



**HAL**  
open science

# Regularizing Common Spatial Patterns to Improve BCI Designs: Theory and Algorithms

Fabien Lotte, Cuntai Guan

► **To cite this version:**

Fabien Lotte, Cuntai Guan. Regularizing Common Spatial Patterns to Improve BCI Designs: Theory and Algorithms. 2010. inria-00476820v1

**HAL Id: inria-00476820**

**<https://inria.hal.science/inria-00476820v1>**

Preprint submitted on 4 May 2010 (v1), last revised 24 Sep 2010 (v4)

**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.

# Regularizing Common Spatial Patterns to Improve BCI Designs: Theory and Algorithms

Fabien LOTTE\*, *Member, IEEE*, Cuntai GUAN, *Senior Member, IEEE*

**Abstract**—One of the most popular feature extraction algorithms for Brain-Computer Interfaces (BCI) is the Common Spatial Patterns (CSP) algorithm. Despite its known efficiency and widespread use, CSP is also known to be very sensitive to noise and prone to overfitting. To address this issue, some groups have recently proposed to regularize CSP. In this paper, we present a simple and unifying theoretical framework to perform such a CSP regularization. We then present a mini-review of existing Regularized CSP (RCSP) algorithms, and describe how to cast them in this framework. We also propose 4 new RCSP algorithms. Finally, we compare the performances of 10 different RCSP algorithms (including these 4 new ones and the original CSP), on EEG data from 17 subjects, from BCI competition data sets. Results showed that the best RCSP methods can outperform CSP by nearly 10% in median classification accuracy and lead to more neurophysiologically relevant spatial filters. They also enable us to perform efficient subject-to-subject transfer. Overall, the best RCSP algorithms on these data were the CSP with Tikhonov Regularization and Weighted Tikhonov Regularization, both newly proposed in this paper.

**Index Terms**—brain-computer interfaces (BCI), EEG, common spatial patterns (CSP), regularization, subject-to-subject transfer

## I. INTRODUCTION

Brain-Computer Interfaces (BCI) are communication systems which enable users to send commands to computers by using brain activity only, this activity being generally measured by ElectroEncephaloGraphy (EEG) [1]. BCI are generally designed according to a pattern recognition approach, i.e., by extracting features from EEG signals, and by using a classifier to identify the user's mental state from such features [1][2]. The Common Spatial Patterns (CSP) algorithm is a feature extraction method which can learn spatial filters maximizing the discriminability of two classes [3][4]. CSP has been proven to be one of the most popular and efficient algorithms for BCI design, notably during BCI competitions [5][6], but also for actual (i.e., online) BCI use [7]. CSP is also the basis for several of the most successful spatio-spectral feature extraction methods for BCI, e.g., [8][4].

Despite the popularity and efficiency of CSP for BCI design, CSP is also known to be highly sensitive to noise and to severely overfit with small training sets [9][10]. In order to cope with these drawbacks, a recent idea was to add a-priori information into the CSP learning process, under the form of regularization terms [11][12][13][14] (see Section IV-A for a mini-review). These Regularized CSP (RCSP) have all been shown to lead to higher performances than the classical CSP. However, all these RCSP algorithms are

expressed with a different formulation and therefore lack a unifying regularization framework. Moreover, they were only compared to standard CSP, and typically with only 4 or 5 subjects [12][11][13], which prevents us from assessing their relative performances. Finally, we believe that a variety of other a-priori could be incorporated into CSP.

Therefore, in this paper we present a simple theoretical framework to obtain RCSP algorithms. We first express the existing RCSP within this framework and then present 4 new RCSP algorithms, based on new a-priori. It should be mentioned that preliminary studies of 2 of these new algorithms have been presented in conference papers [15][14]. Finally, we compare these various algorithms on EEG data from 17 subjects, from publicly available BCI competition data sets.

This paper is organized as follows: Section II describes the CSP algorithm while Section III presents the theoretical framework to regularize it. Then, Section IV expresses existing RCSP algorithms within this framework, as well as presents 4 new algorithms. Finally, Section V describes the EEG data used to compare these various algorithms, as well as the results obtained. The paper ends with a conclusion in Section VI.

## II. THE CSP ALGORITHM

CSP aims at learning spatial filters which can maximize the variance of band-pass filtered EEG signals from one class while minimizing the variance of EEG signals from the other class [4][3]. As the variance of EEG signals filtered in a given frequency band corresponds to the signal power in this band, CSP aims at achieving optimal discrimination for BCI based on band power features [3]. Formally, CSP uses the spatial filters  $w$  which extremize the following function:

$$J(w) = \frac{w^T X_1^T X_1 w}{w^T X_2^T X_2 w} = \frac{w^T C_1 w}{w^T C_2 w} \quad (1)$$

where  $T$  denotes transpose,  $X_i$  is the data matrix for class  $i$  (with the training samples as rows and the EEG electrodes as columns) and  $C_i$  is the spatial covariance matrix of the EEG signals from class  $i$ , assuming a zero mean for the EEG signals. Note that this last assumption is generally met as EEG signals are usually band-pass filtered. This optimization problem can be solved (though this is not the only way) by first observing that the function  $J(w)$  remains unchanged if the filter  $w$  is rescaled. Indeed  $J(k \times w) = J(w)$ , with  $k$  a real constant, which means the rescaling of  $w$  is arbitrary. As such, extremizing  $J(w)$  is equivalent to extremizing  $w^T C_1 w$  subject to the constraint  $w^T C_2 w = 1$  as it is always possible to find a rescaling of  $w$  such that  $w^T C_2 w = 1$ . Using the Lagrange multiplier method, this constrained optimization

problem amounts to extremizing the following function:

$$L(\lambda, w) = w^T C_1 w - \lambda(w^T C_2 w - 1) \quad (2)$$

The filters  $w$  extremizing  $L$  are such that the derivative of  $L$  with respect to  $w$  equals 0:

$$\begin{aligned} \frac{\partial L}{\partial w} &= 2w^T C_1 - 2\lambda w^T C_2 = 0 \\ \Leftrightarrow C_1 w &= \lambda C_2 w \\ \Leftrightarrow C_2^{-1} C_1 w &= \lambda w \end{aligned}$$

We obtain a standard eigenvalue problem. The spatial filters extremizing Eq. 1 are then the eigenvectors of  $M = C_2^{-1} C_1$  which correspond to its largest and lowest eigenvalues. When using CSP, the extracted features are the logarithm of the variance of the EEG signals projected onto these filters  $w$ .

### III. REGULARIZED CSP: THEORY

As mentioned above, in order to overcome the sensitivity of CSP to noise and overfitting, it is worth regularizing it. Adding a-priori information to CSP, and thus regularizing it, can be done at two levels. First, regularization can be performed at the covariance matrix estimation level. Indeed, CSP being based on spatial covariance matrix estimates, these estimates can also suffer from noise or small training sets, and as such can benefit from regularization. This is the most common kind of CSP regularization in the BCI literature. However, another kind of regularization can be designed: regularizing the CSP objective function itself (Eq. 1), by imposing a-priori on the spatial filters to be obtained. The remaining of this section presents these two forms of regularization.

#### A. Regularizing the covariance matrix estimates

Finding spatial filters using CSP requires to estimate the spatial covariance matrices for each class (see previous Section). However, if the available EEG training set is noisy and/or small, these covariance matrices may be poor or non-representative estimates of the mental states involved, and as such lead to poor spatial filters. Therefore, it seems appropriate to add a-priori information to the estimation of these matrices, under the form of regularization terms. Based on [12], such a regularization can be performed as follows:

$$\begin{aligned} \tilde{C}_c &= (1 - \gamma)\hat{C}_c + \gamma I \quad (3) \\ \text{with } \hat{C}_c &= (1 - \beta)s_c C_c + \beta G_c \quad (4) \end{aligned}$$

In these equations,  $C_c$  is the initial spatial covariance matrix for class  $c$ ,  $\tilde{C}_c$  is the regularized estimate,  $I$  is the identity matrix,  $s_c$  is a constant scaling parameter (a scalar),  $\gamma$  and  $\beta$  are two user-defined regularization parameters (with value between 0 and 1) and  $G_c$  is a "generic" covariance matrix (see below for more details). This kind of regularization involves two regularization terms. The first one, associated to the regularization parameter  $\gamma$ , shrinks the initial covariance matrix estimate towards the identity matrix in order to counteract a possible estimation bias due to a small training set. The second term, associated to the regularization parameter  $\beta$ , shrinks the initial covariance matrix estimate towards a

generic covariance matrix, in order to obtain a more stable and reliable estimate. This generic matrix represents a given a-priori on how the covariance matrix for the mental state considered should be. In the literature, this matrix is typically built by using signals from several different subjects whose EEG data has been recorded previously. This has been shown to be an effective way to perform subject-to-subject transfer [13][12][15]. However, it should be mentioned that  $G_c$  could also be defined based only on neurophysiological a-priori.

Learning spatial filters based on this regularization method simply consists in replacing the initial covariance matrices  $C_1$  and  $C_2$  used in CSP by their regularized estimates  $\tilde{C}_1$  and  $\tilde{C}_2$ . Many different regularized CSP algorithms can be designed in this way, depending on whether only one or the two regularization terms are used, and more importantly, depending on how the generic covariance matrix  $G_c$  is built (if the corresponding regularization term is used).

#### B. Regularizing the CSP objective function

Another approach to obtain regularized CSP algorithms consists in regularizing the CSP objective function itself (Eq. 1). More precisely, such method consists in adding a regularization term to the CSP objective function in order to penalize solutions (i.e., resulting spatial filters) that do not satisfy a given a-priori. Formally, the objective function becomes:

$$J_{P_1}(w) = \frac{w^T C_1 w}{w^T C_2 w + \alpha P(w)} \quad (5)$$

where  $P(w)$  is a penalty function measuring how much the spatial filter  $w$  satisfies a given a-priori. The more  $w$  satisfies it, the lower  $P(w)$ . Hence, to maximize  $J_{P_1}(w)$ , we must minimize  $P(w)$ , thus ensuring spatial filters satisfying the a-priori.  $\alpha$  is a user-defined regularization parameter ( $\alpha \geq 0$ , the higher  $\alpha$ , the more satisfied the a-priori). With this regularization, we expect that enforcing specific solutions, thanks to a-priori, will guide the optimization process towards good spatial filters, especially with limited or noisy training data.

In this paper, we focus on quadratic penalties, i.e., terms such that  $P(w) = \|w\|_K^2 = w^T K w$ , where matrix  $K$  encodes the a-priori. Interestingly enough, RCSP with non-quadratic penalties have been proposed [16][17]. They used an  $l_1$  norm penalty to select a sparse set of channels. However, these studies showed that sparse CSP generally gave lower performances than CSP (with all channels), although they require much less channels, hence performing efficient channel reduction. As the focus of this paper is not channel reduction, we only consider quadratic penalties here. Moreover, quadratic penalties lead to a close form solution for optimization (see below), which is more convenient and computationally efficient. With a quadratic penalty term, Eq. 5 becomes:

$$\begin{aligned} J_{P_1}(w) &= \frac{w^T C_1 w}{w^T C_2 w + \alpha w^T K w} \\ &= \frac{w^T C_1 w}{w^T (C_2 + \alpha K) w} \end{aligned}$$

The corresponding Lagrangian is:

$$L_{P_1}(\lambda, w) = w^T C_1 w - \lambda(w^T (C_2 + \alpha K) w - 1) \quad (6)$$

By following the same approach as previously (see Section II), we obtain the following eigenvalue problem:

$$(C_2 + \alpha K)^{-1} C_1 w = \lambda w \quad (7)$$

Thus, maximizing  $J_{P_1}(w)$  consists in taking the eigenvectors corresponding to the largest eigenvalues of  $M_1 = (C_2 + \alpha K)^{-1} C_1$  as the filters  $w$ . With CSP, the eigenvectors corresponding to both the largest and smallest eigenvalues of  $M$  (see Section II) are used as the spatial filters, as they respectively maximize and minimize Eq. 1 [4]. However, for RCSP, we cannot use the eigenvectors corresponding to the lowest eigenvalues of  $M_1$ , as they will minimize Eq. 5, and as such will maximize the penalty term. Therefore, in order to obtain the filters which maximize  $C_2$  while minimizing  $C_1$ , we also need to maximize the following objective function:

$$J_{P_2}(w) = \frac{w^T C_2 w}{w^T C_1 w + \alpha P(w)} \quad (8)$$

which is achieved by using the eigenvectors corresponding to the largest eigenvalues of  $M_2 = (C_1 + \alpha K)^{-1} C_2$  as the filters  $w$ . In other words, with RCSP, the spatial filters used are the eigenvectors corresponding to the largest eigenvalues of  $M_1$  and to the largest eigenvalues of  $M_2$ . With this approach, various regularized CSP algorithms can be designed depending on the knowledge encoded into matrix  $K$ .

### C. Summary

In this Section, we have presented two theoretical approaches to design RCSP algorithms: one at the covariance matrix estimation level and at one the objective function level using quadratic penalties. Naturally, these two approaches are not exclusive and can be combined within the same framework. Table I summarizes this framework and highlight the differences between the standard CSP and RCSP algorithms based on this framework. As can be seen from this table, many different RCSP can be designed depending on 1) which one(s) of the 3 regularization terms (associated to parameters  $\alpha$ ,  $\beta$  and  $\gamma$ ) are used and on 2) how the regularization matrices  $G_c$  and  $K$  are built. The following section presents several such variants, including existing algorithms as well as 4 new ones.

TABLE I  
DIFFERENCES IN OBJECTIVE FUNCTION AND ALGORITHM OPTIMIZATION BETWEEN A STANDARD CSP AND A REGULARIZED CSP (RCSP).

	CSP	RCSP
Objective function	$J(w) = \frac{w^T C_1 w}{w^T C_2 w}$	$J_{P_{\{1,2\}}}(w) = \frac{w^T \tilde{C}_{\{1,2\}} w}{w^T \tilde{C}_{\{2,1\}} w + \alpha P(w)}$ with $P(w) = w^T K w$ $\tilde{C}_c = (1 - \gamma) \hat{C}_c + \gamma I$ $\tilde{C}_c = (1 - \beta) s_c C_c + \beta G_c$
Solutions of the optimization problem	eigenvectors corresponding to the $N_f$ largest and $N_f$ lowest eigenvalues of $M = C_2^{-1} C_1$	eigenvectors corresponding to the $N_f$ largest eigenvalues of $M_1 = (\tilde{C}_2 + \alpha K)^{-1} \tilde{C}_1$ and $M_2 = (\tilde{C}_1 + \alpha K)^{-1} \tilde{C}_2$

## IV. REGULARIZED CSP: ALGORITHMS

### A. Existing RCSP algorithms

Four kinds of RCSP algorithms have been proposed so far: Composite CSP, Regularized CSP with Generic Learning, Regularized CSP with Diagonal Loading and invariant CSP. They are described below within the presented framework.

#### 1) Composite CSP:

The Composite CSP (CCSP) algorithm, proposed by Kang et al, aims at performing subject-to-subject transfer by regularizing the covariance matrices using other subjects' data [13]. Expressed within this paper framework, CCSP uses only the  $\beta$  hyperparameter ( $\alpha = \gamma = 0$ ), and defines the generic covariance matrices  $G_c$  according to covariance matrices of other subjects. Two methods were proposed to build  $G_c$ .

With the first method, denoted here as CCSP1,  $G_c$  is built as a weighted sum of the covariance matrices (corresponding to the same mental state) of other subjects, by de-emphasizing covariance matrices estimated from fewer trials:

$$G_c = \sum_{i \in \Omega} \frac{N_c^i}{N_{t,c}} C_c^i \quad \text{and} \quad s_c = \frac{N_c}{N_{t,c}} \quad (9)$$

where  $\Omega$  is a set of subjects whose data is available,  $C_c^i$  is the spatial covariance matrix for class  $c$  and subject  $i$ ,  $N_c^i$  is the number of EEG trials used to estimate  $C_c^i$ ,  $N_c$  is the number of EEG trials used to estimate  $C_c$  (matrix for the target subject), and  $N_{t,c}$  is the total number of EEG trials for class  $c$  (from all subjects in  $\Omega$  together with the target subject).

With the second method, denote here as CCSP2,  $G_c$  is still built as a weighted sum of the covariance matrices of other subjects, but the weights are here defined according to the Kullback-Leibler (KL) divergence between the data of each subject and the target subject:

$$G_c = \sum_{i \in \Omega} \frac{1}{Z} \frac{1}{KL(i, t)} C_c^i \quad \text{with} \quad Z = \sum_{j \in \Omega} \frac{1}{KL(j, t)} \quad (10)$$

where  $KL(i, t)$  is the KL-divergence between the target subject  $t$  and subject  $i$ , and is defined as follows:

$$KL(i, t) = \frac{1}{2} \left( \log \left( \frac{\det C_c}{\det C_c^i} \right) + \text{tr} (C_c^{-1} C_c^i) - N_e \right) \quad (11)$$

where  $\det$  and  $\text{tr}$  are respectively the determinant and the trace of a matrix, and  $N_e$  is the number of electrodes used. With CCSP2, the scaling constant  $s_c$  is equal to 1.

#### 2) Regularized CSP with Generic Learning:

The RCSP approach with Generic Learning, proposed by Lu et al [12] and denoted here as GLRCSP, is another approach which aims at regularizing the covariance matrix estimation using data from other subjects. GLRCSP uses both the  $\beta$  and  $\gamma$  regularization terms, i.e., it tries to shrink the covariance matrix towards both the identity matrix and a generic covariance matrix  $G_c$ . Similarly to CCSP, this generic matrix is computed from the covariance matrices of other subjects:

$$G_c = s_G \sum_{i \in \Omega} C_c^i \quad (12)$$

where  $s_c = s_G = \frac{1}{(1-\beta)M_{C_c} + \beta \sum_{i \in \Omega} M_{C_c^i}}$ , and  $M_A$  is the number of trials used to compute the covariance matrix  $A$ .

### 3) Regularized CSP with Diagonal Loading:

Another form of covariance matrix regularization that was used in the BCI literature is Diagonal Loading (DL), which consists in shrinking the covariance matrix towards the identity matrix. In other words, it is a form a covariance matrix regularization which uses only the  $\gamma$  regularization parameter ( $\beta = \alpha = 0$ , see Table I). Interestingly enough, in this case the value of  $\gamma$  can be automatically identified using the method of Ledoit and Wolf [18]. We will denote this form of RCSP based on automatic DL as DLCSPauto. In order to see whether this automatic regularization is efficient for discrimination purposes, we will also investigate a classical selection of the regularization parameter using Cross-Validation (CV). We denote the resulting algorithm as DLCSPcv. It should be mentioned that, although covariance matrix regularization based on DL has been used in the BCI literature (see, e.g., [19]), to our best knowledge, it has not been used for CSP regularization, but for regularization of other algorithms such as Linear Discriminant Analysis (LDA).

### 4) Invariant CSP:

The Invariant CSP (iCSP) is an algorithm which regularizes the CSP objective function in order to make the filters invariant to a given source of noise [11] (it therefore uses  $\beta = \gamma = 0$ ). In order to do so, the regularization matrix  $K$  is defined as the covariance matrix of this source of noise, e.g., as the covariance matrix of the changing level of parietal  $\alpha$ -activity. It should be mentioned that, in order to obtain this noise covariance matrix, additional EEG measurements must be performed to acquire the corresponding EEG signals and compute their covariance matrix. As we don't have such measurements for the EEG data sets analyzed here, the iCSP will not be considered in the evaluation section of this paper. However, it is worth noting that this an interesting and efficient approach to make CSP robust against known sources of noise.

## B. New RCSP algorithms

In this section, we proposes 4 new algorithms to incorporate a-priori information into CSP in order to regularize it. These algorithms are a CSP regularized with selected subjects, a Tikhonov Regularized CSP, a weighted Tikhonov Regularized CSP and a spatially Regularized CSP.

### 1) Regularized CSP with Selected Subjects:

This first new RCSP algorithm belongs to the same family of algorithms as CCSP in the sense that it uses data from other subjects to shrink the covariance matrix towards a generic matrix  $G_c$  (it therefore uses only the  $\beta$  regularization parameter,  $\alpha = \gamma = 0$  with this approach). However, contrary to CCSP or GLRCSP, the proposed algorithm does not use the data from all available subjects but only from a subset of them. Indeed, even if data from many subjects is available, it may not be relevant to use all of them, due to potentially large inter-subject variabilities. Thus, we propose to build  $G_c$

from the covariance matrices of a subset of selected subjects. We therefore denote this algorithm as Regularized CSP with Selected Subjects or SSRCSPP. With SSRCSPP, the generic covariance matrix  $G_c$  is built as follows:

$$G_c = \frac{1}{|S_t(\Omega)|} \sum_{i \in S_t(\Omega)} C_c^i \quad (13)$$

where  $|A|$  is the number of elements in set  $A$  and  $S_t(\Omega)$  is a subset of subjects from  $\Omega$ . In order to select an appropriate subset of subjects  $S_t(\Omega)$ , we propose the subject selection algorithm described in Algorithm 1. In this algorithm, the function  $accuracy = trainThenTest(trainingSet, testingSet)$  returns the accuracy obtained when training an SSRCSPP with  $\beta = 1$  (i.e., using only data from other subjects) on the data set  $trainingSet$  and testing them on data set  $testingSet$ . The classifier used here is a LDA. The function  $(best_i, max_{f(i)}) = max_i f(i)$  returns  $best_i$ , the value of  $i$  for which  $f(i)$  is maximum (this maximum value is denoted  $max_{f(i)}$ ). In short, this algorithm sequentially selects the subject to add or to remove from the current subset of subjects, in order to maximize the accuracy obtained when training the BCI on the data from this subset of subjects and testing it on the training data of the target subject. This algorithm has the same structure as the Sequential Forward Floating Search algorithm [20], used to select a relevant subset of features. This ensures the convergence of our algorithm as well as the selection of a good subset of additional subjects. With this subject selection approach, we expect to be able to build better generic covariance matrices  $G_c$  than with CCSP, and thus to reach better performances.

### 2) CSP with Tikhonov Regularization:

As mentioned above, a form of CSP regularization is the regularization of the CSP objective function using quadratic penalties (see Section III-B). In this family of algorithms, we use only the  $\alpha$  regularization parameter ( $\gamma = \beta = 0$  and  $s_c = 1$ ). The first algorithm that we propose in this category is a CSP with Tikhonov Regularization (TR) or TRCSP. Tikhonov Regularization is classical form of regularization, initially introduced for regression problems [21], and which consists in penalizing solutions with large weights. The penalty term is then  $P(w) = \|w\|^2 = w^T w = w^T I w$ . TRCSP is then simply obtained by using  $K = I$  in the proposed framework (see Table I). Such regularization is expected to constrain the solution to filters with a small norm, hence restraining the range of possible solutions and mitigating the influence of artifacts and outliers.

### 3) CSP with Weighted Tikhonov Regularization:

With TRCSP, high weights are penalized equally for each channel. However, in BCI, we know that for a given mental state to be classified, some channels are more important than others. Therefore, it may be interesting to have different penalties for different channels. More precisely, if we have some reasons to think that a given channel is unlikely to have a large contribution in the spatial filters, then we should give a

**Input:**  $D_t$ : training EEG data from the target subject.

**Input:**  $\Omega = \{D_s\}, s \in [0, N_s]$ : set of EEG data from the  $N_s$  other subjects available ( $D_t \ni \Omega$ ).

**Output:**  $S_t(\Omega)$ : a subset of relevant subjects whose data can be used to classify the data  $D_t$  of the target subject.

```

selected0 = {};
remaining0 = Ω;
accuracy0 = 0; n = 1;
while n < Ns do
  Step 1: (bestSubject, bestAccuracy) =
    maxs ∈ remainingn-1 trainThenTest(selectedn-1 +
    {Ds}, Dt);
    selectedn = selectedn-1 + {DbestSubject};
    remainingn = remainingn-1 - {DbestSubject};
    accuracyn = bestAccuracy;
    n = n + 1;
  Step 2: if n > 2 then
    (bestSubject, bestAccuracy) =
    maxs ∈ selectedn trainThenTest(selectedn -
    {Ds}, Dt);
    if bestAccuracy > accuracyn-1 then
      selectedn-1 = selectedn - {DbestSubject};
      remainingn-1 = remainingn + {DbestSubject};
      accuracyn-1 = bestAccuracy;
      n = n - 1;
      go to Step 2;
    else
      go to Step 1;
    end
  end
end
(bestN, selectedAcc) = maxn ∈ [1, Ns] accuracyn;
St(Ω) = selectedbestN;

```

**Algorithm 1:** Subject selection algorithm for the SSRCSPP (Regularized CSP with Selected Subjects) algorithm.

relatively large penalty to this channel, in order to prevent the CSP from assigning it a large contribution (which can happen due to noise or artifacts for instance). On the other hand, if a channel is likely to be useful, we should not prevent CSP from giving it high weights, as this channel is likely to have a genuinely large contribution.

Formally, this leads to a penalty term of the form  $P(w) = w^T D_w w$ , where  $D_w$  is a diagonal matrix such that  $D_w = \text{diag}(w_G)$  and  $w_G(i)$  is the level of penalty assigned to channel  $i$ . The Weighted Tikhonov Regularized CSP (WTRCSP) is then obtained by using  $K = D_w$ . Interestingly enough, these penalty levels  $w_G(i)$  can be defined according to the neurosciences literature, i.e., according to which brain regions (and thus channels) are expected to be useful. However, it may appear as difficult to select manually an appropriate real penalty value for each channel. Therefore, in this paper, we propose to use data from other subjects to obtain  $w_G(i)$ :

$$w_G = \left( \frac{1}{2 \times N_f \times |\Omega|} \sum_{i \in \Omega} \sum_{f=1}^{2 \times N_f} \frac{w_f^i}{\|w_f^i\|} \right)^{-1} \quad (14)$$

where  $w_f^i$  is the  $f^{\text{th}}$  spatial filter obtained using CSP (among the eigenvectors corresponding to the  $N_f$  largest and lowest eigenvalues of  $M$ , see Table I) for the  $i^{\text{th}}$  additional subject available. In other words, the penalty level of a channel is simply set to the inverse of the average weight of this channel in the CSP filters obtained from other

subjects (the less important the average channel weight, the higher the penalty). By doing so, we expect that the degree of usefulness of a given channel would be reflected by its weight in the filters obtained with CSP from other subjects.

#### 4) Spatially Regularized CSP:

The last new algorithm that we propose in this paper is a Spatially Regularized CSP (SRCSP). The motivation behind this algorithm is that despite being used to learn spatial filters, the standard CSP completely ignores the spatial location of the EEG electrodes. The goal of SRCSP is to include spatial information into the learning process of the CSP. More particularly, we would like to obtain spatially smooth filters  $w$ , i.e. filters for which neighboring electrodes have relatively similar weights. Indeed, from a neurophysiological point of view, neighboring neurons tend to have similar functions, which supports the idea that neighboring electrodes should measure similar brain signals (if the electrodes are close enough to each other). To ensure spatial smoothness of the filters  $w$ , we use a Laplacian penalty term  $P(w)$  as in [22], with the following regularization matrix  $K$ :

$$K = D_G - G \quad \text{with} \quad G(i, j) = \exp\left(-\frac{1}{2} \frac{\|v_i - v_j\|^2}{r^2}\right) \quad (15)$$

where  $v_i$  is the vector containing the 3D coordinates of the  $i^{\text{th}}$  electrodes, and  $D_G$  is a diagonal matrix such as  $D_G(i, i) = \sum_j G(i, j)$ . Here,  $r$  is an hyperparameter which defines how far two electrodes can be to be still considered as close to each other. As  $w^T (D_G - G) w = \sum_{i,j} G(i, j) (w_i - w_j)^2$  (see, e.g., [23]), the penalty term  $P(w) = w^T K w$  will be large for non-smooth filters, i.e., for filters in which neighboring electrodes have very different weights.

#### C. Hyperparameter selection

All RCSP algorithms presented here have one or more regularization parameters whose value must be defined by the user (except for DLCSPauto where the hyperparameter value is automatically selected):  $\alpha$ ,  $\beta$  and  $\gamma$  (see Table I). SRCSP has also its own specific hyperparameter:  $r$  which defines the size of the neighborhood considered for smoothing. In [12] and [13], the selection of the regularization parameters for GLRCSP and CCSP was not addressed, and the authors simply presented the results for several values of the hyperparameters. In this paper, we used cross-validation (CV) to select these values. More precisely, we used as the hyperparameter optimal values, the ones which maximized the 10-fold cross validation accuracy on the training set, by using LDA [2] as the classifier. We selected values among the set  $[0, 0.1, 0.2, \dots, 0.9]$  for the parameters  $\beta$  and  $\gamma$ , among the set  $[10^{-10}, 10^{-9}, \dots, 10^{-1}]$  for  $\alpha$ , and among  $[0.01, 0.05, 0.1, 0.5, 0.8, 1.0, 1.2, 1.5]$  for  $r$ .

#### D. The best of two worlds?

So far, all the algorithms presented use only one form of regularization: either they regularize the covariance matrices or they regularize the CSP objective function, but not both at the same time. However, it is easy to imagine an algorithm that would combine these two approaches (see Table I). We

implemented and assessed some of such algorithms, e.g., a SRCSP with Diagonal Loading or a TRCSP with Generic Learning, among others. Unfortunately, none of them reached performances as high as that of the corresponding algorithm with a single form of regularization (results not reported here due to space limitations). It therefore seems that regularizing CSP with a single form of regularization is the simplest and most efficient approach. Thus, only this approach is discussed in the following.

## V. EVALUATION

### A. EEG data sets used for evaluation

In order to assess and compare the RCSP algorithms presented here, we used EEG data of 17 subjects, from 3 publicly available data sets of BCI competitions. These three data sets contain Motor Imagery (MI) EEG signals, i.e., EEG signals recorded while the subjects were performing imagination of limb movements (such as left hand, right, hand or foot movements) [1]. They are described below.

1) *Data set IVa, BCI competition III*: Data set IVa [24], from BCI competition III [6], contains EEG signals from 5 subjects, who performed right hand and foot MI. EEG were recorded using 118 electrodes. A training set and a testing set were available for each subject. Their size was different for each subject. More precisely, 280 trials were available for each subject, among which 168, 224, 84, 56 and 28 composed the training set for subject A1, A2, A3, A4 and A5 respectively, the remaining trials composing their test set.

2) *Data set IIIa, BCI competition III*: Data set IIIa [25], from BCI competition III [6], comprises EEG signals from 3 subjects who performed left hand, right hand, foot and tongue MI. EEG signals were recorded using 60 electrodes. For the purpose of this study, only EEG signals corresponding to left and right hand MI were used. A training and a testing set were available for each subject. Both sets contain 45 trials per class for subject B1, and 30 trials per class for subject B2 and B3.

3) *Data set IIa, BCI competition IV*: Data set IIa [26], from BCI competition IV<sup>1</sup> comprises EEG signals from 9 subjects who performed left hand, right hand, foot and tongue MI. EEG signals were recorded using 22 electrodes. Only EEG signals corresponding to left and right hand MI were used for the present study. A training and a testing set were available for each subject, both sets containing 72 trials for each class.

It should be mentioned that the purpose of this paper is not to compare our results with BCI competition winners. We used BCI competition data sets because they are publicly available and thus enable our results to be reproduced. This paper only deals with feature extraction. However, to enable comparisons with BCI competition winners, one should also carefully design and tune preprocessing and classification algorithms, which is not in the scope of this paper.

### B. Preprocessing

For each data set, EEG signals were classically band-pass filtered in the 8-30 Hz frequency band using a 5<sup>th</sup>

order Butterworth filter. Indeed, this frequency band contains the main frequencies involved in MI [1]. In this work, we considered the discrete classification of the trials, i.e., we assigned a class to each trial and not to each trial sample. For each data set, for each trial, we extracted the features from the time segment located from 0.5s to 2.5s after the cue instructing the subject to perform MI (as done by the winner of BCI competition IV, data set IIa). With each CSP algorithm, we used 3 pairs of filters for feature extraction, as recommended in [4]. we used the filters corresponding to the 3 largest and 3 smallest eigenvalues of  $M$  for CSP, CCSP, DLCSP, GLRCSP and SSRCSPP that are based on covariance matrix regularization, and the filters corresponding to the 3 largest eigenvalues of  $M_1$  and the 3 largest eigenvalues of  $M_2$  for TRCSP, WTRCSP and SRCSP that are based on objective function regularization.

### C. Results and discussion

For each subject, the (R)CSP filters were learnt on the training set available. The log-variances of the spatially filtered EEG signals were then used as the input features to an LDA classifier, one of the most popular and efficient classifier for BCI design [2]. Table II report the classification accuracies (mean, median and standard deviation) obtained when classifying the EEG signals from the test sets.

These results first show that, with the exception of DLCSPPauto and CCSP2, all RCSP algorithms outperformed the classical CSP, often substantially. The best RCSP algorithms outperformed CSP by about 3 or 4% in mean classification accuracy and by nearly 10% in median classification accuracy. This confirms that when using CSP, we should use regularization in order to deal with its non-robust nature. The classification accuracy variance is also lower for the RCSP algorithms. This together with a closer look at the results suggest that RCSP algorithms tend to particularly increase the performances for subjects with poor initial performances (e.g., A3, A5, B2, C5, ...), while leaving performances of already good subjects roughly unchanged. This makes sense as the purpose of the various regularization forms is to deal with noisy or limited data, but not necessarily to improve performance for already good and clean data.

The best RCSP algorithm on these data is the WTRCSP that we proposed in this paper, as it reached both the highest median and mean accuracy. It is only slightly better than the TRCSP, also proposed in this paper. These two algorithms each having only a single hyperparameter to tune ( $\alpha$ ), they also appear as more convenient to use and more computationally efficient than other good RCSP algorithms such as GLRCSP or SRCSP which both have two hyperparameters.

Interestingly enough, it seems that, in general, regularizing the CSP objective function is more rewarding, in terms of obtained classification performances, than regularizing the covariance matrices. An explication could be that regularizing the objective function means incorporating a-priori directly at the spatial filters level (which we aim to improve), whereas regularizing the covariance matrices is only an indirect way of doing so, and therefore may not be optimal.

<sup>1</sup>[http://ida.first.fhg.de/projects/bci/competition\\_iv/](http://ida.first.fhg.de/projects/bci/competition_iv/)

TABLE II

CLASSIFICATION ACCURACIES (%) OBTAINED FOR EACH SUBJECT FOR THE STANDARD CSP AND THE REGULARIZED CSP ALGORITHMS PRESENTED IN THIS PAPER. FOR EACH SUBJECT, THE BEST RESULT IS DISPLAYED IN BOLD CHARACTERS. THE OVERALL MEAN, MEDIAN AND STANDARD DEVIATION (STD) ARE ALSO PROVIDED.

Subject	BCI competition III									BCI competition IV									Overall		
	data set IVa					data set IIIa				data set IIa									Mean	Median	Std
	A1	A2	A3	A4	A5	B1	B2	B3	C1	C2	C3	C4	C5	C6	C7	C8	C9				
CSP	66.07	96.43	47.45	71.88	49.6	95.56	61.67	93.33	<b>88.89</b>	51.39	96.53	70.14	54.86	<b>71.53</b>	81.25	93.75	<b>93.75</b>	75.5	71.9	18.2	
GLRCSP	<b>72.32</b>	96.43	<b>66.84</b>	67.86	<b>89.29</b>	95.56	61.67	90	86.11	58.33	93.75	67.36	55.56	65.28	81.25	93.75	88.19	78.2	<b>81.3</b>	14.3	
CCSP1	66.96	96.43	63.27	71.88	84.92	<b>98.89</b>	45	93.33	86.11	60.42	93.75	56.94	49.31	65.28	81.25	93.75	88.19	76.2	<b>81.3</b>	17.5	
CCSP2	65.18	96.43	45.41	71.88	49.6	95.56	61.67	93.33	<b>88.89</b>	53.47	<b>97.22</b>	70.14	54.17	68.06	79.17	95.14	90.28	75	71.9	18.3	
DLCSPauto	66.96	96.43	46.94	71.43	50	94.44	63.33	95	<b>88.89</b>	51.39	96.53	70.14	56.94	<b>71.53</b>	<b>81.94</b>	93.75	<b>93.75</b>	75.9	71.5	18	
DLCSPcv	64.29	96.43	52.04	71.88	82.54	95.56	<b>78.33</b>	93.33	<b>88.89</b>	50.69	96.53	70.14	55.56	62.5	81.25	93.75	86.81	77.7	<b>81.3</b>	16.1	
SSRCSP	70.54	96.43	53.57	71.88	75.39	95.56	61.67	<b>96.67</b>	<b>88.89</b>	53.47	<b>97.22</b>	70.14	56.25	68.75	79.17	<b>97.22</b>	90.28	77.8	75.4	16.2	
TRCSP	71.43	96.43	63.27	71.88	86.9	<b>98.89</b>	56.67	93.33	<b>88.89</b>	54.17	96.53	<b>70.83</b>	62.5	67.36	81.25	95.87	91.67	79.3	<b>81.3</b>	15.3	
WTRCSP	69.64	<b>98.21</b>	54.59	71.88	85.32	<b>98.89</b>	71.67	93.33	<b>88.89</b>	54.86	96.53	70.14	<b>65.97</b>	61.81	81.25	95.83	90.97	<b>79.4</b>	<b>81.3</b>	15.3	
SRCS	<b>72.32</b>	96.43	60.2	<b>77.68</b>	86.51	96.67	53.33	93.33	<b>88.89</b>	<b>63.19</b>	96.53	66.67	63.19	63.89	78.47	95.83	92.36	79.2	78.5	15.2	

Results obtained by RCSP algorithms based on other subjects data (i.e., CCSP, GLRCSP, SSRCSP and WTRCSP), also showed that subject-to-subject transfer in BCI is possible and valuable. Despite large inter-subject variabilities, knowing that using data from other subjects can improve performances may also benefit to other EEG signal processing and classification algorithms. Among RCSP methods using only data from other subjects to regularize the covariance matrices (i.e., CCSP1, CCSP2 and SSRCSP), SSRCSP reached the highest mean accuracy (although CCSP1 reached an higher median accuracy) and the lowest variance, which suggests that it is worth selecting subjects to build a generic covariance matrix.

Concerning the poor performances of DLCSPauto, it should be mentioned that in one of our previous studies [15], we also observed poor performance with this method when all training data were used. However, when using this approach with a small training set, DLCSPauto proved to be substantially more efficient than CSP. This suggests DLCSPauto is only useful when very few training data is available.

The fact that RCSP algorithms led to lower score than CSP on a few subjects is also surprising. Indeed, if the best results could be obtained without regularization, then we could expect that the hyperparameter selection procedure based on cross-validation would figure it out and set the regularization parameter to 0. However, it seems it was sometimes different. This may suggest that, maybe due to the non-stationarity of EEG, cross-validation is not a very good predictor of generalization performances for BCI. This has been also observed in one subject in [11]. We will investigate this issue in the future.

Figure 1 shows some examples of spatial filters obtained with different (R)CSP algorithms, for different subjects. In general, these pictures show that CSP filters appear as messy, with large weights in several unexpected locations from a neurophysiological point of view. On the contrary, RCSP filters are generally smoother and physiologically more relevant, with strong weights over the motor cortex areas, as expected from the literature [1]. This suggests that, in addition to be more efficient than the standard CSP algorithm, RCSP algorithms also lead to filters that are neurophysiologically more relevant and as such more interpretable.

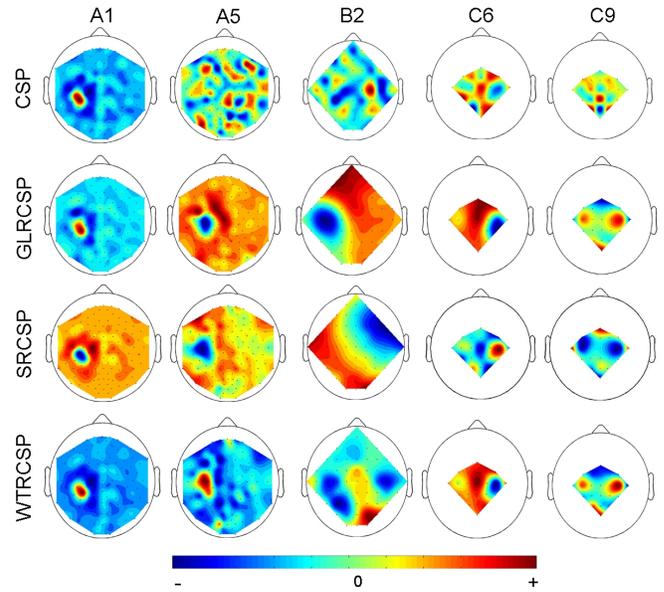


Fig. 1. Electrode weights for corresponding filters obtained with different CSP algorithms (CSP, GLRCSP, SRCSP and WTRCSP), for subjects A1, A5 (118 electrodes), B2 (60 electrodes) and C6, C9 (22 electrodes).

## VI. CONCLUSION

In this paper, we presented a simple theoretical framework to design Regularized CSP. We proposed a mini-review of existing RCSP algorithms, and presented how to cast them in this framework. We also proposed 4 new algorithms within this framework: SSRCSP, TRCSP, WTRCSP and SRCSP. We performed evaluations of 10 different RCSP algorithms (including these 4 new ones and the original CSP), on EEG data from 17 subjects, from BCI competition data sets. Results showed that the best RCSP can outperform CSP by nearly 10% in median classification accuracy and lead to more neurophysiologically relevant spatial filters. They also showed that regularizing CSP was a way of performing efficient subject-to-subject transfer. Overall, the best RCSP algorithms on these data were the CSP with Tikhonov Regularization and Weighted

Tikhonov Regularization, both newly proposed in this paper. Therefore, we would recommend BCI designers using CSP to adopt regularized CSP algorithms, such as TRCSP, in order to obtain more robust and efficient systems. To encourage such an adoption, as well as to ensure the replicability of the present study, a Matlab toolbox containing all the CSP algorithms evaluated in this paper is made freely available to the community at the following location: [http://sites.google.com/site/fabienlotte/code-and-softwares/rcsp\\_toolbox](http://sites.google.com/site/fabienlotte/code-and-softwares/rcsp_toolbox).

Future work could deal with investigating performances of RCSP algorithms - particularly those performing subject-to-subject transfer - with very small training sets, so as to reduce BCI calibration time, in the line of our previous studies [15][27]. It could also be interesting to adapt the presented regularization framework to multiclass CSP approaches based on approximate joint diagonalization, such as [28][29].

#### ACKNOWLEDGMENTS

The authors would like to thank Dr. Schlögl and Dr. Blankertz for providing the electrode coordinates of BCI competition data. We would also like to thank the Berlin and Graz groups for providing the BCI competition data sets used in this study.

#### REFERENCES

- [1] G. Pfurtscheller and C. Neuper, "Motor imagery and direct brain-computer communication," *proceedings of the IEEE*, vol. 89, no. 7, pp. 1123–1134, 2001.
- [2] F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, and B. Arnaldi, "A review of classification algorithms for EEG-based brain-computer interfaces," *Journal of Neural Engineering*, vol. 4, pp. R1–R13, 2007.
- [3] H. Ramoser, J. Müller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial EEG during imagined hand movement," *IEEE Transactions on Rehabilitation Engineering*, vol. 8, no. 4, pp. 441–446, 2000.
- [4] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K.-R. Müller, "Optimizing spatial filters for robust EEG single-trial analysis," *IEEE Signal Proc Magazine*, vol. 25, no. 1, pp. 41–56, 2008.
- [5] 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, 2004.
- [6] B. Blankertz, K. R. Müller, D. J. Krusienski, G. Schalk, J. R. Wolpaw, A. Schlögl, G. Pfurtscheller, J. D. R. Millan, M. Schroder, and N. Birbaumer, "The BCI competition III: Validating alternative approaches to actual BCI problems," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 14, no. 2, pp. 153–159, 2006.
- [7] B. Blankertz, F. Losch, M. Krauledat, G. Dornhege, G. Curio, and K.-R. Müller, "The berlin brain-computer interface: Accurate performance from first-session in BCI-naive subjects," *IEEE Trans Biomed Eng*, vol. 55, no. 10, pp. 2452–2462, 2008.
- [8] K. Ang, Z. Chin, H. Zhang, and C. Guan, "Filter bank common spatial pattern (FBCSP) in brain-computer interface," in *IEEE World Congress on Computational Intelligence & International Joint Conference on Neural Networks (IJCNN)*, 2008, pp. 2391–2398.
- [9] B. Reuderink and M. Poel, "Robustness of the common spatial patterns algorithm in the BCI-pipeline," HMI, University of Twente, Tech. Rep., 2008.
- [10] M. Grosse-Wentrup, C. Liefhold, K. Gramann, and M. Buss, "Beamforming in noninvasive braincomputer interfaces," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 4, pp. 1209 – 1219, 2009.
- [11] B. Blankertz, M. Kawanabe, R. Tomioka, F. Hohlefeld, V. Nikulin, and K.-R. Müller, "Invariant common spatial patterns: Alleviating non-stationarities in brain-computer interfacing," in *Advances in Neural Information Processing Systems 20*, In . MIT Press, Cambridge, MA, 2008.
- [12] H. Lu, K. Plataniotis, and A. Venetsanopoulos, "Regularized common spatial patterns with generic learning for EEG signal classification," in *Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC 2009)*, 2009, pp. 6599 – 6602.
- [13] H. Kang, Y. Nam, and S. Choi, "Composite common spatial pattern for subject-to-subject transfer," *IEEE Signal Processing Letters*, vol. 16, no. 8, pp. 683 – 686, 2009.
- [14] F. Lotte and C. Guan, "Spatially regularized common spatial patterns for eeg classification," in *International Conference on Pattern Recognition (ICPR)*, 2010.
- [15] —, "Learning from other subjects helps reducing brain-computer interface calibration time," in *International Conference on Audio, Speech and Signal Processing (ICASSP'2010)*, 2010, pp. 614–617.
- [16] J. Farquhar, N. Hill, T. Lal, and B. Schölkopf, "Regularised CSP for sensor selection in BCI," in *Proceedings of the 3rd international BCI workshop*, 2006.
- [17] X. Yong, R. Ward, and G. Birch, "Sparse spatial filter optimization for EEG channel reduction in brain-computer interface," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2008, pp. 417–420.
- [18] O. Ledoit and M. Wolf, "A well-conditioned estimator for large-dimensional covariance matrices," *Journal of Multivariate Analysis*, vol. 88, no. 2, pp. 365–411, 2004.
- [19] C. Vidaurre, N. Krämer, B. Blankertz, and A. Schlögl, "Time domain parameters as a feature for EEG-based brain computer interfaces," *Neural Networks*, vol. 22, pp. 1313–1319, 2009.
- [20] P. Pudil, F. J. Ferri, and J. Kittler, "Floating search methods for feature selection with nonmonotonic criterion functions," *Pattern Recognition*, vol. 2, pp. 279–283, 1994.
- [21] A. Tikhonov, "Regularization of incorrectly posed problems," *Soviet Math.*, vol. 4, pp. 1624–1627, 1963.
- [22] Z. Xiang, Y. Xi, U. Hasson, and P. Ramadge, "Boosting with spatial regularization," in *Advances in Neural Information Processing Systems (NIPS)*, 2009.
- [23] D. Cai, X. He, Y. Hu, J. Han, and T. Huang, "Learning a spatially smooth subspace for face recognition," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2007.
- [24] G. Dornhege, B. Blankertz, G. Curio, and K. Müller, "Boosting bit rates in non-invasive EEG single-trial classifications by feature combination and multi-class paradigms," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 6, pp. 993–1002, 2004.
- [25] A. Schlögl, F. Lee, H. Bischof, and G. Pfurtscheller, "Characterization of four-class motor imagery EEG data for the BCI-competition 2005," *Journal of Neural Engineering*, pp. L14–L22, 2005.
- [26] M. Naeem, C. Brunner, R. Leeb, B. Graimann, and G. Pfurtscheller, "Seperability of four-class motor imagery data using independent components analysis," *Journal of Neural Engineering*, vol. 3, pp. 208–216, 2006.
- [27] F. Lotte and C. Guan, "An efficient P300-based brain-computer interface with minimal calibration time," in *Assistive Machine Learning for People with Disabilities symposium (NIPS'09 Symposium)*, 2009.
- [28] M. Grosse-Wentrup and M. Buss, "Multi-class common spatial pattern and information theoretic feature extraction," *IEEE Transactions on Biomedical Engineering*, vol. 55, no. 8, pp. 1991–2000, 2008.
- [29] C. Gouy-Pailler, M. Congedo, C. Brunner, C. Jutten, and G. Pfurtscheller, "Nonstationary brain source separation for multi-class motor imagery," *IEEE transactions on Biomedical Engineering*, vol. 57, no. 2, pp. 469–478, 2010.