AN ADAPTIVE AND EFFICIENT SPATIAL FILTER FOR EVENT-RELATED POTENTIALS

Foad Ghaderi, Sirko Straube

Robotics Group, University of Bremen, Bremen, Germany

ABSTRACT

A major problem in designing brain computer interface (BCI) systems is the variations of data during long sessions, and also from session to session and subject to subject. This demands extensive training sessions to maintain the overall performance of the systems. As an alternative, here we propose an adaptive version of the πSF method. Taking advantage of simple structure, high performance, and low computational complexity of the algorithm, an extension is developed that can adapt the trained transformation to the new instances. Experimental results confirm that the adaptive method ($a\pi SF$) outperforms the non-adaptive versions of πSF and the xDAWN method in an inter-subject train-test schema.

Index Terms— adaptivity, event-related potentials, periodic spatial filter, single-trial ERP detection.

1. INTRODUCTION

The analysis of event-related potentials (ERPs) in the electroencephalogram (EEG) recordings has a long standing history in the study of human perceptual, cognitive and motor functions. The classical method for eliminating noise in ERPs is to average EEG windows of experimental repetitions. However, there are lots of applications, e.g., BCI systems, which rely on single trial detection of ERPs [1, 2].

Spatial filters are a kind of pre-processing methods that transform EEG data to a low-dimensional subspace with high signal-to-noise ratios. Based on the requirements of different applications, specific conditions are optimized by different spatial filters. The common spatial patterns (CSP), e.g., maximizes the temporal variance of one class while minimizing the temporal variance of the other [3]. The CSP method has been successfully used to improve classification performance in motor imagery tasks [4, 5].

Other spatial filters are specifically designed for ERPs, e.g., the methods in [6–9]. The proposed method in [6] maximizes the sum of the squared distances of the components of the mean feature vectors. The xDAWN method [7], maximizes the signal (ERPs) to signal plus noise (ongoing EEG, artifacts and measurement interferences) ratio. Comparative studies reported in [7, 10], confirm the outstanding performance of the method. In [8], tensor decomposition is used to improve the features and reduce dimensions of the data. Recently, periodic spatial filter (πSF) method is proposed in [9]. The method is simple and efficient and is based on enhancing the periodic structure of the data created by concatenating ERP windows of the same type.

A big challenge in developing BCI systems is to reduce the sensitivity of the system to the non-stationarities of the data. ERP waveforms may change from subject to subject, or even from session to session for the same subject. Other parameters e.g., fatigue and electrode resistance changes can also alter the properties of the signals of interest. Compensating the non-stationarities in the EEG data is an essential step towards a robust, high performance BCI system. Furthermore, using such an adapting system would be more comfortable for subjects, since the system can be trained incrementally and long training sessions are avoided. This challenge has been tackled by adapting different modules of BCI systems. Researchers have used adaptive versions of classifiers, e.g., LDA [11–13], QDA [12, 14], SVM [15, 16] to improve the performance of the BCI systems. The problem of unsupervised adaptation of the classifiers in BCI systems has been addressed in [17–19]. As an alternative adaptation approach, [20–22] focus on adapting the CSP method during motor imagery tasks.

In this paper, we propose a variation of the πSF method, the adaptive πSF method ($a\pi SF$), which is designed for ERPs and has the key advantage of having high performance and at the same time being computationally efficient. The adaptation is performed when the true class label is known in the application (usually with some delay). It follows that for each new ERP instance, the previously trained πSF transformation is used and upon arrival of the true label the transformation matrix is updated and used for forthcoming instances.

In spite of lower computational complexity of πSF, experimental results presented here show that the performance of this method is promising. Further we show that $a\pi$SF outperforms the non-adaptive version in a spatial filter transfer scenario, in which the trained model is tested using data from a novel subject.

The paper is structured as follows. In the next section, first a brief overview of adaptive generalized eigenvalue decom-

This work was supported by the German Bundesministerium für Wirtschaft und Technologie (BMWi, grant FKZ 50 RA 1012 and grant FKZ 50 RA 1011).

position and πSF method is presented and then the proposed adaptive spatial filter is addressed. Data description, experiments, and the results are introduced in section 3. Concluding discussions are presented in the last section.

2. PROBLEM FORMULATION

The proposed method is formulated by iteratively estimating the minimizer of a cost function using generalized eigenvalue decomposition (GEVD) of two square matrices. GEVD, as described in sequel, is a powerful statistical tool used in different science and engineering fields.

2.1. Adaptive generalized eigenvalue decomposition

GEVD of two square matrices $\mathbf{A}$ and $\mathbf{B}$ consists of solving the equation $\mathbf{AW} = \mathbf{BW\Lambda}$, where $\mathbf{W}$ and the diagonal matrix $\mathbf{\Lambda}$ are respectively called generalized eigenvector and generalized eigenvalue matrices [23]. Solution to GEVD problem can simultaneously diagonalize the matrices $\mathbf{A}$ and $\mathbf{B}$, i.e., $\mathbf{W}^T\mathbf{AW} = \mathbf{\Lambda}$ and $\mathbf{W}^T\mathbf{BW} = \mathbf{I}$, where $[.]^T$ denotes transpose operator. If $\mathbf{A}$ and $\mathbf{B}$ are symmetric and positive definite, the generalized eigenvalues are real and positive [23].

Online solution of the GEVD problem is an essential requirement in applications that true estimations of $\mathbf{A}$ and $\mathbf{B}$ are not available in advance or they may change over time.

In [24] the solution is proposed as follows. Assume that $\mathbf{w}(k-1)$ is the estimate of an eigenvector at time index $k-1$. Then $\mathbf{w}(k)$ is estimated as:

$$\mathbf{w}(k) = \frac{\mathbf{w}(k-1)\mathbf{B}(k)\mathbf{w}(k-1)}{\mathbf{w}(k-1)\mathbf{A}(k)\mathbf{w}(k-1)}\mathbf{B}^{-1}(k-1)\mathbf{A}(k)\mathbf{w}(k-1) \tag{1}$$

This algorithm only updates the principal generalized eigenvector, $\mathbf{w}_1(k)$. In order to update the other components, one has to deflate $\mathbf{w}_1(k)$ to obtain $\hat{\mathbf{A}}(k)$ and $\hat{\mathbf{B}}(k)$ as

$$\hat{\mathbf{A}}(k) = \left[\mathbf{I} - \frac{\mathbf{A}(k)\mathbf{w}_1(k)\mathbf{w}_1^T(k)}{\mathbf{w}_1^T(k)\mathbf{A}(k)\mathbf{w}_1(k)}\right]\mathbf{A}(k),\ \hat{\mathbf{B}}(k) = \mathbf{B}(k) \tag{2}$$

It can be shown that $\hat{\mathbf{A}}(k)\mathbf{w}_1(k) = 0$ and $\hat{\mathbf{A}}(k)\mathbf{w}_i(k) = \lambda_i\mathbf{B}(k)\mathbf{w}_i(k), i > 1$. Therefore, to update $\mathbf{w}_i(k)$ first $\mathbf{w}_{i-1}(k)$ has to be deflated using (2) to obtain new $\hat{\mathbf{A}}(k)$, then $\mathbf{A}(k)$ in (1) must be substituted with $\hat{\mathbf{A}}(k)$.

2.2. Periodic spatial filter

It is assumed that the EEG data have been recorded using n electrodes at m sample points and each ERP response lasts for m_e samples, the system can be formulated as follows:

$$\mathbf{x}(t) = \mathbf{Ed}(t) + \mathbf{r}(t) \tag{3}$$

where $\mathbf{x}(t)$ is the n dimensional measurement vector, $\mathbf{E}$ is the $n \times m_e$ ERP response, $\mathbf{r}(t)$ is the n dimensional noise vector (including ongoing EEG), and $\mathbf{d}(t)$ is an m_e dimensional vector indicating the time point of the relevant ERPs in the EEG recording. If t_l is the onset of the lth ERP stimulus and $t_l \leq t < t_l + m_e$, then

$$d_j(t) = \begin{cases} 1 & \text{if } j = t - t_l + 1 \\ 0 & \text{else} \end{cases}. \tag{4}$$

Constructing the matrix $\mathbf{D} = [\mathbf{d}(1), \mathbf{d}(2), ..., \mathbf{d}(m)]$ for each ERP class separately, the objective is to estimate ERP response matrices, i.e. $\mathbf{E}$ for each class. It is worth to mention that construction of $\mathbf{D}$ is always possible because the ERPs are time locked to the stimuli and the stimulus markers are stored with EEG data. Assuming $\mathbf{X} = [\mathbf{x}(1), \mathbf{x}(2), ..., \mathbf{x}(m)]$, the least square solution to the problem could be estimated from $\hat{\mathbf{E}} = \arg\min_E ||\mathbf{X} - \mathbf{ED}||^2$ which is

$$\hat{\mathbf{E}} = \mathbf{XD}^T(\mathbf{DD}^T)^{-1}. \tag{5}$$

Equation (5) always has a solution, because $\mathbf{D}$ is a $m_e \times m$ Toeplitz matrix and hence $\mathbf{DD}^T$ is invertible. If we construct $\mathbf{D}$ for one class and $\mathbf{D}'$ for the other, and (using (5)) estimate the ERP matrices, i.e., $\hat{\mathbf{E}}$ and $\hat{\mathbf{E}}'$, matrices $\hat{\mathbf{X}} = \hat{\mathbf{E}}\mathbf{D}$ and $\hat{\mathbf{X}}' = \hat{\mathbf{E}}'\mathbf{D}'$ will contain the enhanced versions of brain responses.

Assume that there are p instances of one ERP class ($\hat{\mathbf{y}}_{T_i}$, for $1 \leq i \leq p$) in $\hat{\mathbf{X}}$ and q instances of the other class ($\hat{\mathbf{y}}_{S_j}$, for $1 \leq j \leq q$) in $\hat{\mathbf{X}}'$. We construct the matrices $\mathbf{y}_T$ and $\mathbf{y}_S$ by horizontally concatenating all instances of $\hat{\mathbf{y}}_{T_i}$ and $\hat{\mathbf{y}}_{S_j}$, respectively. Ideally each row of $\mathbf{y}_T$ and $\mathbf{y}_S$ is a periodic signal with the period equal to the ERP window length, $\tau = m_e$.

The periodic spatial filter (πSF), proposed in [9], is based on enhancing the periodic structure of the data by minimizing $\mathcal{J}(\mathbf{w}, \tau) = \frac{\sum_{t'} |\mathbf{w}^T\mathbf{y}_T(t'+\tau) - \mathbf{w}^T\mathbf{y}_T(t')|^2}{\sum_{t'} |\mathbf{w}^T\mathbf{y}_S(t'+\tau) - \mathbf{w}^T\mathbf{y}_S(t')|^2}$, where $\mathbf{w}$ is the spatial filter vector. This function is further simplified to $\mathcal{J}(\mathbf{w}, \tau) = \frac{\mathbf{w}^T\mathbf{A}_{\mathbf{y}_T}(\tau)\mathbf{w}}{\mathbf{w}^T\mathbf{A}_{\mathbf{y}_S}(\tau)\mathbf{w}}$, where $\mathbf{A}_\mathbf{u}(\tau) = E\{[\mathbf{u}(t'+\tau) - \mathbf{u}(t')][\mathbf{u}(t'+\tau) - \mathbf{u}(t')]^T\} = 2\mathbf{C}_\mathbf{u}(0) - (\mathbf{C}_\mathbf{u}(\tau) + \mathbf{C}_\mathbf{u}(-\tau))$, $E\{.\}$ denotes the expected value of the enclosed term, and $\mathbf{C}_\mathbf{u}(\tau) = E\{\mathbf{u}(t'+\tau)\mathbf{u}(t')^T\}$ is the covariance matrix of $\mathbf{u} \in \{\mathbf{y}_T, \mathbf{y}_S\}$. The minimizer of $\mathcal{J}(\mathbf{w}, \tau)$ is given by the eigenvector corresponding to the smallest generalized eigenvalue of $\mathbf{A}_{\mathbf{y}_T}(\tau)$ and $\mathbf{A}_{\mathbf{y}_S}(\tau)$.

The steps of the πSF method are summarized as follows (for more details about the method see [9]):

1. Using (4), construct matrices $\mathbf{D}$ and $\mathbf{D}'$ for the two classes, respectively,
2. Estimate $\hat{\mathbf{X}}$ and $\hat{\mathbf{X}}'$ using (5): $\hat{\mathbf{X}} = \hat{\mathbf{E}}\mathbf{D} = \mathbf{XD}^T(\mathbf{DD}^T)^{-1}\mathbf{D}$ and $\hat{\mathbf{X}}' = \hat{\mathbf{E}}'\mathbf{D}' = \mathbf{XD}'^T(\mathbf{D}'\mathbf{D}'^T)^{-1}\mathbf{D}'$,
3. Construct $\mathbf{y}_T$ and $\mathbf{y}_S$ by concatenating ERP instances from $\hat{\mathbf{X}}$ and $\hat{\mathbf{X}}'$, respectively,
4. Calculate $\mathbf{A}_{\mathbf{y}_T}(\tau)$ and $\mathbf{A}_{\mathbf{y}_S}(\tau)$,
5. Calculate GEVD of $\mathbf{A}_{\mathbf{y}_T}(\tau)$ and $\mathbf{A}_{\mathbf{y}_S}(\tau)$,
6. The first n_f eigenvectors corresponding to the smallest eigenvalues are selected as the spatial filters.

2.3. Adaptive periodic spatial filter

The idea behind adaptive spatial filters is to compensate the data shifts in BCI applications. A simple case is when the system is trained for a subject but the brain response templates vary over long application intervals or in future sessions. A more complicated case is when the system is trained on a pool of data from different subjects, and the goal is to use the system for a new subject without further training. An adaptive system must be able to extract the data properties and update the trained parameters.

Data shifts in the EEG data would affect the matrices $\mathbf{A_u}$, $\mathbf{u} \in \{\mathbf{y}_T, \mathbf{y}_S\}$. To compensate the shift effects and develop $a\pi$SF, the spatial filer vectors have to be modified using the test instances and the update rule of (1).

Assume that the true class label of $\hat{\mathbf{u}}$, the kth ERP test instance, is availabe with some delay and $\mathbf{u}$(k-1) represents either $\mathbf{y}_T$ or $\mathbf{y}_S$ at the $(k-1)$th adaptation step. That means, depending on the class label of $\hat{\mathbf{u}}$, $\mathbf{u}$(k) will be either $\mathbf{u}(k) = [\mathbf{u}(k-1)\ \hat{\mathbf{u}}]$ or $\mathbf{u}(k) = [\mathbf{u}(k-1)]$.

In the former case, $\mathbf{C}_{\mathbf{u}(k)}(\tau)$ can be written as $\mathbf{C}_{\mathbf{u}(k)}(\tau) = \frac{1}{M+p}[\mathbf{0}\ \mathbf{u}(k-1)\ \hat{\mathbf{u}}][\mathbf{u}(k-1)\ \hat{\mathbf{u}}\ \mathbf{0}]^T$, where M and τ are the lengths of $\mathbf{u}(k-1)$ and $\hat{\mathbf{u}}$, respectively, and $\mathbf{0}$ is a matrix of the same length of $\hat{\mathbf{u}}$ with all zero entries.

We further assume that $\mathbf{u}(k-1) = [\mathbf{u}_0(k-1)\ \mathbf{u}_1(k-1)\ \mathbf{u}_2(k-1)]$ where $\mathbf{u}_0(k-1)$ and $\mathbf{u}_2(k-1)$, are respectively the first and the last ERP instances in $\mathbf{u}(k-1)$. $\mathbf{C}_{\mathbf{u}(k)}(\tau)$ is then simplified as:

$$
\begin{aligned}
\mathbf{C}_{\mathbf{u}(k)}(\tau) =& \frac{1}{M+\tau}[\mathbf{u}_0(k-1)\mathbf{u}_1(k-1)^T + \\
& \mathbf{u}_1(k-1)\mathbf{u}_2(k-1)^T + \mathbf{u}_2(k-1)\hat{\mathbf{u}}^T] \\
=& \frac{1}{M+\tau}[M\mathbf{C}_{\mathbf{u}(k-1)}(\tau) + \mathbf{u}_2(k-1)\hat{\mathbf{u}}^T] \\
=& \frac{M}{M+\tau}\mathbf{C}_{\mathbf{u}(k-1)}(\tau) + \frac{\tau}{M+\tau}\frac{\mathbf{u}_2(k-1)\hat{\mathbf{u}}^T}{\tau} \\
=& (1-\alpha)\mathbf{C}_{\mathbf{u}(k-1)}(\tau) + \alpha\frac{\mathbf{u}_2(k-1)\hat{\mathbf{u}}^T}{\tau}
\end{aligned}
\quad (6)
$$

where $0 < \alpha = \frac{\tau}{M+\tau} < 1$ is the update coefficient. Considering the values of M and τ, the typical values of α have to be small. Similarly one can conclude that

$$
\mathbf{C}_{\mathbf{u}(k)}(0) = (1-\alpha)\mathbf{C}_{\mathbf{u}(k-1)}(0) + \alpha\frac{\hat{\mathbf{u}}\hat{\mathbf{u}}^T}{\tau} \quad (7)
$$

To be able to initialize the adaptive algorithm properly we need to know $\mathbf{A}_{\mathbf{y}_T}$, $\mathbf{A}_{\mathbf{y}_S}$ from the training phase. Since both the matrices are $n \times n$ matrices, it is always possible to save them along with the trained spatial fitlers. On the other hand, $\mathbf{u}_2(k-1)$ is also required to update $\mathbf{C}_{\mathbf{u}(k)}(\tau)$. It is possible to randomly set the value of this matrix for the first iteration of the adaptation. Furthermore, heuristically selecting the proper values for update coefficients α can resolve the need of prior knowledge about M.

The proposed adaptive periodic spatial filter method is summarized as follows:

1. Initialize α with small values (close to zero),
2. Randomly initialize $\mathbf{u}_2(k-1)$ for each ERP type,
3. For each new sample, $\hat{\mathbf{u}}$:
   (a) Based on the class label of $\hat{\mathbf{u}}$, update $\mathbf{A}_{\mathbf{y}_T}(\tau)$ or $\mathbf{A}_{\mathbf{y}_S}(\tau)$ using (6) and (7),
   (b) Set $\mathbf{A} = \mathbf{A}_{\mathbf{y}_T}$, $\mathbf{B} = \mathbf{A}_{\mathbf{y}_S}$. Update $\mathbf{w}_i$ using (1),
   (c) Deflate $\mathbf{w}_i$ using (2),
   (d) Normalize $\mathbf{w}_i$ and repeat from (b) $\forall i$, $i \le n_f$,
   (e) Substitute $\mathbf{u}_2(k-1)$ with $\hat{\mathbf{u}}$.

In order to avoid computational costs of calculating $\mathbf{A}_{\mathbf{y}_S}^{-1}$ in (1), one can use Sheman-Morrison-Woodbury matrix inversion lemma as described in [24].

3. EXPERIMENTS

Performances reported here were evaluated by offline classification of data containing an oddball scenario, i.e. the ERP classes can be divided into an infrequent target and frequent non-target conditions. The experiments were conducted in an inter-subject schema. Experimental setup and recording are described in detail elsewhere [9, 25].

3.1. Data & Classification

The data were recorded from six male subjects (64 electrodes, 10-20 system) participating in two experimental sessions with five runs each. The data from one subject had to be excluded from the analysis due to high noise levels. Each dataset of a single run contains 720 non-target and 120 target instances. The recorded signals were further band-pass filtered between 0.1 Hz and 7 Hz and down-sampled to 100 Hz. Each ERP instance was cut between 0 to 1000 ms after stimulus onset.

In this study, ERP instances were classified using a linear discriminant analysis (LDA) and performance was determined by training spatial filter and LDA on all datasets from 4 subjects and testing on the 10 datasets of the remaining subject, while adapting the spatial filter. Feature vectors were generated from the slopes of lines fitted to overlapping windows cut from the retained channels of the spatial filter (each line fitted to a 400 ms segment with 120 ms overlap) and then passed to the classifier.

3.2. Results

The performances of the $a\pi$SF method for different numbers of retained channels are compared to those of the πSF and xDAWN algorithms in Fig. 1. Reported is the balanced accuracy (BA), which is the average of true positive and true negative rates and therefore unaffected by unbalanced class distributions. The performance values in Fig. 1 are reported

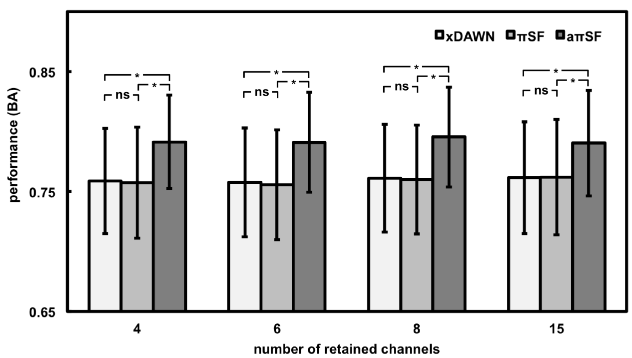

Fig. 1: Classification performances (mean and standard deviations for all subjects) obtained when using xDAWN, πSF and, $a\pi$SF respectively. The results are shown for different numbers of channels retained from these filters. Results of a statistical analysis (see text) are indicated (*=p<0.05; ns=p>0.05).

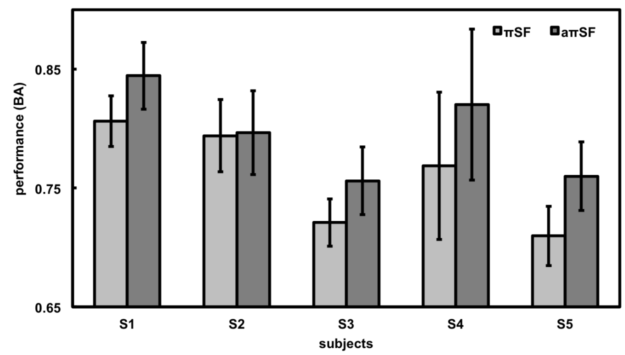

Fig. 2: Single subject performances (mean and standard deviation) obtained for $a\pi$SF and πSF (eight retained channels).

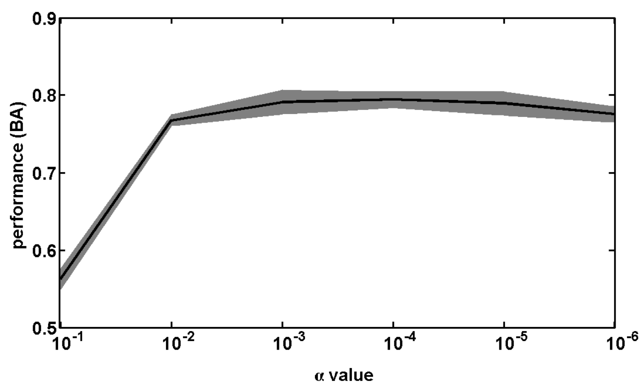

Fig. 3: Effect of α values on classification performance (eight retained channels). The solid black line shows the mean over results of all subjects and the shaded grey area indicates the corresponding standard deviation.

as means and standard deviations over the whole data set (5 subjects $\times$ 2 sessions). The session-wise performance was estimated as the average over all five runs. The differences obtained were compared using two-tailed t-tests for each number of retained channels, respectively. By this, we compared pairwise differences between the applied spatial filter methods. Reported p-values underwent Bonferroni correction (n=3) compensating for multiple testing. For all numbers of retained channels (rc), the use of $a\pi$SF led to a significant improvement in classification performance compared to the two static spatial filters, xDAWN (rc4: $p < 0.05$, all others: $p < 0.01$) and πSF (all: $p < 0.01$). By contrast, the xDAWN and the πSF method alone did not yield different performance levels (all: $p > 0.5$) as has been reported elsewhere [9].

To illustrate how $a\pi$SF improves performances for different subjects, the results for single subjects are depicted in Fig. 2, exemplarily for eight retained channels (i.e., detailed view of the corresponding results presented in Fig. 1). Although the statistical tests confirm that $a\pi$SF improves the performance on average, the results in Fig. 2 demonstrates that the actual amount of this improvement is varying from subject to subject.

For all results presented in Fig. 1 and Fig. 2 we used the same update coefficient, i.e., $\alpha = 10^{-4}$. This parameter is controlling the weight of the new incoming instance (compare Eq. 6 and 7). The effect of the choice of α on the obtained performance, again exemplarily for eight retained channels, is illustrated in Fig. 3. For big α values, small weights are assigned to $\mathbf{C}_{\mathbf{u}(k-1)}$, so that the spatial filter will be trained mostly based on the new coming samples. Therefore, the performance is dropping for large α values. Smaller α values make the spatial filter more conservative and if α approaches zero, the algorithm will converge to the non-adaptive πSF.

4. CONCLUDING DISCUSSIONS

The proposed method was evaluated in an inter-subject train-test schema. Our results confirm the notion that the use of adaptive algorithms for the processing of EEG data is beneficial to obtain better classification performances. The use of $a\pi$SF consistently outperforms the use of the non-adaptive version and the well established xDAWN method (compare Fig. 1). Besides the improve in performance, the usage of adaptive algorithms bears the big advantage that extra training sessions are less required for setting up a BCI system. Furthermore, it has been outlined that the choice of the update parameter α influences the performance outcome and controls the adaptivity characteristics of $a\pi$SF with respect to new incoming samples (Fig. 3).

Within $a\pi$SF the adaptive generalized eigenvalue decomposition is used to update the spatial filter vectors upon arrival of new ERP instances. The method is a fixed point algorithm which is faster than gradient based approaches.

Adaptive classifiers can further improve the performance of BCI systems. A system with simultaneously adapting spatial filter and classifier will be the follow up of this work.

5. REFERENCES

[1] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain-computer interfaces for communication and control," *Clin Neurophysiol*, vol. 113, no. 6, pp. 767–791, 2002.

[2] J. Ibez, J. Serrano, M. del Castillo, L. Barrios, J. Gallego, and E. Rocon, "An EEG-Based design for the online detection of movement intention," *Advances in Computational Intelligence*, pp. 370–377, 2011.

[3] Z. J. Koles, M. S. Lazar, and S. Z. Zhou, "Spatial patterns underlying population differences in the background EEG," *Brain Topogr*, vol. 2, no. 4, pp. 275–84, Jan 1990.

[4] H. Ramoser, J. Muller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial EEG during imagined hand movement," *IEEE Trans Rehabil Eng.*, vol. 8, no. 4, pp. 441–446, Dec 2000.

[5] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K.-R. Muller, "Optimizing spatial filters for robust EEG single-trial analysis," *IEEE Signal Process. Mag.*, vol. 25, no. 1, pp. 41–56, 2008.

[6] U. Hoffmann, J.-M. Vesin, and T. Ebrahimi, "Spatial filters for the classification of event-related potentials," in *Proc the 14th Europ Symp Articial Neural Networks (ESANN)*, Apr 2006, pp. 26–28.

[7] B. Rivet, A. Souloumiac, V. Attina, and G. Gibert, "xDAWN algorithm to enhance evoked potentials: Application to brain-computer interface," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 8, pp. 2035–2043, Aug 2009.

[8] A. Onishi, A. H. Phan, K. Matsuoka, and A. Cichocki, "Tensor classification for P300-based brain computer interface," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSSP)*, 2012, pp. 581–584.

[9] Ghaderi F. and Kirchner E., "Periodic spatial filter for single trial classification of event related brain activity," in *10th Int. Conf. Biomed. Eng. (BioMed)*, 2013.

[10] B. Rivet, H. Cecotti, R. Phlypo, O. Bertrand, E. Maby, and J. Mattout, "EEG sensor selection by sparse spatial filtering in P300 speller brain-computer interface," *Conf Proc IEEE Eng Med Biol Soc*, vol. 2010, pp. 5379–82, Jan 2010.

[11] C. Vidaurre, M. Kawanabe, P. von Büandnau, B. Blankertz, and K.R. Müandller, "Toward unsupervised adaptation of LDA for brain-computer interfaces," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 3, pp. 587–597, Mar 2011.

[12] A. Schlögl, C. Vidaurre, and K.-R. Müller, "Adaptive methods in BCI research – an introductory tutorial," in *Brain-Computer Interfaces*, Brendan Allison, Bernhard Graimann, and Gert Pfurtscheller, Eds., The Frontiers Collection, pp. 331–355. Springer, 2010.

[13] C. Vidaurre, C. Sannelli, K.-R. Müller, and B. Blankertz, "Machine-learning based co-adaptive calibration," *neural comput*, vol. 23, no. 3, pp. 791–816, 2011.

[14] C. Vidaurre, A. Schlogl, R. Cabeza, R. Scherer, and G. Pfurtscheller, "A fully on-line adaptive BCI," *Biomedical Engineering, IEEE Transactions on*, vol. 53, no. 6, pp. 1214–1219, June 2006.

[15] G.G. Molina, "BCI adaptation using incremental-SVM learning," in *3rd Int. IEEE/EMBS Conf Neural Eng, CNE '07.*, May 2007, pp. 337–341.

[16] Y. Huang, D. Erdogmus, M. Pavel, K. E. Hild, and S. Mathan, "Target detection using incremental learning on single-trial evoked response," *Acoustics, Speech, and Signal Processing, IEEE International Conference on*, vol. 0, pp. 481–484, 2009.

[17] S. Lu, C. Guan, and H. Zhang, "Unsupervised brain computer interface based on intersubject information and online adaptation," *IEEE Trans. Neural Syst. Rehabil. Eng*, vol. 17, no. 2, pp. 135–145, Apr 2009.

[18] G. Liu, D. Zhang, J. Meng, G. Huang, and X. Zhu, "Unsupervised adaptation of electroencephalogram signal processing based on fuzzy C-means algorithm," *Int. J Adapt Control and Sig Proc*, vol. 26, no. 6, pp. 482–495, 2012.

[19] A. Llera, V. Gmez, and H. J. Kappen, "Adaptive classification on brain-computer interfaces using reinforcement signals," *Neural Comput.*, vol. 24, no. 11, pp. 482–495, 2012.

[20] Tomioka R., Hill J., B. Blankertz, and K. Aihara, "Adapting spatial filtering methods for nonstationary BCIs," in *Workshop on Information-Based Induction Sciences, IBIS2006*, May 2006, pp. 65–70.

[21] Q. Zhao, L. Zhang, A. Cichocki, and J. Li, "Incremental common spatial pattern algorithm for BCI," in *IEEE World Congress on Computational Intelligence*, June 2008, pp. 2656–2659.

[22] W. Wojcikiewicz, C. Vidaurre, and M. Kawanabe, "Stationary common spatial patterns: Towards robust classification of non-stationary EEG signals," in *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*, May 2011, pp. 577–580.

[23] G. Strang, *Linear Algebra and Its Applications*, Brooks Cole, 3 edition, Feb. 1988.

[24] Y. N. Rao, J. C. Principe, and T. F. Wong, "Fast RLS-like algorithm for generalized eigendecomposition and its applications," *J. VLSI Signal Process. Syst.*, vol. 37, no. 2/3, pp. 333–344, June 2004.

[25] J. H. Metzen, S. K. Kim, and E. A. Kirchner, "Minimizing calibration time for brain reading," in *Proceedings of the 33rd international conference on Pattern recognition*, 2011, DAGM'11, pp. 366–375.