xDAWN algorithm to enhance evoked potentials: application to brain-computer interface.

Bertrand Rivet, Antoine Souloumiac, Virginie Attina, Guillaume Gibert

xDAWN Algorithm to Enhance Evoked Potentials: Application to Brain Computer Interface

Bertrand Rivet[1,2], Antoine Souloumiac[2], Virginie Attina[3] and Guillaume Gibert[3]

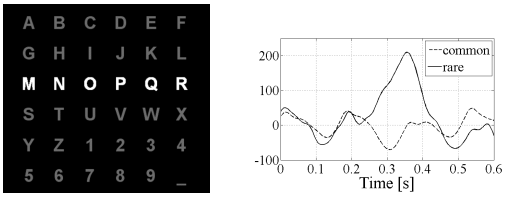
[1] GIPSA-lab, CNRS-UMR 5216, Grenoble Institute of Technology 46 avenue Félix Viallet, 38000 Grenoble, France.

[2] CEA, LIST, Stochastic Processes and Spectra Laboratory, F-91191 Gif-Sur-Yvette, France.

[3] INSERM, U821, Lyon, F-69500, France; IFR 19, Institut Fédératif des Neurosciences, Lyon, F-69000, France; Univ Lyon 1, Lyon, F-69000, France.

***Abstract*—A brain-computer interface (BCI) is a communication system which allows to control a computer or any other devices thanks to the brain activity. The BCI described in this paper is based on the P300 speller BCI paradigm introduced by Farwell and Donchin [1]. An unsupervised algorithm is proposed to enhance P300 evoked potentials by estimating spatial filters; the raw EEG signals are then projected into the estimated signal subspace. Data recorded on three subjects were used to evaluate the proposed method. The results, which are presented using a Bayesian linear discriminant analysis (BLDA) classifier [2], show that the proposed method is efficient and accurate.**

***Index Terms*—Brain computer interface, P300-speller, xDAWN algorithm, spatial enhancement.**



(a) (b)

Fig. 1. Brain-Computer Interface "*P300 speller*". Fig. 1(a): screen display as shown to the subjects with the third highlighted row. Fig. 1(b): time course of the actual signal waveforms at C_z. The continuous line represents the average over rare (i.e. target) stimuli and the dashed line corresponds to the average over common (i.e. non-target) stimuli.

I. Introduction

Brain-Computer Interfaces (BCI) enable direct communication between the user's brain and a computer by analysing brain activities measured with electroencephalogram (EEG) for example [3]. Such human-computer interfaces provides a new non-muscular powerful channel for communicating with the external world. BCIs are thus suitable for people that are incapable of any motor functions: e.g., people with severe neuromuscular disorders or 'locked-in' syndrome patients [4], [5]. Present-day BCIs determine the intent of the user from different electrophysiological signals: for instance, the user may control the modulation of some brain waves (e.g., mu or beta rhythms [6], [7]) or the BCI may exploit natural automatic responses of the brain to external stimuli (e.g., event-related potentials [8], [9], [10], [2]). See [3] for a more detailed review of BCIs.

The BCI problem addressed in this paper concerns the P300 speller introduced by Farwell and Donchin [1], [8]. It enables people to spell words on a computer by sequentially choosing letters from the alphabet without doing any movement: a 6×6 matrix, that includes all the alphabet letters as well as other symbols, was presented to the user on a computer screen (Fig. 1(a)). The procedure used in this BCI is actually derived from the oddball paradigm, in which the subject is asked to distinguish between a common stimulus (also called non-target) and a rare stimulus (also called target) by a mental counting of the target stimulus. As a result of the attentional focus which is enhanced by mental counting, a typical automatic potential is evoked in the brain. Applied to the P300 speller BCI, the stimuli are visual: they consist in the intensification of each of the rows and of the columns of the spelling matrix in a random order. The user focuses her/his attention on the symbol she/he wishes to communicate and mentally counts the number of times the row and the column containing the symbol are intensified. One of the component of the spontaneous response elicited in the user's brain by the target stimuli is known as a P300 evoked potential corresponding to a positive deviation occurring around 300 ms after the stimuli (Fig. 1(b)). The prediction of the desired symbol consists in discriminating between row/column stimuli leading to a P300 evoked potential from row/column stimuli which do not generate a P300 potential: the desired symbol is so determined as the intersection of the row/column targets.

The recorded EEG signals contain P300 potentials as well as other brain activities, muscular and/or ocular artifacts leading to a very low signal-to-noise ratio (SNR) of P300 potential. As a result, the detection of target stimuli is very difficult from a single trial which is defined as a sequence of intensifications in a random order of each of the 6 rows and the 6 columns of the spelling matrix. To increase the classification accuracy, each symbol is spelt several consecutive times and the epochs corresponding to each row/column are averaged over the trials. However, these repetitions decrease the number of symbols spelt per minute: e.g., with 15 repetitions, only 2 characters are typically spelt per minute [1], [8]. Several ways were

proposed to limit the number of necessary repetitions given a high prediction accuracy. The first way is to use a more complicated classifier than a simple average. For instance, Rakotomamonjy *et al.* [11] used as classifier an ensemble of several linear support vector machines (SVM) [12] with an automatic channels selection, and Hoffmann *et al.* [13] proposed a boosting approach. Another way to improve the symbol prediction accuracy is to enhance the P300 evoked potentials by a spatial filtering of the channels. Several methods, based on independent component analysis (ICA) [14], [15], [16], were thus proposed to enhance the SNR and to remove the artifacts, e.g. [17], [10]. However, the major drawback of such methods is that they are not specifically designed to separate brain waves. In most of BCI systems using ICA, after the decomposition in independent components (IC) it is necessary to select (manually or thanks to spatiotemporal prior) the ICs which mainly contained the desired evoked potentials.

In this paper, we address the problem to enhance the P300 evoked potentials for the P300 speller BCI. The proposed method is unsupervised and specifically designed to the P300 speller paradigm: indeed, it only exploits the instants of the visual stimuli. In preliminaries studies [18], [19], we have shown how to automatically estimate P300 subspace from raw EEG signals. P300 evoked potentials are then enhanced by projecting raw EEG on the estimated P300 subspace. In the present study, we extend the proposed algorithm in a more robust way. Moreover, we present in this paper a deeper analysis of the proposed method and more experiments results using a new database confirming the interest of the proposed algorithm.

This paper is organised as follows: Section II describes the evoked subspace estimation and Section III presents the BCI methodology. Section IV presents the results that have been achieved whereas Section V concludes the paper with comments and perspective on the work.

II. xDAWN ALGORITHM TO ENHANCE EVOKED POTENTIALS

The raw EEGs recorded from the user's scalp not only contain the desired P300 evoked potentials but also ongoing activity of the brain and muscular and/or ocular artifacts. As a result, the SNR is very low and the classification task (i.e. the character prediction) is not easy. The aim of this study is thus to provide a simple and unsupervised estimation of the evoked subspace (i.e. the subspace which contains most of the P300 evoked potentials) so that the P300 evoked potentials are enhanced by projecting the raw recorded EEGs on the estimated evoked subspace. The classification between target/non-target stimuli is so simplified leading to a faster spelling device as shown in Section IV.

A training database is used to estimate evoked subspace and then to train the classifier (Section III-B). It consists in a database for which the spelled symbols are known as well as the order of rows/columns intensifications and the corresponding stimulus onsets (i.e. beginning time of illumination).

The proposed method is based on two main ideas:

- the rare events in the oddball paradigm elicit the P300 component of the event-related potential (ERP),
- this synchronous response occupies a small spatial subspace of space spanned by the recorded EEG.

In other words, there exists a typical response synchronised with the target stimuli, and then this synchronous response can be enhanced by a spatial filtering. The proposed method is thus divided into two parts. The synchronous responses are first estimated for each sensor and these responses are then used to estimate spatial filters such that the evoked P300 potentials are enhanced.

Let $x_j(t)$ denote the EEG signal recorded by the j^{th} sensor at time index t and let $X \in \mathbb{R}^{N_t \times N_s}$ be the matrix of recorded EEG signals whose $(i,j)^{th}$ entry is $x_j(i)$. N_s is the number of sensors and N_t the number of temporal samples. Let $a_j(t)$ denote the ERP signal for the j^{th} sensor at time index t, and let $A \in \mathbb{R}^{N_e \times N_s}$ be the matrix of ERP signals whose $(i,j)^{th}$ entry is $a_j(i)$. N_e is the number of temporal samples of the ERP (typically, N_e is chosen to correspond to 600 milliseconds or one second).

The fact that the target stimuli elicit a P300 evoked potential leads to the following model

$$X = D\,A + N. \tag{1}$$

$D \in \mathbb{R}^{N_t \times N_e}$ is the Toeplitz matrix whose first column is defined such that $D_{\tau_k,1} = 1$, where τ_k is the stimulus onset of k^{th} target stimulus ($1 \leq k \leq K$, with K the total number of target stimuli) and such that all the other elements are null. $D\,A$ in (1) thus represents the synchronous response with target stimuli and matrix N the on-going activity of the user's brain as well as the artifacts. The least square estimation of response A is simply performed by

$$\hat{A} = \arg\min_A \left\| X - D\,A \right\|_2^2,$$

whose solution is given by

$$\hat{A} = \left(D^T D\right)^{-1} D^T X, \tag{2}$$

where $.^T$ is the transpose operator. Note that a classical epoching of matrix X to estimate A leads to

$$A^\dagger = D^T X, \tag{3}$$

which could be quite different from (2) if $(D^T D)^{-1}$ is quite different from a diagonal matrix. This is typically the case if synchronous response A extends over several consecutive stimuli: i.e. if $N_e \geq \min \Delta\tau_k$, where

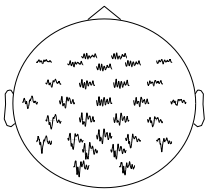
$$\Delta\tau_k = \tau_k - \tau_{k-1} \tag{4}$$

is the interval between two consecutive target stimuli. As copied out in Fig. 2, least squares estimation (2) leads to a very redundant solution which is confirmed by the principal components of $\hat{A}$ (Fig. 2(b)) since in this case the two greatest principal values explain 91% of the total variance of $\hat{A}$.

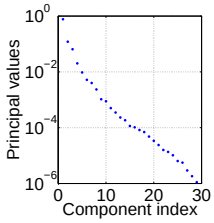
The second idea of the proposed method consists to estimate N_f spatial filters $\mathbf{u}_i$ ($1 \leq i \leq N_f \leq N_s$) such that the synchronous response is enhanced by the spatial filtering

$$X\,U = D\,A\,U + N\,U, \tag{5}$$

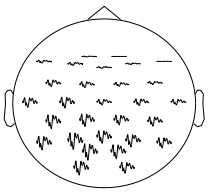
where $U \in \mathbb{R}^{N_s \times N_f}$ is the spatial filters matrix whose i^{th} column is $\mathbf{u}_i$. An intuitive solution should be first to perform



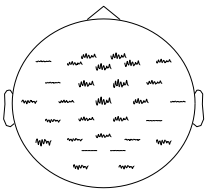
(a) Projection of $\hat{A}$ on user's scalp.



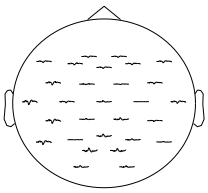
(b) Principal components of $\hat{A}$.



(c) Projection of the first principal component of $\hat{A}$ on user's scalp.



(d) Projection of the second principal component of $\hat{A}$ on user's scalp.



(e) Projection of the third principal component of $\hat{A}$ on user's scalp.

Fig. 2. Least squares estimation of P300 evoked potentials for the first subject of the recorded database (*cf.* Section III-A). Fig. 2(a): projection of $\hat{A}$ (2) on the user's scalp. Each plot corresponds to an ERP of 1 s and plots at the sensor position on the user's head. Fig. 2(b) shows the principal components of $\hat{A}$ (2). These principal components are normalised such that their sum is equal to one. Fig. 2(c), 2(d) and 2(e) projection of the first, second, and third principal component of $\hat{A}$ on the user's scalp, respectively.

a principal component analysis (PCA) of $\hat{A}$ (2) and then to project the recorded signals X on the N_f main components associated with the N_f largest principal values. By doing this thanks to the singular value decomposition (SVD) of $\hat{A}$

$$\hat{A} = \Sigma\,\Delta\,\Pi^T, \tag{6}$$

where Σ and Π are unitary matrices and Δ is a diagonal matrix with nonnegative diagonal elements in decreasing order, and by splitting these matrices into signal and noise subspaces

$$\begin{aligned}\Sigma &= \begin{bmatrix}\Sigma_s & \Sigma_n\end{bmatrix},\\ \Delta &= \begin{bmatrix}\Delta_s & 0\\ 0 & \Delta_n\end{bmatrix},\\ \Pi &= \begin{bmatrix}\Pi_s & \Pi_n\end{bmatrix},\end{aligned}$$

spatial filters U, which are defined as the projector on the signal subspace, are expressed as

$$U_{pca} = \Pi_s. \tag{7}$$

Moreover, one can then rewrite $\hat{A}$ as

$$\hat{A} = \Sigma_s\,\Delta_s\,\Pi_s^T + \Sigma_n\,\Delta_n\,\Pi_n^T. \tag{8}$$

and model (1) is finally formulated as

$$X = D\,A'_{pca}\,W_{pca}^T + N',$$

where $A'_{pca} = \Sigma_s\,\Delta_s$ is the synchronous response of reduced dimensions, $W_{pca} = \Pi_s$ is its spatial distribution over sensors and $N' = N + D\,\Sigma_n\,\Delta_n\,\Pi_n^T$. Even if the PCA enhances the evoked potentials, the major drawback of this solution is that the noise N is not directly taken into account to estimate the spatial filters. Indeed, filtered signals are then obtained by

$$\begin{aligned}\hat{S}_{pca} &= X\,U_{pca}\\ &= D\,A'_{pca} + N'\,U_{pca},\end{aligned}$$

where noisy term $N'\,U_{pca} = N\,U_{pca}$ (since Π_n and Π_s are orthogonal) could also be largely amplified compared to noisy term N in (1).

To overcome this problem, we propose to design spatial filters U such that the signal to signal plus noise ratio is maximised:

$$\hat{U} = \arg\max_U \frac{\mathrm{Tr}\big(U^T\hat{A}^T D^T\,D\hat{A}U\big)}{\mathrm{Tr}\big(U^T X^T\,XU\big)}, \tag{9}$$

where $U \in \mathrm{R}^{N_s\times N_f}$ is the spatial filter matrix whose i^{th} column is equal to $\mathbf{u}_i$ and $\mathrm{Tr}(\cdot)$ is the trace operator. By computing the QR factorisation [20] of X and D respectively and by replacing $\hat{A}$ by (2), criterion (9) can be expressed as

$$\hat{V} = \arg\max_V \frac{\mathrm{Tr}\big(V^T Q_X^T Q_D\,Q_D^T Q_X V\big)}{\mathrm{Tr}\big(V^T\,V\big)}, \tag{10}$$

with $V = R_X\,U$, $X = Q_X\,R_X$ and $D = Q_D\,R_D$, where Q_X and Q_D are orthogonal matrices, and R_X and R_D are upper triangular matrices respectively. Matrix $\hat{V}$ is thus obtained from the Rayleigh quotient (10) whose solution is the concatenation of N_f eigenvectors associated with the N_f largest eigenvalues of matrix $Q_X^T Q_D\,Q_D^T Q_X$ [20]. These vectors can be computed efficiently from the SVD of $Q_D^T Q_X$:

$$Q_D^T\,Q_X = \Phi\,\Lambda\,\Psi^T, \tag{11}$$

where $\Lambda \in \mathbb{R}^{N_s\times N_s}$ is the diagonal matrix of singular values sorted in descending order ($1 \geq \Lambda_{1,1} \geq \cdots \geq \Lambda_{N_s,N_s} \geq 0$), $\Phi \in \mathbb{R}^{N_e\times N_s}$ and $\Psi \in \mathbb{R}^{N_s\times N_s}$ are two column orthonormal matrices. Splitting these matrices into signal and noise subspaces

$$\begin{aligned}\Phi &= \begin{bmatrix}\Phi_s & \Phi_n\end{bmatrix},\\ \Lambda &= \begin{bmatrix}\Lambda_s & 0\\ 0 & \Lambda_n\end{bmatrix},\\ \Psi &= \begin{bmatrix}\Psi_s & \Psi_n\end{bmatrix},\end{aligned}$$

leads to

$$\hat{V} = \Psi_s.$$

The solution of criterion (9) is so obtained by

$$\hat{U} = R_X^{-1}\,\Psi_s. \tag{12}$$

Moreover, one can rewrite $\hat{A}$ (2) as

$$\hat{A} = R_D^{-1}\,\Phi_s\,\Lambda_s\,\Psi_s^T\,R_X + R_D^{-1}\,\Phi_n\,\Lambda_n\,\Psi_n^T\,R_X, \tag{13}$$

thanks to the QR decomposition of D and X, and by using SVD expression of $Q_D^T\,Q_X$ (11). Model (1) is finally expressed as

$$X = D\,A'\,W^T + N', \tag{14}$$

where

$$A' = R_D^{-1}\,\Phi_s\,\Lambda_s \quad (15)$$
$$W = R_X^T\,\Psi_s. \quad (16)$$

A' is the synchronous response of reduced dimensions, W is its spatial distribution over sensors, and $N' = N + D\,R_D^{-1}\,\Phi_n\,\Lambda_n\,\Psi_n^T\,R_X$ is the noise term, respectively. The I dimensional evoked subspace is thus defined by the I couples $\left(\hat{\mathbf{u}}_i, \hat{\mathbf{a}}'_i\right)$ defined by (12) and (15) respectively. Note that this final result is related to the canonical or principal angles notion [20], [21], [22], which is a generalisation of canonical correlation analysis (CCA) [23]. Indeed, it can be shown that the singular value decomposition of $Q_D^T\,Q_X$ provides the principal angles whose cosines are the singular values $\Lambda_{i,i}$ and the associated singular vectors pair (ϕ_i, ψ_i) recursively minimises the quadratic error, for $i = 1, \cdots, N_s$:

$$(\phi_i, \psi_i) = \arg \min_{\substack{\|Q_X\psi\|_2 = \|Q_D\phi\|_2 = 1 \\ Q_X\psi \perp \{Q_X\psi_1, \cdots, Q_X\psi_{i-1}\} \\ Q_D\phi \perp \{Q_D\phi_1, \cdots, Q_D\phi_{i-1}\}}} \|Q_X\psi - Q_D\phi\|_2^2.$$

In this case, $\hat{\mathbf{a}}_i$ is obtained by $R_D^{-1}\phi_i$.

The enhanced signals are then computed by

$$\hat{S} = X\,\hat{U}, \quad = D\,A' + N\,R_X^{-1}\,\Psi_s. \quad (17)$$

Finally, the algorithm to estimate an I dimensional evoked subspace is summarised in Algorithm 1 and is denoted xDAWN regarding model structure (14). Note that the computation of synchronised responses (2) is finally unnecessary to estimate enhancing spatial filters $\hat{\mathbf{u}}_i$ (12).

Algorithm 1 xDAWN algorithm to estimate evoked subspace.

1: Compute QR factorisation of $X \;\Rightarrow\; X = Q_X\,R_X$
2: Compute QR factorisation of $D \;\Rightarrow\; D = Q_D\,R_D$
3: Compute SVD of $Q_D^T\,Q_X \;\Rightarrow\; Q_D^T\,Q_X = \Phi\,\Lambda\,\Psi^T$
4: Select the I couples of singular vectors $\left(\Phi_i, \Psi_i\right)$ associated with the I largest singular values λ_i
5: Finally $\forall\; 1 \leq i \leq I,\; \left(\hat{\mathbf{u}}_i, \mathbf{a}'_i\right) = \left(R_X^{-1}\,\psi_i, R_D^{-1}\,\phi_i\,\lambda_i\right)$
6: Estimate sources: $\forall\; 1 \leq i \leq I,\; \hat{s}_i(t) = \hat{\mathbf{u}}_i^T\,\mathbf{x}(t)$

III. Application to the P300 Speller BCI

In this section, the data acquisition and the preprocessing of the recorded signals are described in Subsection III-A, while the BCI classification problem is presented in Subsection III-B.

A. Data acquisition and preprocessing

1) Data acquisition: Three healthy male subjects (two French and a German, ages 29, 31 and 31 respectively) participated voluntarily in the experiment. They were all free of neurological diseases and had no previous experience with the P300 speller paradigm.

EEG activity was recorded from 29 Ag/AgCl scalp electrodes placed at standard positions of an extended 10-20 international system referenced to the nose and grounded to the forehead. Horizontal and vertical electrooculograms (EOG) were recorded from the right eye. All impedances were kept below 10 kOhms throughout the experiments. Signals were amplified and digitised at a rate of 500 Hz using a BrainAmp amplifier (BrainProducts GmBH, Munich). The EEG was collected and stored using the BCI2000 system with the P300 speller scenario [24].

The subjects were seated in a comfortable chair at a distance of 60 cm from the computer screen in a quiet room. They were watching a 6×6 matrix of letters as shown in Fig. 1(a). The matrix subtended 10.2 °H × 11.9 °W of angular view. The experiment paradigm was controlled by the BCI2000 software and was similar to the one used for the BCI Competition 2003 - P300 Speller data set [25]. Each run corresponded to one word (a French word, or a German word for subject 3) and each word could be composed of two to six characters. Before each run, the entire word to be spelt was indicated at the top of the display. Subjects were asked to focus on the current letter (which was shown after the word in parentheses) and to mentally count the number of times this letter was intensified. Each row and column in the matrix was randomly intensified for 100 ms and the delay between two consecutive intensifications was 80 ms thus leading to an interstimulus interval (ISI) of 180 ms. For each letter, the 12 columns and rows were intensified 15 times. There was a 2.5 second period between each character of a run, allowing the subject to focus the attention on the next character. The subject could make a short break after each run. In total, there were at least about fifty characters for each subject distributed among several runs (73 letters for 19 runs for subject 1, 63 letters for 14 runs for subject 2, and 66 letters for 16 runs for subject 3).

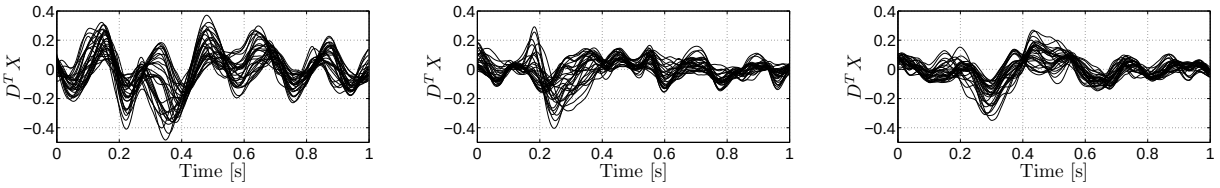
2) Pre-processing: Before estimating the spatial filters by the xDAWN algorithm to enhance the P300 evoked potentials (Section II) the following pre-processing stages were applied. The data were first filtered by a fourth order forward-backward Butterworth bandpass filter. Cut off frequencies were set to 1 Hz and 20 Hz. For each sensor, the bandpass filtered signals were then normalised so that they had zero mean and a standard deviation equal to one. The temporal length of the synchronous response (2) was chosen to be one second.

B. BCI classification

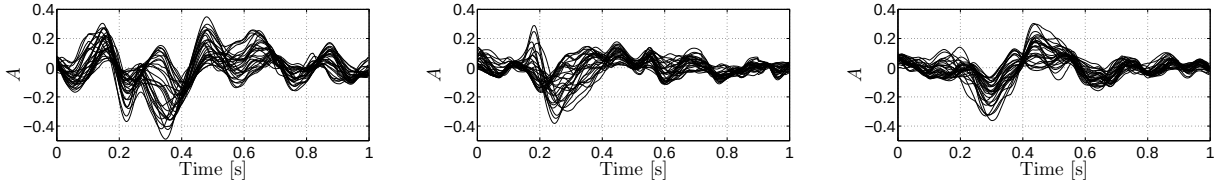
In the P300 speller BCI problem, the spelled character is identified by the detection of a P300 evoked potential related to a given row and to a given column illuminations for each sequence. Feature vector $\mathbf{p}_j$ corresponding to the j^{th} illumination is given by the concatenation of I epochs of estimated sources $\mathbf{f}_i^{(j)} = [f_i^{(j)}(0), \cdots, f_i^{(j)}(N_e)]^T$ defined by

$$f_i^{(j)}(t) = (\hat{s}_i(t) \times \Pi_{N_e}(t - \tau_j)) * \delta(t + \tau_j), \quad (18)$$

where $s_i(t)$ is the estimated sources (17) obtained by the proposed 'xDAWN' algorithm (Section II), τ_j is the stimulus onset of the j^{th} illumination and $\Pi_{N_e}(t)$ is the boxcar function equal to 1 on its support $[0, N_e]$ and equal to 0 elsewhere (typically N_e is chosen to correspond to 600 milliseconds or



(a) Subject 1: classical epoching (b) Subject 2: classical epoching (c) Subject 3: classical epoching



(d) Subject 1: least square estimation (e) Subject 2: least square estimation (f) Subject 3: least square estimation

Fig. 3. Estimations of P300 evoked potentials for the three subjects. First row (Fig. 3(a), 3(b), 3(c)) corresponds to a classical epoching estimation $A^\dagger$ (3). Second row (Fig. 3(d), 3(e), 3(f)) corresponds to the proposed estimation $\hat{A}$ (2) obtained by least squares minimisation. The evoked potentials for all the sensors are stacked on the same plot.

one second). Thus $\mathbf{p}_j$ is defined by

$$\mathbf{p}_j = \left[{\mathbf{f}_1^{(j)}}^T, \cdots, {\mathbf{f}_I^{(j)}}^T \right]^T. \tag{19}$$

Moreover, let t_j denote the associated class with j^{th} illumination, $t_j = 1$ if the j^{th} illumination contains the spelled symbol and $t_j = 0$ otherwise.

Among the proposed classifiers for BCIs, Bayesian linear discriminant analysis (BLDA) [2], [26] is chosen since it was proved to be efficient and was fully automatic (i.e. no hyper-parameters to adjust) [2]. It aims at finding, using a Bayesian framework, a discriminant vector $\mathbf{w}$ such that $\mathbf{w}^T\,\mathbf{p}$ is as close as possible to the class t associated with the corresponding feature vector $\mathbf{p}$. This discriminant vector $\mathbf{w}$ is thus estimated from the set of couples $\{\mathbf{p}_j, t_j\}_j$ obtained from the training database.

Let $h_{r/c}$ denote the output of the classifier corresponding to the illumination of row/column r/c. The score $H_{r/c}(k)$ of row/column r/c after k repetition is given by

$$H_{r/c}(k) = H_{r/c}(k-1) + h_{r/c}, \tag{20}$$

where $H_{r/c}(0) = 0$. After the k^{th} repetition, the recognised symbol is the one with maximal row and column scores.

IV. Results

In this section, evoked subspace estimation results are presented in Subsection IV-A while the complete P300 speller BCI results are presented in Subsection IV-B.

A. P300 subspace estimation

In this set of experiments, the proposed method to estimate the evoked subspace (Section II) was applied to the recorded database.

Estimated synchronous responses $\hat{A}$ (2) for each subject are shown in Figure 3 and compared with the classical epoching estimation (3). For each subfigure, the 29 evoked potentials corresponding to the 29 sensors are stacked on the same plot. One can see that, for each subject, these estimations include a periodic component at 5.6 Hz. This component, whose frequency is the inverse of the ISI (180 ms), is thus due to the visual stimulation (i.e. the intensifications of the rows/columns). The proposed estimation of synchronous response (2) slightly improves the result compared to the classical epoching (3): for instance, one can see that the residual oscillations in Figures 3(d), 3(e), and 3(f) are slightly reduced compared to Figures 3(a), 3(b), and 3(c), respectively. This can mainly seen from 600ms to 1s. These differences, which are due to the fact that $D^T D$ is not diagonal, are the expression of the fact that the temporal length of the estimated evoked potentials (here chosen equal to one second) is sometimes greater than the interval between two consecutive target stimuli $\Delta\tau_k$ (4). Indeed, as shown in Fig. 4, around 50 percent of two consecutive target stimuli leads to overlapped synchronous responses. These figures finally show the importance of taking into account this overlapping for the estimation of the synchronous response.

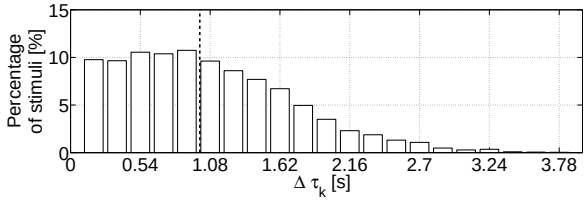
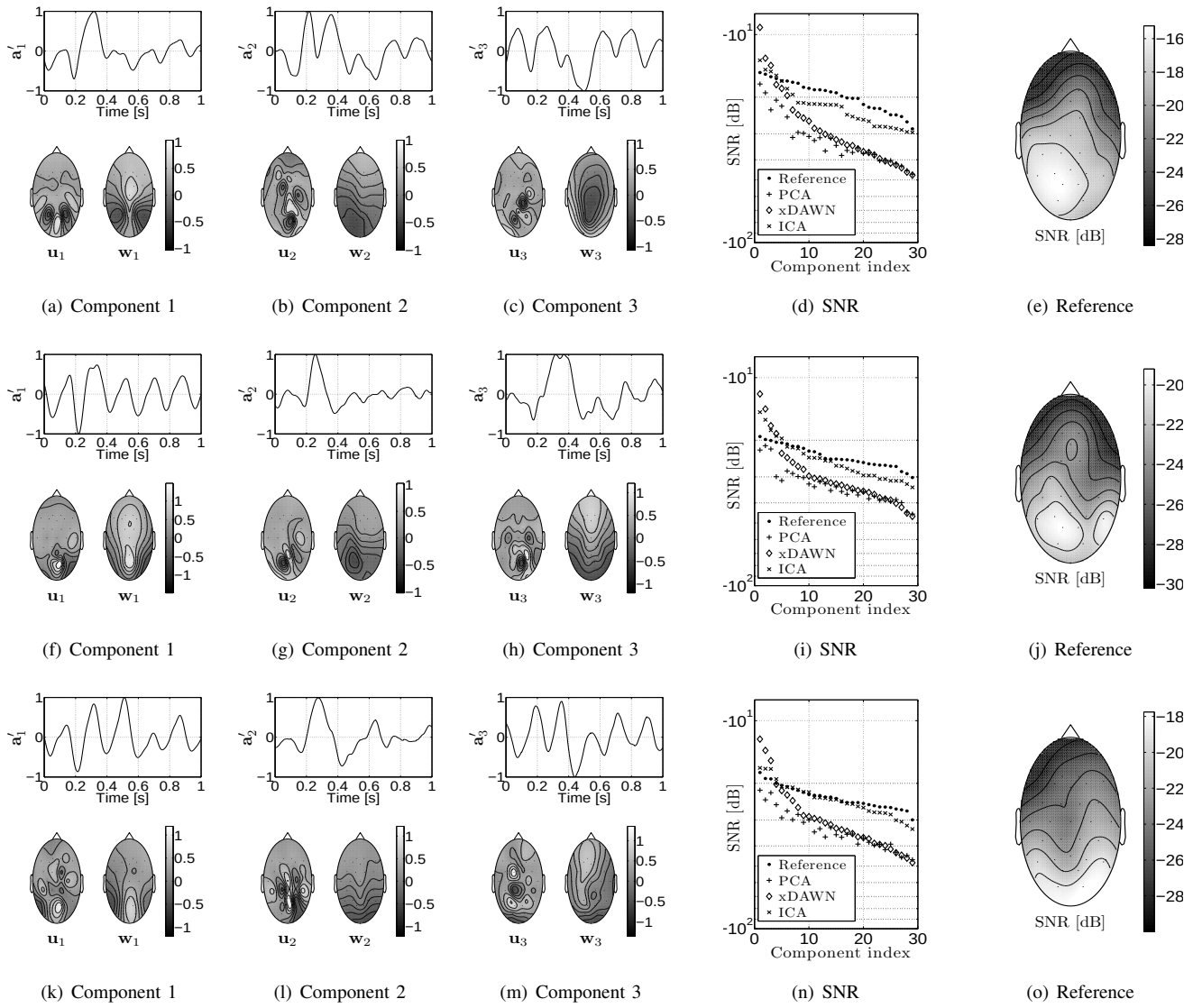


Fig. 4. Histogram of interval between two consecutive target stimuli $\Delta\tau_k$ (4) for the three subjects. The vertical dashed line is located at one second which is the temporal length of the estimated synchronous response A (2).

The proposed xDAWN algorithm (*cf.* Algorithm 1) is then applied to estimate the P300 subspace for the three subjects. The results are presented in Fig. 5, where we plot the enhanced synchronised response $\mathbf{a}_i'$ (15), the spatial filter $\mathbf{u}_i$ (12), and



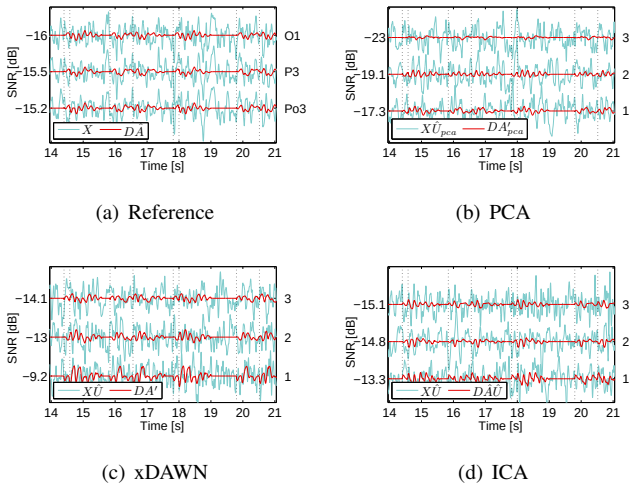
(a) Component 1 (b) Component 2 (c) Component 3 (d) SNR (e) Reference

(f) Component 1 (g) Component 2 (h) Component 3 (i) SNR (j) Reference

(k) Component 1 (l) Component 2 (m) Component 3 (n) SNR (o) Reference

Fig. 5. P300 subspace estimation for the three subjects. Each row corresponds to one subject: Fig. 5(a), 5(b), 5(c), 5(d) and 5(e) are related to subject 1, Fig. 5(f), 5(g), 5(h), 5(i) and 5(j) are related to subject 2, and Fig. 5(k), 5(l), 5(m), 5(n) and 5(o) are related to subject 3. For each subject, Fig. 5(a), 5(b), 5(c), Fig. 5(f), 5(g), 5(h), and Fig. 5(k), 5(l), 5(m) show the triplet: enhanced synchronised response $\mathbf{a}'_i$ (15), spatial filter $\mathbf{u}_i$ (12) and spatial distribution $\mathbf{w}_i$ (16) (top, bottom left and bottom right, respectively) for the three first components estimated by the xDAWN algorithm. Fig. 5(d), 5(i), 5(n) show the SNR obtained by different methods of enhancement for the three subjects: 'reference' means no enhancement ($\hat{U} = I$ and $\hat{A}' = \hat{A}$), 'PCA' corresponds to an enhancement obtained by PCA thanks to (7) and (8), 'xDAWN' results are obtained by the xDAWN algorithm (12) and (15), and 'ICA' results refer to spatial filters $\hat{U}$ estimated by the JADE blind source separation algorithm [27], respectively. Note that 'reference' and 'ICA' SNR are sorted in descending order of SNR, while 'PCA' and 'xDAWN' are sorted in descending order of principal values Δ (6) and in descending order of singular values Λ (11), respectively. Finally, Fig. 5(e), 5(j) and 5(o) show the projection of SNR related to 'reference' over the subjects' scalp.

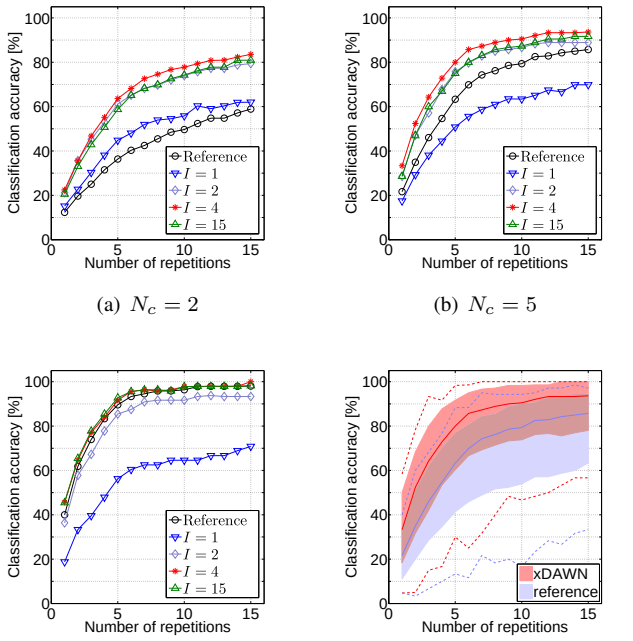
the spatial distribution $\mathbf{w}_i$ (16) for the three first estimated components, as well as the signal-to-signal plus noise ratio (SNR) provided by different methods of enhancement, which is computed as the ratio between the power of the signal defined as $D\mathbf{a}'_i$ and the power of signal plus noise defined as $X\mathbf{u}_i$. Let 'reference' denotes the input SNR computed before spatial enhancement (i.e. $\mathbf{a}'_i = \hat{\mathbf{a}}_i$ where $\hat{\mathbf{a}}_i$ is the i^{th} column of $\hat{A}$ (2) and $\mathbf{u}_i = 1_i$, where 1_i denotes the vector of zeros excepted the i^{th} component equal to 1). First of all, as claimed in the introduction of Section II the input SNR is very low: typically included between -15 dB and -30 dB, depending of the subject and the sensors (see 'reference' in Figs. 5(d), 5(i), and 5(n)). Moreover, even if the PCA analysis (8) allows decomposition of $\hat{A}$ (2) such that its first components combine the maximum of signal, the associated spatial enhancing filters U_{pca} (7) would not provide good output SNRs, as shown in Figs. 5(d), 5(i), and 5(n). Indeed, the output SNRs ('PCA') are even lower than input SNRs ('reference'): this can be explained by the fact that the noise N is not taken into account to design these spatial filters, leading to a poor estimation of enhancing filters. Moreover, this latter fact also leads to not ensure that the first principal components have the best output SNRs: the descending order of SNR is quite different from the descending order of principal values Δ (6) (Fig. 5(d), 5(i) and 5(n)). On the contrary, the proposed xDAWN algorithm provides quite good estimation of spatial enhancing filters U



(a) Reference (b) PCA

(c) xDAWN (d) ICA

Fig. 6. Enhancement of the synchronised response with the target stimuli for subject 1. In the four figures 6(a), 6(b), 6(c), and 6(d) the blue lines correspond to the enhanced signals $X\hat{U}$, and the red lines correspond to the estimated synchronised model DA' estimated by different methods: Fig. 6(a) without enhancement ($\hat{U} = I$ and $A' = \hat{A}$), Fig. 6(b) by PCA using (7) and (8), Fig. 6(c) by the xDAWN algorithm (12) and (15), and finally Fig. 6(d) by the JADE algorithm [27], respectively. The vertical dotted lines correspond to the target stimulus onsets. The left y-axis refers to the SNR (in dB), while the right y-axis refers to the sensor ('reference') or to the index component ('PCA', 'xDAWN' and 'ICA').



(a) $N_c = 2$ (b) $N_c = 5$

(c) $N_c = 20$ (d) Distribution: $N_c = 5,\ I = 4$

Fig. 7. Performance curves. Fig. 7(a), 7(b) and 7(c) show the percentage of good symbol prediction versus the number of repetitions of the same symbol for different numbers of training symbols $N_c = 2,\ 5$or 20, respectively. In each figure, the legend refers to the reference method ('reference') or to the number of kept xDAWN components (I) with the xDAWN algorithm. Fig. 7(d) shows the distribution of the obtained results with $N_c = 5$ and $I = 4$ by the reference method ('reference') and by the xDAWN algorithm ('xDAWN'). The dotted lines correspond to the minimum and maximum classification accuracies. The continuous lines show the median values, while the surfaces extend from 5% to 95% of the classification accuracies.

since the output SNR ('xDAWN') of the first components is larger than the best input SNR ('reference'). The improvement is thus included between 1.2 dB to 7.2 dB for the three first xDAWN components compared to the best input SNR for each subject: by only considering the first xDAWN component, the gain is from 5 dB to 7.2 dB compared to the best input SNR for each subject. Finally, one can see that spatial filters $\hat{U}$ estimated by the JADE blind source separation algorithm [27] (denoted 'ICA') improve the output SNRs compared with no spatial enhancement ('reference'). Indeed, the first 'ICA' components have a higher SNR than the first 'reference' components. However, the best 'ICA' output SNRs are lower than the best SNRs provided by the 'xDAWN' algorithm: the proposed 'xDAWN' algorithm provides a better estimation of the P300 evoked potentials. It is also quite interesting to note that the best input SNRs are related with sensors mainly located at the back of the head rather than sensors located at the top of the head (Fig. 5(e), 5(j) and 5(o)). Finally, for each subject, the three first xDAWN components are plotted (Fig. 5(a), 5(b), 5(c), and 5(f), 5(g), 5(h), and 5(k), 5(l), 5(m)). In some of these components, the presence of a (strong) wave at 5.6 Hz suggests that a specific synchronised response with the visual stimulation should be introduced in model (14). Moreover, it is quite interesting to note that for the three subjects the first xDAWN components are mainly located at the back of the head (e.g. see $\mathbf{w}_1$ and $\mathbf{w}_2$).

Fig. 6 illustrates the enhanced noisy signals $X\mathbf{u}_i$ and the model $D\mathbf{a}_i$ by different methods: without enhancement and with spatial filters provided by PCA, by the proposed xDAWN algorithm or by the JADE blind source separation algorithm [27]. As already mentioned above, the PCA decomposition provides poor estimation of enhancing spatial filters leading thus to a SNR lower than without spatial filtering (Fig. 6(a) and 6(b)). On the other hand, using an appropriate spatial enhancement ('ICA' or 'xDAWN') may improve the SNR (Fig. 6(c) and 6(d)). the xDAWN decomposition leads to really enhance the evoked potentials as shown in Fig. 6(c): even if the SNR is still small (-9.2 dB for the first xDAWN component), the evoked potentials might be seen on the enhanced signal $X\hat{\mathbf{u}}_1$.

B. BCI P300 speller classification

In this set of experiments, the results of BCI classification obtained by different methods are compared: spatial filters U are estimated by different algorithms while the linear classifier $\mathbf{w}$ is estimated by BLDA for each case. In each experiment, N_c symbols are used to train spatial filters U and to train linear classifier $\mathbf{w}$. The BCI performance is then tested on all the others symbols of the same subject. Note that after epoching estimated sources $\hat{s}_i(t)$ (17), $\mathbf{f}_i^{(j)}$ is decimated with a factor 10 before computing feature vector $\mathbf{p}_j$ (19).

The first method uses no spatial filters (i.e. U is chosen equal to the identity) and is denoted 'reference': this corresponds to a simple BLDA on time samples. The second method, denoted 'xDAWN', estimates spatial filters U thanks to the proposed xDAWN method (Section II). The averaged results (for the three subjects) of BCI classification are presented in Fig. 7, which shows the percentage of success

versus the number of repetitions of the same symbol for different numbers of training symbol N_c. For each subject and for each configuration (N_c, I), where I is the number of estimated xDAWN components, 100 different allocations between training and testing database are randomly chosen among the entire set of symbols (i.e. 73 symbols for subject 1, 63 symbols for subject 2 and 66 symbols for subject 3). The curves presented in Fig. 7(a), 7(b) and 7(c) are thus the average results for all 300 different configurations: three subjects times 100 training databases randomly chosen per subject.

As expected, for each configuration the number of repetitions increases the performance. The same positive correlation is found with respect to the number of training symbols N_c. It is quite interesting to consider the case where only a few number of symbols (i.e. $N_c \leq 5$) are used to train both spatial filters U and linear classifier $\mathbf{w}$. Indeed, the proposed 'xDAWN' method outperforms the 'reference' method, except when only the first xDAWN component is considered (Fig. 7(a) and 7(b)). However, with two (or more) main xDAWN components, the proposed method provides better performance than the 'reference' method. Moreover, it seems that a limited number of xDAWN components (around four) is sufficient to mainly gather the evoked potentials: for each configuration, the best performance are obtained by only using four xDAWN components. This latter result confirms the SNR results (Fig. 5(d), 5(i), and 5(n)) where a limited number of xDAWN components differ with a quite higher SNR than other components. Finally, using too many xDAWN components (15 for instance) sightly decreases the performance. This can be explained by the fact that these additional components have poor SNR thus corrupting the feature vector $\mathbf{p}_j$ (19) with additional components $\mathbf{f}_i^{(j)}$ (18) that contain more noise than signal. Furthermore, Fig. 7(d) shows the distribution of classification accuracy, corresponding to the 300 configurations with five training symbols ($N_c = 5$) and with four xDAWN components ($I = 4$). One can see the benefit of using xDAWN algorithm to enhance evoked potentials: the 'xDAWN' results are better than the 'reference' results. For instance, with only five repetitions of the same symbols, 95% percent of the 300 tested configurations provide more than 60% of classification accuracy with the 'xDAWN' method, while only 50% of 'reference' results provide more than 60% of classification accuracy. Finally, with five repetitions (i.e. about 11 seconds), more than 50% of the 'xDAWN' results insure more than 80% of classification accuracy.

In the last set of experiments (Fig. 8) we compare the classification accuracy provided by different spatial filtering methods with five training symbols ($N_c = 5$). 'reference' means that no spatial filtering enhancement is performed: it is a simple BLDA on time samples. 'ICA' refers to spatial filters estimating by JADE [27] blind source separation (BSS) algorithm, and the 15 components kept after BSS are chosen thanks to their high SNR: these 15 components provide the best classification accuracy. 'PCA' corresponds to a principal component analysis of synchronous response (2) given by (7) and (8): the four first components where selected since they represent more than 95% of the total variance of $\hat{A}$ (2). Finally

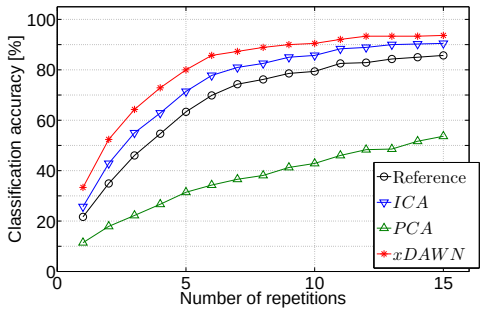


Fig. 8. Comparison of classification accuracy achieved by different spatial enhancement with five training symbols ($N_c = 5$). 'reference' means that no spatial enhancement is performed before BLDA classifier, 'ICA' refers to the JADE blind source separation algorithm [27], PCA corresponds to an enhancement obtained by PCA thanks to (7) and (8), while xDAWN results are obtained by the xDAWN algorithm (12) and (15), respectively.

'xDAWN' refers to the proposed method (Algorithm 1) and four components where selected since the previous set of experiments show that this choice provides the best classification accuracy. Not surprisingly, 'PCA' method provides poor classification accuracy due to the weakness of the method which does not take into account the noise to estimate the spatial filters. Moreover, as one can expected 'ICA' and 'xDAWN' enhancements before BLDA classifier improve the classification accuracy compared to a simple BLDA classifier ('reference'). Finally among the tested methods, the proposed 'xDAWN' algorithm provides the best classification accuracy: with only five symbol repetitions it achieves 80% of classification accuracy while 'ICA' achieves 71%.

V. Conclusions and Perspectives

In this paper, a new unsupervised method to enhance evoked response by target stimuli in an oddball paradigm was presented. Only given the time indexes of rows/columns intensifications, the proposed algorithm estimates the main components of the P300 subspace by providing the best SNR. It was shown to efficiently improve the quality of the evoked responses by taking into account the signal and the noise, as opposed to principal component analysis, which only considers the signal. Using this method to enhance P300 subspace before the BCI classification task speeds up the BCI since less words are required to train the spatial filters and the linear classifier, given a certain percentage of good symbol prediction. Moreover, using this spatial enhancement significantly reduces the dimension of the feature vector used to predict words.

To further improve the performance of the P300-speller BCI, additional work should be considered. For instance, to better estimate the response evoked by target stimuli, a multi-stimuli model should be deemed by assuming that all the non-target stimuli – as well as the target stimuli – evoked specific responses. Finally, sensor selection should also be considered, in order to drastically reduce the number of required EEG electrodes, leading thus to a more ergonomic BCI.

REFERENCES

[1] L. A. Farwell and E. Donchin, "Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials," *Electroencephalography and Clinical Neurophysiology*, vol. 70, no. 6, pp. 510–523, December 1988.

[2] U. Hoffmann, J.-M. Vesin, T. Ebrahimi, and K. Diserens, "An efficient p300-based brain-computer interface for disabled subjects," *Journal of Neuroscience Methods*, vol. 167, no. 1, pp. 115–125, January 2008.

[3] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain-computer interfaces for communication and control." *Clinical Neurophysiology*, vol. 113, no. 6, pp. 767–791, June 2002.

[4] N. Birbaumer, "Brain-computer-interface research: Coming of age," *Clinical Neurophysiology*, vol. 117, no. 3, pp. 479–483, March 2006.

[5] L. R. Hochberg, M. D. Serruya, G. M. Friehs, J. A. Mukand, M. Saleh, A. H. Caplan, A. Branner, D. Chen, R. D. Penn, and J. P. Donoghue, "Neuronal ensemble control of prosthetic devices by a human with tetraplegia," *Nature*, vol. 442, no. 7099, pp. 164–171, July 2006.

[6] J. R. Wolpaw, D. J. McFarland, and T. M. Vaughan, "Brain-Computer Interface Research at the Wadsworth Center," *IEEE Transactions on Rehabilitation Engineering*, vol. 8, no. 2, pp. 222–226, June 2000.

[7] G. Pfurtscheller, C. Neuper, G. R. Müller, B. Obermaier, G. Krausz, A. Schlögl, R. Scherer, B. Graimann, C. Keinrath, D. Skliris, M. Wörtz, G. Supp, and C. Schrank, "Graz-BCI: State of the Art and Clinical Applications," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 11, no. 2, pp. 177–180, June 2003.

[8] E. Donchin, K. Spencer, and R. Wijesinghe, "The mental prosthesis: assessing the speed of a P300-based brain-computer interface," *IEEE Transactions on Rehabilitation Engineering*, vol. 8, no. 2, pp. 174–179, June 2000.

[9] J. Bayliss and D. Ballard, "A virtual reality testbed for brain-computer interface research," *IEEE Transactions on Rehabilitation Engineering*, vol. 8, no. 2, pp. 188–190, June 2000.

[10] H. Serby, E. Yom-Tov, and G. Inbar, "An improved P300-based brain-computer interface," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 13, no. 1, pp. 89–98, March 2005.

[11] A. Rakotomamonjy and V. Guigue, "BCI Competition III: Dataset II-Ensemble of SVMs for BCI P300 Speller," *IEEE Transactions on Biomedical Engineering*, vol. 55, no. 3, pp. 1147–1154, March 2008.

[12] K.-R. Muller, S. Mika, G. Ratsch, K. Tsuda, and B. Scholkopf, "An introduction to kernel-based learning algorithms," *IEEE Transactions on Neural Networks*, vol. 12, no. 2, pp. 181–201, March 2001.

[13] U. Hoffmann, G. Garcia, J.-M. Vesin, K. Diserens, and T. Ebrahimi, "A Boosting Approach to P300 Detection with Application to Brain-Computer Interfaces," in *IEEE EMBS Conference on Neural Engineering*, 2005.

[14] C. Jutten and J. Hérault, "Blind separation of sources. Part I: An adaptive algorithm based on a neuromimetic architecture," *Signal Processing*, vol. 24, no. 1, pp. 1–10, July 1991.

[15] J.-F. Cardoso, "Blind signal separation: statistical principles," *Proceedings of the IEEE*, vol. 86, no. 10, pp. 2009–2025, October 1998.

[16] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. New York: Wiley, 2001.

[17] N. Xu, X. Gao, B. Hong, X. Miao, S. Gao, and F. Yang, "BCI Competition 2003–Data Set IIb: Enhancing P300 Wave Detection Using ICA-Based Subspace Projections for BCI Applications," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 6, pp. 1067–1072, June 2004.

[18] B. Rivet and A. Souloumiac, "Extraction de potentiels évoqués P300 pour les interfaces cerveau-machine," in *Proc. GRETSI*, Troyes, France, September 2007, pp. 625–628, (in French).

[19] ——, "Subspace estimation approcah to P300 detection and application to Brain-Computer Interface," in *Proc. Int. Conf. IEEE Engineering in Medicine and Biology Society (IEEE EMBC)*, Lyon, France, August 2007, pp. 5071–5074.

[20] G. H. Golub and C. F. Van Loan, *Matrix Computation*, 3rd ed. Johns Hopkins University Press, 1996.

[21] G. H. Golub and H. Zha, "The Canonical Correlations of Matrix Pairs and Their Numerical Computation," Stanford University, Tech. Rep. NA-92-13, 1992.

[22] ——, "Perturbation analysis of the canonical correlations of matrix pairs," *Linear Algebra and its Applications*, vol. 210, pp. 3–28, October 1994.

[23] H. Hotelling, "Relations between two set of variables," *Biometrika*, vol. 28, no. 3–4, pp. 321–377, 1936.

[24] G. Schalk, D. J. McFarland, T. Hinterberger, N. Birbaumer, and J. R. Wolpaw, "BCI2000: A General-Purpose Brain-Computer Interface (BCI) System," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 6, pp. 1034–1043, June 2004.

[25] B. Blankertz, K.-R. Müller, G. Curio, T. M. Vaughan, G. Schalk, J. R. Wolpaw, A. Schlögl, C. Neuper, G. Pfurtscheller, T. Hinterberger, M. Schröder, and N. Birbaumer, "The BCI Competition 2003: Progress and Perspectives in Detection and Discrimination of EEG Single Trials," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 6, pp. 1044–1051, June 2004.

[26] D. J. C. MacKay, "Bayesian Interpolation," *Neural Computation*, vol. 4, no. 3, pp. 415–447, 1992.

[27] J.-F. Cardoso and A. Souloumiac, "Blind beamforming for non Gaussian signals," *IEE Proceedings-F*, vol. 140, no. 6, pp. 362–370, December 1993.

Bertrand Rivet was graduated from the École Normale Supérieure de Cachan, France. He received the Agrégation de Physique Appliquée in 2002, the Master's degree from the University of Paris-XI, France, in 2003 and the PhD degree from Grenoble Institute of Technology (GIT), France, in 2006.

He is currently an Associate Professor in signal processing with PHELMA and a member of GIPSA-lab, GIT, France. His research concerns biomedical signal processing, audiovisual speech processing and blind source separation.

Antoine Souloumiac was born in Bordeaux, France, in 1964. He received the M.S. degree and the Ph.D. degree in signal processing from the École Nationale Supérieure des Télécommunications, Paris, France, in 1987 and 1993, respectively.

He is currently with the Department of Technologies of Sensors and Signal (DETeCS), Commissariat à l'Énergie Atomique (CEA), Saclay, France. From 1993 until 2001, he was a Research Scientist with Schlumberger, Montrouge, France. His research interests are in the area of statistical signal processing and its applications, with emphasis on point processes, biomedical signal processing, and independent component analysis (ICA) or blind source separation (BSS).

Virginie Attina received a Ph.D. degree in cognitive sciences from the Institut National Polytechnique de Grenoble, Grenoble, France, in 2005.

She is currently a postdoctoral research fellow at the Brain Dynamics and Cognition unit (U821) (National Institute of Health and Medical Research), Lyon, France. Her general research interests concern human behavior, cognition and communication and methods of rehabilitation of disabled persons.

Guillaume Gibert received a Ph.D. degree in signal, image and speech processing from the Institut National Polytechnique de Grenoble, Grenoble, France, in 2006.

He is currently a postdoctoral research fellow at the Brain Dynamics and Cognition unit (U821) (National Institute of Health and Medical Research), Lyon, France. His general research interests concern signal processing methods applied to commnication systems for disabled persons.