Neuper & Klimesch (Eds.)
Progress in Brain Research, Vol. 159
ISSN 0079-6123

CHAPTER 6

Quantification and visualization of event-related changes in oscillatory brain activity in the time–frequency domain

Bernhard Graimann* and Gert Pfurtscheller

Laboratory of Brain–Computer Interfaces (BCI-Lab), Institute for Knowledge Discovery, Graz University of Technology, Graz, Austria

Abstract: In this chapter we review the traditional approach for ERD/ERS quantification and a more recent approach based on wavelet transform. In particular, we address the visualization of these phenomena and the validation of the results through statistical significance testing. Furthermore, we report on preprocessing using independent component analysis (ICA) and introduce a novel ERD/ERS maximization method.

Keywords: event-related desynchronization (ERD); event-related synchronization (ERS); bootstrap; wavelet; spatial filters

Introduction

The classical approach of quantification and visualization of event-related desynchronization and event-related synchronization (ERD/ERS) is to calculate and display ERD/ERS time courses representing bandpower changes in specific frequency bands (Pfurtscheller and Aranibar, 1977; Kalcher and Pfurtscheller, 1995). Typical ERD/ERS analysis requires the investigation of various channels and frequency bands for different subjects. Even for a modest number of subjects, channels, and frequency bands, the number of time courses generated by the classical approach becomes prohibitively large, which renders a detailed analysis difficult. As a better alternative, time–frequency maps may be used to provide a comprehensive overview of the activity over broad frequency ranges. This approach can greatly facilitate ERD/ERS analysis. A number of methods based on Fourier transform (Makeig, 1993), continuous wavelet transform (Tallon-Baudry and Bertrand, 1999), and matching pursuit (Durka et al., 2001; see also Chapter 8, this volume) have been proposed for the quantification of ERD/ERS. Regardless of which method is employed, appropriate statistical analyses must be performed to ensure the validity of any conclusions made regarding a study. That is, the patterns apparent in the data must be shown to be statistically significant. Failing this, any conclusions made concerning the data could be erroneous.

ERD/ERS analysis also requires appropriate pre-processing of the brain signals to enhance the patterns present. This is particularly true for EEG signals, since they capture the activity from multiple sources occurring within a relatively large brain volume. Owing to volume conduction through scalp, skull, and other layers, the recorded EEG signal can be seen as a "blurred" copy of the underlying cortical activity. To overcome this problem, simple spatial filters like common average

*Corresponding author. E-mail: graimann@tugraz.at

DOI: 10.1016/S0079-6123(06)59006-5

reference (CAR) and orthogonal source derivation are usually used in ERD/ERS analysis (Pfurtscheller, 1988). More sophisticated spatial filters derived from linear transformations such as independent component analysis (ICA) (Makeig et al., 1997; see also Chapter 7) and common spatial patterns (CSPs) (Fukunaga, 1990; Koles et al., 1995) are very promising since they can improve spatial resolution and signal-to-noise ratio (SNR) of the brain signals; however, they have not yet been widely used for the analysis of oscillatory activity in brain signals.

The successful estimation of ERD/ERS requires a framework consisting of appropriate preprocessing of the data, robust quantification, and reliable statistical analysis. In the following, we suggest such a framework. We review the standard method for quantifying ERD/ERS based on band-pass filtering and wavelet analysis, and we show how time–frequency maps representing only statistically significant values can be calculated. Advanced spatial filters derived by ICA and CSP are developed and applied, and the results are compared with the generated by using orthogonal source derivation. The effectiveness of the proposed framework is demonstrated on movement-related EEG and ECoG data.

Quantification of ERD/ERS

The standard ERD/ERS calculation as defined by Pfurtscheller and Aranibar (1979) is performed by band-pass filtering of each trial, squaring of the samples, and subsequent averaging over trials and over sample points. The ERD/ERS is then defined as the proportional power decrease (ERD) or increase (ERS) in relation to the baseline activity during a specific reference interval that is usually selected to be several seconds before the event onset. Since evoked potentials can mask induced activities, it is often useful to subtract the mean of the data for each sample before squaring (Kalcher and Pfurtscheller, 1995). Mathematically, the procedure can be expressed as follows:

$$y_{ij} = (s_{ij} - \bar{s}_j)^2; \quad P_j = \frac{1}{N-1}\sum_{i=1}^{N} y_{ij} \tag{1}$$

$$R = \frac{1}{k+1}\sum_{r_0}^{r_0+k} P_j \tag{2}$$

$$\mathrm{ERDS}_j = \frac{P_j - R}{R} \cdot 100\% \tag{3}$$

where N is the total number of trials (epochs or sweeps), s_{ij} the jth sample of the ith trial of the band-pass filtered data, and $\bar{s}_j$ the mean of the jth sample averaged over all band-pass filtered trials. P_j is the power or intertrial variance at the jth sample. R is the average power in the reference interval $[r_0, r_0+k]$. To increase the statistical reliability, the ERD/ERS values have to be averaged over time intervals. The smaller the number of trials available, the longer will be the time interval required to do this. In standard ERD/ERS analysis with typically 30–80 trials, averaging periods between 125 and 250 ms are used. This choice remains fixed for all frequency bands investigated. Knösche and Bastiaansen (2002) pointed out that the interval length used for averaging should be selected according to the frequency bands being investigated, and in fact using a time interval that is half a period of the frequency under investigation gives the optimal temporal resolution. Equivalently, the Hilbert transform can be employed to calculate the signal envelope, which also leads to optimal temporal resolution (Clochon et al., 1996; Knösche and Bastiaansen, 2002). When considering temporal resolution, it is also necessary to consider frequency resolution. The time and frequency behaviors of a signal are not independent, since their representations are directly linked by the Fourier transform. That is, when the signal has finite time support, its frequency bandwidth must be unlimited, and vice versa. It is not possible for a signal to simultaneously have finite time duration and finite frequency bandwidth (Cohen, 1995). The consequence of this ‘‘uncertainty principle’’ for ERD/ERS analysis is that the widths of the frequency bands investigated have to be changed according to the trade-off between time and frequency resolution. Since the cycle period is long for low-frequency components and short for high-frequency components, low-frequency components have good frequency resolution while high-frequency components have good temporal

resolution. Consequently, the frequency bands used for ERD/ERS analysis should be narrower for lower frequencies and wider for higher frequencies.

As an alternative to the previously described standard method of band-pass filtering, a continuous wavelet approach can be used which is very suitable to account for the trade-off between time and frequency resolution. By using the following complex Morlet wavelet,

$$w(t,f_0) = (\sigma_t\sqrt{\pi})^{-1/2}\ e^{-t^2/2\sigma_t^2}\ e^{i2\pi f_0 t} \qquad (4)$$

the time-varying power values $P(t,f_0)$ of the signal at a specific frequency band with mid-frequency f_0 and bandwidth defined by σ_t is the squared norm of the convolution of the complex wavelet with the signal:

$$P(t,f_0) = \left|w(t,f_0)*s(t)\right|^2 \qquad (5)$$

With $\sigma_t = 1/(2\pi\sigma_{f_0})$ and σ_{f_0} a function of the frequency to analyze determined by

$$\sigma_{f_0} = f_0/c, \qquad (6)$$

the compromise between time and frequency localization is defined by the constant c, which is usually set to 7 (Tallon-Baudry and Bertrand, 1999). Higher values increase the frequency resolution while lower values increase the time resolution.

Both the classical method and the method based on wavelets can be used to construct time–frequency representations for multiple frequencies. In the former case, partially overlapping frequency bands are used while in the latter case the frequency to be analyzed, f_0, is varied over a specific frequency range. The result is a time–frequency map — the so-called ERD/ERS map — representing ERD and ERS patterns covering the entire frequency spectrum of interest. Figure 1 illustrates the construction of such an ERD/ERS map.

ERD values are colored red, while ERS values are shown in blue. The trigger time point (event onset) is marked by a dashed-dotted vertical line. The reference period is indicated by the two dotted vertical lines in the ERD/ERS map.

As mentioned previously, the choice of the frequency range and the frequency bands depends on the actual application. For EEG analysis, a frequency range of 7–40 Hz may be sufficient, while for MEG and EcoG, a broader range would be more appropriate because of the superior SNR of these recording methods.

Figure 2 shows four ERD/ERS maps calculated from ECoG data that was recorded with a sampling rate of 200 Hz during a self-paced index finger movement experiment. As with the ERD/ERS map shown in Fig. 1, the vertical line indicates movement onset. The reference interval, however, was calculated over the entire trial period. This was necessary because the trial length chosen of 6 s resulted in partially overlapping trials making it difficult to select one particular short time period for calculating the baseline activity. The frequency range displayed in Fig. 2 is from 7 to 95 Hz, which provides for the fact that in ECoG data not only alpha and beta frequencies are of interest but also gamma activity. Map A was calculated by the standard method with 2-Hz bands and a Hilbert transform to capture the envelope of the filtered signal. The continuous wavelet approach was used to produce maps B and C. In the former, the trade-off factor between frequency and temporal resolution c was set to 1 (cf. Eq. (6)), while in the latter, it was set to 7. A small value of c results in very good temporal localization at the expense of poor frequency localization. Setting the value to 7 gives reasonable temporal and frequency resolution over the entire frequency range investigated. By contrast, map A shows that using a fixed narrow frequency band for the whole frequency range obscures the gamma activity that is clearly visible in the other maps. Broader frequency bands would be necessary to capture the higher frequency components. A comparison of the alpha and beta frequency ranges from maps A and C shows that for this range both methods produce equivalent results. Thus, the decision concerning which method should be chosen for the frequency range typically used for investigating EEG data is not critical. Map D is equivalent to map C, but depicts only statistically significant ERD/ERS values. This results in an ERD/ERS map that is much clearer and easier to interpret. Without this enhancement, some of the features visible in the maps are not significant despite their appearance. It is therefore crucial to suppress such information since it can lead to the

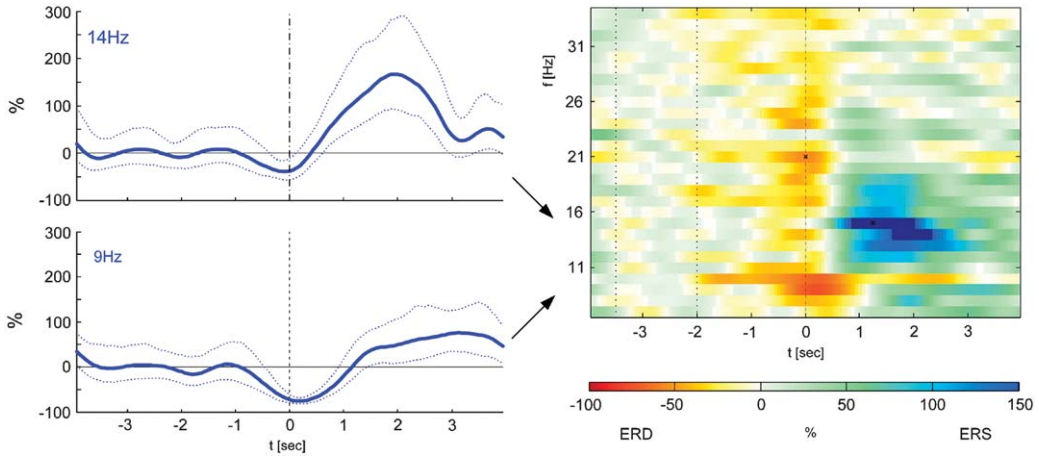

Fig. 1. Assembly of event-related desynchronization and event-related synchronization (ERD/ERS) maps. The time course of the bandpower for each frequency band of interest is calculated and displayed as colors representing the instantaneous proportional change relative to the baseline average. This information (examples on the left) is assembled to form the map on the right. Here the center frequency of each band is denoted on the scale on the left of the map, while the time in seconds appears under the map. As shown in the color legend below the map, values from −100% to +150% are scaled appropriately as various colors. The time courses on the left also show the corresponding confidence intervals that can be used to display only statistically significant values (see Section "Statistical significance of ERD/ERS").

formulation of erroneous conclusions concerning the data. A method for performing this statistical analysis is given below.

Statistical significance of ERD/ERS

Many statistical methods assume either Gaussian distributions of the values to be analyzed or they require a large number of samples (or trials in the case of ERD/ERS analysis). Unfortunately, these assumptions cannot be made for ERD/ERS values. The necessary nonlinearity (squaring) in the quantification procedures yields highly skewed distributions that are no longer approximated by Gaussian distributions. Figure 3 shows the normalized histogram of ERD/ERS values for exemplary alpha and beta bands, respectively. Clearly, both distributions are skewed toward the zero power.

Nonparametric methods that are not based on the Gaussianity assumptions are possible candidates for the statistical assessment of ERD/ERS values. In fact, a simple sign test was proposed by (Kalcher and Pfurtscheller, 1995); however, the sign test reduces the numerical data to a qualitative form (plus or minus), and thus ignores a considerable amount of information. Therefore, the sign test is not very powerful. That is, more data and/or larger differences are necessary to gain sufficient statistical evidence. An efficient approach for calculating statistically significant ERD/ERS values by using parametric tests followed by Box–Cox transforms to approximately normalize the samples was suggested in Zygierewicz et al. (2005). Another powerful way to calculate the confidence and significance of ERD/ERS values is the bootstrap (Graimann et al., 2002). Bootstrap techniques are an alternative to asymptotic methods. The bootstrap is a method for estimating the distribution of a test statistic by resampling the data. The basic idea is simply to replace the unknown population distribution with the known empirical distribution. Properties of the estimator such as confidence intervals are then determined based on the empirical distribution (Davison and Hinkley, 1997). Apart from its general applicability and simplicity, the bootstrap has a number of other advantages. The most important advantage is that there is no need to assume a Gaussian or other parametric distribution on the data. Furthermore, almost any parameter of the bootstrap distribution may serve as a bootstrap

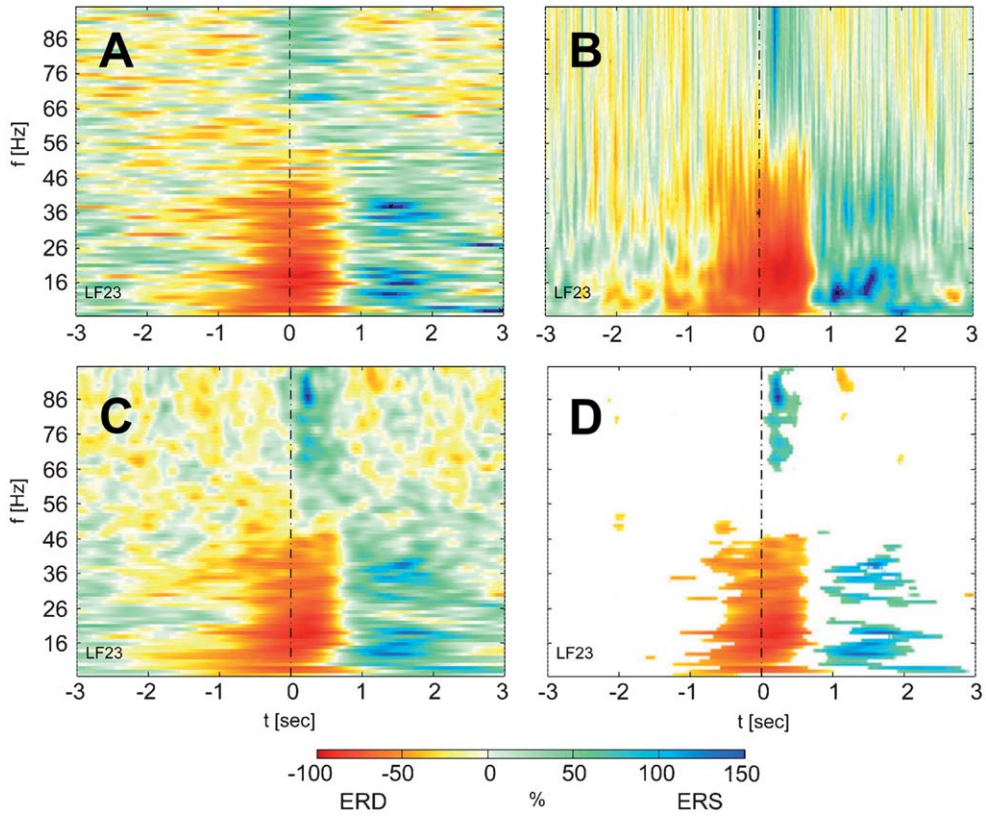

Fig. 2. ERD/ERS maps produced in different ways. All of the above maps represent ERD/ERS recorded from the same electrode. Map A was calculated by the standard method. Maps B and C were calculated using the continuous wavelet approach with a "*c*" constant of 1 and 7, respectively. The former shows higher temporal resolution while the latter shows better frequency resolution. Map C shows most clearly the gamma activity. Map D is the same as map C, but with only the statistically significant features shown. There, it is much clearer where activity of interest exist.

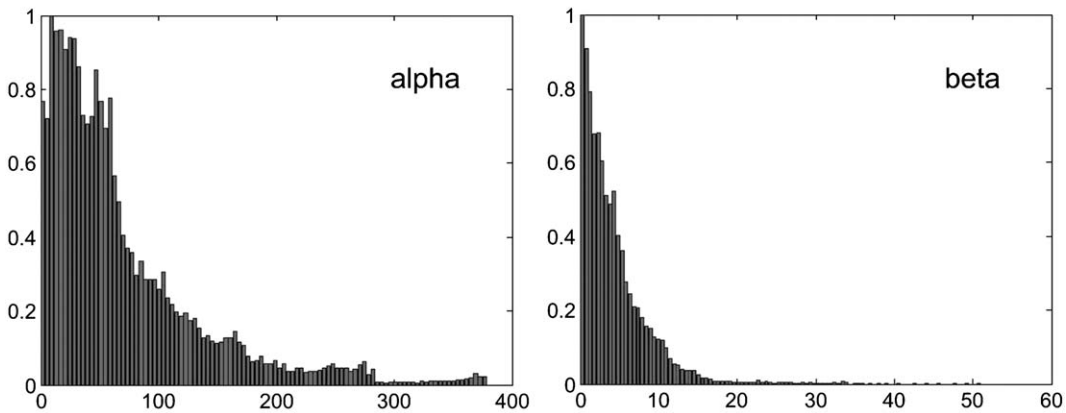

Fig. 3. Normalized power distribution in the alpha and beta bands. The samples are skewed toward zero power for both bands.

estimate of the corresponding population parameter. That is, more robust statistical estimators such as the median or the 95th percentile of the bootstrap distribution can be considered as well (Chernik, 1999). The following procedure shows how a bootstrap procedure, specifically the t-percentile bootstrap (Graimann et al., 2002), may be applied to calculate confidence intervals for the ERD/ERS estimates.

Let N be the number of trials, and B be the number of boostrap resamples. E_j denotes the set of all $e_{ij} = (y_{ij} - R)/R$ of the jth sample and of all trials quantified according to Eqs. (1) and (2) and calculated with the standard ERD/ERS method, the continuous wavelet, or any other suitable method. The sample mean and the sample variance of E_j are denoted by $\bar{e}_j$ and s_j^2, respectively.

For each sample j:

Draw N values from E_j, where each value is selected at random, with replacement. The N drawn values are the bootstrap resample $\hat{E}_j$.
Calculate the mean $\hat{\mu}_{\hat{E}_j}$ and the standard deviation $\hat{\sigma}_{\hat{E}_j}$ of all N values in $\hat{E}_j$.
Calculate $\hat{\mu}_j = \hat{\mu}_{\hat{E}_j} - \bar{y}_j/\hat{\sigma}_{\hat{E}_j}$ Repeat these steps to obtain B boostrap estimates $\hat{\mu}_{j1} \ldots \hat{\mu}_{jB}$. B should be larger than 500.
To approximate the distribution of $\hat{\mu}_j$, sort all estimates so that $\hat{\mu}_{j(1)} \leq \hat{\mu}_{j(2)} \leq \ldots \leq \hat{\mu}_{j(B)}$.
The $100(1-\alpha)\%$ confidence interval is determined by $[\bar{e}_j - s_j\hat{\mu}_{j(k_2)}, \bar{e}_j - s_j\hat{\mu}_{j(k_1)}]$, where $k_1 = B\ \alpha/2$ and $k_2 = B - k_1 + 1$.

Once the confidence intervals are calculated, the assessment of whether or not a value is significant is straightforward. An ERD/ERS value may be considered as significant with $100(1-\alpha)\%$ confidence when both confidence values of this sample show the same sign. That is, an ERS value is significant with, for example, 95% confidence when both 95% confidence limits of this value are positive. And likewise, an ERD value is significant when both confidence limits are negative.

There is one difficulty that arises when attempting to apply the t-percentile bootstrap on a time series since it is designed to capture the statistics of data that are random samples from a distribution. Unfortunately, in a time series the samples are not mutually independent and therefore not random since there is at least a weak dependency between consecutive samples. The block bootstrap (Chernik, 1999) may be used to overcome this problem. In the block bootstrap, blocks of consecutive values of the given data are considered and aligned into a bootstrap sample. Blocks of samples are then resampled to form a new bootstrapped time series. This bootstrap is more difficult to implement, and more computationally intensive than the suggested t-percentile bootstrap. To overcome or at least reduce the independence problem of the t-percentile, the idea of building blocks can be applied by dividing the trials into small overlapping blocks and calculating the mean values of the blocks which are then resampled by the t-percentile bootstrap. This takes the weak dependency of consecutive samples into account.

Spatial Filters

ECoG recordings do not usually require any pre-processing for quantifying ERD/ERS because of their good spatial resolution and high SNR (Huggins et al., 1999; Graimann et al., 2002). By contrast, EEG signals show relatively poor spatial resolution due to the characteristic spatial blurring of its signal, caused by the large distance between signal sources and detectors and volume conduction through scalp, skull, and other layers (Babiloni et al., 2001). In fact, conventional monopolar EEG recordings, which are typically used if the number of electrodes is large, have insufficient spatial resolution to reveal localized ERD/ERS without the application of spatial filtering techniques such as CAR and Laplacian derivations.

In the CAR, the average value of the entire electrode montage (the common average) is subtracted from that of the channel of interest. Because CAR emphasizes signal patterns that are present in a large proportion of the electrode population, it reduces such patterns and thereby functions as a high-pass spatial filter. That is, CAR accentuates components with highly focal distributions (Nunez et al., 1994). For each electrode location, the Laplacian method calculates the second derivative of the instantaneous spatial voltage distribution, and thereby enhances signal patterns

originating in radial sources immediately below each electrode (Nunez et al., 1994). Thus, this method is also a high-pass spatial filter that accentuates localized patterns and moderates more diffuse activity. The value of the Laplacian at each electrode location is calculated by combining the value at that location with the values of a set of surrounding electrodes. The small Laplacian, which is an approximation of the surface Laplacian and is calculated from the four nearest neighbors (Hjorth, 1975), can be seen as the standard pre-processing method in ERD/ERS analysis (Pfurtscheller and Lopes da Silva, 1999). Recently, it was shown that this simple approximation of the surface Laplacian yields equivalent results to spherical spline interpolation (Tandonnet et al., 2005). However, modern signal processing methods like ICA offer an alternative to the conventional spatial filters. In the following, we review and discuss ICA in sufficient detail for ERD/ERS analysis, and we also derive a new spatial filter technique based on CSPs which can be seen as optimal for ERD/ERS analysis.

Independent component analysis

ICA is a statistical signal processing technique that decomposes a multivariate input signal into statistically independent components (Hyvarinen and Oja, 2000). The application of ICA to EEG or other bioelectrical recordings assumes that several conditions are at least approximately fulfilled (Vigario et al., 2000). The most important conditions which are assumed to be met are:

- An EEG signal is a linear mixture of the activities of cerebral and other sources.
- The source signals are statistically independent.
- The mixing of the sources and the sources themselves are stationary.

Since the propagation delays from the sources to the electrodes are negligible and the summation of brain potentials can be considered linear (Jung et al., 2000), the assumption that the mixing is instantaneous is met. The second assumption, statistical independence of the sources, refers solely to the statistical relationships between the probability distributions of the signals involved. It is the basic assumption upon which ICA is based. However, this conjecture does not necessarily imply that the generating neural structures are independent as well. The last assumption, stationarity, is difficult to meet in general because it is well known that EEG is not stationary. In contrast, ICA requires stationary stochastic processes to guarantee a meaningful decomposition of the linearly mixed sources. The easiest way to deal with this problem is to simply ignore the stationarity condition and hope that despite this, ICA will be able to yield reasonable results. Another way is to set up an experimental situation in which the recorded EEG signal can be viewed as stationary. This is in fact the case in the experimental paradigm of ERD/ERS analysis. The controlled conditions of the experimental paradigm in which a task or event (e.g., motor activity) is repeated several times implies that quasi-stationary EEG segments are recorded some seconds prior to and some seconds after the task or event. Therefore, it can be concluded that all three conditions underlying the ICA model are quite reasonable for ERD/ERS analysis.

Owing to the weak assumptions that are made for the ICA model, the order and the scale (including the sign) of the independent components are undetermined. The scale is unimportant for ERD/ERS analysis because it is quantified as a relative measure. However, a completely undetermined order of the components requires analyzing all components because no a priori information about the importance of each component is available. If the number of channels is large, this might become a problem, which could be alleviated by using principal component analysis (PCA) to reduce the dimensionality of the problem before ICA is applied (Jung et al., 2000; Vigario et al., 2000).

In addition to the temporal information of the independent components, ICA also provides information about the topography of the components. The independent components are calculated by $\mathbf{Y} = \mathbf{WX}$, where $\mathbf{X}$ denotes the EEG and $\mathbf{W}$ is the unmixing matrix (the output of the ICA algorithm). The equivalent expression $\mathbf{X} = \mathbf{AY}$,

with **A** the mixing matrix (inverse of **W**), determines the projection of the independent components onto the EEG channels. More specifically, the projection of the *i*th independent component onto the EEG channels is given by the outer product of the *i*th row of **Y** with the *i*th column of the mixing matrix **A**. That is, the columns of **A** are the projection strengths (or weights) of the independent components onto the original signals. This topographic information can be employed to verify the physiological origins of the components.

It has been shown that ICA can be successfully applied to reduce artifacts in EEG (Jung et al., 2000; Vigario et al., 2000). Since artifacts are quite independent from the rest of the signal, ICA is able to separate a wide variety of artifacts from EEG data by linear decomposition. Ocular artifacts, for example, can be easily eliminated by ICA. Muscle activity, however, is more difficult because, in general, the statistical properties of muscle artifacts do not correspond to the basic assumptions of the ICA model. Nonetheless, this capability of ICA has an important implication for the analysis of ERD/ERS. Since the quantification of ERD/ERS requires artifact-free trials, the EEG data has to be visually inspected and trials with artifacts have to be precluded from further analysis. This can result in a considerable loss of information. Since ICA can separate most of the artifacts, a selection of artifact-free trials is not necessary and the full amount of information available (all trials) can be used to quantify ERD/ERS, which can increase the statistical significance of the results.

Maximization of ERD/ERS

Independent component analysis is a so-called unsupervised method. That is, ICA does not use explicit timing information such as which samples of the recording belong to the baseline periods and which belong to the activity periods. In ERD/ERS analysis, however, this information is easily available from the experimental paradigm used. In fact, this information is necessary for the quantification of ERD/ERS. Thus one might want to find a spatial filter that maximizes or minimizes the power ratio in the baseline and activity periods. In other words, to extract ERD we want to find a linear transformation of the data that has maximal power in the baseline period and minimal power in the activity period. For ERS, we want minimal power in the baseline period and maximal power in the activity period. Mathematically, this optimization criterion can be expressed as

$$\mathbf{w} = \arg\max = \frac{\sum\limits_{p_1} y_i^2}{\sum\limits_{p_2} y_i^2} \tag{7}$$

where **w** is the weighting vector (the first row in the spatial filter matrix), and p_k, $k \in [1, 2]$ denotes the baseline or reference period and the activity period, respectively. The general solution for this problem is the principal component of the jointly pre-whitened samples of the two periods under consideration (Fukunaga, 1990), which is also known as the method of CSPs (Koles et al., 1995; Muller-Gerking et al., 1999). A derivation of the solution in the context of ERD/ERS analysis is given in the appendix.

The result of this method is again a spatial filter matrix **W** that decomposes the multivariate EEG signal into components that are optimal (according to the optimization criterion formulated in Eq. (7)) for ERD/ERS analysis. Since the optimization criterion is directly formulated to optimize ERD or ERS, we call this method $\mathrm{ERD}_{\max}$. As with ICA and all the other spatial filter methods discussed so far, the $\mathrm{ERD}_{\max}$ approach does not need reference-free EEG data. It can be applied to monopolar derived EEG or other bioelectrical signals such as ECoG or MEG.

Application of ERD/ERS analysis to movement-related data

To demonstrate how well the Laplacian, ICA, and $\mathrm{ERD}_{\max}$ spatial filters perform in the quantification of ERD/ERS, we present an analysis of a brisk finger movement task. For the sake of clarity of the comparison, we focus here only on the ability of the three different methods to reveal ERD. We will also include in our comparison the result of attempting to produce ERD/ERS maps directly from raw monopolar data to emphasize the

necessity of applying some sort of spatial filtering technique when quantifying ERD/ERS. We will see that of all of the spatial filters investigated, ERD_{max} stands out as the superior approach.

EEG data

The EEG data used for this demonstration was recorded with a sampling rate of 256 Hz from a grid of 34 Ag/AgCl scalp electrodes referenced to the left mastoid (see Fig. 4). The signal was band-pass filtered between 0.5 and 50 Hz.The electrodes were placed around Cz with regular inter-electrode distances of ~2.5 cm.Twelve subjects performed self-paced brisk movements of the right index finger. Each subject performed about 68 finger movements (trials). In addition to the EEG, an EMG signal was recorded to generate a trigger channel indicating movement onset and offset. Details about the experimental paradigm and data recording can be found in Pfurtscheller et al. (2000). Normally trials containing artifacts are excluded from ERD/ERS analysis. For this study, however, all trials were used to test the ability of the algorithms to deal with artifacts.

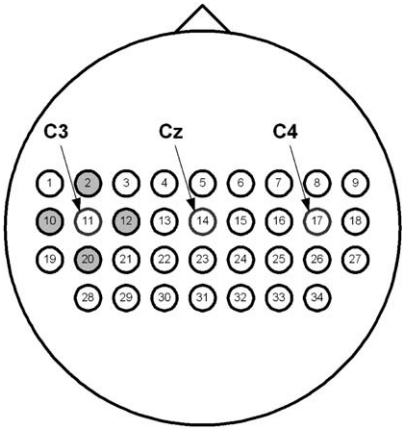

Fig. 4. Electrode positions of the 34-channel montage. C3, Cz, and C4 are located according to the international 10–20 system. The shadowed area gives an example of the electrodes involved in the calculation of the small Laplacian (orthogonal source derivation).

Methods

To demonstrate the difference between orthogonal source derivation (small Laplacian), ICA, and ERD_{max}, the EEG signals were pre-processed in the following three different ways:

1. *Orthogonal source derivation (small Laplacian)*: Since the electrodes were located in a quadratic grid, the orthogonal source derivation was calculated by the following equation:

$$e_i^{\text{lap}} = e_i - \frac{1}{4}\sum_{j \in S_i} e_j,$$

where S_i denotes the four nearest neighbors. That is, to obtain the spatial filtered signal for electrode i, the average of the signals of the neighboring electrodes is subtracted from the signal of electrode i. For boundary electrodes, the three nearest neighbors were used to calculate an approximation of the Laplacian derivation.

2. *Independent component analysis*: A variety of different ICA algorithms are available. For this study, Infomax (Bell and Sejnowski, 1997) was used because it has been proven to be useful in a wide range of biomedical applications (Makeig et al., 1997, 1999; Jung et al., 2001). For this study, we used the implementation of Infomax as provided in the freely available Matlab toolbox EEGlab (Delorme and Makeig, 2004). Before Infomax was applied, the signals were band-pass filtered between 5 and 40 Hz.This was done because band-pass filtering can improve the performance of ICA algorithms (Hyvarinen and Oja, 2000). The result of ICA, the unmixing matrix $\mathbf{W}$, was used to decompose the EEG $\mathbf{X}$ into a sum of temporally independent and spatially fixed independent components,

$$\mathbf{Y} = \mathbf{WX}$$

The independent components (rows of $\mathbf{Y}$) were then subjected to the subsequent ERD/ERS analysis. The projection strengths of the components contained in the columns of the inverse of the unmixing matrix $\mathbf{W}$ together

with the topographic information of the scalp electrodes were used to calculate topographic maps by means of cubic spline interpolation. These maps were individually normalized so that the largest absolute value was 1. This was done not only to increase the color contrast, but also to obtain a single color legend that could be used for all topographic maps.

3. *ERD_{max}*: To achieve ERD/ERS maps with maximal ERD, ERD_{max} was applied to the data according to the algorithm defined in the appendix. The reference period was selected to be from 3.5 to 2 s before trigger offset, which is the same reference period used for calculating the ERD/ERS maps (see below). The optimization time frame for the ERD samples was selected to be 0.5 s prior and 0.5 s after trigger offset. Reference and ERD samples were band-pass filtered between 7 and 13 Hz.Hence, the optimization process was focused on this limited frequency band and the 1-s time period around the trigger. Similar to ICA, the columns of the inverse of the ERD_{max} unmixing matrix were used to calculate topographic maps. Again, the topographic maps were individually normalized to an absolute value of 1 to increase the color contrast.

All channels pre-processed by the three different spatial filters were subjected to ERD/ERS analysis according to the standard procedure outlined in Section "Quantification of ERD/ERS". EEG segments (trials) of a length of 8 s and starting at 4 s prior to movement offset were extracted. ERD/ERS maps comprised frequency bands of 2 Hz with 1-Hz overlap in a frequency range of 7–34 Hz were calculated. Filtering was done in the frequency domain using Hamming windows. Filtering in the frequency domain had the advantage that band-pass filtering and calculation of the Hilbert transform of the filtered signal (by setting the corresponding Hermitian part of the signal to zero) could be combined. By back-transforming the signal in the time domain and calculating the absolute value, the envelope of the band-pass filtered signal was obtained which was then squared to obtain power values. The reference interval was selected to be from 3.5 to 2 s before trigger onset. The *t*-percentile bootstrap with 1000 bootstrap repetitions were performed, to determine ERD/ERS values with a significance of $p = 0.01$.

Results

ERD/ERS analyses for all 12 subjects using monopolar data and data filtered by the three different spatial filters were performed. To conserve space, we show topographically arranged ERD/ERS maps for only one particular subject. This dataset was chosen because not only it showed the characteristic movement-related patterns but also the data were contaminated by artifacts. This provides the opportunity to observe how the algorithms work in the presence of artifacts. Figure 5 shows the ERD/ERS maps that are topographically arranged for monopolar (Fig. 5A) and Laplacian data (Fig. 5B). The spatial resolution has been considerably increased by the application of the Laplacian spatial filter. Pronounced ERD activity that is almost nonexistent in the monopolar data is clearly visible in the maps that are derived from electrodes that are contralateral to the movement side (electrodes C3 and neighboring electrodes), and — although not so prominent — ipsilateral (electrode C4). Laplacian filtering also reveals artifacts in electrodes 1, 7, and the bordering right-hand electrodes.

Figure 6 depicts the ERD/ERS maps for the same subject after ICA spatial filtering (Fig. 6A) and ERD_{max} spatial filtering (Fig. 6B). An important difference here from the maps shown in Fig. 5 is the fact that they are not topographically arranged anymore. Since these two methods do not employ a simple spatial relationship of the EEG channels like the Laplacian derivation, a priori information about the topography of the components is not available. Hence, the number in the lower left corner of the maps does not indicate the channel number, but the component of the corresponding spatial filter method. Consequently, the order in which relevant ERD components can occur is widely arbitrary for ICA, while for ERD_{max}, relevant ERD components should be contained in the first few components, because the

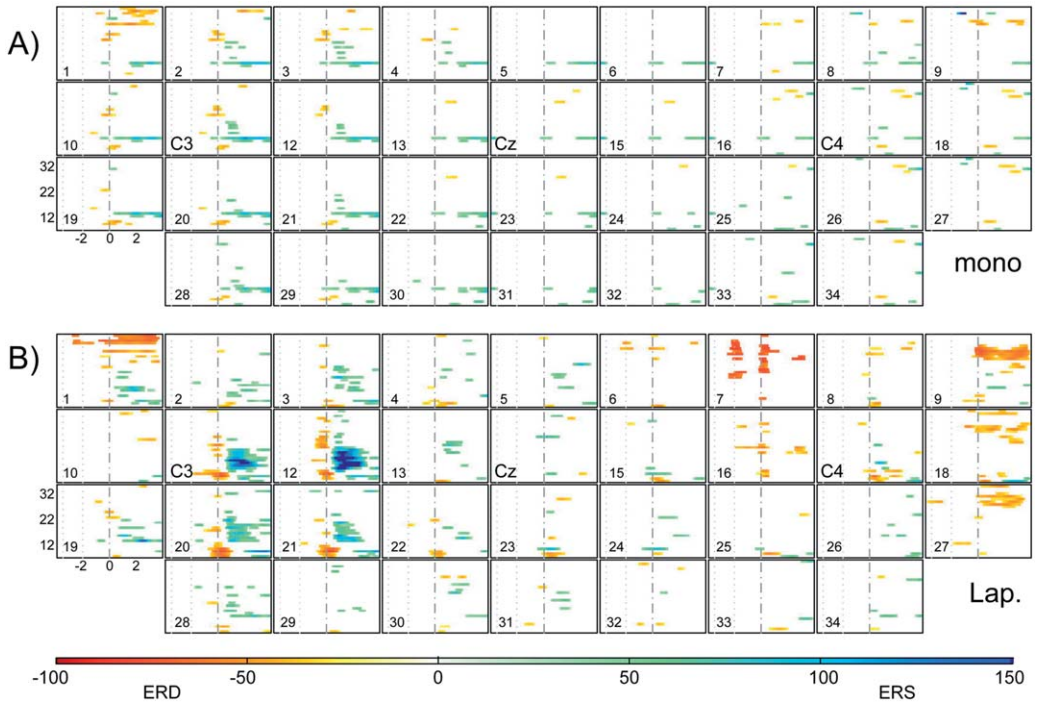

Fig. 5. ERD/ERS maps derived from monopolar data (A) and Laplacian spatially filtered data (B). The maps are topographically arranged. In the cases where the electrode locations are compatible with the international 10/20-placement system, the appropriate electrode names (C3, Cz, and C4) are shown. The vertical dashed-dotted line denotes the movement onset. The dotted lines indicate the reference period. ERD is colored in red; ERS is colored in blue.

ERD_{max} components are sorted according to the value of the eigenvalues that appear in the ERD_{max} algorithm. In fact, for all 12 datasets investigated, the desynchronization pattern contralateral to the movement side, that is the pattern that should be most prominent in a movement-related task like this, was almost always found in the first and in few cases (due to artifacts) in the second ERD_{max} component.

Since the maps generated by ICA and ERD_{max} do not contain information about the topography of the sources of the components, it becomes necessary to produce this missing information in the form of topographic maps. Eight such maps depicting interesting topographic distributions or being associated with components showing prominent ERD/ERS are presented immediately below the respective ERD/ERS maps. They can be used to identify the topographic origin of the corresponding components. In Fig. 6A, for instance, ICA component 18 can be easily identified as the most prominent contralateral activity next to components 3 and 21, which have a similar origin but show less pronounced activity. ICA component 16 is the ipsilateral activity, and components 6, 22, 32, and 34 are artifactual components. Similar results can be found displayed in Fig. 6B for the data spatially filtered according to ERD_{max}. The two most prominent activities (both are contralateral) are found (as expected) in the two first maps. ERD_{max} component 7 shows ipsilateral activity, and components 3, 24, and 34 show artifactual activity. To permit a better comparison between the ERD/ERS maps derived from the three different spatial filters, enlarged ERD/ERS maps that are associated with contralateral activity for the same subject are shown in Fig. 7A, and for a different exemplary dataset in Fig. 7B. The numbers in the lower left corner indicate the channel numbers and the component numbers. In addition

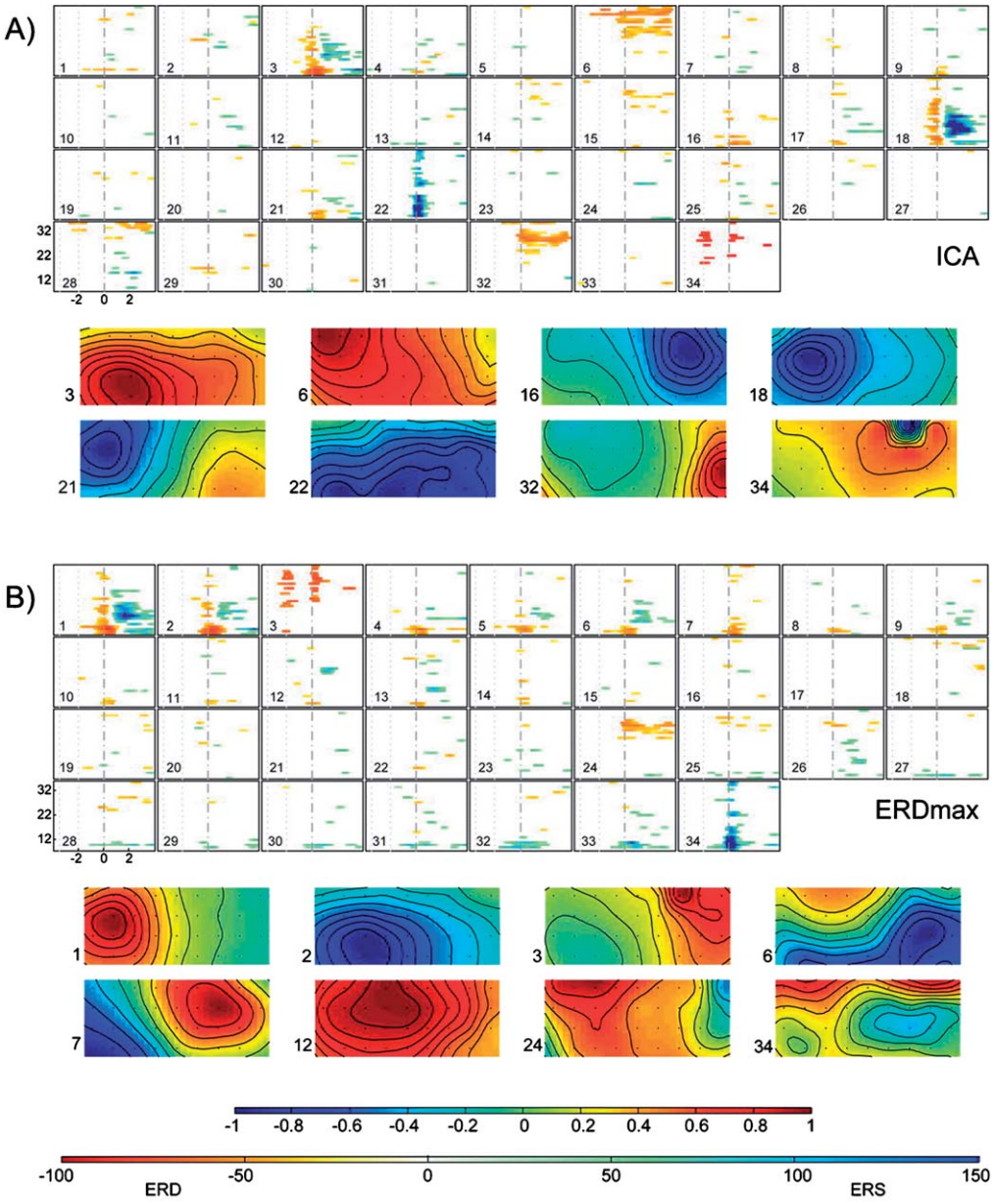

Fig. 6. ERD/ERS maps derived from independent component analysis (ICA, A) and ERD_{max} (B). Since the components are no longer associated directly with the physical locations of the electrodes, normalized topographic maps are provided for a selected number of maps to help establish the source locations of these components. The number on the left-hand side of each topographic map indicates the number of the corresponding ERD/ERS map. Color legends for the ERD/ERS maps and for the topographic maps (small color bar) are shown.

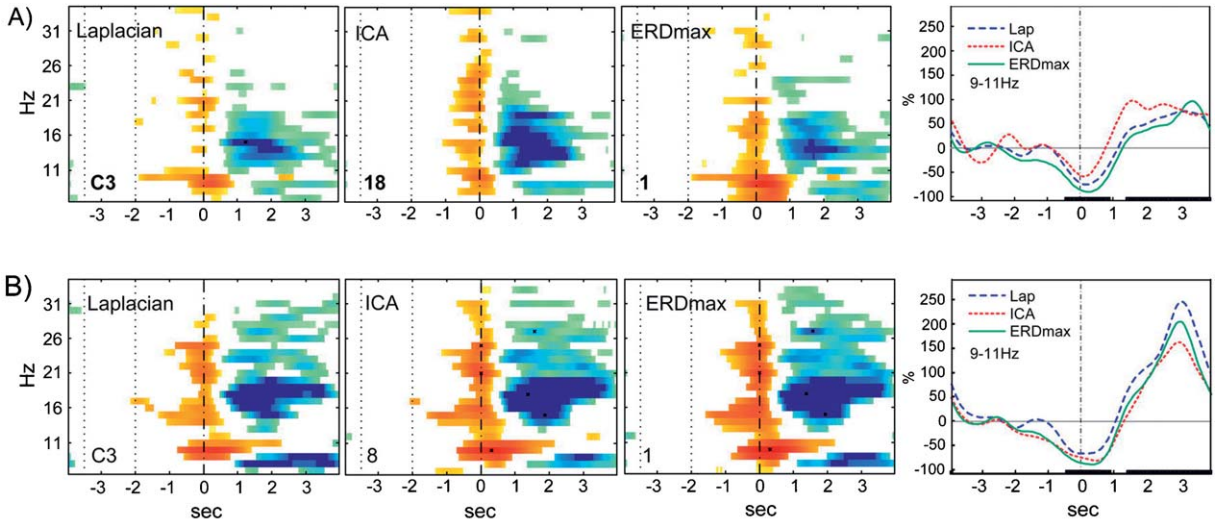

Fig. 7. Enlarged ERD/ERS maps for all three spatial filters for subject S1 (A), subject S2 (B). These maps depict the most prominent contralateral ERD for the corresponding method and subject. The channel numbers for the Laplacian maps and the component numbers for the other spatial filter maps are shown in the lower left corners. ERD/ERS time courses for the 9–11-Hz frequency band are shown on the far right. The color coding is done in the same way as in the previous figures.

to the maps, ERD/ERS time courses for the 9–11 Hz frequency band are displayed. This specific band was chosen because it was in the frequency range used for optimizing the ERD_{max} spatial filter, and it showed pronounced ERD and ERS activity for most of the datasets analyzed.

All three maps in Fig. 7A show basically the same activity patterns over the whole frequency range. ICA produced components are very pronounced in the beta range, but produced only short-lasting ERD components in the alpha range. The ERD/ERS curves show that the ERD indicated for each filtering method is maximal for ERD_{max}. Unexpectedly, Laplacian filtering shows more prominent ERD than ICA filtering does; however, this is an anomaly compared with the rest of the data analyzed. This might be explained in the following way. In the vast majority of the other datasets, ICA decomposed the contralateral activity into only two components, but in this particular case, three components were produced. Since the energy present in the contralateral activity was divided into three components rather than only two, each component contained less energy offering a possible explanation for the reduced ERD.

For the dataset used to produce the maps in Fig. 7B, the situation was different in that both ERD_{max} and ICA decomposed the contralateral activity into only two components. Consequently, the maps for ICA and ERD_{max} look very similar (and also similar to Laplacian). The time courses now show that the ICA curve captures more ERD than the Laplacian. The ERD_{max} curve, however, still captures the greatest amount of ERD activity.

To test if there is a significant difference for the quantification of ERD between the three spatial filters, the grand average of the most prominent contralateral ERD/ERS curves for all 12 datasets in the frequency band from 9 to 11 Hz was calculated, and is shown in Fig. 8. For each curve, the ERD value in the range of 0.5 s before to 0.5 s after the trigger was integrated. The resulting 12 values were used to test the following three null hypotheses: (1) The integrated ICA ERD value is less pronounced or equal to the integrated Laplacian value. (2) The integrated ERD_{max} value is less pronounced or equal to the integrated ICA value. (3) The integrated ERD_{max} value is less pronounced or equal to the integrated Laplacian value. Using a *t*-test, all three hypotheses could be rejected. That is, for the 12 datasets analyzed, the

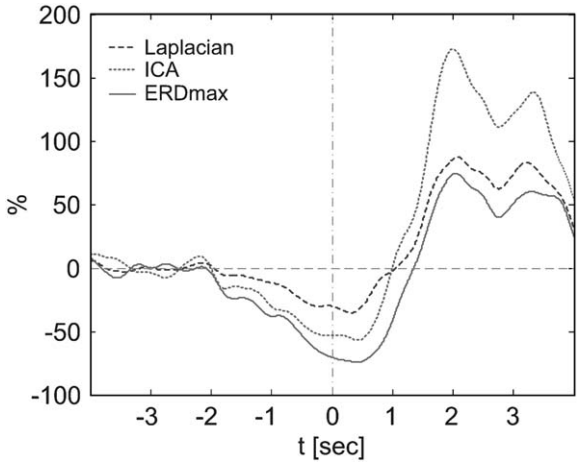

Fig. 8. The grand average over all 12 subjects was combined to produce time courses corresponding to the 9–11-Hz frequency band for all three spatial filters from the most prominent contralateral components.

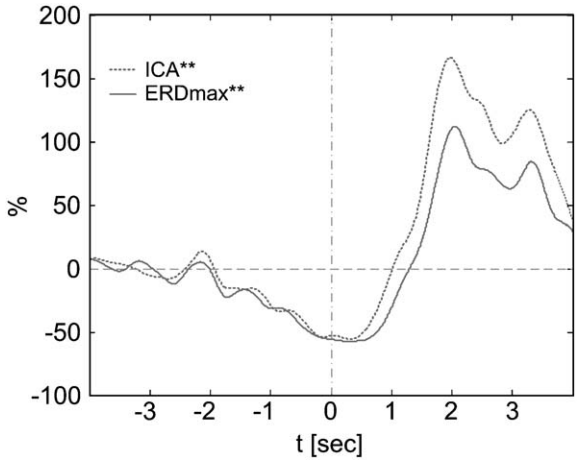

Fig. 9. The grand average over all 12 subjects for the most prominent contralateral components for the spatial filters ICA** and ERD_{max}**. ICA** denotes an ICA spatial filter constructed from pre-filtered data in the narrow frequency band of 7–13 Hz.ERD_{max}** denotes an ERD_{max} spatial filter constructed from pre-filtered data in the 5–40 Hz band.

contralateral ICA components showed more pronounced ERD in this specific frequency range and time period than the signal filtered with the small Laplacian method ($p = 0.032$). The ERD_{max} component showed more pronounced ERD than the ICA component ($p = 0.029$), and the ERD_{max} component showed more pronounced ERD than the corresponding Laplacian signal ($p = 0.003$).

Since the construction of the ERD_{max} spatial filter involves narrow pre-filtering in the frequency band of interest, it is interesting to know how significant this step alone is in yielding the good results produced by the algorithm. To test that question, ICA spatial filters denoted as ICA** were constructed from data that was filtered in the frequency band of 7–13 Hz (the same band used for ERD_{max}). Additionally, ERD_{max}** filters were calculated from data filtered between 5 and 40 Hz (the same band used for ICA). The grand averages of the 9–11 ERD/ERS time courses produced by the application of these spatial filters for all 12 datasets are depicted in Fig. 9. The grand average of the ICA** filtered time courses is very similar to the grand average of the ICA filtered time courses in Fig. 8, demonstrating that narrow band-pass filtering is not required for ICA. In contrast to the result shown in Fig. 8, the grand average achieved with ERD_{max}** is now similar to the result achieved by ICA**. In fact, the same statistical analysis of the integrated ERD values as done before did not reveal any statistical difference between these two spatial filters. Consequently, to maximize the performance of ERD_{max} filters, it is important to filter the data in the frequency band of interest.

Discussion

The ERD/ERS maps derived from all spatial filters investigated shows the characteristic contralateral patterns of mu and beta ERD prior to and during movement followed by postmovement ERS. Compared with the Laplacian, ICA, and ERD_{max} generally produced maps with more localized ERD/ERS. It is especially interesting that both ERD_{max} and ICA seem to decompose the ERD/ERS patterns that are contralateral to the movement side and focused at C3 (channel 11) in the Laplacian maps into two independent components, whereas one of the two components is localized slightly anterior to C3 and the other more posterior to C3. The more posterior component principally represents the mu ERD, whereas the anterior component pre-dominantly contains beta ERS. The separation of mu ERD and beta

ERS is of special interest because equivalent source detection in MEG data revealed the source of the mu rhythm in the post-central gyrus and the source of the central beta oscillation in the pre-central gyrus (Salmelin and Hari, 1994). A similar separation between mu ERD and beta ERS in a self-paced movement task EEG recording was reported in Pfurtscheller et al. (1994).

The ERD/ERS curves in Fig. 8 demonstrate that ERD/ERS patterns of all spatial filter methods are able to reveal similar ERD/ERS time courses. However, the ERD/ERS patterns resulting from ICA and ERD_{max} pre-processing are significantly more prominent than those from Laplacian pre-processing.

The ICA and ERD_{max} ERD/ERS maps should be always interpreted together with their topographic maps. It should be noted, however, that spatial filters cannot solve the inverse problem. The objective of ICA is to decompose the signal into statistically independent signals, and not to identify physically discrete neuronal networks. That precludes the general conclusion that statistically independent signals identified by ICA have their origin in independent cortical areas. Likewise, the objective of ERD_{max} is to find a decomposition of the multivariate signal for which the ratio between the energy in the reference period and the energy in the activity period (where ERD or ERS occurs) is optimal in terms of the optimization criteria, which is directly formulated to maximize ERD. Although the optimization criteria can be extended to simultaneously optimize ERD and ERS, the results presented here suggest that the simpler criteria of optimizing ERD in a specific frequency band are sufficient to achieve very good results. Initially, it seems surprising that optimizing solely ERD without considering ERS reveals patterns that contain both ERD and ERS patterns that are very similar to the ERD/ERS patterns of a Laplacian derivation. The explanation for this, however, is very simple. It is a consequence of the fact that the ERD and ERS patterns of a map are not uncorrelated. That is, ERD/ERS patterns revealed by ERD_{max} have the same statistical source. It is important to note that the term "source" is used here to refer to an uncorrelated component determined by ERD_{max}. A direct relationship of such a component to cortical sources cannot be assumed in general. In other words, as noted earlier for ICA, ERD_{max} cannot be used to solve the inverse problem.

Summary and conclusion

We have reviewed the traditional approach for ERD/ERS quantification and a more recent approach based on wavelet transform, and discussed visualization of these phenomenon and validation of the results through statistical significance testing. For broader frequency ranges such as those found in ECoG data, the wavelet approach is preferable since it easily provides a good compromise between time and frequency resolution. For the narrower frequency range of EEG, this issue is not as critical, and the classical band-pass approach and wavelet approach are equivalent.

We also discussed the importance of verifying the statistical significance of results. Doing this yields clearer maps that reveal only significant results upon which conclusions may be drawn more easily. Since the distribution of ERD/ERS values is unknown and highly skewed, nonparametric statistical methods are required. The solution suggested was a nonparametric bootstrap test. This technique is robust and easily applicable, which permits reliable results to be obtained for even a small number of trials.

Spatial filters are essential for ERD/ERS analysis of EEG data due to the limited spatial resolution. Recently, various multivariate signal methods have been proposed. These methods linearly combine channels to generate improved representation in terms of SNR and spatial resolution. They accomplish this based on specific optimization criteria. We investigated ICA and a method based on CSPs in the context of ERD/ERS analysis. Since the latter can be used to directly maximize ERD or ERS we named this approach ERD_{max}.

Both ICA and ERD_{max} can produce similar results, but because of their different optimization criteria they are not equivalent. ICA decomposes the signal with the goal of achieving statistical independence producing components containing ERD/ERS as only an indirect consequence of this

procedure, while ERD_{max} is directly focused on optimizing ERD/ERS.

From the practical point of view, ERD/ERS analysis with ERD_{max} has the following advantages.

1. The transformation yielded by ERD_{max} can be seen as optimal in terms of maximizing ERD or ERS.
2. The order of the components is fixed. The first component contains most of the ERD or ERS.
3. ERD_{max} can be performed with all channels or a small subset of channels (e.g., the nearest neighbors) independent from the channel topography.
4. The algorithm is very fast.

ICA has already been applied to the analysis of event-related potentials in EEG. Makeig et al. (1997, 1999) have been shown that ICA can effectively decompose multiple overlapping components from selected sets of related ERP averages. Little previous work has been done to investigate the usefulness of ICA in the analysis of oscillatory activity. A complete functional interpretation of the different independent components was not within the scope of this study; however, it seems that both ERD_{max} and ICA could reveal new insights into the functional interpretation of brain oscillations. The preliminary results of this study suggest that further research on this topic should be undertaken.

Appendix

In this section, the new algorithm for finding the optimal linear transformation to maximize ERD/ERS is derived. The goal is to find a matrix $\mathbf{W}$ that transforms the multivariate input $\mathbf{X}$ into components that contain maximal ERD/ERS. In matrix notation:

$$\mathbf{Y} = \mathbf{W}\mathbf{X} \tag{A1}$$

Following the terminology of ICA, $\mathbf{W}$ is the unmixing matrix and $\mathbf{Y}$ is the transformed input signal. The rows of $\mathbf{Y}$ are the components that contain maximal ERD/ERS.

To simplify the derivation, only the maximization of ERD and the projection onto a single vector $\mathbf{w}$ is considered at first. That is, the task is to find a vector $\mathbf{w}$ that linearly combines the input signals so that the resulting component $\mathbf{y}$, where

$$\mathbf{y} = \mathbf{w}^{\mathrm{T}}\mathbf{X}, \tag{A2}$$

contains most of the ERD activity.

For maximizing ERD, the minimum of the following equation has to be found (see also Eq. (3) in Section "Quantification of ERD/ERS").

$$E_{\mathrm{ERD}}(\mathbf{w}) = \frac{A - R}{R} \tag{A3}$$

with

$$A = E\{(\mathbf{w}^{\mathrm{T}}\mathbf{X}_{\mathrm{a}})^2\} \quad \text{and} \quad x_{ij}^{\mathrm{a}} \in S_{\mathrm{a}} \tag{A4}$$

$$R = E\{(\mathbf{w}^{\mathrm{T}}\mathbf{X}_{\mathrm{r}})^2\} \quad \text{and} \quad x_{ij}^{\mathrm{r}} \in S_{\mathrm{r}} \tag{A5}$$

where $E\{\cdot\}$ denotes the expectation value, S_{r} denotes the set of all samples in the reference period, and S_{a} is the set of all samples in the time frame where ERD can be expected (e.g., half a second prior and after the onset of movement tasks).

A further simplification of the problem can be achieved if $\mathbf{X}_{\mathrm{a}}$ and $\mathbf{X}_{\mathrm{r}}$ are jointly whitened (sphered). Whitening means the input matrix is transformed so that the rows are uncorrelated and have unit variances. There are many ways to whiten the input data (Hyvarinen and Oja, 2000). A possible linear whitening transform is given by

$$\mathbf{V} = \mathbf{D}^{-1/2}\mathbf{E}^{\mathrm{T}} \tag{A6}$$

where $\mathbf{D} = diag(d_1 \ldots d_n)$ is the diagonal matrix of the eigenvalues, and $\mathbf{E}$ the matrix whose columns are the unit-norm eigenvectors of the covariance matrix $E\{\mathbf{X}_{\mathrm{a,r}}^{\mathrm{T}}\mathbf{X}_{\mathrm{a,r}}\}$ with $x_{ij}^{\mathrm{a,r}} \in \{S_{\mathrm{a}}, S_{\mathrm{r}}\}$.

After this transformation, the channels are uncorrelated and have unit variance. Thus, the optimization problem can be reduced to minimizing

$$A = E\{(\mathbf{w}^{\mathrm{T}}\mathbf{X}_{\mathrm{a}})^2\} \quad \text{and} \quad x_{ij}^{\mathrm{a}} \in S_{\mathrm{a}} \tag{A7}$$

or equivalently to maximizing

$$R = E\{(\mathbf{w}^{\mathrm{T}}\mathbf{X}_{\mathrm{r}})^2\} \quad \text{and} \quad x_{ij}^{\mathrm{r}} \in S_{\mathrm{r}} \tag{A8}$$

Since only the direction of $\mathbf{w}$ is interesting, the constraint that the norm of $\mathbf{w}$ is unity has to be made.

Then, the constraint optimization problem can be formulated by the technique of the Lagrange method (Luenberger, 1968):

$$L(\mathbf{w},\lambda) = E\{(\mathbf{w}^{\mathrm{T}}\mathbf{X}_{\mathrm{a}})^2\} \; - \lambda(||w|| - 1) \qquad \text{(A9)}$$

The equation consists of the part that is to be optimized and the constraint equation $g(\mathbf{w}) = ||w|| - 1 = \mathbf{w}^{\mathrm{T}}\mathbf{w} - 1$ multiplied by the Lagrange multiplier λ. The constraint optimization problem can be solved by setting the gradient of (A9) to zero.

$$\frac{\partial L(\mathbf{w},\lambda)}{\partial \mathbf{w}} = 2E\{\mathbf{w}^{\mathrm{T}}\mathbf{X}_{\mathrm{a}}^{\mathrm{T}}\mathbf{X}_{\mathrm{a}}\} \; - 2\lambda\mathbf{w}^{\mathrm{T}} = 0 \qquad \text{(A10)}$$

or equivalently

$$\frac{\partial L(\mathbf{w},\lambda)}{\partial \mathbf{w}} = 2E\{\mathbf{X}_{\mathrm{a}}^{\mathrm{T}}\mathbf{X}_{\mathrm{a}}\mathbf{w}\} \; - 2\lambda\mathbf{w} = 0 \qquad \text{(A11)}$$

With $\mathbf{C}_a = E\{\mathbf{X}_{\mathrm{a}}^{\mathrm{T}}\mathbf{X}_{\mathrm{a}}\}$ being the covariance matrix of the ERD samples (samples in the time frame where ERD can occur), Eq. (A11) can be written as:

$$\mathbf{C}_{\mathrm{a}}\mathbf{w} - \lambda\mathbf{w} = (\mathbf{C}_{\mathrm{a}} - \lambda\mathbf{I})\mathbf{w} = 0 \qquad \text{(A12)}$$

Thus, the problem can be solved by calculating the eigenvalues and eigenvectors of the covariance matrix $\mathbf{C}_{\mathrm{a}}$. The eigenvector $\mathbf{w}_{\mathrm{s}}$ corresponding to the smallest eigenvalue is the solution of the problem. In fact, the new transformation is given by combining the whitening transform $\mathbf{V}$ and the eigenvector $\mathbf{w}_{\mathrm{s}}$:

$$\mathbf{w}_{\mathrm{s}}^{\mathrm{T}} \; \mathbf{V} \qquad \text{(A13)}$$

The component that contains maximal ERD can then be calculated by

$$\mathbf{y} = (\mathbf{w}_{\mathrm{s}}^{\mathrm{T}} \; \mathbf{V}) \; \mathbf{X} \qquad \text{(A14)}$$

In fact, the optimization problem described by the Eqs. (A7)–(A12) is equivalent to PCA. That is, the new method can be seen as a two-step procedure. First, the samples of the reference period and the ERD samples are jointly whitened by a whitening transform (any transform that yields uncorrelated data and unity variance can be used). Second, PCA is applied to the ERD samples. The resulting linear transformation that gives a component with maximal ERD is obtained by combining the whitening transform and the eigenvector corresponding to the smallest eigenvalue. Since reference and ERD samples are jointly whitened, ERD maximization can also be achieved by combining the whitening transform and the largest eigenvector of the covariance matrix of $\mathbf{X}_{\mathrm{r}}$ (reference samples). The first approach gives a transformation with minimal variance in the ERD time frame; the latter gives a transformation with maximal variance in the reference period. As all samples were jointly whitened before PCA is applied, both approaches are equivalent. More than one component can be easily derived by using the other eigenvectors. So, for example, the component containing second most ERD activity can be derived by using the eigenvector corresponding to the second smallest eigenvalue of the covariance matrix $\mathbf{X}_{\mathrm{a}}$.

The same derivation can be made for the maximization of ERS. The only difference is that samples of a time frame where ERS is expected have to be used (e.g., samples in a time frame of 1–2 s after movement onset). To increase the SNR, the signal can be band-pass filtered before the optimization procedure is applied. This would result in optimal ERD or ERS time courses for a specific frequency band. Although the derivation presented here considers only one reactive time frame (either ERD or ERS) against the baseline activity in the reference period, a combination of ERD and ERS is also possible. For example, a combined optimization of ERD and ERS in different frequency bands may be useful. Applying the Lagrange method again, the problem could be formulated as

$$L(\mathbf{w},\lambda_1,\lambda_2) = E\{(\mathbf{w}^{\mathrm{T}}\mathbf{X}_{\mathrm{b}})^2\} - \lambda_1 E\{(\mathbf{w}^{\mathrm{T}}\mathbf{X}_{\mathrm{c}})^2\} \\ - \lambda_2(||w|| - 1) \qquad \text{(A15)}$$

where $\mathbf{X}_{\mathrm{b}}$ and $\mathbf{X}_{\mathrm{c}}$ denote the samples of different time frames and/or frequency bands; λ_1 and λ_2 are the Lagrange multipliers. The solution of this equation, however, does not result in a simple eigenvalue problem as before. More sophisticated methods like gradient descent, Newton iteration, or genetic algorithms have to be applied to find a solution for this optimization problem.

The algorithm calculates two matrices: $\mathbf{W}$, the transformation matrix or unmixing matrix, and $\mathbf{A}$, the inverse of $\mathbf{W}$. Similar to the mixing matrix of ICA, the columns of $\mathbf{A}$ can be used to display the topographic distribution of the new components.

References

Babiloni, F., Cincotti, F., Carducci, F., Rossini, P.M. and Babiloni, C. (2001) Spatial enhancement of EEG data by surface Laplacian estimation: the use of magnetic resonance imaging-based head models. Clin. Neurophysiol., 112: 724–727.

Bell, A.J. and Sejnowski, T.J. (1997) The "independent components" of natural scenes are edge filters. Vision Res., 37: 3327–3338.

Chernik, M.R. (1999) Bootstrap Methods: a Practitioner's Guide. Wiley, New York.

Clochon, P., Fontbonne, J.M., Lebrun, N. and Etevenon, P. (1996) A new method for quantifying EEG event-related desynchronization: amplitude envelope analysis. Electroencephalogr. Clin. Neurophysiol., 98: 126–129.

Cohen, L. (1995) Time-Frequency Analysis. Prentice Hall PTR, Englewood Cliffs, NJ.

Davison, A.C. and Hinkley, D.V. (1997) Bootstrap Methods and Their Application. Cambridge University Press, Cambridge, New York, NY, USA.

Delorme, A. and Makeig, S. (2004) EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J. Neurosci. Methods, 134: 9–21.

Durka, P.J., Ircha, D., Neuper, C. and Pfurtscheller, G. (2001) Time-frequency microstructure of event-related electro-encephalogram desynchronisation and synchronisation. Med. Biol. Eng. Comput., 39: 315–321.

Fukunaga, K. (1990) Introduction to Statistical Pattern Recognition. Academic Press, Boston.

Graimann, B., Huggins, J.E., Levine, S.P. and Pfurtscheller, G. (2002) Visualization of significant ERD/ERS patterns in multichannel EEG and ECoG data. Clin. Neurophysiol., 113: 43–47.

Hjorth, B. (1975) An on-line transformation of EEG scalp potentials into orthogonal source derivations. Electroencephalogr. Clin. Neurophysiol., 39: 526–530.

Huggins, J.E., Levine, S.P., BeMent, S.L., Kushwaha, R.K., Schuh, L.A., Passaro, E.A., Rohde, M.M., Ross, D.A., Elisevich, K.V. and Smith, B.J. (1999) Detection of event-related potentials for development of a direct brain interface. J. Clin. Neurophysiol., 16: 448–455.

Hyvarinen, A. and Oja, E. (2000) Independent component analysis: algorithms and applications. Neural Netw., 13: 411–430.

Jung, T.P., Makeig, S., Westerfield, M., Townsend, J., Courchesne, E. and Sejnowski, T.J. (2000) Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects. Clin. Neurophysiol., 111: 1745–1758.

Jung, T.P., Makeig, S., Westerfield, M., Townsend, J., Courchesne, E. and Sejnowski, T.J. (2001) Analysis and visualization of single-trial event-related potentials. Hum. Brain Mapp., 14: 166–185.

Kalcher, J. and Pfurtscheller, G. (1995) Discrimination between phase-locked and non-phase-locked event-related EEG activity. Electroencephalogr. Clin. Neurophysiol., 94: 381–384.

Knösche, T.R. and Bastiaansen, M.C.M. (2002) On the time resolution of event-related desynchronization: a simulation study. Clin. Neurophysiol., 113: 754–763.

Koles, Z.J., Lind, J.C. and Soong, A.C. (1995) Spatio-temporal decomposition of the EEG: a general approach to the isolation and localization of sources. Electroencephalogr. Clin. Neurophysiol., 95: 219–230.

Luenberger, D.G. (1968) Optimization by Vector Space Methods. Wiley, New York.

Makeig, S. (1993) Auditory event-related dynamics of the EEG spectrum and effects of exposure to tones. Electroencephalogr. Clin. Neurophysiol., 86: 283–293.

Makeig, S., Jung, T.P., Bell, A.J., Ghahremani, D. and Sejnowski, T.J. (1997) Blind separation of auditory event-related brain responses into independent components. Proc. Natl. Acad. Sci. U.S.A., 94: 10979–10984.

Makeig, S., Westerfield, M., Jung, T.P., Covington, J., Townsend, J., Sejnowski, T.J. and Courchesne, E. (1999) Functionally independent components of the late positive event-related potential during visual spatial attention. J. Neurosci., 19: 2665–2680.

Muller-Gerking, J., Pfurtscheller, G. and Flyvbjerg, H. (1999) Designing optimal spatial filters for single-trial EEG classification in a movement task. Clin. Neurophysiol., 110: 787–798.

Nunez, P.L., Silberstein, R.B., Cadusch, P.J., Wijesinghe, R.S., Westdorp, A.F. and Srinivasan, R. (1994) A theoretical and experimental study of high resolution EEG based on surface Laplacians and cortical imaging. Electroencephalogr. Clin. Neurophysiol., 90: 40–57.

Pfurtscheller, G. (1988) Mapping of event-related desynchronization and type of derivation. Electroencephalogr. Clin. Neurophysiol., 70: 190–193.

Pfurtscheller, G. and Aranibar, A. (1977) Event-related cortical desynchronization detected by power measurements of scalp EEG. Electroencephalogr. Clin. Neurophysiol., 42: 817–826.

Pfurtscheller, G. and Aranibar, A. (1979) Evaluation of event-related desynchronization (ERD) preceding and following voluntary self-paced movement. Electroencephalogr. Clin. Neurophysiol., 46: 138–146.

Pfurtscheller, G. and Lopes da Silva, F.H. (1999) Event-Related Desynchronization. Elsevier, New York, Amsterdam.

Pfurtscheller, G., Pregenzer, M. and Neuper, C. (1994) Visualization of sensorimotor areas involved in preparation for hand movement based on classification of mu and central beta rhythms in single EEG trials in man. Neurosci. Lett., 181: 43–46.

Pfurtscheller, G., Neuper, C. and Krausz, G. (2000) Functional dissociation of lower and upper frequency mu rhythms in relation to voluntary limb movement. Clin. Neurophysiol., 111: 1873–1879.

Salmelin, R. and Hari, R. (1994) Spatiotemporal characteristics of sensorimotor neuromagnetic rhythms related to thumb movement. Neuroscience, 60: 537–550.

Tallon-Baudry, C. and Bertrand, O. (1999) Oscillatory gamma activity in humans and its role in object representation. Trends Cogn. Sci., 3: 151–162.

Tandonnet, C., Burle, B., Hasbroucq, T. and Vidal, F. (2005) Spatial enhancement of EEG traces by surface Laplacian estimation: comparison between local and global methods. Clin. Neurophysiol., 116: 18–24.

Vigario, R., Sarela, J., Jousmaki, V., Hamalainen, M. and Oja, E. (2000) Independent component approach to the analysis of EEG and MEG recordings. IEEE Trans. Biomed. Eng., 47: 589–593.

Zygierewicz, J., Durka, P.J., Klekowicz, H., Franaszczuk, P.J. and Crone, N.E. (2005) Computationally efficient approaches to calculating significant ERD/ERS changes in the time-frequency plane. J. Neurosci. Methods, 145: 267–276.