

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/5280002>

# Classification methods for ongoing EEG and MEG signals

Article in *Biological Research* · February 2007

DOI: 10.4067/S0716-97602007000500005 · Source: PubMed

CITATIONS

91

READS

3,487

6 authors, including:



**Michel Besserve**

Max Planck Institute for Intelligent Systems

68 PUBLICATIONS 1,080 CITATIONS

[SEE PROFILE](#)



**Karim Jerbi**

Université de Montréal

150 PUBLICATIONS 5,134 CITATIONS

[SEE PROFILE](#)



**Sylvain Baillet**

McGill University

287 PUBLICATIONS 13,532 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



software [View project](#)



Brainstorm [View project](#)

# Classification methods for ongoing EEG and MEG signals

MICHEL BESSERVE<sup>1,2</sup>, KARIM JERBI\*, FRANCOIS LAURENT<sup>1,2</sup>,  
SYLVAIN BAILLET<sup>1,2</sup>, JACQUES MARTINERIE<sup>1,2</sup> and LINE GARNERO<sup>1,2</sup>

<sup>1</sup> CNRS, UPR 640-LENA, Laboratoire Neurosciences Cognitives et Imagerie Cérébrale, 75013 Paris Cedex 13, France

<sup>2</sup> UPMC Univ Paris 06, F-75005, Paris, France

\* Current affiliation: LPPA, CNRS-URM7152, Collège de France, Paris, & Brain Dynamics and Cognition, INSERM U821, Lyon, France

## ABSTRACT

Classification algorithms help predict the qualitative properties of a subject's mental state by extracting useful information from the highly multivariate non-invasive recordings of his brain activity. In particular, applying them to Magneto-encephalography (MEG) and electro-encephalography (EEG) is a challenging and promising task with prominent practical applications to e.g. Brain Computer Interface (BCI). In this paper, we first review the principles of the major classification techniques and discuss their application to MEG and EEG data classification. Next, we investigate the behavior of classification methods using real data recorded during a MEG visuomotor experiment. In particular, we study the influence of the classification algorithm, of the quantitative functional variables used in this classifier, and of the validation method. In addition, our findings suggest that by investigating the distribution of classifier coefficients, it is possible to infer knowledge and construct functional interpretations of the underlying neural mechanisms of the performed tasks. Finally, the promising results reported here (up to 97% classification accuracy on 1-second time windows) reflect the considerable potential of MEG for the continuous classification of mental states.

**Key terms:** brain computer interface, electroencephalography, magnetoencephalography, visuomotor control, Support Vector Machine.

## 1. INTRODUCTION

The non-invasive detection of task-related neurophysiological changes occurring within the human brain is a significant challenge in biomedical engineering. There is a growing interest in using classification techniques to estimate the mental state of a subject, related to a performed task, from multivariate brain functional imaging signals such as functional magnetic resonance imaging (fMRI) (Carlson et al., 2003; Laconte et al., 2006; Haynes and Rees, 2006) or magneto- or electroencephalography (MEG/EEG) (Kubler et al., 2001; Wolpaw et al., 2002; Lal et al., 2005; Lotte et al., 2007) from ongoing, single-trial epochs. In principle, the related algorithms are statistical tools that

can be trained to estimate a qualitative variable, the *class label*, from a set of quantitative variables. Therefore, classification techniques aim at predicting the qualitative mental state of the human brain and come in addition to the estimation of functional brain responses per se.

There are two major reasons why these methods should be considered: first, classification tools let us evaluate the predictive power of functional signals and thus quantify the information they convey about the tasks being performed. Classification methods come in complement to massively univariate statistical tools like Statistical Parametric Mapping (Frackowiak et al., 1997) by mining the information contents of

multidimensional signals. These approaches are more powerful than the univariate statistical tests since the former seek information in many variables at the same time whereas the latter can only process each variable independently (Carlson et al., 2003). The second reason why classification tools are pertinent to neuroscience is their potential applications to real-time prediction of brain states. Their most popular application is undoubtedly Brain Computer Interfaces (BCI) which has witnessed considerable progress during the past decade (Kubler et al., 2001; Wolpaw et al. 2002; Lebdev and Nicolelis, 2006). A further promising real-time application is *neurofeedback* for the treatment of motor or cognitive dysfunctions such as Attention- Deficit/ Hyperactivity Disorders (ADHD) (Fuchs 2003, Strehl 2006).

For some of these applications, exploiting directly the invasive recordings of neural populations can be more efficient than using fMRI, EEG or MEG. In particular, invasive neural interfaces have known considerable progress (Taylor et al., 2002; Wessberg and Nicolelis, 2004); but a lot of work remains to be done to achieve a safe and stable invasive BCI (Lebdev and Nicolelis, 2006). Thus, even if they provide a slower information transfer, devices using non-invasive imaging modalities are still competitive for this application. Among these imaging techniques, Magnetoencephalography (MEG) and electroencephalography (EEG) provide the highest temporal resolution making these techniques ideal for the aforementioned real-time applications. Several parameters, however, need to be carefully taken into account when using classification methods with magneto-electroencephalographic (MEEG) data. First, the choice and design of the experimental paradigm is an important factor that will partially determine subsequent data processing procedures. The most commonly used paradigms involve stimulus-locked subject responses where the spatial and temporal properties of brain responses linked to a stimulus are investigated and referenced in direct

relationship to the stimulus onset. In the context of BCI, a synchronous set-up is directed toward evoked brain responses triggered by an external stimulus. By contrast, asynchronous BCI set-ups do not require any external stimulation to infer mental states. Therefore, asynchronous approaches are challenging because they are expected to yield a less tedious and more natural and efficient communication device. In general, such systems operate with data acquired while subjects alternate between two (or more) sustained mental tasks. A major difference between continuous data acquisition paradigms and classical stimulus-locked experiments is that the task is maintained on a longer time span (typically tens of seconds). Henceforth, we will refer to this type of paradigm as a *continuous* experiment.

The two types of experimental paradigms, stimulus-locked and continuous, involve different types of data processing and investigate different quantitative measures estimated from the signals. Nevertheless, it is noteworthy that while some techniques like the analysis of evoked potentials are only useful for synchronous BCI applications (Farwell and Donchin, 1988; Wolpaw et al., 2002), other task-related phenomena, like sustained oscillatory activities, can be used both with synchronous and asynchronous paradigms (Keirn et al., 1990; Anderson et al., 1998; Pfurtscheller et al. 1997; Borisoff et al., 2004; Milan et al., 2004; Scherer et al., 2004). To elucidate these task-related brain rhythms, spectral analysis has become a standard procedure with two measurements: (a) power spectral density, (which provides an estimation of signal amplitude at different frequencies or within various frequency bands) and (b) long- range signal correlation in the frequency domain (which can be measured via coherence or phase synchrony). Whether they are performed at the sensor or at the cortical level, power estimations are assumed to reflect the synchronization of local neural populations (Singer et al. 1999) while the long-range signal coupling represents long-distance interactions between two signals recorded in distinct brain regions.

The primary objective of this paper is to provide an overview of the major principles of classification approaches applied to EEG and MEG data, and secondly to investigate the practical behavior of such techniques with real MEG data acquired during a continuous visuomotor task. The results shed some light on important issues such as the selection of a classifier and the evaluation of its performances, as well as the identification of physiologically relevant discrimination features for continuous sensorimotor MEEG data.

The layout of the paper is as follows: In Section 2, we review the general principles of classification methods by considering different classification schemes and various measures generally extracted from MEEG recordings for the purpose of classification. In Section 3, we report on the application of classification tools to MEG data recorded during a continuous visuomotor task and we conclude by discussing the utility, interpretation and limitations of the results obtained.

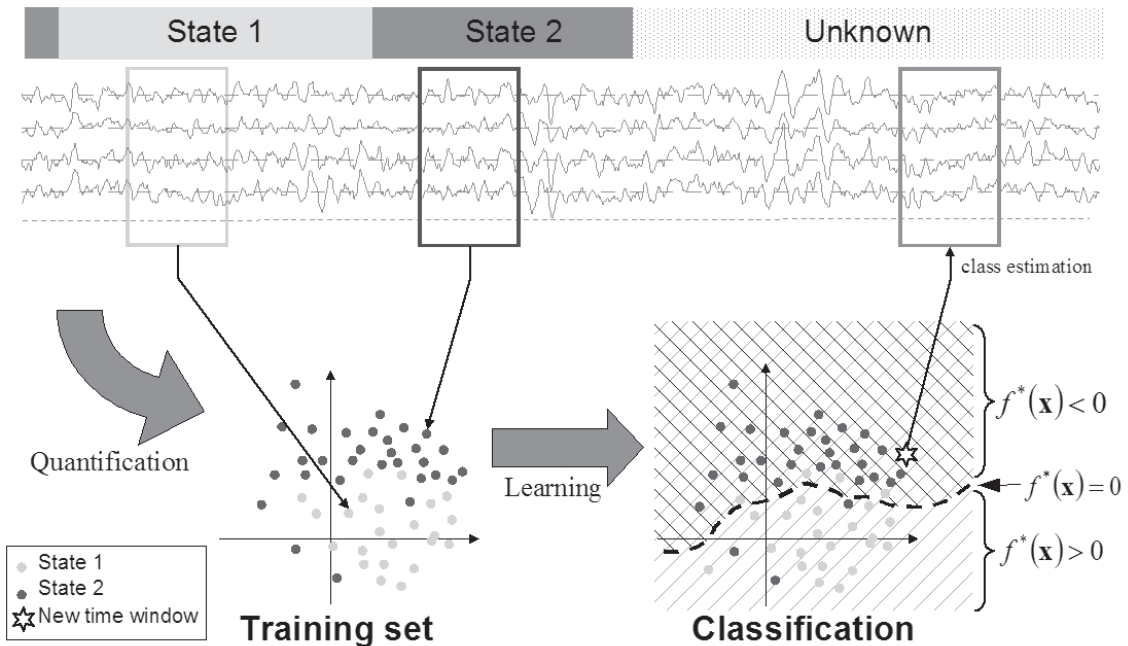
## 2. USING CLASSIFICATION METHODS FOR MEEG

### 2.1 *The different steps*

Applying statistical analysis tools to MEEG signals usually requires temporal segmentation of the recordings according to precise events of the experimental paradigm. For stimulus-locked paradigms, segmentation bounds are set according to the stimulus onset. In particular, a reference time window preceding the stimulus onset is usually taken as a reference epoch (or baseline) to which the temporal evolution of the cortical signals after stimulus onset is compared. Obviously, this scheme can be temporally reversed if the reference point is the subject's response instead of a stimulus (as in movement preparation or anticipation paradigms). Conversely, if continuous paradigms may contain specific events (for example the beginning of the task), they do not systematically impose a temporal ordering of the whole recording. In this

case, if we assume that signal properties are largely invariant over the duration of a given continuous behavioral condition, a long-lasting epoch corresponding to the continuous task can be segmented into successive time windows. This segmentation provides a large number of time windows which can then be processed as multiple trials of the same experimental condition (although all reference-free).

To characterize brain activities from recordings, several features are then computed from the segmented data (details in next section). These features allow representing each segment as a point in a normed vector space, whose entries are feature values. Therefore, by applying this quantification procedure to multiple trials, we obtain a set of distributed points that may form several 'clouds' in the multi-dimensional feature space (Fig. 1). Each cloud of points corresponding to the data of a specific behavioral condition represents a *class*. The next step of the classification procedure consists in exploiting a first dataset as a «training set» to learn to partition the feature space in a way that optimally distinguishes the various classes from each other. Most of the widely used techniques (described below) fit a separation surface, called the *decision boundary*, between domains corresponding to each class. Once the classifier is "trained", new trials can then be classified as belonging to one class or another according to their position in the feature space. In this study, we will restrict the discussion to *binary classification* where only two classes have to be distinguished. In this case, the decision boundary is usually determined using a *discriminant function* (Duda and Hart, 2000)  $f^*(x)$  with the following rule: if  $f^*(x) > 0$  then the trial associated with feature vector  $x$  is classified in class 1, else the trial is attributed to class 2. We will briefly present a few classification methods in addition to common problems that arise from these approaches. A schematic overview of the MEEG classification process in the binary classification case of a continuous paradigm is depicted in Figure 1.



**Figure 1:** Principle of a binary classification of MEEG recordings for a continuous experiment. First, signals are segmented in successive time windows, then features are computed representing each time window as a point in a multidimensional space. Then, the classification algorithm separates the features' space in two domains according to the sign of a discriminant function  $f^*$  of a time window according to its feature vector.  $f^*(x)$ , enabling to predict the class

## 2.2. Characterizing brain activity from MEEG measurements

To quantify task-related cortical activations from recordings, the ongoing signals are usually segmented into time windows on which various computations are performed. Most of the estimated parameters can be divided into families depending on the spatial extent of the physiological phenomenon under investigation: local or long range. Local measurements generally provide a measure of task-related activity picked up at a single sensor or electrode. This measure is assumed to reflect the modulation of neural activity at a focal brain area located next to the detector. By contrast, measurements of long distance interactions quantify the coupling between signals detected at two distinct sensors, possibly, although not necessarily, revealing an information transfer between two distant neural ensembles. The physiological interpretation of such long-range phenomena

is the focus of a large body of literature ranging from realistic models of neural networks to experimental evidence obtained using both invasive and non-invasive recordings, both in human and non-human primates (Rodriguez et al., 1999; Varela et al., 2001; Brovelli et al., 2004.; Jerbi et al., 2007; Lachaux et al., 1999).

Given that the electromagnetic activity recorded on the scalp is an indirect and attenuated signature of the underlying neural processes, the main difference between MEEG and invasive recording techniques, is that the tasks-related effects should be sufficiently strong if they are to be observed non-invasively on the scalp surface. The strength of these signals presumably depends on various parameters including the Signal-to-Noise ratio of the system, the size of the neural populations involved, the experimental paradigm and possibly the alertness or motivation of the subject. To help detecting these signals, a large number of signal processing techniques have been

exploited in non-invasive BCI systems (Bashashati et al., 2007); but most of them make use of the same physiological phenomena and similar measures to those that will be reviewed in the following sections.

### *Local activity measurements*

Given that stimulus locked experimental paradigms have been widely used both for basic cognitive studies and for BCI applications, a large range of analysis tools have been tested and developed over the years to specifically analyze event related activities. The best-studied family of event related activities is Event Related Potentials (ERP). Even if, to date, the physiological processes underlying ERP generation remain debated, many types of ERP induced by specific stimulation set-ups are robustly observed and can be exploited. For instance, a virtual keyboard, called the *P300 speller*, has been developed based on a specific ERP arising after the onset of an attended stimulus (Barett, 1996). Due to baseline fluctuations, ERP can only be detected by averaging data obtained over multiple trials. Although, in theory, this requirement makes ERP unusable for classification of single trial MEEG activities, this problem has been solved in the *P300 speller* paradigm by cumulating multiple trials before classification (Farwell and Donchin, 1988).

A further quantification of local modulations is given by measures known as Event Related Desynchronization (ERD) and Synchronization (ERS) (Pfurtscheller and Lopes da Silva, 1999). ERD (or ERS) represent a decrease (or an increase) of oscillatory power at specific frequencies and in specific locations as compared to the baseline power estimated during a reference period (i.e. a time window prior to stimulus onset or alternatively an epoch during a control condition). Many studies report the presence of task specific rhythms in MEG and EEG recordings; In particular, in the 8-12Hz band occipital alpha rhythms are implied in vision whereas mu rhythms (~11 Hz) are present in the motor cortex in resting states and are suppressed during the execution of a

movement (Salmelin and Hari, 1994). In higher frequency bands, beta rhythms (15-30 Hz) have been shown to be suppressed during movement and to display a strong increase roughly 1-2 seconds after the movement is terminated (a phenomenon known as *beta rebound*). Finally, gamma band power modulations (30-90 Hz) seem to play an important role in wide range of complex cognitive processes (Tallon-Baudry, 1999). Historically, the presence of ERS and ERD in specific frequency bands is interpreted as an activation or deactivation of the neural assemblies, but this theory is questioned according to recent studies: the ERD/ERS phenomena are highly dependent of the task and the brain area (Lopes Da Silva, 2006) and rhythms in the same frequency band can both correspond to deactivation or activation of the underlying network (Pfurtscheller, 2006). ERD in the alpha and beta bands appearing in the motor areas has been extensively used to classify motor imagery tasks (Pfurtscheller, 1997). The usual way to detect ERD/ERS is to make a baseline normalization and then to compute inter trial averages (Grimm and Pfurtscheller, 2006); however studies in BCI protocol have demonstrated their measurability in single trial (Pfurtscheller, 97).

For continuous paradigms, the starting time of the mental process is either not used or not available, which makes ERP too difficult to detect. Moreover the transient nature of ERP is not compatible with the characterization of continuous states. Besides, even though ERD/ERS is usually referenced to a baseline, this phenomenon can be equally considered as induced by an event or as task related activation. Hence a continuous task can also be characterized by ERD/ERS measurements. Technically speaking, the alternative to stimulus-locked ERD/ERS when processing continuous brain activities is to cut original recordings into sliding windows and perform the spectral power measurements on these windows in multiple frequency bands.

The usual tools to extract frequency information from EEG and MEG signals are Fourier transform and band pass filtering. In this case, the feature considered is the spectral power in different frequency bands



$f$  obtained by calculating the average power of the filtered signal  $z_i^f(t)$  of channel  $i$  during each time window of length  $T$ :

$$P_i(f) = \frac{1}{T} \sum_{t=1}^T |z_i^f(t)|^2 \quad (1)$$

These tools can be seen as non parametric spectral analysis methods, as no assumption is made on the power spectrum of the signals. Another method, which has been proved to be efficient for EEG classification, is autoregressive modeling (Anderson, 1998). This is a parametric approach as it assumes that the power spectrum of the signals is described by a few coefficients (typically 3 to 10).

Spectral measurements are also a convenient tool to detect Steady State Visual Evoked Potentials (SSVEP) induced by repetitive visual stimulations in new BCI paradigms (Middendorf et al. 2000).

#### *Long distance interactions*

It is reasonable to assume that any mechanism for brain integration must involve interactions between the participating local neural ensembles (Varela et al., 2001). Quantifying these interactions has been investigated via multiple techniques that quantify the presence of common information in the signals that originate from different brain areas. Assessing this functional relationship between distant signals has been named “functional connectivity” by contrast to anatomical connectivity. Functional connectivity investigates the structure of the brain network at a particular time sample and is supposed to characterize the cooperation between many specialized regions. The most widely used measures for long- range interaction quantification are coherence and phase synchrony (Lachaux et al. 1999).

Coherence measures the interaction of two signals in a particular frequency band by computing the correlation between their Fourier transforms  $Z_1(f)$  and  $Z_2(f)$  of two signals at this frequency (bars in the following equations indicate averaging over successive data segments).

$$Coh_{12}(f) = \frac{Z_1(f)Z_2(f)}{|Z_1(f)|^2 |Z_2(f)|^2}$$

Thus, the resulting quantity can both reflect the correlation between spectral power of the signal (i.e. amplitude) and a temporal synchronization (phase relationship) of the signals.

By contrast, phase synchrony concentrates on assessing the common phase information between two signals. It is calculated by converting the filtered data of each channel  $i$  in the frequency band  $f$ ,  $z_i^f(t)$  into an analytical signal  $a_i^f(t)$  by the Hilbert transform  $H$  with  $a_i^f(t) = z_i^f(t) + H[z_i^f(t)]$ . This allows to define the instantaneous phase  $\varphi_i^f(t)$  of the filtered signal (Pikovsky et al., 2001) by the modulus-argument decomposition  $a_i^f(t) = |a_i^f(t)|e^{i\varphi_i^f(t)}$ , where  $i$  is the complex number of modulus 1 and argument  $\frac{\pi}{2}$ . The phase locking value between channels 1 and 2 in a given frequency band  $f$  is thus:

$$S_{12}(f) = \frac{1}{T} \left| \sum_{t=1}^T e^{i(\varphi_2^f(t) - \varphi_1^f(t))} \right|, \quad (2)$$

In a previous study (Rodriguez et al., 1999), we estimated phase synchrony on surface EEG recordings performed during a visual task. Task-related phase synchrony modulations between widely separated electrodes showed significantly different patterns between “face perception” and “no face perception”. Additionally, in an intracerebral study (Le Van Quyen et al., 2005) we provided evidence for the use of phase synchrony to anticipate epileptic seizures. Moreover, recent findings in the field of Brain-Computer Interfaces suggest that combining spectral power measurements with phase synchrony improves accuracy when classifying EEG signals (Gysels et al., 2004, 2005; Wei et al., 2007).

In the present study, we also investigate the combination of power and phase synchrony measures for classification purposes applied to MEG visuomotor data (details in Section 3). An argument for using both these tools is that they can be considered within the same biological framework of “Resonant Cell Assemblies”

(Varela, 1995). In recent years, it has become evident that the concept of neuronal synchrony can improve our characterization and our understanding of the global aspect of brain dynamics. Indeed neuroscience has provided abundant evidence of synchronization at all levels of the nervous system (Varela et al. 2001), ranging from individual pairs of neurons to larger scales within or in between different local neural assemblies. Generally, oscillations are thought to reflect local synchronization of such assemblies, the underlying principle being that it is the phase locked neural activity that predominantly gives rise to measurable scalp oscillations. By contrast, the long range phase synchronization between two electrodes is assumed to be a signature of the phase locked activity between distinct brain regions. Therefore these two measures are complementary in characterizing brain activity and essential for the global integration hypothesis.

### 2.3. Classification algorithms

Data classification is an issue of primary importance in the field of data mining: it consists in estimating a qualitative variable the *class label* using a set of other variables. Numerous algorithms and methods have been proposed in order to achieve data classification and to improve its efficiency. Most standard classification methods can be broadly described as a two-step procedure which consists first of a learning phase followed by actual estimation of unknown class labels. During the learning phase, a discriminant function is fitted to a portion of the data generally called the training data set and then, in the second phase, the trained model (achieving optimized separation on the training set) is used to discriminate between the classes from new data sets. Two main considerations can distinguish classification algorithms:

- Their decision boundary can be either linear or non-linear
- The fitted discriminant function can result from a probabilistic model (model-based or generative algorithms)

or be expressed directly using the data points from the training set (data-based or discriminative algorithms).

We will present three algorithms illustrating these fundamental characteristics.

#### *Linear Discriminant Analysis (LDA)*

Linear classification algorithms, which represent a large portion of the available techniques, are based on fitting a linear discriminant function to the data. This linear decision is of the form  $f(x) = wx + b$  where  $b$  is the bias and  $w$  the normal vector to the decision boundary  $f(x) = 0$ . One approach to fit this linear function to the data is Linear Discriminant Analysis which amounts to fitting a Gaussian probabilistic model to the data (Duda and Hart, 2001). In its simplest formulation, assuming the feature vector  $x$  has a Gaussian distribution with different class conditional means  $\mu_1$  and  $\mu_2$  for class 1 and 2 respectively and the same covariance matrix  $\Sigma_1 = \Sigma_2 = \Sigma$  in the two cases, the optimal decision function in a Bayesian framework is of the form:

$$f_{LDA}^*(x) = (\mu_1 - \mu_2)^T \Sigma^{-1} \left( x - \frac{\mu_1 + \mu_2}{2} \right) = w_{LDA}x + b_{LDA}$$

An illustration of this technique on toy data is provided Fig. 2 a), showing the affine decision boundary between the two clouds of points of the training set. LDA is an interesting algorithm for many reasons: it is fast, simple to implement and to understand. As a consequence it has been widely used in the BCI community (Garett 2003, Bostanov 2004, Scherer 2004).

#### *Linear Support Vector Machine (SVM)*

Contrary to LDA, that constructs a probabilistic model for each class using all data points, the underlying principle of linear Support Vector Machines (SVM) is to minimize a linear separation error focusing on neighboring points of the linear separation surface (Vapnik, 1998). Points that lie sufficiently far from the separation surface are thus ignored in the learning

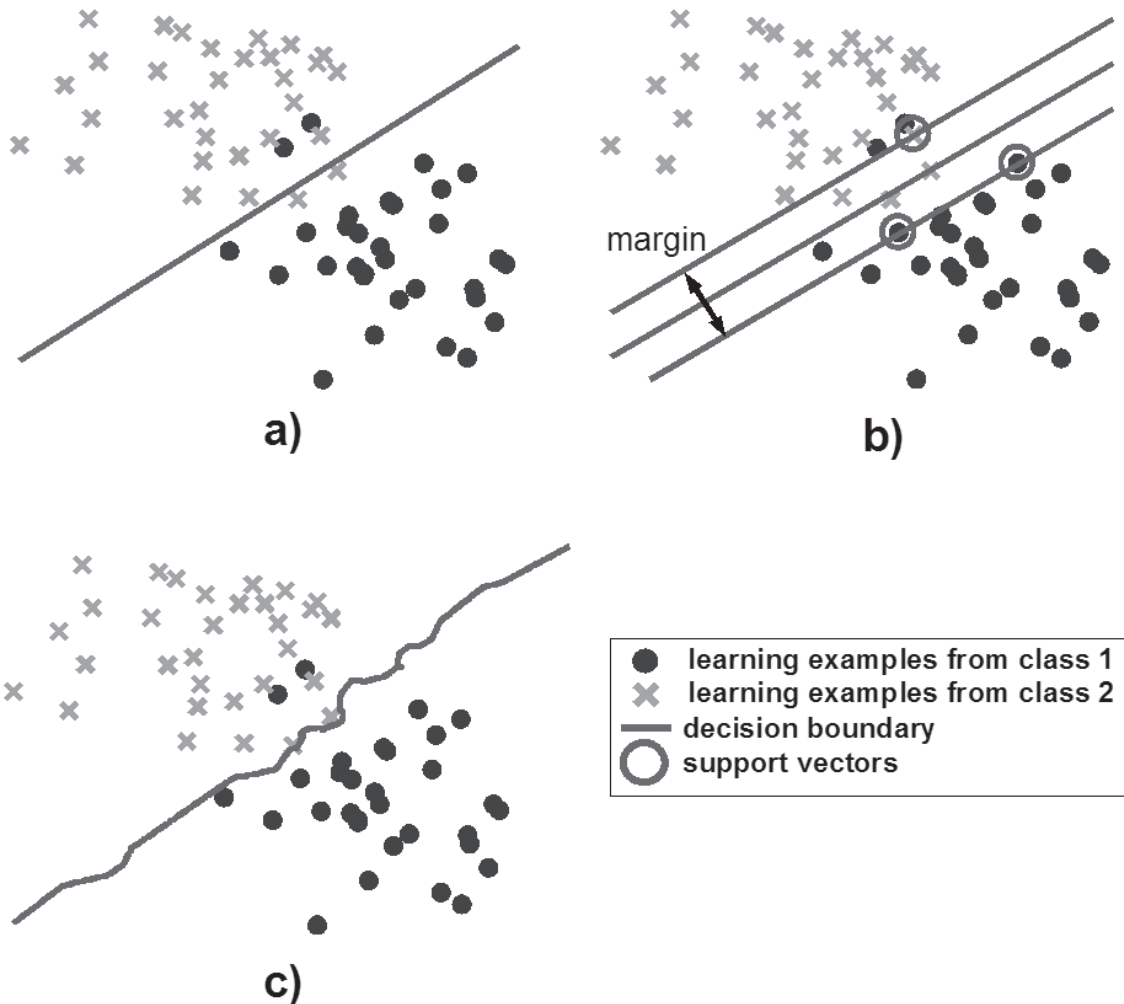


process. More precisely, the linear discriminant function has the same form as in LDA:  $f(x) = wx + b$ ; moreover, we denote by  $(x_i, y_i)$  the  $i$ -th data point of the training set of feature vector  $x$  and of class label  $y_i$  ( $y_i = +1$  if the point belongs to state or class 1,  $y_i = -1$  if the point belongs to class 2). A margin is defined as the domain between two hyperplanes around the

decision boundary of equations  $f(x) = +1$  and  $f(x) = -1$ . The SVM algorithm tends to maximize the width of that margin (of value  $\frac{2}{\|w\|}$ ), which amounts to minimizing the

Euclidian norm of the vector  $w$ , while maintaining most of the data points outside the margin. The optimal discriminant function  $f$  of:

$$(w^*, b^*) = \arg \min_{w, b} C \sum_i \varepsilon_i + \frac{\|w\|^2}{2} \text{ subjected to } \begin{cases} y_i (< w, x_i > +b) \geq 1 - \varepsilon_i, \forall i \\ \varepsilon_i \geq 0, \forall i \end{cases}$$



**Figure 2:** Examples of decision boundaries obtained with the same training set and different classification algorithms: a) LDA classifier, b) SVM classifier, c) classifier with five nearest neighbors using the Euclidian distance.

where  $\sum_i \varepsilon_i$  is the separation error term,  $\frac{\|w\|^2}{2}$

is the regularization term and  $C$  is a user-defined regularization parameter (Muller et al., 2001). As represented Fig. 2 b) some points close to the surface are called support vectors and define the boundaries of the margin on both sides of the surface.

### *K-Nearest Neighbor*

A further data driven method is the k-nearest-neighbors (KNN) algorithm (Hastie et al. 2001). The principle of this third method is as follows: The class label of an unlabelled point is attributed to the predominant class within the  $k$  nearest labeled points belonging to the training set ( $k$  is defined by the experimenter). The KNN algorithm does not rely on any model but rather on the metric used to assess the similarity of any pair of points (usually the Euclidian distance). Using the aforementioned notation  $(x_i, y_i)$  for the points in the learning set the discriminant function associated to KNN is of the form:

$$f_{KNN}^*(x) = \sum_{n \in N_K(x)} y_n$$

where  $N_K(x)$  is the index of the  $K$  nearest neighbors of  $x$  in the training set. From an algorithmic point of view, KNN does not require the iterative learning phase since there is no need to fit a model to the data; however the whole training data set has to be kept in memory to classify the new data. Compared to linear methods the decision boundary  $f_{KNN}(x) = 0$  is strongly nonlinear as shown on an example Fig. 2 c).

Of course, there are a wide variety of other existing techniques that have been used, like neural networks (Hiraiwa et al., 1990, Anderson et al., 1998, Haselsteiner et al., 2000) or Hidden Markov Models (Obermeier et al., 2001). A complete review of the existing techniques exceeds the focus of this paper and the reader can refer to (Duda and Hart, 2001) or (Hastie et al., 2001) for a general view of the field of Pattern Recognition and (Lotte et al., 2007) for an exhaustive review of the algorithms already used for EEG based BCI.

## *2.4. Validating classifier performance*

### *Measuring classification accuracy*

Training a classifier aims at minimizing classification error generally quantified as the ratio of the number of well-classified samples to the total number of samples. In order to interpret it, one should take the chance level as a reference. When trying to classify one sample by randomly distinguishing between two balanced classes, one may expect a mean accuracy of 50%. By analogy, dealing with a three class problem, chance should not go past about 33% asymptotically. However, given that such a level of accuracy might well be considered as insufficient, other criteria have been introduced. One criterion, known as kappa, represents the rate of well-classified examples after subtraction of the asymptotic chance accuracy (Townsend et al., 2006). A further measure of classifier performance is Area Under the ROC curve (AUC). The receiver operating characteristic (ROC) curve represents the evolution of the false positive rate versus the true positive rate which results from thresholding a discriminant function as a function of threshold values (Green and Swets., 1974, Duda and Hart, 2001). A value of the discriminant function above the threshold predicts class 1, whereas a discriminant function below the threshold yields a class 2 prediction. In practice, this measure allows for a control of the acceptable false negative rate of a diagnosis. In summary, the area under the ROC curve is a global measure of the discriminant power of the discriminant function, regardless of the threshold to be chosen and if the area under this curve reaches 1 (i.e. AUC=1) the discrimination is considered to be perfect.

Finally, for BCI applications, it is obvious that the performance of a classifier also depends on the time-length of the data to be classified and the number of classes to be recognized. For example, taking two classifiers with the same accuracies, if the first performs classification on only 1 s time-windows whereas the other uses 2 s windows, then the first can transmit 2 times more information during the same period of

time. Similarly, for equal accuracies, a four class classifier sends two times more binary information than a two class classifier with the same number of trials. To allow comparison of classification results associated to these experimental parameters, the notion of information transfer rate (ITR) has been borrowed from communication theory (Shannon 1964) and can be computed using classifier accuracy  $p$ , the number of classes  $N$  and  $T$  the time length of the trial according to:

$$ITR = \frac{1}{T} (\log_2 N + p \log_2 p + (1-p) \log_2 \frac{(1-p)}{N-1})$$

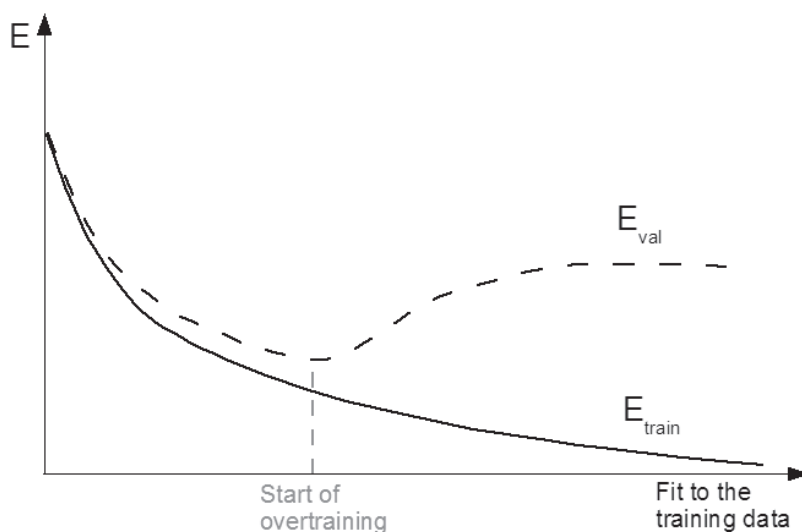
This equation enables the comparison of all kinds of BCI paradigms: according to the literature the current reachable limit of ITR is approximately 25 bits/minute for non invasive BCI's (Krepki et al. 2006).

### Generalization

Generalization is the crucial notion to truly quantify the performance of a classifier: it corresponds to the ability of a trained classifier to classify accurately new data (out of the training set). In order to illustrate that point, consider a set of learning algorithms with a parameter enabling to increase their fit to the training

data (for example by using more and more non-linear discriminant functions). By increasing this fit, classification error on the training data will steadily decrease since the algorithm's aim is to minimize it. But the classifier error on another data set (validation data) will stop decreasing and even start increasing when the increased fit to the training set no longer represents robust properties of the data to classify: this is called overfitting or overtraining. This scenario is illustrated by Fig. 3 where the validation error curve appears useful to select the best step at which the learning process should better be stopped. Due to this issue, classifier performance has always to be computed on a validation or test dataset to assess generalization ability.

Although they appear under various names in the literature, such as cross validation, leave-one-out, leave-k-out, Jack knife (Stone, 1974, Hastie et al., 2001), existing generalization ability assessment frameworks share a common principle. They all consist in depriving the training set of some data samples in order to provide two kinds of sets, training sets and validation sets, with several samplings of the whole data set. For each sampling, trained classifier accuracies are estimated with the items of data that were not used by the learning algorithm.



**Figure 3:** Schematic evolution of learning error ( $E_{\text{learn}}$ ) and validation error ( $E_{\text{val}}$ ) as a function of the fit to learning data. After a regular decrease,  $E_{\text{val}}$  reaches a minimum and a better fit to learning data leads to a worse performance on validation data.

### *Complexity*

A last issue the authors would like to address here is the problem of complexity, a notion related to multiple mathematical quantities. In statistical learning, complexity quantifies the ability of a set of discriminant functions to separate with a perfect accuracy a training set when the clouds of points of the respective classes are very intricate. According to this definition, the discriminant functions of a KNN classifier are clearly more complex than LDA's affine functions. The more complex the discriminant functions of a classifier are, the better they can fit training data and thus the increasing fit to the data can be interpreted as an increase in complexity. As previously mentioned, an overly good fit to the training data may severely impair generalization accuracy of the method. In particular, for affine classifiers such as LDA or linear SVM, complexity increases with the dimension of the feature space. Thus, in MEEG classification, a high dimension due to the huge number of features used to describe each segment of signal may cause problems. This problem is sometimes referred as the curse of dimensionality, arising from the fact that the more dimensions one introduces into the classification problem, the more data is needed to tackle the subsequent increase of possible variability (degrees of freedom). The point is that including an additional dimension would require also adding many more than just one sample to each data set. One should therefore pay attention to the concept of feature selection, which allows us to control the dimension of the feature space. A standard way to select variables is univariate selection: a reduced number of features are drawn from the whole set according to their individual discriminative power. This discriminative power is usually quantified by a Fisher T test. More advanced multivariate selection methods take into account the whole set of features simultaneously in order to choose the optimized subset of variables for the classifier (Garett et al., 2003). Another method that can be used to limit classifier

complexity is Principal Component Analysis (PCA), which reduces the high dimensional data to a few components used for classification.

### 3. APPLICATION TO VISUOMOTOR MEG DATA

We will now illustrate and investigate the use of such procedures by applying linear SVM classification to continuous epochs of MEG data recorded during ongoing visuomotor coordination and during resting states. The aim of this analysis was to assess the ability of the classifier to differentiate between the two conditions and thereby demonstrate its application to real data and indirectly investigate possible physiological processes reflected by the parameters shown to provide best discrimination between the two conditions.

#### *3.1 Data*

##### *Behavioral task*

The visuomotor experiment required subjects to continuously manipulate a trackball to compensate the random rotations of a cube projected on a display screen. The visuomotor (VM) task was alternated with a resting condition (R) during which the subjects relaxed while looking at a motionless cube. The subjects were cued to switch between the two conditions every 8 to 12 s, yielding continuous epochs of steady-state MEG data. All subjects gave informed consent and the study was approved by the local medical ethics committee (Jerbi et al. 2007).

##### *Recordings and Preprocessing*

The cerebral activity was recorded with a whole-head MEG system (151 sensors, VSM MedTech Ltd.) and digitized at 1.25 KHz. The acquired data was first low-pass filtered (100 Hz cut-off), down-sampled to 312.5 Hz and subject to visual inspection. All data segments contaminated by eye-blinks or unwanted swallowing, coughing or movement artifacts were rejected and heart-beat artifacts were corrected using

template-matching. All 8-12s continuous MEG records of each condition (VM or R) were split into non-overlapping 1s- epochs. This yielded about 150 to 300 artifact-free trials for each of the two conditions per subject.

### 3.2 Methods

#### *Feature quantification*

After multiplication by a Hamming window (to avoid edge effects), the 1-s segments of raw MEG signal  $z(t)$  corresponding to the  $k$ -th time window  $W_k$  was band-pass filtered in 6 frequency bands using zero-phase FIR filters (computed by frequency sampling (Oppenheim 89)). The Spectral power was determined for each MEG sensor using Equation 1 in each one of the following standard physiological frequency bands: delta (2-4Hz), theta (5-8Hz), alpha (8-12Hz), beta (15-30Hz), gamma1 (30-60Hz) and gamma2 (60-90Hz). In addition to spectral power, we also computed phase synchrony (i.e. phase locking value, (Lachaux 1999)) between each pair of sensors in each frequency band using Equation 2.

#### *Classification process*

The many types of features that have been computed to quantify the brain state are separated into 12 subsets: 6 subsets containing spectral power features in each of the 6 frequency bands and 6 subsets containing phase synchrony features, also in each frequency band. An exhaustive comparison of these 12 feature subsets has been carried out by using one feature type at a time to predict the brain state (i.e. the behavioral condition) with one of the 3 classifiers : LDA, SVM and KNN. The regularization constant of the SVM was set to  $C=.1$  (based on a preliminary investigation) and the number of neighbors for KNN has been set to 5. The predictive powers of the power and synchrony features were assessed using two methods:

- 1) Ten fold cross-validation (intra-session): the training set of each session of each

subject were split into 10 parts, learning was performed on 9 parts and the classifier was tested on the remaining part.

- 2) Inter-session cross-validation: for each subject, learning was performed on one full session and tested on the remaining session(s). (At least two sessions were available for each subject).

Moreover, the quantity of information about the mental state given by the classifier was evaluated by two quantities:

- 1) Classification accuracy: this is the average between the percentages of good classified points in each class. This measure gives an average rate of 50% for a random guess of the classifier even if the proportion of each class in the dataset is unbalanced.
- 2) Area under the ROC curve (AUR): this quantity gives values of the class information of the discriminant function resulting from the classifier regardless of the threshold to be applied to actually predict the class. The advantage of this method is that it allows for an evaluation of the classifier independently of any possible offset between the features of the test set and the training set. For a random guess the value of AUR is .5 and a perfect (i.e. systematically correct) guess gives  $AUR=1$ . In our results, AUR is computed from the test set to evaluate the generalization ability of the classifier.

### 3.3 Results

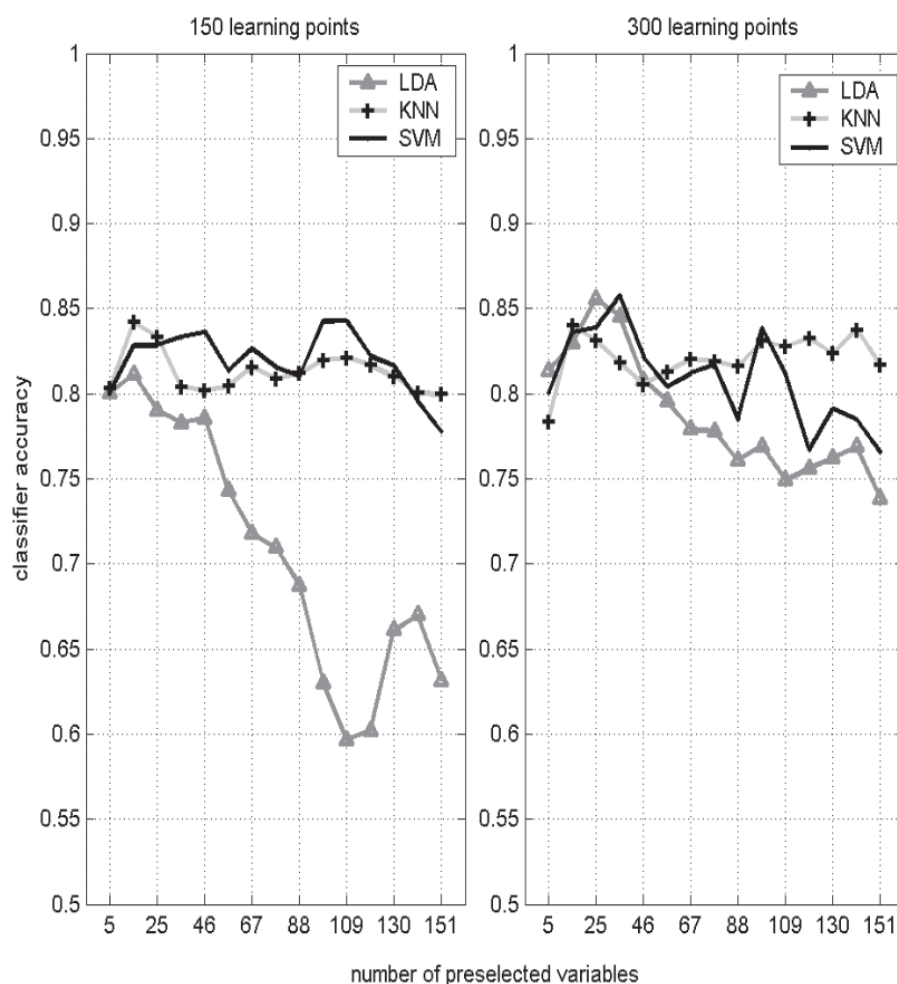
#### *SVM versus LDA and feature selection*

To gain preliminary insight into the effect of the number of features (i.e. possible discrimination parameters) on classification accuracy, we first performed variable selection on the features linked to the beta power. Using an intra-session 10 fold cross validation, only the most discriminative features according to a t-test on the training set were fed into a linear SVM. For comparison, the same procedure was carried out with an LDA classifier and a 5 nearest neighbor classifier (KNN,  $K=5$ ).



The classifier accuracy results of all three algorithms are shown in Fig. 4. Our findings suggest that, when using a small training set (150 time windows), the accuracy of linear SVM and KNN remains stable as the number of selected variables grows. In contrast, LDA accuracy drastically decreases when more than 100 features are selected. Linear SVM and KNN may therefore be considered to be more robust than the LDA approach with this high dimensional data. Our results also indicate that the drop in LDA accuracy is less important when a larger training set

(300 points) is used. This is in line with the idea that high dimensional data is better classified using more data points. Moreover, the results in Fig. 4 also suggest that even when using linear SVM, it might not be necessary to select a high number of features in order to obtain an optimal accuracy with the data at hand. According to these first results from Fig. 4, in the rest of the study investigating the MEG visuomotor data, we chose to use 36 features (with a Fisher t-test) in each subset before the training of a linear SVM classifier.

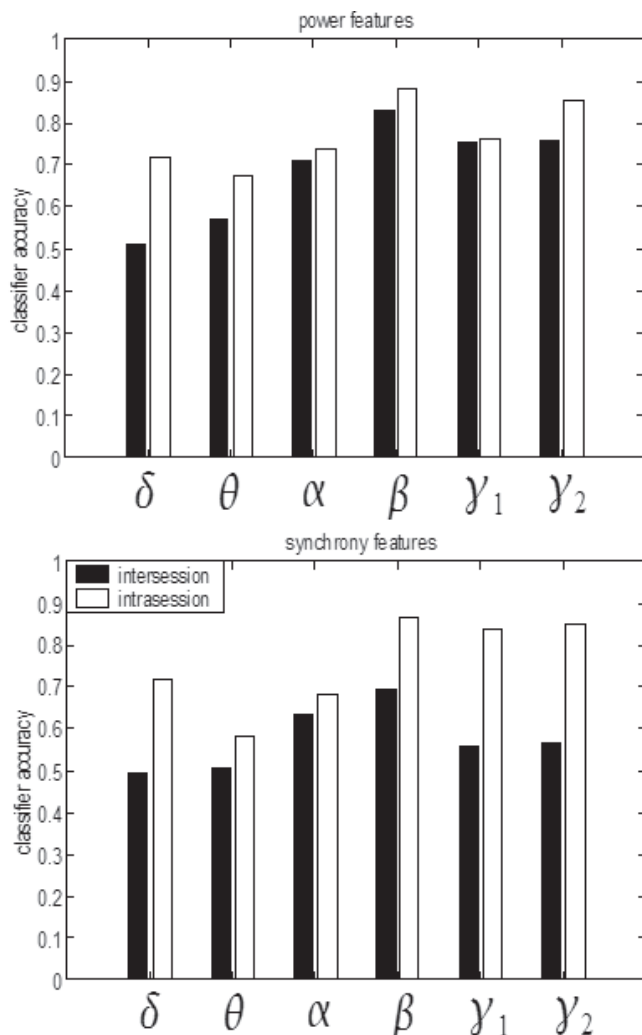


**Figure 4: Classifier accuracy as a function of the number of features and training samples.** Average classifier accuracy in inter-session cross-validation for the 3 subjects as a function of the number of selected power features (from 5 to 151 variables) for LDA, KNN and SVM for two different sizes of the learning set: 150 points (left panel) and 300 points (right panel). Note that these results were obtained by only investigating beta power features as a preliminary analysis aimed at comparing the three techniques.

### Classifier validation

To assess the stability of classifier performance across time, we estimated the classifier accuracy of SVM derived from each feature subset both within the same session and in between sessions using 10 fold cross-validation and inter-session cross-validation respectively. The average classification results across the 3 subjects are given Fig. 5. Our results show that for the MEG visuomotor data used here, the best accuracy was obtained using beta activities as discriminant features. This result did not depend on the type of feature (power or

synchrony) nor the type of validation used (inter or intra session). In particular, the highest score (86%) was obtained for beta power in intra session cross-validation (Fig. 5). Note that nearly all features are less predictive in inter session than in intra session validation. In particular, the classification accuracy based on phase synchrony in the beta and gamma bands was reduced (up to 25%) in inter-session validation. Furthermore, while the predictive power was equivalent for spectral power features and phase synchrony in intra-session validation, spectral power was clearly more predictive in intersession.

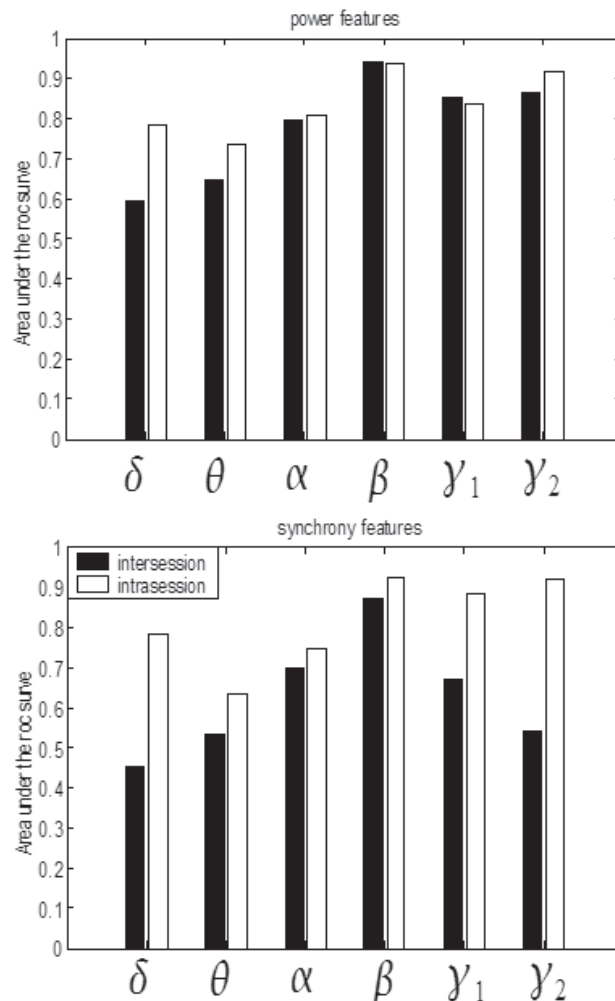


**Figure 5: Inter-session versus intra-session validation.** Classification accuracy of the classifiers derived from power (upper panel) and synchrony (lower panel) features in different frequency bands and for two validation methods: intra-session (using ten fold cross validation) and inter-session (learning on one session and testing on the other).

The information content of each feature set was also investigated with AUR which we computed for the values of the discriminant functions on the test set resulting from each previously computed SVM classifier. The mean AUR across three subjects for each feature subset are given in Fig. 6. The same general tendency as in Fig. 5 is observed except that the ROC area does not drop in intersession compared to intra session in particular for the beta and gamma1 power features. This confirms that, with the data at hand, beta power clearly carries discriminative information which can be used to infer the

mental state from data in another recording session.

The same two cross-validations were also carried out to compute the classification accuracy of feature subsets that combine both spectral power and phase synchrony for a given frequency range. The comparison of the results obtained with classifications based only on power, only on synchrony and on the combination of both are given in Fig. 7. Our findings suggest that whether this combination is beneficial or not to the classifier strongly depends on the frequency band: in the three lower frequency bands, including phase



**Figure 6:** Area under the ROC curve of the decision function of classifiers derived from each feature subset in intra and inter-session cross-validation. Upper panel and lower panels depict the results obtained using power and synchrony features respectively in all six selected frequency bands. The mean classifier accuracy obtained using the selected features from all the frequency bands are also represented on the right.

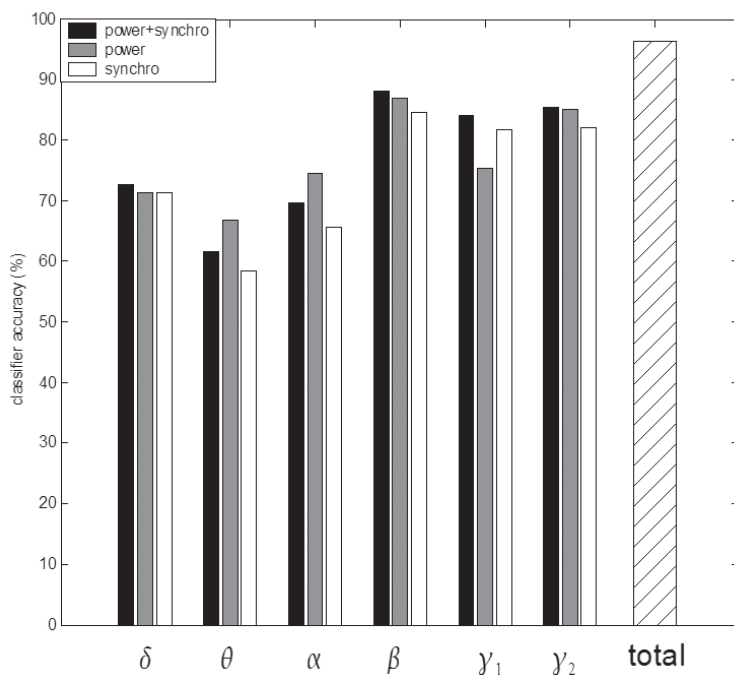
synchrony seems to decrease classifier accuracy achieved with spectral power alone, but conversely the combination of power and synchrony in higher frequency bands outperforms the results obtained with each taken separately. Moreover, combining the selected features from all frequency bands also improves classifier accuracy, reaching 97% as shown on the right of Fig. 7. This accuracy leads to an excellent ITR of 48 bits/minute.

#### *Distribution of classifier coefficients*

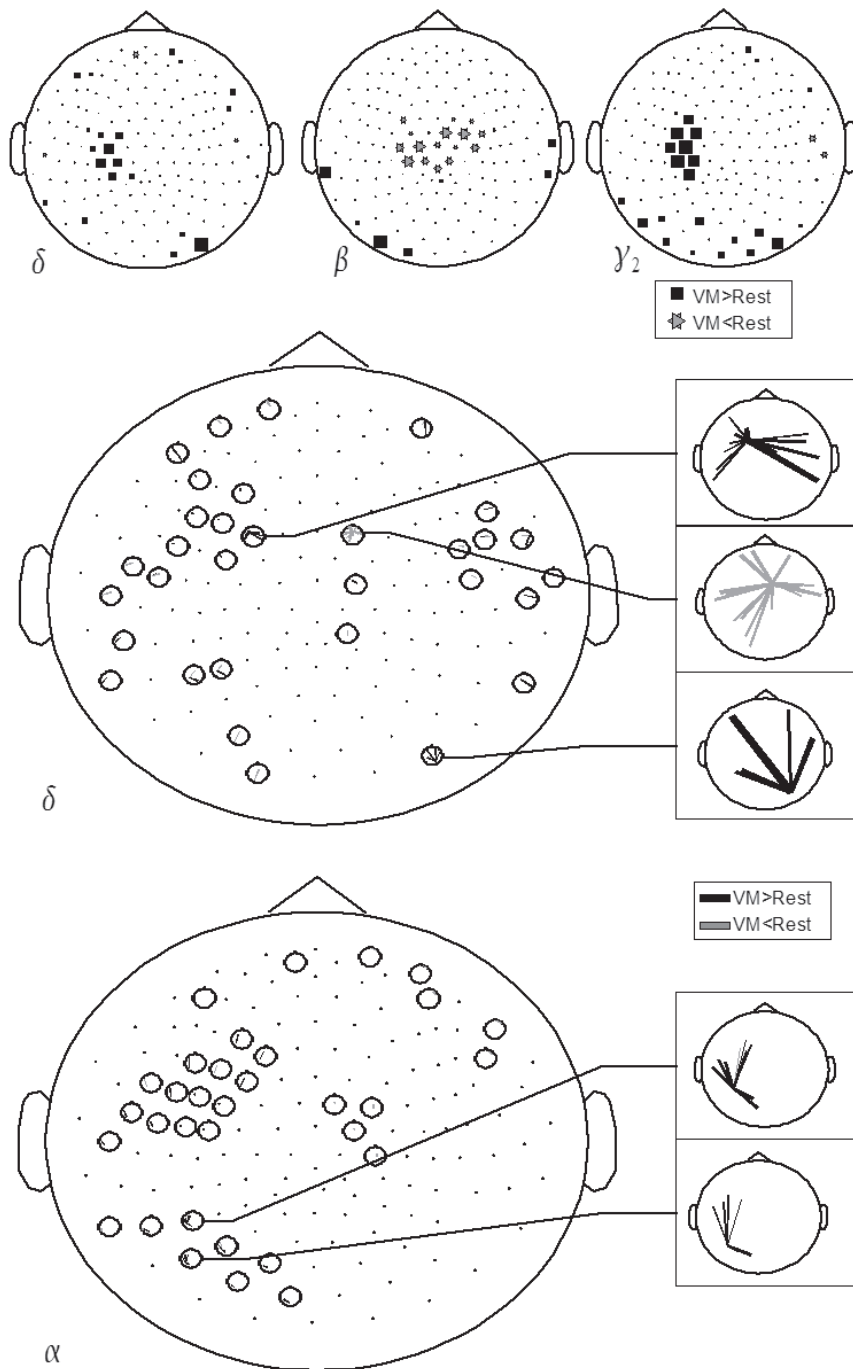
The final point we investigated in this study using the MEG data set is the spatial distribution of classifier coefficients across the sensor-space. When using a binary linear classifier such as LDA or SVM, the discriminant function is an affine function of the features. Therefore when dealing with features of comparable amplitude, one would expect that the features associated to the coefficients with the highest absolute value

are the most discriminant features for the classification. If this coefficient is positive the increase of corresponding feature produces the decision 'class 1' (i.e. VM) while its decrease classifies the data as 'class 2' (Rest). Such interpretations are not completely rigorous as the function is highly multivariate but it may provide useful information on the task-specific features.

To illustrate this with an example, we computed the scalp distribution of the SVM classifier coefficients obtained for delta, beta and gamma 2 power as well as with alpha synchrony for one representative subject on Fig. 8. The upper row shows the topographic distribution of the classifier coefficients related to spectral power features. The beta band power classifier topography shows that this frequency range is associated with negative classifier coefficients in frontal central areas (blue blob in the central topography). In other words, a decrease in beta power over these sensors draws the prediction towards



**Figure 7:** Mean classifier accuracy obtained by combining power and synchrony features, compared to **power-only or synchrony-only classification**. The histograms depict the relationship between classifier accuracy (in all three modes) and the selected frequency band. The bar on the right side shows classification accuracy obtained using the selected features from all the frequency bands combining power and synchrony.



**Figure 8:** Spatial distribution of SVM classifier coefficients for power and synchrony features in various frequency bands for one subject. **Upper row: The distribution of delta (d), beta (b) and gamma 2 (g )** power classifier coefficients: positive coefficients are represented by black squares and negative ones by gray stars (marker size is proportional to amplitude). **Middle and lower panel:** distribution of delta and alpha synchrony classifier coefficients. Note that each small topography on these panel represents the non-zero coefficients of phase synchrony involving the sensor at the current scalp position. Positive coefficients are represented by black lines linking the current sensor and the other sensors, negative ones 2 are represented by grey lines. The close-ups indicate interesting sensors showing many long distance synchronies.



classifying the state as an ongoing visuomotor state (VM) whereas an increase is associated with the resting state (Rest). Moreover, positive coefficients, which correspond to labeling a data sample VM (i.e. a brain state corresponding to visuomotor control) show prominent peaks over frontal areas contralateral to the moving limb in both the delta and gamma2 frequency bands. In contrast to the rather central beta suppression, the location of these delta and high gamma peaks most likely corresponds to sensors most sensitive to neural activation originating in the contralateral sensorimotor cortex. Furthermore, although one might be tempted to interpret the positive coefficients over posterior regions as being task-related and reflecting induced power increase in all three frequency bands during visuomotor coordination, this should be done with great caution. Indeed, gamma power modulations over posterior occipital areas are thought to be involved in visual processing, but one cannot exclude that the posterior modulations at the bottom of the sensor array are not, at least in part, caused by muscle artifacts. Muscle artifacts are known to be broad band (extending also over gamma, beta and even lower frequencies) and may well be selectively present during the visuomotor task but not during the resting condition and thus picked up by our SVM classifier as a discriminant feature. This said, the classifier coefficient peaks over central and contralateral sensorimotor areas are too far from muscle activities to be contaminated by any such artifacts and they clearly represent task-related neural power modulations. Interestingly, these findings are in line with our previous results on task-related power modulation in visuomotor control that were not obtained by any classification procedure but rather by contrasting power topographies across conditions both at sensor level and source level (Jerbi et al. 2004, 2005). This converging evidence from analysis of task-related cortical power modulations on the one hand and linear classification based on multi-frequency power on the other, underline the fact that changes in oscillatory power represent a

strong candidate for classification methods because they represent fundamental modulations in the neurophysiological processes associated with various states. As far as delta and alpha synchrony are concerned (middle and lower panel of Fig. 8), the spatial distribution of positive and negative coefficients seems to be more complex than the power topographies. This may be in part due to the high number of features available. The delta synchrony distribution shows a high number of positive coefficients for synchronization couples between the left motor cortex and other areas, also in line with previous results (Jerbi et al., 2007). Conversely, it is noteworthy that right motor cortex yields negative synchrony coefficients with other distant areas. Whether this reflects the disengagement of the ipsilateral motor cortex (versus the engagement of the contralateral motor cortex) remains to be checked on a complete group of subjects. Also interesting are the positive alpha synchronization coefficients between contralateral parietal cortex and contralateral frontal areas that could reflect the enhanced interactions between visual and frontal areas during the visuomotor task.

### 3.4 Discussion

The results reported here illustrate several important issues related to the classification of MEG states. First of all, we have shown that the number of chosen features can influence classification performance drastically. This is especially true for LDA and any other model-based classifier that requires the estimation of many parameters of the probability distribution of the data. In practice, the number of parameters to estimate has to be small compared to the number of data points in the training set. In LDA for example, all the elements of the covariance matrix of the data have to be estimated: their number is proportional to the square of the number of features. Using LDA in high dimensions thus requires a large amount of data points in the training set. Moreover, even though SVM seems to be less sensitive to the dimension of the

data set used here, this is not necessarily true in general. In fact, if we had chosen a larger regularization constant  $C$ , classification would have dropped due to the high complexity of the classifier. Finally, a non linear classifier like KNN can achieve competitive classification accuracy levels and may very well be chosen for MEEG classification, but does not clearly outperform a linear SVM. It thus seems that, in the high dimensional feature space considered here, linear classifiers are robust and flexible enough to perform the classification task with acceptable accuracy.

After having chosen a reasonable number of parameters and a reliable classifier, the quality of the results is still subject to the validation method. We have shown for example that predicting the mental state of data from one session after training the classifier on another session is more difficult than predicting across the same session using cross-validation. This can be due to many factors related to the reproducibility of the experiment and session specific artifacts or parameter variations (e.g. variable head position within the MEG helmet). This may lead to significant changes in the MEG recordings that could then impair classification results. In some cases, a simple change of the threshold of the discriminant function may be sufficient to improve classification. This is precisely the case we discussed when we observed that in inter-session validation the ROC area did not drop off whereas classifier accuracy did (i.e. ROC accounts for offset variations related to variable baseline levels in between sessions). Such simple corrections might not be sufficient when using phase synchrony measurements as discriminant features, especially in the higher frequency ranges which can be extremely sensitive to artifacts. A further possible cause for the drop in classification accuracy between recording sessions is a change in the state of the subjects themselves, due to learning, a change of strategy or simply alertness and motivation levels. This is particularly relevant for continuous mental states which are internally paced and thus less stable than

exogenous activities triggered by a stimulus. In general, the difficulty to predict one session using another is a crucial and challenging issue which has to be examined carefully for applications such as Brain Computer Interfaces.

In this paper, we discussed the use of power and phase synchrony to perform data classification. A crucial limitation of phase synchrony analysis is that it is extremely difficult to differentiate between true physiological coupling between distinct neural assemblies and false synchronization that appears due to the spreading on the scalp of the electromagnetic field of a unique cortical source. According to Fig. 7 combining power and synchrony features in the upper frequency bands improves classification accuracy which is an argument to say that the recorded synchrony does not carry exactly the same information as power, and thus may not uniquely be explained by diffusion. Whether, and if so to which extent, synchrony between neighboring sensors reflects local amplitude modulations of the same neuronal assembly is not the aim of this specific study and it remains of course an open issue. What is more important, however, is that long-range synchronization between widely separated sensors is less likely to represent a single source especially if one can show that the coupling between the two sensors is not zero-phase (Nolte et al. 2004). Finally, it might be noteworthy to recall that although insight into the dynamics of local and large-scale synchrony and differentiating the role of the two would be well appreciated from a basic physiological stand point, the ultimate goal of a classification method is above all to separate two or more conditions in the most robust and efficient way possible.

The scalp distributions of classifier coefficients obtained here with the visuomotor MEG data set are in agreement with previous work in the field of ERD and ERS analysis in particular in the sensorimotor cortex using non-invasive (Pfurtscheller and Lopes Da Silva, 1999) and invasive (Crone et al. 1998a, 1998b) recordings. Although the spatial distribution of classifier coefficients was

reported here with real visuomotor MEG data in a subset of frequency bands for illustrative purposes, the obtained topographies are in remarkable agreement with our knowledge of the modulation and distribution of cortical oscillatory activity during sensorimotor behavior. First, the distribution of the central negative coefficients for beta power is in line with the known suppression of beta power during movement over bilateral sensorimotor cortices. The power suppression, which seems to be enhanced over the contralateral cortex, probably involves primary sensory, motor and premotor areas, including Brodmann Area 4 (BA4) and most likely the supplementary motor area (SMA) (Jerbi et al. 2004, 2005). Furthermore, in the light of the positive coefficients revealed here in the high gamma range over the contralateral sensorimotor area, and bearing in mind previous findings on the presence of focal gamma activity in the activated motor cortex (Crone et al. 1998b), it is also tempting to suggest that the bilateral beta band activations might be less task-specific or functionally relevant than the gamma power modulations. More studies will be needed in order to systematically address the functional relationship between beta suppression and gamma power increase during motor tasks. Finally, a further frequency range that showed strong positive classification coefficients between visuomotor and resting conditions is the delta (2-4 Hz) band. Although surprisingly low, this frequency band has recently been shown to host slow sensorimotor oscillations related to the ongoing control of hand speed (Jerbi et al. 2007). Whether an increase in delta-range cortical power changes in the motor cortex represents modulation of an intrinsic sensorimotor rhythm or whether it is rather a direct reflection of task parameters is still an open question. Similarly, it would be difficult to draw qualitative conclusions by visual inspection of the complex synchronization topographies (middle and lower panel of Fig. 8), and our observations about enhanced contra-lateral synchronization in the delta and alpha band have to be

confirmed by future work on a larger population of subjects. What matters here, is that the classification method we used identified the sensorimotor delta power as one of the major features that best separates the two conditions, a finding fully in line with a separate analysis of the data. Again, these converging findings from basic research analysis and data classification procedures confirm the fact that successful classification inherently relies on distinguishing task-related physiological phenomena. Moreover, although classification techniques can be improved by taking neuroscience *a priori* knowledge into account (such as frequency bands or regions of interest), they also hold the potential of indirectly enhancing our understanding of brain dynamics. Indeed, the features revealed by such methods to be most relevant for the optimization of classification may lead to new hypotheses about the dynamics of human cognition in specific tasks.

## CONCLUSION

Beyond reviewing the basic principles of BCI-oriented data classification, we used MEG data to illustrate the limitations and differences between several approaches and to shed some light on the effect of various user-defined parameters. Furthermore, the successful MEG data classification reported here in offline analysis, demonstrates that the real-time detection of visuomotor activity using power and synchrony measurements is feasible and suggests that classification of motor or visuomotor imagery (without actual movement) can be obtained using the same methodology developed here. This is particularly the case, since motor imagery is known to activate (although to a lesser extent) predominantly the same areas involved in real movement execution (Beisteiner et al., 1995; Leocani et al., 1999) and the subject's ability to enhance the corresponding oscillatory modulations can also be trained. On the whole, the accuracy of the classification achieved here during a continuous task (i.e. not event-related)

suggests the possibility of achieving a high transmission rate asynchronous BCI. In addition, the agreement between the extracted classification features on one hand and the knowledge of the physiological processes underlying the visuomotor control data on the other, suggests that basic neuroscience research can also benefit from M/EEG data classification methods as they may yield novel insights to the study of the mechanisms underlying ongoing brain dynamics during mental tasks.

#### ACKNOWLEDGEMENTS

The authors would like to thank Dr Mario Chavez and Frédérique Amor for their useful comments on the manuscript.

#### BIBLIOGRAPHY

- ANDERSON C, STOLZ E, SHAMSUNDER S (1998) Multivariate Autoregressive Models for Classification of Spontaneous Electroencephalogram During Mental Tasks. *IEEE Trans Biomed Eng* 45: 277-286
- BARRET G (1996). Event-related potentials (ERPs) as a measure of complex cognitive function. *Electroencephalogr Clin Neurophysiol* 46: 53-63.
- BASHASHATI A, FATOURECHI M, WARD RK, BIRCH GE (2007) A survey of signal processing algorithms in brain-computer interfaces based on electrical brain signals. *J Neural Eng* 4: R32-R57
- BEISTERNER R, HOLLINGER P, LINDINGER G, LANG W, BERTHOZ A. (1995) Mental representations of movements. Brain potentials associated with imagination of hand movements. *Electroenceph Clin Neurophysiol*, 96:183-193
- BORISOFF JF, MASON SG, BASHASHATI A, BIRCH GE (2004) Brain-Computer Interface Design for Asynchronous Control Applications: Improvements to the LF- ASD Asynchronous Brain Switch. *IEEE Trans. Biomed Eng* 51: 985-992
- BOSTANOV V (2004) {BCI} competition 2003-data sets ib and iib: feature extraction from event-related brain potentials with the continuous wavelet transform and the t-value scalogram. *IEEE Trans Biomed Eng* 51: 1057-1061
- BROVELLI A, DING M, LEDBERG A, CHEN Y, NAKAMURA R, BRESSLER S (2004) Beta oscillations in a large-scale sensorimotor cortical network: directional influences revealed by granger causality. *Proc Natl Acad Sci USA* 101: 9849-9854
- CARLSON TA, SCHRATER P, HE S (2003). Patterns of activity in the categorical representation of objects. *J Cogn Neurosci* 15: 704-717
- CRONE N, MIGLIORETTI D, GORDON B, LESSER R (1998a) Functional mapping of human sensorimotor cortex with electrocorticographic spectral analysis. ii. event-related synchronization in the gamma band. *Brain* 121: 2301-2315
- CRONE N, MIGLIORETTI D, GORDON B, SIERACKI J, WILSON M, UEMATSU S, LESSER R (1998b) Functional mapping of human sensorimotor cortex with electrocorticographic spectral analysis. i. alpha and beta event-related desynchronization. *Brain* 121: 2271-2299
- DUDA RO, HART PE, STORK DG (2001) *Pattern Classification* Wiley-Interscience
- FARWELL LA, DONCHIN E (1988) Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroenceph Clin Neurophysiol* 70: 510-523
- FRACKOWIAK RSJ, FRISTON KJ, FRITH CD, DOLAN RJ, MAZZIOTTA JC (1997) *Human Brain Function*. San Diego, Calif. Academic Press
- FUCHS T, BIRBAUMER N, LUTZENBERGER W, GRUZELIER JH, KAISER J (2003) Neurofeedback Treatment for Attention-Deficit/Hyperactivity Disorder in Children: A Comparison with Methylphenidate. *Applied Psychophysiology and Biofeedback* 28: 1-12.
- GARRETT D, PETERSON DA, ANDERSON CW, THAUT MH (2003) Comparison of linear, nonlinear, and feature selection methods for EEG signal classification. *IEEE Trans Neural Syst Rehabil Eng*. 11: 141-144
- GEORGIOPOULOS AP, LANGHEIM FJ, LEUTHOLD AC, MERKLE ANMSPMTISEBR (2005) Magnetoencephalographic signals predict movement trajectory in space. *Exp Brain Res* 167: 132-135.
- GRAIMANN B, PFURTSCHELLER G (2006) Quantification and visualization of event-related changes in oscillatory brain activity in the time-frequency domain In: NEUPER C, KLIMESCH W (eds) *Progress in Brain Research* 159: 79-97
- GREEN DM, SWETS JA (1974). *Signal Detection Theory and Psychophysics*. Huntington, NY: R. E. Krieger
- GYSELS E, CELKA P (2004) Phase Synchronisation for the recognition of Mental Tasks in a Brain Computer Interface. *IEEE Trans Neural Syst Rehabil Eng* 12: 206-415
- GYSELS E, RENEVEY P, CELKA P (2005) SVM-based recursive feature elimination to compare phase synchronization computed from broadband and narrowband EEG signals in Brain-Computer Interfaces. *Signal Processing* 85: 2178-2189
- HASELSTEINER E, PFURTSCHELLER G (2000) Using time-dependent neural networks for EEG classification. *IEEE Trans Rehab Eng* 8: 457-463
- HASTIE T, TIBSHIRANI R, FRIEDMAN J (2001) *The Elements of Statistical Learning* Springer Verlag
- HAYNES JD, REES G (2006) Decoding mental states from brain activity in humans. *Nat Rev Neurosci* 7: 523-534
- HIRAIWA A, SHIMOHARA K, TOKUNAGA Y (1990) EEG topography recognition by neural networks. *IEEE Eng Med Biol Mag* 9: 39-42
- JERBI K, BAILLET S, LACHAUX JP, PANTAZIS D, LEAHY R, GARNERO L (2005) Modulations of Power and Synchronization of Neural Activity during Sustained Visuomotor Coordination, a MEG Study. *Proc. 11th Int Conf on Human Brain Map*
- JERBI K, LACHAUX JP, N'DIAYE K, PANTAZIS D, LEAHY RM, GARNERO L, BAILLET S (2007) Coherent neural representation of hand speed in humans revealed by MEG imaging *Proc Nat Ac Sci USA* 104: 7676-7681
- JERBI K, LACHAUX J-P, BAILLET S, GARNERO L (2004). Imaging Cortical Oscillations during Sustained Visuomotor Coordination In MEG. *Proc. IEEE International Symposium on Biomedical Imaging: From Nano to Macro*.



- KEIRN ZA, AUNON JI (1990) A new mode of communication between man and his surroundings. *IEEE Trans Biomed Eng* 37: 1209-1214
- KREPKI R (2006) Berlin Brain-Computer Interface. The HCI communication channel discovery. *Int J Human-computer Studies* 65: 460-477
- KUBLER A, KOTCHOUBEY B, KAISER J, WOLPAW JR, BIRBAUMER N (2001) Brain-computer communication: unlocking the locked in. *Psychol Bull* 127: 358-375
- LACHAUX JP, RODRIGUEZ E, MARTINERIE J, VARELA FJ (1999) Measuring phase synchrony in brain signals *Hum Brain Mapp* 8: 194-208
- LACONTE S, STROTHER S, CHERKASSKY V, ANDERSON J, HU X (2005) Support vector machines for temporal classification of block design fMRI data. *Neuroimage* 26: 317-329
- LAL TN, SCHRODER M, HILL J, PREISSEL H, HINTERBERGER T, MELLINGER J, BOGDAN M, ROSENSTIEL W, HOFMANN T, BIRBAUMER N, SCHOLKOPF B (2005) A Brain Computer Interface with Online Feedback based on Magnetoencephalography. In: DE RAEDT L, WROBEL S (eds) *Proceedings of the 22nd International Conference on Machine Learning*: 465-472
- LEBEDEV M, NICOLELIS MA (2006) Brain-machine interfaces: past, present and future. *Trends Neurosci* 29: 536-546
- LEOCANI L, MAGNANI G, COMI G (1999) Event Related Desynchronisation during execution, imagination and withholding of movement. *Handbook of Electroencephalography and Clinical Neurophysiology*. PFURTSCHELLER GALDS, F H 6: 245-268
- LOPES DA SILVA FH (2006) Event related oscillations: what about phase? In: NEUPER C, KLIMESCH W (eds) *Progress in Brain Research* 159: 1-17
- LOTTE F, CONGEDO M, LECUYER A, ARNALDI B (2007) A review of classification algorithms for EEG-based Brain-Computer Interfaces. *J Neural Eng* 4: R1-R13
- MIDDENDORF M, MCMILLAN G, CALHOUN G, JONES KS (2000) Brain-computer interfaces based on the steady-state visual-evoked response. *IEEE Trans Rehabil Eng* 8: 211-214
- MILLAN JR, RENKENS F, MOURINO J, GERSTNER W (2004) Noninvasive brain-actuated control of a mobile robot by human EEG. *IEEE Trans Biomed Eng* 51: 1026-33
- MULLER KR, MIKA S, RATSCH G, TSUDA K, SCHOLKOPF B (2001) An introduction to kernel-based learning algorithms. *IEEE Tran Neur Netw* 12: 181-202
- MULLER-GERKING J, PFURTSCHELLER G, FLYVBJERG H (1999) Designing optimal filters for single-trial EEG classification in a movement task. *Clin Neurophysiol* 110: 787-798
- NOLTE G, BAI O, WHEATON L, MARI Z, VORBACH S, HALLETT M (2004) Identifying true brain interaction from EEG data using the imaginary part of coherency. *Clin Neurophysiol* 115: 2292-2307
- OBERMEIER B, GUGER C, NEUPER C, PFURTSCHELLER G (2001) Hidden Markov models for online classification of single trial EEG. *Pattern Recognit Lett* 22: 1299-1309
- OPPENHEIM AV, SCHAEFER RW (1989) Digital signal processing. Prentice Hall, (please replace the date by 1975 for the original edition)
- PFURTSCHELLER G (2006) The cortical activation model (CAM) In: NEUPER C, KLIMESCH W (eds) *Progress in Brain Research* 159: 19-27
- PFURTSCHELLER G, NEUPER C, FLOTZINGER D, PREZINGER M (1997) EEG-based discrimination between imagination of right and left hand Electroencephalogr. *Clin Neurophysiol* 103: 642-651
- PFURTSCHELLER G, LOPES DA SILVA FH (1999) Event-related synchronization and desynchronization: basic movement. EEG/MEG principles *Clin Neurophysiol* 110: 1842-1857
- PIKOVSKI A, ROSENBLUM M, KURTHS J (2001) Synchronization: A Universal Concept in Nonlinear Science
- WEI Q, WANG Y, GAO X, GAO S (2007) Amplitude and phase coupling measures for feature extraction in an EEG-based brain-computer interface. *J Neural Eng* 4: 120-129
- QUYEN MLV, SOSS J, NAVARRO V, ROBERTSON R, CHAVEZ M, BAULAC M, MARTINERIE J (March 2005) Preictal state identification by synchronization changes in long-term intracranial EEG recordings. *Clin Neurophysiol* 116: 559-568
- RODRIGUEZ E, GEORGE N, LACHAUX JP, MARTINERIE J, RENAULT B (1999) Perception shadow: long distance gamma band synchronisation and desynchronisation on the human scalp. *Nature* 397: 430-433
- SALMELIN R, HARI R (1994) Spatiotemporal characteristics of sensorimotor neuromagnetic rhythms related to thumb movement. *Neuroscience* 60: 537-550
- SCHERER R, MÜLLER GR, NEUPER C, GRAIMANN B, PFURTSCHELLER G (2004) An Asynchronously Controlled EEG-Based Virtual Keyboard : Improvement of the Spelling Rate. *IEEE Trans Biomed Eng* 51: 979-984
- SHANNON CE, WEAVER W (1964) *Mathematical Theory of Communication*. Champaign, IL: University Illinois Press
- STONE M (1974) Cross-validatory choice and assesment of statistical predictions. *J Roy Statist Soc* 36: 111-147
- STREHL U, LEINS U, GOTH G, KLINGER C, HINTERBERGER T, BIRBAUMER N (2006) Self-regulation of Slow Cortical Potentials: A New Treatment for Children With Attention-Deficit/Hyperactivity Disorder. *Pediatrics* 118: 1530-1540
- TALLON-BAUDRY C, BERTRAND O (1999) Oscillatory gamma activity in humans and its role in object representation. *Trends Cogn Sci* 3: 151-162
- TALLON-BAUDRY C, BERTRAND O, DELPUECH C, PERMIER J (1997) Oscillatory gamma-band (30-70 Hz) activity induced by a visual search task in humans. *J Neurosci* 17: 722-734
- TAYLOR DM, TILLERY SI, SCHWARTZ AB (2002) Direct cortical control of 3D neuroprosthetic devices. *Science* 296: 1829-1832
- TOWNSEND G, GRAIMANN B, PFURTSCHELLER G (2006) A comparison of common spatial patterns with complex band power features in a four-class BCI experiment *IEEE Trans Biomed Eng* 53: 642- 651
- VAPNIK V (1998) *Statistical Learning Theory*. John Wiley, New York
- VARELA F, LACHAUX JP, RODRIGUEZ E, MARTINERIE J (2001) The Brainweb: Phase synchronisation and Large-scale integration. *Nature Rev Neurosci* 2: 229-239
- VARELA FJ (1995) Resonant cell assemblies : a new approach to cognitive function and neuronal synchrony. *Biol Res* 28: 81-95
- WALTRE DO (1968) Coherence as a mesure of relationship between EEG records. *Electroenceph Clin Neurophysiol* 24: 282



- WESSBERG J, NICOLELIS MA (2004) Optimizing a linear algorithm for real-time robotic control using chronic cortical ensemble recordings in monkeys. *J Cogn Neurosci* 16:1022-1035
- WOLPAW JR, BIRBAUMER N, MCFARLAND DJ, PFURTSCHELLER G, VAUGHAN TM (2002) Brain-computer interfaces for communication and control. *Clin Neurophysiol* 113: 767-791

