Chapter 4
The Analysis of Event-Related Potentials

Marco Congedo

Abstract In this chapter, we provide an introduction to the major methods used for the analysis and classification of Event-Related Potentials (ERPs). We start by considering the problem of estimating ERP ensemble averages in the time domain. An estimator allowing for weights and time shifts for each trial is discussed. Then we consider spatial, temporal and spatio-temporal multivariate filters for improving the estimation, including principal component analysis, the common spatial pattern and blind source separation. Then, we review time-frequency analysis methods. The reader is provided with definitions in order to understand the most commonly used linear and non-linear measures used in the time-frequency domain. We continue with a brief discussion on the importance of the analysis in the spatial domain, including topographic maps and tomographies. Next, we review procedures for applying inferential statistics to ERP studies. Emphasis is given to procedures based on permutation tests, which account for the multiple comparison problem and adapt to the form and degree of correlation between hypotheses. Finally, we consider the problem of classifying ERP single-trials, pointing to recent literature covering the most promising methods currently available, namely, Riemannian geometry, random forests and neural networks.

4.1 Introduction

Event-Related Potentials (ERPs) are a fundamental class of phenomena that can be observed by means of electroencephalography (EEG). They are defined as potential difference fluctuations that are both time-locked and phase-locked to a discrete physical, mental, or physiological occurrence, referred to as the *event*. ERPs are usually described as a number of positive and negative peaks characterized by their polarity, shape, amplitude, latency and spatial distribution on the scalp. All these

M. Congedo (✉)
GIPSA-Lab, Centre National de la Recherche Scientifique (CNRS), Grenoble-INP, Université Grenoble Alpes, Grenoble, France
e-mail: marco.congedo@gmail.com

characteristics depend on the type (class) of event. Each realization of an ERP is named a *sweep* or trial. Important pioneering discoveries of ERPs include the contingent negative variation [96], the P300 [88], the mismatch negativity [68] and the error-related negativity [31]. Another class of time-locked phenomena are the Event-Related De/Synchronizations (ERDs/ERSs, [77]), which are *not* phase-locked. In order to keep a clear distinction between the two, ERD/ERS are referred to as *induced* phenomena, while ERPs are referred to as *evoked* phenomena [89]. Traditionally, ERPs have been conceived as stereotypical fluctuations with approximately fixed polarity, shape, latency, amplitude and spatial distribution. Accordingly, the ERP fluctuations are independent from the ongoing EEG and superimpose to it in a time- and phase-locked fashion with respect to the triggering event. This yields the so-called *additive* generative model. Several observations have challenged this model [23], suggesting the possibility that evoked responses may be caused by a process of *phase resetting*, that is, an alignment of the phase of the spontaneous neuronal activity with respect to the event [44, 57, 62]. According to this model, ERPs result from time/frequency modulations of the ongoing activity of specific neuronal populations. Still another generative model of ERPs was introduced by [65] and [69]. These authors pointed out that ongoing EEG activity is commonly non-symmetric around zero, as can be seen clearly in sub-dural recordings of alpha rhythms [58]. They proposed that averaging amplitude-asymmetric oscillations may create evoked responses with slow components.

In this chapter, we consider several major methods currently used to analyze and classify ERPs. In modern EEG, using a multitude of electrodes is the rule rather than the exception, thus emphasis is given on multivariate methods, since these methods can exploit spatial information and achieve higher signal-to-noise ratio (SNR) as compared to single-electrode recordings. We consider the analysis in the time domain, in the time-frequency domain and in the spatial domain. We also consider inter-trial amplitude and latency variability as well as the case of overlapping ERPs. We then consider useful tools for inferential statistics and classifiers for machine learning specifically targeting ERP data. All the time-domain methods described in this chapter are implicitly based on the additive model, but they may give meaningful results even if the data is generated under other models. Time-frequency domain methods can explicitly study the phase consistency of ERP components. We will show an example analysis for each section. The real data examples in all but the last figure concerns a visual P300 experiments where healthy adults play a brain-computer interface video-game named *Brain Invaders* [21]. This experiment is based on the classical oddball paradigm and yields ERPs pertaining to a *target class,* evoked by infrequent stimuli, and a *non-target class*, evoked by frequent stimuli.

4.2 General Considerations in ERP Analysis

ERP analysis is always preceded by a pre-processing step in which the data is digitally filtered. Notch filters for suppressing power line contamination and band-pass filters are common practice to increase the SNR and remove the direct current level [61]. If the high-pass margin of the filter is lower than 0.5 Hz, the direct current level can be eliminated by subtracting the average potential (baseline) computed on a short window before the ERP onset (typically 250 ms long). Researchers and clinicians are often unaware of the signal changes that can be introduced by a digital signal filter, yet the care injected in this pre-processing stage is well rewarded, since severe distortion in signal shape, amplitude, latency and even scalp distribution can be introduced by an inappropriate choice of digital filter [98].

There is consensus today that for a given class of ERPs only the polarities of the peaks may be considered consistent for a given electrical reference used in the EEG recording; the shape, latency, amplitude and spatial distribution of ERPs are highly variable among individuals. Furthermore, even if within each individual the shape may be assumed stable on average, there may be a non-negligible amplitude and latency inter-sweep variability. Furthermore, the spatial distribution can be considered stable within the same individual and within a recording session, but may vary from session to session, for instance, due to slight differences in electrode positioning. Inter-sweep variability is caused by the combination of several experimental, biological and instrumental factors. Experimental and biological factors may affect both latency and amplitude. Examples of experimental factors are the stimulus intensity and the number of items in a visual search task [61]. Examples of biological factors are the subject's fatigue, attention, vigilance, boredom and habituation to the stimulus. Instrumental factors mainly affect the latency variability; the ERP marking on the EEG recording may introduce a jitter, which may be non-negligible if the marker is not recorded directly on the EEG amplification unit and appropriately synchronized therein, or if the stimulation device features a variable stimulus delivery delay. An important factor of amplitude variability is the ongoing EEG signal; large artifacts and high energy background EEG (such as the posterior dominant rhythm) may affect differently the sweeps, depending on their amplitude and phase, artificially enhancing or suppressing ERP peaks.

Special care in ERP analysis must be undertaken when we record overlapping ERPs, since in this case simple averaging results in biased estimations [85, 99, 100]. ERPs are non-overlapping if the minimum inter-stimulus interval (ISI) is longer than the length of the latest recordable ERP. There is today increasing interest in paradigms eliciting overlapping ERPs, such as some odd-ball paradigms [21] and rapid image triage [104], which are heavily employed in brain-computer interfaces for increasing the information transfer rate [101] and in the study of eye-fixation potentials, where the "stimulus onset" is the time of an eye fixation and saccades follow rapidly [86]. The strongest distortion is observed when the ISI is fixed. Less severe is the distortion when the ISI is drawn at random from an exponential distribution [21, 85].

Amplitude/latency inter-sweep variability as well as the occurrence of overlapping ERPs call for specific analysis methods. In general, such methods result in an improved ensemble average estimation. For a review of such methods, the reader is referred to Congedo and Lopes da Silva [23].

4.3 Time Domain Analysis

The main goal of the analysis in the time domain is to estimate the *ensemble average* of several sweeps and characterize the ERP peaks in terms of amplitude, shape and latency. Using matrix algebra notation, we will denote by $x(t)$, the column vector holding the multivariate EEG recording at N electrodes and at time sample t, whereas N × T matrix X_k will denote a data epoch holding the kth observed sweep for a given class of ERP signals. These sweeps last T samples and start at event time ± an offset that depends on the ERP class. For instance, the ERPs and ERDs/ERSs follow a visual presentation but precede a button press. The sweep onset must therefore be set accordingly adjusting the offset. We will assume along this chapter that T > N, i.e., that the sweeps comprise more samples than sensors. We will index the sweeps for a given class by $k \in \{1, \ldots, K\}$, where K is the number of available sweeps for the class under analysis.

4.3.1 The Additive Generative Model

The *additive* generative model for the observed sweep of a given class can be written as

$$X_k = \sigma_k Q(\tau_k) + N_k, \tag{4.1}$$

where Q is an N × T matrix representing the stereotypical evoked responses for the class under analysis, σ_k are positive scaling factors accounting for inter-sweep variations in the amplitude of Q, τ_k are time-shifts, in samples units, accounting for inter-sweep variations in the latency of Q and N_k are N × T matrices representing the noise term added to the kth sweep. Here by 'noise' we refer to all non-evoked activity, including ongoing and induced activity, plus all artifacts. According to this model, the evoked response in Q is continuously modulated in amplitude and latency across sweeps by the aforementioned instrumental, experimental and biological factors. Therefore, the single-sweep SNR is the ratio between the variance of $\sigma_k Q(\tau_k)$ and the variance of N_k. Since the amplitude of ERP responses on the average is in the order of a few μV, whereas the noise is in the order of several tens of μV, the SNR of single sweeps is very low. The classical way to improve the SNR is averaging several sweeps. This enhances evoked fluctuations by constructive interference, since they are the only time- and phase-locked fluctuations.

4.3.2 Ensemble Average Estimations

The usual *arithmetic ensemble average* of the K sweeps is given by

$$\bar{X} = \tfrac{1}{K} \sum_{k=1}^{K} X_k. \tag{4.2}$$

This estimator is unbiased if the noise term is zero-mean, uncorrelated to the signal, spatially and temporally uncorrelated and stationary. It is actually optimal if the noise is also Gaussian [56]. However, these conditions are never matched in practice. For instance, EEG data are both spatially and temporally correlated and typically contain outliers and artifacts, thus are highly non-stationary. As a rule of thumb, the SNR of the arithmetic ensemble average improves proportionally to the square root of the number of sweeps. In practice, it is well known that the arithmetic mean is an acceptable ensemble average estimator provided that sweeps with low SNR are removed and that enough sweeps are available. A better estimate is obtained by estimating the weights σ_k and shift τ_k to be given to each sweep before averaging. The resulting *weighted and aligned arithmetic ensemble average* is given by

$$\bar{X} = \frac{\sum_{k=1}^{K} (\sigma_k X_k(\tau_k))}{\sum_{k=1}^{K} \sigma_k}. \tag{4.3}$$

Of course, with all weights equal and all time-shifts equal to zero, ensemble average estimation (4.3) reduces to (4.2). Importantly, when ERP overlaps, as discussed above, estimators (4.2) or (4.3) should be replaced by a multivariate regression version, which is given by (1.9) in Congedo et al. [22].

4.3.3 Multivariate Filtering Methods

A large family of multivariate methods have been developed with the aim of improving the estimation of ERP ensemble averages by means of spatial, temporal or spatio-temporal filtering. These filters transform the original time-series of the ensemble average in a number of components, which are linear combinations of the original data. A spatial filter outputs components in the form of time-series, which are linear combinations of sensors for each sample, along with the spatial patterns corresponding to each component. A temporal filter outputs components in the form of spatial maps, which are linear combinations of samples for each sensor, along with the temporal patterns corresponding to each component. A spatio-temporal filter outputs components that are linear combinations of sensor and samples at the same time, along with the corresponding spatial and temporal patterns. Given an ensemble average estimation such as in (4.2) or (4.3), the output of the spatial, temporal, and spatio-temporal filters are the components given by

$$\begin{cases} \bar{Y} = B^T \bar{X} & \text{spatial} \\ \bar{Y} = \bar{X} D & \text{temporal} \\ \bar{Y} = B^T \bar{X} D & \text{spatio - temporal} \end{cases} \qquad . \qquad (4.4)$$

For both the $N \times P$ spatial filter matrix B and the $T \times P$ temporal filter matrix D, we require $0 < P < N$, where P is named the *subspace dimension*. The upper bound for P is due to the fact that for our data $N < T$ and that filtering is achieved effectively by discarding from the ensemble average the N-P components not accounted for by the filters, that is, at least one component must be discarded. The task of a filter is indeed to decompose the data in a small number of meaningful components so as to suppress noise while enhancing the relevant signal. Once designed the matrices B and/or D, the filtered ensemble average estimation is obtained by projecting back the components onto the sensor space, as

$$\begin{cases} \bar{X}' = AB^T \bar{X} & \text{spatial} \\ \bar{X}' = \bar{X} D E^T & \text{temporal} \\ \bar{X}' = AB^T \bar{X} D E^T & \text{spatio - temporal} \end{cases} \qquad , \qquad (4.5)$$

where $N \times P$ matrix A and $T \times P$ matrix E are readily found so as to verify

$$B^T A = E^T D = I. \qquad (4.6)$$

In the spatio-temporal setting the columns of matrix A and E are the aforementioned spatial and temporal patterns, respectively. In the spatial setting, only the spatial patterns in A are available, however the components in the rows of $\bar{Y}$ (spatial) in (4.4) will play the role of the temporal patterns. Similarly, in the temporal setting, only the temporal patterns in E are available, however the components in the columns of $\bar{Y}$ (temporal) in (4.4) will play the role of the spatial patterns. So, regardless the type of chosen filter, in this kind of analysis it is customary to visualize the spatial patterns in the form of scalp topographic or tomographic maps and the temporal pattern in the form of associated time-series. This way one can evaluate the spatial and/or temporal patterns of the components that should be retained and those that should be discarded so as to increase the SNR. Nonetheless, we stress here that in general these patterns bear no physiological meaning. A notable exception are the patterns found by the family of blind source separation methods, discussed below, which, under a number of assumptions, allow such interpretation.

4.3.4 Principal Component Analysis

Principal component analysis (PCA) has been the first multivariate filter of this kind to be applied to ERP data [28, 45] and has been often employed [12, 27, 51]. A long-lasting debate has concerned the choice of the spatial vs. temporal PCA [27, 79],

however we hold here that the debate is resolved by performing a spatio-temporal PCA, combining the advantages of both. The PCA seeks uncorrelated components maximizing the variance of the ensemble average estimation (4.2) or (4.3); the first component explains the maximum of its variance, while the remaining components explain the maximum of its remaining variance, subjected to being uncorrelated to all the previous. Hence, the variance explained by the N-P discarded components explains the variance of the 'noise' that has been filtered out by the PCA. In symbols, the PCA seeks matrices B and/or D with orthogonal columns so as to maximize the variance of $\bar{X}'$. Note that for any choice of $0 < P < N$, the filtered ensemble average estimator $\bar{X}'$ obtained by PCA is the best P-rank approximation to $\bar{X}$ in the least-squares sense, i.e., for any $0 < P < N$, the matrices B and/or D as found by PCA attain the minimum variance of $\bar{X} - \bar{X}'$.

The PCA is obtained as it follows: let

$$\bar{X} = U W V^{T} \tag{4.7}$$

be the singular-value decomposition of the ensemble average estimation, where $N \times T$ matrix W holds along the principal diagonal the N non-null singular values in decreasing order ($w_1 \geq \cdots \geq w_N$) and where $N \times N$ matrix U and $T \times T$ matrix V hold in their columns the left and right singular vectors, respectively. Note that the columns of U and V are also the eigenvectors of $\bar{X}\bar{X}^{T}$ and $\bar{X}^{T}\bar{X}$, respectively, with corresponding eigenvalues in both cases being the square of the singular values in W and summing to the variance of $\bar{X}'$. The spatial PCA is obtained filling B with the first P column vectors of U, the temporal PCA is obtained filling D with the first P column vectors of V and the spatio-temporal PCA is obtained filling them both. The appropriate version of (4.4) and (4.5) then applies to obtain the components and the sought filtered ensemble average estimation, respectively. In all cases $0 < P < N$ is the chosen subspace dimension. Note that since for PCA the vectors of the spatial and/or temporal filter matrix are all pair-wise orthogonal, (4.6) is simply verified by setting $A = B$ and/or $E = D$.

An example of spatio-temporal PCA applied to an ERP data set is shown in Fig. 4.1, using estimator (4.2) in the second column and estimator (4.3) in the fourth column. The ERP of this subject features a typical N1/P2 complex at occipital locations and an oscillatory process from about 50–450 ms, better visible at central and parietal location, ending with a large positivity peaking at 375 ms (the "P300"). We see that by means of only four components the PCA effectively compresses the ERP, retaining the relevant signal; however, eye-related artefacts are also retained (see traces at electrodes FP1 and FP2). This happens because the variance of these artefacts is very high, thus as long as the artefacts are somehow spatially and temporally consistent across sweeps, they will be retained in early components along with the consistent (time and phase-locked) ERPs, even if estimator (4.3) is used. For this reason, artefact rejection is generally necessary before applying a PCA.
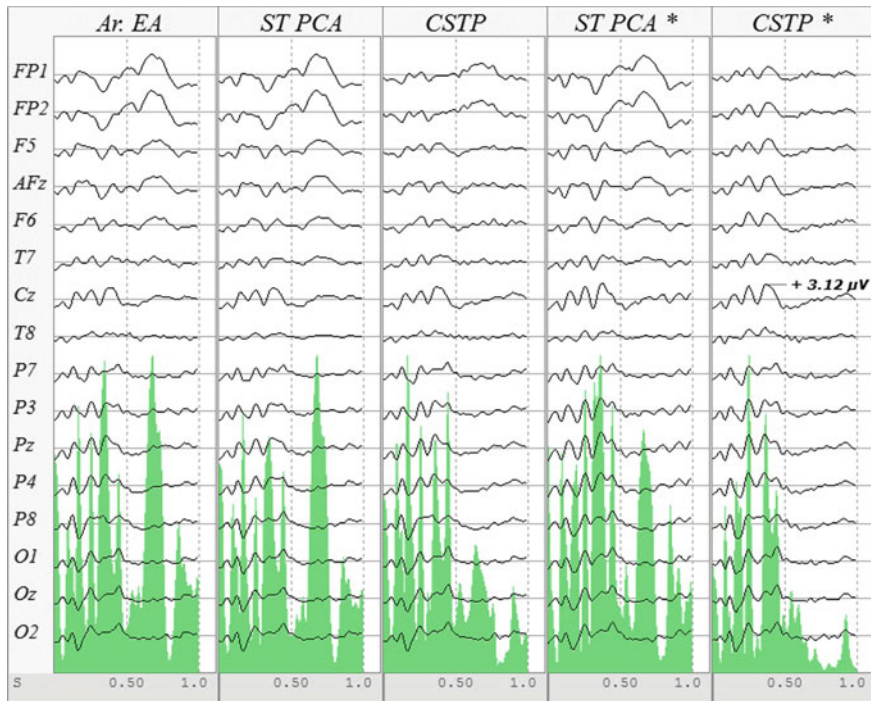
Fig. 4.1 Comparison of several filtered ensemble average estimations via (4.5) using several spatio-temporal filtering methods. One second of data starting at target (infrequent stimulus) presentation averaged across 80 sweeps is displayed. No artifact rejection was performed. The green shaded area is the global field power (Lehmann and Skrandies [52] in arbitrary units, *Legend* "Ar. EA" = non-filtered arithmetic mean ensemble average given by (4.2). "ST PCA" = spatio-temporal PCA with P = 4. "CSTP" = CSTP with P = 12; These two filters have been applied to estimator (4.2). "*" The filters are applied on the weighted and aligned estimator (4.3) using the adaptive method of Congedo et al. [22]. All plots have the same horizontal and vertical scales

4.3.5 The Common Pattern

In order to improve upon the PCA we need to define a measure of the SNR, so that we can devise a filter maximizing the variance of the evoked signal, like PCA does, while also minimizing the variance of the noise. Consider the average spatial and temporal sample covariance matrix when the average is computed across all available sweeps, such as

$$S = \frac{1}{K} \sum_{k=1}^{K} COV(X_k), \ T = \frac{1}{K} \sum_{k=1}^{K} COV\left(X_k^T\right) \tag{4.8}$$

and the covariance matrices of the ensemble averages, namely,

$$\bar{S} = COV(\bar{X}), \ \bar{T} = COV(\bar{X}^T). \tag{4.9}$$

The quantities in (4.8) and (4.9) are very different; in fact S and T hold the covariance of all EEG processes that are active during the sweeps, regardless the fact they are time and phase-locked or not, while in $\bar{S}$ and $\bar{T}$ the non-phase-locked signals have been attenuated by computing the ensemble average in the time domain. That is to say, referring to model (4.1), S and T contain the covariance of the signal plus the covariance of the noise, whereas $\bar{S}$ and $\bar{T}$ contain the covariance of the signal plus an attenuated covariance of the noise. A useful definition of the SNR for the filtered ensemble average estimation is then

$$\text{SNR}(\bar{X}') = \frac{\text{VAR}(AB^T\bar{X}DE^T)}{\frac{1}{K}\sum_{k=1}^{K}\text{VAR}(AB^T\bar{X}_k DE^T)}. \tag{4.10}$$

The common spatio-temporal pattern (CSTP), presented in Congedo et al. [22], is the filtering method maximizing this SNR. It can be used as well when the data contains several classes of ERPs. The sole spatial or temporal common pattern approaches are obtained as special cases. Both conceptually and algorithmically, the CSTP can be understood as a PCA performed on whitened data. So, the PCA can be obtained as a special case of the CSTP by omitting the whitening step. The reader is referred to Congedo et al. [22] for all details and reference to available code libraries. An example of CSTP is shown in Fig. 4.1. In contrast to the spatio-temporal PCA, the CSTP has removed almost completely the eye-related artefact. The last two plots in Fig. 4.1 show the filtered ensemble average estimation obtained by spatio-temporal PCA and CSTP using the adaptive method presented in Congedo et al. [22] for estimating the weights and shift so as to use (4.3) instead of (4.2); the CSTP estimator is even better in this case, as residual eye-related artefacts at electrodes FP1 and FP2 have been completely eliminated.

4.3.6   Blind Source Separation

Over the past 30 years, *Blind Source Separation* (BSS) has established itself as a core methodology for the analysis of data in a very large spectrum of engineering applications such as speech, image, satellite, radar, sonar, antennas and biological signal analysis [17]. In EEG, BSS is often employed for denoising/artifact rejection (e.g., [26]) and in the analysis of continuously recorded EEG, ERDs/ERSs and ERPs. Traditionally, BSS operates by spatially filtering the data. Therefore, it can be casted out in the framework of spatial filters we have previously presented, that is, using the first of the three expressions in (4.4) and (4.5). We have seen that PCA and the common pattern filter seek abstract components optimizing some criterion: the signal variance for PCA and an SNR for the common pattern. In contrast, BSS aims at estimating the *true brain dipolar components* resulting in the observed scalp measurement. For

doing so, BSS makes a number of assumptions. The common one for all BSS methods is that the observed EEG potential results from an instantaneous linear mixing of a number of cortical dipolar electric fields. Although this is an approximation of the physical process of current generation in the brain and diffusion through the head [70], physical and physiological knowledge support such generative model for scalp potentials [9]. In particular, the model fits well low-frequency electrical phenomena with low spatial resolution, which yield the strongest contribution to the recordable EEG. The model reads

$$x(t) = As(t), \tag{4.11}$$

where, as before, $x(t)$ is the observed N-dimensional sensor measurement vector, $s(t)$ the unknown P-dimensional vector holding the true dipolar *source process* (with $0 < P \le N$), and A, also assumed unknown in BSS, is named the *mixing matrix*. BSS entails the estimation of a *demixing matrix B* allowing source process estimation

$$y(t) = B^T x(t). \tag{4.12}$$

We say that the source process can be *identified* if

$$y(t) \approx Gs(t), \tag{4.13}$$

where $P \times P$ matrix $G = B^T A$ is a scaled permutation matrix, i.e., a square matrix with only one non-null element in each row and each column. Matrix G cannot be observed since A is unknown. It enforces a shuffling of the order and amplitude (including possible sign switching) of the estimated source components, which cannot be solved by BSS. Equation (4.13) means that in BSS the actual *waveform* of the source process has been approximately identified, albeit the sign, scaling and order of the estimated source process is arbitrary. Such identification is named *blind* because no knowledge on the source waveform $s(t)$ nor on the mixing process A is assumed. Fortunately, condition (4.13) can be achieved under some additional assumptions relating to the statistical properties of the dipolar source components (see [11, 78]).

Two important families of BSS methods operate by canceling inter-sensor second order statistics (SOS) or higher (than two) order statistics (HOS); the latter family being better known as *independent component analysis* (ICA) (see [17], for an overview). In doing so, both assume some form of *independence among the source processes*, which is specified by inter-sensor statistics that are estimated from the data. The difference between the two families resides in the assumption about the nature of the source process; since Gaussian processes are defined exhaustively by their mean and variance (SOS), ICA may succeed only when at most one of the components is Gaussian. On the other hand, SOS methods can identify the source process components regardless of their distribution, i.e., even if they are all Gaussian, but source components must have a unique power spectrum signature and/or a unique pattern of energy variation across time, across experimental conditions or, in the case of ERPs, across ERP classes (see [20, 24]). For HOS methods the available EEG

can be used directly as input of the algorithms [26]. For BSS methods, either lagged covariance matrices or Fourier co-spectral matrices are estimated on the available data, then the demixing matrix B is estimated as the approximate joint diagonalizer of all these matrices [20]. Details on the application of BSS methods to ERP data can be found in Congedo et al. [24].

Figure 4.2 shows the result of a SOS-based BSS analysis applied to P300 data; here the ensemble averages have been aligned using the method described by Congedo et al. [22]. Analyzing both the temporal course and spatial distribution, we see that the BSS analysis finds two relevant source components: S7 features a topographic map (spatial pattern) with maximum at the vertex and an ERP (temporal pattern) with maximum at 370 ms, clearly describing the P300. S13 features a topographic map with maximum at parietal and occipital bilateral derivations and an ERP with the classical P100/N200 complex describing a visual ERP. Both source components are present only in the target sweeps. Further analysis of these components will be presented in the Sect. 4.4. time-frequency domain analysis. Clearly, BSS has successfully separated the two ERP components.

It is worth mentioning that while traditionally only spatial BSS is performed, a spatio-temporal BSS method for ERPs has been presented in Korczowski et al. [49]. Just as in the case of PCA and common pattern, a spatio-temporal approach is preferable for ERP analysis, thus it should be pursued further (Fig. 4.2).

4.4 Time-Frequency Domain Analysis

Time-Frequency Analysis (TFA) complements and expands the time domain analysis of ERP thanks to a number of unique features. While the analysis in the time domain allows the study of phase-locked ERP components only, TFA allows the study of both phase-locked (evoked) and non-phase-locked (induced) ERP components. In addition to timing, the TFA provides information about the *frequency* (both for evoked and induced components) and about the *phase* (evoked components only) of the underlying physiological processes. This is true for the analysis of a single time series (univariate) as well as for the analysis of the dependency between two time-series (bivariate), the latter not being treated here. In all cases, the time series under analysis may be the sweeps derived at significant scalp derivations or BSS source components with specific physiological meaning as obtained by the methods we have discussed above. In this section we introduce several univariate TFA measures.

A time-frequency analysis (TFA) decomposes a signal in a two dimensional plane, with one dimension being the *time* and the other being the *frequency*. Whereas several possible time-frequency representations exist, nowadays in ERP studies we mainly encounter *wavelets* [50, 89] or the *analytic signal* resulting from the *Hilbert transform* [13, 84, 90]. Several studies comparing wavelets and the Hilbert transform have found that the two representations give similar results [8, 53].

The example we provide below employs the Hilbert transform [37], which is easily and efficiently computed by means of the fast Fourier transform [64]. By applying
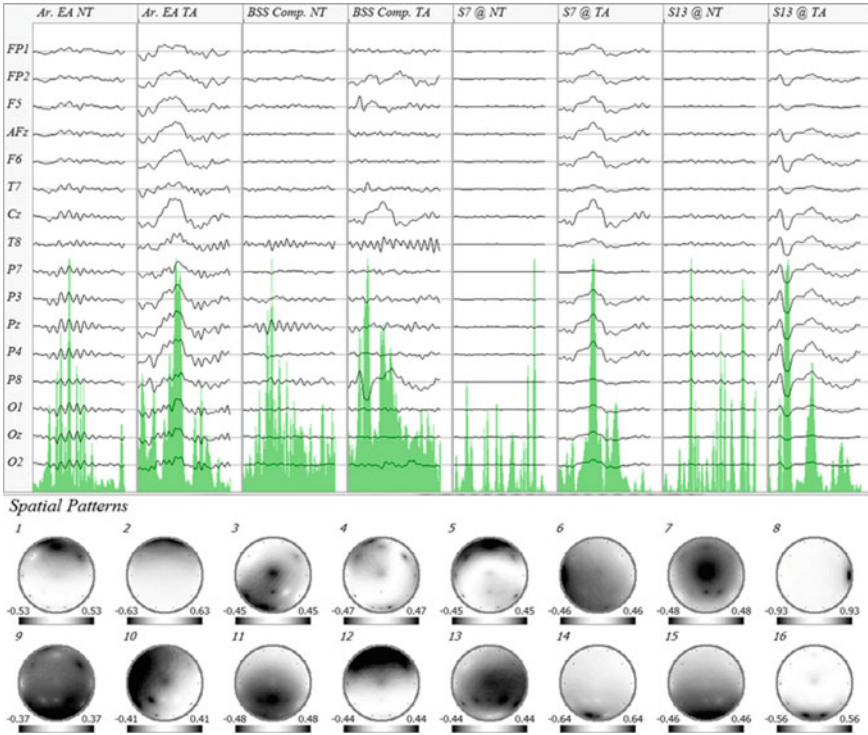
Fig. 4.2 SOS-based blind source separation of ERP. From left to right of the top panel: the weighted and aligned ensemble average (4.3) of the non-target sweeps (Ar. EA NT) and of the target sweeps (Ar. EA TA), the BSS components for non-target (BSS Comp. NT) and target (BSS Comp. TA) ensemble average (obtained via (4.4), first expression), the same filtered ensemble average retaining source component 7 for the non-target (S7 @ NT) and target sweeps (S7 @ TA) and the filtered ensemble average obtained retaining source component 13 for the non-target (S13 @NT) and target sweeps (S13 @ TA). +: arbitrary vertical units for each trace. The bottom panel shows the spatial patterns (columns of the inverse of matrix B) of the BSS components in the form of monochromatic topographic maps. The sign of the potential is arbitrary in BSS analysis. Each map is scaled to its own maximum. Note the separation of two source components: S7 which accounts for the P300, with maximum at the vertex and an ERP with maximum at 370 ms, and S13, which accounts for the classic P100/N200 visual ERP, with maximum at parietal and occipital bilateral derivations. As expected, both source components are present only in the target sweeps, whereas other components are visible in both the target and non-target sweeps

a filter bank to the signal, that is, a series of band-pass filters centered at successive frequencies f (for example, centered at 1 Hz, 2 Hz, …) and by computing the Hilbert transform for each filtered signal, we obtain the analytic signal in the time-frequency representation. Each time-frequency point of the *analytic signal* is a complex number $z_{tf} = a_{tf} + ib_{tf}$ (Fig. 4.3). For each sample of the original signal we obtain from z_{tf} the *instantaneous amplitude* r_{tf}, also known as the *envelope*, as its *modulus* $r_{tf} = |z_{tf}|$ and the *instantaneous phase* φ_{tf} as its *argument* $\varphi_{tf} = \mathrm{Arg}\,(z_{tf})$. The amplitude r_{tf}
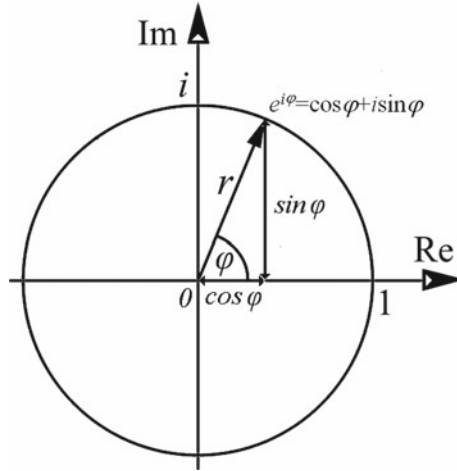
Fig. 4.3 In the complex plane the abscissa is the real line and the ordinate is the imaginary line endowed with the imaginary unit i, which is defined as $i^2 = -1$. A complex number can be represented in Cartesian form as the point $z = a + ib$ in such plane, where a is the real coordinate and ib is the imaginary coordinate. The point can be represented also by a position vector, that is, the vector joining the origin and the point, with length r and angle φ (in the left part of the figure the point is on the unit circle). r and φ are known as the polar coordinates. In trigonometric form the coordinates are $r\cos\varphi$ and $ir\sin\varphi$, therefore, using Euler's formula $e^i = \cos\varphi + i\sin\varphi$, we can also express any complex number as $z = re^i$

is expressed in μV units. The phase φ_{tf} is a cyclic quantity usually reported in the interval $(-\pi, \ldots, \pi]$, but can be equivalently reported in any interval such as $(-1, \ldots, 1]$, $(0, \ldots, 1]$ or in degrees $(0°, \ldots, 360°]$. The physical meaning and interpretation of the analytic signal, the instantaneous amplitude and the instantaneous phase are illustrated in Fig. 4.4. Besides illustrating these concepts, the simple examples in Fig. 4.4 shows how prone to errors may be the interpretation of the analytic signal if a filter bank is not used.

There are two ways of averaging the analytic signal across sweeps. The first is sensitive to evoked (phase-locked) ERP components. The second is sensitive to both evoked and induced (non-phase-locked) components. Thus, we obtain complementary information using the two averaging procedures. In order to study evoked components we average directly the analytic signal at each time-frequency point, such as

$$\bar{z}_{tf} = \frac{1}{K} \sum_k a_{ktf} + i \frac{1}{K} \sum_k b_{ktf} \tag{4.14}$$

from which the *average instantaneous amplitude* (envelope) is given by

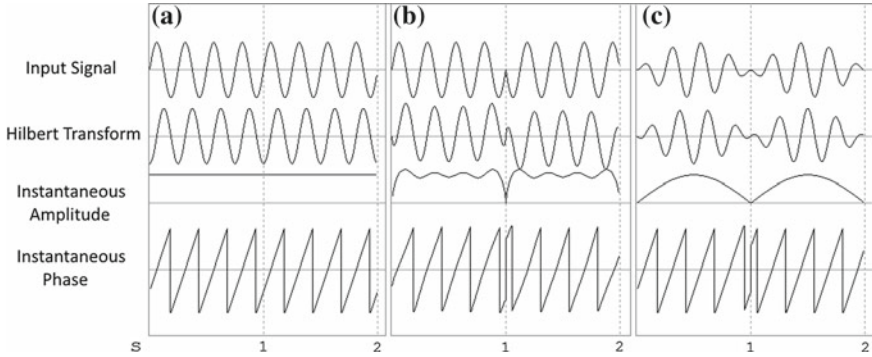$$\bar{r}_{ft} = \left| \bar{z}_{tf} \right| \tag{4.15}$$

Fig. 4.4 Three 2-second signals were generated (input Signal). Time is on the abscissa. The vertical scaling is arbitrary. The Hilbert transform of the input signal is shown in the second traces. The next two traces are the instantaneous amplitude (envelope) and instantaneous phase. Note that the envelope is a non-negative quantity. **a** The input signal is a sine wave at 4 Hz. The instantaneous amplitude is constant in the whole epoch. The phase oscillates regularly in between its bounds at 4 Hz. **b** The input signal is a sine wave at 4 Hz with a phase discontinuity occurring exactly in the middle of the epoch. The instantaneous amplitude now drops in the middle of the epoch. As expected, the instantaneous phase features a discontinuity in the middle of the epoch. **c** the input signal is a sine wave at 4 Hz multiplied by a sine wave at 0.5 Hz with the same amplitude. The result input signal is a sine wave at 4 Hz, which amplitude and phase are modulated by the sine wave at 0.5 Hz. The Instantaneous amplitude is the envelope of the sine at 0.5 Hz. The instantaneous phase is like the one in B, but is now caused by the multiplication with the 0.5 Hz wave

and the *average instantaneous phase* is given by

$$\bar{\varphi}_{tf} = \arg(\bar{z}_{tf}) \tag{4.16}$$

Note that in this case the envelope may be high only if the sweeps at that time-frequency point have a *preferred phase*, whereas if the phase is randomly distributed from sweep to sweep, the average envelope will tend toward zero. This phenomenon is illustrated in Fig. 4.5.

While the Hilbert transform is a linear operator, *non-linear* versions of measures (4.15) and (4.16) may be obtained by adding a simple *normalization* of the analytic signal at each sweep [74]; before computing the average in (4.14), replace a_{ktf} by a_{ktf} / r_{ktf} and b_{ktf} by b_{ktf} / r_{ktf}, where $r_{ktf} = \sqrt{a_{ktf}^2 + b_{ktf}^2}$ is the modulus. This means that at all time-frequency points and for each sweep the complex vector $a_{ktf} + ib_{ktf}$ is stretched or contracted so as to be constrained on the unit complex circle (Fig. 4.6). The average instantaneous amplitude (4.15) and phase (4.16) after the normalization will be actually sensitive to the stability of the phase across sweeps, *regardless of amplitude.* Such non-linear measure is known as *inter-trial phase coherence* (ITPC: [62]), but has been named by different authors also as "inter-trial phase clustering", "phase coherence" among other ways [14].
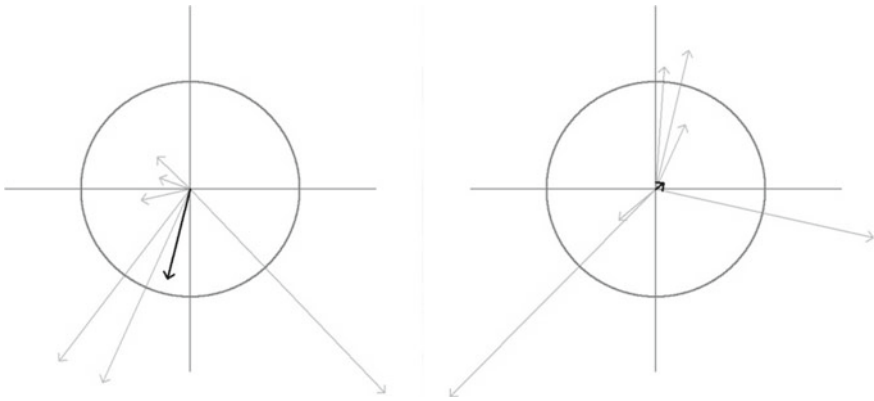
Fig. 4.5 In each diagram six complex numbers are represented as position vectors (gray arrows) in the complex plane (see Fig. 4.3). Consider these vectors as representing the analytic signal for a given time-frequency point estimated on six different sweeps. In each diagram the black arrow is the position vector corresponding to the average of the six complex numbers as per (4.15). In the left diagram the vectors are distributed within one half circle, featuring a preferred direction. In the right diagram the vectors are more randomly distributed around the circle; the resulting mean vector is much smaller, although the average length of the six vectors in the two diagram is approximately equal
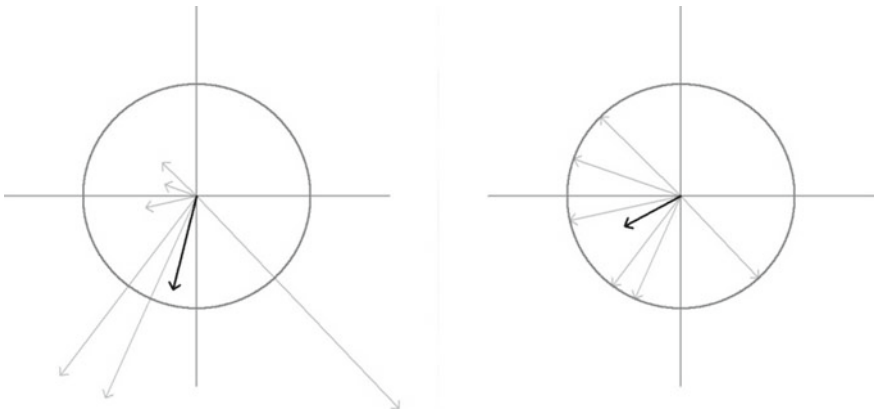

Fig. 4.6 The left diagram is the same as in Fig. 4.5. The vectors in the right diagram have been normalized to unit length (non-linear normalization). Note that the mean vector on the right points in a different direction as compared to the mean vector on the left, albeit the vectors have the same direction in the two diagrams; while on the left diagram the amplitude of the vectors weights the average, on the right diagram the amplitude is ignored

If induced components are of interest, instead of using (4.14) we average the envelope computed on each sweep as

$$\bar{r}_{tf} = \frac{1}{K} \sum_k |z_{ktf}| = \frac{1}{K} \sum_k \sqrt{a_{ktf}^2 + b_{ktf}^2} \qquad (4.17)$$

In this case, the average envelope depends on the amplitude of the coefficients in each sweep and is not affected by the randomness of the analytic signal phase. Note that it does not make sense to average phase values φ_{ktf} estimated at each sweep, as we have done with amplitude in (4.17), since the phase is a circular quantity.[1]

Measures (4.15), (4.16) and their normalized (non-linear) versions can be modified computing a *weighted average* of the normalized analytic signal. Note that the non-normalized average analytic signal is equal to the normalized average analytic signal weighted by its own envelope. Choosing the weights differently, we obtain quite different measures of phase consistency. For instance, weights can be given by experimental or behavioral variables such as reaction time, stimulus luminance, etc. In this way, we can discover phase consistency effects that are specific to certain properties of the stimulus or certain behavioral responses [14, 15]. Taking as weight the envelope of the signal at the frequency under analysis and the analytic signal of *another frequency* (that we name here the *modulating* frequency) we obtain a measure of phase-amplitude coupling named *modulation index* (MI: [10, 14, p. 413]). If the distribution of the modulating phase is uniform, high values of MI reveal dependency between the two frequencies. The modulating frequency is usually lower than the frequency under analysis. Note that by weighting the normalized analytic signal arbitrarily, the obtained average amplitude is no longer guaranteed to be bounded superiorly by 1.0. Furthermore, such measures are subjected to several confounding effects and must be standardized using resampling methods (for details see [10, 14, pp. 253–257, 413–418]). An alternative to the MI measure that does not require such standardization is the *phase-amplitude coupling* (PAC), which is the MI normalized by the amplitude [72]. Measures such as MI and PAC and other variants, along with bivariate counterparts (e.g., [94]), are used to study an important class of phenomena that can be found in the literature under the name of amplitude-amplitude, phase-amplitude and phase-phase nesting (or coupling, interaction, binding…), amplitude modulation and more [16, 36, 54, 55, 73, 93].

Several measures of amplitude and phase in the time-frequency plane are shown in the following real-data example. Figure 4.7 shows a time-frequency analysis of source S7 and S13 of Fig. 4.2. The analysis has been performed on the average of the 80 target sweeps, from -1000 to $+1000$ ms with respect to the flash (visual stimulus), indicated on the abscissa as the time "0". Successively, the first and last 200 ms have been trimmed at both sides to remove edge effects. See the caption of the figure for explanations and the interpretation of results.

---

[1]The time of the day is also a circular quantity and provides a good example. The appropriate average of 22 h and 1 h is 23 h 30, but this is very far from their arithmetic mean. See also Cohen [14, pp. 214–246].
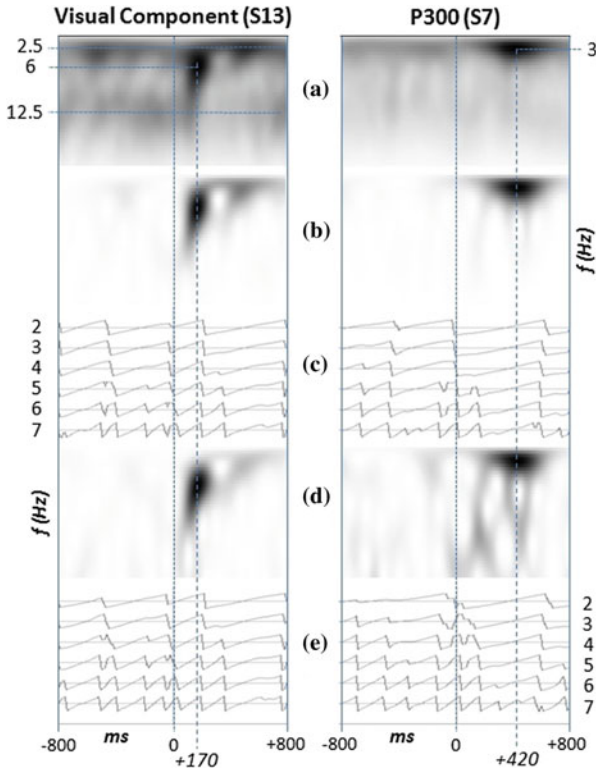
Fig. 4.7 Time-Frequency analysis of the source component S13 (left column) and S7 (right column) shown in Fig. 4.2. **a** Estimated instantaneous amplitude for frequency going from 1 Hz (top of the plot) to 20 Hz (bottom of the plot), in 0.5 Hz steps, computed using (4.17). This method is sensitive to *both phase-locked and non-phase-locked components*. The instantaneous amplitude is color coded, with white coding the minimum and black coding the maximum. The amplitude in **a** features a maximum in the time-frequency plane at around 6 Hz happening 170 ms post-stimulus, corresponding to the P100/N200 complex (see Fig. 4.2). We also notice a sustained activity around 2.5 Hz from about 200 to 700 ms post-stimulus. Note that at 2.5 Hz substantial power is present also *before* the stimulus, but this does not happen at 6 Hz. **b** Estimated instantaneous amplitude obtained with (4.15). This method is sensitive to *phase-locked components*. Note that both post-stimulus maxima at around 2.5 and 6 Hz survive, whereas anywhere else in the time-frequency plot the amplitude becomes negligible, including pre-stimulus activity around 2.5 Hz. Note also that the 2.5 Hz activity post-stimulus now is weaker. Taken together the analyses in **a** and **b** suggest that the activity around 6 Hz may be strictly phase-locked, whereas the activity at 2.5 Hz may be mixed with non-phase-locked components. Plot **c** shows the instantaneous phase of S13 in the closed interval $(-\pi \ldots -\pi]$, for frequencies in the range 2 Hz, …, 7 Hz, in 1 Hz increments. This has been computed using (4.16), hence it is the phase spectrum corresponding to **b**. At about 220 ms post-stimulus, in correspondence to the end of the maximum at 6 Hz, the phase alines at all frequencies in the range 2 Hz, …, 7 Hz. The amplitude spectrum in **d** and corresponding phase spectrum in **e** are the non-linear (normalized) version of **b** and **c**, respectively. The results are very similar to those seen in **b** and **c**, although they appear a bit noisier. For S7, the instantaneous amplitude (4.17) features only one strong maximum at about 3 Hz in between 280 and 570 ms (**a**, right column). This maximum corresponds to the P300 peak (Fig. 4.2). The same activity is seen also in **b** and **d**, although they appear noisier. This analysis suggests that the P300 is strictly phase-locked to the stimulus

We end up this section with some considerations about TFA analysis. The Hilbert transform can be obtained by the FFT algorithm [64]. The use of this algorithm requires the choice of a tapering window in the time domain to counteract spectral leakage due to finite window size (see Harris [39]). As illustrated in Fig. 4.4, the analytic signal does not necessarily represent adequately the phase of the original signal. The study of Chavez et al. [13] has stressed that this is the case in general only if the original signal is a simple oscillator with a narrow-band frequency support. These authors have provided useful measures to check empirically the goodness of the analytic signal representation. Because of this limitation, for a signal displaying multiple spectral power peaks or broad-band behavior, which is the case in general of EEG and ERP, the application of a filter bank to extract narrow-band behavior is necessary. When applying the filter bank, one has to make sure not to distort the phase of the signal. In general, a finite impulse response filter with linear phase response is adopted (see Widmann et al. [98], for a review). The choice of the filters band width and frequency resolution is usually a matter of trials and errors; the band width should be large enough to capture the oscillating behavior and small enough to avoid capturing several oscillators in adjacent frequencies. Also, the use of filter banks engenders edge effects, that is, severe distortions of the analytic signal at the left and right extremities of the time window under analysis [67]. This latter problem is easily solved defining a larger time window centered at the window of interest and successively trimming an adequate number of samples at both sizes, as we have done in the example of Fig. 4.7. The estimation of instantaneous phase for sweeps, time sample and frequencies featuring a low SNR are meaningless; the phase being an angle, it is defined for vectors of any length, even if the length (i.e., the amplitude) is negligible. However, phase measures can be interpreted only where the amplitude is high [7]. The effect is exacerbated if we apply the non-linear normalization, since in this case very small coefficients are weighted as the others in the average, whereas they should better be ignored.

4.5 Spatial Domain Analysis

Scalp topography and tomography (source localization) of ERPs are the basic tools to perform analysis in the *spatial domain* of the electrical activity generating ERPs. This is fundamental for linking experimental results to brain anatomy and physiology. It also represents an important dimension for studying ERP dynamics per se, complementing the information provided in time and/or frequency dimensions [52]. The spatial pattern of ERP scalp potential or of an ERP source component provides useful information to recognize and categorize ERP features, as well as to identify artifacts and background EEG. Early ERP research was carried out using only a few electrodes. Current research typically uses several tens and even hundreds of electrodes covering the whole scalp surface. More and more high-density EEG studies involve realistic head models for increasing the precision of source localization

methods. Advanced spatial analysis has therefore become common practice in ERP research.

In contrast to continuous EEG, ERP studies allow spatial analysis with high-temporal resolution, i.e., they allow the generation of topographical and/or tomographical maps for each time sample. This is due to the SNR gain engendered by averaging across sweeps. Thus, as compared to continuous EEG, ERPs offer an analysis in the spatial domain with much higher temporal resolution. The SNR increases with the number of averaged sweeps. One can further increase the SNR by using a multivariate filtering method, as previously discussed. One can also increase the SNR by averaging spatial information in adjacent samples. The spatial patterns observed at all samples forming a peak in the *global field power*[2] can safely be averaged, since within the same peak the spatial pattern is supposed to be constant [52].

When using a source separation method (see Fig. 4.2 for an example) the spatial pattern related to each source component is given by the corresponding column vector of the estimated mixing matrix, i.e., the pseudo-inverse of the estimated matrix B^T. In fact, a source separation method decomposes the ensemble average in a number of source components, each one having a different and fixed spatial pattern. These patterns are analyzed separately as a topographic map and are fed individually to a source localization method as input data vector. Source localization methods in general perform well when the data is generated by one or two dipoles only, while if the data is generated by multiple dipoles the accuracy of the reconstruction is questionable [95]. BSS effectively decomposes the ensemble average in a number of simple source components, typically generated by one or two dipoles each [25]. As a consequence, spatial patterns decomposed by source separation can be localized with high accuracy by means of source localization methods. Note that applying a generic filtering method such as PCA and CSTP, the components given by the filter are still mixed and so are the spatial patterns held as column vectors by the matrix inverse of the spatial filter, that is, the pseudo-inverse of B^T. This prevents any physiological interpretation of the corresponding spatial patterns. Source separation methods are therefore optimal candidates for performing high-resolution spatial analysis by means of ERPs. An example of topographical analysis is presented in Figs. 4.2 and 4.8. For an example of tomographic analysis refer to Congedo et al. [24].

4.6 Inferential Statistics

As we have seen, in time-domain ERP studies it is of interest to localize experimental effects along the dimension of *space* (scalp location) and *time* (latency and duration of the ERP components). Analysis in the time-frequency-domain involves the study of amplitude and phase in the time-frequency plane. The dimensions retained by

---

[2]The global field power is defined for each time sample as the sum of the squares of the potential difference at all electrodes. It is very useful in ERP analysis to visualize ERP peaks regardless their spatial distribution [52].
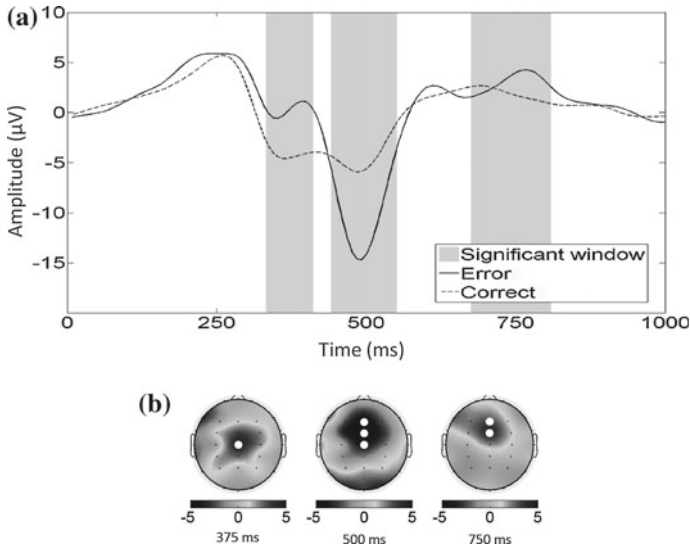
Fig. 4.8 **a** Ensemble average error-related potentials (19 subjects) for "Correct" and "Error" trials at electrode Cz. The supra-threshold cluster size permutation test was applied in the time and spatial dimension, with a = 0.05, to compare the "Correct" and "Error" condition. A significant positivity for error trials was found at time window 320–400 ms at electrode Cz ($p < 0.01$), a significant negativity for error trials at time window 450–550 ms at clustered electrodes Fz, FCz, Cz ($p < 0.01$) and a significant positivity for error trials at time 650–775 ms at clustered electrodes Fz, FCz ($p = 0.025$). Significant time windows are indicated by grey areas in (**a**) and significant clustered derivations by white disks in (**b**). The supra-threshold cluster size test display good power while controlling the FWER. Data is from the study of Congedo et al. [24]

the experimenter for the statistical analysis actually are combined to create a multi-dimensional measurement space. For example, if a time-frequency representation is chosen and amplitude is the variable of interest, the researcher defines a statistical hypothesis at the intersection of each time and frequency measurement point. Typical hypotheses in ERP studies concern differences in central location (mean or median) within and between subjects (t-tests), the generalization of these tests to multiple experimental factors including more than two levels, including their interaction (ANOVA) and the correlation between ERP variables and demographic or behavioral variables such as response-time, age of the participants, complexity of the cognitive task, etc. (linear and non-linear regression, ANCOVA).

The goal of a statistical test is to either reject or accept the corresponding *null hypothesis* for a given type I error (α), which is the a priori chosen probability to reject a null hypothesis when this is indeed true (false discovery). By definition, our conclusion will be wrong with probability α, which is typically set to 0.05. Things becomes more complicated when several tests are performed simultaneously; performing a statistical test independently for each hypothesis inflates the type I error rate proportionally to the number of tests. This is known as the *multiple-comparison*

problem [41, 97] and is very common in ERP studies, where several points in time, space and frequency are to be investigated. Let M be the number of hypotheses to be tested and M_0 be the number of true null hypotheses. Testing each hypothesis independently at the α level, the expectation of false discoveries is $M_0 \times \alpha$. Thus, if all null hypotheses are actually true, i.e., $M_0 = M$, we expect to commit on the average $(100 \times \alpha) \%$ false discoveries. This is, of course, an unacceptable error rate. Nonetheless, the more hypotheses are false and the more they are correlated, the more the error rate is reduced. ERP data is highly correlated along adjacent time points, spatial derivations and frequency. Therefore, special care should be undertaken in ERP statistical analysis to ensure that the error rate is controlled while preserving statistical power, that is, while preserving an acceptable chance to detect those null hypotheses that are false. Two families of statistical procedures have been employed in ERP studies with this aim: those controlling the *family-wise error rate (FWER)* and those controlling the *false-discovery rate (FDR)*.

The *family-wise error rate* (FWER) is the probability of making one or more false discoveries among all hypotheses. A procedure controlling the FWER at the α level ensures that the probability of committing even only one false discovery is less than or equal to α, regardless the number of tests and how many null hypotheses are actually true. The popular Bonferroni procedure belongs to this family; each hypothesis is tested at level α/M instead that at level α. Sequential Bonferroni-like procedures like the one proposed by Holm [42] also control the FWER, while featuring higher power. However, all Bonferroni-like procedures fail to take into consideration explicitly the correlation structure of the hypotheses, thus they are unduly conservative, the more so the higher the number of hypotheses to be tested.

An important general class of test procedures controlling the FWER is known as *p-min permutation tests* [75, 97], tracing back to the seminal work of Fisher [35] and Pitman [80–82]. Permutation tests are able to account adaptively for any correlation structure of hypotheses, regardless of its form and degree. Also, they do not need a distributional model for the observed variables, e.g., Gaussianity, as required by t-tests, ANOVA etc. [6, 30, 35, 43, 46, 75, 80–82, 91, 92, 97]. Even more appealing, one may extract whatever variable from the data and perform a valid test, thus we are not limited to test on central location, correlation, etc. Depending on the experimental design, even the random sampling assumption may be relaxed [30]. Given these characteristics, permutation tests are ideal options for testing hypotheses in ERP studies and have received much attention in the neuroimaging community [1, 43, 76].

Permutation tests are available for classical correlation, within- and between-subject mean difference tests, as well as for testing the main effects in ANOVA designs [30]. However, a straightforward permutation test for interaction effects in ANOVA designs does not exist, although some solutions have been proposed [75]. This is a major limitation if more than one independent variable is manipulated in the experiment. Also, like other resampling methods such as bootstrap and Monte Carlo, permutation tests require intense computations. For large data sets, permutation tests may be time consuming, although this is rarely a concern with modern computers and the typical size of data sets in ERP analysis.

Another kind of FWER-controlling permutation test in ERP analysis is the supra-threshold cluster size test [43]. A variant of this test has been implemented in the EEG toolbox Fieldtrip [71], following the review of Maris and Oostenveld [63]. This procedure assesses the probability to observe a concentration of the effect simultaneously along one or more studied dimensions. For example, in testing the mean amplitude difference of a P300 ERP, one expects the effect to be concentrated both along time, around 300–500 ms, and along space, at midline central and adjacent parietal locations. This leads to a typical correlation structure of hypothesis in ERP data; under the null hypothesis the effect would instead be scattered all over both time and spatial dimensions. An example of the supra-threshold cluster size test applied in the time-space ERP domain is shown in Fig. 4.8.

Another family of testing procedures controls the *false discovery rate* (FDR). The FDR is the expected proportion of falsely rejected hypotheses [3]. Indicating by R the number of rejected hypotheses and by F the number of those that have been falsely rejected, the FDR controls the expectation of the ratio F/R. This is clearly a less stringent criterion as compared to the FWER, since, as the number of discoveries increases, we allow proportionally more errors. The original FDR procedure of Benjamini and Hochberg [3] assumes that all hypotheses are independent, which is clearly not the case in general for ERP data. A later work has extended the FDR procedure to the case of arbitrary dependence structure among variables [4], however, contrary to what one would expect, the resulting procedure is more conservative, yielding low power in practice. The FDR procedure and its version for dependent hypotheses have been the subject of several improvements (e.g., [38, 87]). Recent research on FDR-controlling procedures attempts to increase their power by sorting the hypotheses based on a priori information [32]. Such sorting may be guided by previous findings in similar experiments, by the total variance of the variables when using central location tests, or by any criterion that is independent to the test-statistics. Another trend in this direction involves arranging the hypotheses in hierarchical trees prior to testing [102] and in analyzing experimental replicability [40]. The FDR procedures tend to be unduly conservative when the number of hypotheses is very large, although much less so than Bonferroni-like procedures. In contrast to FWER-controlling procedures, FDR-controlling procedures are much simpler and faster to compute. They offer, however, a much looser guarantee against the actual type I error rates and, like Bonferroni-like procedures, do not take explicitly into consideration the correlation structure of ERP data.

4.7 Single-Sweep Classification

The goal of a classification method is to automatically estimate the class to which a single-sweep belongs. The task is challenging because of the very low amplitude of ERPs as compared to the background EEG. Large artifacts, the non-stationary nature of EEG and inter-sweep variability exacerbate the difficulty of the task. Although single-sweep classification has been investigated since a long time [29], it has recently

received a strong impulsion thanks to development of ERP-based brain computer interfaces (BCI: [101]). In fact, a popular family of such interfaces is based on the recognition of the P300 ERP. The most famous example is the *P300 Speller* [34], a system allowing the user to spell text without moving, but just by focusing attention on symbols (e.g., letters) that are flashed on a virtual keyboard.

The fundamental criterion for choosing a classification method is the achieved accuracy for the data at hand. However, other criteria may be relevant. In BCI systems, the training of the classifier starts with a calibration session carried out just before the actual session. Such calibration phase makes the usage of BCI system impractical and annoying. To avoid this, there are at least two other desirable characteristics that a classification method should possess [60]: its ability to *generalize* and its ability to *adapt*. Generalization allows the so-called *transfer learning*, thanks to which data from other sessions and/or other subjects can be used to initialize a BCI system so as to avoid the calibration phase. Transfer learning may involve using data from previous sessions of the same subject ("cross-session") and/or data from other subjects ("cross-subject"). The continuous (on-line) adaptation of the classifier [47, 48] ensures that optimal performance is achieved once the initialization is obtained by transfer learning [18]. Taken together, generalization and on-line adaptation ensure also the stability of the system in adverse situations, that is, when the SNR of the incoming data is low and when there are sudden environmental, instrumental or biological changes during the session. This is very important for effective use of a BCI outside the controlled environment of research laboratories.

Classification methods differ from each other in the way they define the *set of features* and in the *discriminant function* they employ. Traditionally, the classification approaches for ERPs have given emphasis to the optimization of either one or the other aspect in order to increase accuracy. The approaches emphasizing the definition of the set of features try to increase the SNR of single-sweeps by using multivariate filtering, as those we have encountered in the section on time domain analysis, but specifically designed to increase the separation of the classes in a reduced feature space where the filter projects the data [83, 104]. For data filtered in this way, the choice of the discriminant function is not critical, in the sense that similar accuracy is obtained using several types of discriminant functions. In general, these approaches perform well even if the training set is small, but generalize poorly across sessions and across subjects because the spatial filters are optimal only for the session and subject on whom they are estimated. Instead, the approaches emphasizing the discriminant function use sharp machine learning algorithms on raw data or on data that has underwent little-preprocessing. Many machine learning algorithms have been tried in the BCI literature for this purpose [59, 60]. The three traditional approaches that have been found effective in P300 single-sweep classification are the support-vector machine, the stepwise linear discriminant analysis and the Bayesian linear discriminant analysis. In general, those require large training sets and have high computational complexity, but generalize fairly well across sessions and across subjects. The use of a *random forest* classifier is currently gaining popularity in the BCI community, incited by good accuracy properties [33]. However, its generalization and adaptation capability have not been established yet. The *deep neural networks*

learning [5] has recently been shown to be very promising in other fields of research. Studies testing its performance on ERP data are not conclusive so far. An approach that features at the same time good accuracy, good generalization and good adaptation capabilities in the case of ERP data has been recently borrowed from the field of *differential geometry*. This approach makes use of the *Riemannian geometry* on the manifold of symmetric positive definite (SPD) matrices. Covariance matrices are of this kind. A very simple classifier can be obtained based on the minimum distance to mean (MDM) method [2]: every sweep is represented as a covariance matrix, i.e., as a point on the multidimensional space of SPD matrices. The training set is used to estimate the *center of mass* of training points for each class, i.e., a point best representing the class. An unlabeled sweep is then simply assigned to the class the center of mass of which is the closest to the unlabeled sweep. This approach as well as other classifiers based on Riemannian geometry have been shown to possess good accuracy, generalization and robustness properties [19, 66, 103, 105]

References

1. S. Arndt, T. Cizadlo, N.C. Andreasen et al., Tests for comparing images based on randomization and permutation methods. J. Cereb. Blood Flow Metab. **16**, 1271–1279 (1996)
2. A. Barachant, S. Bonnet, M. Congedo et al., Multi-class brain computer interface classification by riemannian geometry. IEEE Trans. Biomed. Eng. **59**, 920–928 (2012)
3. Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. Roy. Stat. Soc. Ser. B Stat. Methodol. **57**, 289–300 (1995)
4. Y. Benjamini, D. Yukutieli, The control of the false discovery rate in multiple testing under dependency. Ann. Stat. **29**, 1165–1188 (2001)
5. Y. Bengio, Learning deep architectures for AI. Found. Trends Mach. Learn. **2**, 1–12 (2009)
6. R.C. Blair, J.F. Troendle, R.W. Beck, Control of familywise errors in multiple assessments via stepwise permutation tests. Stat. Med. **15**, 1107–1121 (1996)
7. P. Bloomfield, *Fourier Analysis of Time Series. An Introduction*, 2nd edn. (Wiley, Hoboken, New Jersey, 2000), p. 261
8. A. Burns, Fourier-, Hilbert- and wavelet-based signal analysis: are they really independent approaches? J. Neurosci. Methods **137**, 321–332 (2004)
9. G. Buszáki, C.A. Anastassiou, C. Koch, The origin of extracellular fields and currents—EEG, ECoG, LFP and spikes. Nat. Rev. Neurosci. **13**, 407–420 (2012)
10. R.T. Canolty, E. Edwards, S.S. Dalal et al., High gamma power is phase-locked to theta oscillations in human neocortex. Science **313**, 1626–1628 (2006)
11. J.-F. Cardoso, Blind signal separation: statistical principles. Proc. IEEE **86**, 2009–2025 (1998)
12. R.M. Chapman, J.W. McCrary, EP component identification and measurement by principal component analysis. Brain Cogn. **27**, 288–310 (1995)
13. M. Chavez, M. Besserve, C. Adam et al., Towards a proper estimation of phase synchronization from time series. J. Neurosci. Methods **154**, 149–160 (2006)
14. M.X. Cohen, *Analyzing Neural Time Series Data: Theory and Practice* (The MIT Press, Cambridge, Massachusetts, 2014), p. 600
15. M.X. Cohen, J.F. Cavanagh, Single-trial regression elucidates the role of prefrontal theta oscillations in response conflict. Front. Psychol. **2**, 30 (2011)
16. L.L. Colgin, Theta-gamma coupling in the entorhinal-hippocampal system. Curr. Opin. Neurobiol. **31**, 45–50 (2015)
17. P. Comon, C. Jutten (eds.), *Handbook of Blind Source Separation, Independent Component Analysis and Applications* (Academic Press, Cambridge, MA, 2010)

18. M. Congedo, *EEG Source Analysis*. Dissertation, University of Grenoble Alpes, 2013
19. M. Congedo, A. Barachant, R. Bhatia, Riemannian geometry for EEG-based brain-computer interfaces; a primer and a review. BCI **4**, 155–174 (2017)
20. M. Congedo, C. Gouy-Pailler, C. Jutten, On the blind source separation of human electroencephalogram by approximate joint diagonalization of second order statistics. Clin. Neurophysiol. **119**, 2677–2686 (2008)
21. M. Congedo, M. Goyat, N. Tarrin et al., in *"Brain Invaders": A Prototype of an Open-Source P300- Based Video Game Working with the OpenViBE*. 5th International Brain-Computer Interface Conference, Graz, Austria, September 2011. (2011), pp. 280–283
22. M. Congedo, L. Korczowski, A. Delorme et al., Spatio-temporal common pattern; a reference companion method for ERP analysis. J. Neurosci. Methods **267**, 74–88 (2016)
23. M. Congedo, F.H. Lopes da Silva, Event-related potentials: general aspects of methodology and quantification, in *Niedermeyer's Electroencephalography, Basic Principles, Clinical Applications, and Related Fields*, ed. by D.L. Schomer, F.H. Lopes da Silva (Oxford University Press, Oxford, 2017)
24. M. Congedo, S. Rousseau, C. Jutten, An introduction to EEG source analysis with an illustration of a study on error-related potentials, in *Guide to Brain-Computer Music Interfacing*, ed. by E. Miranda, J. Castet (Springer, London, 2014), p. 313
25. A. Delorme, J. Palmer, J. Onton et al., Independent EEG sources are dipolar. PLoS One **7**, e30135 (2012)
26. A. Delorme, T. Sejnowski, S. Makeig, Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis. Neuroimage **34**, 1443–1449 (2007)
27. J. Dien, Evaluating two-step PCA of ERP data with Geomin, Infomax, Oblimin, Promax, and Varimax rotations. Psychophysiology **47**, 170–183 (2010)
28. E. Donchin, A multivariate approach to the analysis of average evoked potentials. IEEE Trans. Biomed. Eng. **3**, 131–139 (1966)
29. E. Donchin, Discriminant analysis in average evoked response studies: the study of single trial data. Electroencephalogr. Clin. Neurophysiol. **27**, 311–314 (1969)
30. E.S. Edgington, *Randomization Tests*, 3rd edn. (Marcel Dekker, New York, 1995)
31. M. Falkenstein, J. Hohnsbein, J. Hoormann et al., Effects of crossmodal divided attention on late ERP components. II. Error processing in choice reaction tasks. Electroencephalogr. Clin. Neurophysiol. **78**, 447–455 (1991)
32. A. Farcomeni, L. Finos, FDR control with pseudo-gatekeeping based on possibly data driven order of The hypotheses. Biometrics **69**, 606–613 (2013)
33. F. Farooq, P. Kidmose, in *Random Forest Classification for P300 Based Brain Computer Interface Applications*. 21th European Signal Processing Conference, Marrakech, Morocco, September 2013. (2013) pp. 1–5
34. L.A. Farwell, E. Donchin, Talking off the top of your head: toward a mental prothesis utilizing event-related brain potentials. Electroencephalogr. Clin. Neurophysiol. **70**, 510–523 (1988)
35. R.A. Fisher, *Design of Experiments* (Oliver and Boyd, Edinburgh, 1935)
36. W.J. Freeman, Mechanism and significance of global coherence in scalp EEG. Curr. Opin. Neurobiol. **31**, 199–205 (2015)
37. D. Gabor, Theory of communication. J. IEE (London) **93**, 429–457 (1946)
38. W. Guo, M.B. Rao, On control of the false discovery rate under no assumption of dependency. J. Stat. Plan. Interference **138**, 3176–3188 (2008)
39. F.J. Harris, On the use of windows for harmonic analysis with the discrete Fourier transform. Proc. IEEE **66**, 51–83 (1978)
40. R. Heller, D. Yekutieli, Replicability analysis for genome-wide association studies. Ann. Appl. Stat. **8**, 481–498 (2014)
41. Y. Hochberg, A.C. Tamhane, *Multiple Comparison Procedures* (Wiley, Hoboken, NJ, 1987)
42. S. Holm, A simple sequentially rejective multiple test procedure. Scand. J. Stat. **6**, 65–70 (1979)

43. A.P. Holmes, R.C. Blair, J.D.G. Watson et al., Nonparametric analysis of statistic images from functional mapping experiments. J. Cereb. Blood Flow Metab. **16**, 7–22 (1996)
44. B.H. Jansen, G. Agarwal, A. Hedge et al., Phase synchronization of the ongoing EEG and auditory EP generation. Clin. Neurophysiol. **114**, 79–85 (2003)
45. E.R. John, D.S. Ruchkin, J. Vilegas, Experimental background: signal analysis and behavioral correlates of evoked potential configurations in cats. Ann. N. Y. Acad. Sci. **112**, 362–420 (1964)
46. W. Karniski, R.C. Blair, A.D. Snider, An exact statistical method for comparing topographic maps, with any number of subjects and electrodes. Brain Topogr. **6**, 203–210 (1994)
47. P.-J. Kindermans, D. Verstraeten, B. Schrauwen, A Bayesian Model for Exploiting Application Constraints to Enable Unsupervised Training of a P300-based BCI. PLoS ONE **7**, e33758 (2012)
48. P.-J. Kindermans, M. Schreuder, B. Schrauwen et al., True zero-training brain-computer interfacing—an online study. PLoS One **9**, e102504 (2014)
49. L. Korczowski, F. Bouchard, C. Jutten et al., in *Mining the Bilinear Structure of Data with Approximate Joint Diagonalization*. 24th European Signal Processing Conference, Budapest, Hungary, August 2016. (2016) pp. 667–671
50. J.-P. Lachaux, E. Rodriguez, J. Martinerie et al., Measuring phase synchrony in brain signals. Hum. Brain Mapp. **8**, 194–208 (1999)
51. T.D. Lagerlund, F.W. Sharbrough, N.E. Busacker, Spatial filtering of multichannel electroencephalographic recordings through principal component analysis by singular value decomposition. J. Clin. Neurophysiol. **14**, 73–82 (1997)
52. D. Lehmann, W. Skrandies, Reference-free identification of components of checkerboard-evoked multichannel potential fields. Electroencephalogr. Clin. Neurophysiol. **48**, 609–621 (1980)
53. M. Le Van Quyen, J. Foucher, J.-P. Lachaux et al., Comparison of Hilbert transform and wavelet methods for the analysis of neuronal synchrony. J. Neurosci. Methods **111**, 83–98 (2001)
54. J.E. Lisman, O. Jensen, The theta-gamma neural code. Neuron **77**, 1002–1016 (2013)
55. R.R. Llinas, The intrinsic electrophysiological properties of mammalian neurons: insights into central nervous system function. Science **242**, 1654–1664 (1988)
56. J.M. Lęski, Robust weighted averaging. IEEE Trans. Biomed. Eng. **49**, 796–804 (2002)
57. F.H. Lopes da Silva, Event-related neural activities: what about phase? Prog. Brain Res. **159**, 3–17 (2006)
58. F.H. Lopes da Silva, J.P. Pijn, D. Velis et al., Alpha rhythms: noise, dynamics, and models. Int. J. Psychophysiol. **26**, 237–249 (1997)
59. F. Lotte, M. Congedo, A. Lécuyer et al., A review of classification algorithms for EEG-based brain-computer interfaces. J. Neural Eng. **4**, R1–R13 (2007)
60. F. Lotte, L. Bougrain, A. Cichocki et al., A review of classification algorithms for EEG-based brain-computer interfaces: a 10-year update (Manuscript submitted for publication) (2018)
61. S. Luck, *An Introduction to the Event-Related Potential Technique*, 2nd edn. (The MIT Press, Cambridge, MA, 2014)
62. S. Makeig, M. Westerfield, T.P. Jung et al., Dynamic brain sources of visual evoked responses. Science **295**, 690–694 (2002)
63. E. Maris, R. Oostenveld, Nonparametric statistical testing of EEG- and MEG-data. J. Neurosci. Methods **164**, 177–190 (2007)
64. S.L. Marple, Computing the discrete-time analytic signal via FFT. IEEE Trans. Signal Process. **47**, 2600–2603 (1999)
65. A. Mazaheri, O. Jensen, Rhythmic pulsing: linking ongoing brain activity with evoked responses. Front. Hum. Neurosci. **4**, 177 (2010)
66. L. Mayaud, S. Cabanilles, A. Van Langhenhove et al., Brain-computer interface for the communication of acute patients: a feasibility study and a randomized controlled trial comparing performance with healthy participants and a traditional assistive device. BCI **3**, 197–215 (2016)

67. F. Mormann, K. Lehnertz, P. David et al., Mean phase coherence as a measure for phase synchronization and its application to the EEG of epilepsy patients. Physica D **144**, 358–369 (2000)
68. R. Näätänen, A.W.K. Gaillard, S. Mäntysalo, Early selective-attention effect on evoked potential reinterpreted. Acta Psychol (Amst) **42**, 313–329 (1978)
69. V.V. Nikulin, K. Linkenkaer-Hansen, G. Nolte et al., Non-zero mean and asymmetry of neuronal oscillations have different implications for evoked responses. Clin. Neurophysiol. **121**, 186–193 (2010)
70. P.L. Nunez, R. Srinivasan, *Electric Field of the Brain: The Neurophysics of EEG* (Oxford University Press, New York, 2006)
71. R. Oostenveld, P. Fries, E. Maris et al., FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput. Intell. Neurosci. **2011**, 1 (2011)
72. T.E. Özkurt, A. Schnitzler, A critical note on the definition of phase-amplitude cross-frequency coupling. J. Neurosci. Methods **201**, 438–443 (2011)
73. S. Palva, J.M. Palva, Discovering oscillatory interaction networks with M/EEG: challenges and breakthroughs. Trends Cogn Sci **16**, 219–230 (2012)
74. R.D. Pascual-Marqui, Instantaneous and lagged measurements of linear and nonlinear dependence between groups of multivariate time series: frequency decomposition. arXiv:07 11.1455 (2007)
75. F. Pesarin, *Multivariate Permutation Tests* (Wiley, Hoboken, NJ, 2001)
76. K.M. Petersson, T.E. Nichols, J.-B. Poline et al., Statistical limitations in functional neuroimaging II. Signal detection and statistical inference. Philos. Trans. Roy. Soc. Lond. **354**, 1261–1281 (1999)
77. G. Pfurtscheller, F.H. Lopes da Silva, Event-related EEG/MEG synchronization and desynchronization: basic principles. Clin. Neurophysiol. **110**, 1842–1857 (1999)
78. D.-T. Pham, J.-F. Cardoso, Blind separation of instantaneous mixtures of non stationary sources. IEEE Trans. Signal Process. **49**, 1837–1848 (2001)
79. T.W. Picton, S. Bentin, P. Berg et al., Guidelines for using human event-related potentials to study cognition: recording standards and publication criteria. Psychophysiology **37**, 127–152 (2000)
80. E.J.G. Pitman, Significance tests which may be applied to samples from any population. J. Roy. Stat. Soc. Ser. B Stat. Methodol. **4**, 119–130 (1937)
81. E.J.G. Pitman, Significance tests which may be applied to samples from any population. II. The correlation coefficient. J. Roy. Stat. Soc. Ser. B Stat. Methodol. **4**, 225–232 (1937)
82. E.J.G. Pitman, Significance tests which may be applied to samples from any population. III. The analysis of Variance test. Biometrika **29**, 322–335 (1938)
83. B. Rivet, A. Souloumiac, V. Attina et al., xDAWN algorithm to enhance evoked potentials: application to brain-computer interface. IEEE Trans. Biomed. Eng. **56**, 2035–2043 (2009)
84. M.G. Rosenblum, A.S. Pikovsky, J. Kurths, Phase synchronization of chaotic oscillators. Phys. Rev. Lett. **76**, 1804–1807 (1996)
85. D.S. Ruchkin, An analysis of average response computatios based upon aperiodic stimuli. IEEE Trans. Biomed. Eng. **12**, 87–94 (1965)
86. S.C. Sereno, K. Rayner, Measuring word recognition in reading: eye-movements and event-related potentials. Trends Cogn. Sci. **7**, 489–493 (2003)
87. J.D. Storey, A direct approach to false discovery rate. J. Roy. Stat. Soc. Ser. B Stat. Methodol. **4**, 479–498 (2002)
88. S. Sutton, M. Braren, J. Zubin et al., Evoked-potential correlates of stimulus uncertainty. Science **150**, 1187–1188 (1965)
89. C. Tallon-Baudry, O. Bertrand, C. Delpuech et al., Stimulus specificity of phase-locked and non-Phase-locked 40 Hz visual responses in human. J. Neurosci. **16**, 4240–4249 (1996)
90. P. Tass, M.G. Rosenblum, J. Weule et al., Detection of n:m phase locking from noisy data: application to magnetoencephalography. Phys. Rev. Lett. **81**, 3291–3294 (1998)

91. J.F. Troendle, A stepwise resampling method of multiple hypothesis testing. J. Am. Stat. Assoc. **90**, 370–378 (1995)
92. J.F. Troendle, A permutation step-up method of testing multiple outcomes. Biometrics **952**, 846–859 (1996)
93. F. Varela, J.-P. Lachaux, E. Rodriguez et al., The brainweb: phase synchronization and large-scale integration. Nat. Rev. Neurosci. **2**, 229–239 (2001)
94. M. Vinck, R. Oostenveld, M. van Wingerden et al., An improved index of phase synchronization for electrophysiological data in the presence of volume-conduction, noise and sample-size bias. NeuroImage **55**, 1548–1565 (2011)
95. M. Wagner, M. Fuchs, J. Kastner, Evaluation of sLORETA in the presence of noise and multiple sources. Brain Topogr. **16**, 277–280 (2004)
96. W.G. Walter, R. Cooper, V.J. Aldridge et al., Contingent negative variation: an electric sign of sensorimotor association and expectancy in the human brain. Nature **203**, 380–384 (1964)
97. P.H. Westfall, S.S. Young, *Resampling-Based Multiple Testing. Examples and Methods for p-Values Adjustment* (Wiley, Hoboken, NJ, 1993)
98. A. Widmann, E. Schröger, B. Maess, Digital filter design for electrophysiological data—a practical approach. J. Neurosci. Methods **250**, 34–46 (2014)
99. M. Woldorff, Adjacent response overlap during the ERP averaging process and a technique (Adjar) for its estimation and removal. Psychophysiology **25**, 490 (1988)
100. M. Woldorff, Distortion of ERP averages due to overlap from temporally adjacent ERPs: analysis and correction. Psychophysiology **30**, 98–119 (1993)
101. J. Wolpaw, E.W. Wolpaw (eds.), *Brain-Computer Interfaces: Principles and Practice* (Oxford University Press, New York, 2012), p. 424
102. D. Yekutieli, Hierarchical false discovery rate-controlling methodology. J. Am. Stat. Assoc. **103**, 309–316 (2012)
103. F. Yger, M. Berar, F. Lotte, Riemannian approaches in brain-computer interfaces: a review. IEEE Trans. Neural Syst. Rehabil. Eng. **25**, 1753–1762 (2016)
104. K. Yu, K. Shen, S. Shao et al., Bilinear common spatial pattern for single-trial ERP-based rapid serial visual presentation triage. J. Neural Eng. **9**, 046013 (2012)
105. P. Zanini, M. Congedo, C. Jutten et al., Transfer learning: a riemannian geometry framework with applications to brain-computer interfaces. IEEE Trans. Biomed. Eng. (2017 in press)