

Published in final edited form as:

J Neural Eng. 2013 February ; 10(1): 016002. doi:10.1088/1741-2560/10/1/016002.

ADAPTIVE LAPLACIAN FILTERING FOR SENSORIMOTOR RHYTHM-BASED BRAIN-COMPUTER INTERFACES

Jun Lu¹, Dennis J. McFarland², and Jonathan R. Wolpaw²

¹Guangdong University of Technology, Guangzhou, China 510006

²Laboratory of Neural Injury and Repair, Wadsworth Center, New York State Department of Health and State University of New York, Albany, NY 12201

Abstract

Objective—Sensorimotor rhythms (SMRs) are 8–30 Hz oscillations in the EEG recorded from the scalp over sensorimotor cortex that change with movement and/or movement imagery. Many brain-computer interface (BCI) studies have shown that people can learn to control SMR amplitudes and can use that control to move cursors and other objects in one, two, or three dimensions. At the same time, if SMR-based BCIs are to be useful for people with neuromuscular disabilities, their accuracy and reliability must be improved substantially. These BCIs often use spatial filtering methods such as common average reference (CAR), Laplacian (LAP) filter or common spatial pattern (CSP) filter to enhance the signal-to-noise ratio of EEG. Here we test the hypothesis that a new filter design, called an “adaptive Laplacian (ALAP) filter,” can provide better performance for SMR-based BCIs.

Approach—An ALAP filter employs a Gaussian kernel to construct a smooth spatial gradient of channel weights, and then simultaneously seeks the optimal kernel radius of this spatial filter and the regularization parameter of linear ridge regression. This optimization is based on minimizing leave-one-out cross-validation error through a gradient descent method, and is computationally feasible.

Main results—Using a variety of kinds of BCI data from a total of 22 individuals, we compare the performances of ALAP filter to CAR, small LAP, large LAP and CSP filter. With a large number of channels and limited data, ALAP performs significantly better than CSP, CAR, small LAP and large LAP both in classification accuracy as well as in mean squared error. Using fewer channels restricted to motor areas, ALAP is still superior to CAR, small LAP and large LAP, but equally matched to CSP.

Significance—Thus, ALAP may help to improve the accuracy and robustness of SMR-based BCIs.

Index Terms

Brain computer interface (BCI); brain-machine interface (BMI); spatial filter; electroencephalogram (EEG); leave-one-out (LOO) cross-validation; assistive communication

1 INTRODUCTION

A principal aim of brain-computer interface (BCI) research is to establish new communication channels that translate brain signals into control commands for output

devices such as computer applications or neuroprosthesis to be used by people with severe neuromuscular disabilities (Wolpaw and Wolpaw 2012). In past decades, various noninvasive (Blankertz *et al* 2007, Ortner *et al* 2011, Tianyou *et al* 2012) and invasive BCIs (Taylor *et al* 2002, Hochberg *et al* 2006, Leuthardt *et al* 2011) have been developed. For noninvasive BCIs, brain signals can be acquired by scalp-recorded electroencephalogram (EEG) from a human who tries to convey his/her intentions according to some defined paradigms. These paradigms can be exogenous such as steady-state visual evoked potentials (SSVEP)-based BCIs and P300-based BCIs (Wu and Yao 2008, Aloise *et al* 2011), or endogenous such as slow cortical potentials (SCP)-based BCIs and sensorimotor rhythms (SMRs)-based BCIs (Hinterberger *et al* 2004, McFarland *et al* 2010). Relatively, exogenous BCIs require minimal training and can be set-up easily, but they need permanent attention to external stimuli which may be fatiguing for some users. On the other hand, endogenous BCIs take much longer time for training, but they are independent of any stimulation and can be operated at will (Pfurtscheller and Scherer 2010, Nicolas-Alonso and Gomez-Gil 2012).

Here we focus on SMR-based BCIs using EEG recording. SMRs (8–14 Hz μ rhythm and 14–30 Hz β rhythm) are recorded over sensorimotor cortex. The amplitudes of SMRs decrease (event-related desynchronization) and/or increase (event-related synchronization) in association with movement and movement imagery (Pfurtscheller and da Silva 1999). People with or without motor disabilities can learn to change SMR amplitudes and can use that control to operate a BCI (Wolpaw and McFarland 2004, Liao *et al* 2007, Bai *et al* 2008, McFarland *et al* 2010, Blankertz *et al* 2010, Zhang and Guan 2010). However, the accuracies and reliability of these BCIs are reduced by unrelated brain signals (e.g., visual α rhythms) and non-brain artifacts (e.g., electromyogram (EMG) and electrooculogram (EOG)). Thus, the design of effective algorithms for extracting SMRs and translating them into BCI outputs is challenging (e.g., Blankertz *et al* 2006). Many SMR-based BCIs seek to increase performance by using spatial filtering methods such as the common average reference (CAR), Laplacian (LAP) filtering, or common spatial pattern (CSP) algorithms (e.g., Wolpaw *et al* 2002, McFarland *et al* 1997, Blankertz *et al* 2008). And each method has its own set of advantages and disadvantages. The idea behind the CAR is to remove the averaged brain activity, which can be seen as EEG noise (Dien 1998). But since CAR is a constant global spatial filter, the noise from a local area may diffuse to all the other channels (i.e., electrodes). LAP subtracts the average of its neighbors from the channel of interest. LAP enhances local activity from local sources, and reduces widely distributed activity, including that from distant sources (e.g., EMG, eye movements and visual α rhythm) (Mouriño *et al* 2001). But LAP filters need to determine the optimal spacing of the neighbor channels, which requires time-consuming ad hoc manual tuning, such as small Laplacian (SLAP) and large Laplacian (LLAP) filter (McFarland *et al* 1997). Since both CAR and LAPs are unsupervised algorithms which do not involve any class information, they are simple but robust and generally can be used in SMR-based BCIs at the beginning of an experiment. CSP is a supervised data driven algorithm. It optimizes spatial filters to project multichannel EEG into a linear subspace so that the difference between the average variances of two-class mental tasks is maximized (Ramoser *et al* 2000). CSP requires a considerable amount of training data in order to insure a stable and good performance (Sannelli *et al* 2010). It usually performs better than CAR and LAP in the later phase of the experiment. But when the training data is limited and the signal-to-noise ratio (SNR) of EEG is poor, CSP is prone to overfitting that make it performs worse than LAP (Sannelli *et al* 2010). Based on this fact, some robust variations of CSP have been proposed such as CSP patches (Sannelli *et al* 2011) and regularized CSP (Lotte and Guan 2011). These variations reduce the complexity (or flexibility) of CSP model by restricting CSPs to the subsets of channels or by use of a penalty function based on the prior information. Despite their successes, CSPs are not directly designed for solving a regression problem. With the

developments of SMR-based BCIs, regression is preferable to classification because it is better suited to controlling continuous cursor movements in real time and it generalizes more readily to novel target locations (McFarland and Wolpaw 2005, Fruitet *et al* 2010). Unlike CSP, CAR and LAPs can not only be used for classification but also for regression.

Therefore, based on the models of CAR and LAP, we proposed a supervised data driven algorithm, called adaptive Laplacian (ALAP) filtering algorithm with the expectation to keep a proper tradeoff between learning ability and model complexity. The ALAP algorithm is robust with small training sets, and can address prediction problems for classification and regression. Specifically, ALAP employs the Gaussian kernel to construct a continuous slope and then models the instantaneous voltage of each channel by subtracting the weighted average of all channels. ALAP behaves like a high pass filter in the spatial domain. Both CAR and SLAP filters can be derived from the ALAP framework by adopting a particular kernel parameter. In contrast to CSP, ALAP unifies the optimizations of spatial filtering and the prediction algorithm towards the same goal (i.e., minimizing the error of leave-one-out (LOO) cross validation (CV)). Here we used linear ridge regression (LRR) (Hoerl and Kennard 1970) as the prediction algorithm. The whole model has only two parameters (i.e., the kernel radius of ALAP and the regularization coefficient of LRR). To accelerate the optimization procedure, we design a gradient descent method that employs the closed form of LRR LOO error (Cook and Weisberg 1982) and explicitly compute the derivatives about the kernel radius and the regularization coefficient. This optimization technique is inspired by the study of Bo *et al* (2006). However, we extend it from the feature scaling for kernel Fisher discriminate analysis (KFDA) to the radius selection for ALAP spatial filtering. The circular Laplacian filter (Song and Epps, 2006) is perhaps most closely related to ALAP. The circular Laplacian filter is based on Perrin's spherical model (Perrin *et al.*, 1989). It re-references EEG using interpolated values on a circle around an electrode and behaves like a band-pass filter in the spatial domain. Perrin *et al.* (1989) suggested that a circular Laplacian filter could produce better classification results than LAPs with subject-specific radius, although they did not show how to select the optimal radius.

This paper is organized as follows. Section 2 describes CAR, LAP, and CSP filters. Section 3 describes the ALAP and indicates its relationships to the other filters. Section 4 describes the data used for comparing the filters, the comparison procedure and the results. Finally, section 5 summarizes the properties of the ALAP filter and considers directions for further research.

2 CAR, LAP, AND CSP SPATIAL FILTERS

2.1 Common Average Reference (CAR)

The CAR subtracts the average value of the entire channel montage from each channel of interest, so that noise common to a large proportion of the channels is reduced (Bertrand *et al* 1985). If the entire head is covered by equally spaced channels and the potential on the head is generated by point sources, the CAR results in a spatial voltage distribution with a mean of zero (Bertrand *et al* 1985). While the assumptions of uniform and complete channel coverage as well as point sources are usually not met completely in practice, the CAR provides EEG recordings that are nearly reference-free. Because the common average emphasizes EEG components that are present in a large proportion of the EEG channels, the subtraction of the common average reduces these components and thereby functions as a high-pass spatial filter (i.e., accentuates EEG components with highly focused spatial distributions). On the other hand, a component that is present in some channels but absent or minimal in a channel of interest may appear in that channel in inverted form (i.e., as a “ghost potential”) (Desmedt *et al* 1990). In brief, CAR is a constant global filter method that does not optimize the local spatial structure of multichannel EEG signals.

The CAR is computed according to the formula,

$$\mathbf{x}_i^{CAR} = \mathbf{x}_i^{ER} - \sum_{j=1}^C \mathbf{x}_j^{ER} / C \quad (1)$$

where $\mathbf{x}_i^{ER} \in R^{1 \times T}$ is the potential between channel i and the reference, T is the number of sampled time points in a trial and C is the number of channels in the montage.

2.2 Laplacian (LAP)

The LAP calculates for each channel the second derivative of the instantaneous spatial voltage distribution, and thereby emphasizes activity originating in radial sources immediately below the electrode. Thus, it is a high-pass spatial filter that accentuates localized activity and reduces more diffuse activity. In practice, the value of the LAP for each channel is calculated as the difference between that channel and a weighted average of the surrounding channels (Nunez *et al* 1994). Thus, the selection of surrounding channels affects the result. As their distances from the channel of interest decrease, the LAP becomes more sensitive to the components with higher spatial frequencies.

Using data obtained in the study of SMR-based cursor control, McFarland *et al* (1997) compared LAP filters calculated with two different sets of surrounding channels: nearest neighbor channels (i.e., SLAP) and next-nearest neighbor channels (i.e., LLAP). They found that the LLAP generally performed better than the SLAP. This result suggested that SMRs were not highly focused and/or that their scalp locations varied over time. In any case, SMR spatial distributions and the distributions of unrelated brain and non-brain activity are likely to be complex and highly variable across and even within individuals. In addition, channel spacing differs with different EEG montages. In sum, the fixed selection of surrounding channels may weaken the overall performance of a LAP filter.

The LAP is computed according to the formula,

$$\mathbf{x}_i^{LAP} = \mathbf{x}_i^{ER} - \sum_{j \in S_i} g_{ij} \mathbf{x}_j^{ER} \quad (2)$$

where $g_{ij} = \frac{1/d_{ij}}{\sum_{j \in S_i} 1/d_{ij}}$, S_i is the set of channels surrounding channel i , and d_{ij} is the distance between channel i and channel j , $j \in S_i$.

2.3 Common Spatial Pattern (CSP)

The CSP method optimizes the spatial filter to map multichannel EEG signals to the linear subspace so that the variance of one class is maximized while the variance of the other class is minimized (Ramoser *et al* 2000). CSP has been proven to be one of the most popular and efficient algorithms for SMR-based BCIs, notably during BCI competitions (Blankertz *et al* 2006). However, since the number of estimated parameters of a CSP filter grows with the square of the number of channels, CSP is very sensitive to the noise and prone to overfitting when available training trials are limited (Grosse-Wentrup *et al* 2009, Lotte and Guan 2011). In addition, CSP is not directly designed for solving a regression problem. With the developments of SMR-based BCIs, regression is preferable to classification because it is better suited to controlling continuous cursor movements in real time and it generalizes more readily to novel target locations (McFarland and Wolpaw 2005, Fruitet *et al* 2010). The CSP algorithm is formulated as follows.

Let C by C matrices \mathbf{R}_a and \mathbf{R}_b denote the averaged normalized spatial covariance matrices of class a and b , respectively. The composite covariance matrix is $\mathbf{R} = \mathbf{R}_a + \mathbf{R}_b$. As \mathbf{R} is a symmetrical matrix, it can be factored into its eigenvectors by SVD.

$$\mathbf{R} = \mathbf{U}_0 \mathbf{\Lambda} \mathbf{U}_0^T$$

The whitening transformation of the composite covariance matrix \mathbf{P} is:

$$\mathbf{P} = \mathbf{\Lambda}^{-1/2} \mathbf{U}_0^T$$

By \mathbf{P} , the individual covariance matrices \mathbf{R}_a and \mathbf{R}_b are transformed to:

$$\mathbf{S}_a = \mathbf{P} \mathbf{R}_a \mathbf{P}^T \quad \mathbf{S}_b = \mathbf{P} \mathbf{R}_b \mathbf{P}^T$$

\mathbf{S}_a and \mathbf{S}_b share the same eigenvectors, $\mathbf{S}_a = \mathbf{B} \mathbf{\Psi}_a \mathbf{B}^T$, $\mathbf{S}_b = \mathbf{B} \mathbf{\Psi}_b \mathbf{B}^T$ and $\mathbf{\Psi}_a + \mathbf{\Psi}_b = \mathbf{I}$ where \mathbf{I} is the identity matrix. As a consequence, the projection matrix $\mathbf{W} = \mathbf{B}^T \mathbf{P}$ gives the mapping of each EEG trial $\mathbf{X} \in R^{C \times T}$; $\mathbf{Z} = \mathbf{W} \mathbf{X}$. The signals \mathbf{Z}_p ($p = 1, \dots, 2m$) with the maximum differences of variances are associated with the largest eigenvalues $\mathbf{\Psi}_a$ and $\mathbf{\Psi}_b$. These signals are the m first and m last rows of \mathbf{Z} due to the calculation of \mathbf{W} . The feature vector \mathbf{f} of each trial is extracted as:

$$\mathbf{f}_p = \log \left(\text{var}(\mathbf{Z}_p) / \sum_{i=1}^{2m} \text{var}(\mathbf{Z}_i) \right)$$

where \mathbf{f}_p is the p th component of \mathbf{f} , the log-transformation serves to make the data more closely approximate the normal distribution.

3 ADAPTIVE LAPLACIAN (ALAP) FILTER

An ALAP filter might improve SMR-based BCI performance because it is not subject to certain limitations of CAR and LAP filters. The CAR is a constant global spatial filter, and thus it cannot capture the local spatial distributions of EEG components such as SMRs. Although the LAP is a local spatial filter, it cannot adapt to the differences across channels and across subjects in the spatial distributions of unrelated brain signals or non-brain artifacts. Because they are not adaptive, CAR and LAP have no capacity to optimize the performances.

3.1 Framework of ALAP Filter

The ALAP filter can focus on the local spatial distributions of EEG components. A Gaussian kernel is used to construct a continuous slope and to model the instantaneous voltage of each EEG channel with its neighbors. An ALAP filter can be seen as a generalization of CAR and LAP filters, since both of them can be derived from the ALAP framework by specifying a particular kernel parameter. Unlike CSP where the spatial filter and the translation algorithm are optimized separately with different objectives, ALAP is a the wrapper method which means the spatial filter and the translation algorithm are both simultaneously optimized with the same goal, (i.e., minimizing the LOO error of prediction).

First, the local spatial filtering model based on a Gaussian kernel is formulated as:

$$\mathbf{x}_i^{ALAP} = \sum_{j=1}^C \frac{\omega_{ij}}{z_i} (\mathbf{x}_i^{ER} - \mathbf{x}_j^{ER}) \quad (3)$$

with:

$$\omega_{ij} = \exp(-\theta \|\mathbf{v}_i - \mathbf{v}_j\|_2^2), \theta \geq 0 \quad (4)$$

where \mathbf{v}_i and \mathbf{v}_j are the positions of channel i and channel j respectively; $z_i = \sum_{j=1}^C \omega_{ij}$ is the normalization item. Although the Gaussian kernel is usually defined as $\exp(-\|\mathbf{v}_i - \mathbf{v}_j\|_2^2 / \sigma^2)$, we employ the alternative form in Eq.(4) to facilitate the derivative calculations later in section 3.2. The purpose of the kernel weighting is to smoothly control the influence of each pair of channels depended on their distance. The kernel parameter θ can be seen as the scale factor of these distances. If θ is small, the value ω_{ij} of is less sensitive to the distance $\|\mathbf{v}_i - \mathbf{v}_j\|_2^2$. Thus, the ALAP becomes similar to a global spatial filter. Conversely, if θ is large, the ALAP becomes similar to a local spatial filter.

Second, since the power change of μ and β rhythms reflects the ERD/ERS, ALAP extracts the power features of band-pass filtered EEG defined as:

$$f_{ik} = \log(\|\mathbf{x}_{ik}^{ALAP}\|_2^2) \quad (5)$$

where f_{ik} is the feature from channel i and trial k . The log-transformation serves to approximate a normal distribution of the data, since the least square solution of LRR is optimal when the features belong to the normal distribution. Then f_{ik} is centered with zero mean,

$$f'_{ik} = f_{ik} - \sum_{k=1}^N f_{ik} / N \quad (6)$$

where N is the number of trials. Here we used LRR (Hoerl and Kennard 1970) as the prediction algorithm. LRR minimizes the penalized sum-of-squares error function by incorporating the l_2 -norm regularization to control the model complexity and improve the generalization performance. The prediction model is constructed as:

$$\min_{\alpha, \theta, \lambda} \frac{1}{2} \left(\|\mathbf{F}^T \alpha - \mathbf{y}\|_2^2 + \lambda \|\alpha\|_2^2 \right) \quad (7)$$

where $\mathbf{F} \in R^{C \times N}$ is the feature matrix $(\mathbf{F})_{ik} = f'_{ik}$, $\alpha \in R^C$ is the linear projection, $\mathbf{y} \in R^N$ is the centered label vector, and $\lambda \geq 0$ is the regularization coefficient controlling the bias-variance trade-off. Given θ and λ , the solution is:

$$\alpha = (\mathbf{F}\mathbf{F}^T + \lambda \mathbf{I})^{-1} \mathbf{F}\mathbf{y} \quad (8)$$

where \mathbf{I} denotes the $C \times C$ identity matrix.

3.2 Optimization with the LOO Error

Eq. (7) and (8), indicate that θ and λ have important roles in the ALAP filtering model and should be derived from EEG data of SMR-based BCI trials. One popular approach for model selection is CV (Stone 1974). The n -fold CV splits the data into n parts, and uses each

alternatively to train and to validate the model. The final performance is the mean of the performances on the n different validation sets. CV maximizes the total number of validation trials and potentially helps to protect against overfitting. When n equals the number of trials, n -fold CV is the LOO CV.

The LOO error is considered to be an almost unbiased estimator of the generalization error (Luntz and Brailovsky 1969) and has been used frequently in model selection (Ojeda *et al* 2008, Arlot and Celisse 2011, Yuan *et al* 2012). Moreover, the development of closed-form solutions for performing LOO CV in certain learning algorithms such as LRR (Cook and Weisberg 1982) and KFDA (Cawley and Talbot 2003) significantly reduces the computational complexities. Thus, the present study uses the LOO error as the optimization criterion in tuning the parameters of ALAP filtering. Specifically, we explicitly compute the derivatives of LOO error with respect to θ and λ , and then seek them by a gradient descent method. This procedure is based largely on the studies of feature scaling in kernel learning algorithms (Bo *et al* 2006, Cawley and Talbot 2007). The difference is that these authors focus on adjusting the kernel matrices, whereas we extend the method to control the spatial filtering model.

In accord with Cook and Weisberg (1982), the closed form of LOO error for ALAP is:

$$\mathbf{r} = (\mathbf{I} - \mathbf{H}) \mathbf{y} \odot \mathbf{E} \quad (9)$$

where \odot denotes an element-wise division, $\mathbf{r} \in \mathbb{R}^N$ is the residual error vector, the hat matrix $\mathbf{H} = \mathbf{F}^T (\mathbf{F} \mathbf{F}^T + \lambda \mathbf{I})^{-1} \mathbf{F}$, $\mathbf{E} = \mathbf{1} - \text{diag}(\mathbf{H})$ and $\mathbf{1}$ is a column vector of N ones. The aim of ALAP is to minimize the LOO error.

$$\min_{\theta, \lambda} J = \frac{1}{2} \|\mathbf{r}\|_2^2 \quad (10)$$

Then, according to the chain rule, the derivative of J with respect to θ can be expressed as

$$\frac{\partial J}{\partial \theta} = \mathbf{r}^T \frac{\partial \mathbf{r}}{\partial \theta} \quad (11)$$

And based on Eq. (9), the derivative of \mathbf{r} with respect to θ is given by Bo *et al* (2006):

$$\frac{\partial \mathbf{r}}{\partial \theta} = (\mathbf{I} - \mathbf{H}) \mathbf{y} \odot \mathbf{E} \odot \mathbf{E} \otimes \left(\text{diag} \left(\frac{\partial \mathbf{H}}{\partial \theta} \right) \right) - \left(\frac{\partial \mathbf{H}}{\partial \theta} \right) \mathbf{y} \odot \mathbf{E} \quad (12)$$

where \otimes denotes an element-wise product. Let $\mathbf{C} = \mathbf{F} \mathbf{F}^T + \lambda \mathbf{I}$, then the derivative of \mathbf{H} with respect to θ is computed as:

$$\frac{\partial \mathbf{H}}{\partial \theta} = \frac{\partial \mathbf{F}^T}{\partial \theta} \mathbf{C}^{-1} \mathbf{F} + \mathbf{F}^T \frac{\partial \mathbf{C}^{-1}}{\partial \theta} \mathbf{F} + \mathbf{F}^T \mathbf{C}^{-1} \frac{\partial \mathbf{F}}{\partial \theta} \quad (13)$$

Based on Selby (1974), the derivative of \mathbf{C}^{-1} with respect to θ is:

$$\begin{aligned} \frac{\partial \mathbf{C}^{-1}}{\partial \theta} &= -\mathbf{C}^{-1} \frac{\partial \mathbf{C}}{\partial \theta} \mathbf{C}^{-1} \\ &= -\mathbf{C}^{-1} \left(\frac{\partial \mathbf{F}}{\partial \theta} \mathbf{F}^T + \mathbf{F} \frac{\partial \mathbf{F}^T}{\partial \theta} \right) \mathbf{C}^{-1} \end{aligned} \quad (14)$$

Based on Eq. (6), the derivative of \mathbf{F} with respect to θ is:

$$\frac{\partial(\mathbf{F})_{ik}}{\partial\theta} = \frac{\partial f_{ik}}{\partial\theta} - \frac{1}{N} \sum_{k=1}^N \frac{\partial f_{ik}}{\partial\theta} \quad (15)$$

According to the log-power computation Eq. (5), the derivative of f_{ik} with respect to θ is:

$$\frac{\partial f_{ik}}{\partial\theta} = 2 \frac{\partial (\mathbf{x}_{ik}^{ALAP})}{\partial\theta} (\mathbf{x}_{ik}^{ALAP})^T / \|\mathbf{x}_{ik}^{ALAP}\|_2^2 \quad (16)$$

According to spatial filtering model Eq. (3), the derivative of \mathbf{x}_{ik}^{ALAP} with respect to θ is:

$$\frac{\partial (\mathbf{x}_{ik}^{ALAP})}{\partial\theta} = \sum_{j=1}^C \left(\frac{\partial (\omega_{ij}/z_i)}{\partial\theta} \right) (\mathbf{x}_{ik}^{ER} - \mathbf{x}_{jk}^{ER}) \quad (17)$$

where $\frac{\partial (\omega_{ij}/z_i)}{\partial\theta} = \frac{\partial \omega_{ij}}{\partial\theta} / z_i - \omega_{ij} z_i^{-2} \left(\sum_{j=1}^C \frac{\partial \omega_{ij}}{\partial\theta} \right)$. Furthermore, according to the kernel function Eq. (4):

$$\frac{\partial \omega_{ij}}{\partial\theta} = -\omega_{ij} \|\mathbf{v}_i - \mathbf{v}_j\|_2^2 \quad (18)$$

Combining Eq. (11)–(18) yields the derivative of the LOO error with respect to θ . The derivative of \mathbf{H} with respect to λ is:

$$\frac{\partial \mathbf{H}}{\partial \lambda} = -\mathbf{F}^T \mathbf{C}^{-1} \mathbf{C}^{-1} \mathbf{F} \quad (19)$$

Upon that, the derivative of J with respect to λ can be computed in a similar way. Since both θ and λ are positive, the parameterizations $(\theta, \lambda) = (e^\xi, e^\psi)$ are used to avoid these constraints, and then $\theta = e^\xi$, $\lambda = e^\psi$. The gradient decent method is implemented with the Quasi-Newton BFGS algorithm from the optimization toolbox of MATLAB. The step length is determined by the cubic polynomial line search procedure. Table 1 lists the pseudo-code of the ALAP filtering algorithm. The computational complexity of ALAP is estimated as:

$$L \times O(C^2 \times N + C^3 + C^3 \times N) + O(C^3)$$

where L is the iteration number, $C^2 \times N$ comes from the computation of the spatial filtering and feature extraction for all the channels and trials, the cubic item C^3 is the computational complexity of inverting matrix $\mathbf{C} = \mathbf{F}\mathbf{F}^T + \lambda \mathbf{I}$, $C^3 \times N$ comes from the computation of the derivative of \mathbf{F} with respect to θ , and $O(C^3)$ is the computational complexity of calculating \mathbf{a} (i.e., the last step in table 1).

3.3 ALAP Compared to Other Spatial Filter Approaches

3.3.1 Comparison to CSP—ALAP and CSP are both supervised algorithms. In the SMR-based BCI applications, they generally use as signal features the power in specific

frequency bands. However, the criteria and optimizations of them are quite different. First, ALAP minimizes the LOO error of prediction whereas CSP maximizes the ratio of average variances extracted from the EEG signal of different mental tasks. Second, ALAP employs a local spatial filtering model on all the channels, and cooperates with LRR to search for the optimal parameters, while CSP optimizes the spatial filters without the local constraint and performs as an independent feature extraction step. Third, the ALAP has only one filter parameter, while the CSP filter vector is equal in length to the number of channels. Thus, when the channel number is great and the sample size is small, ALAP may be more robust than CSP parameter. In addition, ALAP can solve regression problem, which plays an important role in more flexible BCI paradigms, such as multi-dimensional cursor control with multiple targets in each dimension (McFarland *et al* 2010).

3.3.2 Comparison to CAR and LAP—Although ALAP is a supervised method and CAR is an unsupervised method, CAR can be regarded as a special case of ALAP. With respect to the formula of ALAP (i.e., Eq. (3)), if $\theta = 0$, then $\omega_{ij} = 1$, and

$\mathbf{x}_i^{ALAP} = \sum_{j=1}^C (\mathbf{x}_i^{ER} - \mathbf{x}_j^{ER}) / C$. ALAP and LAP are both local spatial filter algorithms. But their distance measures and scale are different. This is illustrating by rewriting Eq. (3) as:

$$\mathbf{x}_i^{ALAP} = A_i \left(\mathbf{x}_i^{ER} - \sum_{j \neq i} h_{ij} \mathbf{x}_j^{ER} \right) \quad (20)$$

where the scale coefficient: $A_i = 1 - \frac{1}{1 + \sum_{j \neq i} \exp(-\theta \|\mathbf{v}_i - \mathbf{v}_j\|_2^2)}$, and the weight of channel j :

$h_{ij} = \frac{\exp(-\theta \|\mathbf{v}_i - \mathbf{v}_j\|_2^2)}{\sum_{j \neq i} \exp(-\theta \|\mathbf{v}_i - \mathbf{v}_j\|_2^2)}$. In contrast to the LAP filter formula (i.e., Eq. (2)), ALAP uses $\exp(-\theta \|\mathbf{v}_i - \mathbf{v}_j\|_2^2)$ as the distance between channels i and j where parameter θ smoothly controls the neighbor of channel i . If $\theta \rightarrow +\infty$, only the nearest channels are used as the reference and the ALAP is similar to the SLAP except that the scale is smaller. Note that LLAP cannot be obtained within the framework of ALAP, because the nearest neighbor channels are not used in the LLAP.

Figure 1 illustrates the spatial frequency responses of the CAR, SLAP, LLAP and ALAP filter. Suppose 25 channels are equally spaced as a 5×5 matrix, the nearest neighbor distance is 1, and the channel of interest is at the center. By Eq. (1)-(4), the channel weights of these spatial filters can be computed. Then the two-dimensional Fourier transform is performed. The resulting response functions of the filters are displayed in figure 1. Figure 1 shows that the SLAP focuses more on the high frequency components than does the CAR, and that, as θ increases, ALAP changes from CAR to a SLAP and its response magnitude becomes smaller. Unlike ALAP, the LLAP is a band-pass filter that also suppresses the high frequency components.

4 THE BCI PERFORMANCES OF THE FOUR DIFFERENT SPATIAL FILTERS

4.1 EEG Data

In order to assess the utility of the ALAP algorithm, we compared it to CSP, CAR, SLAP, and LLAP using the EEG data of 22 subjects. These subjects were from three publicly available data sets of BCI competitions and from this laboratory. Their affiliations are listed in table 2. The prediction tasks of all the data sets are based on single-trial structure. And the data sets are described as follows.

4.1.1 Data Set IV, BCI Competition II—The goal of Data set IV, from BCI competition II (Blankertz *et al* 2002), was to predict the laterality of upcoming finger movements (left vs. right hand) 130 ms before a key press. There were 416 trials (316 trials for training and 100 trials for test) acquired from one subject (A1), and the EEG was recorded with 28 channels mainly covering the primary or sensorimotor cortices bilaterally. This data set did not provide the two-dimensional channel position file but declared that the channels followed the international 10/20-system. Thus, we used the standard 10/20-system position file from EEGLAB for the channels (Delorme and Makeig 2004).

4.1.2 Data Set IVa, BCI Competition III—Data set IVa, from BCI competition III (Dornhege *et al* 2004), consists of EEG signals from five subjects who performed right hand and foot motor imagery (MI). The EEG was recorded with 118 channels. A training set and a test set are available for each subject. The relative sizes of these two sets were different for each subject. More precisely, a total of 280 trials were available for each subject, and the training set sizes were 168, 224, 84, 56 and 28 for subjects B1–B5 respectively. The test set consisted of the remaining trials.

4.1.3 Data Set I, BCI Competition IV—Data set I, from BCI competition IV (Blankertz *et al* 2007), contains 59-channel EEG signals recorded from 4 healthy subjects (C1–C4) while they performed two-class MIs (i.e., left hand/right hand or hand/foot). The original goal of this data set was to apply a classifier to continuous EEG. For the purpose of the present study, only the calibration data with cue information were used. More specifically, the calibration data (consisting of two runs totaling 200 trials, balanced between the two classes) for each subject were split into two parts, the first 100 trials for training, and the next 100 trials for test. In addition, we changed the class labels from $\{-1, 1\}$ to $\{1, 2\}$ to compare prediction performance with the other data sets in this study.

4.1.4 Data Set of Cursor Movement Control—Our own data set contains EEG signals from 12 subjects (D1–D12) who modulated mu or beta rhythm to control vertical cursor movement toward one of two targets located at different heights along the right edge of the video screen. The details of the online control protocol are given in McFarland and Wolpaw (2008). The EEG was recorded with 64 channels covering the whole scalp followed the international 10/20-system. A training set and a test set were available for each subject. Each set contained 50–60 trials for each target. For the purpose of this study, an offline analysis was performed.

4.2 The Results

4.2.1 Preprocessing and Measures of Filter Performance—In this paper, we considered the discrete classification of single-trial EEG. For finger movement prediction and cursor movement control datasets, we used all the sampled time points of each trial; for the MI classification datasets, we used the sampled time points located from 0.5s to 2.5s after the cue instructing the subject to perform the MI task (Lotte and Guan 2011). The EEG data of all the subjects were individually band-pass filtered with 5th-order Butterworth filters. In the Appendix, we give a heuristic approach for frequency band selection. As spatial filtering, with CSP, we used m pairs of the most discriminative filters, where m is determined by 5-fold cross validation (Blankertz *et al* 2008). That is, we used the filters corresponding to the m largest and m smallest eigenvalues for CSP; with the SLAP (LLAP), we calculated the reference using the nearest (next-nearest) neighboring channels for all the datasets. After frequency filtering and spatial filtering, the log-variances of the filtered EEG time series were extracted as the features. Then LRR was used as the translation algorithm. The regularization coefficient of LRR was tuned by LOO CV. Considering that the output of LRR is a continuous value, for classification we categorized each trial based on the distance

between the prediction and class labels. Suppose that the class labels are c_1 and c_2 , and \hat{z}_i is the prediction for trial i . If $\|\hat{z}_i - c_1\|_2^2 \leq \|\hat{z}_i - c_2\|_2^2$ trial i is classified as class c_1 , otherwise it is classified as class c_2 . The performance of each filter is given by its classification accuracy (CA) and its mean squared errors (MSE). These two indices are defined as:

$$CA = (1 - M/N) \times 100\%$$

$$MSE = E[(z - \hat{z})^2]$$

where M is the number of incorrect trial classifications, z and \hat{z} are the true labels and predictions respectively, and E is the expectation function.

4.2.2 Convergence of the ALAP Algorithm—Using all the available channels for each subjects, we tested the ALAP algorithm with the SMR-based BCI data described in section 4.1 and observed the number of iterations it took before the stop criterion was met (i.e., before the decrease of the LOO error J expressed in Eq.(10) was less than 10^{-3}). For the optimization, we tried different initial values of kernel parameter and regularization coefficient, and then choose those that lead to the lowest LOO errors. Figure 2 plots the LOO errors calculated over iterations for all the subjects. Note that $J = MSE \times N/2$ (N is the number of training trials), the scales of LOO errors for different subjects could be much different due to the various sizes of training sets and different SNRs of EEG data.

Figure 2 shows that, for most subjects, the LOO errors decreased markedly in the early iterations, and rapidly converged to a final stable value. Across all of the 22 subjects, the ALAP algorithm took 3–17 iterations to meet the stop criterion, with a mean of 6.09 and a standard deviation of 3.01 iterations. The computation of the ALAP algorithm was performed using Matlab R2011b running on Windows 7 Professional SP1 64bit with Intel Core i7-2640M CPU 2.8GHz. The time for ALAP optimization for each subject ranged between 0.91 to 27.6s. Among them, the fastest was with 28 channels and 316 training trials, and the slowest was with 118 channels and 224 training trials. The time for ALAP prediction per trial for each subject was less than 5 ms. That indicates that we can use the optimized ALAP model for BCI prediction in real time, and update the ALAP model between runs or sessions (about 300 trials) within 1 s, if the number of channels is less than 28.

4.2.3 Optimized Parameters of the ALAP—Figure 3 demonstrates the optimized kernel parameter θ of ALAP for each subject with all the available channels. It indicates that, even in the same datasets (i.e., with the same channel montages and distance units) θ s differ across subjects. As previously noted, for small values of θ , the ALAP approximates the CAR and for large values of θ it approximates the SLAP (see the analysis in Section 3.3.2)). For example, based on the channel positions of cursor movement control dataset according to subjects D1-D12, figure 4 demonstrates the scalp maps of $-h_{ij}$ in Eq.(20) for channel C3 with different θ s. From figure 4 we can see that, when $\ln(\theta) = 0$, ALAP tended to subtract the weighted average components from all the other channels from channel C3 (i.e., performed near to CAR). When $\ln(\theta) = 6$, ALAP was close to SLAP but not exactly, since the channels were not absolutely evenly placed and the distances between a channel and its nearest four neighbors may not be uniform. Clearly, a single fixed value θ cannot perform optimally across all subjects. That is the motivation for developing the ALAP algorithm.

4.2.4 Comparison of the Scalp Topographies of the Spatial filters—Figure 5 shows for representative subjects with each filter the r^2 topographies for the correlation

between the training labels and the SMR features. For the ALAP as compared to the other filters, the r^2 values are higher and more sharply focused over sensorimotor areas. Furthermore, for the ALAP as compared to the other filters, r^2 values over other scalp areas are uniformly low.

Figure 6 presents the most discriminative spatial patterns of the CSP algorithm (Blankertz *et al* 2008) with all the available channels for the same group of subjects as shown in figure 5. Although the spatial patterns are not directly comparable to r^2 values, figure 6 demonstrates that the CSP algorithm learns brain patterns both within and outside the sensory motor areas. These brain pattern were probably influenced by artifacts and noises, since the classification accuracies obtained by CSP for these subjects were poor, especially for B3, C2 and D5 even no more than 70% (see table 3). This might due to the limited number of training trials which made CSP overfitted the artifacts and noises.

4.2.5 Comparison of BCI Performances—Table 3 and table 4 show the classification accuracies (CAs) and mean squared errors (MSEs) obtained for each subject with different spatial filters. On average, ALAP achieved the best performances both in CA (84.65%) and in MSE (0.1346). A one-way ANOVA with repeated measures indicated that the five filters differed significantly in both CA and MSE (d.f. =4, $F=4.6$, $p=2.1\times 10^{-3}$ and d.f. =4, $F=2.85$, $p=2.86\times 10^{-2}$, respectively). Additional two-group ANOVAs showed that the ALAP was significantly better than each of the other filters in both CA ($p=9.10\times 10^{-3}$) and MSE ($p=3.05\times 10^{-2}$). Moreover, using the same data with randomized labels, we calculated the CAs for these spatial filters with 20 different randomizations and found that the average level was between 49.59% and 50.22% for each filter. This test indicates that these methods were not biased. The performance of CSP was not markedly better than those of CAR in either CA or MSE, despite the presumed advantage of its supervised feature extraction. This is probably because we used a large number of channels for the MI classification experiments (118 and 59) and cursor movement control experiment (64), whereas most of the training sets contained no more than 120 trials. This lack of difference in performance may reflect overfitting of the CSP filter (Grosse-Wentrup *et al* 2009, Lotte and Guan 2011). In contrast, with the same number of channels and trials, the ALAP filter performed better, which may be attributed to its having fewer model parameters. The relative poor performances of the SLAP and LLAP filters probably resulted from inter-subject differences in the spatial frequencies and distributions of the SMR components and the noises. SLAP and LLAP always use the nearest or next-nearest neighbor channels and thus cannot adapt to inter-subject differences.

As a further investigation, we used 17 channels from motor areas (i.e., FC3, FC1, FCz, FC2, FC4, C5, C3, C1, Cz, C2, C4, C6, CP3, CP1, CPz, CP2, CP4) to compare the performances of the ALAP, CAR, CSP and SLAP filters. In this experiment, the LLAP filter was not included because the channels are few and none of them represent a full montage of next-nearest neighbors as described in McFarland *et al* (1997). Table 5 and table 6 show the CAs and MSEs obtained for each subject with different spatial filters using 17 channels. On average, CSP achieved the best performances in CA (85.02%), ALAP filter achieved the best performances in MSE (0.1304). A one-way ANOVA with repeated measures indicated that the four filters differed significantly in both CA and MSE (d.f. =3, $F=5.49$, $p=2.1\times 10^{-3}$ and d.f. =3, $F=3.63$, $p=1.74\times 10^{-2}$, respectively). Additional two-group ANOVAs showed that the ALAP filter was significantly better than CAR and SLAP in both CA ($p=8.70\times 10^{-3}$) and MSE ($p=3.82\times 10^{-2}$). There was little difference between ALAP and CSP either in CA ($p=0.52$) or MSE ($p=0.83$). Comparing table 5, 6 and table 3, 4, we can find that the performance of CSP was markedly improved when the channels were fewer and focused on motor areas. This was probably due to the fact that the complexity of CSP was reduced and potential noise from other brain area were reduced. Although the performance

of ALAP was still significantly better than CAR and SLAP, this advantage was decreased. The reason could be that the limited number of channels restricts the spatial resolution of the ALAP and makes the ALAP, CAR and SLAP more similar. For an extreme example of only 2 channels, it can be deduced that ALAP, CAR and SLAP are same based on Eq. (1)-(4). Comparing tables 5, 6 with tables 3, 4, we can find the average performances of the ALAP filter was more stable than the other filters. This may be due to the proper tradeoff between learning ability and model complexity.

5 DISCUSSION

In this paper, we analyzed the spatial filtering algorithms including CAR, LAPs as well as CSP, and then proposed the adaptive Laplacian (ALAP) filtering algorithm to address the small training set problem of SMR-BCIs. ALAP can vary between CAR-like and SLAP-like by adjusting the radius of the Gaussian kernel in order to translate each individual's EEG signals into BCI outputs as accurately as possible. For real BCI applications, many factors could influence the selection of a proper kernel radius. These include the channel montages of datasets, the spatial characteristics of noise (i.e., unrelated brain activity and non-brain activity) as well as the spatial characteristics of the SMR. Use of a fixed kernel radius could not adapt to these variations. We evaluated the spatial filters with EEG data from 22 subjects through offline analysis. The results showed that with the large number of channels and limited data, ALAP can outperform CSP, CAR, SLAP and LLAP both in classification accuracy as well as in mean squared error. With fewer channels, ALAP is still superior to CAR, SLAP and LLAP, but equally matched CSP.

An ALAP filter has three promising properties. First, ALAP integrates the optimizations of spatial filtering and the prediction algorithm to minimize the error of LOO CV. It improves BCI accuracy while reduces the risk of overfitting. Second, the LOO error of ALAP is expressed in a closed form, and the parameter derivatives can be explicitly calculated. This makes ALAP computationally feasible. Third, ALAP can address not only classification problems but also regression problems. This allows ALAP to provide greater flexibility, which is needed for BCI applications such as multi-dimensional cursor control with more than two targets in each dimension.

The most significant difficulty of the ALAP filter described here is the local minimum problem of gradient descent searching. A better solution might be to use evolutionary computation techniques, such as a genetic algorithm (Srinivas and Patnaik 1994) or particle swarm optimization (Kennedy and Eberhart 1995). However, these techniques would add computational complexity. As shown in section 3.2, the computation complexity of the ALAP filter is a cubic function of the channel numbers, which suggests that selecting fewer channels can dramatically accelerate the computation. Therefore, combining evolutionary computation and a channel selection technique is a potential approach to improving the ALAP filter.

The LOO error used here as the optimization criterion of the ALAP filter is a squared loss function (i.e., Eq.(10)). That is sensitive to extreme data points. Furthermore, it was originally designed for solving regression problems. In dealing with classification problems, it penalizes correctly classified data points if their output values are larger than the corresponding labels. To overcome this disadvantage, techniques to reshape the error, such as a sigmoid function (Bo *et al* 2006) or hinge loss (Bartlett and Wegkamp 2008), have been proposed. In the future, such methods may enhance the robustness of ALAP and provide a version specialized for classification problems.

In BCI experiments, small variations in the placement of the electrodes may occur. As a result, EEG signals are recorded from slightly different positions. This is a problem for all

spatial filter methods. For ALAP, when θ is small, the slight change (noise) can be reduced by averaging more channels, but more channels will be influenced by this noise; when θ is great, the influence of the noise will be restricted to the smaller area, but more influenced by each neighbor channel. In addition, for different channels (e.g., within/beyond the motor area), the variations in the placement may have a different influence on spatial filtering, and it is hard to analyze how sensitive the ALAP is to these effects. At present, we think that if the variations in the placement of the electrodes exists between the sessions within the training set, the ALAP filter will seek the optimal θ to minimize the estimation of generalization error (i.e., LOO error) in contrast to a fix θ , such as that of SLAP and CAR. If variations exists between the training sessions and test sessions, then information of the test set should be explored (e.g., use of a semi-supervised approach or a transfer learning approach). This problem should be examined in future work.

The time, frequency and spatial characters of SMR have close relationships with each other. Some studies have demonstrated that embedding the selection of time/frequency- dependent parameters into spatial filter algorithm may further enhance BCI performance (Lemm *et al* 2005, Dornhege *et al* 2006, Wu *et al* 2008, Wu *et al* 2011). Generally, it can be considered as an optimization problem in a border parameter space. Although the overfitting risk could be greater, it is another promising direction to improve the ALAP filter.

Acknowledgments

This work was supported in part by NIH NIBIB Bioengineering Research Partnership under Grant EB00856, National Natural Science Foundation of China under Grant 61273192, Natural Science Foundation of Guangdong Province under Grant S2011030002886 (team project), Natural Science Foundation of Guangdong Province under Grant 8351009001000002, Program for New Century Excellent Talents in University under Grant NCET-11-0911 and Special Scientific Funds approved in 2011 for the Recruited Talents by Guangdong Provincial universities.

References

- Arlot S, Celisse A. Segmentation of the mean of heteroscedastic data via cross-validation. *Statist Comput.* 2011; 21(4):613–32.
- Aloise F, Schettini F, Aricò P, Leotta F, Salinari S, Mattia D, Babiloni F, Cincotti F. P300-based brain–computer interface for environmental control: an asynchronous approach. *J Neural Eng.* 2011; 8:025025. [PubMed: 21436520]
- Bai O, Lin P, Vorbach S, Floeter MK, Hattori N, Hallett M. A high performance sensorimotor beta rhythm-based brain–computer interface associated with human natural motor behavior. *J Neural Eng.* 2008; 5:24–35. [PubMed: 18310808]
- Bartlett PL, Wegkamp MH. Classification with a reject option using a hinge loss. *J Mach Learn Res.* 2008; 9:1823–40.
- Bertrand O, Perrin F, Pernier J. A theoretical justification of the average reference in topographic evoked potential studies. *Electroenceph Clin Neurophysiol.* 1985; 62:462–64. [PubMed: 2415344]
- Blankertz B, Curio G, Müller K-R. Classifying single trial EEG: towards brain computer interfacing. In: T. G. Diettrich and S. Becker and Z. Ghahramani (eds.). *Advances in Neural Inf Proc Systems.* 2002; 14(NIPS 01)
- Blankertz B, Müller K-R, Krusienski DJ, Schalk G, Wolpaw JR, Schlogl A, Pfurtscheller G, Millan JR, Schroder M, Birbaumer N. The BCI competition III: validating alternative approaches to actual BCI problems. *IEEE Trans Neural Syst Rehabil Eng.* 2006; 14:153–9. [PubMed: 16792282]
- Blankertz B, Dornhege G, Krauledat M, Müller K-R, Curio G. The non-invasive Berlin brain-computer interface: fast acquisition of effective performance in untrained subjects. *NeuroImage.* 2007; 37(2):539–550. [PubMed: 17475513]
- Blankertz B, Tomioka R, Lemm S, Kawanabe M, Müller K-R. Optimizing spatial filters for robust EEG single-trial analysis. *IEEE Signal Proc Mag.* 2008; 25(1):41–56.

- Blankertz B, Sannelli C, Halder S, Hammer Eva M, Kübler A, Müller KR, Curio G, Dickhaus T. Neurophysiological predictor of SMR-based BCI performance. *NeuroImage*. 2010; 51(4):1303–9. [PubMed: 20303409]
- Bo L, Wang L, Jiao L. Feature scaling for kernel Fisher discriminant analysis using leave-one-out cross validation. *Neural Comput*. 2006; 18(4):961–78. [PubMed: 16494697]
- Cawley GC, Talbot NLC. Efficient leave-one-out cross validation of kernel Fisher discriminant classifiers. *Pattern Recognition*. 2003; 36:2585–92.
- Cawley GC, Talbot NLC. Preventing over-fitting in model selection via Bayesian regularisation of the hyper-parameters. *J Mach Learn Res*. 2007; 8:841–861.
- Cook, RD.; Weisberg, S. Residuals and influence in regression monographs on statistics and applied probability. Chapman and Hall; New York: 1982.
- Delorme A, Makeig S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics. *J Neurosci Methods*. 2004; 134:9–21. [PubMed: 15102499]
- Desmedt JE, Chalkin V, Tomberg C. Emulation of somatosensory evoked potential (SEP) components with the 3-shell head model and the problem of ‘ghost potential fields’ when using an average reference in brain mapping. *Electroenceph Clin Neurophysiol*. 1990; 77:243–58. [PubMed: 1695136]
- Dien J. Issues in the application of the average reference: review, critiques, and recommendations. *Behav Res Methods*. 1998; 30(1):34–43.
- Dornhege G, Blankertz B, Curio G, Müller K-R. Boosting bit rates in non-invasive EEG single-trial classifications by feature combination and multi-class paradigms. *IEEE Trans Biomed Eng*. 2004; 51(6):993–1002. [PubMed: 15188870]
- Dornhege G, Blankertz B, Krauledat M, Losch F, Curio G, Müller K-R. Combined optimization of spatial and temporal filters for improving brain-computer interfacing. *IEEE Trans Biomed Eng*. 2006; 53(11):2274–81. [PubMed: 17073333]
- Fruitet J, McFarland DJ, Wolpaw JR. A comparison of regression techniques for a two-dimensional sensorimotor rhythm-based brain–computer interface. *J Neural Eng*. 2010; 7:016003.
- Grosse-Wentrup M, Liefhold C, Gramann K, Buss M. Beamforming in noninvasive brain computer interfaces. *IEEE Trans Biomed Eng*. 2009; 56(4):1209–1219. [PubMed: 19423426]
- Hinterberger T, Schmidt S, Neumann N, Mellinger J, Blankertz B, Curio G, Birbaumer N. Brain-computer communication and slow cortical potentials. *IEEE Trans Biomed Eng*. 2004; 51(6): 1011–18. [PubMed: 15188872]
- Hochberg LR, Serruya MD, Friehs GM, Mukand JA, Saleh M, Caplan AH, Branner A, Chen D, Penn RD, Donoghue JP. Neuronal ensemble control of prosthetic devices by a human with tetraplegia. *Nature*. 2006; 442:164–171. [PubMed: 16838014]
- Hoerl AE, Kennard RW. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*. 1970; 42(1):80–86.
- Kennedy, J.; Eberhart, R. Particle swarm optimization. *Proceedings of IEEE Int. Conf. on Neural Networks IV*; 1995. p. 1942–48.
- Lemm S, Blankertz B, Curio G, Müller K-R. Spatio-spectral filters for improved classification of single trial EEG. *IEEE Trans Biomed Eng*. 2005; 52(9):1541–48. [PubMed: 16189967]
- Leuthardt EC, Gaona C, Sharma M, Szrama N, Roland J, Freudenberg Z, Solis J, Breshears J, Schalk G. Using the electrocorticographic speech network to control a brain-computer interface in humans. *J Neural Eng*. 2011; 8:036004. [PubMed: 21471638]
- Liao X, Yao D, Wu D, Li C. Combining spatial filters for the classification of single-trial EEG in a finger movement task. *IEEE Trans Biomed Eng*. 2007; 54(5):821–31. [PubMed: 17518278]
- Lotte F, Guan C. Regularizing common spatial patterns to improve BCI designs: unified theory and new algorithms. *IEEE Trans Biomed Eng*. 2011; 58(2):355–62. [PubMed: 20889426]
- Luntz A, Brailovsky V. On estimation of characters obtained in statistical procedure of recognition (In Russian). *Techicheskaya Kibernetika*. 1969; 3
- McFarland DJ, McCane LM, David SV, Wolpaw JR. Spatial filter selection for EEG-based communication. *Electroenceph Clin Neurophysiol*. 1997; 103:386–94. [PubMed: 9305287]

- McFarland DJ, Wolpaw JR. Sensorimotor rhythm-based brain–computer interface (BCI): feature selection by regression improves performance. *IEEE Trans Neural Syst Rehabil Eng.* 2005; 14:372–9. [PubMed: 16200760]
- McFarland DJ, Wolpaw JR. Sensorimotor rhythm-based brain–computer interface (BCI): model order selection for autoregressive spectral analysis. *J Neural Eng.* 2008; 5:155–62. [PubMed: 18430974]
- McFarland DJ, Sarnacki WA, Wolpaw JR. Electroencephalographic (EEG) control of three-dimensional movement. *J Neural Eng.* 2010; 7:36007.
- Mourino, J.; Millán, JdR; Cincotti, F.; Chiappa, S.; Jané, R.; Babiloni, F. Spatial filtering in the training process of a brain computer interface. *Proc. 23rd Annu. Int. Conf. Engineering in Medicine and Biology Soc.; Istanbul, Turkey.* 2001; 2001. p. 639-642.
- Nicolas-Alonso LF, Gomez-Gil J. Brain computer interfaces, a review. *Sensors.* 2012; 12:1211–79. [PubMed: 22438708]
- Nunez PL, Silberstein RB, Cadusch PJ, Wijesinghe RS, Westdorp AF, Srinivasan R. A theoretical and experimental study of high resolution EEG based on surface Laplacians and cortical imaging. *Electroenceph Clin Neurophysiol.* 1994; 90:40–57. [PubMed: 7509273]
- Ojeda F, Suykens JAK, Moor BD. Low rank updated LS-SVM classifiers for fast variable selection. *Neural Networks.* 2008; 21:437–49. [PubMed: 18343309]
- Ortner R, Allison BZ, Korisek G, Gaggl H, Pfurtscheller G. An SSVEP BCI to control a hand orthosis for persons with tetraplegia. *IEEE Trans Rehabil Eng.* 2011; 19(1):1–5.
- Perrin F, Pernier J, Bertrand O, Echallier JF. Spherical splines for scalp potential and current density mapping. *Electroencephalogr Clin Neurophysiol.* 1989; 72(2):184–87. [PubMed: 2464490]
- Pfurtscheller, G.; Scherer, R. Brain-computer interfaces used for virtual reality control. *Proc. 1st Int. Conf. on Appl. Bionics Biomech.; Venice, Italy.* October 14–16; 2010.
- Pfurtscheller G, da Silva FHL. Event-related EEG/MEG synchronization and desynchronization: Basic principles. *Clin Neurophysiol.* 1999; 110:1842–57. [PubMed: 10576479]
- Ramoser H, Muller-Gerking J, Pfurtscheller G. Optimal spatial filtering of single trial EEG during imagined hand movement. *IEEE Trans Rehabil Eng.* 2000; 8(4):441–46. [PubMed: 11204034]
- Sannelli, C.; Vidaurre, C.; Muller, K-R.; Blankertz, B. Common spatial pattern patches—an optimized filter ensemble for adaptive brain-computer interfaces. *Proc. 32nd Annu. Int. Conf. Engineering in Medicine and Biology Soc.; Buenos Aires, Argentina.* 2010; 2010. p. 4351-54.
- Sannelli C, Vidaurre C, Müller KR, Blankertz B. Common spatial pattern patches –an optimized filter ensemble for adaptive brain-computer interfaces. *J Neural Eng.* 2011; 8(2):025012. [PubMed: 21436539]
- Schalk G, Leuthardt EC. Brain-computer interfaces using electrocorticographic signals. *IEEE Trans Biomed Eng.* 2011; 4:140–54.
- Selby, SM. *Standard mathematical tables.* CRC Press; 1974.
- Song, L.; Epps, J. Improving separability of EEG signals during motor imagery with an efficient circular Laplacian. *Proc. 31st IEEE Int. Conf. Acoustics, Speech and Signal Processing; Toulouse, France.* 2006; 2006. p. 14-19.
- Srinivas M, Patnaik L. Adaptive probabilities of crossover and mutation in genetic algorithms. *IEEE Trans Syst Man Cyber.* 1994; 24(4):656–67.
- Stone M. Cross-validatory choice and assessment of statistical predictions. *J R Stat Soc.* 1974; B 36:111–47.
- Taylor DM, Tillery SIH, Schwartz AB. Direct cortical control of 3D neuroprosthetic devices. *Science.* 2002; 296(5574):1829–32. [PubMed: 12052948]
- Tianyou Y, Yuanqing L, Jinyi L, Zhenghui G. Surfing the internet with a BCI mouse. *J Neural Eng.* 2012; 9:036012. [PubMed: 22626911]
- Wolpaw JR, Birbaumer N, McFarland DJ, Pfurtscheller G, Vaughan TM. Brain-computer interfaces for communication and control. *Clin Neurophysiol.* 2002; 113:767–91. [PubMed: 12048038]
- Wolpaw JR, McFarland DJ. Control of a two-dimensional movement signal by a noninvasive brain–computer interface in humans. *Proc Natl Acad Sci.* 2004; 101:17849–54. [PubMed: 15585584]
- Wolpaw, JR.; Wolpaw, EW. *Brain-computer interfaces: principles and practice.* Oxford Univ. Press; 2012.

- Wu, B.; Wang, Y.; Chen, W.; Zheng, X. Time-frequency optimized spatial patterns for movement-related EEG decoding. *Proceedings of the 5th IEEE/EMBS Int. Conf. on Neural Eng.*; 2011. p. 84-7.
- Wu W, Gao X, Hong B, Gao S. Classifying single-trial EEG during motor imagery by iterative spatio-spectral patterns learning (ISSPL). *IEEE Trans Biomed Eng.* 2008; 55(6):1733–43. [PubMed: 18714838]
- Wu Z, Yao D. Frequency detection with stability coefficient for steady-state visual evoked potential (SSVEP)-based BCIs. *J Neural Eng.* 2008; 5(1):36–43. [PubMed: 18310809]
- Yuan J, Liu X, Liu C. Leave-one-out manifold regularization. *Expert Syst Appl.* 2012; 39(5):5317–24.
- Zhang H, Guan C. A maximum mutual information approach for constructing a 1D continuous control signal at a self-paced brain–computer interface. *J Neural Eng.* 2010; 7:56009.

Appendix

Frequency band selection for spatial filtering

Here we provide a heuristic approach to select the frequency band as a general preprocessing step for all the spatial filters evaluated in the experiment. For this approach, we used the channels from motor areas of the hands and/or feet most involved in the mental tasks used in the different datasets. These 13 channels included C3, Cz, C4, FC3, FCz, FC4, C5, C1, C2, C6, CP3, CPZ, CP4. The frequency band that can obtain lowest LOO error with LRR using log-variance features from these channels were chosen. The pseudo code of this approach is listed in table 7. The selected frequency band for each subject is listed in table 8.

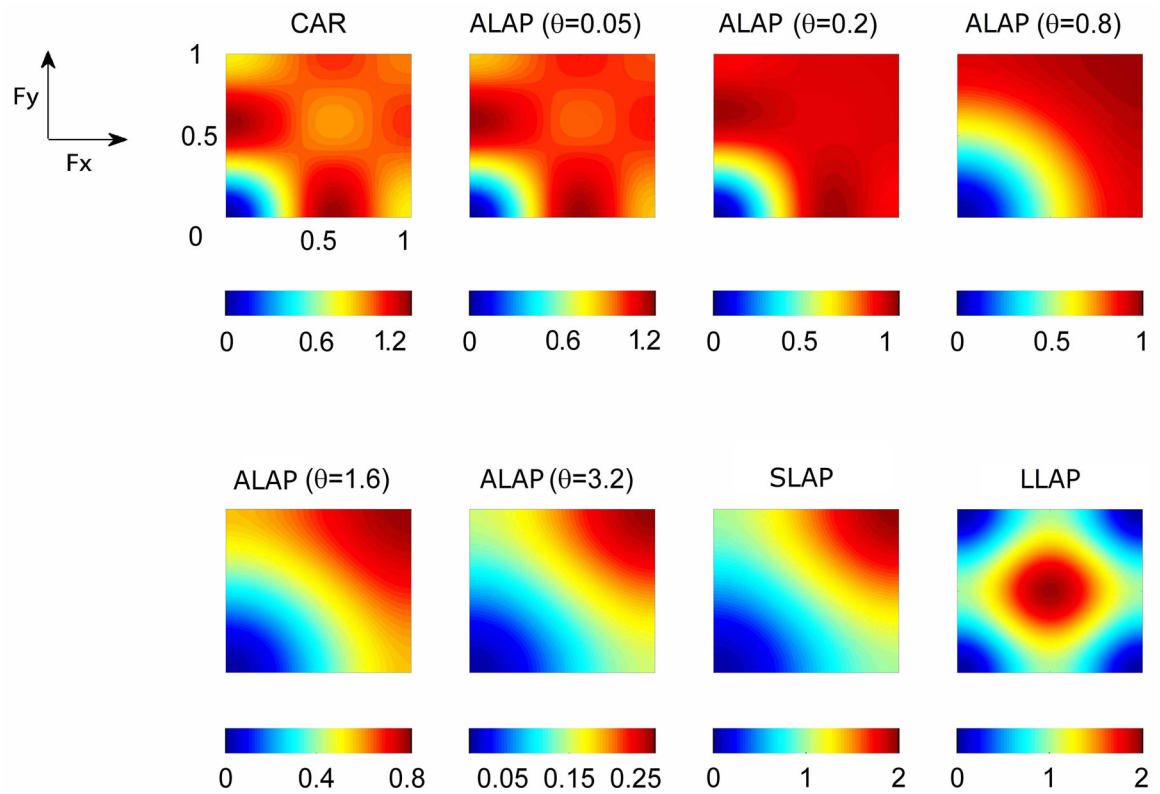


Figure 1.

The frequency response of CAR, SLAP, LLAP and ALAP filters with different kernel parameters in a two dimensional spatial domain. F_x (F_y) is the normalized frequency in the horizontal (vertical) direction. The color bar under each plot shows the response magnitude. See text for full description.

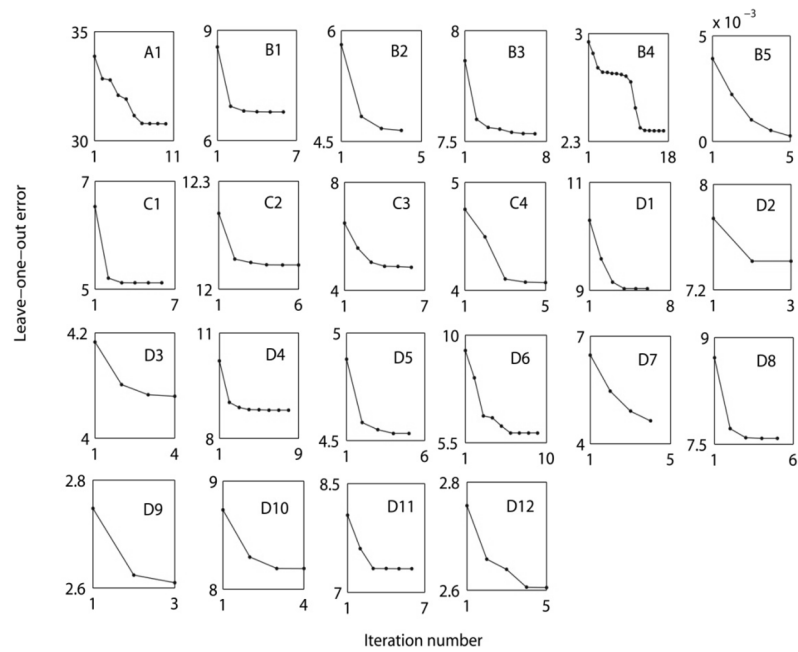


Figure 2.

Convergence of the gradient-based optimization procedure of the ALAP algorithm. The graph shows LOO errors computed (see section 3.2) over the iterations. Each curve corresponds to one subject.

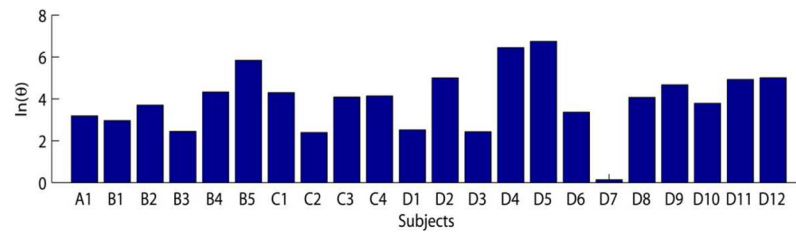


Figure 3. The optimized kernel parameter θ of ALAP for each subject. Here labels A, B, C and D denote the subjects from different datasets. This graph shows the variation in θ across subjects.

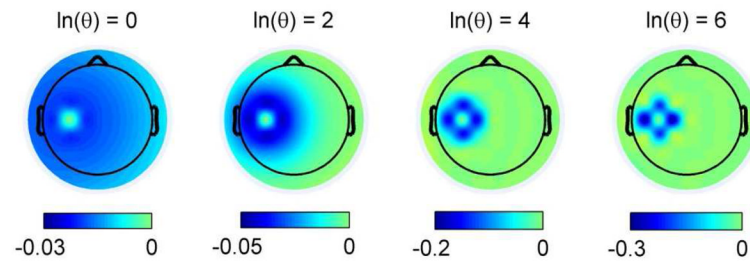


Figure 4. Scalp maps of $-h_{ij}$ in Eq.(20) for channel C3 (the green point in the blue areas) with different θ s. These scalp maps are based on the channel positions of cursor movement control dataset (i.e., subjects D1–D12).

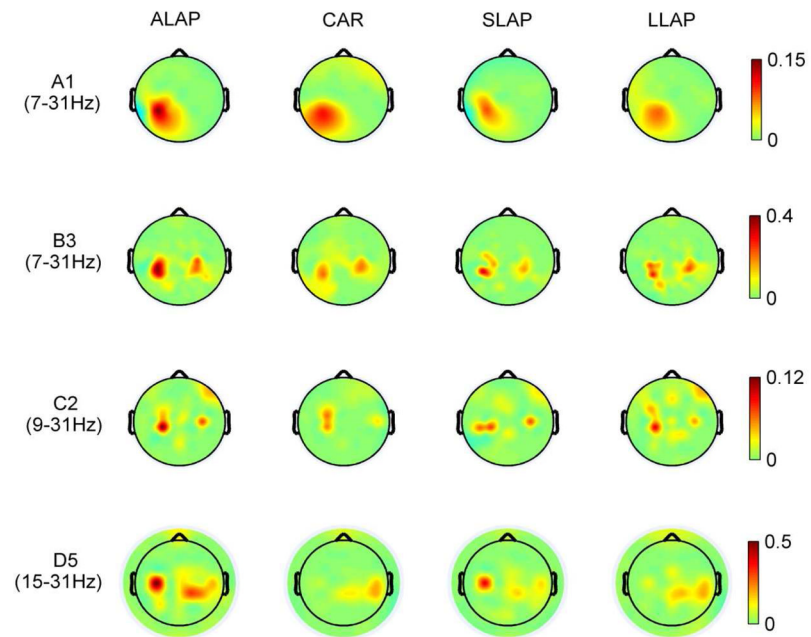


Figure 5. Scalp maps from representative subjects of the correlation r^2 values between training labels and the features extracted from each channel with different spatial filters.

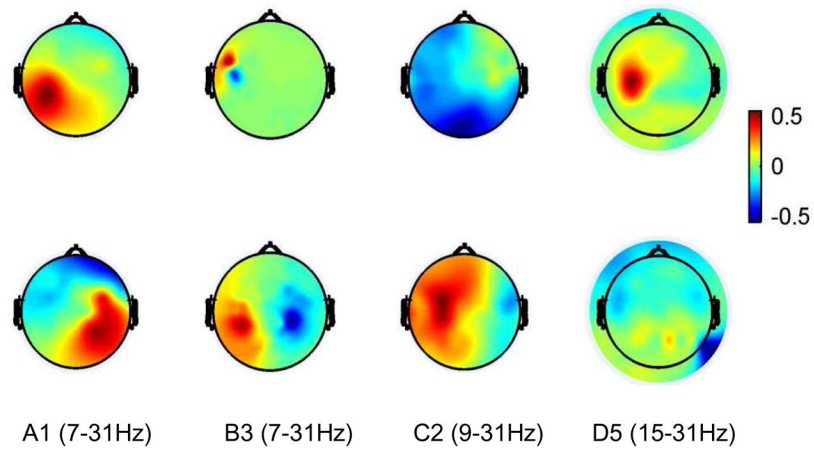


Figure 6.

The most discriminative spatial patterns obtained by CSP algorithm for subjects A1, B3, C2 and D5. Each column corresponds to one subject.

Table 1

ALAP filtering algorithm

1	Initialization: given multichannel EEG dataset $\{(\mathbf{X}_i, y_i)\}_{i=1}^N$ ($\mathbf{X}_i \in R^{C \times T}$, $y_i \in R$), two-dimensional channel position $\mathbf{v} \in R^{C \times 2}$, set initial parameters $\theta^{(0)}$ and $\lambda^{(0)}$, iteration number L , stop criterion δ .
2	for $t = 1: L$
3	Filter the EEG signal for each channel with respect to the dataset $\{\mathbf{X}_i\}_{i=1}^N$, kernel radius $\theta^{(t-1)}$ and channel position \mathbf{v} , by Eq. (3), (4);
4	Extract the power-features by Eq.(5),(6), then get the feature matrix \mathbf{F} ;
5	Calculate the LOO error $J(\theta^{(t-1)}, \lambda^{(t-1)})$ by Eq. (9), (10);
6	Calculate $\frac{\partial J}{\partial \theta}$ by Eq. (11)–(18), and compute $\frac{\partial J}{\partial \lambda}$;
7	Update $(\theta^{(t-1)}, \lambda^{(t-1)})$ to $(\theta^{(t)}, \lambda^{(t)})$ with $\left(\frac{\partial J}{\partial \theta}, \frac{\partial J}{\partial \lambda}\right)$;
8	Calculate $J(\theta^{(t)}, \lambda^{(t)})$ by Eq. (9), (10);
9	If $ J(\theta^{(t)}, \lambda^{(t)}) - J(\theta^{(t-1)}, \lambda^{(t-1)}) < \delta$, break, end ;
10	end ;
11	Calculate \mathbf{a} by Eq. (8) with the optimized θ and λ , then output them.

Table 2

The affiliations between the subjects and datasets.

Data sets	Subjects
Data Set IV, BCI Competition II	A1
Data Set IVa, BCI Competition III	B1–B5
Data Set I, BCI Competition IV	C1–C4
Data Set of Cursor Movement Control	D1–D12

Classification accuracies (%) obtained for each subject with each filter using all the available channels. The averages and standard deviations for all the subjects are also provided. The best results for each subject and the best average are in bold italics.

Table 3

Sub.	CAR	SLAP	LLAP	CSP	ALAP
A1	69.00	64.00	65.00	76.00	75.00
B1	84.82	83.93	86.61	74.11	87.50
B2	94.64	100.00	98.21	98.21	98.21
B3	63.27	66.84	70.92	65.31	72.96
B4	71.88	78.13	72.32	85.71	83.93
B5	80.56	79.37	66.67	75.00	82.54
C1	84.00	86.00	89.00	83.00	88.00
C2	50.00	50.00	48.00	67.00	69.00
C3	77.00	82.00	83.00	88.00	85.00
C4	84.00	90.00	85.00	94.00	91.00
D1	67.31	55.77	67.31	78.85	68.27
D2	74.04	80.77	79.81	76.92	81.73
D3	90.38	86.54	84.62	83.65	89.42
D4	76.92	83.65	87.50	75.96	88.46
D5	73.00	87.00	77.00	70.00	89.00
D6	78.00	81.00	86.00	72.00	82.00
D7	87.96	71.30	87.96	82.41	88.89
D8	79.63	84.26	79.63	72.22	79.63
D9	95.83	98.96	97.92	94.79	98.96
D10	80.83	85.83	87.50	85.00	85.83
D11	70.83	70.83	77.08	77.08	79.17
D12	92.39	98.91	96.74	98.91	97.83
Avg.	78.47	80.23	80.63	80.64	84.65
Std.	10.94	12.93	12.08	9.67	8.50

Table 4

Mean squared errors obtained for each subject with each filter using all the available channels. The averages and stand deviations for all the subjects together are also provided. The best results for each subject and the best average are in bold italics.

Sub.	CAR	SLAP	LLAP	CSP	ALAP
A1	0.2084	0.2080	0.2178	<i>0.1715</i>	0.1954
B1	0.1173	0.1361	0.1072	0.1743	<i>0.0961</i>
B2	0.0791	<i>0.0283</i>	0.0442	0.0421	0.0288
B3	0.2071	0.2502	<i>0.1860</i>	0.2113	0.1961
B4	0.1913	0.1995	0.2077	0.1362	<i>0.1217</i>
B5	0.2082	0.2670	0.3714	<i>0.1675</i>	0.2047
C1	0.1287	0.1264	0.1235	0.1391	<i>0.1189</i>
C2	0.2500	0.2500	0.2500	0.2397	<i>0.1965</i>
C3	<i>0.1488</i>	0.1887	0.1408	0.1532	0.1892
C4	0.1258	0.0928	0.1157	<i>0.0785</i>	0.0823
D1	0.2124	0.2823	0.2749	<i>0.1795</i>	0.2233
D2	0.1896	0.1890	<i>0.1566</i>	0.1940	0.1712
D3	<i>0.1125</i>	0.1640	0.1540	0.1382	0.1218
D4	0.1471	0.1232	<i>0.1145</i>	0.1558	0.1197
D5	0.2370	0.1206	0.2083	0.2121	<i>0.1055</i>
D6	0.1473	0.1469	0.1327	0.1960	<i>0.1294</i>
D7	0.1235	0.2205	<i>0.1054</i>	0.1398	0.1162
D8	0.1791	<i>0.1426</i>	0.1638	0.1848	0.1632
D9	0.0647	<i>0.0452</i>	0.0495	0.0644	0.0453
D10	0.1433	0.1268	0.1249	0.1474	<i>0.1228</i>
D11	0.2032	0.1993	<i>0.1567</i>	0.1859	0.1643
D12	0.1165	<i>0.0472</i>	0.0629	0.0546	0.0481
Avg.	0.1609	0.1616	0.1577	0.1530	<i>0.1346</i>
Std.	0.0505	0.0715	0.0766	0.0526	0.0548

Table 5

Classification accuracies (%) obtained for each subject with each filters using 17 channels from motor areas. The averages and stand deviations for all the subjects together are also provided. The best results for each subject and the best average are in bold italics.

Sub.	CAR	SLAP	CSP	ALAP
A1	69.00	68.00	<i>75.00</i>	72.00
B1	<i>87.50</i>	83.04	86.61	<i>87.50</i>
B2	96.43	98.21	<i>100.00</i>	98.21
B3	68.88	66.33	<i>69.90</i>	68.88
B4	83.48	<i>86.16</i>	85.71	<i>86.16</i>
B5	85.71	91.27	<i>91.67</i>	90.48
C1	80.00	86.00	86.00	<i>88.00</i>
C2	59.00	65.00	<i>76.00</i>	65.00
C3	80.00	79.00	<i>88.00</i>	80.00
C4	91.00	90.00	91.00	<i>92.00</i>
D1	<i>78.85</i>	57.69	77.88	77.88
D2	61.54	<i>66.35</i>	63.46	63.46
D3	91.35	85.58	<i>93.27</i>	91.35
D4	83.65	79.81	79.81	<i>85.58</i>
D5	81.00	88.00	89.00	<i>91.00</i>
D6	83.00	73.00	<i>86.00</i>	83.00
D7	<i>93.52</i>	88.89	87.04	<i>93.52</i>
D8	77.78	83.33	77.78	<i>85.19</i>
D9	97.92	98.96	<i>100.00</i>	97.92
D10	67.50	81.67	<i>84.17</i>	83.33
D11	78.13	76.04	<i>83.33</i>	78.13
D12	95.65	<i>100.00</i>	98.91	98.91
Avg.	81.40	81.47	<i>85.02</i>	84.43
Std.	11.00	11.64	9.34	10.22

Table 6

Mean squared errors obtained for each subject with each filter using 17 channels from motor area. The averages and stand deviations for all the subjects together are also provided. The best results for each subject and the best average are in bold italics.

Sub.	CAR	SLAP	CSP	ALAP
A1	0.2084	0.2139	0.1903	<i>0.1869</i>
B1	0.1100	0.1257	0.1210	<i>0.1094</i>
B2	0.0710	0.0494	<i>0.0358</i>	0.0488
B3	0.2006	0.2102	<i>0.1928</i>	0.2006
B4	0.1387	0.1138	0.1287	<i>0.1111</i>
B5	0.1470	0.0956	<i>0.1020</i>	0.1141
C1	0.1434	0.1215	<i>0.1019</i>	0.1093
C2	0.2198	0.2292	<i>0.1917</i>	0.2183
C3	0.1635	<i>0.1467</i>	0.1579	0.1629
C4	0.1021	0.0947	<i>0.0943</i>	0.1013
D1	0.1658	0.2807	0.1709	<i>0.1657</i>
D2	0.2430	<i>0.2233</i>	0.2351	0.2303
D3	<i>0.0927</i>	0.1679	0.0960	0.0928
D4	0.1138	0.1200	0.1527	<i>0.0984</i>
D5	0.1933	0.1348	<i>0.1158</i>	0.1259
D6	0.1346	0.1693	<i>0.1174</i>	0.1357
D7	<i>0.0813</i>	0.0934	0.0847	0.0814
D8	0.1731	<i>0.1588</i>	0.1863	0.1644
D9	0.0486	0.0470	0.0627	<i>0.0435</i>
D10	0.2047	0.1505	<i>0.1476</i>	0.1510
D11	0.1794	0.1990	<i>0.1764</i>	0.1802
D12	0.0714	0.0344	<i>0.0250</i>	0.0377
Avg.	0.1457	0.1445	0.1312	<i>0.1304</i>
Std.	0.0544	0.0640	0.0543	0.0545

Table 7**Frequency band selection algorithm**

-
- 1 **Input** multichannel EEG dataset $\{(\mathbf{X}_i, y_i)\}_{i=1}^N$ ($\mathbf{X}_i \in R^{C \times T}$, $y_i \in R$), where C denotes the number of channel set, T denotes the number of sampled time points in a trial.
 - 2 Filter EEG data \mathbf{X}_i into narrow bands $\mathbf{X}_i^{f_k} \in R^{C \times T}$ channel by channel, where $f_k \in [7 + 2(k-1), 7 + 2k]$ Hz, $k = 1, 2, \dots, 12$ and extract the log-variance of $\mathbf{X}_i^{f_k}$ as features $\mathbf{Z}_i^{f_k} \in R^C$.
 - 3 Train linear ridge regression model with feature set $\{(\mathbf{Z}_i^{f_k}, y_i)\}_{i=1}^N$ and seek the optimal regularization coefficient with lowest LOO error. Then the negative of this lowest LOO error is used as the score of frequency band f_k , denoted by S_{f_k} .
 - 4 $k^* \leftarrow \arg \max_k S_{f_k}$, record the initial best score and frequency band $S^* \leftarrow S_{f_{k^*}}$, $f^* \leftarrow [7 + 2(k^* - 1), 7 + 2k^*]$ Hz, and get the initial bounds $k_0 \leftarrow k^* - 1$, $k_1 \leftarrow k^*$.
 - 5 **for** $j = 1:12$
 - 6 Extend the frequency band:
 - a. If $k_0 = 1$ and $k_1 = 12$, then break.
 - b. If $k_0 > 1$ and $k_1 = 12$, then $k_0 = k_0 - 1$; If $k_0 = 1$ and $k_1 < 12$, then $k_1 = k_1 + 1$; Similar to step (2) and step (3), but compute the score of a wider frequency band $[7 + 2k_0, 7 + 2k_1]$ Hz as S_j .
 - c. If $1 < k_0$, $k_1 < 12$, there are two candidate bands: $[7 + 2(k_0 - 1), 7 + 2k_1]$ Hz and $[7 + 2k_0, 7 + 2(k_1 + 1)]$ Hz. Compute the scores of them respectively as S_j^L and S_j^R . If $S_j^L > S_j^R$, then $S_j \leftarrow S_j^L$, $k_0 = k_0 - 1$; else, $S_j \leftarrow S_j^R$, $k_1 = k_1 + 1$.
 - 7 If $S_j > S^*$, then $S^* \leftarrow S_j$ and $f^* \leftarrow [7 + 2k_0, 7 + 2k_1]$ Hz.
 - end**
 - 8 **Output** the frequency band f^* .
-

Table 8

Selected frequency band for each subject.

Sub.	Selected frequency band
A1	7–31 Hz
B1	11–17 Hz
B2	7–29 Hz
B3	7–31 Hz
B4	9–15 Hz
B5	7–31 Hz
C1	7–31 Hz
C2	9–31 Hz
C3	7–25 Hz
C4	9–29 Hz
D1	9–13 Hz
D2	7–23 Hz
D3	19–27 Hz
D4	9–15 Hz
D5	15–31 Hz
D6	7–23 Hz
D7	7–27 Hz
D8	7–29 Hz
D9	7–21 Hz
D10	7–17 Hz
D11	11–15 Hz
D12	19–31 Hz