Subject-specific channel selection for classification of motor imagery electroencephalographic data

Yuan Yang, Olexiy Kyrgzov, Joe Wiart, Isabelle Bloch

SUBJECT-SPECIFIC CHANNEL SELECTION FOR CLASSIFICATION OF MOTOR IMAGERY ELECTROENCEPHALOGRAPHIC DATA

Yuan Yang[1,3]*, Olexiy Kyrgyzov*[1,3]*, Joe Wiart*[2,3]*, Isabelle Bloch*[1,3]

[1]Institut Mines-Télécom, Télécom ParisTech/CNRS LTCI, Paris, France
[2]Orange Labs R&D, Issy les Moulineaux, France
[3]Whist Lab, France

ABSTRACT

Brain-computer interfaces (BCIs) are systems that record brain signals and then classify them to generate computer commands. Keeping a minimal number of channels (electrodes) is essential for developing portable BCIs. Unlike existing methods choosing channels without optimization of time segment for classification, this work proposes a novel subject-specific channel selection method based on a criterion derived from Fisher's discriminant analysis to realize the parametrization of both time segment and channel positions. The experimental results show that the method can efficiently reduce the number of channels (from 118 channels to no more than 11), and shorten the training time, without a significant decrease of classification accuracy on a standard dataset.

***Index Terms*—** Brain computer interfaces, electroencephalography, biomedical signal processing, machine learning

1. INTRODUCTION

Brain-computer interface (BCI) systems capture subject's intentions by recording his brain signals and then translate them into commands to build a direct communication between brain and computer [1]. The most popular brain signal used in BCI is electroencephalogram (EEG) because of its low cost and high time resolution [2]. EEG studies show that imaginary movements of different body parts can cause a power attenuation of sensorimotor rhythms, i.e. μ and β rhythms, called event-related desynchronization (ERD), at corresponding "active" cortex areas [3]; meanwhile, a power enhancement of sensorimotor rhythms called event-related synchronization (ERS) might be observed at other "idling" areas [3]. As the left brain controls the right body, the right hand typically elicits ERD at hand representation area of the left brain, while a foot movement may cause an ERS at hand representation areas of one or both side(s) of the brain [3]. Thus, motor imagery of different body parts can be identified by classifying ERD/ERS patterns, which gives birth to motor imagery BCI [2]. However, ERD/ERS patterns are typically short-lasting (half to few seconds) and occur in specific frequency bands [4]. The poor signal-to-noise ratio (SNR) of raw EEG signals and the mixture of different sources of brain activities (e.g. visual-related activities and motor imagery) make it difficult to extract discriminative features for BCI classification [5]. Although multi-channel recording with a large number of channels (e.g. 118) and spatial filtering algorithms (e.g. common spatial patterns (CSP) [5]) can improve the SNR and extract discriminative features from overlapping signals, this setting reduces the portability and practicability of BCI so that it represents a main drawback for final users [4]. Moreover, their effectiveness depends on the choice of the frequency band and the time segment of the EEG data [5].

To develop an easy to use system, several algorithms were proposed to reduce the number of channels in BCI [6, 7, 8]. However, they simply addressed the issue of channel selection by spatial information, disregarding the potential impact of time and frequency information. As a result, the optimal combination of time, frequency and channel position may not be achieved in a BCI design. Although a recent study showed that a broad frequency band (8-30Hz) that covers both μ (8-12Hz) and β (18-25Hz) bands can generally be used when employing a feature, called time domain parameters (TDPs), the existing methods mainly work with the popular band power (BP) feature, which is sensitive to frequency band and time segment.

Here, we propose a channel selection method for TDP features. In this method, a novel criterion based on Fisher's dicriminant analysis is proposed to measure the discrimination power of TDP features extracted from different channels and different time segments, so as to find the optimal time segment and subset of channels for BCI design. The paper is organized as follows: Section 2 describes the experimental data. A brief introduction to TDP is given in Section 3. A novel criterion for measuring the discrimination power of TDP features is proposed in Section 4. The time segment optimization and channel selection procedures are presented in Section 5. Experimental results are shown in Section 6.

2. DATA DESCRIPTION

The dataset IVa [9] from BCI competition III is used in this study. As it consists of EEG signals recorded using 118 electrodes, this dataset is very suitable for a fine selection of EEG channels. Five subjects, denoted "aa", "al", "av", "aw" and "ay", have performed 280 trials of cue-driven motor imagery (right hand: 140 trials, right foot: 140 trial) during the recording. The acquisition process was driven by visual cues, presented during 3.5s, and separated by randomly chosen intervals, ranging from 1.75 to 2.25s. Subjects were required to perform the corresponding motor imagery task during the presentation of a cue and to relax in the intermission. Ground truth is available for all subjects in this dataset. The aim of the experiment is to perform classification of the signal, for each subject, into two classes (right hand, right foot), with as few electrodes as possible. As the training data are recorded before the testing data in real applications, the first 70 trials for each class were used for training, and the remaining ones for the independent testing in this study.

3. TIME DOMAIN PARAMETERS

The EEG signals are bandpass filtered between 8 and 30 Hz using a 5th order Butterworth filter. For one channel (electrode) and one

This work was partially supported by grants from China Scholarship Council and Orange Labs. The authors would like to thank Dr. Sylvain Chevallier (Université de Versailles St-Quentin) for some useful discussions.

trial, we denote by $x(t)$ the filtered EEG signal in a time segment $[t_0, t_0 + T - 1]$. Time domain parameters (TDPs) are a set of broad band (i.e. 8-30Hz) EEG features defined in the time domain [10]:

$$TDP^{(p)} = \log(\underset{t\in[t_0,t_0+T-1]}{var}(\frac{d^p x(t)}{dt^p})), p = 0, 1, 2, ... \quad (1)$$

The logarithm is applied here to make the distribution of TDPs approximately normal, since the linear classifier we use here typically assumes that the input features follow Gaussian distributions [1]. Note that the TDP of order 0, $A = TDP^{(0)}$, is the logarithmic band power (BP) of the filtered signal. It characterizes the EEG pattern in terms of amplitude. As EEG signal can be considered as a mixture of sinusoidal waves, the derivative provides the information on frequency. The TDP of order 1, $M = TDP^{(1)}$ is a feature that reflects the EEG pattern in terms of frequency, and the TDP of order 2, $C = TDP^{(2)}$, reflects the change in frequency. We use these three TDPs, $[A, M, C]$, in this work, since they carry more information than the only BP feature, and have clearer physical meanings than TDPs of higher orders in BCI research.

4. A CRITERION BASED ON FISHER'S DISCRIMINANT

Fisher's discriminant analysis (Fisher's LDA) is a very popular classification algorithm in BCI research [1], because it has a very low computational cost and usually yields good results for motor imagery BCIs [11]. It projects high-dimensional data onto a direction and performs a linear classification in this one-dimensional space. The optimal projection is found by maximizing the separation between two classes. Let us assume that we have two classes of observations, h and f. In a one-dimensional feature space, the separation between two classes is defined using the Fisher criterion [1]:

$$FC = \frac{(\mu^h - \mu^f)^2}{(\sigma^h)^2 + (\sigma^f)^2} \quad (2)$$

where μ^h and μ^f are the mean values of the feature over all trials for classes h and f, respectively, $(\sigma^h)^2$ and $(\sigma^f)^2$ are the variances of the feature.

In feature selection, FC can be used to evaluate the discrimination power of each single feature [1]. However, it is not suitable to evaluate the discrimination power of a group of features. Thus, we propose a novel and simplified criterion based on Fisher's discriminant, called F score, $\hat{F}$, and we use it to estimate the discrimination power of a group of features (here TDPs):

$$\hat{F} = \frac{\left\|\vec{\mu}^h - \vec{\mu}^f\right\|_2^2}{tr(\Sigma^h) + tr(\Sigma^f)} \quad (3)$$

where $\|\cdot\|_2$ denotes the $L2$-norm (Euclidean norm), and $tr(\cdot)$ the trace of a matrix. Compared to FC, $\hat{F}$ is a derived version relying on the Euclidean distance between class centers, $\left\|\vec{\mu}^h - \vec{\mu}^f\right\|_2$, to estimate the difference between classes, and employing the trace of the covariance matrix to evaluate the variance within a class. Note that this simple expression avoids estimating a projection direction as required by the general multi-dimensional expression of LDA.

5. TIME-SPATIAL OPTIMIZATION FOR CHANNEL SELECTION

This method aims to find the optimal time segment and subset of channels for classification. The general scheme of the method is shown in Fig. 1. First, TDPs, $[A_e^\chi(i), M_e^\chi(i), C_e^\chi(i)]$, are computed from five overlapping time segments $[t_n, t_n+T-1]$, $n = 1, 2, ..., N$ ($N = 5$) of 0-2.0s, 0.5-2.5s, 1.0-3.0s, 1.5-3.5s and 2.0-4.0s after cue

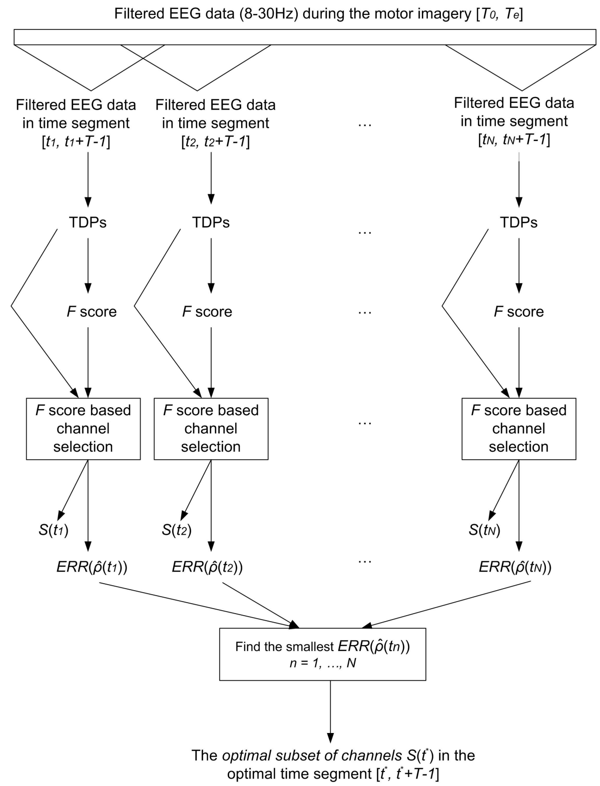

Fig. 1. General scheme of our method.

on-set ($t_n = 0, 0.5, 1.0, 1.5, 2.0$s, $T = 2$s) for each single trial i at channel e for class χ ($\chi \in \{h, f\}$). Then, the discrimination power of channel e during one time segment is estimated by the F score:

$$\hat{F}_e = \frac{(\bar{A}_e^h - \bar{A}_e^f)^2 + (\bar{M}_e^h - \bar{M}_e^f)^2 + (\bar{C}_e^h - \bar{C}_e^f)^2}{\tilde{A}_e^h + \tilde{A}_e^f + \tilde{M}_e^h + \tilde{M}_e^f + \tilde{C}_e^h + \tilde{C}_e^f} \quad (4)$$

with

$$\overline{TDP}_e^\chi = \frac{1}{K^\chi}\sum_{i=1}^{K^\chi} TDP_e^\chi(i) \quad (5)$$

$$\widetilde{TDP}_e^\chi = \frac{1}{K^\chi - 1}\sum_{i=1}^{K^\chi} (TDP_e^\chi(i) - \overline{TDP}_e^\chi)^2 \quad (6)$$

where K^χ is the number of training trials for class χ.

We first select the optimal subset of channels in each time segment $[t_n, t_n + T - 1]$. Existing methods typically determine the number of selected channels based on user's experience [8] or exhaustive searching strategy [6, 7], which is either arbitrary or time-consuming. Here, we propose an automatic approach, by considering the properties of both features and classifier to determine the size of the subset of selected channels.

Let $\hat{F}_m$ be the largest F score among all channels in the time segment $[t_n, t_n + T - 1]$:

$$\hat{F}_{max} = \max\left\{\hat{F}_e \mid e \in \{1, ..., 118\}\right\} \quad (7)$$

The relative discrimination power of each channel e is defined as:

$$\rho_F(e) = \frac{\hat{F}_e}{\hat{F}_{max}} \quad (8)$$

The value of $\rho_F(e)$ is between 0 and 1. A larger $\rho_F(e)$ indicates a larger relative discrimination power. Thus, a threshold $\hat{\rho}$ can be set to extract the channels with $\rho_F(e) > \hat{\rho}$ to be used for classification. A lower value of $\hat{\rho}$ tends to pick out more channels. In practice, it would be better to have training trials with five times as many as the dimensionality of features to guarantee a good performance of the classifier we are using [12]. As each channel yields three TDPs, the range of $\hat{\rho}$ can be shrunk to $[P, 1.0]$ to feed the classifier, where P is obtained by:

$$\min_P Num(P) \; s.t. \qquad (9)$$
$$P \in [0, 1.0], Num(P) \geq K/3R$$

where $Num(P)$ is the number of selected channels with $\rho_F(e) > P$, K is the number of trials for training, and R is the ratio of the number of trials to the number of features for a specific classifier (here $R = 5$). Then, different subsets of channels according to different $\hat{\rho} \in [P, 1.0]$ are used to train the classifier. The optimal $\hat{\rho}$ is obtained by seeking the subset with the lowest training error (ERR) in the classifier training. The training error is defined as the observed overall disagreement between classification outputs and true classes. Let $\hat{\rho}^*$ be the optimal $\hat{\rho}$, so it is obtained by:

$$ERR(\hat{\rho}^*) = \min\{ERR(\hat{\rho}) \mid \hat{\rho} \in [P, 1.0]\} \qquad (10)$$

If there are more than one $\hat{\rho}^*$ obtained by Equation (10), we use the largest $\hat{\rho}^*$ as the optimal one.

For each time segment $[t_n, t_n + T - 1]$, the optimal subsets of channels $S(t_n)$ is obtained by using $\hat{\rho}^*(t_n)$, the optimal $\hat{\rho}$ in the time segment $[t_n, t_n+T-1]$. Denote by $ERR(\hat{\rho}^*(t_n))$ the training error achieved by $S(t_n)$. The optimal time segment $[t^*, t^* + T - 1]$ is found by seeking the lowest training error $ERR(\hat{\rho}^*(t_n))$ among all time segments:

$$ERR(\hat{\rho}^*(t^*)) = \min_{t_n}\{ERR(\hat{\rho}^*(t_n))\} \qquad (11)$$

so as to obtain the optimal subset of channels $S(t^*)$ in the optimal time segment $[t^*, t^* + T - 1]$.

6. RESULTS

The optimal time segment and subset of selected channels for each subject are shown in Fig. 2. The numbers of selected channels are listed in Table 1. The number of selected electrodes, which is no more than 11 (see Table 1), is less than that of commercial BCI system Emotiv EPOC, which has 14 electrodes. Thus, the number of electrodes selected by our method is still reasonable and acceptable for general application (e.g. in a game environment). The computational time for finding the optimal combination of time segment and subset of channels depends on the full number of channels and the number of time segments. Here, for 118 channels and 5 time segments, it only needs 11 seconds (Matlab 7.10.0, Window 7 Professional 64bits, CPU 2.66GHz, RAM 2.0Go).

For most subjects, the selected channels are mainly around the hand representation area of the left brain (C3), because motor imagery of the right hand typically elicits strong ERD in this area (see Fig. 3). However, the selected channels may also exist in the right hand representation area for some subjects. Subject "aw" is an example, where the selected channels are mainly in the right brain. Further examination of the ERD/ERS maps for this subject shows that motor imagery of the right foot elicits very strong ERS in the hand representation area of the right brain (C4) (see Fig. 4), which explains why the channels with large discrimination power are mainly in the right brain.

Table 1. Number of selected channels for each subject.

User	aa	al	av	aw	ay
Number of channels	8	6	11	10	11

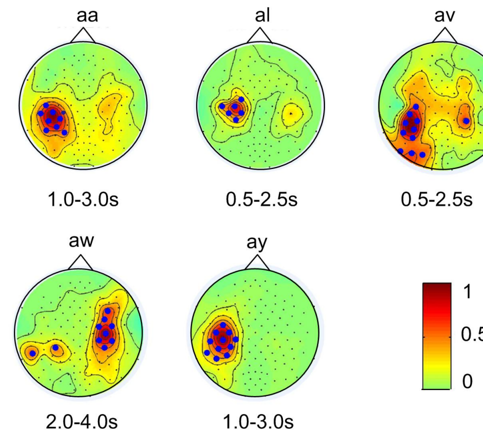

Fig. 2. Channel distribution of the F score and selected channels (marked by bold points) for the subject-specific optimal segment for each subject.

The TDPs are extracted from the optimal time segment and selected channels for the classifier training and the independent testing. In the testing, the classification results are evaluated by the classification accuracy (ACC), which is defined as the observed overall agreement between classification outputs and true classes. The mean classification accuracy for all subjects is denoted by $\overline{ACC}$. The classification results are compared with those obtained by using both BP and TDPs features extracted from all channels with CSP (3 pairs of spatial filters [13]) and from just three commonly used channels (C3, Cz, C4) at foot and hand representation areas (see Table 2). In the comparison, the time segment is from the cue on-set to the cue ending. The features are computed in this time segment. The paired-sample t-test was employed to reveal the statistical significance of the difference between the results of different methods. From Table 2, we can see that the results obtained by using TDPs are better than those using BP features for most subjects (even if the difference is not significant $p > 0.05$), indicating the interest of using TDPs in motor imagery BCI. The results obtained using our method ($\overline{ACC} = 0.78$) are significantly better than using three commonly used channels i.e. C3, Cz, C4 ($\overline{ACC} = 0.72$ when using TDPs, $\overline{ACC} = 0.71$ when using BP, all $p < 0.05$). Compared to the results obtained using all channels with CSP, the mean classification accuracy of our method is better than using full-cap CSP with BP features ($\overline{ACC} = 0.76$, not significant with $p > 0.05$) and equal to using full-cap CSP with TDP features ($\overline{ACC} = 0.78$ when using TDPs). For some subjects ("aa" and "ay"), our method even yields higher ACC than full-cap CSP. Thus, it meets the goal of largely reducing the number of electrodes (from 118 channels to no more than 11), without a major loss of classification performance. Moreover, we use a relatively shorter time segment (2s length) than other methods (3.5s length). For most subjects (except "aw"), the classification outputs were obtained before the ending of cue, which indicates that less time (here, less than 3.5s) is required for recording training data from those subjects.

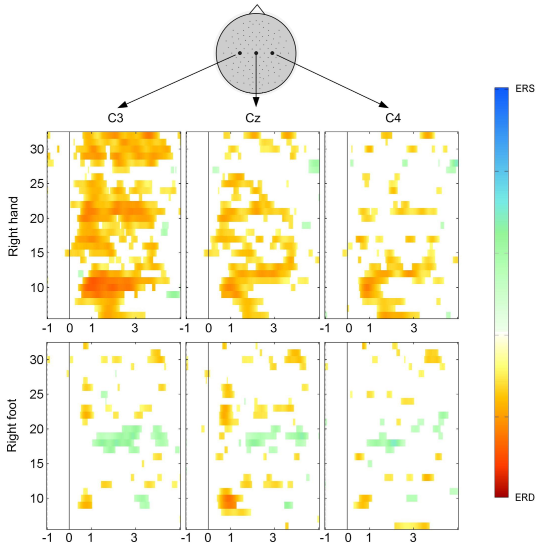

Fig. 3. Time-frequency visualization of ERD/ERS for Subject "ay", using the time interval between -1s (i.e. 1s before cue on-set) and 0s (i.e. cue on-set) as the baseline.

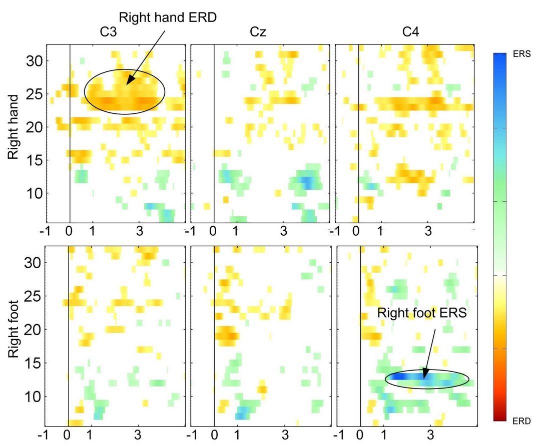

Fig. 4. Time-frequency visualization of ERD/ERS for Subject "aw", using the time interval between -1s (i.e. 1s before cue on-set) and 0s (i.e. cue on-set) as the baseline.

7. CONCLUSION

In this paper, we proposed a novel method and a new measure of discrimination power, relying on subject-specific time-spatial analysis for channel selection. The results show that our method can largely reduce the number of channels (from 118 channels to no more than 11), and shorten the training time, without a significant decrease of classification performance on a standard dataset (BCI competition III dataset IVa). The number of electrodes selected by our method is less than the one of a commercial BCI system, Emotiv EPOC, so the number is still reasonable and acceptable for general public applications, such as BCI games. This method can be used in designing BCI systems using few channels (electrodes) for subject-specific applications. Our approach can also be used to let the user decide on the best compromise between accuracy, easy use and portability, according to his needs. In the future, we will evaluate the robustness of this method to data evolution by random selection of training and testing data, and also perform experiments on more datasets.

Table 2. Evaluation results for different methods.

	CSP		C3+Cz+C4		Our method
User	BP	TDPs	BP	TDPs	TDPs
aa	0.46	0.47	0.64	0.59	0.67
al	0.94	0.94	0.79	0.81	0.88
av	0.68	0.69	0.58	0.59	0.61
aw	0.94	0.94	0.73	0.78	0.81
ay	0.75	0.84	0.81	0.82	0.92
mean	0.76	0.78	0.71	0.72	0.78

8. RELATION TO PRIOR WORK

Although earlier studies have presented the need for channel selection in reducing the electrodes required in a BCI system [6, 7, 8], they simply addressed the issue based on spatial information, disregarding the potential impact of temporal information. This contribution, with the proposition of a novel algorithm, emphasizes the potential effects of the chosen time segment on channel selection. Meanwhile, a novel criterion derived from Fisher's criterion is proposed to evaluate the discrimination power of a group of features, and applied on time domain parameters (TDP), which overcomes the disadvantage of Fisher's criterion [1] on TDP feature selection.

9. REFERENCES

[1] K.R. Müller, M. Krauledat, G. Dornhege, G. Curio, and B. Blankertz, "Machine learning techniques for brain-computer interfaces," *Biomed Tech (Berlin)*, vol. 49, no. 1, pp. 11–22, 2004.

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

[3] G. Pfurtscheller, C. Brunner, A. Schlogl, and FH Lopes da Silva, "Mu rhythm (de)synchronization and EEG single-trial classification of different motor imagery tasks," *NeuroImage*, vol. 31, no. 1, pp. 153–159, 2006.

[4] Y. Yang, S. Chevallier, J. Wiart, and I. Bloch, "Time-frequency selection in two bipolar channels for improving the classification of motor imagery EEG," in *Proc. IEEE EMBC*, 2012, pp. 2744–2747.

[5] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K.R. Müller, "Optimizing spatial filters for robust EEG single-trial analysis," *IEEE Signal Process. Mag.*, vol. 25, no. 1, pp. 41–56, 2008.

[6] A. Barachant and S. Bonnet, "Channel selection procedure using Riemannian distance for BCI applications," in *Proc. IEEE/EMBS NER*, 2011, pp. 348–351.

[7] T.N. Lal, M. Schröder, T. Hinterberger, J. Weston, M. Bogdan, N. Birbaumer, and B. Schölkopf, "Support vector channel selection in BCI," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 1003–1010, 2004.

[8] Y. Wang, S. Gao, and X. Gao, "Common spatial pattern method for channel selelction in motor imagery based brain-computer interface," in *Proc. IEEE EMBS*, 2006, pp. 5392–5395.

[9] B. Blankertz, G. Dornhege, M. Krauledat, K.R. Muller, V. Kunzmann, F. Losch, and G. Curio, "The Berlin brain-computer interface: EEG-based communication without subject training," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 14, no. 2, pp. 147–152, 2006.

[10] C. Vidaurre, N. Kramer, B. Blankertz, and A. Schlogl, "Time domain parameters as a feature for EEG-based brain-computer interfaces," *Neural Netw.*, vol. 22, no. 9, pp. 1313–1319, 2009.

[11] F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, and B. Arnaldi, "A review of classification algorithms for EEG-based brain–computer interfaces," *J. Neural Eng.*, vol. 4, pp. R1–R14, 2007.

[12] A. Jain and D. Zongker, "Feature selection: Evaluation, application, and small sample performance," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 2, pp. 153–158, 1997.

[13] Y. Yang, S. Chevallier, J. Wiart, and I. Bloch, "Automatic selection of the number of spatial filters for motor-imagery BCI," in *Proc. ESANN*, 2012, pp. 109–114.