

A periodic spatio-spectral filter for event-related potentials

Foad Ghaderi*

*Human Machine Interaction Lab.
Faculty of Electrical and Computer Engineering, Tarbiat Modares University, Tehran, Iran
E-mail: fghaderi@modares.ac.ir*

Su Kyoung Kim

Robotics Innovation Center (RIC), German Research Center for Artificial Intelligence (DFKI), Bremen, Germany

Elsa Andrea Kirchner

*Robotics Innovation Center (RIC), German Research Center for Artificial Intelligence (DFKI), Bremen, Germany & Robotics Group,
University of Bremen, Bremen, Germany*

Abstract

With respect to single trial detection of event-related potentials (ERPs), spatial and spectral filters are two of the most commonly used pre-processing techniques for signal enhancement. Spatial filters reduce the dimensionality of the data while suppressing the noise contribution and spectral filters attenuate frequency components that most likely belong to noise subspace. However, the frequency spectrum of ERPs overlap with that of the ongoing electroencephalogram (EEG) and different types of artifacts. Therefore, proper selection of the spectral filter cutoffs is not a trivial task. In this research work, we developed a supervised method to estimate the spatial and finite impulse response (FIR) spectral filters, simultaneously. We evaluated the performance of the method on offline single trial classification of ERPs in datasets recorded during an oddball paradigm. The proposed spatio-spectral filter improved the overall single-trial classification performance by almost 9% on average compared with the case that no spatial filters were used. We also analyzed the effects of different spectral filter lengths and the number of retained channels after spatial filtering.

Keywords: Electroencephalogram, event-related potential (ERP) detection, brain-computer interface (BCI), spatial filters, spatio-spectral filters.

Introduction

Researchers and clinicians use patterns extracted from brain in order to identify the cognitive states of participants for different purposes, e.g., for developing brain-computer interface (BCI) systems. Considering the high temporal resolution, low costs, and ease of use (compared with other acquisition techniques), EEG data is one of the most commonly used sources of information about brain activity.

Because of high levels of noise and artifacts, extracting brain patterns from EEG signals is usually a challenging task [1, 2]. Although the traditional method for eliminating noise in ERP signals is to average the ERP windows of experimental repetitions [1], this approach, called averaging, is not applicable in systems that rely on single trial detection of ERPs [3]. By definition, the single trial ERP classification problem has concerned detecting specific
10 brain patterns within the recorded EEG, whereby either an internal event or an external stimulus elicited those patterns [4, 5]. Researchers and engineers for example use ERP classifiers to identify whether the participants are mentally involved in a particular task or not. To this end, the system has to be able to discriminate between the two cases of existence or absence of the patterns of interest in the data. System designers use the information about the existence or absence of the these patterns, for example, in BCI systems to issue appropriate control commands or
15 to communicate with the external world [6, 7, 8].

There has been a long debate about the neural origins of the ERPs. On one hand, in evoked model for ERPs, it is believed that ERPs are generated by additive evoked responses which are independent of the ongoing EEG [9]. On the other hand, in phase reset model, researchers show evidence that ERPs are generated from changes in dynamics of ongoing EEG, i.e., early ERP components are the superposition of ongoing EEG oscillations that reset
20 phases in response to stimuli [10]. Despite the high number of research work on the topic, neither of the approaches could be proved or disproved [11] and results from deep brain recording suggest that indeed both processes may contribute at least to a subgroup of ERPs [12].

Different pre-processing methods have been proposed so far to increase the signal-to-noise ratio of EEG signals [1, 13]. Spatial and spectral filters are two important approaches, among others, which are widely used by
25 engineers and researchers for this purpose. Exploiting the redundancy in the data, spatial filters transfer the multi-channel measurements to low-dimensional subspaces while increasing the signal-to-noise ratio. Considering the requirements of different applications, specific conditions are optimized by spatial filters. For example, the common spatial patterns (CSP) algorithm maximizes the temporal variance of one class while minimizing that of the other class [14]. This method has been successfully used to improve classification performance in motor imagery
30 tasks [15, 16].

Considering that event-related potentials (ERPs) are time-locked to some stimuli, specific spatial filters are designed for signal enhancement. The proposed method in [17] maximizes the sum of the squared distances of the components of the mean feature vectors. The xDAWN method [18] constructs an algebraic model of the recorded signals that is composed of three parts: 1) the responses on targets, 2) the response common to all
35 stimuli (including targets and non-targets), and 3) the residual noise. Then, a defined relation for the signal-to-signal-plus-noise ratio (SSNR) is maximized through the Rayleigh quotient after applying QR and singular value decomposition. Comparative studies reported in [18, 19] confirm the outstanding performance of the method. Tensor decomposition improves the features and reduces the dimensionality of the data, e.g., [20]. A spatio-

temporal linearly-constrained minimum variance (LCMV) beamformer was used in [21] for ERP detection.

40 Abandoning the debate about the neural origins of the ERPs, if one ignores the noise in EEG data and concatenates the ERP instances of one class, the resulting signal is a periodic multi-channel signal. However, noise and artifacts contaminate the EEG data and hence the periodic structure deteriorates. The periodic spatial filter (π SF) method, presented in [22, 23], is based on enhancing the periodic structure of the data created by concatenating the ERP windows.

45 In general, spatial filters can be seen as linear transformations that transform raw multi-channel recordings into the task relevant neural sources. Beamformer techniques estimate sources that have specific waveforms, or originate from specific points on the cortex [21]. Contrastingly, with the PCA, ICA, xDAWN, π SF, and π SSF methods, multivariate linear methods in which, given a set of measurements, the approach has attempted to find inverse transforms (spatial filters) estimating the original sources. These methods are based on QR and SVD algorithms, 50 which decompose the data into a set of eigenvalues and eigenvectors¹. For these methods no assumption about the location or waveform template of the sources is required. It is shown in [24] that only if the measurements are uncorrelated the spatial filters are equal to the mixing matrix, and hence the sources can be estimated. However, the data captured from the sensors are usually correlated, and hence one has to be careful to avoid interchangeably using spatial filters and activation patterns. The latter reflect the contributing neural processes [24].

55 On the other hand, researchers and clinicians usually assume that the most discriminating features of the signals of interest are in some specific spectral bands [6]. Therefore, they use prior information about properties of signals to filter out the irrelevant parts of the recordings, i.e., different types of noise and artifacts [25]. However, there is no specific guideline in the literature for selecting the appropriate filtering pass band for ERP analysis. Considering that P300, an ERP component elicited in the process of target recognition, is mostly associated with activities in the delta band, Jansen *et al.* used the data in the 0-4 Hz band for classifying the ERPs [26]. However, P300 is present in 60 other spectral bands as well [27, 28]. Researchers reported different pass bands in P300 classification applications, e.g., 0.1-30 Hz [29], 0.1-20 Hz [30], 0.1-25 Hz [31], 1-20 Hz [18], 0.1-15 Hz [32], 0.1-12 Hz [28], 0.5-12 Hz [33, 34], 0.1-7 Hz [22], 0.1-4 Hz [26], 1-40 Hz [35], 1-10.66 Hz [3] and 0.5-15 Hz [21]. Farquhar and Hill showed in [5] that in a special design, in which no spatial filter is used, different spectral filtering cutoff frequencies would 65 not affect the performance of a regularized linear discriminant classifier if the EEG data is spatially whitened in advance.

It has been shown that it is possible to jointly optimize the specification of the spectral and the spatial filters. Common spatio-spectral pattern (CSSP) [36], filter bank common spatial pattern (FBCSP) [37], and sub-band common spatial pattern (SBCSP) [38] are some spectral extensions of the CSP method. Moreover, Zhang *et al.*

¹An eigenvector of a linear transformation is a non-zero vector that its direction will not change by applying the transformation. Eigenvalues are a special set of scalars associated with the eigenvectors.

70 used mutual information to optimize the parameters of a sequence of filter banks and spatial filters to generate log power features in motor imagery tasks [39]. In two separate works, Suk and Lee [40] and Mahanta *et al.* [41] developed spatio-spectral extensions of the CSP method using probabilistic approaches. In [42] channel-specific finite impulse response (FIR)² filters and the common spatial patterns are jointly estimated. Jointly optimized spatial and temporal filters have been proposed in e.g., [43] [44].

75 For the sake of improving the performance of π SF, we extended the method by designing a spectral FIR filter for every spatial filter. Here, we present the basic and the extended variants of the periodic spatial filters, called the periodic spatio-spectral filter (π SSF). Both approaches and their corresponding algorithms are deeply discussed to show the underlying concepts and their benefits compared with those of the xDAWN method, which provides outstanding performance [18, 19]. For evaluation of the presented spatial filters, we address three research
80 questions: Firstly, is there an effect of the proposed approaches on the classification performance? Based on this research question, it is investigated whether there are any differences in the classification performance depending on the approach type (no spatial filter, xDAWN, π SF, π SSF). Secondly, does the number of retained channels of the spatial filters have an impact on the classification performance? Finally, does the FIR filter length have an effect on the classification performance? In order to give more insight into the actual effects of the spatial filters,
85 we also visualized some examples of the data produced at different parts of the system.

The paper is structured as follows. In the next section, the details of the proposed methods are presented. In Methods section, the data, experiments and results are addressed. Discussions and conclusions are presented in the last two sections.

Problem formulation

Assume that EEG data, $\mathbf{s}(t)$, is recorded using n electrodes at m sample points. It is known that there are instances of two different ERP classes in the data and the objective is to use spatial and spectral filters to increase the discriminability of the two classes. The solution to this problem, which is composed of two linear transformations, can be formulated as follows:

$$\mathfrak{z}_i(t) = \mathbf{w}_i^T \mathbf{x}_i(t) = \mathbf{w}_i^T \left(\sum_0^L h_{i_l} \mathbf{s}(t-l) \right) \quad (1)$$

90 where \mathbf{w}_i is the i th, $i \leq n_f$, n dimensional spatial filter vector, $\mathbf{h}_i = [h_{i_0}, h_{i_1}, \dots, h_{i_L}]^T$ is the vector of FIR filter coefficients associated with \mathbf{w}_i , $\mathfrak{z}_i(t)$ is the i th output channel of the spatio-spectral filter, n_f is the overall number of retained channels after spatial filtering the data, L is the length of the FIR filter vector, and $[\cdot]^T$ denotes the transpose operator, respectively. Based on this model, for each spatial filter vector, \mathbf{w}_i , a separate FIR filter, \mathbf{h}_i , is estimated.

²A filter whose response to any finite length input is of finite duration.

95 The problem is addressed by estimating \mathbf{w}_i and \mathbf{h}_i in two different scenarios. First, in the π SF method, a fixed FIR filter is assumed to be used for all spatial filter vectors and the objective is to find the appropriate spatial filters. Next, in the π SSF method, we alternate between estimating the two filters by fixing one and optimizing the other. The details of the methods are presented in what follows. For the sake of simplicity, we omit the index i in the rest of the paper.

100 *Periodic spatial filter*

EEG data include brain-related electrical activity, non-brain physiological processes, and non-physiological noise. ERPs are a part of brain-related electrical activity, which are induced time-locked to internal events or external stimuli. Based on evoked model, the ERPs are additive evoked responses with a determined waveform in each channel. Phase reset model claims that EEG oscillations reset their phases in occurrence of stimuli. Regardless
 105 of the neural origins of the ERPs, ignoring the noise in the recorded EEG would result in a multi-channel data in which the ERP instances of the same class have similar waveforms. Therefore, by concatenating the instances of each class³, new multi-channel signals will be generated that are periodic. However, it is not possible to ignore the noise in EEG data, thus concatenated ERP windows are the superpositions of multi-channel periodic signals and noise. Discriminating the ERP instances can be achieved by enhancing the periodic structure of the two con-
 110 catenated multi-channel signals. For the sake of generality, in sequel we use class labels $c \in \{1, 2\}$ for the ERP classes.

Considering (1), assume that a pre-defined FIR filter is used and $\mathbf{x}(t) = \sum_0^L h_i \mathbf{s}(t-l)$ is spectrally filtered data. If m and m_e are the lengths of the whole data and each ERP window instance, respectively, the recorded data can be modeled as follows:

$$\mathbf{x}(t) = \mathbf{E}_1 \mathbf{d}_1(t) + \mathbf{E}_2 \mathbf{d}_2(t) + \mathbf{r}(t) \quad (2)$$

where $\mathbf{x}(t)$ is the n dimensional measurement vector, \mathbf{E}_1 and \mathbf{E}_2 are the $n \times m_e$ ERP response for the two classes, $\mathbf{r}(t)$ is the n dimensional noise vector (including ongoing EEG), and $\mathbf{d}_1(t) = [d_{11}(t), d_{12}(t), \dots, d_{1m_e}(t)]^T$ and $\mathbf{d}_2(t) = [d_{21}(t), d_{22}(t), \dots, d_{2m_e}(t)]^T$ are m_e dimensional vectors indicating the times of occurrence of the ERPs of each class. For $c \in \{1, 2\}$, if t_k is the time index of the onset of the k th stimulus in class c and $t_k \leq t < t_k + m_e$ then

$$d_{cl}(t) = \begin{cases} 1 & \text{if } l = t - t_k + 1 \\ 0 & \text{else} \end{cases} . \quad (3)$$

³Each ERP instance is a window of data cut in the neighborhood of the corresponding stimulus onset, e.g. between 100 ms before and 900 ms after the onset.

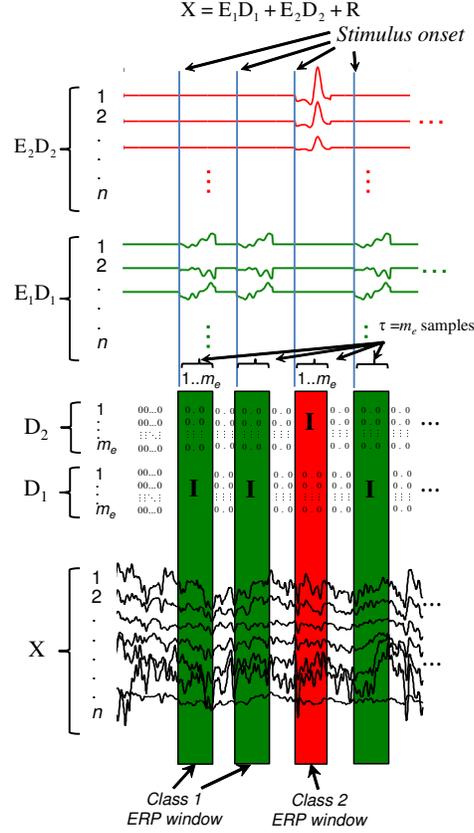


Figure 1: Graphical illustration of the model used for developing the proposed spatial filtering algorithm. There are two types of ERPs in the data time-locked to the onsets of specific stimuli (each type illustrated with a different color). \mathbf{X} is the $n \times m$ matrix of multi-channel EEG recording, \mathbf{E}_1 and \mathbf{E}_2 are the $n \times m_e$ matrices of ERP waveforms (see text for more details), and \mathbf{D}_1 and \mathbf{D}_2 are the $m_e \times m$ Toeplitz matrices characterizing the occurrence time of each ERP instance for the two classes, respectively. Therefore, all the elements of \mathbf{D}_1 and \mathbf{D}_2 are zero, except for the onset of a related stimulus where an identity matrix (\mathbf{I}) appears. The matrix \mathbf{R} represents all artifacts, noise and ongoing EEG contributions. According to this model, all the values in $\mathbf{E}_1 \mathbf{D}_1$ and $\mathbf{E}_2 \mathbf{D}_2$ are zero, except for the time points that an ERP exists.

Constructing the matrices $\mathbf{D}_1 = [\mathbf{d}_1(1), \mathbf{d}_1(2), \dots, \mathbf{d}_1(m)]$ and $\mathbf{D}_2 = [\mathbf{d}_2(1), \mathbf{d}_2(2), \dots, \mathbf{d}_2(m)]$, the objective is to estimate ERP response matrices, i.e., \mathbf{E}_1 and \mathbf{E}_2 (see Fig. 1 for a graphical illustration of the model). It is worth to mention that since the ERPs are time-locked to the stimuli and the stimulus markers are stored with the EEG data, the \mathbf{D}_1 and \mathbf{D}_2 matrices are always available. Assuming $\mathbf{X} = [\mathbf{x}(1), \mathbf{x}(2), \dots, \mathbf{x}(m)]$, the least square solutions to the problems could be estimated from $\hat{\mathbf{E}}_1 = \arg \min_{E_1} \|\mathbf{X} - \mathbf{E}_1 \mathbf{D}_1\|^2$ and $\hat{\mathbf{E}}_2 = \arg \min_{E_2} \|\mathbf{X} - \mathbf{E}_2 \mathbf{D}_2\|^2$, i.e.,

$$\begin{aligned} \hat{\mathbf{E}}_1 &= \mathbf{X} \mathbf{D}_1^T (\mathbf{D}_1 \mathbf{D}_1^T)^{-1} \\ \hat{\mathbf{E}}_2 &= \mathbf{X} \mathbf{D}_2^T (\mathbf{D}_2 \mathbf{D}_2^T)^{-1} \end{aligned} \quad (4)$$

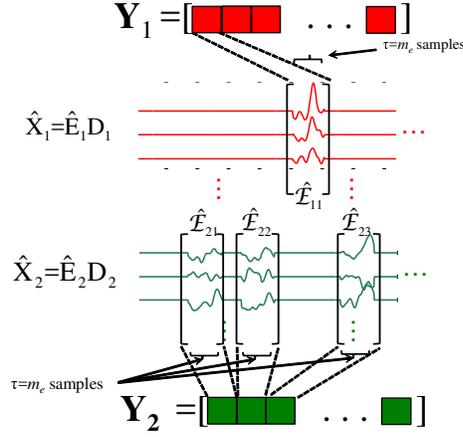


Figure 2: \mathbf{Y}_1 and \mathbf{Y}_2 matrices are constructed by horizontally concatenating the estimated ERP instances from $\hat{\mathbf{E}}_1 \mathbf{D}_1$ and $\hat{\mathbf{E}}_2 \mathbf{D}_2$, respectively.

115 Considering that \mathbf{D}_1 and \mathbf{D}_2 are $m_e \times m$ Toeplitz matrices defined using (3), it can be shown that $\mathbf{D}_1 \mathbf{D}_1^T$ and $\mathbf{D}_2 \mathbf{D}_2^T$ are invertible⁴ and hence (4) always have solutions. Therefore, the matrices $\hat{\mathbf{X}}_1 = \hat{\mathbf{E}}_1 \mathbf{D}_1$ and $\hat{\mathbf{X}}_2 = \hat{\mathbf{E}}_2 \mathbf{D}_2$ contain the enhanced brain responses. However, exploiting the redundancy in the data, our goal is to use only subspaces of $\hat{\mathbf{X}}_1$ and $\hat{\mathbf{X}}_2$ that contain more ERP contributions.

120 Assume that there are p instances of ERP class 1 ($\hat{\mathcal{E}}_{1i}$, for $1 \leq i \leq p$) in $\hat{\mathbf{X}}_1$ and q instances of the other class ($\hat{\mathcal{E}}_{2j}$, for $1 \leq j \leq q$) in $\hat{\mathbf{X}}_2$. As illustrated in Fig. 2, the matrices \mathbf{Y}_1 and \mathbf{Y}_2 are constructed by horizontally concatenating all instances of $\hat{\mathcal{E}}_{1i}$ and $\hat{\mathcal{E}}_{2j}$, respectively. Assuming that the ERPs of each class have similar waveforms, in an ideal case each row of \mathbf{Y}_1 and \mathbf{Y}_2 is a periodic signal with the period equal to the ERP window length, i.e., $\tau = m_e$.

Inspired by the periodicity measure defined in [45], the periodic spatial filter method is based on enhancing the periodic structure of one ERP class compared with the other one by minimizing

$$\mathcal{J}(\mathbf{w}, \tau) = \frac{\sum_t |\mathbf{w}^T \mathbf{Y}_1(t + \tau) - \mathbf{w}^T \mathbf{Y}_1(t)|^2}{\sum_t |\mathbf{w}^T \mathbf{Y}_2(t + \tau) - \mathbf{w}^T \mathbf{Y}_2(t)|^2} \quad (5)$$

where \mathbf{w} is the spatial filter vector. Expanding (5), it can be shown that $\mathcal{J}(\mathbf{w}, \tau)$ is equivalent to:

$$\mathcal{J}(\mathbf{w}, \tau) = \frac{\mathbf{w}^T \mathbf{P}_{\mathbf{Y}_1}(\tau) \mathbf{w}}{\mathbf{w}^T \mathbf{P}_{\mathbf{Y}_2}(\tau) \mathbf{w}} \quad (6)$$

⁴The inverse of a square matrix \mathbf{A} is the matrix \mathbf{A}^{-1} such that $\mathbf{A} \mathbf{A}^{-1} = \mathbf{I}$, where \mathbf{I} is the identity matrix. A matrix possessing an inverse is called invertible.

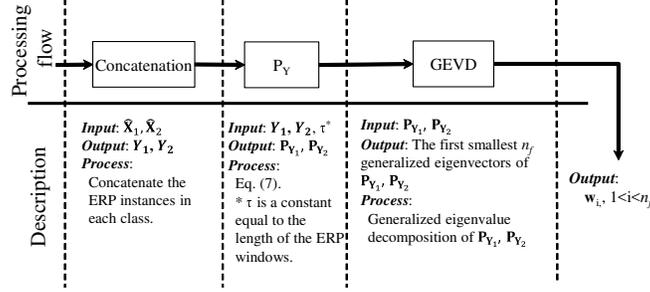


Figure 3: Processing flow and description of each step of π SF algorithm. Only steps 3 to 6 of the algorithm are illustrated. τ is a constant equal to the ERP window length (m_e).

where

$$\begin{aligned} \mathbf{P}_U(\tau) &= E\{[\mathbf{U}(t + \tau) - \mathbf{U}(t)][\mathbf{U}(t + \tau) - \mathbf{U}(t)]^T\} \\ &= 2\mathbf{C}_U(0) - \mathbf{C}_U(\tau) - \mathbf{C}_U(-\tau), \end{aligned} \quad (7)$$

$E\{\cdot\}$ denotes the expected value of the enclosed term, and $\mathbf{C}_U(\tau) = E\{\mathbf{U}(t + \tau)\mathbf{U}(t)^T\}$ is the delayed covariance matrix of $\mathbf{U} \in \{\mathbf{Y}_1, \mathbf{Y}_2\}$.

It can be shown that $\mathbf{C}_U(-\tau) = \mathbf{C}_U(\tau)^T$. Therefore, $\mathbf{P}_U(\tau)$ is symmetric positive-definite, and hence the generalized eigenvalues are real and positive [46]. Using the Rayleigh-Ritz theorem of linear algebra, the minimizer of (6) is given by the eigenvector corresponding to the smallest generalized eigenvalue of $\mathbf{P}_{Y_1}(\tau)$ and $\mathbf{P}_{Y_2}(\tau)$. Each eigenvector \mathbf{w}_i is called a spatial filter, and assigns different weights to the EEG electrodes.

By reordering the generalized eigenvectors in ascending order of the corresponding generalized eigenvalues, the obtained signals will be sorted based on the ratio of their periodicity at period τ . Therefore, selecting only the first few generalized eigenvectors leads to a subspace with the most relevant features for discriminating the two ERP classes. Steps of the π SF method are described in Algorithm 1, and diagrammatically illustrated in Fig. 3.

Algorithm 1 π SF

Require: Data is low-pass filtered at a predefined cutoff frequency.

- 1) Using (3), construct matrices \mathbf{D}_1 and \mathbf{D}_2 for the two classes, respectively,
- 2) Estimate $\hat{\mathbf{X}}_1$ and $\hat{\mathbf{X}}_2$ using (4), i.e., $\hat{\mathbf{X}}_1 = \hat{\mathbf{E}}_1\mathbf{D}_1 = \mathbf{X}\mathbf{D}_1^T(\mathbf{D}_1\mathbf{D}_1^T)^{-1}\mathbf{D}_1$ and $\hat{\mathbf{X}}_2 = \hat{\mathbf{E}}_2\mathbf{D}_2 = \mathbf{X}\mathbf{D}_2^T(\mathbf{D}_2\mathbf{D}_2^T)^{-1}\mathbf{D}_2$,
- 3) Construct \mathbf{Y}_1 and \mathbf{Y}_2 by concatenating instances of the two ERP classes from $\hat{\mathbf{X}}_1$ and $\hat{\mathbf{X}}_2$, respectively,
- 4) Calculate $\mathbf{P}_{Y_1}(\tau)$ and $\mathbf{P}_{Y_2}(\tau)$ using (7),
- 5) Calculate GEVD* of $\mathbf{P}_{Y_1}(\tau)$ and $\mathbf{P}_{Y_2}(\tau)$,
- 6) The first n_f eigenvectors corresponding to the smallest eigenvalues are selected as the spatial filters.

* GEVD: generalized eigenvalue decomposition

Periodic spatio-spectral filter

135 In developing the π SF method, we assumed that all EEG channels are low-pass filtered using a pre-designed FIR filter. In this section, the π SF method is enhanced by estimating a customized FIR filter for each estimated spatial filter. Similar to spatial filter vectors, the FIR filter coefficients have to enhance the periodic structure by minimizing the cost function (6).

The same formulation as in (2) is used to model the unfiltered EEG recording, i.e., $\mathbf{s}(t) = \mathbf{E}_{s1}\mathbf{d}_1(t) + \mathbf{E}_{s2}\mathbf{d}_2(t) +$
 140 $\mathbf{r}_s(t)$, where \mathbf{E}_{s1} and \mathbf{E}_{s2} are the ERP responses and $\mathbf{r}_s(t)$ is the noise vector in the raw data, respectively. Following the procedure used for constructing \mathbf{Y}_1 and \mathbf{Y}_2 , we construct \mathbf{Z}_1 and \mathbf{Z}_2 from the ERP instances in $\hat{\mathbf{S}}_1 = \hat{\mathbf{E}}_{s1}\mathbf{D}_1 = \mathbf{S}\mathbf{D}_1^T(\mathbf{D}_1\mathbf{D}_1^T)^{-1}\mathbf{D}_1$ and $\hat{\mathbf{S}}_2 = \hat{\mathbf{E}}_{s2}\mathbf{D}_2 = \mathbf{S}\mathbf{D}_2^T(\mathbf{D}_2\mathbf{D}_2^T)^{-1}\mathbf{D}_2$, respectively, (the procedure is similar to that of π SF except that instead of $\mathbf{x}(t)$ the raw data, $\mathbf{s}(t)$, is used).

Assume that \mathbf{w} is known and \mathbf{h} is the unknown variable, we have to modify (6) to be a function of \mathbf{h} . Considering (7), the cost function in (6) is composed of the terms $\mathbf{w}^T\mathbf{C}_U(\tau)\mathbf{w}$ for different τ values. Assuming $\mathbf{U}(t) = \sum_0^L h_l\mathbf{V}(t-l)$, this term can be reformulated as:

$$\mathbf{w}^T\mathbf{C}_U(\tau)\mathbf{w} = \mathbf{h}^T\mathbf{B}_V(\tau)\mathbf{h} \quad (8)$$

where $\mathbf{V} \in \{\mathbf{Z}_1, \mathbf{Z}_2\}$, and $\mathbf{B}_V(\tau)_{i,j} = \mathbf{w}^T\mathbf{C}_V(\tau-i+j)\mathbf{w}$ (see Appendix for the details). Therefore, using (8) one can rewrite (6) as:

$$\mathcal{J}(\mathbf{h}, \tau) = \frac{\mathbf{h}^T\mathbf{Q}_{Z_1}(\tau)\mathbf{h}}{\mathbf{h}^T\mathbf{Q}_{Z_2}(\tau)\mathbf{h}} \quad (9)$$

where $\mathbf{Q}_V(\tau) = 2\mathbf{B}_V(0) - \mathbf{B}_V(\tau) - \mathbf{B}_V(-\tau)$.

145 Similar to the approach used to solve (6), the minimizer of (9) can be estimated using the generalized eigenvalue decomposition of $\mathbf{E}_{Z_1}(\tau)$ and $\mathbf{E}_{Z_2}(\tau)$. Therefore, in order to solve the spatio-spectral filtering problem, we have to alternate between solving (6) and (9) to find the minimizer vectors. To this end, we first assume that \mathbf{h} is known and estimate \mathbf{w} , then keeping \mathbf{w} fixed, \mathbf{h} is updated. Until convergence, this procedure has to be repeated for all n_f spatial filters. This way, for each spatial filter a different customized spectral filter is estimated. The proposed
 150 spatio-spectral filtering approach is summarized in Algorithm 2. Processing flow and description of each sub-step of step 4 of the π SSF algorithm are illustrated in Fig. 4.

Remarks:

1. The main concept of the π SF and the π SSF methods are similar. The difference between the two algorithms is that in π SSF we estimate the specific spectral filter for each spatial filter, separately.
- 155 2. To avoid filtering distortions, the ERP class instances have to be cut with a size larger than τ . After filtering, only a portion of the data, which is not affected by the filtering artifacts, will be used to represent the ERPs.

Algorithm 2 π SSF

- 1) Using (3), construct matrices \mathbf{D}_1 and \mathbf{D}_2 for the two classes,
- 2) Estimate $\hat{\mathbf{S}}_1$ and $\hat{\mathbf{S}}_2$ using (4) *, i.e., $\hat{\mathbf{S}}_1 = \hat{\mathbf{E}}_{s1} \mathbf{D}_1 = \mathbf{S} \mathbf{D}_1^T (\mathbf{D}_1 \mathbf{D}_1^T)^{-1} \mathbf{D}_1$ and $\hat{\mathbf{S}}_2 = \hat{\mathbf{E}}_{s2} \mathbf{D}_2 = \mathbf{S} \mathbf{D}_2 (\mathbf{D}_2 \mathbf{D}_2^T)^{-1} \mathbf{D}_2$,
- 3) Initialize $\mathbf{h} = [1 \ 0 \ 0 \ \dots \ 0]^T$,
- 4) For each i , $1 \leq i \leq n_f$:
 - i) Repeat until convergence
 - a) Construct \mathbf{Z}_1 and \mathbf{Z}_2 from $\hat{\mathbf{S}}_1$ and $\hat{\mathbf{S}}_2$, respectively,
 - b) $\mathbf{Y}_1(t) = \sum_0^L h_l \mathbf{Z}_1(t-l)$, $\mathbf{Y}_2(t) = \sum_0^L h_l \mathbf{Z}_2(t-l)$,
 - c) Calculate $\mathbf{P}_{Y_1}(\tau)$ and $\mathbf{P}_{Y_2}(\tau)$,
 - d) Estimate \mathbf{w}_i , the i th generalized eigenvector of $\mathbf{P}_{Y_1}(\tau)$ and $\mathbf{P}_{Y_2}(\tau)$,
 - e) Calculate $\mathbf{Q}_{Z_1}(\tau)$ and $\mathbf{Q}_{Z_2}(\tau)$,
 - f) Estimate \mathbf{h}_i , the smallest GEV of $\mathbf{Q}_{Z_1}(\tau)$ and $\mathbf{Q}_{Z_2}(\tau)$,

* Raw data is used, i.e., $\mathbf{S} = [s(1), s(2), \dots, s(m)]$

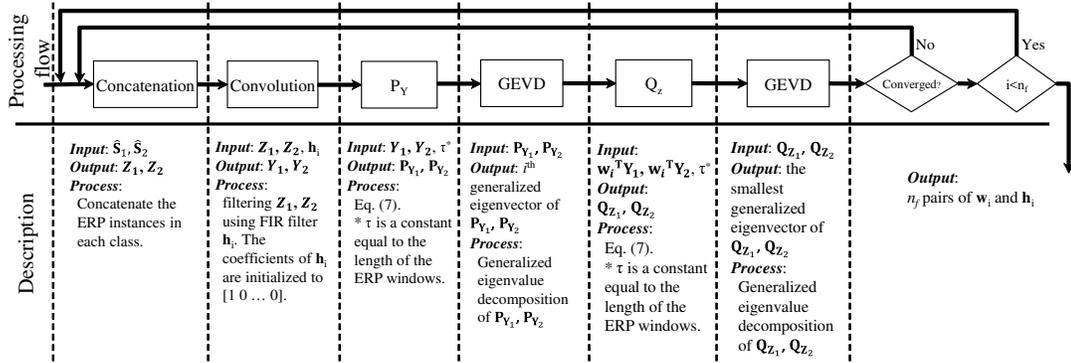


Figure 4: Processing flow and description of step 4 of π SSF algorithm. τ is a constant equal to the ERP window length (m_e).

3. $\mathbf{Q}_{Z_1}(\tau)$ and $\mathbf{Q}_{Z_2}(\tau)$ have to be calculated in step e of the algorithm at each iteration. Reestimating these matrices in every iteration of the algorithm is a computationally expensive task. However, $\mathbf{B}_V(\tau)$ is a function of \mathbf{w} and $\mathbf{C}_V(\tau)$ for different τ values. Although \mathbf{w} is modified in each iteration, $\mathbf{C}_{Z_1}(\tau)$ and $\mathbf{C}_{Z_2}(\tau)$ are constant and have to be evaluated only once for each dataset.

160

Materials and Methods

Experimental setup and testbed

We evaluated the methods using data recorded in a previous study [47] in which a testbed was used that requested the participants to play a labyrinth game and to respond to infrequent target stimuli (see Fig. 5). The testbed allows to investigate the EEG activity of an operator who is controlling a device while reacting to incoming infrequent information at the same time.⁵ Two sessions were recorded on two different days with at least one day

165

⁵For visualization of the task and setup see <https://www.youtube.com/watch?v=PJIHJTWS7AQ>



Figure 5: Experimental setup and view of the participants through the head mounted display (HMD) on the virtualized game and presented infrequent first target and frequent non-target stimuli.

break in between. Each session consisted of five repetitions (called “runs”) of the Labyrinth Oddball paradigm. After each of the five runs there was a break of 10 minutes. In each dataset (recorded during each run), there were 120 and 720 target and non-target instances, respectively, presented with an inter-stimulus interval (ISI) of 1000 ms with a random jitter of ± 100 ms.

The labyrinth game is a virtualized BRIO[®] labyrinth game which was displayed to the participants via a head mounted display (HMD). The virtualized game was controlled by using a real BRIO[®] labyrinth game. Playing the game was a complex sensorimotor task. The requested response was to press a buzzer after recognizing a task-relevant target stimulus. Participants were instructed to respond to all target stimuli even in case they were uncertain. They were in a competition to miss as few as possible targets while achieving good performance in the game. The task-relevant infrequent target stimuli were interleaved with task-irrelevant frequent stimuli. Target stimuli were under-represented compared to non-target stimuli (ratio 1 : 6). Hence, the participants had to distinguish two different kinds of stimuli: task-relevant infrequent targets and task-irrelevant frequent non-targets, while playing the labyrinth game. Stimuli were presented in random order. Since the manipulation task was very demanding, a rather long response time from 200 ms to approximately 2000 ms after target stimulus presentation was allowed before a second target stimulus was presented. In case there was no response within the given time period on the first target stimulus, the trial was labeled as missed target. In case of a missed target stimulus it was expected that the participant did not see or recognize the target stimulus or was unable to respond because of very high workload (see [47] for discussion). On the second target a response time of 200 ms to 1000 ms was allowed.

Frequent and infrequent first target stimuli were very similar in shape and occurrence in order to avoid differences in early visual processing of the stimuli (see Fig. 5). For this work only EEG segments after frequent non-target stimuli and infrequent target stimuli of the type of first target were used. EEG segments after missed target stimuli and second target stimuli were excluded from analysis. With respect to muscle artifacts it must be

stated that the subject did continuously move their hands and arms all through the runs to play the game.

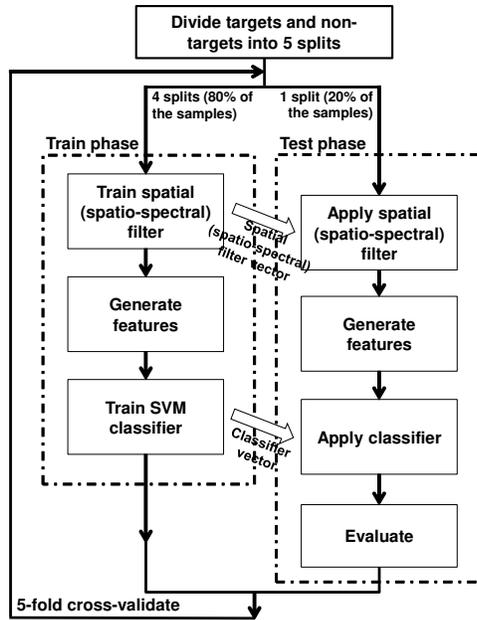


Figure 6: The processing flow for single trial ERP classification is evaluated by 5-fold cross-validation. Keeping the ratio of targets vs. non-targets in each dataset, the ERP instances are divided into 5 splits, where 4 of them are used for training spatial filter and classifier and the last one is used for testing. The average of the performances is reported as the overall performance. The same processing flow has been used for xDAWN, π SF, and π SSF methods.

190 *Participants*

EEG data were recorded from 6 male participants (mean age 27.5, standard deviation 2.1; right-handed, and normal or corrected-to-normal vision) performing in the "Labyrinth Oddball" scenario [47]. All participants were trained to play the virtualized labyrinth game of the testbed.

Data acquisition procedure

195 Data was continuously recorded at 1 kHz sampling rate from 64 electrodes (in accordance with the extended 10-20 system with reference at electrode FCz, data recorded at electrodes at TP7 and TP8 was excluded), using an actiCap system (Brain Products GmbH, Munich, Germany) and amplified by two 32 channel BrainAmp DC amplifiers (Brain Products GmbH, Munich, Germany). The impedance was kept below 5 k Ω .

Ethics statement

200 The study has been conducted in accordance with the Declaration of Helsinki and approved by the ethics committee of the University of Bremen. Participants had given informed written consent to participate.

Data processing

In order to assess our spatio-spectral filter, we conducted separate evaluations using xDAWN, π SF, and π SSF. For all the methods, we down-sampled the recorded data to 200 Hz, and high-pass filtered at 0.1 Hz. For the π SF and xDAWN methods, we used a low-pass FIR filter of length 20 and a cutoff frequency of 7 Hz. For the π SSF method, we first low-pass filtered the data at 45 Hz before training the spectral filters.

To avoid filtering artifacts, we cut the ERP instances between 1000 ms before and 2000 ms after the stimulus onset. After applying the spectral filters to the data, we used only 1000 ms of the data starting at stimulus onset for further analysis. In order to have a fair comparison, this procedure was used for all methods. That means, for the π SF and π SSF methods $\tau = m_e = 1$ s.

We generated feature vectors by concatenating the data from all spatially filtered channels. Although the method enhances the periodic structure of one class compared with the other one, it is not trivial to exploit the periodicity of the filtered data when extracting features. This is because the classification is performed on single instances of the ERPs. Therefore, feature vectors were constructed by concatenating all channels and normalizing the resulted time series⁶.

We used a support vector machine (SVM) classifier with linear kernel in the next step to discriminate the two ERP classes. To overcome overfitting of the classifier, we found the optimum values for the regularization parameter of the SVM classifier for each dataset separately using grid search technique from a set of predefined values, i.e., $\{0, 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}, 10^{-6}\}$. We also used the threshold optimization technique [48, 49] to adjust the separation level between the two classes at the output of the classifier. This technique is based on evaluating the histograms of the classifier’s prediction scores for the two classes and choosing a threshold value that is optimal for a given evaluation metric on the training data. As illustrated in Fig. 6, we then randomly divided the targets and non-targets in each dataset into 5 disjoint splits. In a 5-fold cross-validation schema, we trained the spatial filters on the data from 4 splits. Independent of this training process, we trained the classifier using the features generated from the spatially filtered data. We then tested the obtained model, including the spatial (spatio-temporal) filter and the classifier, using the remaining split of the data. The final result for each dataset is the average over 5 values. For experiments with the π SSF method, the spatial and spectral filters were jointly estimated before being used in the feature generation module.

After conducting all the experiments, we obtained 60 samples: 6 subjects x 2 sessions (2 different days) x 5 runs. The result of each experiment is reported as balanced accuracy (BA), which is the average of true positive and true negative rates and therefore unaffected by the unbalanced class distributions. All evaluations were conducted using pySPACE, which is an open source software framework for signal processing and machine learning [49].

⁶The time series have zero mean and unit variance after normalization.

Statistical analysis

The obtained results were statistically analyzed based on three research questions (hypotheses): Firstly, do
235 the investigated approaches (spatial filters) lead to an improvement in the classification performance compared to
the baseline (no spatial filter)? Here, we check whether there are any differences between different approaches
(no spatial filter, xDAWN, π SF, π SSF). When we find any effect on the classification performance depending on
approach type, a post-hoc analysis (pairwise comparisons between different approaches) is performed to determine
e.g., whether the presented approach π SSF improves the classification performance more than the other approaches
240 (" π SF" and "xDAWN"). Secondly, does the number of retained channels of spatial filters have an impact on the
classification performance? Again, it is first investigated whether there are any differences between the different
numbers of retained channels in the classification performance and then a post-hoc analysis is performed to find
how many numbers of retained channels lead to the highest classification performance. Finally, does the filter
length of the π SSF affect the classification performance? Here, we investigate whether there are any differences
245 between the different filter lengths. Again, a post-hoc analysis is performed to determine which range of filter
length is optimal for the system.

Based on the research questions, we investigated three effects on the classification performance: 1) effect of
approach type (no spatial filter, xDAWN, π SF, π SSF), 2) effect of the number of retained channels (2, 3, 4, 6,
8, 15, 30), and 3) effect of filter length (5, 6,...20). For statistical analysis, we selected three different designs
250 depending on the approach type, since it is not possible to evaluate all independent variables (approach type,
number of retained channels, filter length) within one evaluation design for statistical analysis. Actually, some
levels of independent variable (xDAWN, π SF) do not contain filter length (5, 6,...20) and our baseline (no use of
spatial filter) contains neither filter length (5, 6,...20) nor number of retained channels (2, 3, 4, 6, 8, 5, 30). In
contrast, π SSF contains both filter length (5, 6,...20) and number of retained channels (2, 3, 4, 6, 8, 5, 30). Thus,
255 we evaluate the results based on three different evaluation designs.

First, we investigated the effect of the number of retained channels on classification performance. To this end,
we performed a two-way repeated measures ANOVA with approach type (xDAWN, π SF) and number of retained
channels (2, 3, 4, 6, 8, 15, 30) as within-subjects factors (*Design 1*). We excluded the approach π SSF in this
analysis, since the π SSF additionally contains filter length.

260 Thus, the classification performance obtained by using the π SSF was separately analyzed to investigate the
effect of filter length and number of retained channels on the classification performance. To this end, we performed
a two-way repeated measures ANOVA with number of retained channels (2, 3, 4, 6, 8, 15, 30) and filter length (5,
6,...20) as within-subjects factors (*Design 2*).

For comparison of approaches (xDAWN, π SF, π SSF) with the baseline (no use of spatial filter), we fixed the
265 number of retained channels and filter length, since the baseline contains neither number of retained channels nor

filter length as factors. Here, we performed a one-way repeated measures ANOVA with approach type (xDAWN, π SF, π SSF, no spatial filter) as a within-subjects factor (*Design 3*). For all statistical analyses, Greenhouse–Geisser correction was performed where necessary and Bonferroni-Holm correction was performed for multiple comparisons.

		Number of retained channels after spatial filtering							No spatial filter
		2	3	4	6	8	15	30	
S1	xDAWN	0.870 ± 0.023	0.882 ± 0.026	0.887 ± 0.021	0.895 ± 0.021	0.888 ± 0.022	0.886 ± 0.021	0.881 ± 0.027	0.795 ± 0.035
	π SF	0.887 ± 0.018	0.886 ± 0.021	0.894 ± 0.025	0.889 ± 0.016	0.892 ± 0.026	0.893 ± 0.021	0.893 ± 0.017	
S2	xDAWN	0.825 ± 0.027	0.827 ± 0.022	0.818 ± 0.022	0.823 ± 0.025	0.841 ± 0.023	0.809 ± 0.033	0.835 ± 0.038	0.768 ± 0.024
	π SF	0.825 ± 0.031	0.830 ± 0.029	0.821 ± 0.024	0.840 ± 0.021	0.829 ± 0.033	0.826 ± 0.029	0.841 ± 0.034	
S3	xDAWN	0.864 ± 0.034	0.863 ± 0.029	0.850 ± 0.033	0.866 ± 0.033	0.862 ± 0.036	0.835 ± 0.040	0.837 ± 0.035	0.794 ± 0.042
	π SF	0.854 ± 0.034	0.867 ± 0.028	0.860 ± 0.032	0.862 ± 0.036	0.857 ± 0.058	0.037 ± 0.047	0.864 ± 0.026	
S4	xDAWN	0.789 ± 0.023	0.796 ± 0.038	0.805 ± 0.026	0.788 ± 0.031	0.764 ± 0.041	0.793 ± 0.033	0.808 ± 0.030	0.752 ± 0.044
	π SF	0.801 ± 0.022	0.813 ± 0.030	0.821 ± 0.033	0.798 ± 0.032	0.775 ± 0.045	0.768 ± 0.032	0.812 ± 0.031	
S5	xDAWN	0.916 ± 0.023	0.918 ± 0.025	0.916 ± 0.021	0.922 ± 0.017	0.929 ± 0.018	0.917 ± 0.021	0.905 ± 0.027	0.856 ± 0.019
	π SF	0.911 ± 0.028	0.921 ± 0.021	0.922 ± 0.014	0.925 ± 0.021	0.925 ± 0.023	0.908 ± 0.027	0.911 ± 0.032	
S6	xDAWN	0.875 ± 0.023	0.870 ± 0.020	0.871 ± 0.021	0.860 ± 0.027	0.856 ± 0.015	0.870 ± 0.021	0.868 ± 0.017	0.805 ± 0.016
	π SF	0.871 ± 0.025	0.872 ± 0.023	0.874 ± 0.022	0.861 ± 0.031	0.850 ± 0.015	0.863 ± 0.028	0.856 ± 0.020	
Avg	xDAWN	0.857 ± 0.044	0.859 ± 0.043	0.858 ± 0.042	0.859 ± 0.048	0.857 ± 0.055	0.852 ± 0.048	0.856 ± 0.035	0.795 ± 0.036
	π SF	0.858 ± 0.040	0.863 ± 0.039	0.863 ± 0.040	0.862 ± 0.043	0.855 ± 0.052	0.852 ± 0.050	0.863 ± 0.036	

Table 1: Target versus non-target classification performance (balanced accuracy) in Labyrinth Oddball scenario. Results obtained from using xDAWN and π SF and also the configuration that no spatial filter was used are presented. ERP instances were randomly divided into 5 splits (with almost equal ratio of number of targets vs. non-targets, i.e., 1:6). Different numbers of spatially filtered channels were retained to evaluate the performance of the spatial filters. The xDAWN and π SF and the classifier were trained using 4 splits and tested using the last one. The numbers in each cell of the table represent the average and standard deviation of the experiments. Mean and margin of errors are reported (95% confidence interval (CI) = mean ± margin of errors)

270 Results

Table 1 presents the classification performances of the xDAWN and π SF methods compared with the case that no spatial filter was used. Results obtained from all datasets of each participant and also the overall averages are reported. To see the effect of the number of spatially filtered channels on the performance, we repeated the experiments for each dataset by selecting different numbers of retained channels. Figure. 7.a shows the mean classification performance for different numbers of retained channel across xDAWN and π SF (see section statistical analysis, Design 1). We found no main effect of the number of retained channels [$F(6, 354) = 1.39, p = n.s., \eta_{partial}^2 = 0.02$ (small effect)].

Figure. 7.b shows the mean classification performance for each approach across all types of retained channels. There was a main effect of approach type, i.e., the π SF achieved a higher performance compared to xDAWN [$F(1, 59) = 4.91, p < 0.032, \eta_{partial}^2 = 0.08$ (medium effect)]. Figure 7.c shows the mean classification performance for the different numbers of retained channels for both approaches (xDAWN and π SF). There was no interaction between approach type and number of retained channels [$F(6, 354) = 0.89, p = n.s., \eta_{partial}^2 =$

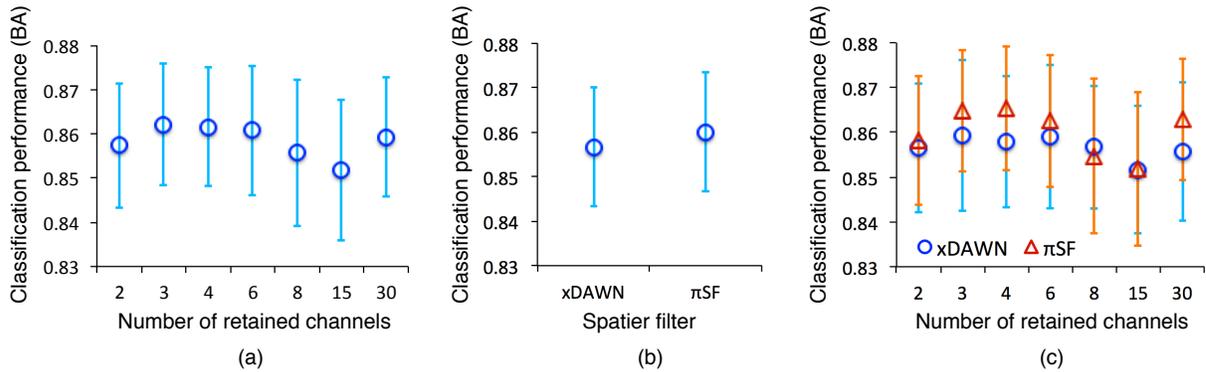


Figure 7: Comparison between different numbers of retained channels and different approaches. (a) Classification performance for each retained channel over xDAWN and π SF, (b) Classification performance for each approach (xDAWN, π SF) over all types of retained channels and (c) Classification performance for each approach (xDAWN, π SF) and each retained channel. Error bars indicate the 95% CI.

0.02 (small effect)]. For both approaches, there were no significant differences between the numbers of retained channels.

285 Figure. 8 shows the averages of the BAs obtained from the experiments with the periodic spatio-spectral π SSF filter. Different FIR filter lengths were used and the results are reported for experiments with different numbers of retained channels. Both of these factors affected the classification performance. Figure. 8.a shows the mean classification performance for the different numbers of retained channels across all types of filter lengths (see section statistical analysis, Design 2). We found a main effect of number of retained channels [$F(6, 354) = 36.08$, $p < 0.001$, $\eta^2_{partial} = 0.38$ (large effect size)]. The classification performance was slightly higher for π SF compared to xDAWN [xDAWN: mean = 0.857 (confidence interval (CI): 0.843, 0.870), π SF: mean = 0.860 (CI: 0.847, 0.873, mean difference between π SF and xDAWN = 0.003 (CI: 0.006, 0.000)]. The classification performance was reduced with increased number of retained channels. We obtained the highest performance with 4 retained channels, which significantly differed from all types of retained channels except for 3 retained channels [4 vs. 2: $t(59) = 4.36$, $p < 0.001$, mean difference: 0.011 (CI: 0.018, 0.003), $r = 0.20$ (small effect); 4 vs. 3: $t(59) = 2.05$, $p = n.s.$, mean difference: 0.003 (CI: 0.008, 0.001), $r = 0.09$ (very small effect); 4 vs. 6: $t(59) = 2.91$, $p < 0.001$, mean difference: 0.006 (CI: 0.010, 0.002), $r = 0.18$ (small effect); 4 vs. 8: $t(59) = 4.52$, $p < 0.001$, mean difference: 0.013 (CI: 0.018, 0.008), $r = 0.34$ (medium effect); 4 vs. 15: $t(59) = 7.70$, $p < 0.001$, mean difference: 0.026 (CI: 0.031, 0.0192), $r = 0.45$ (medium effect); 4 vs. 30: $t(59) = 9.26$, $p < 0.001$, mean difference: 0.035 (CI: 0.042, 0.027), $r = 0.51$ (larger effect)].

On the other hand, with 30 retained channels the worst performance was achieved compared to all other numbers of retained channels, followed by 15 retained channels [30 vs. all: $p < 0.001$, 15 vs. all: $p < 0.001$]. The reduction in performance for 15 and 30 retained channels was further observed for most filter lengths (see, Fig. 8.c)

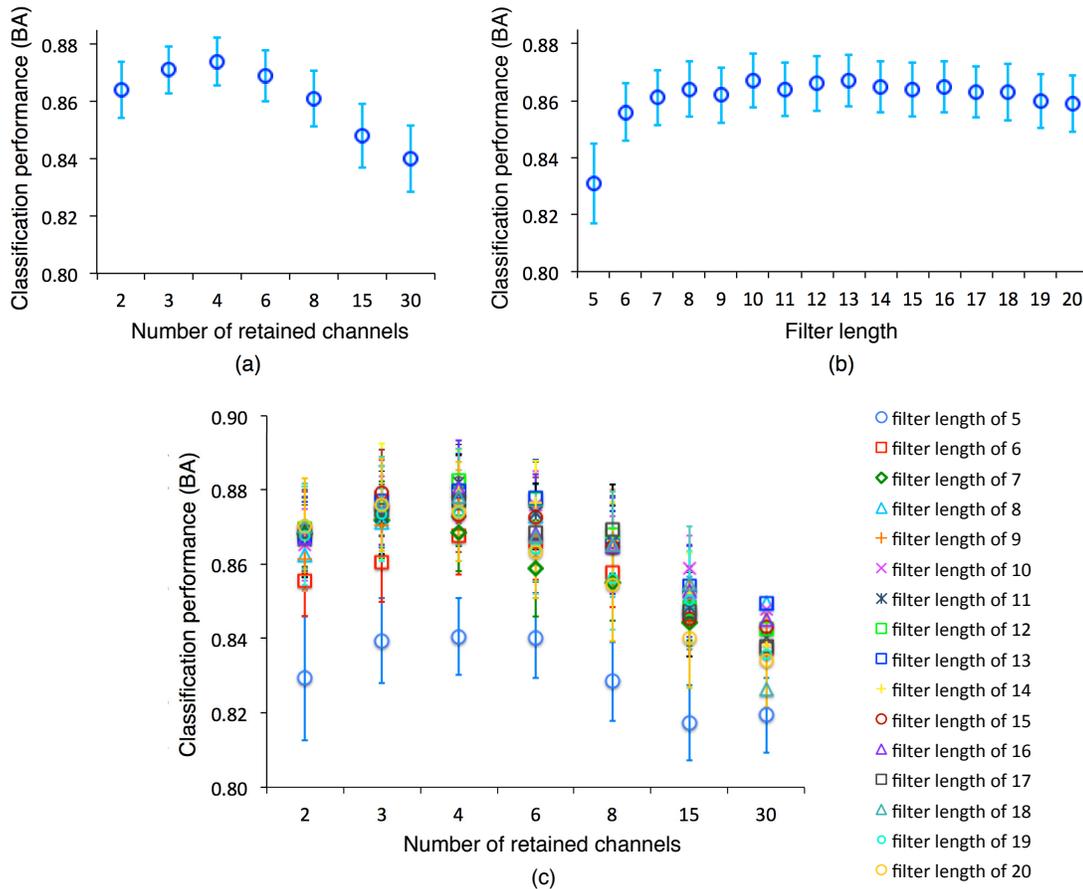


Figure 8: Comparison between different numbers of retained channels and different filter types, when π SSF is used. (a) Classification performance for each retained channel over all filter types, (b) Classification performance for each filter over all types of retained channels and (c) Classification performance for each retained channel and each filter type. Error bars indicate the 95% CI.

Figure 8.b shows the mean classification performance for the different filter lengths across all types of retained channels. We found a main effect of filter length [$F(15, 885) = 14.82, p < 0.001, \eta_{partial}^2 = 0.20$ (large effect)]. We observed that using very short FIR filters degrades the performance of the system, e.g., the filter length of 5 achieved the worst performance compared to the other filter lengths [5 vs. all: $p < 0.001$]. Further, the significant reduction caused by the usage of a very short FIR filter length could be shown for all types of retained channels [5 vs. all: $p < 0.029$] except for 30 retained channels (see Fig. 8.c). Such reduction was also shown for very long FIR filters (i.e., filter length of 18, 19, 20). Thus, we observed no difference between the filter length of 5 and the filter length of 18, 19, and 20 in the classification performance [5 vs. 18: $p = n.s.$, 5 vs. 19: $p = n.s.$, 5 vs. 20: $p = n.s.$]. The obtained results for short filters (less than 5) were not consistent, and hence not included in the figure.

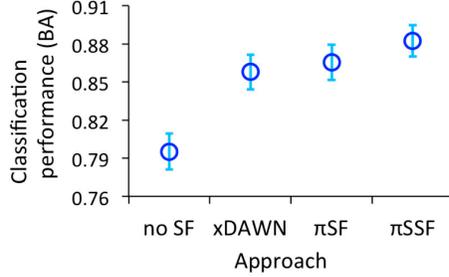
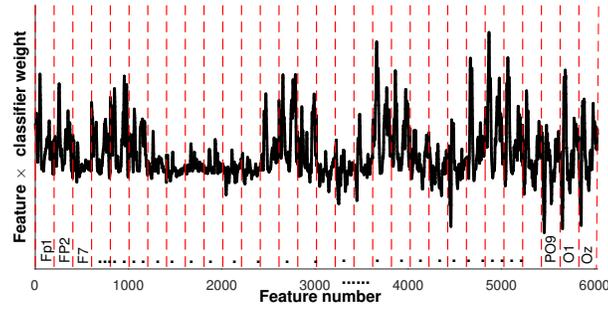


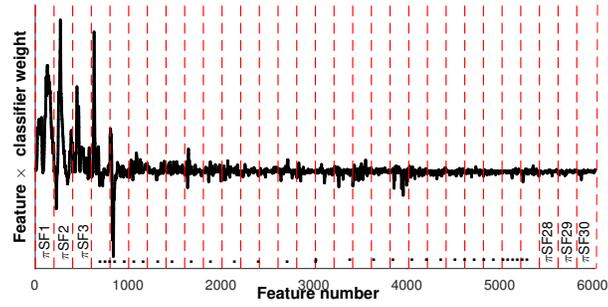
Figure 9: Mean classification performance of different approaches across all subjects. Error bars indicate the 95% CI.

Balanced accuracies obtained from the three approaches, i.e., xDAWN, π SF, and π SSF are presented together with the baseline (no use of spatial filter) in Fig. 9 (see section statistical analysis, Design 3). For all approaches, 4 channels are retained and the length of the FIR filter for the experiments with π SSF is 12. We found a main effect of approach type [$F(3, 177) = 101.71, p < 0.001, \epsilon = 0.0737, p < 0.001, \eta_{partial}^2 = 0.63$ (large effect size)]. The π SSF achieved the best classification performance [mean: 0.882 (CI: 0.870, 0.895)] followed by the π SF [mean: 0.856 (CI: 0.852, 0.879)], xDAWN [mean: 0.858 (CI: 0.844, 0.872)], and no spatial filter [mean: 0.795 (CI: 0.781, 0.809)]. The π SSF significantly outperformed all other approaches [π SSF vs. no spatial filter: $t(59) = 12.99, p < 0.001$, mean difference: 0.087 (CI: 0.069, 0.106), $r = 0.64$ (large effect); π SSF vs. xDAWN: $t(59) = 5.86, p < 0.001$, mean difference: 0.025 (CI: 0.013, 0.036), $r = 0.35$ (medium effect); π SSF vs. π SF: $t(59) = 3.98, p < 0.004$, mean difference: 0.017 (CI: 0.003, 0.031), $r = 0.21$ (medium effect)]. However, the π SF did not significantly differ from the xDAWN [$t(59) = -2.29, p = n.s.$, mean difference: 0.008 (CI: 0.001, 0.017), $r = 0.15$ (small effect)]. No use of spatial filter achieved the worst performance, which significantly differed from all other approaches [no spatial filter vs. xDAWN: $t(59) = 11.04, p < 0.001$, mean difference: 0.063 (CI: 0.078, 0.047), $r = 0.58$ (large effect); no spatial filter vs. π SF: $t(59) = 11.06, p < 0.001$, mean difference: 0.070 (CI: 0.088, 0.053), $r = 0.58$ (large effect); no spatial filter vs. π SSF: $t(59) = 12.99, p < 0.001$, mean difference: 0.087 (CI: 0.106, 0.069), $r = 0.64$ (large effect)].

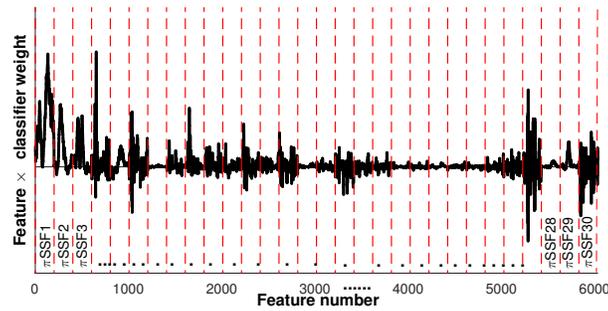
Training the classifier would result in a linear transformation, i.e., \boldsymbol{w} and b which are the classifier weight vector and the offset of the classifier, respectively. In the test phase, the classifier predicts the label of every input data based on the inner product of the two vectors $[\boldsymbol{w}, b]$ and $[\boldsymbol{f}, 1]$, where \boldsymbol{f} is the feature vector. In order to investigate the effect of using spatial filters on the behavior of classifiers, we illustrated the averages of the Hadamard products of \boldsymbol{w} and \boldsymbol{f} of all target instances of one of the datasets in Fig. 10. In this way, the spatial filters and the classifier are merged in a single linear transformation [24] and their effect on the data can be seen together. The red dashed lines in Fig. 10 separate the features extracted from different channels.



(a) No spatial filter used.



(b) π SF used



(c) π SSF used

Figure 10: Element-wise multiplication of classifier coefficient vector and the normalized feature vectors for (a) raw data and (b), (c) the data that was spatially filtered using π SF and π SSF, respectively (1s of data from 30 retained spatially filtered channels, sampled at 200 Hz). Red dashed lines separate the features taken from different channels. Illustrated are the averages calculated from all target instances from one of the datasets in the test phase. Similar patterns were observed for target and non-target instances from all participants. π SF $_i$ (π SSF $_i$) represent the i th channel retained after application of π SF (π SSF) to the raw data. The FIR filter length for π SSF was 12.

Figure 10a shows the averaged products for the case that no spatial filter was used. In other words, the raw data from the first thirty electrodes from the 10-20 system (FP1, FP2, F7, ..., O1, Oz) is used in this plot. It is difficult to say which electrodes have the most contribution to the final prediction value. However, from the average plot, it can be seen that the data from some electrodes in frontal, central, parietal and occipital regions are more involved. Figures. 10b and 10c show the Hadamard products for the cases that π SF and π SSF are applied. Here, the features are generated from the filtered data, i.e., π SF1, π SF2, ..., π SF30 and π SSF1, π SSF2, ..., π SSF30, respectively. In contrast to Fig. 10a, the most important elements are condensed in the left part of the plot in Fig. 10b, which correspond to the first few spatially filtered channels. Remember that the spatial filters are sorted in ascending order based on their corresponding eigenvalues. Therefore, from the classifier's point of view the first few retained channels have more discriminating features. From Fig. 10c one can see that there are still significant contributions at the first few channels, however, there are also some features with high weights at the other spatio-spectrally filtered channels.

In order to investigate the effect of the designed filters on the ERP waveforms, we illustrated the grand averages of the targets and non-targets in the output channels of the xDAWN, π SF and π SSF in Fig. 11. We also plotted the spatial filter coefficients on topographic maps to visualize the importance of each electrode to the filtered data. If the i th spatial filter assigns higher weights to an EEG electrode, that means the data from that electrode has more contribution to the i th retained channel. As expected, the largest filter coefficients are mostly located in the central, parietal, and occipital regions of the brain which correspond to the electrode positions that show P300 activity, as one of the most pronounced ERP activities evoked in our experimental setup by target stimuli compared to non-target stimuli. The waveforms and the topographic maps in Fig. 11.q - 11.r are much different from those in Fig. 11.e - 11.f and Fig. 11.k - 11.l. The difference is more obvious in the 6th retained channel where the low pass components are filtered out.

Fig. 12 illustrates the frequency responses of the FIR filters estimated by the π SSF method for one of the datasets. The length of all the filters is 12, and the illustrated filters are associated with the first six spatial filters. From the frequency responses, one can conclude that all the filters (except the last one) are band-stop filters, however, since the data was low-pass filtered in advance (the cutoff frequency was 45 Hz), in practice the filters operate like low-pass filters. In contrast to the first 5 filters, the 6th one is different from the others and is a high-pass filter and affects the output of the spatio-spectral filter. This effect can be clearly seen in the Fig. 11.r.

Discussions

The results confirm that all spatial filters can consistently improve the single trial classification performance. In particular, the significant superior classification performance can be achieved with the extended version of spatial filter (π SSF) compared with other spatial filters (π SF, xDAWN). This confirms the hypothesis that there are

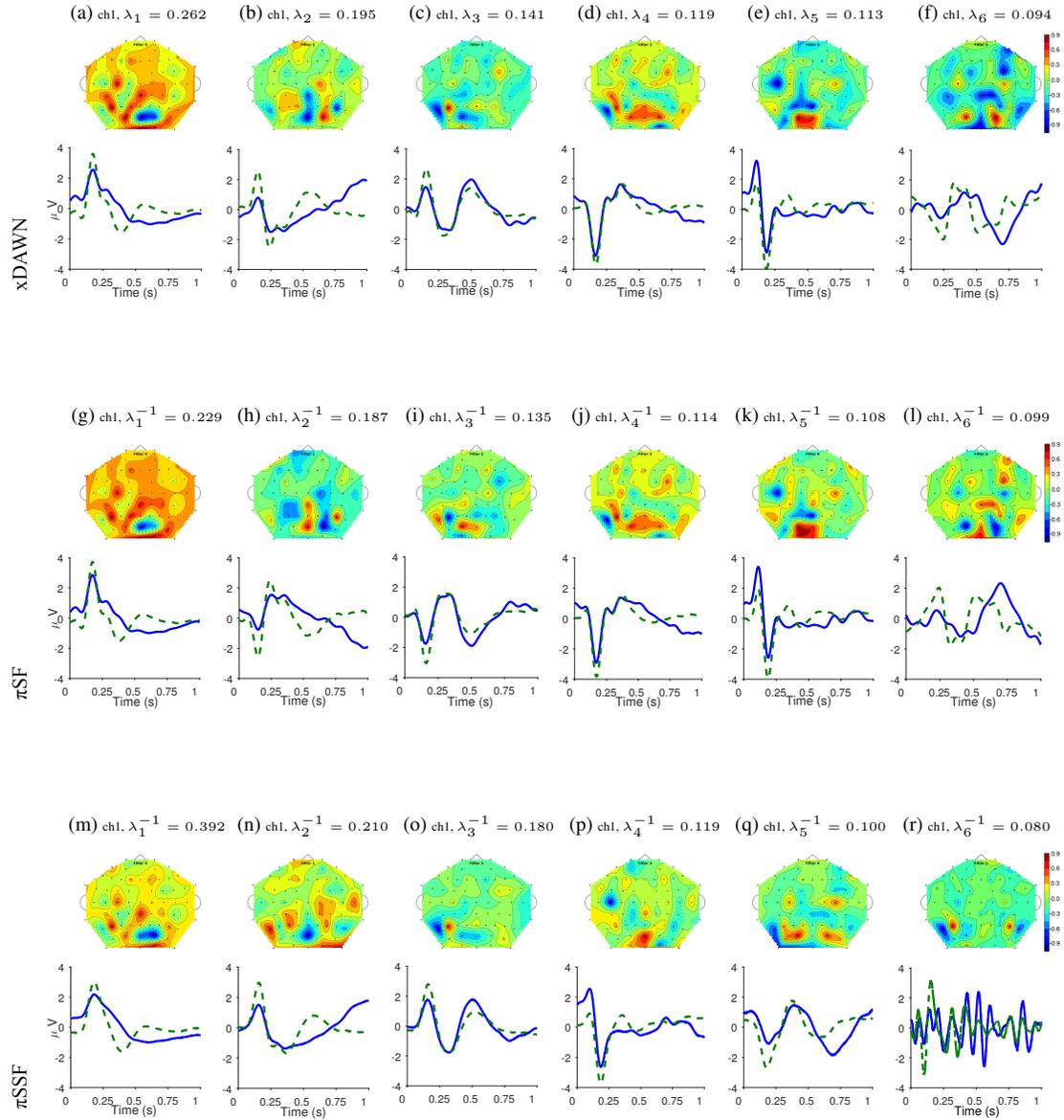


Figure 11: Grand averages of normalized targets and non-targets extracted from the first 6 spatially/spatio-spectrally filtered channels along with the corresponding spatial filter coefficients plotted on topographic maps over the scalp (solid blue lines: targets, dashed green lines: non-targets). The grand averages are normalized to unit variance and zero mean and the results of xDAWN, π SF, and π SSF algorithms are illustrated separately. The corresponding singular/eigenvalues are also listed.

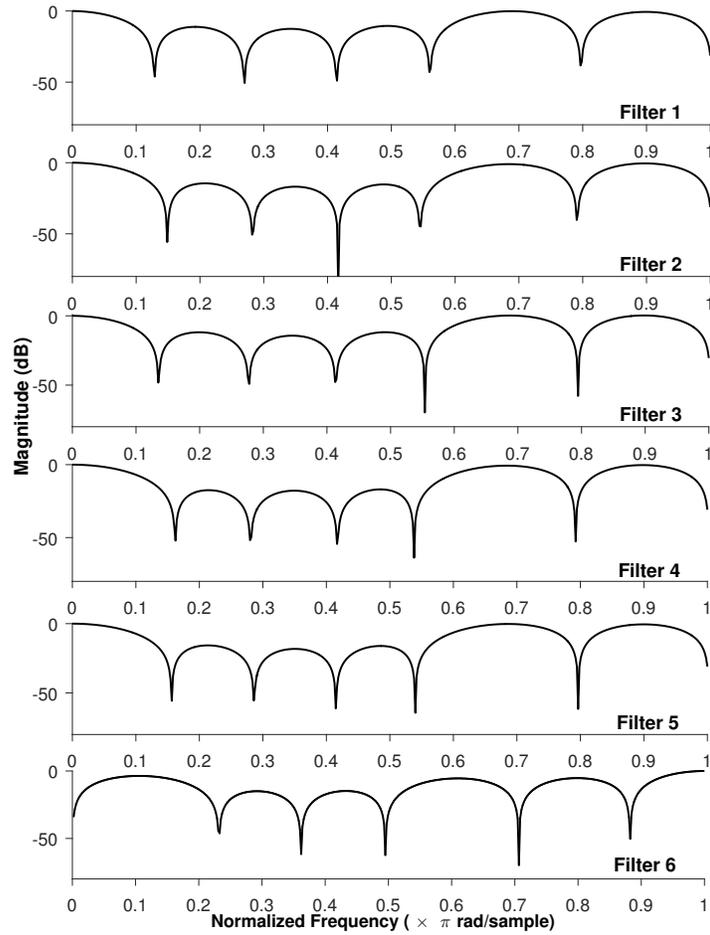


Figure 12: Frequency responses of the first 6 FIR filters estimated for the same dataset as used for preparation of the results shown in Fig. 11. All filters, except the last one, behave like low-pass filters on the EEG data that has been low-pass filtered at 45 Hz in advance. The last filter is performing as a high-pass filter. The effect of this filter can be seen in Fig. 11x. Sampling frequency was 200 Hz.

differences depending on approach type.

370 One of the objectives of this work is to show how proper selection of the cutoff frequencies of the spectral filters can affect the performance of single trial ERP classification. In the case of predesigned filters, we filtered all data at the fixed cutoff frequency of 7 Hz with an FIR filter of length 20. We selected the filter length arbitrarily, however, if the filter is not too short it will not affect the performance significantly. For the π SSF method, individual cutoffs were optimized for each spatial filter. That means, although the filters have the properties of linear transformations, 375 the design process for them is different. Designing the filters in π SSF can be seen as a data-driven approach that extracts extra information about the particular dataset at hand. This information is transparent to the predesigned filtering approach and is the key to get better performance by π SSF. Here, we used 7 Hz for low-pass filtering

the data, however, one could select a different cutoff frequency (list of different selections has been reported in Introduction section). As mentioned before, there is no general guideline for selecting the best cutoff frequency. We showed that selecting appropriate cutoffs based on the properties of particular datasets, rather than using a fixed cutoff, can improve the performance.

Results obtained from the application of the π SSF method confirm that by jointly estimating the spatial filters and the FIR filter coefficients, the classification performance can be improved by almost 9 percent compared with the case that no spatial filters are used. However, the performance is affected by two factors, i.e., the number of retained channels and the length of the FIR filters, as discussed below.

1) Retaining more channels can cause performance degradation for the π SSF method, (Fig. 8.c), which confirms the hypothesis that there are performance differences depending on the number of retained channels of the spatial filters. This phenomena is not observed for π SF (see Table 1). The spatial filters are calculated using generalized eigenvalue decomposition. In other words, the eigenvectors capture the relevant energy of the signals of interest from different directions. Therefore, retaining more channels will still add more data from new directions. This conclusion is in general true only if the eigenvectors are orthogonal. In the case of π SF, most of the contribution of relevant data can be collected from the first few channels. However, alternating between spatial filter estimation, and spectral filter estimation in step i of Algorithm 2, creates a moving target for each filter estimation step. That means, although the eigenvectors are in proper directions at the first iterations of the π SSF algorithm, they may deviate from the optimal orthogonal directions later, and hence allow more contributions of the noise subspace. This effect might be relatively small for the first few directions (which have most of the energy of the signals of interest), but may increase for the rest of the eigenvectors. Therefore, retaining more channels when using π SSF can potentially increase noise in the extracted feature vectors. This finding can also be confirmed by considering the spectrum of the *feature* \times *classifier weight* in Fig. 10c. Although the plot depicts the average over all targets in a single dataset, the data from some retained channels show high frequency components, which is a sign of using the noise subspace.

2) This problem can also be analyzed from another point of view. For the π SF and xDAWN methods, the FIR filters of length 20 were designed in advance to low-pass filter the data at some specific cutoff frequencies. However, \mathbf{h} , the FIR filter in the π SSF method, is not explicitly designed as a low-pass filter. In other words, \mathbf{h} is a convolutive transformation that minimizes the cost function (9) and would not necessarily yield a low-pass filter. From Fig. 10b it can be seen that the contribution of the signals of interest (the ERP activity) is higher in the first few output channels of the π SF method. The ERPs are by nature smooth activities, therefore, discriminating between the two classes can be achieved by low-pass filters. From the other side, since the contribution of noise increases in the rest of the channels, minimizer of the cost function might be a high-pass filter (compare Fig. 10c, Fig. 11.r, and Fig. 12). The FIR filter corresponding to the 6th spatial filter in Fig. 12 is clearly a high-pass filter

and hence the filtered data in Fig. 11x has high frequency components for both targets and non-targets. This can also mislead the classifier. Constraining the optimization problem in a way that the FIR filter coefficients represent only low-pass filters may improve the performance of π SSF by avoiding overfitting.

Conclusions

415 A novel spatial filter for single trial detection of event-related potentials is presented in this paper. The method is extended to jointly estimate spectral filter coefficients for each spatial filter and the outputs are compared with those of the xDAWN method.

The results imply that there is no significant change in performance when changing the number of retained channels. Relying on fewer retained channels is actually the main advantage of using spatial filters. As explained
420 before, spatial filters can reduce computational costs by reducing the dimensionality of the data. We showed that the first few output channels of the π SF method have the main contribution to the ERP classification task. That means, adding more channels would not affect the performance, significantly.

Considering the results presented in Table 1 and Fig. 11, one can see that besides the close performance on ERP detection, the outcome of xDAWN and π SF methods are similar with respect to the grand averages of the
425 targets/non-targets. The xDAWN method maximizes the signal-to-signal plus noise ratio (signal and noise denote respectively the ERPs, and the ongoing EEG plus noise), while the π SF algorithm maximizes the periodic structure of one ERP class compared with the other one. It turns out that both defined cost functions have similar optimum points. The main advantage of the π SF method over xDAWN is the simpler algorithm and lower computational costs. This advantage is critical when using the spatial filters in mobile BCI applications [50, 51, 52], in which
430 computational resources are limited, or online adaptive BCI applications [8, 53], in which the objective is to avoid the calibration phase and improve the performance by letting the spatial filters adapt to the upcoming samples. The π SF is simple and efficient and can be easily updated using adaptive generalized eigenvalue technique (see [53] for more details about the adaptive π SF method).

To calculate spectral filters using π SSF, we have to first calculate covariance matrices for each filter and alter-
435 nate between estimating the spatial filters and the spectral filters until convergence. This process is time consuming if repeated in each iteration, however, the filters are estimated only once in the training phase, which is usually off-line.

Noteworthy is that the main reason for selecting 45 Hz as the lowpass cutoff frequency for the π SSF method was to provide the opportunity for the algorithm to converge to a solution in a wide range between 0.1 Hz and
440 45 Hz. However, investigations about the effect of different cutoff frequencies could be the subject of future studies.

The π SF and π SSF algorithms are based on the key assumption that the ERPs of each class have similar waveforms. A challenging argument can come from the fact that in extended recording sessions with variable arousal states this assumption can be violated. However, in development of the methods it is not mandatory that
445 all the target and non-target ERPs have exactly the same waveform. It is known that the ERPs from different responses to the same stimulus are highly correlated, and the averaging method is based on this fact [1]. Therefore, high correlation of the ERP samples would be enough for the method to work.

Spatial filters are used to enhance the quality of signal of interest and hence improve the classification performance in BCI applications. We showed that joint estimation of spatial and spectral filters improves the clas-
450 sification accuracy significantly. Although the spatio-temporal filter training takes more time, all the training computation is conducted in offline mode and does not hamper the normal operation of the BCI system. The main requirement in BCI systems is higher performance, which is obtained by the proposed spatio-spectral filtering method. Application of the proposed spatial filters is not limited to BCI systems. They can be used for supervised analysis of any multi-channel recording in case that there is redundancy in data and the objective is to increase the
455 discriminability of the instances of two different classes.

Appendix

In developing the π SSF algorithm, we need to reformulate the cost function (6) in terms of the FIR filter coefficients vector, \mathbf{h} . The term $\mathbf{w}^T \mathbf{C}_U(\tau) \mathbf{w}$ can be expanded as:

$$\begin{aligned}
& \mathbf{w}^T \mathbf{C}_U(\tau) \mathbf{w} \\
&= \mathbf{w}^T \sum \mathbf{U}(t + \tau) \mathbf{U}(t)^T \mathbf{w} = \mathbf{w}^T \sum \sum_0^L h_l \mathbf{V}(t + \tau - l) \sum_0^L h_l \mathbf{V}(t - l) \mathbf{w} \\
&= \mathbf{w}^T \sum [h_0 \mathbf{V}(t + \tau) + \dots + h_l \mathbf{V}(t + \tau - l)] [h_0 \mathbf{V}(t) + \dots + h_l \mathbf{V}(t - l)]^T \mathbf{w} \\
&= \mathbf{w}^T [\sum \mathbf{V}(t + \tau) \mathbf{V}(t)^T [h_0 h_0 + h_1 h_1 + \dots + h_l h_l] + \\
&\quad \sum \mathbf{V}(t + \tau) \mathbf{V}(t - 1)^T [h_0 h_1 + h_1 h_2 + \dots + h_{l-1} h_l] + \\
&\quad \sum \mathbf{V}(t + \tau - 1) \mathbf{V}(t)^T [h_1 h_0 + h_2 h_1 + \dots + h_l h_{l-1}] + \\
&\quad \dots \\
&\quad \sum \mathbf{V}(t + \tau) \mathbf{V}(t - l - 1)^T [h_0 h_{l-1} + h_1 h_l] + \\
&\quad \sum \mathbf{V}(t + \tau - l - 1) \mathbf{V}(t)^T [h_{l-1} h_0 + h_l h_1] + \\
&\quad \sum \mathbf{V}(t + \tau) \mathbf{V}(t - l)^T h_0 h_l + \\
&\quad \sum \mathbf{V}(t + \tau - l) \mathbf{V}(t)^T h_l h_0] \mathbf{w} \\
&= h_0 h_0 \mathbf{w}^T \mathbf{C}_V(\tau) \mathbf{w} + h_1 h_0 \mathbf{w}^T \mathbf{C}_V(\tau + 1) \mathbf{w} + h_2 h_0 \mathbf{w}^T \mathbf{C}_V(\tau + 2) \mathbf{w} + \\
&\quad \dots + h_{l-1} h_0 \mathbf{w}^T \mathbf{C}_V(\tau + l - 1) \mathbf{w} + h_0 h_1 \mathbf{w}^T \mathbf{C}_V(\tau - 1) \mathbf{w} + \\
&\quad h_1 h_1 \mathbf{w}^T \mathbf{C}_V(\tau) \mathbf{w} + h_2 h_1 \mathbf{w}^T \mathbf{C}_V(\tau + 1) \mathbf{w} + \\
&\quad \dots + h_{l-1} h_1 \mathbf{w}^T \mathbf{C}_V(\tau + l - 2) \mathbf{w} + h_0 h_2 \mathbf{w}^T \mathbf{C}_V(\tau - 2) \mathbf{w} + \\
&\quad h_1 h_2 \mathbf{w}^T \mathbf{C}_V(\tau - 1) \mathbf{w} + h_2 h_2 \mathbf{w}^T \mathbf{C}_V(\tau) \mathbf{w} + \\
&\quad \dots + h_{l-1} h_2 \mathbf{w}^T \mathbf{C}_V(\tau + l - 3) \mathbf{w} + \dots \\
&\quad h_0 h_{l-1} \mathbf{w}^T \mathbf{C}_V(\tau - l + 1) \mathbf{w} + h_1 h_{l-1} \mathbf{w}^T \mathbf{C}_V(\tau - l + 2) \mathbf{w} + \\
&\quad \dots + h_{l-1} h_{l-1} \mathbf{w}^T \mathbf{C}_V(\tau) \mathbf{w}
\end{aligned}$$

Factorizing h_i s and simplifying the equation, the following will be resulted:

$$\mathbf{w}^T \mathbf{C}_U(\tau) \mathbf{w} = \mathbf{h}^T \mathbf{B}_V(\tau) \mathbf{h}$$

where

$$\mathbf{B}_V(\tau) = \begin{bmatrix} \mathbf{w}^T \mathbf{C}_V(\tau) \mathbf{w} & \mathbf{w}^T \mathbf{C}_V(\tau+1) \mathbf{w} & \cdots & \mathbf{w}^T \mathbf{C}_V(\tau+l-1) \mathbf{w} \\ \mathbf{w}^T \mathbf{C}_V(\tau-1) \mathbf{w} & \mathbf{w}^T \mathbf{C}_V(\tau) \mathbf{w} & \cdots & \mathbf{w}^T \mathbf{C}_V(\tau+l-2) \mathbf{w} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{w}^T \mathbf{C}_V(\tau-l+1) \mathbf{w} & \mathbf{w}^T \mathbf{C}_V(\tau-l+2) \mathbf{w} & \cdots & \mathbf{w}^T \mathbf{C}_V(\tau) \mathbf{w} \end{bmatrix},$$

or in a more compact form $\mathbf{B}_V(\tau)_{i,j} = \mathbf{w}^T \mathbf{C}_V(\tau-i+j) \mathbf{w}$. For simplicity, the divisors in the definition of the sample covariance matrices are removed.

460 Acknowledgment

The authors are thankful to Dr. S. Straube from Bremen University and Dr. S. Sanei from University of Surrey for fruitful discussions and their comments on this paper. This work was supported by the German Bundesministerium für Wirtschaft und Technologie (BMWi, grant FKZ 50 RA 1012 and grant FKZ 50 RA 1011).

References

- 465 [1] S. Sanei, Adaptive Processing of Brain Signals, Wiley, 2013.
- [2] F. Ghaderi, K. Nazarpour, J. McWhirter, S. Sanei, Removal of ballistocardiogram artifacts using the cyclostationary source extraction method, *IEEE Trans Biomed Eng* 57 (11) (2010) 2667–2676.
- [3] A. R. Marathe, A. J. Ries, V. J. Lawhern, B. J. Lance, J. Touryan, K. McDowell, H. Cecotti, The effect of target and non-target similarity on neural classification performance: A boost from confidence, *Front Neurosci* 9 (270) (2015) 1–11.
- 470 [4] J. Polich, Updating P300: An integrative theory of P3a and P3b, *Clin Neurophysiol* 118 (10) (2007) 2128–2148.
- [5] J. Farquhar, N. Hill, Interactions between pre-processing and classification methods for event-related-potential classification, *Neuroinformatics* 11 (2) (2013) 175–192.
- 475 [6] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, T. M. Vaughan, Brain-computer interfaces for communication and control, *Clin Neurophysiol* 113 (6) (2002) 767–791.
- [7] E. A. Kirchner, S. K. Kim, M. Tabie, H. Woehrle, M. Maurus, F. Kirchner, An intelligent man-machine interface - multi-robot control adapted for task engagement based on single-trial detectability of P300, *Front Hum Neurosci* 10 (291) (2016) 1–21.

- 480 [8] H. Woehrle, E. A. Kirchner, Online classifier adaptation for the detection of P300 target recognition processes in a complex teleoperation scenario, in: H. P. da Silva, A. Holzinger, S. Fairclough, D. Majoe (Eds.), In International Conference on Physiological Computing Systems, Springer Berlin Heidelberg, 2014, pp. 105–118.
- [9] V. Mäkinen, H. Tiitinen, P. May, Auditory event-related responses are generated independently of ongoing
485 brain activity, *NeuroImage* 24 (4) (2005) 961 – 968.
- [10] P. Sauseng, W. Klimesch, W. Gruber, S. Hanslmayr, R. Freunberger, M. Doppelmayr, Are event-related potential components generated by phase resetting of brain oscillations? a critical discussion, *Neuroscience* 146 (4) (2007) 1435 – 1444.
- [11] M. Xu, Y. Jia, H. Qi, Y. Hu, F. He, X. Zhao, P. Zhou, L. Zhang, B. Wan, W. Gao, D. Ming, Use of a steady-
490 state baseline to address evoked vs. oscillation models of visual evoked potential origin, *NeuroImage* 134 (2016) 204 – 212.
- [12] J. Fell, T. Dietl, T. Grunwald, M. Kurthen, P. Klaver, P. Trautner, C. Schaller, C. E. Elger, G. Fernández, Neural bases of cognitive ERPs: more than phase reset., *J Cogn Neurosci* 16 (9) (2004) 1595–1604.
- [13] A. Bashashati, M. Fatourehchi, R. K. Ward, G. E. Birch, A survey of signal processing algorithms in brain-
495 computer interfaces based on electrical brain signals, *J Neural Eng* 4 (2) (2007) 32–57.
- [14] Z. J. Koles, M. S. Lazar, S. Z. Zhou, Spatial patterns underlying population differences in the background EEG, *Brain Topogr* 2 (4) (1990) 275–284.
- [15] H. Ramoser, J. Muller-Gerking, G. Pfurtscheller, Optimal spatial filtering of single trial EEG during imagined hand movement, *IEEE Trans Rehabil Eng.* 8 (4) (2000) 441–446.
- 500 [16] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, K.-R. Muller, Optimizing spatial filters for robust EEG single-trial analysis, *IEEE Signal Process. Mag.* 25 (1) (2008) 41–56.
- [17] U. Hoffmann, J.-M. Vesin, T. Ebrahimi, Spatial filters for the classification of event-related potentials, in: Proc the 14th Europ Symp Artificial Neural Networks (ESANN), 2006, pp. 26–28.
- [18] B. Rivet, A. Souloumiac, V. Attina, G. Gibert, xDAWN algorithm to enhance evoked potentials: Application
505 to brain-computer interface, *IEEE Trans. Biomed. Eng.* 56 (8) (2009) 2035–2043.
- [19] E. Kristensen, A. Guerin-Dugué, B. Rivet, Comparison between adjar and xDawn algorithms to estimate Eye-Fixation Related Potentials distorted by overlapping, in: 7th International IEEE/EMBS Conference on Neural Engineering (NER), 2015, pp. 976–979.

- [20] A. Onishi, A. H. Phan, K. Matsuoka, A. Cichocki, Tensor classification for P300-based brain computer interface, in: Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP), 2012, pp. 581–584.
510
- [21] M. van Vliet, N. Chumerin, S. De Deyne, J. Wiersema, W. Fias, G. Storms, M. Van Hulle, Single-trial ERP component analysis using a spatiotemporal LCMV beamformer, *IEEE Trans. Biomed. Eng.* 63 (1) (2016) 55–66.
- [22] F. Ghaderi, E. Kirchner, Periodic spatial filter for single trial classification of event related brain activity, in: Proc. 10th Int. Conf. Biomed. Eng. (BioMed), 2013.
515
- [23] F. Ghaderi, Joint spatial and spectral filter estimation for single trial detection of event related potentials, in: IEEE International Workshop on Machine Learning for Signal Processing, (MLSP), 2013.
- [24] F. Biessmann, S. Daehne, F. C. Meinecke, B. Blankertz, K. Goergen, K. R. Mueller, S. Haufe, On the Interpretability of Linear Multivariate Neuroimaging Analyses: Filters, Patterns and their Relationship, in: NIPS workshop on Machine Learning and Inference in Neuroimaging 2012, 2012.
520
- [25] F. Ghaderi, S. K. Kim, E. A. Kirchner, Effects of eye artifact removal methods on single trial P300 detection, a comparative study, *J Neurosci Methods* 221 (2014) 41–47.
- [26] B. Jansen, A. Allam, P. Kota, K. Lachance, A. Osho, K. Sundaresan, An exploratory study of factors affecting single trial P300 detection, *IEEE Trans. Biomed. Eng.* 51 (6) (2004) 975–978.
- [27] V. Kolev, T. Demiralp, J. Yordanova, A. Ademoglu, Ü. Isoglu-Alka, Time-frequency analysis reveals multiple functional components during oddball P300, *NeuroReport* 8 (8) (1997) 2061–2065.
525
- [28] J. Jin, B. Z. Allison, T. Kaufmann, A. Kübler, Y. Zhang, X. Wang, A. Cichocki, The changing face of P300 BCIs: A comparison of stimulus changes in a P300 BCI involving faces, emotion, and movement, *PLoS ONE* 7 (11) (2012) 1–10.
- [29] T. Kaufmann, E. Hammer, A. Kübler, ERPs contributing to classification in the P300 BCI, in: 5th Int BCI Conf 2011, 2011.
530
- [30] Q. Zhao, L. Zhang, J. Cao, A. Cichocki, Higher-order PLS for classification of ERPs with application to BCIs, in: Signal Information Processing Association Annual Summit and Conference (APSIPA ASC), 2012 Asia-Pacific, 2012, pp. 1–5.
- [31] F. Akram, H. S. Han, T. S. Kim, A P300-based brain computer interface system for words typing, *Computers in Biology and Medicine* 45 (2014) 118–125.
535

- [32] E. W. Sellers, D. J. Krusienski, D. J. McFarland, T. M. Vaughan, J. R. Wolpaw, A P300 event-related potential brain-computer interface (BCI): The effects of matrix size and inter stimulus interval on performance, *Biological Psychology* 73 (3) (2006) 242–252.
- 540 [33] R. Panicker, S. Puthusserypady, Y. Sun, An asynchronous P300 BCI with SSVEP-based control state detection, *IEEE Trans Biomed Eng* 58 (6) (2011) 1781–1788.
- [34] A.-M. Brouwer, J. B. van Erp, A tactile P300 brain-computer interface, *Front Neurosci* 4 (2010) 1–11.
- [35] R. Roy, S. Bonnet, S. Charbonnier, P. Jallon, A. Campagne, A comparison of ERP spatial filtering methods for optimal mental workload estimation, in: 37th Conf Proc IEEE Eng Med Biol Soc (EMBC), 2015, pp. 7254–7257.
- 545 [36] S. Lemm, B. Blankertz, G. Curio, K. Muller, Spatio-spectral filters for improving the classification of single trial EEG, *IEEE Trans Biomed Eng* 52 (9) (2005) 1541–1548.
- [37] K. K. Ang, Z. Y. Chin, H. Zhang, C. Guan, Filter Bank Common Spatial Pattern (FBCSP) in Brain-Computer Interface, in: *Neural Networks, 2008. IJCNN 2008. (IEEE World Congress on Computational Intelligence)*.
550 *IEEE International Joint Conference on, 2008*, pp. 2390–2397.
- [38] Q. Novi, C. Guan, T. H. Dat, P. Xue, Sub-band Common Spatial Pattern (SBCSP) for Brain-Computer Interface, in: *3rd International IEEE/EMBS Conference on Neural Engineering, (CNE), 2007*, pp. 204–207.
- [39] H. Zhang, Z. Y. Chin, K. K. Ang, C. Guan, C. Wang, Optimum spatio-spectral filtering network for brain-computer interface, *IEEE Trans Neural Netw* 22 (1) (2011) 52–63.
- 555 [40] H. Suk, S.-W. Lee, A probabilistic approach to spatio-spectral filters optimization in brain-computer interface, in: *IEEE Int Conf Systems, Man, and Cybernetics (SMC), 2011*, pp. 19–24.
- [41] M. Mahanta, A. Aghaei, K. Plataniotis, A bayes optimal matrix-variate LDA for extraction of spatio-spectral features from EEG signals, in: *Annual Int Conf IEEE Engineering in Medicine and Biology Society (EMBC), 2012*, pp. 3955–3958.
- 560 [42] K. Yu, Y. Wang, K. Shen, X. Li, The synergy between complex channel-specific FIR filter and spatial filter for single-trial EEG classification, *PLoS ONE* 8 (10) (2013) e76923.
- [43] B. Rivet, A. Souloumiac, Optimal linear spatial filters for event-related potentials based on a spatio-temporal model: Asymptotical performance analysis, *Signal Processing* 93 (2) (2013) 387–398.

- [44] J. Lu, K. Xie, D. McFarland, Adaptive spatio-temporal filtering for movement related potentials in EEG-based brain computer interfaces, *Neural Systems and Rehabilitation Engineering, IEEE Transactions on* 22 (4) (2014) 847–857.
- [45] L. K. Saul, J. B. Allen, Periodic component analysis: An eigenvalue method for representing periodic structure in speech, in: *NIPS'00, 2000*, pp. 807–813.
- [46] B. N. Parlett, *The symmetric eigenvalue problem*, Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1998.
- [47] E. A. Kirchner, S. K. Kim, S. Straube, A. Seeland, H. Wöhrle, M. M. Krell, M. Tabie, M. Fahle, On the applicability of brain reading for predictive human-machine interfaces in robotics, *PLoS ONE* 8 (12) (2013) e81732.
- [48] J. H. Metzen, E. A. Kirchner, Rapid adaptation of brain reading interfaces based on threshold adjustment, in: *Proceedings of the 2011 Conference of the German Classification Society, 2011*, pp. 138–138.
- [49] M. M. Krell, S. Straube, A. Seeland, H. Wöhrle, J. Teiwes, J. H. Metzen, E. A. Kirchner, F. Kirchner, pySPACE - a signal processing and classification environment in Python, *Front Neuroinform* 7 (40) (2013) 1–11.
- [50] S. Debener, R. Emkes, M. De Vos, M. Bleichner, Unobtrusive ambulatory EEG using a smartphone and flexible printed electrodes around the ear, *Scientific reports* 5 (2015) 1–11.
- [51] Q. T. Obeidat, T. A. Campbell, J. Kong, Introducing the edges paradigm: A P300 brain–computer interface for spelling written words, *IEEE Trans Hum Mach Syst* 45 (6) (2015) 727–738.
- [52] Q. Obeidat, T. Campbell, J. Kong, The zigzag paradigm: a new P300-based brain computer interface, in: *Proceedings of the 15th ACM on International conference on multimodal interaction, ACM, 2013*, pp. 205–212.
- [53] F. Ghaderi, S. Straube, An adaptive and efficient spatial filter for event-related potentials, in: *Proc. Europ Sig Proc Conf (EUSIPCO), 2013*.