Spatio-temporal common pattern: A companion method for ERP analysis in the time domain

Marco Congedo, Louis Korczowski, Arnaud Delorme, Fernando Lopes da Silva

HAL Id: hal-01343026
https://hal.science/hal-01343026

Submitted on 7 Jul 2016

Spatio-Temporal Common Pattern;

a Companion Method for ERP Analysis in the Time Domain

Marco CONGEDO[a*] , Louis KORCZOWSKI[a], Arnaud DELORME[b,c,d], Fernando LOPES DA SILVA[e]

a GIPSA-lab, CNRS and Grenoble Alpes University, Grenoble, France
b. Université de Toulouse; UPS; Centre de Recherche Cerveau et Cognition; Toulouse, France
c. CNRS; CerCo; France
d. Swartz Center for Computational Neurosciences, UCSD, La Jolla, CA, USA
e Center of Neuroscience, Swammerdam Institute for Life Sciences, Amsterdam, The Netherlands

*Corresponding author.
Team ViBS (Vision and Brain Signal Processing)
GIPSA-lab (Grenoble Images Parole Signal Automatique)
CNRS (National Center for Scientific Research)
11 rue des Mathématiques
Domaine universitaire - BP 46 -
38402, Grenoble, France.
Tel: +33 (0) 4 76 82 62 52
1

---

[1] Abbreviations: AEA = Arithmetic Ensemble Average, CSTP = Common Spatio-Temporam Pattern, ACSTP=Adaptive CSTP.

Abstract

Background

Already used at the incept of research on event-related potentials (ERP) over half a century ago, the arithmetic mean is still the benchmark for ERP estimation. Such estimation, however, requires a large number of sweeps and/or a careful rejection of artifacts affecting the electroencephalographic recording.

New Method

In this article we propose a method for estimating ERPs as they are naturally contaminated by biological and instrumental artifacts. The proposed estimator makes use of multivariate spatio-temporal filtering to increase the signal-to-noise ratio. This approach integrates a number of relevant advances in ERP data analysis, such as single-sweep adaptive estimation of amplitude and latency and the use of multivariate regression to account for ERP overlapping in time.

Results

We illustrate the effectiveness of the proposed estimator analyzing a dataset comprising 24 subjects involving a visual odd-ball paradigm, without performing any artifact rejection.

Comparison with Existing Method(s)

As compared to the arithmetic average, a lower number of sweeps is needed. Furthermore, artifact rejection can be performed roughly using permissive automatic procedures.

Conclusion

The proposed ensemble average estimator yields a reference companion to the arithmetic ensemble average estimation, suitable both in clinical and research settings. The method can be applied equally to event related fields (ERF) recorded by means of magnetoencephalography. In this article we describe all necessary methodological details to promote testing and comparison of this proposed method by peers. Furthermore, we release a MATLAB toolbox, a plug-in for the EEGLAB software suite and a stand-alone executable application.

Keywords

1.1 Introduction

A substantial amount of studies using electroencephalography (EEG) concerns event-related potentials (ERPs). ERPs are electrical potential fluctuations displaying stable time relationship to some physical, mental, or physiological occurrence, referred to as the *event*. They are often described in the time-domain as a number of positive and negative waves (peaks) characterized by their polarity, shape, amplitude, latency and spatial distribution on the scalp. All these characteristics depend on the type (class) of event and constitute the object of the so-called time-domain ERP analysis (Lopes da Silva, 2010; Picton et al., 2000). Influential discoveries on human cognitive processes have been done within this framework starting half a century ago with the pioneering observations of the contingent negative variation (Walter et al., 1964), the P300 (Sutton et al., 1965), the readiness potential (or Bereitschaspotential: Deecke et al., 1969) and the mismatch negativity (Näätänen et al., 1978). Thanks to these discoveries ERPs today are of paramount importance in the EEG research as a whole and find highly specific clinical applications.

For decades ERPs have been conceived as stereotyped fluctuations of electrical potentials time and phase-locked to an event. By stereotyped we mean that they have been considered as having fixed polarity, shape, latency, amplitude and spatial distribution. According to this view, the ERP is considered independent from the ongoing EEG and simply add to the latter (*additive model*). Recently, other generative models for ERP have been proposed. Based on the seminal findings of Sayers and Beagley (1974), Sayers et al. (1974) and McClelland and Sayers (1983), it has been proposed that evoked activity may consist, at least partially, of an enhanced alignment of phase components (*phase resetting model*) of the ongoing neuronal activity (Jansen et al., 2003; Lopes da Silva, 2006; Makeig et al., 2002, 2004). From another perspective, Mazaheri and Jensen (2010) pointed out that ongoing EEG activity is commonly non-symmetric around zero, as can be seen clearly in sub-dural recordings of alpha rhythms (Lopes da Silva et al., 1997). They proposed that this kind of amplitude asymmetric oscillation may be considered as resulting from rhythmic bouts of inhibition constituting non-sinusoidal waveforms, specifically in the 8-13Hz frequency band; averaging these waveforms may create evoked responses with slow components (*amplitude asymmetry model*). Along these lines, it has been emphasized that EEG oscillations with non-zero mean may cause a baseline shift that remains after averaging, and thus may contribute to evoked responses (Nikulin et al., 2010). In an editorial where these and other fundamental aspects of the generation of evoked responses are analyzed, de Munck and Bijma (2010) stressed the need for a mathematical framework enabling the comparison of different models. We may add that a specific understanding of the biophysics underpinning the generation of ERPs is an essential prerequisite for the conception of an adequate signal processing method to analyze ERP data. In any case, it is necessary to have a comprehensive analysis method to be able to detect and characterize ERPs in the time domain. In this paper we present an approach contributing to this general

objective, which, while making the working assumption that the evoked ERP components are additive, can also be applied if they are generated according to other models, as long as their *spatial and temporal patterns are prevalently stable* across sweeps.

The amplitude of ERP peaks amounts to a few μV, whereas the amplitude of the on-going (spontaneous) EEG signal may be as high as several tens of μV. Thus, even ignoring the unavoidable presence of artifacts in EEG recordings, the signal-to-noise variance ratio (SNR) of a single sweep may be extremely low. This is a major challenge in time-domain ERP analysis, which therefore requires dedicated filtering and/or ensemble averaging methods (Luck, 2014; Regan, 1989). It should be noted that for a given class of ERP, only the polarities of the peaks may be considered consistent for a given EEG recording montage; the shape, latency, amplitude and spatial distribution of ERPs are highly variable among individuals. Furthermore, even if within an individual the shape may be assumed stable, there may be a non-negligible amplitude and latency variability from sweep to sweep. The sweep-to-sweep latency and amplitude variability of ERPs springs from the combination of several instrumental, experimental and biological factors.

An additional difficulty in ERP analysis arises when we record overlapped ERPs, which causes a bias in the usual ensemble average estimations (Ruchkin, 1965; Smith and Kutas, 2015; Woldorff, 1988, 1993). ERP 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 (Congedo et al., 2011) and rapid image triage (Yu et al., 2012), which are heavily employed in brain-computer interfaces for increasing the information transfer rate (Wolpaw and Wolpaw, 2012), and in the study of eye-fixation potentials, where the event is the time of an eye-fixation in between saccades and the fixations rapidly follow each other (Sereno and Rayner, 2003).

Despite the fact that the literature on ERP estimation is extensive, still today in the community there is no consensus on a standard methodology for addressing the estimation of the ensemble average (Lopes da Silva, 2010; Picton et al., 2000). As a matter of fact, only a few proposed techniques have enjoyed a long-lasting and widespread popularity. This is the case of Woody's adaptive latency correction (Cabasson and Meste, 2008, Gasser et al., 2003; Picton et al., 2000; Souloumiac and Rivet, 2013; Woody, 1967), ERP filtering based on spatial or temporal principal component analysis (Chapman and McCrary, 1995; Dien, 2010; Lagerlund et al., 1997) and of the arithmetic ensemble averaging (Lopes da Silva, 2010). Indeed, an efficient way to increase the SNR of ERPs is to apply a spatio-temporal filter. In this spirit, principal component analysis (PCA, see Cichocki and Amari, 2002) has been applied in ERP analysis since half a century to improve the arithmetic average (Donchin, 1966; John et al. 1964). Given an estimation of the average ERP response, one may seek a linear combination of the derivations (spatial PCA) or a linear combination of the samples (temporal PCA), discarding those spatial or temporal components that are irrelevant for the estimation (noise). A long-lasting debate has concerned

the choice of the spatial or temporal PCA (Dien, 2010; Picton et al, 2010). Actually, in the case of ERPs the debate is pointless since one can perform a spatial and temporal PCA at the same time. However, PCA in general is not very effective in suppressing ERP noise. This is due to the fact that PCA neglects the spatio-temporal structure of the noise, which in ERP data is far from being white (uncorrelated), neither spatially, nor temporally. Furthermore, PCA is sensitive to high-amplitude artifacts such as eye-movements and some instrumental artifacts. A much sharper filter can be obtained by first pre-whitening the data and thereafter applying a spatio-temporal (bilinear) PCA. This is what we obtain by adapting to ERPs the popular common spatial pattern (CSP) filter and extending it to a bilinear setting.

First introduced in the EEG literature by Koles (1991), Koles and Soong (1998) and Ramoser et al. (2000), the CSP has since enjoyed a widespread popularity and has been extended in several ways (e.g., Dornhege et al., 2006; Lemm et al., 2005; Lotte et al., 2010; Townsend et al., 2006; Wang and Zheng, 2008). The CSP is commonly used for optimal separation of two classes of data. This is achieved finding the spatial filters that maximize the variance of one class with respect to the other (whereas the PCA maximizes the signal variance). None of these early works, however, has addressed the use of CSP for ERP data. The adaptation of the CSP to ERP data must be specifically designed in order to maximize the SNR between the estimated ensemble average ERP and the background noise, including all electrical potentials that are not stable (time- and phase-locked) across sweeps. Such a CSP specifically conceived for the analysis of ERPs has been proposed by Rivet et al. (2009), who also formalized an appropriate ensemble average estimation for overlapping ERPs introducing the use of a multivariate regression model. A later extension of this method accounted for latency variability as well (Souloumiac and Rivet, 2013). Yu et al. (2012) extended the idea to the spatio-temporal filtering setting (CSTP). However, their approach does not treat the overlapping case and does not consider the latency variability, left alone the amplitude variability of ERP peaks. Also, these latter authors resort to an iterative solution, whereas, as we will show, a closed-form algebraic solution exists also for the bilinear case. Accordingly, in this article we provide a closed-form solution to the common spatio-temporal pattern (CSTP) for ERPs, accounting at the same time for both amplitude and latency variability and treating appropriately the case of overlap between ERPs. Moreover, in contrast to previous attempts, we provide a heuristic procedure to estimate the optimal number of spatio-temporal components, yielding an automatic algorithm. Our method effectively combines and extends a number of previous propositions in an integrated framework. Since the algorithm is fully adaptive (data-driven), it automatically adjusts to the data set of interest. For instance, it is able to automatically remove outliers and common sources of artifacts, even if they are large and tend to be time-locked to the stimulus, which is sometimes the case of eye movements and blinks (Picton, 2000). However, when the data are clean of such artifacts, the CSTP estimation is very close to the arithmetic average (Fig. 2), as we require. Conceptually, our approach remains rather simple, albeit sharp from a signal processing perspective; it amounts to perform a spatio-temporal PCA on data that have been spatio-temporally pre-whitened, estimating adaptively

5

optimal time-shifts (latency) and weights (amplitude) for each sweep. As far as computational complexity is concerned, it is moderate. Thanks to these characteristics, we are confident that the proposed framework may provide a useful companion method for ERP analysis in the time domain, complementing the arithmetic mean. To encourage the testing of our proposition we provide a MATLAB toolbox (available at : http://louis-korczowski.org/) a stand-alone executable application (available at https://sites.google.com/site/marcocongedo) and a plug-in for the widespread suite for EEG data analysis EEGLAB (Delorme and Makeig, 2004).

1.2 Material and Methods

The case of overlapping ERPs leads to a general formulation, which reduces to a simpler one for the non-overlapping case. For clarity of exposition we present first the simplification engendered by the non-overlapping nature of the recorded ERPs and the general case thereafter; the first is simpler to understand, the second is what needs to be implemented using a multivariate regression framework. In the sequel we denote matrices by upper case italic characters (A), vectors, variables and integer indices by lower case italic characters (a) and constants by upper case characters (A). A set of objects is enclosed in curly brackets such as for index $n \in \{1,\ldots,N\}$. For the sake of notation simplicity, the column vectors of matrix A and its elements are both indicated by the same letter used for the matrix, but in lower case, i.e., $A=[a_1\ a_2\ \ldots]$, and a_{ij} is the (i,j) element of A. For a diagonal matrix D, the nth diagonal element will be denoted simply as d_n. The same symbol will be used for indices and for the upper bound in summation, the latter being omitted if there is no possible confusion, thus Σ_n will always indicate a summation over all elements in the set $n \in \{1,\ldots,N\}$. We denote by $(\cdot)^T$, $tr(\cdot)$, $(\cdot)^{-1}$, $\|\cdot\|_F$ and $\lambda_n(\cdot)$ the transpose of a matrix, its trace, inverse, Frobenius norm and n^{th} eigenvalue, respectively. The identity matrix of dimension P·P is denoted by I_P. The rest of the notation should be clear from the context.

1.2.1 Data Model

In this article we treat the analysis of ERP at the single-subject level, thus we will not introduce an index for the subjects. In the sequel we will denote by $z \in \{1,\ldots,Z\}$ the index of Z ERP recorded classes (for example, in an odd-ball P300-based experiment we have Z=2 because we have a target and a non-target class of ERP) and by $k \in \{1,\ldots,K_z\}$ the index of the K_z sweeps recorded for class z.

1.2.2 Model for non-overlapping ERPs

In the case of non-overlapping ERPs we employ an additive multivariate ERP generation model taking into account the single-sweep amplitude and latency variability, such as

$$X_{zk} = s_{zk} \, Q_z\!\left(t_{zk}\right) + N_{zk} \, . \tag{1.1}$$

In the model above, $X_{zk} \in \mathfrak{R}^{N \cdot T}$ is the kth observed sweep belonging to the zth class, comprised of N derivations (non-reference electrodes) and T>N samples, $N_{zk} \in \mathfrak{R}^{N \cdot T}$ holds the residual EEG signal, modeled here as *additive noise* and Q_z is a stereotyped ERP response for a given class, which amplitude s_{zk} and latency t_{zk} are continuously modulated across sweeps by instrumental, experimental and biological factors. Notice that the 'noise' term here includes background EEG signal plus instrumental, environmental and biological artifacts (Congedo et al., 2008), which explains the overwhelming majority of the observed single-sweep variance. According to this model, the single-sweep SNR (signal-to-noise ratio) is the ratio between the variance of $s_{zk} \, Q_z\!\left(t_{zk}\right)$ and the variance of N_{zk}.

1.2.3 Model for overlapping ERPs

It is known since a long time that model (1.1) is no more adequate in the case of overlapping ERPs (Ruchkin, 1965; Woldorff, 1988). Following Souloumiac and Rivet (2013), instead we use the more generic model formulation

$$X = Q_1 T_1 + \cdots + Q_Z T_Z + N \, , \tag{1.2}$$

where $X \in \mathfrak{R}^{N \cdot L}$ is the matrix holding the entire EEG recording of the session, with L the total number of samples, matrices $Q_1,...,Q_Z \in \mathfrak{R}^{N \cdot T}$ hold the fixed stereotyped ERP response for the Z classes as in model (1.1), matrices $T_1,...,T_Z \in \mathfrak{R}^{T \cdot L}$ are Toeplitz matrices and $N \in \mathfrak{R}^{N \times L}$ holds the residual additive noise, as in model (1.1), but now altogether for the whole recording. The elements of the Toeplitz matrices $T_1,...,T_Z \in \mathfrak{R}^{T \cdot L}$ are zero everywhere, except on the diagonals starting at the K_z columns corresponding to the ERP events for the corresponding class, shifted by t_{zk} samples, diagonals on which their element is s_{zk}. Hence, the Toeplitz matrices have as many non-zero diagonals as many sweeps there are in their class. Their role in (1.2) is just to extract the sweeps from the stream of data with the appropriate time shifts and weights. As we will see, they allow a least-square ensemble average estimation based on multivariate regression of the sweeps even if they overlap; this overlap may be of any form, that is, any overlapping of ERPs can be accounted for, either of the same as well as of different classes. This model can be written in compact form as (Souloumiac and River, 2013)

$$X = QT + N \, , \tag{1.3}$$

where $Q = [Q_1 \cdots Q_Z] \in \Re^{N \cdot ZT}$ is obtained stacking horizontally the Q_z matrices and $T = \left[T_1^T \cdots T_Z^T \right]^T \in \Re^{ZT \cdot L}$, named the *design matrix* (for instance, in fMRI studies), is obtained stacking vertically the T_z matrices.

Note that both models (1.1) and (1.3) assume the knowledge of latencies t_{zk} and amplitudes s_{zk}, but these are unknown. We will consider then the adaptive estimation of corresponding time shifts and weights coefficients, which we will indicate by τ_{zk} and σ_{zk}, respectively. We precede initializing the time shifts to zero and the weights to $1 / \|X_{zk}\|_F$, then adaptively estimating them[2]. This yields a tractable model, yet, it ensures that both the time shifts and weights are appropriately estimated from the data. Hereafter let

$$X'_{zk} = \sigma_{zk} X_{zk} (\tau_{zk}) \tag{1.4}$$

be the kth sweep belonging to class z, weighted by σ_{zk} and time-shifted by τ_{zk}, where the weight is a strictly positive real number and the time-shift is an integer in EEG sample units.

1.2.4 Spatio-Temporal Sweep Filtering

For any given subject our objective is to find *bilinear transformations of the sweeps* maximizing the SNR of the ERP, that is, to find for a given class *z,* a pair of matrices $B_z \in \Re^{N \cdot P_z}$ and $D_z \in \Re^{T \cdot P_z}$ transforming the sweeps such as

$$Y_{zk} = B_z^T X'_{zk} D_z \in \Re^{P_z \cdot P_z} . \tag{1.5}$$

In the above expression B_z is a collection of P_z *spatial filters,* i.e., acting as a linear combination of the data across derivations, while D_z is a collection of P_z *temporal filters,* i.e., acting as a linear combination of the data across samples. They act together in pairs, constituted by the corresponding columns of B_z and D_z, seeking spatio-temporal combinations of the data. The integer $P_z \le N$ is named the *subspace dimension*, where the subscript *z* reminds that the optimal value P_z is possibly different in different classes (1.5). Note that we may require the temporal subspace dimension to be superior to the spatial subspace dimension, or vice versa. In this article, however, we require them being equal.

---

[2] Such initialization for the weights exploits the fact that the noise part of X_{zk} is much larger than the ERP part.

Together with spatial and temporal filters we also obtain the corresponding spatial and temporal *patterns* $A_z \in \mathfrak{R}^{\text{N·P}_z}$ and $E_z \in \mathfrak{R}^{\text{T·P}_z}$, which are found by (pseudo-) inverting the filters, i.e., so as to verify

$$B_z^T A_z = D_z^T E_z = I_{\text{P}_z}.$$ (1.6)

These patterns can be plotted to visualize the time course and scalp map of the components, respectively. Everything we have said about spatial and temporal filters in the case of non-overlapping ERPs applies as well in the case of overlapping ERPs. In this case however, we need to subtract from (1.5) the other ERPs overlapping within the time window of sweep X_{zk}.

1.2.5 Single-Sweep Estimation

The filtered sweep in the original data space, that is, the single-sweep estimation, is given by

$$\hat{X}_{zk} = A_z Y_{zk} E_z^T \in \mathfrak{R}^{\text{N·T}},$$ (1.7)

where Y_{zk} is estimated using (1.5). The role of pattern matrices A_z and E_z (1.6) is to project the filtered components Y_{zk} back in the original sensor space.

1.2.6 Unfiltered Ensemble Average Estimations

1.2.6.1 For non-overlapping ERPs

For any given subject

$$\bar{X}_z^{'} = \frac{1}{\sum_{k=1}^{\text{K}_z} \sigma_k} \sum_k X_{zk}^{'} \in \mathfrak{R}^{\text{N·T}}$$ (1.8)

is the *weighted and aligned[3] arithmetic mean* (arithmetic ensemble average: AEA) of observed sweeps for class *z*. This is a least-squares estimation of Q_z in (1.1)[4]. Using equal weights, estimator (1.8) 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 (Lęski, 2002). However these conditions are never matched, since EEG data is both spatially and temporally correlated and contains large outliers and artifacts. In practice, it is well known that the arithmetic mean is an acceptable

---

[3] that is, computed from weighted and time-shifted sweeps as per (1.4).

[4] We say "least-squares" because $\bar{X}_z^{'}$ is the data matrix solving optimization problem $\arg\min_{\bar{X}_z^{'}} \sum_{k=1}^{\text{K}_z} \left\| X_{zk}^{'} - \bar{X}_z^{'} \right\|_F^2$.

ensemble average estimator only if sweeps with low SNR are removed and/or a large number of sweeps is available. The estimator we propose overcomes this limitation. Notice that a sweep features low SNR if the noise part of the signal is high (for example, in the case of large artifacts superimposed to the sweep), but also if the ERP amplitude is low for whatever instrumental, experimental or biological reason affecting that sweep.

1.2.6.2 For overlapping ERPs

In the case of overlapping ERPs the least-square estimation of the ensemble average ERP can be obtained using multivariate regression techniques. Extending the result of Souloumiac and Rivet (2013) in order to allow the estimation of weighted and aligned arithmetic means, this is obtained by

$$
\left[\bar{X}_1' \cdots \bar{X}_Z' \right] = XT^T \left(TT^T \right)^{-1} \begin{bmatrix} j_1 I_{\mathrm{T}} & \cdots & 0 \\ & \ddots & \\ 0 & & j_Z I_{\mathrm{T}} \end{bmatrix} \in \Re^{N \cdot ZT} , \tag{1.9}
$$

where, $j_z = \sum_{k=1}^{K_z} \sigma_{zk}^2 \Big/ \sum_{k=1}^{K_z} \sigma_{zk}$ and X, T have been defined in (1.3). The estimation of the ensemble average for each class is obtained as the corresponding partition of $\left[\bar{X}_1' \cdots \bar{X}_Z' \right]$ (see Appendix A).

1.2.7 Filtered Ensemble Average Estimations

Using the spatio-temporal filtering proposed in this article, the filtered, weighted and aligned ensemble average provides an estimation of the average stereotyped response Q_z and will be given by

$$
\hat{Q}_z = A_z B_z^T \bar{X}_z' D_z E_z^T \tag{1.10}
$$

no matter if we employ ensemble average estimation (1.8) or (1.9).

1.2.8 Bilinear Common Pattern (Common Spatio-Temporal pattern)

For class z let

$$
\bar{C}_{z(S)} = COV\left(\bar{X}_z' \right) \in \Re^{N \cdot N}, \quad \bar{C}_{z(T)} = COV\left(\bar{X}_z'^T \right) \in \Re^{T \cdot T} \tag{1.11}
$$

be the spatial (left-hand side) and temporal (right-hand side) sample covariance matrices of the ensemble average ERP estimation (1.8) or (1.9), depending on whether we are working with overlapping or non-overlapping ERPs, respectively, and let

$$C_{(S)} = \frac{1}{\sum_z K_z} \sum_z \sum_{k=1}^{K_z} COV(X_{zk}) \in \Re^{N \cdot N}, \quad C_{(T)} = \frac{1}{\sum_z K_z} \sum_z \sum_{k=1}^{K_z} COV(X_{zk}^T) \in \Re^{T \cdot T} \quad (1.12)$$

be an estimation of the spatial (left-hand side) and temporal (right-hand side) sample covariance matrix of the data in all classes. Note that in (1.11) we compute the covariance matrices of the sweep average, whereas in (1.12) we are averaging the covariance matrices of the sweeps. Thus the estimations in (1.11) and in (1.12) are not at all equivalent, even if the data comprise one class only and regardless of the use of weights and time-shifts. In particular, because of sweep averaging, the variance in (1.11) can be decomposed as the variance of the ERP response plus the variance of the noise reduced proportionally to the number of sweeps by averaging. On the other hand, the variance in (1.12) can be decomposed as the variance of the ERP response plus the total variance of the noise. Thus, the variance in (1.11) concerns mostly phase-locked (evoked) potentials, since the non-phase locked potentials are reduced by averaging, while the variance in (1.12) concerns all processes, regardless whether they are phase-locked or not, i.e., it concerns both evoked and induced activity (see Tallon-Baudry et al., 1996 and Başar-Eroglu et al., 1992) as well as ongoing EEG and artifacts. Accordingly, the relevant definition of SNR for the average, phase-locked ERP to the event of class z is

$$SNR(\bar{X}_z) = \frac{var(\bar{X}_z')}{\frac{1}{\sum_z K_z} \sum_z \sum_{k=1}^{K_z} var(X_{zk})}, \quad (1.13)$$

where the sum in the denominator is carried out over all sweeps of all classes and where *var* indicates the total variance of the data matrix argument, e.g., $var(X) = tr(XX^T) = \|X\|_F^2$. Our objective here is to find a bilinear transformation of the data maximizing this ratio, i.e., to find two matrices $B_z \in \Re^{N \cdot P_z}$ and $D_z \in \Re^{T \cdot P_z}$ solving, for a given subspace dimension P_z and given latency and time-shift correction, the following optimization problem

$$\underset{B_z, D_z | P_z, \sigma_{zk}, \tau_{zk}}{\arg \max} \frac{var(B_z^T \bar{X}_z' D_z)}{\frac{1}{\sum_z K_z} \sum_z \sum_{k=1}^{K_z} var(B_z^T X_{zk} D_z)}. \quad (1.14)$$

This problem is an extension to the weighted and aligned spatio-temporal setting of the aligned common spatial pattern problem addressed in Souloumiac and Rivet (2013). Note that these authors include in the denominator the spatial covariance matrix of the whole EEG recording, while here, in concordance with the bilinear setting, we include in the denominator only ERP data segments. It is also an extension of the problem addressed by Yu et al. (2012) regarding the weighted, time-shifted and overlapping ERPs setting. These latter authors have proposed an iterative algorithm alternating between the update of B_z and of D_z using the relation between the SVD (singular-value decomposition) and the EVD

(eigenvalue-eigenvector decomposition). With such an approach the computational cost is much higher as compared to an analytic solution. In appendix B we show that an algebraic solution to the bilinear common pattern is possible with a two-step procedure. Accordingly, the CSTP algorithm is summarized here below:

CSTP Algorithm

Input: For class z,
- a set of weights σ_{zk} summing to K_z (optional, they may be all equal to one)
- a set of time shifts τ_{zk} (optional, they may be all zero).
- a set of K_z sweeps $\{X_{z1}, \ldots, X_{zK_z}\}$
- the ensemble average ERP estimation $\overline{X}'_z \in \Re^{\text{N·T}}$, obtained by (1.8) or (1.9), depending whether we are in the non-overlapping or overlapping case, respectively, with N the number of electrodes and T>N the number of samples.
- matrices $C_{(S)} = \frac{1}{\sum_z K_z} \sum_z \sum_{k=1}^{K_z} COV\left(X_{zk}\right), \qquad C_{(T)} = \frac{1}{\sum_z K_z} \sum_z \sum_{k=1}^{K_z} COV\left(X_{zk}^T\right)$
- the subspace dimension P_z<N.

Output: For class z, a spatial and temporal filter $(B_z, A_z) \in \Re^{\text{N·P}_z}$ and pattern $(D_z, E_z) \in \Re^{\text{T·P}_z}$

Algorithm:

Let $\quad U\Phi U^T$ and $V\Psi V^T$ be the EVD of $C_{(S)}$ and $C_{(T)}$, respectively. $\hfill (1.17)$

Define $\quad F_{(S)} \leftarrow \left(\phi_1^{-\frac{1}{2}}u_1 \cdots \phi_Q^{-\frac{1}{2}}u_Q\right) \in \Re^{\text{N·Q}}, \qquad F_{(T)} \leftarrow \left(\psi_1^{-\frac{1}{2}}v_1 \cdots \psi_R^{-\frac{1}{2}}v_R\right) \in \Re^{\text{T·R}}$ and $\hfill (1.18)$

$\quad G_{(S)} \leftarrow \left(u_1 \phi_1^{\frac{1}{2}} \cdots u_Q \phi_Q^{\frac{1}{2}}\right) \in \Re^{\text{N·Q}}, \qquad G_{(T)} \leftarrow \left(v_1 \psi_1^{\frac{1}{2}} \cdots v_R \psi_R^{\frac{1}{2}}\right) \in \Re^{\text{T·R}}, \hfill (1.19)$

where Q≤N and R≤N are the largest indices verifying ϕ_Q, $\psi_R > \frac{1}{10^6}\lambda_1\left(C_{(S)}\right)$, respectively.

Let $\quad \Pi_z W_z \Xi_z^T$ be the SVD of $\overline{Z}'_z = F_{(S)}^T \overline{X}'_z F_{(T)}$. $\hfill (1.22)$

$\quad B_z \leftarrow F_{(S)}\tilde{\Pi}_z, \qquad D_z \leftarrow F_{(T)}\tilde{\Xi}_z, \qquad A_z \leftarrow G_{(S)}\tilde{\Pi}_z, \qquad E_z \leftarrow G_{(T)}\tilde{\Xi}_z, \hfill (1.23)$

where $\tilde{\Pi}_z$ and $\tilde{\Xi}_z$ hold in columns the first P_z<N column vectors of Π_z and Ξ_z, respectively.

In the CSTP algorithm above, the weights σ_{zk} and the time-shifts τ_{zk} are provided by the user along with the subspace dimension P_z. Next, we show how all these quantities can be estimated adaptively from the data to achieve a fully adaptive and fully unsupervised algorithm.

1.2.9 Estimating the Optimal Subspace Dimension

The subspace dimension P_z should be large enough to include all spatio-temporal components describing the ERP for class z and small enough to effectively suppress noise. Note that if a temporal

filter is employed, even setting $P_z=N$ we filter out some noise, specifically, the noise explained by the T-P_z temporal components not included to construct the temporal filter E_z in (1.23). Still, we will require $P_z<N$, i.e., the algorithm should always remove at least one spatio-temporal component. The unsupervised estimation of the optimal subspace dimension is known to be a difficult task (Cichocki and Amari. 2002). Hereby we propose a solution by injecting some a-priori information, as an educated guess, about where in space and time the ERP under analysis produces noticeable potentials. For doing so, let us consider a subset of derivations and time samples in the ERP data. We name such a subset a *mask*. For example, if we aim at estimating a visual P300 ERP, the mask will typically include time samples around 300-500 ms post stimulus. The locations, typically including mid-central, parietal and occipital derivations, are highly subject-dependent. The filtered estimation of the ensemble average $\hat{Q}_z$ is given by (1.10). Let $\Upsilon_z\left(\hat{Q}_z\right)$ be the matrix $\hat{Q}_z$ with all elements that do not belong to the mask set to zero. We can then look for a value of P_z maximizing locally

$$SNR\left(\Upsilon_z\left(\hat{Q}_z\right)\right) = \left\|\Upsilon_z\left(\hat{Q}_z\right)\right\|_F^2 \Big/ \left\|\hat{Q}_z\right\|_F^2 \, , \tag{1.15}$$

which is the ratio between the variance of the ensemble average within the mask (the spatio-temporal window of interest) and the variance in the whole ensemble average. First, we smooth the estimates (1.15) by a 3-point moving average, then we estimate the optimal P_z as the highest value corresponding to a local maximum of the SNR (1.15) not inferior to 66% of its global maximum, in the set {2,…, N-1}. Please note that here and hereafter for whatever function, by "*valid local maximum*" we will refer to local maxima found by this procedure.

1.2.10 Amplitude Estimation and Sweep Adaptive Weighting

For a given class *z*, the amplitude variability can be addressed adaptively using an appropriate measure of the signal-to-noise ratio (SNR) for each sweep. Given the estimation of the spatio-temporal filters, such a measure is

$$SNR\left(X_{zk}\right) = \left\|A_z B_z^T X_{zk} D_z E_z^T\right\|_F \Big/ \left\|X_{zk} - A_z B_z^T X_{zk} D_z E_z^T\right\|_F \, . \tag{1.16}$$

Referring to model (1.1), the closer is $A_z B_z^T X_{zk} D_z E_z^T$ to $s_{zk} Q_z\left(t_{zk}\right)$, the higher (1.16) will be. Also, ratio (1.16) drops if the sweep contains large artifacts or outliers, since artifacts and outliers (and in general, the noise part of the data) is present in the denominator, but is reduced in the nominator, due to filtering. Thus, ratio (1.16) is a comprehensive and adaptive choice for weighting the individual sweeps and the ensuing weighted ensemble average estimation. It is particularly effective to suppress artifacts

and outliers. Although not necessary, in this article the weights are always normalized so that their average is 1.0 (see step (*B*) here below). The algorithm for estimating the weights is described below:

WEIGHTS ESTIMATION

Input: For class z,

- the K_z raw sweeps X_{kz}

- the spatio-temporal filters and patterns A_z, B_z, D_z, E_z estimated by the CSTP.

Output: For class *z*, the set of weights σ_{zk}, for $k \in \{1,\dots,K_z\}$

Algorithm:

(*A*) - For every sweep do (for *k*:=1 to K_z) $\sigma_{zk} \leftarrow \left\| A_z B_z^T X_{zk} D_z E_z^T \right\|_F \Big/ \left\| X_{zk} - A_z B_z^T X_{zk} D_z E_z^T \right\|_F$

(*B*) - For every sweep do (for *k*:=1 to K_z) $\sigma_{zk} \leftarrow \left(K_z \sigma_{zk} \right) \Big/ \sum_{k=1}^{K_z} \sigma_{zk}$

1.2.11 Sweep Adaptive Latency Correction

In order to account for the latency variability we use an improved version of the classical Woody's adaptive filtering method (Woody, 1967). The original method iteratively estimates a time shift τ_{zk} for each sweep X_{zk}, attempting to align the sweeps as much as possible to improve the SNR of the sweeps ensemble average estimation. At each iteration, the sweeps average is estimated and the shift for each sweep is updated as the shift resulting in the maximal correlation between the sweep and the average. The iterations are repeated until the shifts become negligible. Woody's method has been used extensively and is known to perform fairly well provided that the SNR of the sweeps is high enough.

Three relevant improvements have been proposed since the classical proposition of Woody (1967) and have proved effective: the first is to use the maximum covariance instead of the maximum correlation for estimating the shift (Jaśkowski and Verleger, 1999). The second consists in applying the method on spatially filtered data instead of on raw data, so as to place the method in a more favorable SNR condition (Souloumiac and Rivet, 2013). The third, springing from a maximum likelihood analysis, improves the shift estimation for each sweep by computing the covariance between the current sweep and the ensemble average estimated eliminating the current sweep (Cabasson and Meste, 2008). Here we combine all these improvements and we extend the method to the bivariate filtering case. Furthermore, we can refine the sweep alignment by using the mask operator Υ_z already introduced for estimating the optimal subspace dimension (1.15). Yet, the use of the mask here is optional and can be omitted. The algorithm for estimating adaptively the time shifts is described here below:

The time shift estimation algorithm has two free parameters, i and E. The maximum number of iterations (i) can be blindly set as indicated without affecting much the result. On the other hand, the maximum allowed time-shift (E) should be chosen taking into consideration the specific characteristics of the data, namely the instrumental jitter and the class of ERPs to be estimated. We will come back to this in the discussion section.

1.2.12 The ACSTP Algorithm

Having described all necessary building blocks, we can finally present the fully adaptive ERP estimation algorithm as a whole. We name the fully adaptive algorithm ACSTP, which stands for *Adaptive Common Spatio-Temporal Pattern*. An estimation of the weights and time shifts will improve the ensemble average estimation, hence the CSTP estimation of the filters, which depends on those

estimates (note that the SVD in (1.22) depends on the weighted and aligned ensemble average estimation in (1.21)). In addition, the time-shift estimation will depend upon the weights, but not vice versa, so we always search for the time-shift after estimating the weights. For a given P_z (subspace dimension), the elementary deterministic sequence of operations in our algorithm is then

I)   the estimation of the CSTP filters (initialized with all τ_{zk} set to zero and all $\sigma_{zk} \leftarrow 1/\|X_{zk}\|_F$)

II)   the estimation of the weights σ_{zk}

III)   the estimation of the time shifts τ_{zk} (using the σ_{zk} just estimated and, optionally, the mask)

IV)   a final estimation of the CSTP filters (using time shift τ_{zk} and weights σ_{zk} as just estimated)

We shall run simply this sequence if P_z is selected by the researcher. If an unsupervised estimation of the optimal subspace dimension P_z is sought by using the a mask (1.15), the ACSTP algorithm is as follows:

ACSTP Algorithm *(using a mask)*

Input: the input of algorithms CSTP, WEIGHT ESTIMATION and TIME SHIFT ESTIMATION, initialized and updated as indicated below.

Output: a spatial filter and pattern $(B_z, A_z) \in \Re^{\text{N·Pz}}$, a temporal filter and pattern $(D_z, E_z) \in \Re^{\text{T·Pz}}$,
the set of weights σ_{zk} and time-shifts τ_{zk} estimated on the data,
the automatically chosen subspace dimension P_z.

Algorithm:

FOR p : 2 TO N DO
{

   1) CSTP (with P_z=p, all time-shifts τ_{zk} set to zero and weights initialized as $\sigma_{zk} \leftarrow 1/\|X_{zk}\|_F$)
   2) WEIGHTS ESTIMATION
   3) TIME SHIFTS ESTIMATION (using the mask operator Υ_z, and the weights σ_{zk} as just estimated)
   4) CSTP (with P_z=p, time-shifts τ_{zk} and weights σ_{zk} as just estimated)

   $SNR(p) \leftarrow \left\| \Upsilon_z \left(\hat{Q}_z \right) \right\|_F^2 / \left\| \hat{Q}_z \right\|_F^2$, using (1.10) to estimate $\hat{Q}_z$
}
Smooth the *SNR(p)* value by a 3-point moving average. Let *max* and *min* be the maximum and minimum of the *SNR(p)* so obtained. Normalize the *SNR*s in the interval [0…1], as $SNR(p) \leftarrow (SNR(p) - min)/max$, for *p*: 2 to N. If it exists at least one valid local maximum of the normalized SNR(*p*) (see Section 1.2.9) in the range $p \in \{3,…,N-1\}$, then set P_z to the highest value of *p* corresponding to such a valid local maximum, otherwise set P_z=N-1. Set the ACSTP output to the "4) CSTP" output obtained with such P_z.

Notice that if one does not wish to use weights it is sufficient to omit the corresponding weights estimation step, while if one does not wish to apply latency correction it suffices to omit the time-shift estimation step. If one does not wish to apply neither latency nor amplitude correction, it suffices to omit both steps and to call the CSTP algorithm only once within the *For* loop.

1.2.13 Experimental Data

We have tested the algorithm on a data set comprised of 24 subjects. The data set concerns an experiment carried out at GIPSA-lab in Grenoble using the *Brain Invaders* P300-based brain-computer interface (Congedo et al., 2011). The experiment has been carried out in accordance with the code of ethics of the World Medical Association (Declaration of Helsinki). Participants signed an informed consent form prior to participating to the experiment. The experimental procedure was approved by the ethical committee of the University of Grenoble Alpes (CERNI). The data we use here concerns only the training sessions of this experiment, which is a typical visual odd-ball paradigm. Each training session consisted of five trials comprising eight repetitions of a sequence of 12 flashes. The subjects looked at thirty-six aliens arranged in a 6x6 regular grid slowly moving together from one side of the screen to the other as in the original video game *Space Invaders* (1978, Saito, Japan). A sequence of flashes consisted of 12 flashes of groups of six aliens (Fig. 1B) chosen in such a way that after a sequence of 12 flashes each alien had flashed two times. At each trial the subjects focused on a target alien chosen at random by the computer among the 36 aliens. The target was indicated by a red circle at the beginning of each sequence of flashes. The subject was instructed to mentally count the number of times the target had flashed in each trial (attended deviant odd-ball paradigm, see Kotchoubey, 2015). The screen background color was black. Aliens were all displayed in dark grey color except the target alien, which was always displayed in red color. During a flash, the color of non-target aliens was switched to light grey and the color of the target alien to cyan (the opposite of red on the color circle). Each flash lasted 60 ms. The inter-stimulus interval (ISI), i.e., the interval between two flashes, was randomly drawn from an exponential distribution with mean 1.0, multiplied by 100 to obtain a mean ISI of 100 ms. and bounded in the range {20…500} ms by drawing a random number until it fell in this range. An exponential ISI minimizes the distortion due to the overlap (Ruchkin, 1965). Since there were 40 repetitions of 12 flash sequences in total, the training session resulted in the recording of 400 non-target ERPs and 80 target ERPs. *No artifact rejection whatsoever was performed on the data before entering the analysis.*

Data were acquired using the *Porti* TMSi EEG amplification unit (Twente Medical Systems International B.V., Enschede, the Netherlands) by means of 16 silver/silver-chloride electrodes positioned at location FP1, FP2, F5, AFz, F6, T7, Cz, T8, P7, P3, Pz, P4, P8, O1, Oz, O2 according to

the 10/20 international system, using as reference the left ear-lobe (Jasper, 1958). After acquisition, a IIR notch filter to suppress 50Hz power-line contamination and a non-causal forward-backward FIR Butterworth band-pass filter (4th order) with zero phase distortion in the band-pass region 1-20 Hz were applied. Then data were down-sampled from 512 (the original sampling rate during acquisition) to 128 samples per second using a standard filtering/moving average method. The common-mode rejection of the Porti EEG machine enforces a hardware common-average reference to the EEG potentials. The maximum number of time-shift samples allowed was ±4. As mask Υ_z for the P300 target ERP, five masks accommodated all 24 subjects, as established upon inspection of the arithmetic ensemble average of each subject (table 1). We report our considerations about the choice of a mask in the discussion section.

Table 1. Masks used to estimate the optimal subspace dimension P_z. The AEA (ensemble average estimation) for subjects indicated using bold number characters, which has guided the mask assignment, is shown in Fig. 2.

Subject	Electrode	Time Interval
1, **8**, 9, **10**, 13, 15, 19, 20, 21	Cz, P7, P3, Pz, P4, P8, O1, Oz, O2	50, 550
6, **7**, 22, 23	FP1, FP2, Afz, F3, F4, Cz, P7, P3, Pz, P4, P8	50, 550
2, 3, 5, **11**, 14, 17, 18	FP1, FP2, Afz, F3, F4, Cz, P7, P3, Pz, P4, P8, 01, 0z, 02	50, 550
4, 12, 16	Cz, P7, P3, Pz, P4, P8	50, 550
24	T7, Cz, T8, P7, P3, Pz, P4, P8	50, 550
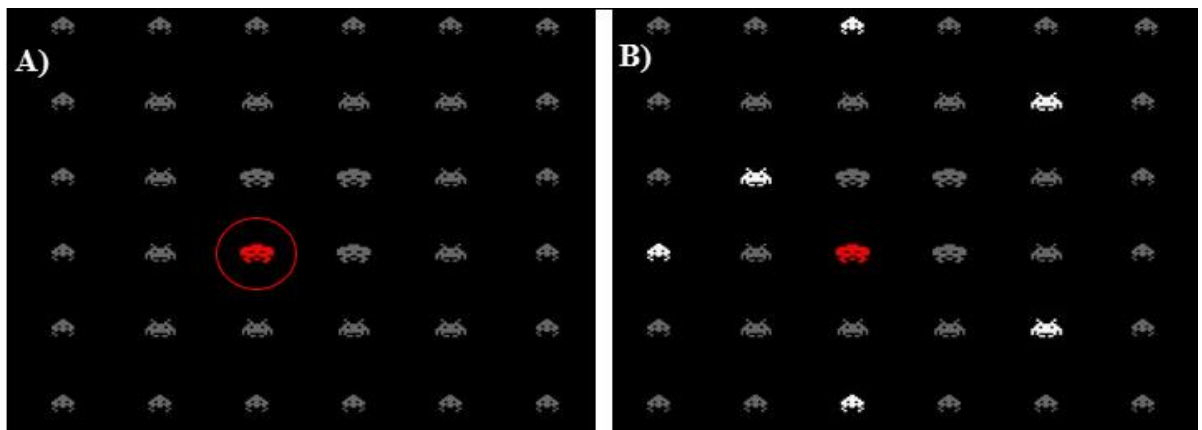
Figure 1. Screen Shots of the "Brain Invaders". Two screen shots of the Brain Invaders in training mode, used to run the experiment. A): at the beginning of each trial the target alien is circled in red; B): random flashing of six aliens (shown in white), in this case, not including the target (hence producing a non-target ERP).

1.3 Results

We first consider the estimation of the target (TA) and non-ratget (NT) class ERP ensemble average. Figure 2 shows the arithmetic ensemble average (AEA: white panels) and the estimation given by the proposed ACSTP algorithm (ACSTP: grey panels) for four representative subjects (ss 8, 10, 7 and 11). Note that for the AEA we do not apply amplitude and latency correction. The subspace dimension chosen by the ACSTP algorithm (P_z) is reported within parenthesis. Concerning the target class, for all four subjects the filtering process preserves or only slightly attenuates the P300. Subjects 7 and 11 are representative of subject featuring a clean AEA; in this case the ACSTP estimation is very close to the AEA estimation, as one would expect. This is reflected in a similar global field power profile for the AEA and ACSTP estimation[5]. Subjects 8 and 10 present with large eye-related artifacts surviving the arithmetic averaging process, better visible at frontal pole locations FP1 and FP2. For these subjects the ACSTP successfully removes the artifacts with good specificity. This can be appreciated, for instance, for ss 10, where the ACSTP estimation at frontal electrodes before 500 ms. has preserved early peaks that one can see at electrodes Cz as well, demonstrating that they are of cerebral origin and as such should not be removed.

The way the automatic choice of the subspace dimension P_z is carried out is illustrated in Fig. 3. The plots show for ss 8, 10, 7 and 11 the smoothed SNR (1.15), the threshold fixed at 66% of the normalized SNR, the global maximum in the range [2,.., N] (cross) and the chosen local maximum estimating P_z found according to the criteria specified in the ACSTP algorithm box section 1.2.12.

The effect of the ACSTP filter is analyzed in Fig. 4, which shows the 16 CSTP temporal components in the form of time-series and the spatial components in the form of topographic maps, again for ss. 8, 10, 7 and 11. These components are, respectively, the columns of matrix E and A of the CSTP filter (1.10). The subspace dimension chosen by the algorithm for subject 8 is 14 (Fig. 2), thus component 15 and 16 have been removed. The temporal maps in Fig. 4 show that these two components consist of very slow potential variations unrelated to the ERP of interest. For subject 10 Pz=13, thus components 14, 15 and 16 have been removed. The spatial maps suggest that these components receive maximal contribution from the frontal pole, thus they are most likely generated by the eyes. A similar analysis and similar conclusions can be carried out for ss 7 and 11. Notice that the components produced by CSTP should not be interpreted directly in terms of physiological phenomena, as the CSTP is not a source separation method. Indeed the components may represent mixtures of physiological sources just as scalp data time-series and topography are. Here we show these components only to illustrate the relevance of the automatic choice of the subspace dimension obtained by using a mask.

---

[5] The global filed power (Lehmann and Skrandies, 1980) is the average absolute amplitude across derivations at each sample. It is very useful to detect phase-locked dipolar fields generating ensemble average estimations.
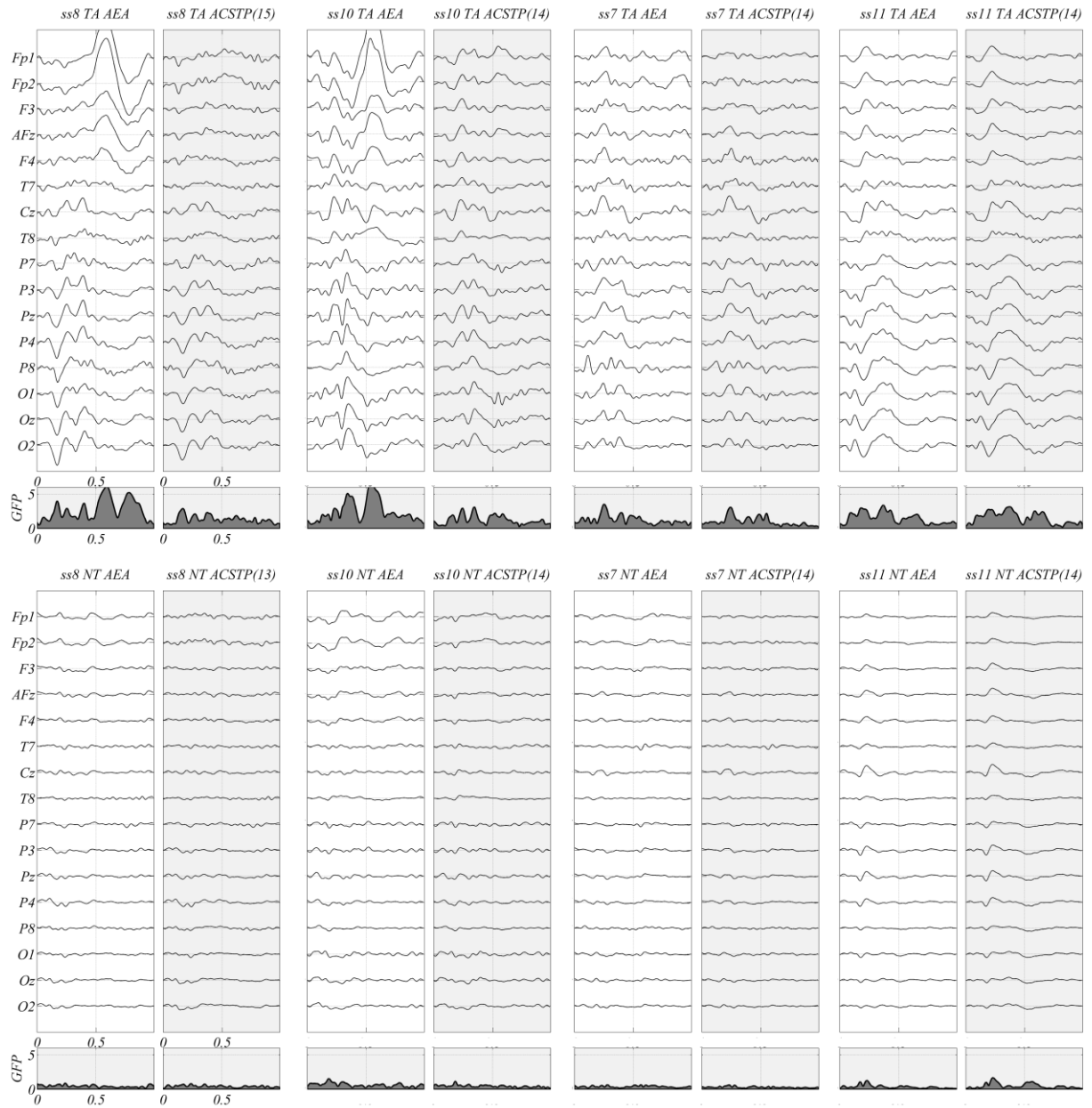
Figure 2. Ensemble Average Estimations. Unweighted and unaligned arithmetic ensemble average (AEA: white panels) and weighted and aligned ensemble average estimated by ACSTP (grey panels) of the 80 target sweeps (TA, top row) and 400 non-target (NT, bottom row) for four representative subjects (ss 8, 10, 7 and 11). Upward deflections with respect to the horizontal lines (zero potential) indicate positive potentials. Each plot comprises one second of data starting at target flashing. For each subject the voltage scaling is arbitrary, but it is the same for all plots within each subject. The shaded area plot is the global field power (GFP).
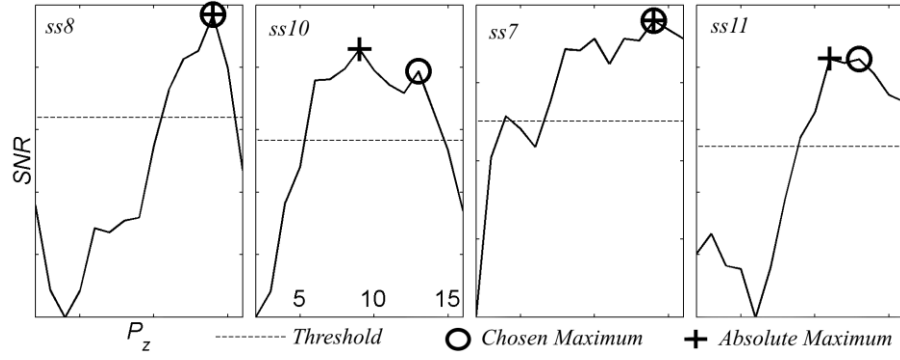
Figure 3. Smoothed Normalized SNR (1.15) used by the ACSTP algorithm to select automatically the optimal subspace dimension Pz. The smoothed and normalized SNR in arbitrary scale obtained for dimensions [2…16], the threshold, defined as 66% of the total SNR, the absolute maximum (cercle) and the maximum chosen according to criteria specified in the ACSTP algorithm box in section 1.2.12, for subjects 8, 10, 7 and 11.
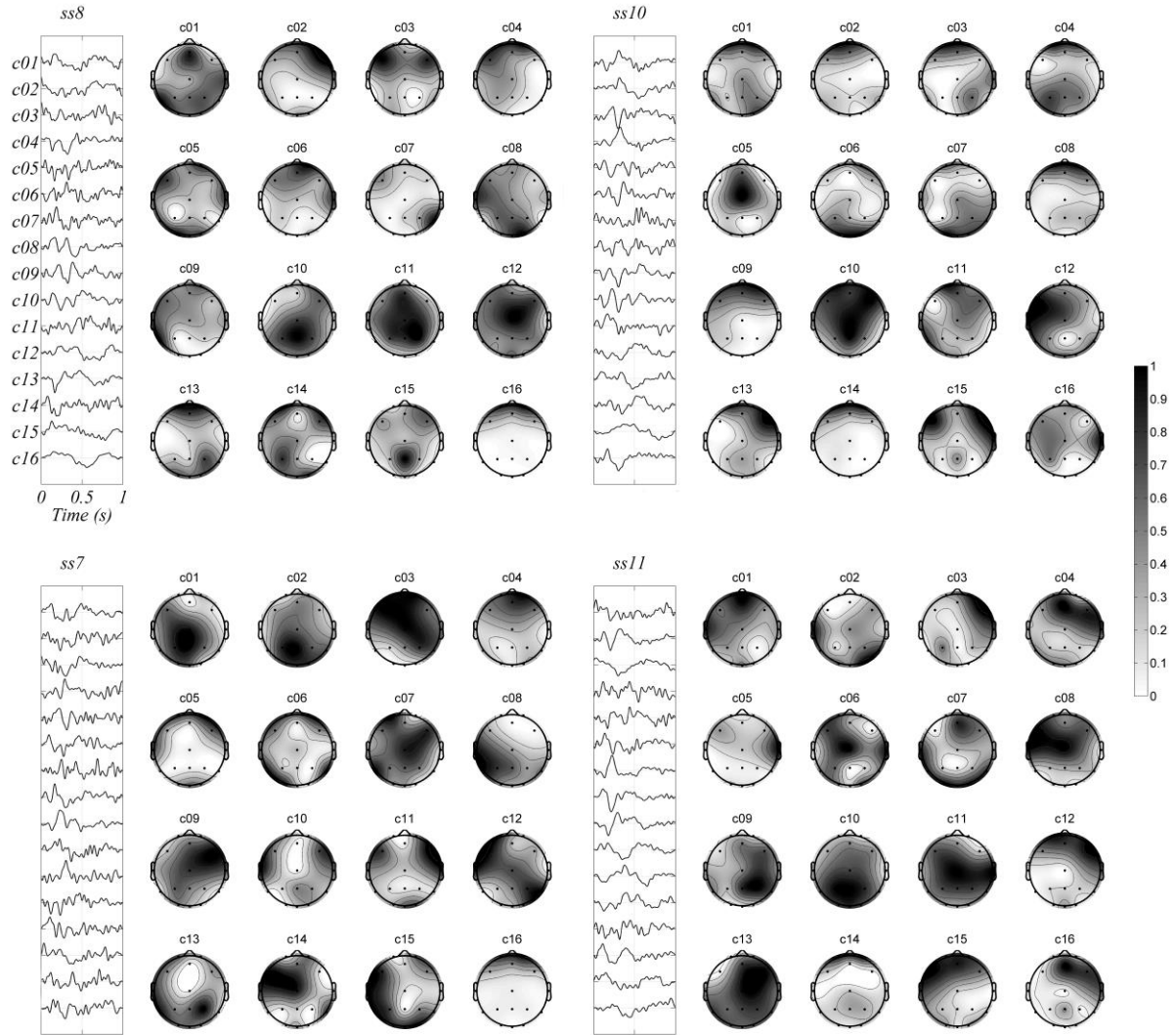
Figure 4. Spatial and Temporal Patterns of the CSTP. The 16 temporal components in the form of time-series and spatial components in the form of scalp topographic maps as found by the CSTP algorithm for ss 8, 10, 7 and 11. Both temporal and spatial maps are displayed on an arbitrary scale. The spatial maps are normalized to the maximum.

The ACSTP algorithm provides an estimation of the ensemble average, thus, as for any estimator, it is important to study its variability. We study this variability as a function of the number of available sweeps. For doing so we use a bootstrap procedure: we randomly draw 600 times 5%, 10%, 15%, 20%, 25% and 50% of the available 80 target and 400 non-target sweeps. This yields a bootstrap sample size of 4, 8, 12, 16, 20 and 40 target sweeps and of 20, 40, 60, 80, 100 and 200 non-target sweeps. For each sample size we compute the arithmetic average of the 600 bootstraps and the root mean square error (RMSE) of each bootstrap ERP vs. the arithmetic average of the 600 bootstrap ERPs. Thus, we obtain a mean and standard deviation RMSE across the 600 bootstraps for each subject and for each bootstrap sample size. The RMSE is defined as the square root of the mean squared difference, computed at all electrodes and all time points; since the RMSE is computed between a bootstrap and the average of the 600 bootstraps, the higher the RMSE the higher the variability of the estimator. The results for the AEA and ACSTP estimators are given for all subjects in Fig. 5. The RMSE is consistently lower for ACSTP as compared to the AEA both for the target (TA) and non-target (NT) class. As compared to AEA, the RMSE of ACSTP tends to increase slower when decreasing the sample size. For large sample sizes the RMSE of the two estimators is very similar. These results show that the ACSTP estimator features a significantly lower variability, the more so the smaller the sample size, approaching the AEA estimator asymptotically (as the sample size goes to infinity).

Figure 6 show butterfly ERP plots of the 600 bootstraps at electrodes AFz, Cz and Pz for different values of sample size, along with the mean and the 95% confidence interval of the 600 bootstraps for subject 8. The abscissa and ordinate of all plots are the same and with the same scaling. For the target class (top of the figure) we notice the presence of ocular artefacts peaking at about 500 ms post-stimulus, visible in the AEA estimator at frontal electrode AFz (see also Fig. 2). The ACSTP estimator reduces them substantially in all cases and even when only 10% of the sweeps are used to obtain the estimate (8 sweeps only).
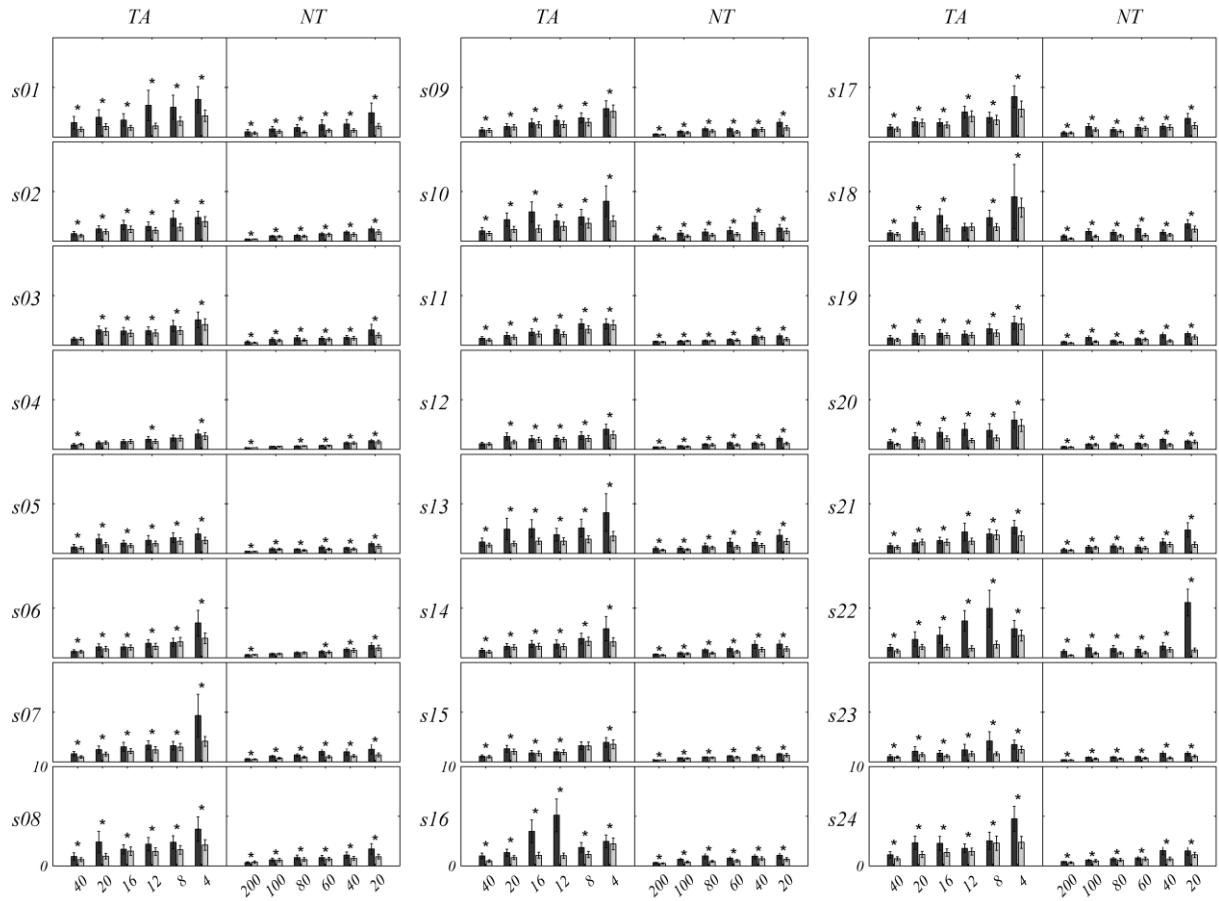
Figure 5. AEA and ACSTP Estimator Variability. Mean and 3 sd root mean square error (RMSE) of the 600 bootstraps as a function of the bootstrap sample size (x-axis), for target (TA) and non-target (NT) sweeps and each subject separately (*s01* to *s24*). Dark Grey: AEA (Arithmetic Ensemble Average). Light Gray: ACSTP (Adaptive Common Spatio-Temporal Pattern). A two-tailed paired student t-test was used to compare the RMSE of the AEA and of the ACSTP for each sample size. The p-values obtained were corrected using the false discovery rate procedure (with expected proportion of false discovery q=0.05: Benjamini and Hochberg, 1995). Significant p-values after correction are indicated by an asterisk.
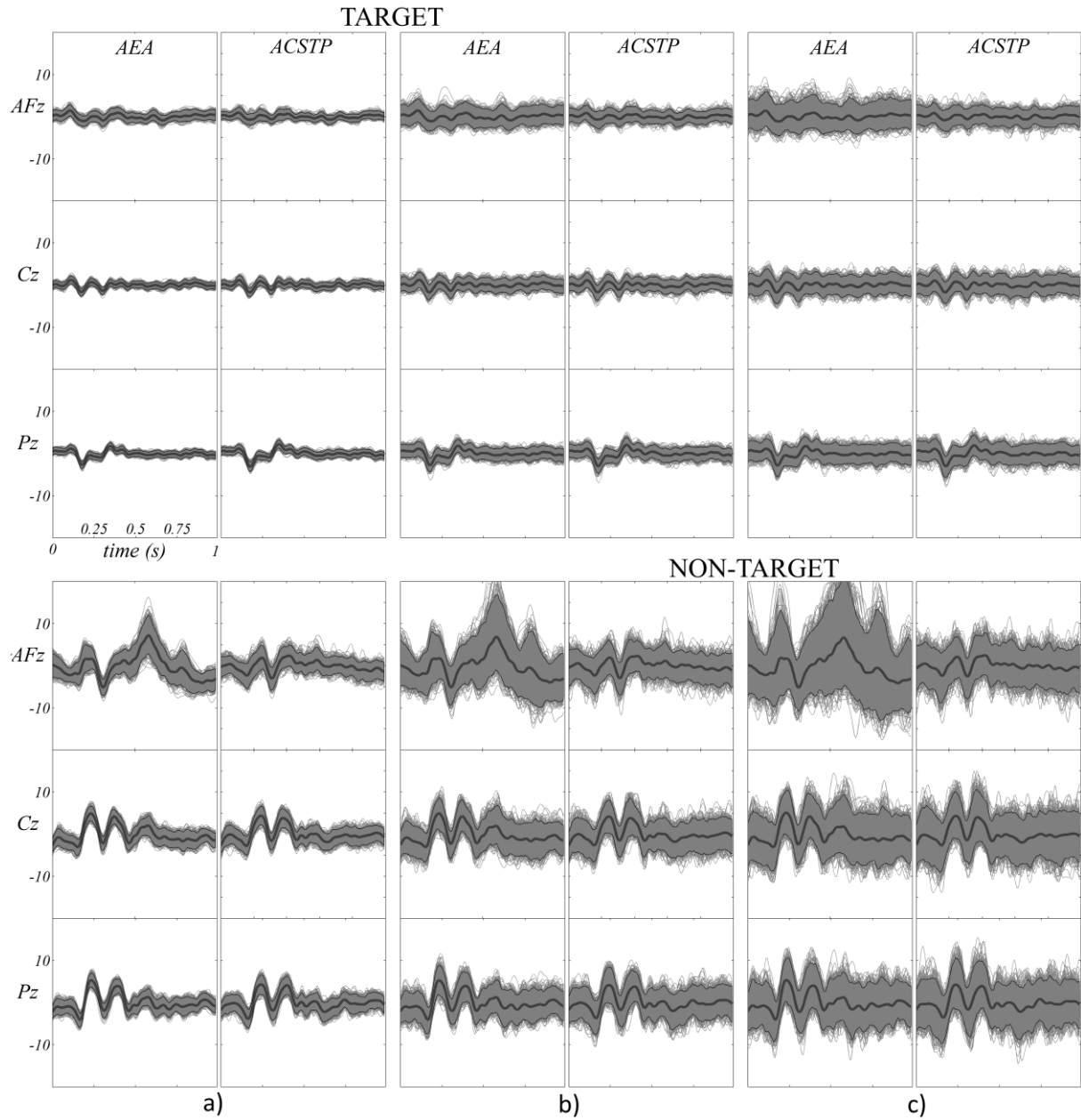
Figure 6. Butterfly plot of bootstrapped ERPs for Subject 8. The thin lines are plots of the 600 bootstrap ERPs estimated by AEA and ACSTP at electrodes AFz, Cz and Pz for three different values of sample size (a)= 50%, b)=20%, c)=10% of the available sweeps). The mean (thick dark grey line) and the 95% confidence interval (grey shaded area) of the 600 bootstraps are also shown. Upward deflections indicate positive potentials. In all plots the horizontal axis represents time, with zero corresponding to the time of flashing, and the vertical axis represents voltage in microVolts. The scaling is the same for all plots so as to allow comparisons. Top: target (TA) class. Bottom: non-target (NT) class.

1.4 Conclusions and Discussion

The goal of this article was to provide a reference method for the time-domain analysis of ERPs. The CSTP algorithm itself has a major free parameter, the subspace dimension (P_z). We have proposed one procedure for estimating this parameter automatically, involving the use of a mask. Together with adaptive time-shift and amplitude correction this yields the ACSTP algorithm. In the present study, five different mask definitions have been necessary to accommodate the principal spatial patterns of the P300 potentials in a 24-subject sample. According to our experience, a small number of masks, as found here, may accommodate the quasi-totality of the adult healthy population working with P300 data. While the proposed procedure for selecting the subspace dimension does not guarantee optimality, it provides a useful and workable guess, as our real-data analysis witness. For the sake of estimating an ensemble average one may be content with the output of the ACSTP algorithm. In addition, one can refine the search by running the CSTP algorithm for some values of P_z above and below the one chosen by the ACSTP algorithm. Yet, an inappropriate choice of the mask may jeopardize the ensemble average estimation and is a supervised step itself. A fully unsupervised method for selecting the optimal subspace dimension is possible when the estimated means are to be used for statistical analysis (e.g., contrasting the means of two experimental conditions). In this case one may treat the subspace dimension P_z as an independent variable, that is, estimate means in a wide range of P_z values such as [4, N-1] and perform statistical tests on the means for all values of P_z in the range. Using a p-min permutation test strategy (Westfall and Young, 1992) one may take the value of p yielding the strongest evidence against the null hypothesis, provided that the omnibus hypothesis is rejected.

The use of time-shift adaptive estimation has a free parameter, the width of the maximal time-shift allowed (E), which may affect the result considerably. The width of the maximal time-shift allowed should not be set automatically because it strictly depends on the amount of instrumental jitter, the nature of the experiment and the class of ERP under analysis. It can be fixed once though for any given equipment and experiment. Experimental results on the expected amount of physiological ERP latency are available (e.g., Gaspar et al., 2011). In addition to exploit previous knowledge, one should be aware of the necessity to remove any artifacts with a strong periodic character before running any time-shift adaptive estimation, since the algorithm may try to align the sweeps based on such oscillations instead of on true time-locked responses. Power-line contaminations constitute the most important kind of such artifacts. Similarly, ERP containing very low frequencies may result in distorted correlation between each sweep and the ensemble average, biasing the time-shift correction. Therefore, both an efficient notch filter (at the power-line frequency) and a band-pass filter with a high-pass in between 0.1 Hz and 1 Hz should be applied to the data if time-shift correction is sought. In addition, the use of a mask for time-shift estimation is strongly recommended.

Typically, ERP studies estimate the ensemble averages as the arithmetic average of a large number of sweeps after removing EEG epochs with excessive artifacts. This procedure, besides being time consuming, is problematic in the case of overlapping ERPs, as EEG segments holding all ERPs overlapping with the artifact, even if only partially, must be removed (see the discussion in Smith and Kutas, 2015). This difficulty is not encountered by artifact filtering methods based on eye-related activity regression and blind source separation, although these methods may filter and distort the signal in unpredictable ways. Thanks to the combination of a sharp spatio-temporal filtering and single-sweep weighting, the method we propose is able to process EEG data contaminated by common artifacts such as those related to eye blinks, eye movements, facial muscle contractions, slow head movements, loss of electrode contact, etc. In this article we have analyzed a complete data set of 24 subjects without any artifact removal or artifact filtering and we have obtained an enhancement of the SNR over the arithmetic ensemble average in all cases. The ability to perform analysis without any previous artifact rejection is a remarkable advantage of the proposed method over the arithmetic mean and other standard methods presented in the literature. Of course, very large artifacts should be removed from the stream of data, but this can be done without supervision with an amplitude thresholding method using a rather permissive threshold such as $\pm 200 \mu V$ (assuming a high-pass filter has been applied to remove the DC offset). Other possible extreme EEG data distortions such as loss of electrode contact can also be easily detected and removed automatically using sensitive methods such as the Riemannian Potato (Barachant et al., 2013) with a rather permissive threshold.

The design matrix T introduced in (1.3) and used in (1.9) for computing the ensemble average estimations may be a very large matrix, apparently posing problems of numerical efficiency for the ensemble average estimations in (1.9). Suppose an experiment involving the study of ten classes of overlapping ERP, lasting 30 minutes total, recording EEG with 128 electrodes at 256 samples per seconds and defining ERP windows of one second. The data matrix X would be of dimension 128x460800, requiring about 472Mb of memory for encoding the data with a double precision. On the other hand, the design matrix T would be of size 2560x460800, requiring for the same precision more than 9.4Gb of memory. Indeed with long experiments, long ERP windows or many ERP classes the efficiency of the computations in (1.9) may become a matter of concern. Fortunately however, one does not need to store in memory matrix T at all. In fact, both products XT^T and TT^T in (1.9) can be computed very efficiently just knowing the vectors of weights and time-shifts, data matrix X and the event times for each class. Regarding the inversion of the matrix TT^T in (1.9), which in our example would be a matrix of dimension 2560x2560, it so happens that this matrix is sparse (i.e., it has many zero entries), so efficient sparse matrix inversion routines can be employed; since the last term in (1.9) is just a scaling, overall the ensemble average estimations in (1.9) can be obtained very efficiently on standard commercial computers.

A key factor for the success of CSTP estimations is the goodness of the noise estimation. If only a few sweeps are available, the estimation of the spatial and temporal noise covariance matrices (1.12) is poor. In this case one can use also random segments of the EEG data (not only actual sweeps) to construct these matrices. This is a simple and effective strategy, allowing proper noise estimation as soon as about 40 seconds of EEG recording is available. Another possible strategy is to regularize the estimation of these matrices by known methods based on statistical estimation theory (Engemann and Gramfort, 2015) or transfer learning (Lotte and Guan, 2011).

It should be stressed here that the CSTP is not a source separation approach; the filter is not designed to estimate the waveform and spatial distribution on the scalp of the actual source processes generating the ERPs, rather, it only seeks spatio-temporal combinations of the data maximizing the SNR. As a consequence, temporal and spatial patterns as those presented in Fig. 4 should not be interpreted as representing physiological phenomena. The same is true in general for all methods based on PCA or SNR enhancement. The only family of methods that allows such interpretation is waveform-preserving source separation (Congedo et al., 2014; Makeig et al., 2002).

In summary, we have presented a reference companion method for ERP analysis in the time domain based on a spatio-temporal multivariate filtering of the ensemble average estimation. Besides performing an effective spatio-temporal filtering, this method adaptively estimates weights and time-shifts for each sweep. Furthermore, the method considers appropriately the case of overlapping ERPs. It is in the combination of all these features that the generality of the ACSTP algorithm resides. We have proposed a solution to the problem of estimating the optimal subspace dimension for the filter. The method can be applied on data where only segments containing very extreme artifacts are discarded; common source of artifacts can be analyzed. Given these characteristics, the method proposed here can be used routinely to obtain an improved ensemble average estimation to complement and enrich the information provided by the standard arithmetic average and to obtain an ensemble average estimation with a reduced number of sweeps. It can also be used as a pre-processing step before entering statistical analysis, for example by the LIMO EEG Toolbox (Pernet et al., 2011). The ACSTP algorithm has been described in all relevant details, allowing peers to replicate exactly its implementation. Furthermore, a MATLAB toolbox (available at : http://louis-korczowski.org/) a stand-alone executable application (available at https://sites.google.com/site/marcocongedo) and a plug-in for the EEGLAB suite (Delorme and Makeig, 2004) have been made available to the community.

Acknowledgements

Appendix A

Referring to (1.9), note that if the ERPs do not overlap, $\left(TT^T\right)^{-1}$ reduces to a block diagonal matrix equal to

$$
\begin{pmatrix} \frac{1}{\sum_{k=1}^{K_1}\sigma_{1k}^2}I_{\mathrm{T}} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{\sum_{k=1}^{K_Z}\sigma_{Zk}^2}I_{\mathrm{T}} \end{pmatrix} \quad \text{and} \quad XT^T = \left[\sum_{k=1}^{K_1}\sigma_{1k}X_{1k}\left(\tau_{1k}\right)\cdots\sum_{k=1}^{K_Z}\sigma_{Zk}X_{Zk}\left(\tau_{Zk}\right) \right] \quad \text{holds the}
$$

weighted sum of the sweeps for each class stacked horizontally one next to the other. In this case each average estimation in (1.9) reduces to the weighted arithmetic mean (1.8), for whatever set of non-negative weights. Furthermore, if all weights are equal to one and the time shifts are equal to zero, (1.9) reduces to the arithmetic ensemble average. This way we see precisely why estimation (1.9) is a generalization of (1.8).

Appendix B

Let

$$
C_{(S)} = U\Phi U^T, \quad C_{(T)} = V\Psi V^T \tag{1.17}
$$

be the eigenvalue-eigenvector decompositions of the noise covariance estimations in (1.12), with $U \in \mathfrak{R}^{N \cdot N}$, $V \in \mathfrak{R}^{T \cdot T}$ orthogonal matrices holding in columns the eigenvectors $\left[u_1 \cdots u_N\right]$, $\left[v_1 \cdots v_T\right]$ and $\Phi \in \mathfrak{R}^{N \cdot N}$, $\Psi \in \mathfrak{R}^{T \cdot T}$ diagonal matrices holding the (non-negative) eigenvalues in decreasing order of magnitude, i.e., such that $\phi_1 \geq \cdots \geq \phi_N$ and $\psi_1 \geq \cdots \geq \psi_T$. Let

$$
F_{(S)} = \left(\phi_1^{-\frac{1}{2}}u_1 \cdots \phi_Q^{-\frac{1}{2}}u_Q\right) \in \mathfrak{R}^{N \cdot Q}, \quad F_{(T)} = \left(\psi_1^{-\frac{1}{2}}v_1 \cdots \psi_R^{-\frac{1}{2}}v_R\right) \in \mathfrak{R}^{T \cdot R} \tag{1.18}
$$

which transpose have right-inverses

$$
G_{(S)} = \left(u_1\phi_1^{\frac{1}{2}} \cdots u_Q\phi_Q^{\frac{1}{2}}\right) \in \mathfrak{R}^{N \cdot Q}, \quad G_{(T)} = \left(v_1\psi_1^{\frac{1}{2}} \cdots v_R\psi_R^{\frac{1}{2}}\right) \in \mathfrak{R}^{T \cdot R}, \tag{1.19}
$$

where right-inverse means that these matrices verify

$$
F_{(S)}^T G_{(S)} = I_Q, \quad F_{(T)}^T G_{(T)} = I_R. \tag{1.20}
$$

In (1.18)-(1.20) Q≤N and R≤T are the largest integer indices for which ϕ_Q and ψ_R are non-null, respectively. In practice (with empirical data), we choose Q and R so that the eigenvalues ϕ_Q and ψ_R are larger than a threshold explaining a *very* small portion of the variance of the noise (say, one million times smaller than the first eigenvalue of the noise term $C_{(S)}$). In this way the final result will not be sensitive at all to the exact choice of Q and R and these parameters can be set automatically without affecting the result. Note that we can set Q=N and R=T as long as matrices $C_{(S)}$ and $C_{(T)}$ in (1.12), respectively, are positive definite, which for $C_{(T)}$ will be the case in general if several times more than T/N sweeps are averaged (remember that without loss of generality we are assuming that T>N). Finally,

note that the threshold must be set very low in any case because ERP components may explain very little variance as compared to the variance of large EEG artifacts, thus *we must make sure we do not remove ERP components in the whitening step*.

We can now obtain a bilinear (spatio-temporal) whitening of the data with the pair of matrices $F_{(S)}$ and $F_{(T)}$; let us apply the whitening transformation to the ensemble average estimation for class z, such as

$$\bar{Z}_z^{'} = F_{(S)}^T \bar{X}_z^{'} F_{(T)} \in \mathfrak{R}^{Q \cdot R}, \tag{1.21}$$

where we use (1.8) or (1.9) to estimate $\bar{X}_z^{'}$ depending whether we are in the non-overlapping or overlapping case, respectively. The transformation is a *spatio-temporal whitening* of the noise term since it verifies $F_{(S)}^T C_{(S)} F_{(S)} = I_Q$ and $F_{(T)}^T C_{(T)} F_{(T)} = I_R$. Now, let

$$\bar{Z}_z^{'} = \Pi_z W_z \Xi_z^T, \tag{1.22}$$

be the SVD of the whitened ERP ensemble average estimation (1.21), with Π and Ξ holding in columns the left and right singular vectors, respectively, and W holding in the diagonal the singular values in decreasing order, as usual. Finally, the bilinear common filters and patterns are given by

$$B_z = F_{(S)} \tilde{\Pi}_z, \qquad D_z = F_{(T)} \tilde{\Xi}_z, \qquad A_z = G_{(S)} \tilde{\Pi}_z, \qquad E_z = G_{(T)} \tilde{\Xi}_z, \tag{1.23}$$

where $\tilde{\Pi}_z = \left[\pi_{z1} \cdots \pi_{zP_z} \right]$ and $\tilde{\Xi} = \left[\xi_{z1} \cdots \xi_{zP_z} \right]$ hold in columns the first $P_z \leq N$ column vectors of Π_z and Ξ_z (1.22), respectively, $F_{(S)}^T, F_{(T)}^T$ are given in (1.18) and $G_{(S)}, G_{(T)}$ are given in (1.19). What we have achieved is that, by construction, B_z and D_z verify

$$\begin{cases} B_z^T C_{(S)} B_z = I_{P_z} \\ D_z^T C_{(T)} D_z = I_{P_z} , \\ B_z^T \bar{X}_z^{'} D_z = \Lambda_z \end{cases} \tag{1.24}$$

where $\Lambda_z \in \mathfrak{R}^{P_z \cdot P_z}$ is a diagonal matrix with diagonal elements sorted by magnitude. These elements hold the square root of the SNR variance ratio (1.14) attained by the corresponding vectors of B_z and D_z. Note that they are just the first P_z elements of diagonal matrix W_z in (1.22), whereas the remaining elements of W_z explain the square root of the variance filtered out by the bilinear transformation. Thus, effective SNR enhancement is obtained simply retaining $P_z < N$ vectors to construct the filter (1.23). Note that by setting $F_{(T)} = G_{(T)} = D_z = E_z = I_T$ we obtain the *common spatial pattern* (CSP), while by setting $F_{(S)} = G_{(S)} = B_z = A_z = I_N$ we obtain the *common temporal pattern* (CTP). Also, by setting $F_{(S)}^T = G_{(S)} = I_N$ and $F_{(T)}^T = G_{(T)} = I_T$ (i.e., omitting the whitening step) the CSTP reduces to the bilinear PCA; in this way the CSTP can be seen as a bilinear PCA applied to whitened data.

References

Barachant, A., Andreev, A., Congedo, M., 2013. The Riemannian Potato: an automatic and adaptive artifact detection method for online experiments using Riemannian geometry. TOBI Workshop lV, Sion, Switzerland.

Başar-Eroglu, C., Başar, E., Demiralp, T., Schürmann, M., 1992. P300-response: possible psychophysiological correlates in delta and theta frequency channels. A review. Int. J. Psychophysiol. 13(2), 161-79.

Benjamini, Y., Hochberg, Y., 1995. Controlling the False discovery Rate: A Practical and Powerful Approach to Multiple Testing, J. R. Stat. Soc. Series B, 57 (1), 289-300.

Cabasson, A., Meste, O., 2008. Time Delay Estimation: A New Insight Into the Woody's Method. IEEE Signal Process. Lett. 15, 573-576.

Chapman, R.M., McCrary, J.W., 1995. EP Component Identification and Measurement by Principal Component Analysis. Brain and Cognition 27, 288-310.

Cichocki, A., Amari, S.I., 2002. Adaptive Blind Signal and Image Processing. Learning Algorithms and Applications. John Wiley and Sons, London, 586 pp.

Congedo, M., Goyat, M., Tarrin, N., Ionescu, G., Rivet, B., Varnet, L., et al., 2011. "Brain Invaders": a prototype of an open-source P300- based video game working with the OpenViBE, Proc. IBCI Conf., Graz, Austria, 280-283.

Congedo, M., Gouy-Pailler, C., Jutten C., 2008. On the Blind Source Separation of Human Electroencephalogram by Approximate Joint Diagonalization of Second Order Statistics, Clin. Neurophysiol. 119, 2677–2686.

Congedo, M., Rousseau, S., Jutten, C. 2014. An Introduction to EEG Source Analysis with an illustration of a study on Error-Related Potentials. In "Guide to Brain-Computer Music Interfacing"(Chapter 8), E. Miranda and J. Castet (Eds), Srpinger, London, 313 p.

Deecke, L., Scheid, P., Kornhuber, H.H., 1969. Distribution of readiness potential, pre-motion positivity and motion potential of the human cerebral cortex preceding voluntary finger movements. Exp. Brain Res. 7, 158-168.

de Munck, J.C., Bijma, F. 2010. How are evoked responses generated? The need for a unified mathematical framework. Clin. Neurophysiol. 121, 127-129.

Delorme, A., Makeig, S., 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J. Neurosci. Methods. 134(1), 9-21.

Dien, J., 2010. Evaluating two-step PCA of ERP data with Geomin, Infomax, Oblimin, Promax, and Varimax rotations. Psychophysiol. 47, 170-183.

Donchin, E., 1966. A multivariate approach to the analysis of average evoked potentials. IEEE Trans. Biomed. Eng. 3, 131-139.

Dornhege, G., Blankertz, B., Krauledat, M., Losch, F., Curio, G., Müller K.R., 2006. Combined optimization of spatial and temporal filters for improving brain-computer interfacing. IEEE Trans. Biomed. Eng. 53(11), 2274-81.

Engemann, D.A., Gramfort, A., 2015. Automated model selection in covariance estimation and spatial whitening of MEG and EEG signals. Neuroimage, 108, 328-342.

Gaspar, C.M., Rousselet, G.A., Pernet, C.R., 2011. Reliability of ERP and single-trial analyses. Neuroimage, 58, 620-629.

Gasser, T., Möcks, J., Verleger, R., 2003. SELAVCO: a method to deal with trial-to-trial variability of evoked potentials. Electroencephalogr Clin Neurophysiol. 55(6), 717-23.

Jansen, B.H., Agarwal. G., Hedge. A., Boutros, N.N., 2003. Phase synchronization of the ongoing EEG and auditory EP generation. Clin. Neurophysiol. 114, 79-85.

Jaśkowski, P., Verleger, R., 1999. Amplitudes and latencies of single-trial ERP's estimated by a maximum-likelihood method. IEEE Trans. Biomed. Eng. 46(8), 987-93.

Jasper, H.H., 1958. The ten–twenty electrode system of the International Federation. Electroencephalogr. Clin. Neurophysiol. 10, 367–380.

John, E. R., Ruchkin, D. S., Vilegas, J., 1964. Experimental background: signal analysis and behavioral correlates of evoked potential configurations in cats. Ann. N.Y. Acad. Sci., 112, 362-420.

Koles, Z.J., 1991. The Quantitative extraction and Topographic Mapping of the Abnormal Components in the Clinical EEG. Electroencephalogr. Clin. Neurophysiol. 79, 440-447.

Koles, Z.J., Soong, A., 1998. EEG Source Localization: Implementing the Spatio-Temporal Decomposition Approach. Electroencephalogr. Clin. Neurophysiol., 107, 343-352.

Kotchoubey B., 2015. Event-Related Potentials in Disorders of Consciousness. in Rossetti, A.O., Laureys, S. (Eds.), Clinical Neurophysiology in Disorders of Consciousness, Springer, London, 161 p.

Lagerlund, T.D., Sharbrough, F.W., Busacker, N.E., 1997. Spatial filtering of multichannel electroencephalographic recordings through principal component analysis by singular value decomposition. J. Clin. Neurophysiol. 14(1), 73-82.

Lehmann, D., Skrandies, W., 1980. Reference-free identification of components of checkerboard-evoked multichannel potential fields. Electroencephalogr. Clin. Neurophysiol. 48(6), 609–621.

Lemm, S., Blankertz, B., Curio, G., Müller, K.R., 2005. Spatio-spectral filters for improving the classification of single trial EEG. IEEE Trans. Biomed. Eng. 52(9), 1541-8.

Lęski, J.M., 2002. Robust Weighted Averaging, IEEE Trans. Biomed. Eng. 49(8), 796-804.

Lopes da Silva, F.H., 2006. Event-related neural activities: what about phase? Prog. Brain Res. 159, 3-17.

Lopes da Silva, F.H., Pijn, J.P., Velis, D., Nijssen, P.C.G. 1997. Alpha rhythms: noise, dynamics, and models. Int. J. Psychophysiol. 26, 237-249.

Lopes Da Silva, F.H., 2010. Event-related Potentials: General Aspects of Methodology and Quantification. In "Niedermeyer's Electroencephalography: Basic Principles, Clinical Applications, and Related Fields". D. Schomer and F. Lopes da Silva (Eds.), 6th Edition, Lippincott Williams & Wilkins, london, 1275 pp.

Lotte, F., Guan, C.T., 2011. Regularizing Common Spatial Patterns to Improve BCI Designs: Unified Theory and New Algorithms, IEEE Trans. Biomed. Eng. 58(2), 355-362.

Luck, S., 2014. An Introduction to the Event-Related Potential Technique. 2nd Ed. MIT Press, Boston, 416 pp.

Makeig S., Westerfield M., Jung T.P., Enghoff S., Townsend J., Courchesne E., Sejnowski T.J., 2002. Dynamic brain sources of visual evoked responses. Science, 295(5555), 690-4.

Makeig, S., Debener, S., Onton, J., Delorme A., 2004. Mining event-related brain dynamics. Trends Cogn. Sci., 8(5), 204–210

Mazaheri A., Jensen O. (2010) Rhythmic pulsing: linking ongoing brain activity with evoked responses. Front. Hum. Neurosci. 4, 177.

McClelland, R.J., Sayers, B.M. 1983. Towards fully objective evoked responses audiometry. British J. Audiology, 17, 263-270.

Näätänen, R., Gaillard, A.W.K., Mäntysalo, S., 1978. Early selective-attention effect on evoked potential reinterpreted. Acta Psychologica, 42, 313-329.

Nikulin, V.V., Linkenkaer-Hansen, K., Nolte, G., Curio, G. 2010. Non-zero mean and asymmetry of neuronal oscillations have different implications for evoked responses. Clin. Neurophysiol. 121, 186-193.

Pernet C.R., Chauveau, N. Gaspar, C., Rousselet, G.A. (2011) LIMO EEG : A Toolbox for Hierarchical LInear MOdeling of ElectroEncephalographic Data, Comput. Intell. Neurosci., 831409.

Picton, T.W., Bentin, S., Berg, P., Donchin, E., Hillyard, S.A., Johnson, R. Jr., Miller, G.A., Ritter, W., Ruchkin, D.S., Rugg, M.D., Taylor, M.J., 2000. Guidelines for using human event-related potentials to study cognition: recording standards and publication criteria. Psychophysiol. 37(2), 127-52.

Ramoser, H., Muller-Gerking, J., Pfurtscheller, G., 2000. Optimal Spatial Filtering of single trial EEG during Imagined Hand Movement. IEEE Trans Rehabil Eng 8(4), 441-446.

Regan, D., 1989. Human Brain Electrophysiology: Evoked Potentials and Evoked Magnetic Fields in Science and Medicine. New York, Elsevier, 672 pp.

Rivet, B., Souloumiac, A., Attina, V., Gibert, G., 2009. xDAWN algorithm to enhance evoked potentials: application to brain-computer interface. IEEE Trans. Biomed. Eng. 56(8), 2035-43.

Ruchkin, D.S., 1965. An Analysis of Average response Computatios based Upon Aperiodic Stimuli. IEEE Trans. Biomed. Eng., 12(2), 87-94.

Sayers, B.M., Beagley, H.A., Henshall, W.R., 1974. The mechansim of auditory evoked EEG responses. Nature. 247(5441), 481-3.

Sayers, B.M., Beagley, H.A., 1974. Objective evaluation of auditory evoked EEG responses.
Nature. 251(5476), 608-9.

Sereno, S.C., Rayner, K., 2003. Measuring word recognition in reading: eye-movements and event-related potentials. Trends Cogn. Sci., 7(11), 489-493.

Smith, N.J., Kutas, M., 2015. Regression-based estimation of ERP waveforms: II. Nonlinear effects, overlap correction, and practical considerations. Psychophysiol. 52(2), 169-81.

Souloumiac, A., Rivet B., 2013. Improved estimation of EEG evoked potentials by jitter compensation and enhancing spatial filters. Proc. ICASSP, Vancouver, Canada, 1222-1226.

Sutton, S., Braren, M., Zubin, J., John, E.R., 1965. Evoked-potential correlates of stimulus uncertainty. Science, 150(3700), 1187-8.

Tallon-Baudry C., Bertrand O., Delpuech C., Pernier J. (1996) Stimulus Specificity of Phase-Locked and Non-Phase-Locked 40 Hz Visual Responses in Human. Journal of Neuroscience, 16, 4240-4249.

Townsend, G., Graimann, B., Pfurtscheller, G., 2006. A comparison of common spatial patterns with complex band power features in a four-class BCI experiment. IEEE Trans. Biomed. Eng. 53(4), 642-51.

Walter, W.G., Cooper, R., Aldridge, V.J., McCallum, W.C., Winter, A.L., 1964. Contingent Negative Variation: an electric sign of sensorimotor association and expectancy in the human brain. Nature 203, 380-4.

Wang, H., Zheng W., 2008. Local temporal common spatial patterns for robust single-trial EEG classification. IEEE Trans. Neural Syst. Rehabil. Eng. 16(2), 131-9.

Woldorff, M., 1988. Adjacet Response overlap during the ERP averaging process and a technique (Adjar) for its estimation and removal. Psychophysiol. 25, 490.

Woldorff, M., 1993. Distortion of ERP averages due to overlap from temporally adjacent ERPs: analysis and correction. Psychophysiol. 30(1), 98-119.

Wolpaw, J., Wolpaw, E.W., (Eds) 2012. Brain-Computer Interfaces. Principles and Practice, Oxford Press, London, 424 pp.

Woody, C.D., 1967. Characterization of an adaptive filter for the analysis of variable latency neuroelectric signals. Med. Biol. Eng. 5(6), 539-554.

Yu, K., Shen, K., Shao, S., Ng, W.C., Li, X., 2012. Bilinear common spatial pattern for single-trial ERP-based rapid serial visual presentation triage. J. neural Eng. 9(4), 046013.

Westfall P.H., Young, S.S., 1992. Resampling-based multiple testing: examples and methods for P-value adjustment., John Wiley & sons, New York.