A Tutorial on EEG Signal Processing Techniques for Mental State Recognition in Brain-Computer Interfaces

Fabien Lotte

Chapter 7

A Tutorial on EEG Signal Processing Techniques for Mental State Recognition in Brain-Computer Interfaces

Fabien LOTTE

Abstract This chapter presents an introductory overview and a tutorial of signal processing techniques that can be used to recognize mental states from electroencephalographic (EEG) signals in Brain-Computer Interfaces. More particularly, this chapter presents how to extract relevant and robust spectral, spatial and temporal information from noisy EEG signals (e.g., Band Power features, spatial filters such as Common Spatial Patterns or xDAWN, etc.), as well as a few classification algorithms (e.g., Linear Discriminant Analysis) used to classify this information into a class of mental state. It also briefly touches on alternative, but currently less used approaches. The overall objective of this chapter is to provide the reader with practical knowledge about how to analyse EEG signals as well as to stress the key points to understand when performing such an analysis.

7.1 Introduction

One of the critical steps in the design of Brain-Computer Interface (BCI) applications based on ElectroEncephaloGraphy (EEG) is to process and analyse such EEG signals in real-time, in order to identify the mental state of the user. Musical EEG-based BCI applications are no exception. For instance, in (Miranda et al, 2011), the application had to recognize the visual target the user was attending to from his/her EEG signals, in order to execute the corresponding musical command. Unfortunately, identifying the user's mental state from EEG signals is no easy task, such signals being noisy, non-stationary, complex and of high dimensionality (Lotte et al, 2007). Therefore, mental state recognition from EEG signals requires specific signal processing and machine learning tools. This chapter aims at providing the reader with a basic knowledge about how to do EEG signal processing and the kind

Fabien LOTTE

Inria Bordeaux Sud-Ouest / LaBRI, 200 avenue de la vieille tour, 33405, Talence Cedex, France, e-mail: fabien.lotte@inria.fr

of algorithms to use to do so. This knowledge is - hopefully - presented in an accessible and intuitive way, by focusing more on the concepts and ideas than on the technical details.

This chapter is organized as follows: Section 7.2 presents the general architecture of an EEG signal processing system for BCI. Then, Section 7.3 describes the specific signal processing tools that can be used to design BCI based on oscillatory EEG activity while Section 7.4 describes those that can used for BCI based on Event Related Potentials (ERP), i.e., brain responses to stimulus and events. Section 7.5 presents some alternative tools, still not as popular as the one mentioned so far but promising, both for BCI based on oscillatory activity and those based on ERP. Finally, Section 7.6 proposes a discussion about all the tools covered and their perspectives while Section 7.7 concludes the paper.

7.2 General EEG signal processing principle

In BCI design, EEG signal processing aims at translating raw EEG signals into the class of these signals, i.e., into the estimated mental state of the user. This translation is usually achieved using a pattern recognition approach, whose two main steps are the following:

- **Feature Extraction:** The first signal processing step is known as "feature extraction" and aims at describing the EEG signals by (ideally) a few relevant values called "features" (Bashashati et al, 2007). Such features should capture the information embedded in EEG signals that is relevant to describe the mental states to identify, while rejecting the noise and other non-relevant information. All features extracted are usually arranged into a vector, known as a feature vector.
- **Classification:** The second step, denoted as "classification" assigns a class to a set of features (the feature vector) extracted from the signals (Lotte et al, 2007). This class corresponds to the kind of mental state identified. This step can also be denoted as "feature translation" (Mason and Birch, 2003). Classification algorithms are known as "classifiers".

As an example, let us consider a Motor Imagery (MI)-based BCI, i.e., a BCI that can recognized imagined movements such left hand or right hand imagined movements (see Figure 7.1). In this case, the two mental states to identify are imagined left hand movement on one side and imagined right hand movement on the other side. To identify them from EEG signals, typical features are band power features, i.e., the power of the EEG signal in a specific frequency band. For MI, band power features are usually extracted in the μ (about $8 - 12$ Hz) and β (about $16 - 24$ Hz) frequency bands, for electrode localized over the motor cortex areas of the brain (around locations C3 and C4 for right and left hand movements respectively) (Pfurtscheller and Neuper, 2001). Such features are then typically classified using a Linear Discriminant Analysis (LDA) classifier.
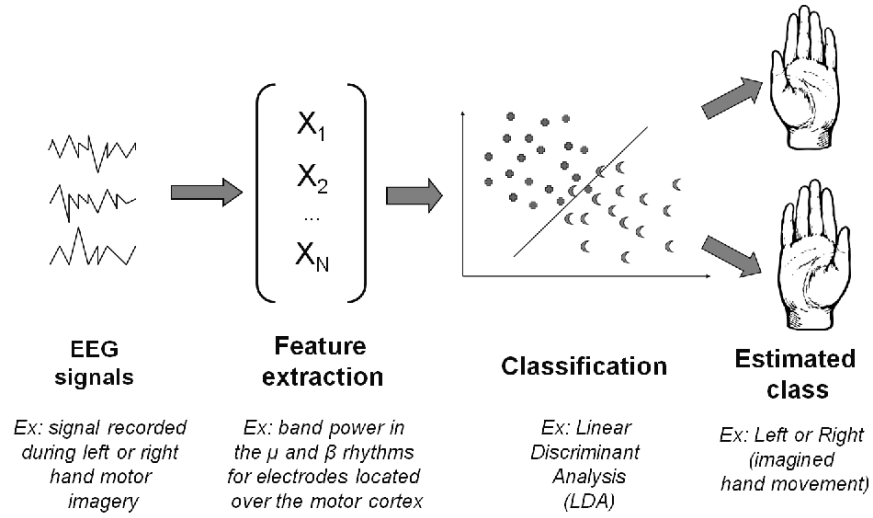
EEG signals	**Feature extraction**	**Classification**	**Estimated class**
Ex: signal recorded during left or right hand motor imagery	*Ex: band power in the μ and β rhythms for electrodes located over the motor cortex*	*Ex: Linear Discriminant Analysis (LDA)*	*Ex: Left or Right (imagined hand movement)*

Fig. 7.1 A classical EEG signal processing pipeline for BCI, here in the context of a motor imagery-based BCI, i.e., a BCI that can recognized imagined movements from EEG signals.

It should be mentioned that EEG signal processing is often built using machine learning. This means the classifier and/or the features are automatically tuned, generally for each user, according to examples of EEG signals from this user. These examples of EEG signals are called a training set, and are labeled with their class of belonging (i.e., the corresponding mental state). Based on these training examples, the classifier will be tuned in order to recognize as appropriately as possible the class of the training EEG signals. Features can also be tuned in such a way, e.g., by automatically selecting the most relevant channels or frequency bands to recognized the different mental states. Designing BCI based on machine learning (most current BCI are based on machine learning) therefore consists of 2 phases:

- **Calibration** (a.k.a., training) phase: This consists in 1) Acquiring training EEG signals (i.e., training examples) and 2) Optimizing the EEG signal processing pipeline by tuning the feature parameters and/or training the classifier.
- **Use** (a.k.a., test) phase: This consists in using the model (features and classifier) obtained during the calibration phase in order to recognize the mental state of the user from previously unseen EEG signals, in order to operate the BCI.

Feature extraction and classification are discussed in more details hereafter.

7.2.1 Classification

As mentioned above, the classification step in a BCI aims at translating the features into commands (McFarland et al, 2006) (Mason and Birch, 2003). To do so, one can use either regression algorithms (McFarland and Wolpaw, 2005) (Duda et al, 2001) or classification algorithms (Penny et al, 2000) (Lotte et al, 2007), the classification algorithms being by far the most used in the BCI community (Bashashati et al, 2007) (Lotte et al, 2007). As such, in this chapter, we focus only on classification algorithms. Classifiers are able to learn how to identify the class of a feature vector, thanks to training sets, i.e., labeled feature vectors extracted from the training EEG examples.

Typically, in order to learn which kind of feature vector correspond to which class (or mental state), classifiers try either to model which area of the feature space is covered by the training feature vectors from each class - in this case the classifier is a generative classifier - or they try to model the boundary between the areas covered by the training feature vectors of each class - in which case the classifier is a discriminant classifier. For BCI, the most used classifiers so far are discriminant classifiers, and notably Linear Discriminant Analysis (LDA) classifiers.

The aim of LDA (also known as Fisher's LDA) is to use hyperplanes to separate the training feature vectors representing the different classes (Duda et al, 2001) (Fukunaga, 1990). The location and orientation of this hyperplane is determined from training data. Then, for a two-class problem, the class of an unseen (a.k.a., test) feature vector depends on which side of the hyperplane the feature vector is (see Figure 7.2). LDA has very low computational requirements which makes it suitable for online BCI system. Moreover this classifier is simple which makes it naturally good at generalizing to unseen data, hence generally providing good results in practice (Lotte et al, 2007). LDA is probably the most used classifier for BCI design.
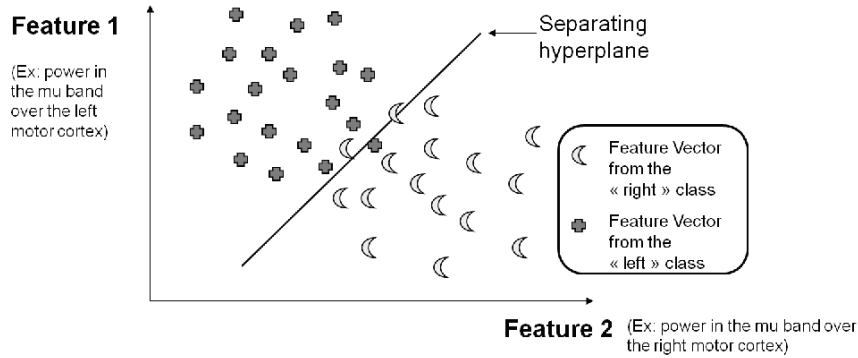

Fig. 7.2 Discriminating two types of motor imagery with a linear hyperplane using a Linear Discriminant Analysis (LDA) classifier.

Another very popular classifier for BCI is the Support Vector Machine (SVM) (Bennett and Campbell, 2000). An SVM also uses a discriminant hyperplane to identify classes (Burges, 1998). However, with SVM, the selected hyperplane is the one that maximizes the margins, i.e., the distance from the nearest training points, which has been found to increase the generalization capabilites (Burges, 1998) (Bennett and Campbell, 2000).

Generally, regarding classification algorithms, it seems that very good recognition performances can be obtained using appropriate off-the-shelf classifiers such as LDA or SVM (Lotte et al, 2007). What seems to be really important is the design and selection of appropriate features to describe EEG signals. With this purpose, specific EEG signal processing tools have been proposed to design BCI. In the rest of this chapter we will therefore focus on EEG feature extraction tools for BCI. For readers interested to learn more about classification algorithms, we refer them to (Lotte et al, 2007), a review paper on this topic.

7.2.2 Feature extraction

As mentioned before, feature extraction aims at representing raw EEG signals by an ideally small number of relevant values, which describe the task-relevant information contained in the signals. However, classifiers are able to learn from data which class corresponds to which input features. As such, why not using directly the EEG signals as input to the classifier? This is due to the so-called "curse-of-dimensionality", which states that the amount of data needed to properly describe the different classes increases exponentially with the dimensionality of the feature vectors (Jain et al, 2000) (Friedman, 1997). It has been recommended to use from 5 to 10 times as many training examples per class as the input feature vector dimensionality[1] (Raudys and Jain, 1991). What would it mean to use directly the EEG signals as input to the classifier? Let us consider a common steup with 32 EEG sensors sampled at 250Hz, with one trial of EEG signal being 1 second long. This would mean a dimensionality of $32*250 = 8000$, which would require at least 40000 training examples. Obviously we cannot ask the BCI user to perform each mental task 40000 times to calibrate the BCI before he/she could use it. A much more compact representation is therefore needed, hence the necessity to perform some form of feature extraction.

With BCI, there are 3 main sources of information that can be used to extract features from EEG signals:

- **Spatial information:** Such features would describe where (spatially) the relevant signal comes from. In practice, this would mean selecting specific EEG channels, or focusing more on specific channels than on some other. This amounts to focusing on the signal originating from specific areas of the brain.

---

[1] note that this was estimated before SVM were invented, and that SVM are generally less sensitive - although not completely immune - to this curse-of-dimensionality

- **Spectral (frequential) information:** Such features would describe how the power in some relevant frequency bands varies. In practice, this means that the features will use the power in some specific frequency bands.
- **Temporal information:** Such features would describe how the relevant signal varies with time. In practice this means using the EEG signals values at different time points or in different time windows.

Note that these three sources of information are not the only ones, and alternatives can be used (see Section 7.5). However, they are by far the most used one, and, at least so far, the most efficient ones in terms of classification performances. It should be mentioned that so far, nobody managed to discover nor to design a set of features that would work for all types of BCI. As a consequence, different kinds of BCI currently use different sources of information. Notably, BCI based on oscillatory activity (e.g., BCI based on motor imagery) mostly need and use the spectral and spatial information whereas BCI based on event related potentials (e.g., BCI based on the P300) mostly need and use the temporal and spatial information. The next sections detail the corresponding tools for these two categories of BCI.

7.3 EEG signal processing tools for BCI based on oscillatory activity

BCI based on oscillatory activity are BCI that use mental states which lead to changes in the oscillatory components of EEG signals, i.e., that lead to change in the power of EEG signals in some frequency bands. Increase of EEG signal power in a given frequency band is called an Event Related Synchronisation (ERS), whereas a decrease of EEG signal power is called an Event Related Desynchronisation (ERD) (Pfurtscheller and da Silva, 1999). BCI based on oscillatory activity notably includes motor imagery-based BCI (Pfurtscheller and Neuper, 2001), Steady State Visual Evoked Potentials (SSVEP)-based BCI (Vialatte et al, 2010) as well as BCI based on various cognitive imagery tasks such as mental calculation, mental geometric figure rotation, mental word generation, etc. (Friedrich et al, 2012) (Millán et al, 2002). As an example, imagination of a left hand movement leads to a contralateral ERD in the motor cortex (i.e., in the right motor cortex for left hand movement) in the μ and β bands during movement imagination, and to an ERS in the β band (a.k.a., beta rebound) just after the movement imagination ending (Pfurtscheller and da Silva, 1999). This section first describes a basic design for oscillatory activity-based BCI. Then, due to the limitations exhibited by this design, it exposes more advanced designs based on multiple EEG channels. Finally, it presents a key tool to design such BCIs: the Common Spatial Pattern (CSP) algorithm, as well as some of its variants.

7.3.1 Basic design for an oscillatory activity-based BCI

Oscillatory activity-based BCI are based on change in power in some frequency bands, in some specific brain areas. As such, they naturally need to exploit both the spatial and spectral information. As an example, a basic design for a motor-imagery BCI would exploit the spatial information by extracting features only from EEG channels localized over the motor areas of the brain, typically channels C3 for right hand movements, Cz for foot movements and C4 for left hand movements. It would exploit the spectral information by focusing on frequency bands μ ($8-12$ Hz) and β ($16-24$ Hz). More precisely, for a BCI that can recognize left hand MI versus right hand MI, the basic features extracted would be the average band power in $8-12$ Hz and $16-24$ Hz from both channels C3 and C4. Therefore, the EEG signals would be described by only 4 features.

There are many ways to compute band power features from EEG signals (Herman et al, 2008) (Brodu et al, 2011). However, a simple, popular and efficient one is to first band-pass filter the EEG signal from a given channel into the frequency band of interest, then to square the resulting signal to compute the signal power, and finally to average it over time (e.g., over a time window of 1 s). This is illustrated in Figure 7.3.
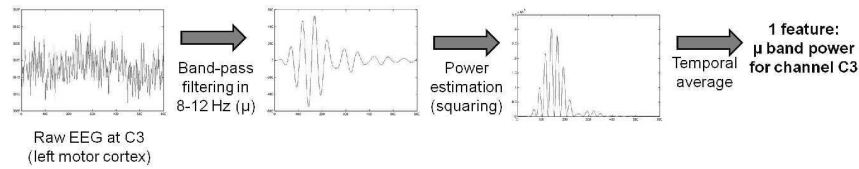

Fig. 7.3 Signal processing steps to extract band power features from raw EEG signals. The EEG signal displayed here was recorded during right hand motor imagery (the instruction to perform the imagination was provided at $t = 0$ s on the plots). The contralateral ERD during imagination is here clearly visible. Indeed, the signal power in channel C3 (left motor cortex) in 8-12 Hz clearly decreases during this imagination of a right hand movement.

Unfortunately, this basic design is far from being optimal. Indeed, it uses only two fixed channels. As such, relevant information, measured by other channels might be missing, and C3 and C4 may not be the best channels for the subject at hand. Similarly, using the fixed frequency bands $8-12$ Hz and $16-24$ Hz may not be the optimal frequency bands for the current subject. In general, much better performances are obtained when using subject-specific designs, with the best channels and frequency bands optimized for this subject. Using more than two channels is also known to lead to improved performances, since it enables to collect the relevant information spread over the various EEG sensors.

7.3.2 Towards advanced BCI using multiple EEG channels

Both the need to use subject-specific channels and the need to use more than 2 channels lead to the necessity to design BCI based on multiple channels. This is confirmed by various studies which suggested that, for motor imagery, 8 channels is a minimum to obtain reasonnable performances (Sannelli et al, 2010) (Arvaneh et al, 2011), with optimal performances achieved with a much larger number, e.g., 48 channels in (Sannelli et al, 2010). However, simply using more channels will not solve the problem. Indeed, using more channels means extracting more features, thus increasing the dimensionality of the data and suffering more from the curse-of-dimensionality. As such, just adding channels may even decrease performances if too little training data is available. In order to efficiently exploit multiple EEG channels, 3 main approaches are available, all of which contribute to reducing the dimensionality:

- **Feature selection** algorithm: These are methods to select automatically a subset of relevant features, among all the features extracted.
- **Channel selection** algorithms: These are similar methods that select automatically a subset of relevant channels, among all channels available.
- **Spatial Filtering** algorithms: These are methods that combine several channels into a single one, generally using weighted linear combinations, from which features will be extracted.

They are described below.

7.3.2.1 Feature selection:

Feature selection are classical algorithms widely used in machine learning (Guyon and Elisseeff, 2003) (Jain and Zongker, 1997) and as such also very popular in BCI design (Garrett et al, 2003). There are too main families of feature selection algorithms:

- *Univariate* algorithms: They evaluate the discriminative (or descriptive) power of each feature individually. Then, they select the N best individual features (N needs to be defined by the BCI designer). The usefulness of each feature is typically assessed using measures such as Student t-statistics, which measures the feature value difference between two classes, correlation based measures such as R^2, mutual information, which measures the dependence between the feature value and the class label, etc. (Guyon and Elisseeff, 2003). Univariate methods are usually very fast and computationally efficient but they are also suboptimal. Indeed, since they only consider the individual feature usefulness, they ignore possible redundancies or complementarities between features. As such, the best subset of N features is usually not the N best individual features. As an example, the N best individual features might be highly redundant and measure almost the

same information. As such using them together would add very little discriminant power. On the other hand, adding a feature that is individually not very good but which measures a different information from that of the best individual ones is likely to improve the discriminative power much more.

- *Multivariate* algorithms: They evaluate subsets of features together, and keep the best subset with N features. These algorithms typically use measures of global performance for the subsets of features, such as measures of classification performances on the training set (typically using cross-validation (Browne, 2000)) or multivariate mutual information measures, see, e.g., (Hall, 2000) (Pudil et al, 1994) (Peng et al, 2005). This global measure of performance enables to actually consider the impact of redundancies or complementarities between features. Some measures also remove the need to manually select the value of N (the number of features to keep), the best value of N being the number of features in the best subset identified. However, evaluating the usefulness of subsets of features leads to very high computational requirements. Indeed, there are many more possible subsets of any size than individual features. As such there are many more evaluations to perform. In fact, the number of possible subsets to evaluate is very often far too high to actually perform all the evaluations in practice. Consequently, multivariate methods usually rely on heuristics or greedy solutions in order to reduce the number of subsets to evaluate. They are therefore also suboptimal but usually give much better performances than univariate methods in practice. On the other hand, if the initial number of features is very high, multivariate methods may be too slow to use in practice.

7.3.2.2 Channel selection:

Rather than selecting features, one can also select channels and only use features extracted from the selected channels. While both channel and feature selection reduce the dimensionality, selecting channels instead of features has some additional advantages. In particular using less channels means a faster setup time for the EEG cap and also a lighter and more comfortable setup for the BCI user. It should be noted, however, that with the development of dry EEG channels, selecting channels may become less crucial. Indeed the setup time will not depend on the number of channel used, and the BCI user will not have more gel in his/her hair if more channels are used. With dry electrodes, using less channels will still be lighter and more comfortable for the user though.

Algorithms for EEG channel selection are usually based or inspired from generic feature selection algorithm. Several of them are actually analogous algorithms that assess individual channel useufulness or subsets of channels discriminative power instead of individual features or subset of features. As such, they also use similar performance measures, and have similar properties. Some other channel selection algorithms are based on spatial filter optimization (see below). Readers interested to know more about EEG channel selection may refer to the following papers and

associated references (Schröder et al, 2005) (Arvaneh et al, 2011) (Lal et al, 2004) (Lan et al, 2007), among many other.

7.3.2.3 Spatial filtering:

Spatial filtering consists in using a small number of new channels that are defined as a linear combination of the original ones:

$$\tilde{x} = \sum_i w_i x_i = wX \qquad (7.1)$$

with $\tilde{x}$ the spatially filtered signal, x_i the EEG signal from channel i, w_i the weight given to that channel in the spatial filter and X a matrix whose i^{th} row is x_i, i.e., X is the matrix of EEG signals from all channels.

It should be noted that spatial filtering is useful not only because it reduces the dimension from many EEG channels to a few spatially filtered signals (we typically use much less spatial filters than original channels), but also because it has a neurophysiological meaning. Indeed, with EEG, the signals measured on the surface of the scalp are a blurred image of the signals originating from within the brain. In other words, due to the smearing effect of the skull and brain (a.k.a., volume conduction effect), the underlying brain signal is spread over several EEG channels. Therefore spatial filtering can help recovering this original signal by gathering the relevant information that is spread over different channels.

There are different ways to define spatial filters. In particular, the weights w_i can be fixed in advance, generally according to neurophysiological knowledge, or they can be data driven, that is, optimized on training data. Among the fixed spatial filters we can notably mention the bipolar and Laplacian which are local spatial filters that try to locally reduce the smearing effect and some of the background noise (McFarland et al, 1997). A bipolar filter is defined as the difference between 2 neighboring channels, while a Laplacian filter is defined as 4 times the value of a central channel minus the values of the 4 channels around. For instance, a bipolar filter over channel C3 would be defined as $C3_{bipolar} = FC3 - CP3$, while a Laplacian filter over C3 would be defined as $C3_{Laplacian} = 4C3 - FC3 - C5 - C1 - CP3$, see also Figure 7.4. Extracting features from bipolar or Laplacian spatial filters rather than from the single corresponding electrodes has been shown to significantly increase classification performances (McFarland et al, 1997). An inverse solution is another kind of fixed spatial filter (Michel et al, 2004) (Baillet et al, 2001). Inverse solutions are algorithms that enable to estimate the signals originating from sources within the brain based on the measurements taken from the scalp. In other words, inverse solutions enable us to look into the activity of specific brain regions. A word of caution though: inverse solutions do not provide more information than what is already available in scalp EEG signals. As such, using inverse solutions will NOT make a non-invasive BCI as accurate and efficient as an invasive one. However, by focusing on some specific brain areas, inverse solutions can contribute to reducing background noise, the smearing effect and irrelevant information originating from

other areas. As such, it has been shown than extracting features from the signals spatially filtered using inverse solutions (i.e., from the sources within the brain) leads to higher classification performances than extracting features directly from scalp EEG signals (Besserve et al, 2011) (Noirhomme et al, 2008). In general, using inverse solutions has been shown to lead to high classification performances (Congedo et al, 2006) (Lotte et al, 2009b) (Qin et al, 2004) (Kamousi et al, 2005) (Grosse-Wentrup et al, 2005). It should be noted that since the number of source signals obtained with inverse solutions is often larger than the initial number of channels, it is necessary to use feature selection or dimensionality reduction algorithms.
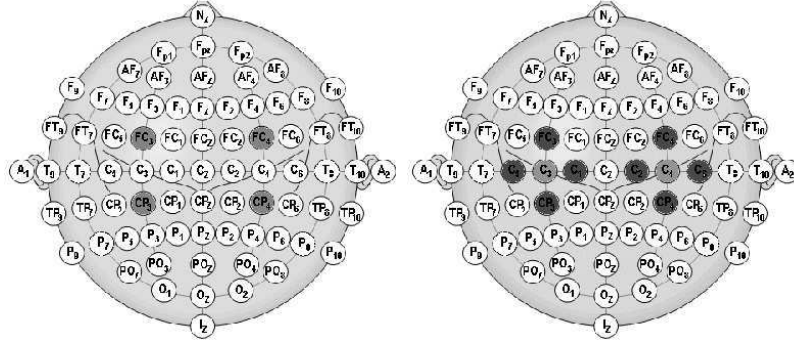

Fig. 7.4 Left: channels used in bipolar spatial filtering over channels C3 and C4. Right: channels used in Laplacian spatial filtering over channels C3 and C4.

The second category of spatial filters, i.e., data driven spatial filters, are optimized for each subject according to training data. As any data driven algorithm, the spatial filter weights w_i can be estimated in an unsupervised way, that is without the knowledge of which training data belongs to which class, or in a supervised way, with each training data being labelled with its class. Among the unsupervised spatial filters we can mention Principal Component Analysis (PCA), which finds the spatial filters that explain most of the variance of the data, or Independent Component Analysis (ICA), which find spatial filters whose resulting signals are independent from each other (Kachenoura et al, 2008). The later has been shown rather useful to design spatial filters able to remove or attenuate the effect of artifacts (EOG, EMG, etc. (Fatourechi et al, 2007)) on EEG signals (Tangermann et al, 2009) (Xu et al, 2004) (Kachenoura et al, 2008) (Brunner et al, 2007). Alternatively, spatial filters can be optimized in a supervised way, i.e., the weights will be defined in order to optimize some measure of classification performance. For BCI based on oscillatory EEG activity, such a spatial filter has been designed: the Common Spatial Patterns (CSP) algorithm (Ramoser et al, 2000) (Blankertz et al, 2008b). This algorithm has greatly contributed to the increase of performances of this kind of BCI, and, thus, has become a standard tool in the repertoire of oscillatory activity-based BCI de-

signers. It is described in more details in the following section, together with some of its variants.

7.3.3 Common Spatial Patterns and variants

Informally, the CSP algorihtm finds spatial filters w such that the variance of the filtered signal is maximal for one class and minimal for the other class. Since the variance of a signal band-pass filtered in band b is actually the band-power of this signal in band b, this means that CSP finds spatial filters that lead to optimally discriminant band-power features since their values would be maximally different between classes. As such, CSP is particularly useful for BCI based on oscillatory activity since their most useful features are band-power features. As an example, for BCI based on motor imagery, EEG signals are typically filtered in the $8 - 30$ Hz band before being spatially filtered with CSP (Ramoser et al, 2000). Indeed this band contains both the μ and β rhythms.

Formally, CSP uses the spatial filters w which extremize the following function:

$$J_{CSP}(w) = \frac{wX_1X_1^Tw^T}{wX_2X_2^Tw^T} = \frac{wC_1w^T}{wC_2w^T} \tag{7.2}$$

where T denotes transpose, X_i is the training band-pass filtered signal matrix for class i (with the samples as columns and the channels as rows) and C_i the spatial covariance matrix from class i. In practice, the covariance matrix C_i is defined as the average covariance matrix of each trial from class i (Blankertz et al, 2008b). In this equation, wX_i is the spatially filtered EEG signal from class i, and $wX_iX_i^Tw^T$ is thus the variance of the spatially filtered signal, i.e., the band-power of the spatially filtered signal. Therefore, extremizing $J_{CSP}(w)$, i.e., maximizing and minimizing it, indeed leads to spatially filtered signals whose band-power is maximally different between classes. $J_{CSP}(w)$ happens to be a Rayleigh quotient. Therefore, extremizing it can be solved by Generalized Eigen Value Decomposition (GEVD). The spatial filters w that maximize or minimize $J_{CSP}(w)$ are thus the eigenvectors corresponding to the largest and lowest eigenvalues, respectively, of the GEVD of matrices C_1 and C_2. Typically, 6 filters (i.e., 3 pairs), corresponding to the 3 largest and 3 lowest eigenvalues are used. Once these filters obtained, a CSP feature f is defined as follows:

$$f = \log(wXX^Tw^T) = \log(wCw^T) = \log(var(wX)) \tag{7.3}$$

i.e., the features used are simply the band power of the spatially filtered signals. CSP requires more channels than fixed spatial filters such as Bipolar or Laplacian, however in practice, it usually leads to significantly higher classification performances (Ramoser et al, 2000). The use of CSP is illustrated in Figure 7.5. In this figure, the signals spatially filtered with CSP clearly show difference in variance (i.e., in band power) between the two classes, hence ensuring high classification performances.
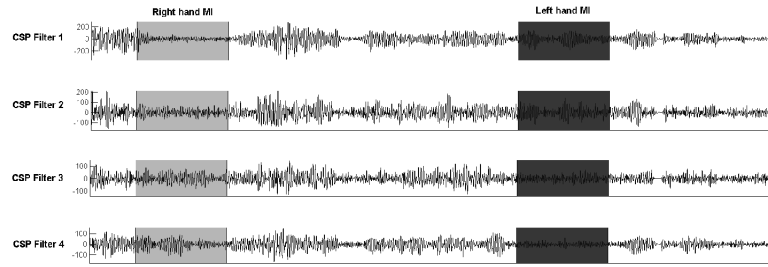
Fig. 7.5 EEG signals spatially filtered using the CSP (Common Spatial Patterns) algorithm. The first two spatial filters (top filters) are those maximizing the variance of signals from class "Left Hand Motor Imagery" while minimizing that of class "Right Hand Motor Imagery". They correspond to the largest eigen values of the GEVD. The last two filters (bottom filters) are the opposite, they maximize the variance of class "Right Hand Motor Imagery" while minimizing that of class "Left Hand Motor Imagery" (They correspond to the lowest eigen values of the GEVD). This can be clearly seen during the periods of right or left hand motor imagery, in light and dark grey respectively.

The CSP algorithm has numerous advantages: first, it leads to high classification performances. CSP is also versatile, since it works for any ERD/ERS BCI. Finally, it is computationally efficient and simple to implement. Altogether this makes CSP one of the most popular and efficient approach for BCI based on oscillatory activity (Blankertz et al, 2008b).

Nevertheless, despite all these advantages, CSP is not exempt from limitations and is still not the ultimate signal processing tool for EEG-based BCI. In particular, CSP has been shown to be non-robust to noise, to non-stationarities and prone to overfitting (i.e., it may not generalize well to new data) when little training data is available (Grosse-Wentrup and Buss, 2008) (Grosse-Wentrup et al, 2009) (Reuderink and Poel, 2008). Finally, despite its versatility, CSP only identifies the relevant spatial information but not the spectral one. Fortunately, there are ways to make CSP robust and stable with limited training data and with noisy training data. An idea is to integrate prior knowledge into the CSP optimization algorithm. Such knowledge could represent any information we have about what should be a good spatial filter for instance. This can be neurophysiological prior, data (EEG signals) or meta-data (e.g., good channels) from other subjects, etc. This knowledge is used to guide and constraint the CSP optimization algorithm towards good solutions even with noise, limited data and non-stationarities (Lotte and Guan, 2011). Formally, this knowledge is represented in a regularization framework that penalizes unlikely solutions (i.e., spatial filters) that do not satisfy this knowledge, therefore enforcing it. Similarly, prior knowledge can be used to stabilize statistical estimates (here, covariance matrices) used to optimize the CSP algorithm. Indeed, estimating covariance matrices from few training data usually leads to poor estimates (Ledoit and Wolf, 2004).

Formally, a Regularized CSP (RCSP) can be obtained by maximizing both equation 7.4 and 7.5:

$$J_{RCSP1}(w) = \frac{w\tilde{C}_1 w^T}{w\tilde{C}_2 w^T + \lambda P(w)} \tag{7.4}$$

$$J_{RCSP2}(w) = \frac{w\tilde{C}_2 w^T}{w\tilde{C}_1 w^T + \lambda P(w)} \tag{7.5}$$

with

$$\tilde{C}_i = (1 - \gamma)C_i + \gamma G_i \tag{7.6}$$

In these equations, $P(w)$ is the penalty term that encodes the prior knowledge. This a positive function of the spatial filter w, whose value will increase if w does not satisfy the knowledge encoded. Since the filters are obtained by maximizing J_{RCSPi}, this means that the numerator (which is positive) must be maximized and the denominator (which is also positive) must be minimized. Since $P(w)$ is positive and part of the denominator, this means that $P(w)$ will be minimized as well, hence enforcing that the spatial filters w satisfy the prior knowledge. Matrix G_i is another way of using prior knowledge, in order to stabilize the estimates of the covariance matrices C_i. If we have any idea about how these covariance matrices should be, this can be encoded in G_i in order to define a new covariance matrix $\tilde{C}_i$ which is a mix of the matrix C_i estimated on the data and of the prior knowledge G_i. We will present below what kind of knowledge can be encoded in $P(w)$ and G_i.

For the penalty term $P(w)$, a kind of knowledge that can be used is spatial knowledge. For instance, from a neurophysiological point of view, we know that neighboring neurons tend to have similar functions, which supports the idea that neighboring electrodes should measure similar brain signals (if the electrodes are close enough to each other), notably because of the smearing effect. Thus neighboring electrodes should have similar contributions in the spatial filters. In other words, spatial filters should be spatially smooth. This can be enforced by using the following penalty term:

$$P(w) = \sum_{i,j} Prox(i, j)(w_i - w_j)^2 \tag{7.7}$$

Where $Prox(i, j)$ measures the proximity of electrodes i and j, and $(w_i - w_j)^2$ is the weight difference between electrodes i and j, in the spatial filter. Thus, if two electrodes are close to each other and have very different weights, the penalty term $P(w)$ will be high, which would prevent such solutions to be selected during the optimization of the CSP (Lotte and Guan, 2010b). Another knowledge that can be used is that for a given mental task, not all the brain regions are involved and useful. As such, some electrodes are unlikely to be useful to classify some specific mental tasks. This can be encoded in $P(w)$ as well:

$$P(w) = wDw^T \quad with \quad D(i,j) = \begin{cases} channel \quad i \quad \text{``uselessness''} \quad if \quad i = j \\ 0 \qquad\qquad\qquad\qquad\qquad\qquad otherwise \end{cases}$$
(7.8)

Basically, the value of $D(i,i)$ is the penalty for the ith channel. The higher this penalty, the less likely this channel will have a high contribution in the CSP filters. The value of this penalty can be defined according to neurophysiological prior knowledge for instance, large penalties being given to channels unlikely to be useful and small or no penalty being given to channels that are likely to genuinely contribute to the filter. However, it may be difficult to precisely define the extent of the penalty from the literature. Another alternative is the use data previously recorded from other subjects. Indeed, the optimized CSP filters already obtained from previous subject give information about which channels have large contributions on average. The inverse of the average contribution of each channel can be used as the penalty, hence penalizing channels with small average contribution (Lotte and Guan, 2011). Penalty terms are therefore also a nice way to perform subject-to-subject transfer and re-use information from other subjects. These two penalties are examples that have proven useful in practice. This usefulness is notably illustrated in Figure 7.6, in which spatial filters obtained with the basic CSP are rather noisy, with strong contributions from channels not expected from a neurophysiological point of view. On the contrary the spatial filters obtained using the two RCSP penalties described previously are much cleaner, spatially smoother and with strong contributions localized in neurophysiologically relevant areas. This in turns led to higher classification performances, with CSP obtaining 73.1% classification accuracy versus 78.7% and 77.6% for the regularized versions (Lotte and Guan, 2011). It should be mentioned, however, that strong contributions from non-neurophysiologically relevant brain areas in a CSP spatial filter may be present to perform noise-cancellation, and as such does not mean the spatial filter is bad per se (Haufe et al, 2014). It should also be mentioned that other interesting penalty terms have been proposed, in order to deal with known noise sources (Blankertz et al, 2008a), non-stationarities (Samek et al, 2012) or to perform simultaneous channel selection (Farquhar et al, 2006) (Arvaneh et al, 2011).

Matrix G_i in equation 7.6 is another way to add prior knowledge. This matrix can notably be defined as the average covariance matrix obtained from other subjects who performed the same task. At such it enables to define a good and stable estimate of the covariance matrices, even if few training EEG data is available for the target subject. This has been shown to enable us to calibrate BCI system with 2 to 3 times less training data than with the basic CSP, while maintaining classification performances (Lotte and Guan, 2010a).

Regularizing CSP using a-priori knowledge is thus a nice way to deal with some limitations of CSP such as its sensitivity to overfitting and its non-robustness to noise. However, these regularized algorithms cannot address the limitation that CSP only optimizes the use of the spatial information, but not that of the spectral one. In general, independently of the use of CSP, there are several ways to optimize the use of the spectral information. Typically, this consists in identifying, in one way or another, the relevant frequency bands for the current subject and mental tasks per-
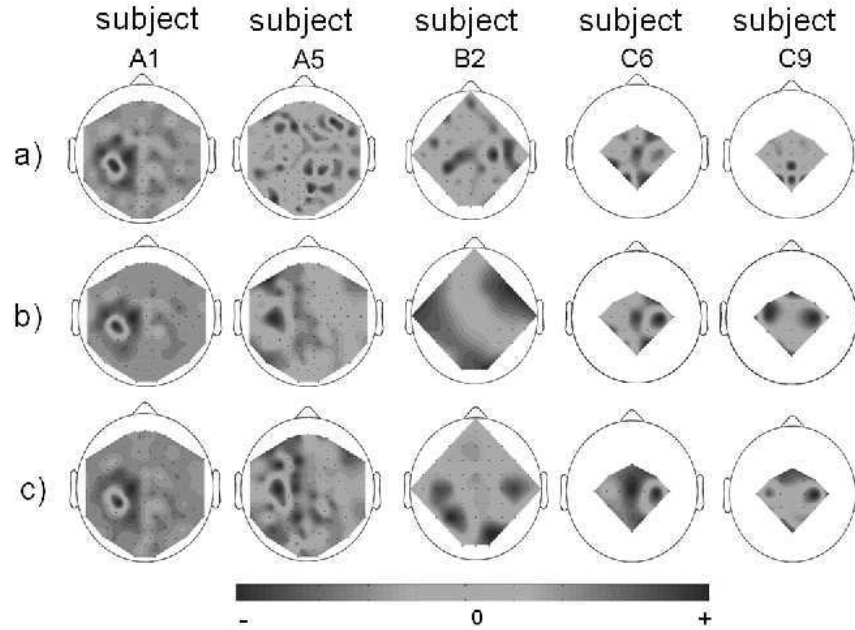
Fig. 7.6 Spatial filters (i.e., weight attributed to each channel) obtained to classify left hand versus right hand motor imagery. The electrodes, represented by black dots, are here seen from above, with the subject nose on top. a) basic CSP algorithm, b) RCSP with a penalty term imposing spatial smoothness, c) RCSP with a penalty term penalizing unlikely channels according to EEG data from other subjects.

formed. For instance, this can be done manually (by trial and errors), or by looking at the average EEG frequency spectrum in each class. In a more automatic way, possible methods include extracting band power features in multiple frequency bands and then selecting the relevant ones using feature selection (Lotte et al, 2010), by computing statistics on the spectrum to identify the relevant frequencies (Zhong et al, 2008), or even by computing optimal band-pass filters for classification (Devlaminck, 2011). These ideas can be used within the CSP framework in order to optimize the use of both the spatial and spectral information. Several variants of CSP has been proposed in order to optimize spatial and spectral filters at the same time (Lemm et al, 2005) (Dornhege et al, 2006) (Tomioka et al, 2006) (Thomas et al, 2009). A simple and computationally efficient method is worth describing: the Filter Bank CSP (FBCSP) (Ang et al, 2012). This method, illustrated in Figure 7.7, consists in first filtering EEG signals in multiple frequency bands using a filter bank. Then, for each frequency band, spatial filters are optimized using the classical CSP algorithm. Finally, among the multiple spatial filters obtained, the best resulting features are selected using feature selection algorithms (typically mutual information-based feature selection). As such, this selects both the best spectral and

spatial filters since each feature corresponds to a single frequency band and CSP spatial filter. This algorithm, although simple, has proven to be very efficient in practice. It was indeed the algorithm used in the winning-entries of all EEG data sets from the last BCI competition[2] (Ang et al, 2012).
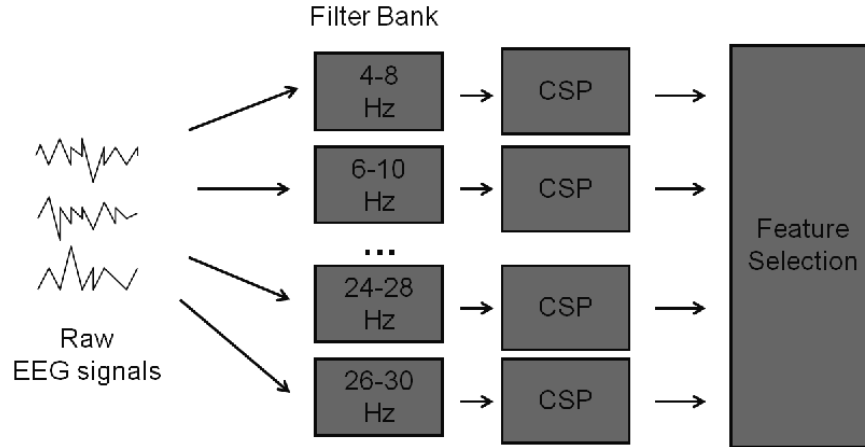

Fig. 7.7 Principle of Filter Bank Common Spatial Patterns (FBCSP): 1) band-pass filtering the EEG signals in multiple frequency bands using a filter bank; 2) optimizing CSP spatial filter for each band; 3) selecting the most relevant filters (both spatial and spectral) using feature selection on the resulting features.

7.3.4 Summary for oscillatory activity-based BCI

In summary, when designing BCI aiming at recognizing mental states that involve oscillatory activity, it is important to consider both the spectral and the spatial information. In order to exploit the spectral information, using band power features in relevant frequency bands is an efficient approach. Feature selection is also a nice tool to find the relevant frequencies. Concerning the spatial information, using or selecting relevant channels is useful. Spatial filtering is a very efficient solution for EEG-based BCI in general, and the Common Spatial Patterns (CSP) algorithm is a must-try for BCI based on oscillatory activity in particular. Moreover, there are several variants of CSP that are available in order to make it robust to noise, non-stationarity, limited training data sets or to jointly optimize spectral and spatial filters. The next section will address the EEG signal processing tools for BCI based

---

[2] BCI competitions are contests to evaluate the best signal processing and classification algorithms on given brain signals data sets. See http://www.bbci.de/competition/ for more info.

on evoked potentials, which are different from the ones described so far, but share some general concepts.

7.4 EEG signal processing tools for BCI based on event related potentials

An Event Related Potential (ERP) is a brain responses due to some specific stimulus perceived by the BCI user. A typical ERP used for BCI design is the P300, which is a positive deflection of the EEG signal occurring about 300ms after the user perceived a rare and relevant stimulus (Fazel-Rezai et al, 2012) (see also Figure 7.8).
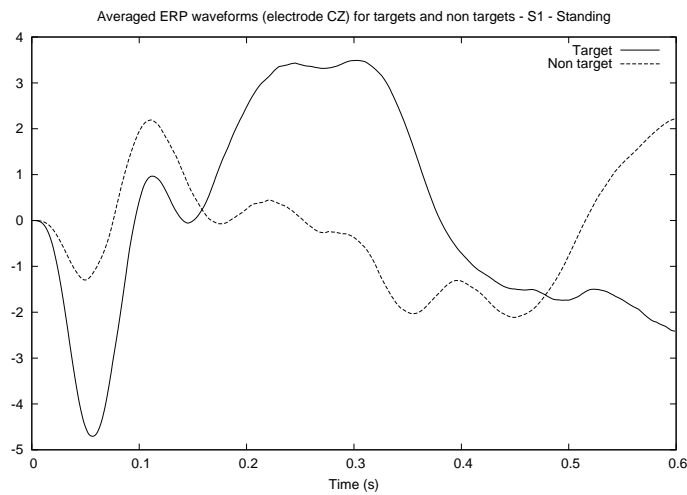

Fig. 7.8 An exemple of an average P300 ERP after a rare and relevant stimulus (Target). We can clearly observe the increase in amplitude about 300ms after the stimulus, as compared to the non-relevant stimulus (Non target).

ERP are characterized by specific temporal variations with respect to the stimulus onset. As such, contrary to BCI based on oscillatory activity, ERP-based BCI exploit mostly a temporal information, but rarely a spectral one. However, as for BCI based on oscillatory activity, ERP-based can also benefit a lot from using the spatial information. Next section illustrates how the spatial and temporal information is used in basic P300-based BCI designs.

7.4.1 Basic signal processing tools for P300-based BCI

In P300-based BCI, the spatial information is typically exploited by focusing mostly on electrodes located over the parietal lobe (i.e., by extracting features only for these electrodes), where the P300 is know to originate. As an example, Krusienski et al recommand to use a set of 8 channels, in positions Fz, Cz, P3, Pz, P4, PO7, Oz, PO8 (see Figure 7.9) (Krusienski et al, 2006).
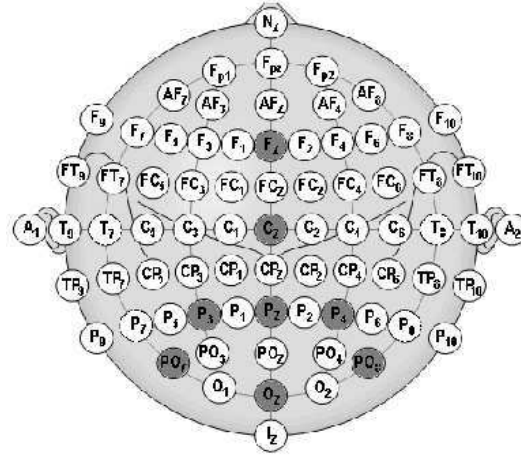

Fig. 7.9 Recommended electrodes for P300-based BCI design, according to (Krusienski et al, 2006).

Once the relevant spatial information identified, here using, for instance, only the electrodes mentioned above, features can be extracted for the signal of each of them. For ERP in general, including the P300, the features generally exploit the temporal information of the signals, i.e., how the amplitude of the EEG signal varies with time. This is typically achieved by using the values of preprocessed EEG time points as features. More precisely, features for ERP are generally extracted by 1) low-pass or band-pass filtering the signals (e.g., in 1-12 Hz for the P300), ERP being generally slow waves, 2) downsampling the filtered signals, in order to reduce the number of EEG time points and thus the dimensionality of the problem and 3) gathering the values of the remaining EEG time points from all considered channels into a feature vector that will be used as input to a classifier. This process is illustrated in Figure 7.10 to extract features from channel Pz for a P300-based BCI experiment.

Once the features extracted, they can be provided to a classifier which will be trained to assigned them to the target class (presence of an ERP) or to the non-target class (absence of an ERP). This is often achieved using classical classifiers such as LDA or SVM (Lotte et al, 2007). More recently, automatically regularized LDA have been increasingly used (Lotte and Guan, 2009) (Blankertz et al, 2010), as well
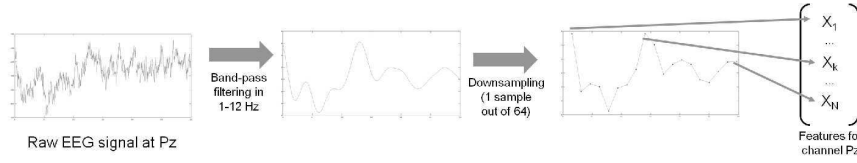
Fig. 7.10 Typical process to extract features from a channel of EEG data for a P300-based BCI design. On this picture we can see the P300 becoming more visible with the different processing steps.

as Bayesian LDA (Hoffmann et al, 2008) (Rivet et al, 2009). Both variants of LDA are specifically designed to be more resistant to the curse-of-dimensionality through the use of automatic regularization. As such, they have proven to be very effective in practice, and superior to classical LDA. Indeed, the number of features is generally higher for ERP-based BCI than for those based on oscillatory activity. Actually, many time points are usually needed to describe ERP but only a few frequency bands (or only one) to describe oscillatory activity. Alternatively, feature selection or channel selection techniques can also be used to deal with this high dimensionality (Lotte et al, 2009a) (Rakotomamonjy and Guigue, 2008) (Krusienski et al, 2006). As for BCI based on oscillatory activity, spatial filters can also prove very useful.

7.4.2 Spatial filters for ERP-based BCI

As mentionned above, with ERP the number of features is usually quite large, with many features per channel and many channels used. The tools described for oscillatory activity-based BCI, i.e., feature selection, channel selection or spatial filtering can be used to deal with that. While feature and channel selection algorithms are the same (these are generic algorithms), spatial filtering algorithms for ERP are different. One may wonder why CSP could not be used for ERP classification. This is due to the fact that a crucial information for classifying ERP is the EEG time course. However, CSP completely ignores this time course as it only considers the average power. Therefore, CSP is not suitable for ERP classification. Fortunately, other spatial filters have been specifically designed for this task.

One useful spatial filter available is the Fisher spatial filter (Hoffmann et al, 2006). This filter uses the Fisher criterion for optimal class separability. Informally, this criterion aims at maximizing the Between class-variance, i.e., the distance between the different classes (we want the feature vectors from the different classes to be as far apart from each other as possible, i.e., as different as possible) while minimizing the within class-variance, i.e., the distance between the feature vectors from the same class (we want the feature vectors from the same class to be as similar as possible). Formally, this means maximizing the following objective function:

$$J_{Fisher} = \frac{tr(S_b)}{tr(S_w)} \tag{7.9}$$

with

$$S_b = \sum_{k=1}^{N_c} p_k(\bar{x}_k - \bar{x})(\bar{x}_k - \bar{x})^T \tag{7.10}$$

and

$$S_w = \sum_{k=1}^{N_c} p_k \sum_{i \in C_k} (x_i - \bar{x}_k)(x_i - \bar{x}_k)^T \tag{7.11}$$

In these equations, S_b is the between-class variance, S_w the within-class variance, N_c is the number of classes, x_i is the ith feature vector, $\bar{v}$ is the average of all vectors v, C_k is the kth class and p_k the probability of class k.

This criterion is widely used in machine learning in general (Duda et al, 2001), and can be used to find spatial filters such that the resulting features maximize this criterion, and thus the discriminability between the classes. This is what the Fisher spatial filter does. It finds the spatial filters such that the spatially filtered EEG time course (i.e., the feature vector) is maximally different between classes, according to the Fisher criterion. This is achieved by replacing x_i (the feature vector) by wX_i (i.e., the spatially filtered signal) in equations 7.10 and 7.11. This gives an objective function of the form $J(w) = \frac{w\hat{S}_b w^T}{w\hat{S}_w w^T}$, which, like the CSP algorithm, can be solved by GEVD. This has been showed to be very efficient in practice (Hoffmann et al, 2006).

Another option, that has also proved very efficient in practice, is the xDAWN spatial filter (Rivet et al, 2009). This spatial filter, also dedicated to ERP classification, uses a different criterion from that of the Fisher spatial filter. xDAWN aims at maximizing the signal to signal plus noise ratio. Informally, this means that xDAWN aims at enhancing the ERP response, at making the ERP more visible in the middle of the noise. Formally, xDAWN finds spatial filters that maximize the following objective function:

$$J_{xDAWN} = \frac{wADD^T A^T w^T}{wXX^T w^T} \tag{7.12}$$

where A is the time course of the ERP response to detect for each channel (estimated from data, usually using a Least Square estimate) and D is a matrix containing the positions of target stimuli that should evoke the ERP. In this equation, the numerator represents the signal, i.e., the relevant information we want to enhance. Indeed, $wADD^T A^T w^T$ is the power of the time course of the ERP responses after spatial filtering. On the contrary, in the denominator, $wXX^T w^T$ is the variance of all EEG signals after spatial filtering. Thus, it contains both the signal (the ERP) plus the noise. Therefore, maximizing J_{xDAWN} actually maximizes the signal, i.e., it enhances the ERP response, and simultaneously minimizes the signal plus the noise, i.e., it makes the noise as small as possible (Rivet et al, 2009). This has indeed been shown to lead to much better ERP classification performance.

In practice, spatial filters have proven to be useful for ERP-based BCI (in particular for P300-based BCI), especially when little training data is available. From a theoretical point of view, this was to be expected. Actually, contrary to CSP and Band Power which extract non-linear features (the power of the signal is a quadratic operation), features for ERP are all linear and linear operations are commutative. Since BCI classifiers, e.g., LDA, are generally also linear, this means that the classifier could theoretically learn the spatial filter as well. Indeed, both linearly combining the original features X for spatial filtering ($F = WX$), then linearly combining the spatially filtered signals for classification ($y = wF = w(WX) = \hat{W}X$) or directly linearly combining the original features for classification ($y = WX$) are overall a simple linear operation. If enough training data is available, the classifier, e.g., LDA, would not need spatial filtering. However, in practice, there is often little training data available, and first performing a spatial filtering eases the subsequent task of the classifier by reducing the dimensionality of the problem. Altogether, this means that with enough training data, spatial filtering for ERP may not be necessary, and leaving the classifier learn everything would be more optimal. Otherwise, if few training data is available, which is often the case in practice, then spatial filtering can benefit a lot to ERP classification (see also (Rivet et al, 2009) for more discussion of this topic).

7.4.3 Summary of signal processing tools for ERP-based BCI

In summarry, when designing ERP-based BCI, it is important to use the temporal information. This is mostly achieved by using the amplitude of preprocessed EEG time points as features, with low-pass or band-pass filtering and downsampling as preprocessing. Feature selection algorithms can also prove useful. It is also important to consider the spatial information. To do so, either using or selecting relevant channels is useful. Using spatial filtering algorithms such as xDAWN or Fisher spatial filters can also prove a very efficient solution, particularly when little training data is available. In the following, we will briefly describe some alternative signal processing tools that are less used but can also prove useful in practice.

7.5 Alternative methods

So far, this chapter has described the main tools used to recognize mental states in EEG-based BCI. They are efficient and usually simple tools that have become part of the standard toolbox of BCI designers. However, there are other signal processing tools, and in particular other kinds of features or information sources that can be exploited to process EEG signals. Without being exhaustive, this section briefly presents some of these tools for interested readers, together with corresponding ref-

erences. The alternative EEG feature representations that can be used include the following 4 categories:

- **Temporal** representations: temporal representations measure how the signal varies with time. Contrary to basic features used for ERP, which simply consist in the EEG time points over time, some measures have been developed in order to characterize and quantify those variations. The corresponding features include Hjorth parameters (Obermeier et al, 2001) or Time Domain Parameters (TDP) (Vidaurre et al, 2009). Recent research results have even suggested that TDP could be more efficient that the gold-standard Band Power features (Vidaurre et al, 2009) (Ofner et al, 2011).
- **Connectivity** measures: they measure how much the signal from two channels are correlated, synchronized or even if one signal may be the cause of the other one. In other words, connectivity features measure how the signal of two channels are related. This is particularly useful for BCI since it is known that, in the brain, there are many long distance communications between separated areas (Varela et al, 2001). As such, connectivity features are increasingly used for BCI and seem to be a very valuable complement to traditional features. Connectivity features include coherence, phase locking-values or Directed Transfer Function (DFT) (Krusienski et al, 2012) (Grosse-Wentrup, 2009) (Gouy-Pailler et al, 2007) (N. Caramia, 2014).
- **Complexity** measures: they naturally measure how complex the EEG signal may be, i.e., they measure its regularity or how predictable it can be. This has also been shown to provide information about the mental state of the user, and also proved to provide complementary information to classical features such as band-power features. The features from this category used in BCI include approximate entropy (Balli and Palaniappan, 2010), predictive complexity (Brodu et al, 2012) or waveform length (Lotte, 2012).
- **Chaos theory**-inspired measures: another category of features that has been explored is chaos-related measures, which assess how chaotic the EEG signal can be, or which chaotic properties it can have. This has also been shown to extract relevant information. Examples of corresponding features include fractal dimension (Boostani and Moradi, 2004) or multi-fractal cumulants (Brodu et al, 2012).

While these various alternative features may not be as efficient as the standards tools such as Band Power features, they usually extract a complementary information. Consequently, using band power features together with some of these alternative features has led to increase classification performances, higher that the performances obtained with any of these features used alone (Dornhege et al, 2004) (Brodu et al, 2012) (Lotte, 2012).

It is also important to realize that while several spatial filters have been designed for BCI, they are optimized for a specific type of feature. For instance, CSP is the optimal spatial filter for Band Power features and xDAWN or Fisher spatial filters are optimal spatial filters for EEG time points features. However, using such spatial filters with other features, e.g., with the alternative features described above, would be clearly suboptimal. Designing and using spatial filters dedicated to these

alternative features is therefore necessary. Results with waveform length features indeed suggested that dedicated spatial filters for each feature significantly improve classification performances (Lotte, 2012).

7.6 Discussion

Many EEG signal processing tools are available in order to classify EEG signals into the corresponding user's mental state. However, EEG signal processing is a very difficult task, due to the noise, non-stationarity, complexity of the signals as well as due to the limited amount of training data available. As such, the existing tools are still not perfect, and many research challenges are still open. In particular, it is necessary to explore and design EEG features that are 1) more *informative*, in order to reach better performances, 2) *robust* to noise and artifacts, in order to use the BCI outside laboratories, potentially with moving users, 3) *invariant*, to deal with non-stationarity and session-to-session transfer and 4) *universal*, in order to design subject-independent BCI, i.e., BCI that can work for any user, without the need for individual calibration. As we have seen, some existing tools can partially address, or at least, mitigate such problems. Nevertheless, there is so far no EEG signal processing tool that has simultaneously all these properties and that is perfectly robust, invariant and universal. Therefore, there are still exciting research works ahead.

7.7 Conclusion

In this chapter, we have provided a tutorial and overview of EEG signal processing tools for users' mental state recognition. We have presented the importance of the feature extraction and classification components. As we have seen, there are 3 main sources of information that can be used to design EEG-based BCI: 1) the spectral information, which is mostly used with band power features; 2) the temporal information, represented as the amplitude of preprocessed EEG time points and 3) the spatial information, which can be exploited by using channel selection and spatial filtering (e.g., CSP or xDAWN). For BCI based on oscillatory activity, the spectral and spatial information are the most useful, while for ERP-based BCI, the temporal and spatial information are the most relevant. We have also briefly explored some alternative sources of information that can also complement the 3 main sources mentioned above.

This chapter aimed at being didactic and easily accessible, in order to help people not already familiar with EEG signal processing to start working in this area or to start designing and using BCI in their own work or activities. Indeed, BCI being such a multidisciplinary topic, it is usually difficult to understand enough of the different scientific domains involved to appropriately use BCI systems. It should also be mentioned that several software tools are now freely available to help users

design BCI systems, e.g., Biosig (Schlögl et al, 2007), BCI2000 (Mellinger and Schalk, 2007) or OpenViBE (Renard et al, 2010). For instance, with OpenViBE, it is possible to design a new and complete BCI system without writing a single line of code. With such tools and this tutorial, we hope to make BCI design and use more accessible, e.g., to design musical BCI.

7.8 Questions

Please find below 10 questions to reflect on this chapter and try to grasp the essential messages:

1. Do we need feature extraction? In particular why not using the raw EEG signals as input to the classifier?
2. What part of the EEG signal processing pipeline can be trained/optimized based on the training data?
3. Can we design a BCI system that would work for all users (a so-called subject-indepedent BCI)? If so, are BCI designed specifically for one subject still relevant?
4. Are univariate and multivariate feature selection methods both suboptimal in general? If so, why using one type or the other?
5. By using an inverse solution with scalp EEG signals, can I always reach a similar information about brain activity as I would get with invasive recordings?
6. What would be a good reason to avoid using spatial filters for BCI?
7. which spatial filter to you have to try when designing an oscillatory activity-based BCI?
8. Let us assume that you want to design an EEG-based BCI, whatever its type: can CSP be always useful to design such a BCI?
9. Among typical features for oscillatory activity-based BCI (i.e., band power features) and ERP-based BCI (i.e., amplitude of the preprocessed EEG time points), which ones are linear and wich ones are not (if applicable)?
10. Let us assume you want to explore a new type of features to classify EEG data: could they benefit from spatial filtering and if so, which one?

References

Ang K, Chin Z, Wang C, Guan C, Zhang H (2012) Filter bank common spatial pattern algorithm on bci competition iv datasets 2a and 2b. Frontiers in Neuroscience 6

Arvaneh M, Guan C, Ang K, Quek H (2011) Optimizing the channel selection and classification accuracy in eeg-based bci. IEEE Transactions on Biomedical Engineering 58:1865–1873

Baillet S, Mosher J, Leahy R (2001) Electromagnetic brain mapping. IEEE Signal Processing Magazine 18(6):14–30

Balli T, Palaniappan R (2010) Classification of biological signals using linear and nonlinear features. Physiological measurement 31(7):903

Bashashati A, Fatourechi M, Ward RK, Birch GE (2007) A survey of signal processing algorithms in brain-computer interfaces based on electrical brain signals. Journal of Neural engineering 4(2):R35–57

Bennett KP, Campbell C (2000) Support vector machines: hype or hallelujah? ACM SIGKDD Explorations Newsletter 2(2):1–13

Besserve M, Martinerie J, Garnero L (2011) Improving quantification of functional networks with eeg inverse problem: Evidence from a decoding point of view. Neuroimage

Blankertz B, Kawanabe M, Tomioka R, Hohlefeld F, Nikulin V, Müller KR (2008a) Invariant common spatial patterns: Alleviating nonstationarities in brain-computer interfacing. In: Advances in Neural Information Processing Systems 20, In . MIT Press, Cambridge, MA

Blankertz B, Tomioka R, Lemm S, Kawanabe M, Müller KR (2008b) Optimizing spatial filters for robust EEG single-trial analysis. IEEE Signal Proc Magazine 25(1):41–56

Blankertz B, Lemm S, Treder M, Haufe S, Müller KR (2010) Single-trial analysis and classification of ERP components a tutorial. Neuroimage

Boostani R, Moradi MH (2004) A new approach in the BCI research based on fractal dimension as feature and adaboost as classifier. Journal of Neural Engineering 1(4):212–217

Brodu N, Lotte F, Lécuyer A (2011) Comparative study of band-power extraction techniques for motor imagery classification. In: Computational Intelligence, Cognitive Algorithms, Mind, and Brain (CCMB), 2011 IEEE Symposium on, IEEE, pp 1–6

Brodu N, Lotte F, Lécuyer A (2012) Exploring two novel features for EEG-based brain-computer interfaces: Multifractal cumulants and predictive complexity. Neurocomputing 79(1):87–94

Browne MW (2000) Cross-validation methods. Journal of Mathematical Psychology 44(1):108–132

Brunner C, Naeem M, Leeb R, Graimann B, Pfurtscheller G (2007) Spatial filtering and selection of optimized components in four class motor imagery eeg data using independent components analysis. Pattern Recognition Letters 28(8):957 – 964, DOI DOI: 10.1016/j.patrec.2007.01.002, URL http://www.sciencedirect.com/science/article/B6V15-4MV74WJ-1/2/525b7adff6f9a8a71984d1a2e083e365

Burges CJC (1998) A tutorial on support vector machines for pattern recognition. Knowledge Discovery and Data Mining 2:121–167

Congedo M, Lotte F, Lécuyer A (2006) Classification of movement intention by spatially filtered electromagnetic inverse solutions. Physics in Medicine and Biology 51(8):1971–1989

Devlaminck D (2011) Optimization of brain-computer interfaces. PhD thesis, University of Ghent

Dornhege G, Blankertz B, Curio G, Müller K (2004) Boosting bit rates in non-invasive EEG single-trial classifications by feature combination and multi-class paradigms. IEEE Transactions on Biomedical Engineering 51(6):993–1002

Dornhege G, Blankertz B, Krauledat M, Losch F, Curio G, Müller KR (2006) Combined optimization of spatial and temporal filters for improving brain-computer interfacing. IEEE Trans Biomed Eng 53(11):2274–2281

Duda RO, Hart PE, Stork DG (2001) Pattern Recognition, second edition. WILEY-INTERSCIENCE

Farquhar J, Hill N, Lal T, Schölkopf B (2006) Regularised CSP for sensor selection in BCI. In: Proceedings of the 3rd international BCI workshop

Fatourechi M, Bashashati A, Ward R, Birch G (2007) EMG and EOG artifacts in brain computer interface systems: A survey. Clinical Neurophysiology 118(3):480–494

Fazel-Rezai R, Allison B, Guger C, Sellers E, Kleih S, Kübler A (2012) P300 brain computer interface: current challenges and emerging trends. Frontiers in Neuroengineering 5(14)

Friedman JHK (1997) On bias, variance, 0/1-loss, and the curse-of-dimensionality. Data Mining and Knowledge Discovery 1(1):55–77

Friedrich E, Scherer R, Neuper C (2012) The effect of distinct mental strategies on classification performance for brain-computer interfaces. International Journal of Psychophysiology (0), DOI 10.1016/j.ijpsycho.2012.01.014, URL http://www.sciencedirect.com/science/article/pii/S0167876012000165

Fukunaga K (1990) Statistical Pattern Recognition, second edition. ACADEMIC PRESS, INC

Garrett D, Peterson DA, Anderson CW, Thaut MH (2003) Comparison of linear, nonlinear, and feature selection methods for EEG signal classification. IEEE Transactions on Neural System and Rehabilitation Engineering 11:141–144

Gouy-Pailler C, Achard S, Rivet B, Jutten C, Maby E, Souloumiac A, Congedo M (2007) Topographical dynamics of brain connections for the design of asynchronous brain-computer interfaces. In: Proc. Int. Conf. IEEE Engineering in Medicine and Biology Society (IEEE EMBC), pp 2520–2523

Grosse-Wentrup M (2009) Understanding brain connectivity patterns during motor imagery for brain-computer interfacing. In: Advances in neural information processing systems (NIPS) 21

Grosse-Wentrup M, Buss M (2008) Multi-class common spatial pattern and information theoretic feature extraction. IEEE Transactions on Biomedical Engineering 55(8):1991–2000

Grosse-Wentrup M, Gramann K, Wascher E, Buss M (2005) EEG source localization for brain-computer-interfaces. In: 2nd International IEEE EMBS Conference on Neural Engineering, pp 128–131

Grosse-Wentrup M, Liefhold C, Gramann K, Buss M (2009) Beamforming in non invasive brain computer interfaces. IEEE Transactions on Biomedical Engineering 56(4):1209 – 1219

Guyon I, Elisseeff A (2003) An introduction to variable and feature selection. Journal of Machine Learning Research 3:1157–1182

Hall M (2000) Correlation-based feature selection for discrete and numeric class machine learning. In: Proc. 17th International Conf. on Machine Learning, pp 359–366

Haufe S, Meinecke F, Görgen K, Dähne S, Haynes JD, Blankertz B, Bießmann F (2014) On the interpretation of weight vectors of linear models in multivariate neuroimaging. NeuroImage 87:96–110

Herman P, Prasad G, McGinnity T, Coyle D (2008) Comparative analysis of spectral approaches to feature extraction for eeg-based motor imagery classification. Neural Systems and Rehabilitation Engineering, IEEE Transactions on 16(4):317–326

Hoffmann U, Vesin J, Ebrahimi T (2006) Spatial filters for the classification of event-related potentials. In: European Symposium on Artificial Neural Networks (ESANN 2006)

Hoffmann U, Vesin JM, Ebrahimi T, Diserens K (2008) An efficient P300-based brain-computer interface for disabled subjects. Journal of Neuroscience Methods 167:115–125

Jain A, Zongker D (1997) Feature selection: evaluation, application, and small sample performance. IEEE Transactions on Pattern Analysis and Machine Intelligence 19(2):153–158

Jain A, Duin R, Mao J (2000) Statistical pattern recognition : A review. IEEE Transactions on Pattern Analysis and Machine Intelligence 22(1):4–37

Kachenoura A, Albera L, Senhadji L, Comon P (2008) ICA: A potential tool for BCI systems. IEEE Signal Processing Magazine 25(1):57–68

Kamousi B, Liu Z, He B (2005) Classification of motor imagery tasks for brain-computer interface applications by means of two equivalent dipoles analysis. IEEE Transactions on Neural Systems and Rehabilitation Engineering 13(2):166–171

Krusienski D, Sellers E, Cabestaing F, Bayoudh S, McFarland D, Vaughan T, Wolpaw J (2006) A comparison of classification techniques for the P300 speller. Journal of Neural Engineering 3:299–305

Krusienski D, McFarland D, Wolpaw J (2012) Value of amplitude, phase, and coherence features for a sensorimotor rhythm-based brain–computer interface. Brain research bulletin 87(1):130–134

Lal T, Schröder M, Hinterberger T, Weston J, Bogdan M, Birbaumer N, Schölkopf B (2004) Support vector channel selection in BCI. IEEE TBME 51(6):10031010

Lan T, Erdogmus D, Adami A, Mathan S, Pavel M (2007) Feature and channel selection for cognitive state estimation using ambulatory EEG. Computational Intelligence and Neuroscience

Ledoit O, Wolf M (2004) A well-conditioned estimator for large-dimensional covariance matrices. Journal of Multivariate Analysis 88(2):365–411

Lemm S, Blankertz B, Curio G, Mller KR (2005) Spatio-spectral filters for improving classification of single trial EEG. IEEE Trans Biomed Eng 52(9):1541–1548

Lotte F (2012) A new feature and associated optimal spatial filter for EEG signal classification: Waveform length. In: International Conference on Pattern Recognition (ICPR), pp 1302–1305

Lotte F, Guan C (2009) An efficient P300-based brain-computer interface with minimal calibration time. In: Assistive Machine Learning for People with Disabilities symposium (NIPS'09 Symposium)

Lotte F, Guan C (2010a) Learning from other subjects helps reducing brain-computer interface calibration time. In: International Conference on Audio, Speech and Signal Processing (ICASSP'2010), pp 614–617

Lotte F, Guan C (2010b) Spatially regularized common spatial patterns for EEG classification. In: International Conference on Pattern Recognition (ICPR)

Lotte F, Guan C (2011) Regularizing common spatial patterns to improve BCI designs: Unified theory and new algorithms. IEEE Transactions on Biomedical Engineering 58(2):355–362

Lotte F, Congedo M, Lécuyer A, Lamarche F, Arnaldi B (2007) A review of classification algorithms for EEG-based brain-computer interfaces. Journal of Neural Engineering 4:R1–R13

Lotte F, Fujisawa J, Touyama H, Ito R, Hirose M, Lécuyer A (2009a) Towards ambulatory brain-computer interfaces: A pilot study with P300 signals. In: 5th Advances in Computer Entertainment Technology Conference (ACE), pp 336–339

Lotte F, Lécuyer A, Arnaldi B (2009b) FuRIA: An inverse solution based feature extraction algorithm using fuzzy set theory for brain-computer interfaces. IEEE transactions on Signal Processing 57(8):3253–3263

Lotte F, Langhenhove AV, Lamarche F, Ernest T, Renard Y, Arnaldi B, Lécuyer A (2010) Exploring large virtual environments by thoughts using a brain-computer interface based on motor imagery and high-level commands. Presence: teleoperators and virtual environments 19(1):54–70

Mason S, Birch G (2003) A general framework for brain-computer interface design. IEEE Transactions on Neural Systems and Rehabilitation Engineering 11(1):70–85

McFarland DJ, Wolpaw JR (2005) Sensorimotor rhythm-based brain-computer interface (BCI): feature selection by regression improves performance. IEEE Transactions on Neural Systems and Rehabilitation Engineering 13(3):372–379

McFarland DJ, McCane LM, David SV, Wolpaw JR (1997) Spatial filter selection for EEG-based communication. Electroencephalographic Clinical Neurophysiology 103(3):386–394

McFarland DJ, Anderson CW, Müller KR, Schlögl A, Krusienski DJ (2006) BCI meeting 2005-workshop on BCI signal processing: feature extraction and translation. IEEE Transactions on Neural Systems and Rehabilitation Engineering 14(2):135 – 138

Mellinger J, Schalk G (2007) Toward Brain-Computer Interfacing, in: g. dornhege, j.r. millán et al. (eds.) edn, MIT Press, chap BCI2000: A General-Purpose Software Platform for BCI Research, pp 372–381. 21

Michel C, Murray M, Lantz G, Gonzalez S, Spinelli L, de Peralta RG (2004) EEG source imaging. Clin Neurophysiol 115(10):2195–2222

Millán J, no JM, Franzé M, Cincotti F, Varsta M, Heikkonen J, Babiloni F (2002) A local neural classifier for the recognition of eeg patterns associated to mental tasks. IEEE transactions on neural networks 13(3):678–686

Miranda E, Magee W, Wilson J, Eaton J, Palaniappan R (2011) Brain-computer music interfacing (bcmi) from basic research to the real world of special needs. Music and Medicine 3(3):134–140

N Caramia SR F Lotte (2014) Optimizing spatial filter pairs for eeg classification based on phase synchronization. In: International Conference on Audio, Speech and Signal Processing (ICASSP'2014)

Noirhomme Q, Kitney R, Macq B (2008) Single trial EEG source reconstruction for brain-computer interface. IEEE Transactions on Biomedical Engineering 55(5):1592–1601

Obermeier B, Guger C, Neuper C, Pfurtscheller G (2001) Hidden markov models for online classification of single trial EEG. Pattern recognition letters pp 1299–1309

Ofner P, Muller-Putz G, Neuper C, Brunner C (2011) Comparison of feature extraction methods for brain-computer interfaces. In: International BCI Conference 2011

Peng H, Long F, Ding C (2005) Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. IEEE Transactions on Pattern Analysis and Machine Intelligence 27(8):1226–1238

Penny W, Roberts S, Curran E, Stokes M (2000) EEG-based communication: a pattern recognition approach. IEEE Transactions on Rehabilitation Engineering 8(2):214–215

Pfurtscheller G, Neuper C (2001) Motor imagery and direct brain-computer communication. proceedings of the IEEE 89(7):1123–1134

Pfurtscheller G, da Silva FHL (1999) Event-related EEG/MEG synchronization and desynchronization: basic principles. Clinical Neurophysiology 110(11):1842–1857

Pudil P, Ferri FJ, Kittler J (1994) Floating search methods for feature selection with nonmonotonic criterion functions. Pattern Recognition 2:279–283

Qin L, Ding L, He B (2004) Motor imagery classification by means of source analysis for brain computer interface applications. Journal of Neural Engineering 1(3):135–141

Rakotomamonjy A, Guigue V (2008) BCI competition III: Dataset II - ensemble of SVMs for BCI P300 speller. IEEE Trans Biomedical Engineering 55(3):1147–1154

Ramoser H, Muller-Gerking J, Pfurtscheller G (2000) Optimal spatial filtering of single trial EEG during imagined hand movement. IEEE Transactions on Rehabilitation Engineering 8(4):441–446

Raudys SJ, Jain AK (1991) Small sample size effects in statistical pattern recognition: Recommendations for practitioners. IEEE Transactions on Pattern Analysis and Machine Intelligence 13(3):252–264

Renard Y, Lotte F, Gibert G, Congedo M, Maby E, Delannoy V, Bertrand O, Lécuyer A (2010) OpenViBE: An open-source software platform to design, test and use

brain-computer interfaces in real and virtual environments. Presence: teleoperators and virtual environments 19(1):35–53

Reuderink B, Poel M (2008) Robustness of the common spatial patterns algorithm in the BCI-pipeline. Tech. rep., HMI, University of Twente

Rivet B, Souloumiac A, Attina V, Gibert G (2009) xdawn algorithm to enhance evoked potentials: Application to brain computer interface. IEEE Transactions on Biomedical Engineering 56(8):2035–2043

Samek W, Vidaurre C, Müller KR, Kawanabe M (2012) Stationary common spatial patterns for braincomputer interfacing. Journal of Neural Engineering 9(2)

Sannelli C, Dickhaus T, Halder S, Hammer E, Müller KR, Blankertz B (2010) On optimal channel configurations for SMR-based brain-computer interfaces. Brain Topography

Schlögl A, Brunner C, Scherer R, Glatz A (2007) Towards Brain-Computer Interfacing, g. dornhege, j.r. millán, t. hinterberger, d.j. mcfarland, k.-r. müller edn, MIT press, chap BioSig - an open source software library for BCI research, pp 347–358. 20

Schröder M, Lal T, Hinterberger T, Bogdan M, Hill N, Birbaumer N, Rosenstiel W, Schölkopf B (2005) Robust EEG channel selection across subjects for brain-computer interfaces. EURASIP J Appl Signal Process pp 3103–3112

Tangermann M, Winkler I, Haufe S, Blankertz B (2009) Classification of artifactual ICA components. Int J Bioelectromagnetism 11(2):110–114

Thomas K, Guan C, Chiew T, Prasad V, Ang K (2009) New discriminative common spatial pattern method for motor imagery brain computer interfaces. IEEE Transactions on Biomedical Engineering 56(11)

Tomioka R, Dornhege G, Aihara K, Mller KR (2006) An iterative algorithm for spatio-temporal filter optimization. In: Proceedings of the 3rd International Brain-Computer Interface Workshop and Training Course 2006, pp 22–23

Varela F, Lachaux J, Rodriguez E, Martinerie J (2001) The brainweb: phase synchronization and large-scale integration. Nature reviews neuroscience 2(4):229–239

Vialatte F, Maurice M, Dauwels J, Cichocki A (2010) Steady-state visually evoked potentials: Focus on essential paradigms and future perspectives. Progress in Neurobiology, 90:418–438

Vidaurre C, Krämer N, Blankertz B, Schlögl A (2009) Time domain parameters as a feature for EEG-based brain computer interfaces. Neural Networks 22:1313–1319

Xu N, Gao X, Hong B, Miao X, Gao S, Yang F (2004) BCI competition 2003–data set IIb: enhancing P300 wave detection using ICA-based subspace projections for BCI applications. IEEE Transactions on Biomedical Engineering 51(6):1067–1072

Zhong M, Lotte F, Girolami M, Lécuyer A (2008) Classifying EEG for brain computer interfaces using gaussian processes. Pattern Recognition Letters 29:354–359