[[1]](#footnote-1)

Optimizing Spatial Filters for Robust EEG Single-Trial Analysis

Alessandro Napoli

*Abstract*—Spatial filters are more and more used in EEG single-trial analysis. Their main purpose is to reduce the signal-to-noise ratio of the recorded channels in order to improve the performance of BCI systems. During the last years several techniques have been proposed to implement such filters in order to optimize their coefficients to the characteristics of different subjects. Many of the proposed techniques make use of filter coefficients that are based on head geometrical models. These coefficients are usually manually selected by the operators of the BCI systems according to their own experience and skills in analyzing preliminary data from the subjects. Moreover filter coefficients selection may be a time consuming operation that must be performed each time a subject performs a BCI experiment. In this view an automated data-driven parameters selection, based on statistical analysis of the training data sets, seems valuable to automatically select the best subject’s parameters. In this paper we review the common spatial pattern (CSP) technique and its mathematical background that has been proven to be a powerful offline analysis tool to implement optimal spatial filters that can be used to improve BCI systems performance.

# INTRODUCTION

In recent years many studies have shown that by recording brain signals and applying specific experimental paradigms, it is possible to transform the brain signals into control signals. These extracted command signals have given some users the opportunity to control different applications using signals coming from their brain. [1],[2],[3]

Such a system is called brain-computer interface (BCI) and it has been proven to be a great instrument to help paralyzed people or those who suffer from motor disabilities to better interact with the world.

Traditionally BCI research has moved towards the creation of very sophisticated systems that run on specific algorithms and make use of advanced technologies. These systems can be adapted to different user needs and characteristics and they represent a valid tool to improve the quality of life of people with compromised motor abilities. [5],[6]

Unfortunately BCI systems proved to be very challenging from different perspectives: the highly variable and uncontrollable nature of brain signals, which represent the input to this kind of systems, makes the signal processing necessary to extract some useful correlation between commands and user intents very complex. At the same time, the high subject-to-subject variability and operating conditions do not allow researchers to focus on specific experimental paradigms or processing techniques. [4],[6]

In order to address these issues and to improve the signal-to-noise ratio in BCI experiments, different processing methods have been proposed, some of which have been derived from machine learning techniques and statistical analysis.[8] In this view particular care has been dedicated to the design of optimal spatial filters, which instead of being based on fixed data, related to the geometry of the acquisition system, are based on statistical properties derived from each subject signals.[9],[12] More and more popular in BCI applications is becoming the use of methods called Common Spatial Patterns (CSP) [11] to build powerful spatial filters that are successful in detecting variations in localized oscillatory neural activity, also called Event-Related Desynchronization (ERD) or Event-Related Synchronization (ERS).

In the following we will briefly describe the general BCI processing techniques and then we will focus on CSP methods and how their use can improve BCI performance. [10]

# Processing Techniques Overview

In this section we will focus on some basic aspects of noninvasive BCI systems and some of the most important aspects of BCI signal processing.

## BCI Components

Brain Computer Interfaces (BCI) are special systems whose aim is to control a device using some subject's brain signals that are related to his intentions. In this view BCIs main goal is to be able to relate EEG signals, which are the result of neural firing in the brain, to human intentions using specific techniques of feature extraction and dedicated algorithms.

A classic example of a BCI system design is shown in Figure 1, where the different modules of a general system are shown.



Figure 1: General BCI design.

EEG signals are very complex, since they are generated as superposition of the simultaneous activity of different systems spatially distributed in a conductive volume. Moreover this conductive volume presents a high degree of inhomogeneity, because of the different tissues that are involved, such as the brain, the scull, softer tissues and the skin.

The main challenges in dealing with this kind of signals stem from the inaccessibility of the EEG signal sources: the brain and its structures. This implies that the signals we acquire and deal with are the result of the overall brain activity and they are neither exclusively the expression of activation of a specific area nor related to the contribution of a specific cerebral activity.

BCI processing chain is composed of different components connected together with the goal of converting raw brain data into control command signals as shown in Figure 2, where some of the common modules are presented.



Figure 2: BCI Signal Processing Components

## Feature Extraction

The first step of the processing chain shown in Figure 2 is the feature extraction and it involves both spatial filtering and spectral analysis modules (Figure 3).



**Spectral Analysis**

**Spatial Filter**

Figure 3: Feature Extraction Section

The spatial filter reduces the effects of spatial blurring, which are typical of brain signals. This spatial blurring that affects brain signals is a consequence of the brain and head anatomy, of the distance between the sensors (electrodes) and the signal sources (neurons) in the brain. We need to emphasize the importance of this problem with EEG signals, because of the high inhomogeneity present in brain tissues.

Over the years, different approaches have been presented to cope with the above issues. In particular, the spatial filter improves the signal-to-noise ratio of the EEG channels by filtering out those signal components not related to a specific task or subject’s motor intent. These filters are usually implemented by subtracting a weighted sum of a subset of channels from the electrodes of interest, as shown in (1).

$$s\_{i},\_{t}=\sum\_{j}^{}w\_{i},\_{j}\*u\_{j},\_{t}$$

**Equation 1: Weighted sum of the raw signals**

Where $u\_{j},\_{t}$ is the input matrix of the system (raw EEG signals) with j indicating the channels and t the time samples; $s\_{i},\_{t}$ are the filtered EEG signals using the filter coefficients $w\_{i},\_{j}$ .

The type of the implemented spatial filter determines the values of the matrix coefficients $w\_{i},\_{j}$ and traditionally these weights are fixed, such as in Laplacian Spatial Filters and in Common Average Reference Filters (CAR). [14]

Laplacian filters discretize approximations of the second spatial derivative of a two dimensional Gaussian distribution on the scalp surface and attempt to invert the processes that blurred brain signals detected on the scalp. In practice these filters are often implemented simply by subtracting the average of the four next nearest neighbor electrodes from the center location.

CAR filters are instead implemented by re-referencing the voltage read from every electrode at each time point to an estimated reference that is calculated by averaging the signals from all recorded electrodes. In practice these filters compute the signal amplitude that is common to all electrodes and subtract it from the signal at each location.

Spatial filters can also be implemented using different approaches such as data-driven methods. For example, spatial filters based on methods such as principal components or independent components have data driven weights. It is important to notice that both these methods are unsupervised. In contrast, the method of common spatial patterns is both supervised and data-driven. In [13] several independent components algorithms, Laplacian and bipolar derivations, and common spatial patterns on data derived from a four-class motor imagery task have been compared. They found that the Laplacian and independent components methods were comparable, but the method of common spatial patterns resulted in the best classification.

The filter coefficients selection is a very important step in BCI processing, in fact the effectiveness of the filtered signals and their signal-to-noise ratio depend on this selection. The high variability of the acquired signals, both from subject-to-subject and from acquisition-to-acquisition, and the diverse experimental conditions strongly influence the characteristics of the signals. In this view it seems valuable to use filters whose coefficients are data-driven and may be adjusted to any experimental condition and subject’s characteristics. In fact data-driven methods show better performance in improving the signal-to-noise ratio when compared to fixed approaches that struggle to keep up with the high variability present in the system.

The second step of the feature extraction is the spectral analysis, whose function is to project the input signals into a new domain, in which the brain signal features modulated by the user, are best expressed. This allows to separate some physiological artifacts, present in the original signals, from user intent related features. Traditional techniques suggest to transform time windowed version of selected spatial filtered EEG channels (their selection depends on the characteristic of the subject) from the time domain into the frequency domain, as shown in (2).

$$f\_{t}= \left\{F\left(TW\left(i\_{1},t,k\right),w\_{1}\right),F\left(TW\left(i\_{2},t,k\right),w\_{2}\right),…\right\}^{T}$$

**Equation 2: Computation of the feature vector** $f\_{t}$

In (2) we compute the Fourier Transform, in the band w, of the time windowed version of the signals TW(i,t,k), obtained by selecting the last k samples of the i-th channel at time t.

Feature extraction with frequency-domain signals has involved a wide variety of techniques.[15] These include methods that are time-based, space-based, and time-space methods. Time-based methods include band-pass filtering, Fourier-based spectral analysis, parametric methods such as autoregressive spectral methods, and use of wavelets. Space-based methods include Laplacian filters, principal components, independent components, and common spatial patterns. Time-space-based methods include component analysis in time and space and multivariate autoregressive models.

These different techniques have been then compared in several studies. For example, in [16] the authors report no difference between band-power analysis techniques, based on digital filtering and adaptive autoregressive parameters obtained by means of Kalman filtering. In [17] the authors compared band-power analysis, carried out using Hjorth parameters, and Fractal Dimensions as possible features for classifying motor imagery data. Band-power methods yielded the best performance for four of five subjects, but the authors concluded that fractal dimension could be considered as an alternative to band-power. In [18] spectral bands based on AR (autoregressive) models, the FFT, and a matched filter are compared. In this case the matched filter outperformed other methods.

Although BCI signal extraction is usually described as involving two distinct phases, spatial and temporal filtering, it is also possible to include both in a single process. For example, in [19] the authors used common spatial patterns with time-delay embedding. This method resulted in a single-step method producing a better classification than a method implemented by combining band-pass filtering and common spatial patterns.

Also in this step of the processing several settings, such as frequency bands, time window duration and channels to analyze, are selected manually and chosen to maximize the performance of the system for a specific subject, according to the implemented algorithms. For example, when working with imagined movement detection, the EEG features we try to extract are mu and beta rhythms, which show an oscillatory behavior at certain frequency bands (8-12Hz) and then these are usually extracted in the frequency domain. [2] At the same time in order to focus on the correct oscillatory EEG components, generated by imagined motor movements we need only to focus on the activity recorded by those electrodes located in proximity of the motor cortex area. The last parameter in this analysis is the time duration of the time windows we select. In doing this we need to keep in mind that shorter time windows allow to obtain feature values more quickly and so we can update the application control more often. On the other hand, longer windows allow to carry out a more accurate analysis, but they assure slower responses.

## Translation Algorithm

The second processing macro step is a translation algorithm, usually accomplished using conventional classification or regression procedures. The translation algorithm aims at translating the extracted features from the previous step into device commands.



**Normalizer**

**Classifier**

Figure 4: Translation Algorithm

Usually the classifier may be implemented creating a linear combination of the extracted features. The vector $f\_{t}$ contains those amplitudes of different frequency bands and at different scalp locations [20] that are linearly combined.

$$x\_{t}=\sum\_{i}^{}φ\_{i}\*f\_{i},\_{t}$$

**Equation 3: Linear combination of the features vector using the coefficients** $φ\_{i}$

In (3) the coefficients $φ\_{i}$ are traditionally chosen manually, by offline inspection of the training data recorded from the subject.

Different solutions have been studied to implement this step of the processing and they vary from easier linear analysis methods (LDA, linear regression) [15] to more complex neural networks and support vector machines. [20]

Currently linear methods still represent the most used options for the classifier design.

The normalizer is another important step in the processing; in fact it is used to compensate for spontaneous changes in brain signal statistics (nonstationarity).

$$C\_{t}=\frac{x\_{t}- μ }{σ}$$

**Equation 4: Normalization equation**

Where $μ$ is the predicted mean of the signals, estimated using recent trials and $σ$ is the standard deviation.

Moreover the translation algorithm may also include a whitening procedure (linear transformation) that produces signals with zero mean and a defined variance. In this way the output device does not have to account for changes in brain signal characteristics that are not related to the specific task.

# Offline Analysis

After briefly describing the general online signal processing for BCI applications, in this section we present the offline analysis, which is necessarily performed every time prior to the actual BCI online experiment. In fact it is of critical importance for the correct online operation of a BCI system as its performance is based on the effectiveness of the offline analysis. As shown in Figure 5, the real-time operation of the system depends upon the parameters and signal features obtained from the offline analysis, which is performed on specific data sets, called training data, recorded while the subject is performing predefined tasks specified by the implemented experimental paradigm. These data sets then allow us to obtain, for any subject, signal related features that can be correlated to his specific intent or performed task.

 In other words acquiring significant and accurate brain signals during the training data is fundamental to the subsequent BCI operation. Really important in this context is also the experimental paradigm adopted for the training data acquisition.

 Considering what said above, the signals of the training data would influence the whole BCI processing and performance. In fact, as shown in Figure 5 all the parameters and even the EEG features that we subsequently use in the processing are derived from the training data set.

 This data is often analyzed manually by an operator who selects features and parameters applying some statistical method. In our experience this process is fundamental for the success of a BCI application; in fact the user might not be able to use the application or he could perform poorly if the processing parameters are not selected accurately.



**Figure 5: BCI System where features and parameters used during the online experiment are derived using statistical analysis of training data recorded offline.**

# Statistical Approaches

The processing method presented in the previous sections allows researchers to implement fast and efficient algorithms suitable for real-time applications, with narrow time constraints.

Nonetheless such an approach to BCI applications strongly relies on offline parameters selection form the training data sets.

Often parameters and features, used during online experiments, are derived from offline analysis manually. This implies that the system operator has to spend a consistent amount of time looking at the recorded training data for a specific subject, in order to select the best features and parameters that can be used in the subsequent online experiment.

This is not a negligible aspect because such an operation must be performed before any experiment and for each subject involved in the experiments.

In our view, carrying out manually the offline analysis ends up being both time consuming and operator dependent. The effectiveness of the selected parameters depends on the ability of the operator of the BCI system to analyze the training data set. Moreover such an approach is time consuming, especially if we are willing to increase the number of users who can benefit from the use of BCI systems.

Considering what said above, it is easy to see the importance of features and parameters selection. We think that this key aspect for the successful operation of a BCI cannot just rely on the skills of the system operator. This is why we propose the introduction of statistical approaches, applied to the training data sets, to perform an automated selection of the subject related parameters and features. This automated analysis would help the researchers get the best set of parameters for any user and under any experimental condition. During the last years many studies have shown that different analysis techniques can be implemented for this purpose, here we focus on Common Spatial Patterns analysis.

## CSP Analysis

A very interesting approach that allows us to select optimal weights to be used in our spatial filter is a statistical analysis technique called Common Spatial Patterns (CSP).

This technique is really useful to analyze multichannel data sets that are the result of experimental paradigms where the subject can perform one out of two possible tasks (or movements). These kinds of experiments are generally called two class experiments: we have two possible states (classes) and we need to identify the state of the system by observing some output signals somehow correlated to the two conditions.

CSP approach allows to statistically evaluate the elements of a *n x n* spatial filter matrix (where n is the number of acquired channels). Using this matrix to filter the raw brain signals we can obtain a data-driven decomposition of the original data space into a new signal space described by (5):

$$x\_{csp}(t)= W^{T}x\left(t\right)$$

Equation 5

Where each row of the vector $x\_{csp}$ represents an original channel filtered using the matrix **W**. The elements of the matrix **W** are computed using an optimization criterion that, in CSP techniques, tends to maximize the variance of the spatially filtered signals for one of the two possible experimental conditions (or classes) while at the same time minimizing the variance of the signals for the remaining condition. Such a method allows to obtain filtered signals, where those components uncorrelated to specific spatial patterns have been filtered out.

Each column of the matrix **W** is a vector $w\_{j}$called spatial filter and each column of the matrix $A=\left(W^{-1}\right)^{T}$ is the vector $a\_{j}$called spatial pattern.

Now let us take a deeper look at how this method can be applied to BCI systems to give a powerful tool to discriminate among different mental states characterized by different ERD or ERS phenomena.

## Technical Approaches to CSP Analysis

CSP methods are based on the estimation of as many covariance matrices as the number of conditions that we want to determine. Once these matrices have been computed, this approach proposes to maximize a matrix at a time while minimizing the other ones with respect to a single class of the experiment.

For example focusing on the case of a two class experiment, say a right and left imagined hand movement we can write:

$$COV\_{c}=\frac{1}{\left|I\_{c}\right|}\sum\_{i}^{}X\_{i}X\_{i}^{T}$$

Equation 6

Where the subscript c indicates one of the two possible experimental classes. $I\_{c}$ is the set of indices corresponding to trials belonging to each condition (in this example: right and left). Equation 6 represents a pooled estimate of the covariance in each condition, which means that it computes an estimate of the covariance obtained using several different samples recorded in different conditions. This is always true if we assume that each **X** is scaled and centered.

 Now using this definition we can start CSP analysis, by simultaneously diagonalizing the covariance matrices for the two movements as follows:

$$\left\{\begin{array}{c}W^{T}COV\_{left}W= D\_{left}\\ W^{T}COV\_{right}W= D\_{right}\end{array}\right.$$

Equation 7

 Where **D** represents a diagonal matrix. Equation 7 allows to compute those filter weight matrices **W** such that

$$D\_{left}+D\_{right}=I$$

Equation 8

 Mathematically (7) can be solved by solving the generalized eigenvalue problem given by

$$COV\_{right} w= λ COV\_{left} w$$

Equation 9

 In order to solve (7) we can use matrices **W** composed of the vectors $w\_{j}$as column vectors and $λ\_{j}\_{c}=w\_{j}^{T}COV\_{c}w\_{j}$ are the diagonal elements of $D\_{c}$with $λ$being equal to $ \frac{λ\_{right}}{λ\_{left}}$ .

 What stated above is very important and allows us to state that when the value of a $λ\_{j}\_{c}$ is really close to one then the corresponding spatial filters $w\_{j}$present a high variance towards the condition indicated by the subscript c and a low variance towards the opposite condition. Thus these filters allow us to discriminate between these two classes.

 Another way to emphasize the above concept is by looking at the effects that the transformation in (5) has on the original data. In fact the original strong correlation between the two original axes is removed and now both distributions are simultaneously de-correlated with respect to two classes. Moreover, as we can see from the example shown in Figure 6 the two distributions are now maximally dissimilar along these new axes.



Figure 6: Example of CSP filtering in 2-D

 Figure 6 makes use of sample data to show how CSP filters operate on two sets of 2-D samples drawn from two Gaussian distributions. In part (a) is shown the original distribution, before filtering, while in (b) the filtering results are shown. From this graph we can see how the two filtered distributions are now uncorrelated. Moreover they are now distributed along two axis, a vertical axes which maximizes the variance for the class represented by circles and minimizes the variance for the other class represented by crosses and vice versa.

## CSP as Generative Approach

The method shown in the previous section can be also seen as an application of Generative Approach.

In order to underline effects of CSP spatial filterinf on the raw data, let us consider a linear mixing model generated by nonstationary sources, say $s\_{c}$, for simplicity we can consider them as composed of two classes c = (+,-) and Gaussian distributed, with zero mean and variance $D\_{c}$for the two classes, we can also assume that these distributions are uncorrelated.

If the estimates of their covariance matrices are close to the true matrices, when applying the simultaneous diagonalization, as described in the previous section, we obtain a maximum likelihood estimator of the backward model given by

$$W=\left(A^{-1}\right)^{T}$$

Equation 10

In this view we can show the discriminative power of this method by writing

$$\left\{\begin{array}{c}s\_{d}=COV\_{+}-COV\_{-}\\s\_{c}=COV\_{+}+COV\_{-}\end{array}\right.$$

Equation 11

Where $s\_{d}$ represents the discriminative activity and $s\_{c}$ represents the common activity. Now we need to find a way to maximize (11). A solution to this problem is given by the Rayleigh Coefficient that can be obtained by solving the same generalized eigenvalue problem that we have discussed previously, that is we need to maximize the following ratio with respect to $w\_{j}$

$$\frac{w^{T}s\_{d}w}{w^{T}s\_{c}w}$$

Equation 12

Again we can assume that $λ\_{+}+ λ\_{-}=1$ as we did earlier, so we can see that large values of (12) give us large responses in the positive condition and the opposite is true for the other condition.

 In practice when we deal with a classification setting, especially during an online experiment, we need carefully to choose the number of components (eigenvectors) that we want to use. In fact if the number of components is too small the classifier performance would be poor because such few components would not guarantee an effective discrimination between different classes. On the other hand selecting a large number of components would be worthless too. In fact it would only increase the computational and power resources necessary to apply the CSP method, without giving any increment in the discriminative power of the system.

 In practice the number of eigenvectors to use is usually determined by looking at the performance of the actual system. In fact if the system performance does not benefit from an increase in the number of eigenvectors, we stop taking new ones into account. Alternatively the number of used coefficients can be selected using other criteria.

The previous examples and original applications of CSP analysis were limited to binary classification problems. This does not mean that the proposed technique cannot be applied to multiclass experiments, in fact CSP can still be carry out on binary sub-problems, derived from the general data set. For example we can divide multiple decisions in pairs or we can look for discrimination between any class and the rest state.

An alternative approach to the application of CSP to multitask problems is to carry out an approximate simultaneous diagonalization of the full filter matrix with respect to the possible movements. [25]

# Results

CSP approaches have been used for years to address the problem of spatial filtering EEG signals and to get the best discriminative features to control BCI applications.

In particular CSP approach seems to guarantee better BCI performance, especially when compared to other traditional techniques that require longer training period for the users before they can control the system. CSP tries to transfer the whole learning burden from the subject to the system, by selecting a system setup that best adapts to the user needs and characteristics.

In order to evaluate the actual performance of CSP analysis in [11] they have proposed to test its accuracy and information transfer rate during online experiments. In Table 1 and Table 2 are shown the results the authors got by testing CSP algorithm on healthy subjects.



Table 1



Table 2

In the last column of Table 1 is shown the Information transfer Rate (ITR) defined according to (7)

$$ITR=\frac{\# decisions}{duration in minutes} \*( p log\_{2}\left(p\right)+$$

$$+ \left(1-p\right)log\_{2}\left(\frac{1-p}{N-1}\right)+log\_{2}\left(N\right) )$$

Equation 13

 Where *p* is the accuracy of the system to understand the user decisions between N targets. The equation in (13) describes the capacity of a symmetric communication channel that makes mistakes with equal probability $\frac{1-p}{N-1}$ with respect to all the other N-1 possible classes divided by the required time to communicate the total amount of information.

 ITR depends upon the accuracy of the classifier but also on the chosen design for the feedback application which converts the classifier decisions into command signals.

## CSP Performance

Another interesting way to appreciate the performance of CSP techniques in BCI is to take a look at the effects that this technique has on brain signals when used to spatial filter EEG signals. In Figure 7 are shown the results of a spatial filtering implemented using four CSP filters applied to continuous band-pass filtered EEG signals. [11] The first two filters maximize those signal components that correspond to an imagined left hand movement while minimizing those components corresponding to an imagined right hand movement. The opposite scenario can be observed for the last two filters in Figure 7.



Figure 7: Effects of four CSP filters on four EEG channels

Another important aspect of CSP approaches is that while selecting the filter coefficients we can visually represent them using some topology plots. In other words, we can recreate head topologies where instead of plotting the original EEG channel activations we can plot the values of the filter coefficients derived using CSP. This allows us to test the effectiveness of those coefficients while performing the analysis and before moving to the online experiments. In fact those coefficients that are not physiologically plausible can easily be disregarded, this gives us an opportunity to check the results of the statistical analysis even before applying the filter. In Figure 8 is shown an example of CSP and how we can plot filter coefficients and patterns to represent source projections on the scalp. In Figure 8 we can see how the filter coefficients can be used to project the original sources of activity. It is easy to note that the components with minimum variance during an imagined hand movement are contralateral to the hand the subject imagines to move, as expected.



Figure 8: Topographical scalp maps displaying filters and patterns derived from CSP analysis of raw EEG signals

 In order to emphasize the effectiveness of CSP methods in spatial filtering the EEG signals, we can compare the frequency response of the filtered signals to frequency responses obtained applying other well-known spatial filter techniques.

 In Figure 9 are shown the spectra of left versus right hand motor imagery. All the plots are obtained from the same data set and using the same frequency analysis, [11] such as the different responses we can see are entirely related to the different performances of the applied filters.



Figure 9: Spectral analysis of a CP4 EEG channel filtered using different spatial filters.

 From Figure 9 it is easy to note that the CAR and CSP represent the best filters, in fact they offer the best discrimination between the two different movements. In reality from the figure we can see that CSP offers a better discriminative effect than the CAR filter, in fact only when applying CSP we can see two distinct frequency peaks relative to the right movement.

## Merits and Limitations

CSP techniques have proven to be successful in online BCI applications. Their main advantage is probably related to the fact they are not applied as black-box method, in fact as we have already mentioned, CSP results can be visualized using topography maps and this allows us to check the efficacy of the analysis before even testing its results online.

It is important to underline that CSP is neither a source localization method nor a source separation. On the other hand, each filter is built to optimize two different effects: it maximizes the variance along one class while minimizing it along the other class.

An important limitation of this method is the number of parameters we need to select before performing a real analysis. Parameters such as frequency band-pass, time intervals and the number of CSP filters actually used may not always be easy to determine. Usually some general settings are adopted, for example frequency band 7-30 Hz; time intervals starting 1 second after the stimulus onset; two or three spatial filters for each side of the eigenvalue problem. In this view it is important to emphasize that when trying to improve online performances these parameters should be carefully selected on the base of subject’s characteristics.

Another important issue regarding CSP is that their discriminative capabilities only hold with respect to the separation of the mean power of two classes. Such separation may not be sufficient to obtain a real discrimination, especially for those samples that are close to the decision boundaries. In addition, the mean may be sensitive to outliers, which in this context may be represented by signal artifacts, such as blinking, eye and muscle movements. Anyway this effect has been shown to be less severe than previously thought. In fact the features corresponding to such CSP artifacts get a near-zero weight in the classification step and is thereby neglected.

# Conclusions

In this paper we have reviewed some of the general BCI processing techniques, with particular attention to CSP spatial filter technique. This method uses the covariance matrices of the raw signals to find the best discriminative filter coefficients for a two class experiment.

 This technique has proven to be successful in increasing the performance of two online BCI applications. In this paper we have presented also CSP advantages such as low computational costs and interpretability as well as some limitations such as model selection and sensitiveness to outliers.

In order to improve this method and make it more robust for future applications we need to make use of new spatial-temporal filtering methods that allow more accurate and interpretable classification even for nonstationary and noisy signals like those read on a subject’s scalp.

References

1. Jonathan R. Wolpaw and Dennis J. McFarland. “Control of a two-dimensional movement signal by a noninvasive brain–computer interface in humans”. Edited by Emilio Bizzi, Massachusetts Institute of Technology, Cambridge, MA, and approved November 2, 2004 (received for review May 17, 2004).
2. Christa Neuper, Reinhold Scherer, Selina Wriessnegger, Gert Pfurtscheller . “Motor imagery and action observation: Modulation of sensorimotor brain rhythms during mental control of a brain–computer interface”. Clinical neurophysiology, February 2009 (volume 120 issue 2 Pages 239-247 DOI: 10.1016/j.clinph.2008.11.015).
3. Jonathan R Wolpaw, Niels Birbaumer, Dennis J McFarland, Gert Pfurtscheller, Theresa M Vaughan. “Brain–computer interfaces for communication and control”. Clinical Neurophysiology - June 2002 (Vol. 113, Issue 6, Pages 767-791).
4. Schalk, G.; McFarland, D.J.; Hinterberger, T.; Birbaumer, N.; Wolpaw, J.R.;  “BCI2000: a general-purpose brain-computer interface (BCI) system”. IEEE Transactions on Biomedical Engineering, June 2004 (Vol. 51, Issue 6, Pages 1034 - 1043).
5. Gert Pfurtscheller, Gernot R. Muller, Jorg Pfurtscheller, Hans Jurgen Gerner, Rudiger Rupp, “ ‘Thought' - control of functional electrical stimulation to restore hand grasp in a patient with tetraplegia “. Neuroscience Letters, Volume 351, Issue 1, 6 November 2003, Pages 33-36, ISSN 0304-3940, DOI: 10.1016/S0304-3940(03)00947-9.
6. Mikhail A. Lebedev, Miguel A.L. Nicolelis, Brain-machine interfaces: past, present and future, Trends in Neurosciences, Volume 29, Issue 9, September 2006, Pages 536-546, ISSN 0166-2236, DOI: 10.1016/j.tins.2006.07.004.
7. Georg E. Fabiani, Dennis J. McFarland, Jonathan R. Wolpaw, and Gert Pfurtscheller*. “Conversion of EEG Activity Into Cursor Movement by a Brain–Computer Interface (BCI)* ”. IEEE Transactions on Neural Systems and Rehabilitation Engineering, VOL. 12, NO. 3, September 2004.
8. J. d. R. Millán, R. Rupp, G. R. Müller-Putz, R. Murray-Smith, C. Giugliemma,M. Tangermann, C. Vidaurre, F. Cincotti, A. Kübler, R. Leeb, C. Neuper, K.-R. Müller, and D. Mattia. “Combining Brain–Computer Interfaces and Assistive Technologies: State-of-the-Art and Challenges ”. Front Neurosci. 2010; 4: 161. Published online 2010 September 7. Prepublished online 2010 March 28.
9. Johannes Müller-Gerking, Gert Pfurtscheller, Henrik Flyvbjerg. “Designing optimal spatial filters for single-trial EEG classification in a movement task ”. Clinical Neurophysiology - 1 May 1999 (Vol. 110, Issue 5, Pages 787-798).
10. Dornhege, G.;   Blankertz, B.;   Krauledat, M.;   Losch, F.;   Curio, G.;   Muller, K.-R.; “Combined Optimization of Spatial and Temporal Filters for Improving Brain-Computer Interfacing”. IEEE Transactions on Biomedical Engineering, Nov 2006 (Vol. 53, Issue 11, Pages 2274 – 2281).
11. Blankertz, B.;   Tomioka, R.;   Lemm, S.;   Kawanabe, M.;   Muller, K.-R.; “Optimizing Spatial filters for Robust EEG Single-Trial Analysis ”. IEEE Signal Processing Magazine, 2008. (Vol. 25, Issue 1, Pages 41-56).
12. G. Pfurtscheiler, C. Guger, H. Ramoser. “EEG-based Brain-Computer Interface Using Subject-Specific Spatial Filters ”. Engineering Applications of Bio-Inspired Artificial Neural Networks, Lecture Notes in Computer Science, 1999. Springer Berlin / Heidelberg.
13. Naeem, M., C. Brunner, R. Leeb, B. Graimann, and G. Pfurtscheller. 2006. “Seperability of four-class motor imagery data using independent components analysis”. *J. Neural Eng.* 3:208–216.
14. McFarland, D.J., L.M. McCane, S.V. David, and J.R. Wolpaw. 1997. “Spatial filter selection for an EEG-based brain-computer interface”. *Electroencephalography and Clinical Neurophysiology* 103:386–394.
15. McFarland, D.J., C.W. Anderson, K-R. Muller, A. Schlogl, and D.J. Krusienski. 2006. BCI meeting 2005-workshop on “BCI signal processing: feature extraction and translation”. *IEEE Trans. Neural Syst. Rehab. Eng.* 14:135–138.
16. Vidaurre, C., A. Schlogl, R. Cabeza, R. Scherer, and G. Pfurtscheller. 2005. “Adaptive on-line classification for EEG-based brain computer interfaces with AAR parameters and band power estimates”. *Biomed. Technik* 50:350–354.
17. Boostani, R. and M.H. Moradi. 2004.” A new approach in the BCI research based on fractal dimension as feature and Adaboost as classifier”. *J. Neural Eng.* 1:212–217.
18. Krusienski, D.J., G. Schalk, D.J. McFarland, and J.R. Wolpaw. 2007. “A mu-rhythm matched filter for continuous control”.
19. Lemm, S., B. Blankertz, G. Curio, and R-K. Muller. 2005. “Spatio-spectral filters for improving the classification of single trial EEG”. *IEEE Trans. Biomed. Eng.* 52:1541–1548.
20. Millan, J.R., F. Renkens, J. Mourino, and W. Gerstner. 2002. “Noninvasive brain-actuated control of a mobile robot by human”. EEG. *IEEE Trans. Biomed. Eng.* 51:1026–1033.
21. Rezaei, S., K. Tavakolian, A.M. Nasrabadi, and S.K. Setarehdan. 2006. “Different classification techniques considering brain-computer interface applications”. *Journal of Neural Engineering* 3:139–144.
22. B. Blankertz, G. Dornhege, M. Krauledat, G. Curio, and K.-R. Müller, “The noninvasive Berlin Brain-Computer Interface: Fast acquisition of effective performance in untrained subjects,” NeuroImage, vol. 37, no. 2, pp. 539–550, 2007.
23. B. Blankertz, G. Dornhege, M. Krauledat, K.-R. Müller, V. Kunzmann, F. Losch, and G. Curio, “The Berlin Brain-Computer Interface: EEG-based communication without subject training,” IEEE Trans. Neural Sys. Rehab. Eng., vol. 14, no. 2 pp. 147–152, 2006.
1. [↑](#footnote-ref-1)