EEG sensor selection by sparse spatial filtering in P300 speller brain-computer interface
Bertrand Rivet, Hubert Cecotti, Ronald Phlypo, Olivier Bertrand, Emmanuel Maby, Jérémie Mattout

► To cite this version:

HAL Id: hal-00517388
https://hal.archives-ouvertes.fr/hal-00517388
Submitted on 14 Sep 2010

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.
EEG sensor selection by sparse spatial filtering in P300 speller brain-computer interface

Bertrand Rivet, Hubert Cecotti, Ronald Phlypo, Olivier Bertrand, Emmanuel Maby and Jérémie Mattout

Abstract—A Brain-Computer Interface (BCI) is a specific type of human-machine interface that enables communication between a subject/patient and a computer by direct control from decoding of brain activity. This paper deals with the P300-speller application that enables to write a text based on the oddball paradigm. To improve the ergonomics and minimize the cost of such a BCI, reducing the number of electrodes is mandatory. We propose a new algorithm to select a relevant subset of electrodes by estimating sparse spatial filters. A $l_1$-norm penalization term, as an approximation of the $l_0$-norm, is introduced in the xDAWN algorithm, which maximizes the signal to signal-plus-noise ratio. Experimental results on 20 subjects show that the proposed method is efficient to select the most relevant sensors; from 32 down to 10 sensors, the loss in classification accuracy is less than 5%.

I. INTRODUCTION

Brain-Computer Interfaces (BCI) are new kinds of devices, which enable direct communication between the user’s brain and a computer [1], [2]: such systems allow people to communicate through direct measurements of brain activity, without requiring any movement. BCIs usually exploit cerebral activities measured by electroencephalography (EEG) and could thus be useful to patients who suffer from severe neuromuscular disorders (e.g. “locked in” patients). Current BCIs determine the user’s intent from different electrophysiological signals: for instance the user may actively control some brain waves (e.g., mu or beta rhythms) or the BCI may exploit natural passive response of the brain to external stimuli (e.g., event related potentials) [1].

The BCI problem addressed in this paper concerns the P300-speller [3], [4] which enables people to write a text on a computer. It is based on the oddball paradigm: the task is to discriminate between epochs containing a P300 potential evoked by a (rare) target stimulus and epochs associated with the (frequent) non-target stimuli. To gather enough evidence and yield robust classification, each symbol/stimulus is presented/flashed several times. However, this is done at the expense of reducing the speed of the system: e.g., with 15 repetitions, about two symbols can be spelt per minute only [3]. Moreover, the signal-to-noise ratio (SNR) of EEG signals is very low, and signals often contain muscular and/or ocular artifacts. To increase SNR and hence classification accuracy, several methods based on spatial filtering have been proposed such as those based on independent component analysis (ICA) [5], [6] or, more recently, the xDAWN algorithm [7], [8]. These studies have shown that enhancing the P300 response using spatial filters enables to reduce the size of the required training set in order to reach a certain level of classification accuracy in the test phase. As a consequence, the ergonomics is improved as the training phase becomes shorter. Yet to further improve the BCI ergonomics, a crucial need is to reduce the number of electrodes. Indeed, using a large number of electrodes is undesirable since it increases the discomfort of the patient, the time needed to install the whole system as well as its global cost. However, only a few studies have focused on sensor selection so far. Some studies did compared a priori selected subsets in terms of subsequent classification performance (e.g., [9], [10]) while others have applied by blind approaches to select the most relevant sensors (e.g., [11]). The major drawback of the latter is its need for K-fold cross validation, which requires a large number of training symbols. In [12], we proposed a faster and efficient sensor selection approach based on backward elimination. In the current study, the xDAWN algorithm is modified by supplementing the SNR criterion with a $l_1$-norm penalization term which leads to estimate sparse spatial filters, that is with only a few non-zero coefficients.

This paper is organized as follows. Section II describes the P300 subspace estimation and the associated classification issue. Section III presents the obtained results on empirical data from twenty subjects. Finally Section IV concludes the paper with comments and perspectives for future work.

II. METHODOLOGY

A. Data description and preprocessing

The P300-speller enables the user to select symbols (letters, digits, images). The target plays the role of an oddball and thus elicits a P300 response each time it is emphasized on screen. In our setting, a $6 \times 6$ matrix that contains all the available symbols is displayed on a computer screen (Fig. 1). To spell one symbol, the user has to focus on the one s/he wants to spell. When the user focuses his attention on a cell of the matrix, it is possible to detect a P300 (a positive deflection in voltage at a latency of about 300 ms in the EEG) after the cell was intensified. Symbols are not intensified or flashed individually but together with the symbols that belong to the same matrix row or column. Row and column intensification is randomized by blocks of 12 (6 rows and 6 columns). The sets of 12 intensifications is repeated $N_{\text{epoch}}$ times for each symbol. Therefore, $2N_{\text{epoch}}$ P300 responses might be detected for to identify the target recognition: each

This work was partially supported by ANR-Tecsan RoBik, ANR-Blanc GAZEE&EEG and ANR-DEFIS CoAdpaf.

Bertrand Rivet, Hubert Cecotti and Ronald Phlypo are with GIPSA-lab, CNRS-UMR 5226, Grenoble Institute of Technology, France bertrand.rivet@gipsa-lab.grenoble-inp.fr

Olivier Bertrand, Emmanuel Maby and Jérémie Mattout are with INSERM U821, Institut Federatif des Neurosciences, Lyon, France
symbol is uniquely defined by the intersection of one row and one column.

Data were recorded from 20 healthy subjects [13]. Two sessions were considered: one for training the classifier, the other for testing. For the training and test sessions, the subject had to write 50 and 60 predetermined symbols respectively. Each row and column in the spelling matrix was randomly intensified for 100ms and the delay between two consecutive intensifications was 80ms, leading to an interstimulus interval (ISI) of 180ms. For each symbol there was 10 consecutive epochs (Nepoch = 10).

The EEG signals were initially sampled at 100Hz. Data preprocessing included a fourth order bandpass filter between 1Hz and 20Hz. Finally, each sensor signal was centered and reduced so that they had a zero mean value and a standard deviation equal to one.

B. xDAWN algorithm to enhance P300 potentials

In this section, the principle of the xDAWN algorithm is briefly reprised [7], [8]. Its goal is to enhance the signal to signal-plus-noise (SSNR). This method relies on two assumptions: i) signal is made of two typical responses, one evoked by the targets and one evoked by all stimuli (target and non-target ones); ii) responses evoked by the target stimuli could be enhanced by spatial filtering. Let \( X \in \mathbb{R}^{N_t \times N_s} \) denote the recorded signals, where \( N_t \) and \( N_s \) are the number of samples and sensors, respectively. We denote the responses synchronized with the target and non-target stimuli by \( A_1 \in \mathbb{R}^{N_t \times N_s} \) and \( A_2 \in \mathbb{R}^{N_t \times N_s} \), respectively. Thus,

\[
X = D_1 A_1 + D_2 A_2 + N,
\]

where \( D_1 \in \mathbb{R}^{N_t \times N_t} \) and \( D_2 \in \mathbb{R}^{N_t \times N_t} \) are Toeplitz matrices whose first column entries are set to zero except for those that correspond to target stimuli and all stimuli, respectively. \( N_t \) and \( N_s \) indicate the number of samples in responses \( A_1 \) and \( A_2 \), respectively. \( N \in \mathbb{R}^{N_t \times N_s} \) denotes the residual noise.

The aim of the xDAWN algorithm is to estimate \( N_f \) spatial filters \( U_1 \in \mathbb{R}^{N_s \times N_f} \) so that the SSNR defined by

\[
g(U) = \frac{\text{Tr}(U^T \Sigma_1 U)}{\text{Tr}(U^T \Sigma_X U)}
\]

is maximum

\[
\hat{U}_1 = \arg \max_{U_1} g(U_1)
\]

with \( \text{Tr}(\cdot) \) is the trace operator and \( \hat{\Sigma}_1 = \hat{A}^T_1 D_1^T D_1 \hat{A}_1 \), \( \Sigma_X = X^T X \). Note that \( \hat{A}_1 \) is the least mean square estimation of the unknown target evoked response \( A_1 \). Since \( D_1 A_1 \) and \( D_2 A_2 \) could overlap, \( \hat{A}_1 \) is estimated from

\[
\begin{pmatrix}
\hat{A}_1 \\
\hat{A}_2 
\end{pmatrix}
= \left( D^T D \right)^{-1} D^T X,
\]

where \( D = [D_1, D_2] \). To estimate the spatial filters \( \hat{U}_1 \), one computes the QR decompositions \( X = Q_X R_X \) and \( D_1 = Q_1 R_1 \) [14]. Spatial filters then obtains from

\[
\hat{U}_1 = R_X^{-1} \Psi_{1:N_f}, \quad \Psi_{1:N_f} \text{ is the concatenation of the } N_f \text{ singular vectors } \Psi_i \text{ associated with the } \lambda_i \text{ maximum singular values provided by the singular value decomposition of } R_1 B_1^T Q_X = \Phi \Lambda \Psi^T. \Phi \text{ and } \Psi \text{ are unitary matrices, } \lambda \text{ is a diagonal matrix } [14] \text{ and } A_1 = B_1^T X. \text{ Finally, the enhanced signals are given by}

\[
\hat{S}_1 \triangleq X \hat{U}_1 = D_1 A_1 \hat{U}_1 + D_2 A_2 \hat{U}_1 + N \hat{U}_1.
\]

C. Sparse estimation of spatial filters to select sensors

A substantial drawback of the xDAWN algorithm (3) is that it provides a non-sparse spatial filters \( \hat{U}_1 \), leading thus to consider that all the sensors are necessary to achieve a suitable enhancement of response \( A_1 \) evoked by the target. However we have already shown that based on this approach, the number of sensors could still be reduced (e.g., from 32 down to 8 sensors) while keeping a high classification accuracy (85% instead of 92%) [12]. Nevertheless, in this study, we describe and evaluate a more generic and simpler method to select relevant sensors. It consists of introducing a \( l_1 \)-norm penalization term in the xDAWN criterion (3) as follows:

\[
\hat{v}_1 = \arg \max_{\|v_1\|_2 = 1} f(v_1, \lambda),
\]

with \( f(v_1, \lambda) \) defined by

\[
f(v_1, \lambda) = g(v_1) - \lambda \frac{\|v_1\|_1}{\|v_1\|_2},
\]

where hyperparameter \( \lambda \geq 0 \) is a penalization multiplier adjusted to control the sparsity of the spatial filter \( v_1 \). Indeed, the \( l_1 \)-norm penalization is well known to be a good candidate to solve sparse problem as it could be seen as an approximation of the \( l_0 \)-norm which is the number of non-zero elements of a vector [15], [16]. Note that the constraint \( \|v_1\|_2 = 1 \) is required since the Rayleigh quotient (first term in criterion (7)) is invariant by a scaling factor. To maximize (6), a gradient method is used:

\[
v_1^{(k+1)} = \frac{v_1^{(k)} + \rho^{(k)} \partial v_1 f(v_1^{(k)}, \lambda)}{\|v_1^{(k)} + \rho^{(k)} \partial v_1 f(v_1^{(k)}, \lambda)\|_2},
\]

where \( \partial v_1 f(v_1^{(k)}, \lambda) \) is a subgradient of \( f(v_1, \lambda) \) defined by

\[
\partial v_1 f(v_1^{(k)}, \lambda) = \frac{\partial}{\partial v_1} \left( v_1^T \hat{\Sigma}_1 v_1 \right) - \lambda \partial v_1 \left( \frac{\|v_1\|_1}{\|v_1\|_2} \right),
\]
with
\[
\frac{\partial}{\partial \mathbf{v}_1} \left( \mathbf{v}_1^T \hat{\Sigma}_1 \mathbf{v}_1 \right) = \frac{2}{\mathbf{v}_1^T \Sigma_X \mathbf{v}_1} \left( \hat{\Sigma}_1 \mathbf{v}_1 - \mathbf{v}_1^T \hat{\Sigma}_1 \mathbf{v}_1 \right),
\]
and
\[
\frac{\partial}{\partial \mathbf{v}_1} \left( \frac{\|\mathbf{v}_1\|_1}{\|\mathbf{v}_1\|_2} \right) = \frac{S(\mathbf{v}_1) 1_{N_v}}{\|\mathbf{v}_1\|_2} - \frac{\|\mathbf{v}_1\|_1}{(\mathbf{v}_1^T \mathbf{v}_1)^{3/2}} \mathbf{v}_1,
\]
where $1_{N_v} \in \mathbb{R}^{N_v}$ is a vector of ones and $S(\mathbf{v}_1)$ is a diagonal matrix such that
\[
S(\mathbf{v}_1)_{ii} = \begin{cases} 
1, & \text{if } \mathbf{v}_1(i) > 0 \\
0, & \text{if } \mathbf{v}_1(i) = 0 \\
-1, & \text{if } \mathbf{v}_1(i) < 0.
\end{cases}
\]
Note that i) this is only a subgradient since the $l_1$-norm is undifferentiable; ii) the division of $\|\mathbf{v}_1\|_1$ by the $l_2$-norm $\|\mathbf{v}_1\|_2$ in the penalization term in criterion (7) is to ensure that iteration $\mathbf{v}_1^{(k)} + \rho^{(k)} \partial_l f(\mathbf{v}_1^{(k)}, \lambda)$ has a $l_2$-norm as close as possible to one before projection onto the $l_2$-ball of unit radius in order to speed up the convergence of the gradient algorithm. In this study, parameter $\lambda$ that controls the sparsity of the estimated spatial filter $\hat{\mathbf{v}}_1$ is chosen manually based on the number $N_{ss}$ of selected sensors. Finally, the $N_{ss}$ selected sensors corresponding to the non-zero entries in $\hat{\mathbf{v}}_1$ (6). The final sparse spatial filters $\hat{\mathbf{v}}_1^*$ are then computed by applying the classical (unconstrained) xDAWN algorithm (3) to those selected sensors only.

**D. Classification**

Among the proposed classifiers that are considered for BCIs, Bayesian linear discriminant analysis (BLDA) [9] is chosen since it has proved efficient and is fully automated (i.e. no need to adjust some hyperparameter) [9]. BLDA consists of finding a discriminant vector $\mathbf{w}$ such that $\mathbf{w}^T \mathbf{p}$ is as close as possible to the class index $t$ associated with the corresponding feature vector $\mathbf{p}$ obtained from the concatenation of time-course samples of enhanced signals. The discriminant vector $\mathbf{w}$ is estimated from the set of pairs $\{\mathbf{p}_j, t_j\}_{1 \leq j \leq 12N_v, N_{epochs}}$ obtained from the $N_v$ symbols of the training database.

**III. RESULTS**

In this section we present the results obtained from the above described experiment using the proposed new approach for sparse spatial filtering, in comparison with classical filtering using the xDAWN algorithm. We considered epochs of 1s duration for the target response $A_1$ and 1.80s for non-target response $A_2$. Moreover, in this preliminary study, the penalization parameter $\lambda$ was optimized manually and independently for each subject.

As $\lambda$ increases, the sparsity of the estimated spatial filters $\hat{\mathbf{v}}_1$ (6) grows (Fig. 2). This penalization parameter must thus be chosen carefully with respect to the desired number of sensors to be kept. On figure 2, we see the interesting evolution of the SSNR as a function of hyperparameter $\lambda$. The sparse SSNR corresponds to $g(\mathbf{v}_1)$, where $\mathbf{v}_1$ are the sparse spatial filters computed using the penalized criterion. The classical SSNR refers to $g(\hat{\mathbf{v}}_1^*)$, where $\hat{\mathbf{v}}_1^*$ are the spatial filters computed using the unconstrained criterion, once the sensors have been selected. We observe that the sparse SSNR follows a continuous decrease, while the classical SSNR follows a step-wise decrease where each step corresponds to a drop of one or several sensors. For the latter, the more sensors have been dropped simultaneously, the larger the drop in SSNR. This clear relationship suggests that one could fairly easily adjust the penalization parameter given a desired final number of sensors.

Fig. 3 shows an example of spatial filters estimated by classical xDAWN (Fig. 3(a)) and by sparse xDAWN (Fig. 3(b)), respectively. With $\lambda = 0.4$, only 5 out of 32 sensors ends up with non-zero entries. Importantly, it appears that the remaining sensors do not necessarily correspond to those with the largest entries in $\mathbf{u}_1$. Hence simple projection onto the $l_1$-ball or soft/hard-thresholding would not have resulted in the same optimum set of sensors.

Figure 4 depicts the classification accuracy as a function of the number of selected sensors ($N_{ss}$). As expected, the more sensors, the higher the accuracy. However, from 10 to 32 sensors the gain in average classification accuracy is smaller than 5%. The same effect obtains when only considering 5 epochs per symbol. Finally, we observed that for about 75% of the subjects, reducing the number of electrodes down to 8 has not a profound impact on classification accuracy (Fig. 4(b)). This is highly promising in the aim of improving the ergonomics of such BCIs.

Although the selected electrodes are specific to each subject, preliminary comparisons between subjects show that
Further works will include a deeper analysis of the influence of the penalization parameter $\lambda$ on the criterion to propose a fully automated scheme. Then we will be able to focus on the identification of universal sensor subsets that could ensure high accuracy for every subject.

**References**


[16] D. L. Donoho, "For most large underdetermined systems of equations, the minimal $\ell_1$,$\ell_2$,$\ell_\infty$-norm near-solution approximates the sparsest near-solution," *Communications on Pure and Applied Mathematics*, vol. 59, no. 7, pp. 907–934, 2006.
