceed 75.35% as soon as $p' > 5$.

650

By analysing Table 2 and Figure 21, results seem to level off between $p' = 9$ and $p' = 12$. A statistical test is used to determine if these p' values give results that are significantly different. Given that CCR results are not following Gaussian distributions (Shapiro Wilk test with a significance level equal to 0.05), the Kruskal-Wallis one-way analysis of variance test (KW) is used. In this non-parametric test, the null hypothesis corresponds to " \mathcal{H}_0 : There is no difference between the CCR of the group of p' values tested" and the alternative hypothesis is " \mathcal{H}_1 : The group of p' values have different CCR results". When the KW test is applied, a p -value is obtained. When it is superior to the significance level, it leads to accept \mathcal{H}_0 . In this work, the significance level is set to 0.05. According to the KW test, there are no difference between results of $p' = 9, 10, 11$ and 12 (p -value equal to 0.116). However, according to the KW test, a difference exists between CCR results of $p' = 8, 9, 10, 11$ and 12 (p -value equal to $1.6 * 10^{-4}$) or CCR results of $p' = 9, 10, 11, 12$ and 13 (p -value equal to $7.31 * 10^{-3}$).

5.2.2. Choice of the p' value

The first aim of our work is to find a small subset that allows us to obtain a good prediction for the alertness of a subject (CCR). It is necessary to choose the best compromise between the subset size (p' value) and the quality of the obtained prediction (CCR). p' value choice must be done considering only the results obtained during the GA search. In Figure 21 and Table 2, three values of p' seem to represent a good compromise:

- An important increase occurs between $p' = 5$ and $p' = 6$. From $p' = 6$, result are about 80%. Thus, the smallest value of p' that gives satisfactory result is $p' = 6$.
- In Figure 21, it appears that results keep improving between $p' = 6$ and $p' = 14$. However, according to the KW test, there is no difference between the CCR results of $p' = 9, 10, 11$ and 12 . Thus, the choice $p' = 9$ can represent a good compromise.
- The best result obtained between $p' = 1$ and 15 is reached for $p' = 14$.

Thus, the p' choice depend on the goal of the study. In our case, the best compromise seems to be $p' = 9$. This value is a compromise between a

satisfactory model with a small number of electrode ($p' = 6$) and the more
685 accurate model with a costly number of electrodes. That is why, we choose to consider $p' = 9$ as the definitive value of p' (bold line in the table 2). Only the 15 best genomes found by the GA (one by run) for $p' = 9$ are tested on the external test set. On the trial level, the mean of the 15 CCR obtained on the external test set is 76.29% with a standard deviation of 4.35. The
690 best genome found over the 15 runs of the GA for $p' = 9$ is kept. It allows to obtain an external CCR of 71.59% on the trials and 75% on the subjects.

5.2.3. Selected electrodes

For each value of p' , we obtain 15 subsets of variables of length p' (1
695 per run of the GA) corresponding to the best genomes found by the GA. Figure 22 represents the evolution of these subsets of electrodes for different values of p' . Graph 22(a) shows the occurrence and the localization of the selected electrodes in the best subset found by the GA for $p' = 3$. Thus, for example an electrode which was selected in each best subset by the GA
700 for a given p' value will have a score equal to 15 and will appear in black. Similar graphs are proposed in 22(b), 22(c), 22(g) and 22(h) for respectively $p' = 6, 9, 12$ and 15 . By comparing the figures 22(b) and 22(c), it appears that the selected electrodes (which appears in black) for $p' = 6$ are kept for
705 $p' = 9$. Same conclusions can be observed for $p' = 12$ and 15 . However, for $p' = 3$, result are less stable. The most selected electrode is chosen 5 times over the 15 runs. This instability is explained by the fact that several subset of 3 electrodes give the same results. Thus, an electrode which is often selected for a given value of p' is usually kept for higher values of p' . This result is very important because each run of the GA is independent of the
710 other runs. In particular, the initialisation of the GA does not use the results of another run. Thus, this result shows that signals of the electrodes selected by the GA contain useful informations for the classification task. Moreover, for $p' = 6, 9, 12$ and 15 , the electrode subset seems very stable. In fact, most of selected electrodes appear in black.

715 Figure 22(d) is a 3D plot which represents each best subset of variable in function of p' . For a given electrode and a given run in 22(d), a value equal to 1 means that this electrode was selected in this run. Once again, similar graphs are proposed in 22(e), 22(f), 22(i) and 22(j) for respectively
720 $p' = 6, 9, 12$ and 15 . These figures allows us to visualize that the selected

electrodes subsets are very stable over the different runs for a given value of p' .

For a given value of p' and a given electrode, Figure 23 represents the number of selection of this electrode in best subsets obtained by the GA. It shows that when an electrode is selected for a given value of p' , this electrode is kept for higher values of p' . Moreover, for a given value of p' , peaks that exceed 10 selections are obtained. Then, for a given p' value, subsets seems stable.

Figure 24 represents the subset of selected electrodes given by the best genome obtained in the best run for $p' = 9$. This subset contains pre-frontal (FPz and FP2), frontal (F3, F5 and F7), fronto-central (FC5), temporal (T3), central (C5) and temporal-parietal (TP7) electrodes. In our work, the selected electrodes are on the edge of the cap.

735 5.2.4. Prediction of the alertness state of a new subject

Given the results, the best compromise is to keep a subset of 9 electrodes (represented in Figure 24). Finally, we construct a CSP filter matrix W and a decision rule (using a LDA classifier) with these 9 electrodes on the 19 subjects.

740 In a future application of this work, the EEG of a new individual can be recorded using only the 9 selected electrodes. This signal will be band-pass filtered between 7 and 30 Hz. Our method needs at least a signal of 8 second length. If the recorded signal is larger than 16 seconds length, it will be split 745 into trials. If the signal length is between 8 and 16 seconds, an unique trial will be considered. The trail(s) of the unseen individuals will be projected using the W matrix learned on the 19 subjects and a log var transformation will be used. Then, the decision rule constructed on the 19 subjects will be applied on each trial (or on the unique trial) of the new individual. A 750 majority affection (or a simple affection if only one trial is available) will be used on all the trial(s) to predict the alertness of the new subject.

755 It is important to understand that even if the search of the best subset of electrodes to use was costly in computation time, our model provides a fast prediction for the alertness state of a new subject. Indeed, the prediction of one trial for an unseen subject only requires $4.21 * 10^{-4}$ seconds in our implementation (MATLAB running on a Windows with $8 * 2.8$ GHz Intel

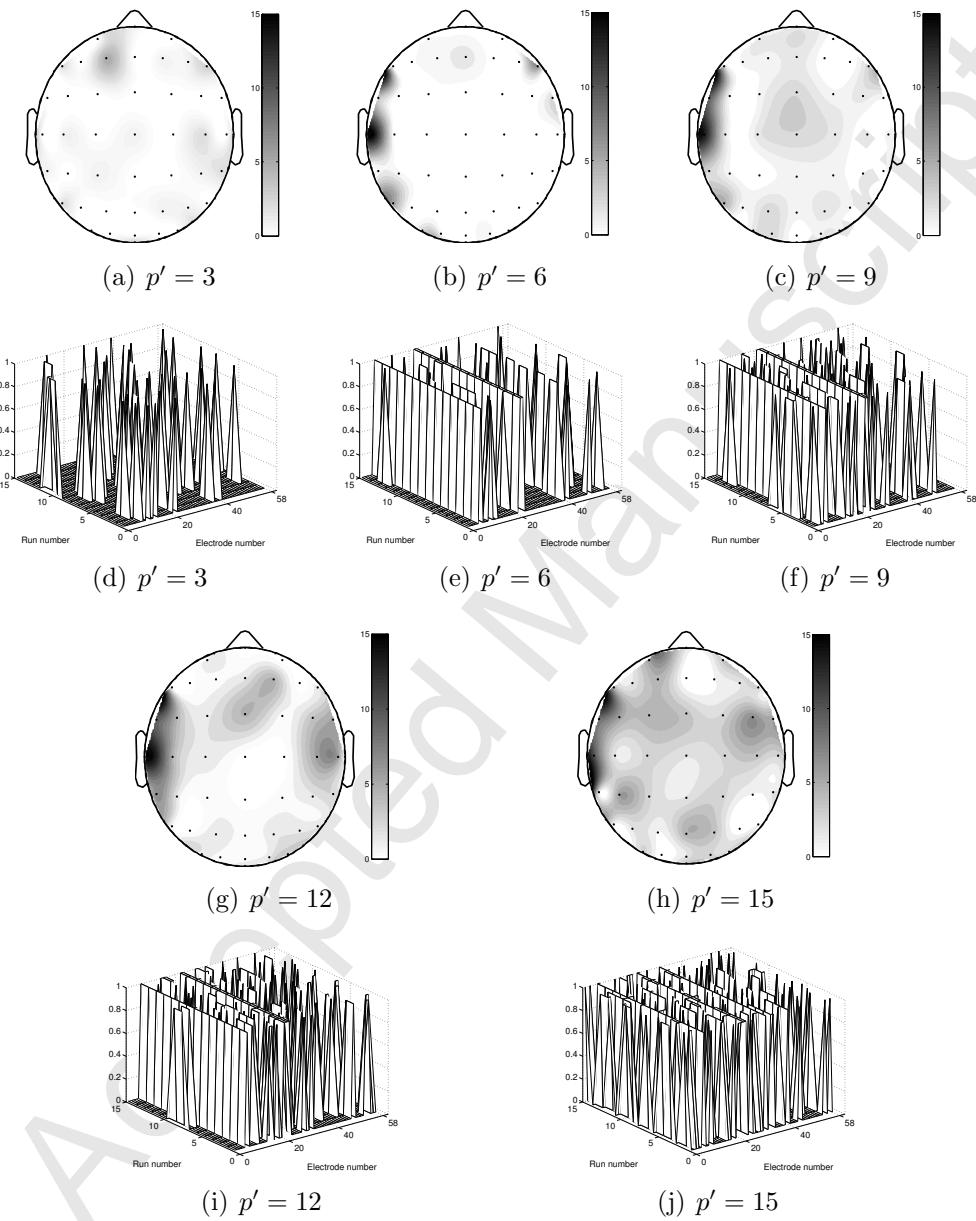


Figure 22: Figures a, b, c, g and h represents the number of selection of each electrode in the 15 best genomes (1 per run) found by the GA respectively for $p' = 3, 6, 9, 12$ and 15 . Figures d, e, f, i and j are 3D plots which represents the best genome found for each run of the GA with respectively $p' = 3, 6, 9, 12$ and 15 .

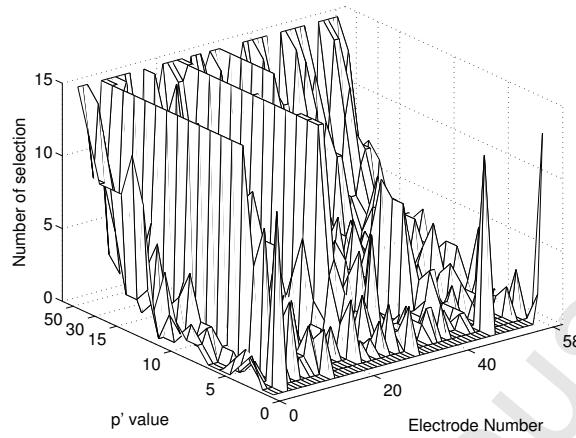


Figure 23: 3D plot which represents the number of selection of each electrode

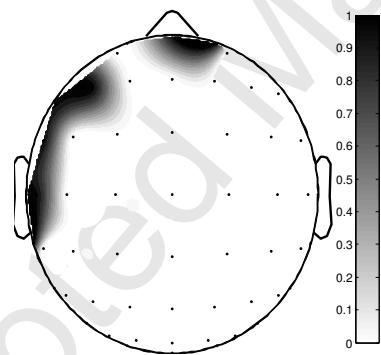


Figure 24: Best combination of 9 electrodes obtained by the best of the 15 runs of the GA. The selected electrodes appears in black. Figure obtained by using the Visual RCSP toolbox [11].

Core i7 2920XM processors and 16 GB 2.5 GHz RAM). For this reason, the obtained model (decision rule) allows us to consider real time applications.

760 6. Conclusion

In this paper, a method to predict the state of alertness of humans using their brain activity was studied. A genetic algorithm was used, for a given number p' , to find the best combination of p' electrodes to predict the alert-

ness state. On each combination of p' electrodes, a CSP method coupled to
 765 a linear discriminant analysis was used to build a decision rule (model) and to obtain a correct classification rate. Using this algorithm, a subset of 9 electrodes was kept. It allowed us to obtain a reliable model (average of correct classification rate equal to 71.59% on an external test set) to predict the alertness state of an unseen person. Our approach is costly in computation
 770 time (GA search) but it allows to construct a decision rule that provides a fast prediction of the alertness state of a new individual. Thus, this decision rule can be used in real time applications in order to automatically detect alertness changes for persons who have to stay vigilant (pilots or surgeons for example). The good classification rates obtained in this paper show that the
 775 CSP method can be used to extract synthetic signals from trial of different individuals. The methodology detailed in this paper can be easily adapted on other data in three dimensions (time, subjects and variables) like for example electrocardiogram (ECG) signals in order to predict a binary variable.

780 An exchange with neurobiologists now seems necessary to link the results obtained by the genetic algorithm to human physiology. Moreover, it is possible to extend the work proposed in this paper. In fact, the fitness function can be easily modified in order to consider the number of electrode kept (p'). Additionally, other linear and non-linear classification methods can
 785 be used to replace the LDA method. We are also trying to develop a sparse CSP method to perform variable selection in the CSP calculation (without a GA). This method could represent a faster way to perform variable selection using the CSP.

7. Appendix: Propositions and proofs

Proposition 7.1. *Let A and B be two symmetric and invertible matrices. Let \mathbf{w} be a \mathbb{R}^p vector. Then the solution of the following optimisation problem:*

$$\max_{\mathbf{w} \in \mathbb{R}^p} \frac{\mathbf{w}' A \mathbf{w}}{\mathbf{w}' B \mathbf{w}}$$

790 *is the eigenvector of $B^{-1}A$ which corresponds to the largest eigenvalue.*

Proof. We are looking for

$$\max_{\mathbf{w} \in \mathbb{R}^p} f(\mathbf{w}) = \max_{\mathbf{w} \in \mathbb{R}^p} \frac{\mathbf{w}' A \mathbf{w}}{\mathbf{w}' B \mathbf{w}}.$$

If $\frac{\partial f}{\partial \mathbf{w}} = 0$, \mathbf{w}_1 is a local optimum of $J(\mathbf{w})$. However,

$$\frac{\partial f}{\partial \mathbf{w}} = \frac{\left(\frac{\partial}{\partial \mathbf{w}} (\mathbf{w}' A \mathbf{w}) \right) (\mathbf{w}' B \mathbf{w}) - (\mathbf{w}' A \mathbf{w}) \left(\frac{\partial}{\partial \mathbf{w}} (\mathbf{w}' B \mathbf{w}) \right)}{(\mathbf{w}' B \mathbf{w})^2} = 0$$

$$\Leftrightarrow 2A\mathbf{w}(\mathbf{w}' B \mathbf{w}) - (\mathbf{w}' A \mathbf{w}) 2B\mathbf{w} = 0$$

$$\begin{aligned} &\Leftrightarrow A\mathbf{w} = \frac{\mathbf{w}' A \mathbf{w}}{\mathbf{w}' B \mathbf{w}} B \mathbf{w} \\ &\Leftrightarrow B^{-1} A \mathbf{w} = \frac{\mathbf{w}' A \mathbf{w}}{\mathbf{w}' B \mathbf{w}} \mathbf{w} \end{aligned}$$

Thus, \mathbf{w}_1 is a local optimum of $J(\mathbf{w})$ if and only if \mathbf{w}_1 is an eigenvector of $B^{-1}A$ corresponding to eigenvalue $\frac{\mathbf{w}_1' A \mathbf{w}_1}{\mathbf{w}_1' B \mathbf{w}_1} = J(\mathbf{w}_1)$. However, we want to maximize $f(\mathbf{w})$ so \mathbf{w}_1 is the eigenvector of $B^{-1}A$ which corresponds to the largest eigenvalue. ■

800 Proposition 7.2. Let A and B be symmetric and positive-definite matrices. Then AB is a positive-definite matrix.

Proof. As A is a symmetric and positive-definite matrix, it exists an unique positive-definite matrix C such that $A = C^2$. Thus, $AB = C^2 B = C C B C C^{-1} = C(CBC)C^{-1} = CDC^{-1}$ with $D = CBC$. This is similar to D which is symmetric and positive-definite. Indeed, $\forall \mathbf{x} \in \mathbb{R}^n$, we have $\mathbf{x}' D \mathbf{x} = \mathbf{x}' C B C \mathbf{x} = \mathbf{x}' C' B C \mathbf{x} = \tilde{\mathbf{x}}' B \tilde{\mathbf{x}}$, where $\tilde{\mathbf{x}} = C\mathbf{x} \in \mathbb{R}^n$. However $\tilde{\mathbf{x}}' B \tilde{\mathbf{x}} > 0$ because B symmetric and positive-definite.

Thus, like D , AB have all its eigenvalues real and positive. ■

810

Proposition 7.3. Let A be a $p \times p$ positive-definite matrix. Let $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p$ be the eigenvectors of A respectively belonging to eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p > 0$. Then, A^{-1} , the inverse of A , has for eigenvectors $\mathbf{x}_p, \mathbf{x}_{p-1}, \dots, \mathbf{x}_1$ respectively belonging to eigenvalues $\frac{1}{\lambda_p} \geq \frac{1}{\lambda_{p-1}} \geq \dots \geq \frac{1}{\lambda_1} > 0$.

Proof. A is positive-definite thus A is invertible. Moreover, $\forall i \in [1, p]$, we have

$$\begin{aligned} A\mathbf{x}_i &= \lambda_i \mathbf{x}_i \\ \Leftrightarrow A^{-1}A\mathbf{x}_i &= A^{-1}\lambda_i \mathbf{x}_i \\ \Leftrightarrow \mathbf{x}_i &= \lambda_i A^{-1}\mathbf{x}_i \\ \Leftrightarrow \frac{1}{\lambda_i} \mathbf{x}_i &= A^{-1}\mathbf{x}_i \end{aligned}$$

Then, \mathbf{x}_i is the eigenvector of A^{-1} belonging to $\frac{1}{\lambda_i}$. ■

Proposition 7.4. Let C_1 and C_2 be two symmetric and positive-definite matrices. Then, $(C_2 + C_1)^{-1}C_1$, $(C_2 + C_1)^{-1}C_2$, $C_2^{-1}C_1$ and $C_1^{-1}C_2$ have same eigenvectors belonging to different eigenvalues.

Proof. To begin, we will notice that if C_2 and C_1 are symmetric and positive-definite matrices, then $(C_2 + C_1)^{-1}C_1$, $(C_2 + C_1)^{-1}C_2$, $C_2^{-1}C_1$ and $C_1^{-1}C_2$ are also positive-definite matrices. In fact, the inverse of a positive-definite matrix is a positive-definite matrix. Moreover, Proposition 7.2 allows us to know that the product of two symmetric and positive-definite matrices is a positive-definite matrix. Thus, let \mathbf{u} be the eigenvector of $(C_1 + C_2)^{-1}C_1$ belonging to the eigenvalue λ . By definition, it comes:

$$(C_1 + C_2)^{-1} C_1 \mathbf{u} = \lambda \mathbf{u} \quad (7)$$

$$C_1 \mathbf{u} = \lambda (C_1 + C_2) \mathbf{u}$$

$$\frac{1}{\lambda} \mathbf{u} = C_1^{-1} (C_1 + C_2) \mathbf{u}$$

$$\frac{1}{\lambda} \mathbf{u} = (I + C_1^{-1} C_2) \mathbf{u}$$

$$\left(\frac{1}{\lambda} - 1\right) \mathbf{u} = C_1^{-1} C_2 \mathbf{u}$$

$$\left(\frac{1-\lambda}{\lambda}\right) \mathbf{u} = C_1^{-1} C_2 \mathbf{u} \quad (8)$$

$$\left(\frac{1-\lambda}{\lambda}\right) C_1 \mathbf{u} = C_2 \mathbf{u}$$

$$C_2^{-1} C_1 \mathbf{u} = \left(\frac{\lambda}{1-\lambda}\right) \mathbf{u} \quad (9)$$

$$C_2^{-1} C_1 \mathbf{u} = \left(\frac{\lambda}{1-\lambda} - 1 + 1\right) \mathbf{u}$$

$$(I + C_2^{-1} C_1) \mathbf{u} = \left(\frac{\lambda}{1-\lambda} + 1\right) \mathbf{u}$$

$$C_2^{-1} (C_2 + C_1) \mathbf{u} = \left(\frac{1}{1-\lambda}\right) \mathbf{u}$$

$$(C_2 + C_1) \mathbf{u} = C_2 \left(\frac{1}{1-\lambda}\right) \mathbf{u}$$

$$(1 - \lambda) \mathbf{u} = (C_2 + C_1)^{-1} C_2 \mathbf{u}. \blacksquare \quad (10)$$

(11)

Acknowledgements

Funding provided by FP7-PEOPLE-2013-IRSES project ACOBSEC financed by the European Commission, contract No. 612689. Additional funding provided by the CONACYT (Mexico) Basic Science Research Project No. 835 178323. The authors would like to thank Julien Clauzel, Nidal El Yacoubi, Emilie Drouineau and Vérane Faure, interns in the team, for their help and support during the acquisition of the EEG recordings.

References

References

- 840 [1] E. Lughofe, J.E. Smith, P. Caleb-Solly, M.A. Tahir, C. Eitzinger, D. Sannen, M. Nuttin, Human-Machine Interaction Issues in Quality Control Based on On-Line Image Classification, *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, vol. 39 (5), pp. 960–971, 2009.
- 845 [2] E. Lughofe, Human-Inspired Evolving Machine — The Next Generation of Evolving Intelligent Systems?, *SMC Newsletter*, vol. 36, 2011.
- [3] A. Subasi, M. Akin, K. Kiymik, O. Erogul, Automatic recognition of vigilance state by using a wavelet-based artificial neural network, *Neural Comput and Applic* 14 (2005) 45–55.
- 850 [4] N. Hazarika, J. Chen, C. Tsoi, A. Sergejew, Classification of EEG signals using the wavelet transform, *Signal Processing* 59 (1997) 61–72.
- [5] K. Ben Khalifa, M. Bédoui, M. Dogui, F. Alexandre, Alertness states classification by SOM and LVQ neural networks, *International Journal of Information Technology* 1 (2005) 131–134.
- 855 [6] H. Cecotti, A. Graeser, Convolutional neural network with embedded fourier transform for EEG classification, *International Conference on Pattern Recognition*, Tampa, Florida (2008) 1–4.
- [7] K. Fukunaga, W. L. G. Koontz, Application of the karhunen-loève expansion to feature selection and ordering, *IEEE Transactions on Computers* C-19 (4) (1970) 311–318.
- 860 [8] Z. J. Koles, M. S. Lazar, S. Z. Zhou, Spatial patterns underlying population differences in the background EEG., *Brain topography* 2 (1990) 275–284.
- [9] G. Pfurtscheller, C. Neuper, D. Flotzinger, M. Pregenzer, EEG-based discrimination between imagination of right and left hand movement, *Electroencephalography and Clinical Neurophysiology* 103 (1997) 642 – 651.

- [10] J. Müller-Gerking, G. Pfurtscheller, H. Flyvbjerg, Designing optimal spatial filters for single-trial EEG classification in a movement task, Clinical Neurophysiology 110 (1999) 787–798.
- [11] F. Lotte, C. Guan, Regularizing common spatial patterns to improve BCI designs: unified theory and new algorithms., IEEE Transactions on biomedical Engineering 58 (2) (2011) 355–362.
- [12] A. Vuckovic, V. Radivojevic, A. Chen, D. Popovic, Automatic recognition of alertness and drowsiness from EEG by an artificial neural network, Medical Engineering and Physics 24 (2002) 349–360.
- [13] M. Yeo, X. Li, K. Shen, E. Wilder-Smith, Can SVM be used for automatic EEG detection of drowsiness?, Safety Science 47 (2009) 115–124ra.
- [14] C. Anderson, Z. Sijercic, Classification of EEG signals from four subjects during five mental tasks, Proceedings of the Conference on Engineering Applications in Neural Networks, London, United Kingdom (1996) 407–414.
- [15] B. Obermaier, C. Guger, C. Neuper, G. Pfurtscheller, Hidden markov models for online classification of single trial EEG data, Pattern recognition letters 22 (2001) 1299–1309.
- [16] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, K.-R. Müller, Optimizing spatial filters for robust eeg single-trial analysis, in: IEEE Signal Proc. Magazine, 2008, pp. 581–607.
- [17] L.-C. Shi, Y. Li, R.-H. Sun, B.-L. Lu, A sparse common spatial pattern algorithm for brain-computer interface, in: Proceedings of the 18th international conference on Neural Information Processing - Volume Part I, ICONIP'11, Springer-Verlag, Berlin, Heidelberg, 2011, pp. 725–733.
- [18] G. Chatrian, E. Lettich, P. Nelson, Ten percent electrode system for topographic studies of spontaneous and evoked eeg activity, Journal of Clinical Neurophysiology 25 (1985) 83–92.
- [19] J. H. Schultz, Le Training autogne, PUF, 1958.
- [20] E. Jacobson, Biologie des motions. Les bases thoriques de la relaxation., 1974.

- 900 [21] W. G. Walter, R. Cooper, V. Aldridge, W. C. McCallum, A. Winter,
Contingent negative variation: An electric sign of sensorimotor association and expectancy in the human brain., *Nature* 203 (1964) 380–384.
- [22] W. Rosenblith, Some quantifiable aspects of the electrical activity of the nervous system (with emphasis upon responses to sensory stimuli), *Rev. Mod. Physics* 31 (1959) 532–545.
- 905 [23] J. J. Tecce, A CNV rebound effect., *Electroencephalography and Clinical Neurophysiology* 46 (1979) 546–551.
- [24] P. Naitoh, L. C. Johnson, A. Lubin, Modification of surface negative slow potential (CNV) in the human brain after total sleep loss., *Electroencephalography and Clinical Neurophysiology* 30 (1971) 17–22.
- 910 [25] M. Timsit-Berthier, A. Gerono, H. Mantanus, Inversion de polarité de la variation contingente négative au cours d'état d'endormissement., *EEG Neurophysiol* 11 (1981) 82–88.
- 915 [26] H. Ramoser, J. Müller-Gerking, G. Pfurtscheller, Optimal spatial filtering of single trial EEG during imagined hand movement, *IEEE Transactions on Rehabilitation Engineering* 8 (4) (2010) 441–446.
- [27] M. Arvaneh, C. Guan, K. K. Ang, C. Quek, Optimizing the channel selection and classification accuracy in EEG-based BCI, *IEEE Transactions on Biomedical Engineering* 58 (2011) 1865–1873.
- 920 [28] T. Hastie, R. Tibshirani, J. Friedman, *The Elements of Statistical Learning : Data Mining, Inference, and Prediction.*, 2nd Edition, Springer, 2009.
- [29] A. De Jong, K, An analysis of the behavior of a class of genetic adaptive systems, Ph.D. thesis, University of Michigan (1975).
- 925 [30] H. Holland, J., *Adaptation in Natural and Artificial Systems*, Ann Arbor, University of Michigan Press, 1975.

Laurent Vézard is a PhD student at the third year. His advisors are Dr. Marie Chavent, Dr. Pierrick Legrand and Dr. Frédérique Faïta-Aïnseba. He is bio statistical Engineer by the National Institute of Applied Sciences of Toulouse (2010). He is also member of the CQFD team of Inria Bordeaux Sud-Ouest. His current main research interests are classification, dimension reduction and feature extraction from data using genetic algorithms.

Pierrick Legrand received his PhD in applied mathematics from "Ecole centrale de Nantes" and from Nantes university in December 2004. In 2005 and 2006, he received two post-doctoral positions (Evovision group at CI-CESE research center (Ensenada, México) and Inria COMPLEX Team (Rocquencourt, France). On September 2006, he became associate professor at the University of Bordeaux and researcher at the IMB (Institut de Mathématiques de Bordeaux, UMR CNRS 5251). He was a researcher in the Inria ALEA team from 2010 to 2014. Since 2014, he is a member of the CQFD research team at Inria Bordeaux Sud-Ouest.

Marie Chavent is assistant professor in statistics at University of Bordeaux (France). She is member of the Probability and Statistics team at Mathematics Institute of Bordeaux (IMB, UMR CNRS 5251). She is also member of the CQFD research team of Inria Bordeaux Sud-Ouest. Her research interests are dimension reduction, clustering and data analysis.

Frédérique Faïta-Aïnseba is assistant professor and researcher in Cognitive Sciences at the Bordeaux University in France. She obtained her PhD in Neuroscience in Marseille (France) in 1995. After her graduation, she held a post-doctoral position at Montreal University (Canada). Since her arrival to Bordeaux, she dedicates her researches to the study of cognitive processes involve in language, reading and musical listening.

Leonardo Trujillo received an electronic engineering (2002) and a masters in computer science (2004) from the Technical Institute of Tijuana in México. He then received a doctorate in computer science from CICESE research center, México (2008). He is currently professor at the Technical Institute of Tijuana in México (ITT). Dr. Trujillo is involved in interdisciplinary research in the fields of evolutionary computation, computer vision, image analysis, pattern recognition and autonomous robotics.