Available online at www.sciencedirect.com
ScienceDirect
Pattern Recognition Letters 28 (2007) 957–964
Pattern Recognition Letters
www.elsevier.com/locate/patrec

Spatial filtering and selection of optimized components in four class motor imagery EEG data using independent components analysis

Clemens Brunner *, Muhammad Naeem, Robert Leeb, Bernhard Graimann, Gert Pfurtscheller

Laboratory of Brain–Computer Interfaces, Institute for Knowledge Discovery, Graz University of Technology, Krenngasse 37, 8010 Graz, Austria

Received 10 May 2006; received in revised form 15 December 2006
Available online 17 January 2007

Communicated by G. Borgefors

Abstract

Three independent components analysis (ICA) algorithms (Infomax, FastICA and SOBI) have been compared with other preprocessing methods in order to find out whether and to which extent spatial filtering of EEG data can improve single trial classification accuracy. As reference methods, common spatial patterns (CSP) (a supervised method, whereas all ICA algorithms are unsupervised), bipolar derivations and the original raw monopolar data were used. In addition to only performing ICA, the number of components was reduced with PCA before calculating a spatial filter for Infomax and FastICA.

The multichannel data (22 channels) of eight subjects, consisting of two sessions recorded on different days, was analyzed. The task was to perform motor imagery of the left hand, right hand, foot or tongue, respectively, during predefined time slices (cued paradigm). For a measure of fitness, classification accuracies for both cross-validated results using data from just one session as well as simulated online results (representing the session-to-session transfer) were calculated. In the latter case, the spatial filters and classifiers were computed for one session and applied to the completely unseen second session.

For the data analyzed in this study, Infomax outperformed the other two ICA variants by far, both in the cross-validated as well as in the simulated online case. CSP, on the other hand, yielded significantly lower classification accuracies than Infomax for the cross-validated results, whereas there is no statistically significant difference when it comes to simulated online data. Performing PCA before ICA improved the results in the case of FastICA, whereas the classification accuracies dropped significantly for Infomax.

Keywords: Spatial filtering; Independent components analysis (ICA); Common spatial patterns (CSP); Principal components analysis (PCA); Electroencephalogram (EEG); Brain–computer interface (BCI); Motor imagery

1. Introduction

Independent components analysis (ICA) is an unsupervised statistical method used for decomposing a complex mixture of signals into independent sources (Vigário et al., 2000). It is especially suitable for preprocessing multichannel electroencephalographic (EEG) data in brain–computer interface (BCI) research because it can remove a number of different artifacts such as electromyogram (EMG) or electrooculogram (EOG) signals (Jung et al., 2000a,b). It can also be used to separate different rhythmic EEG components, such as right- and left-hemispheric mu rhythms, from ongoing EEG (Makeig et al., 2004).

In this study, a feature selection algorithm automatically selected a small number of ICA components that are optimally suitable to differentiate between different brain states associated with four motor imagery tasks in a BCI experiment. The main goal behind this strategy was to improve the single trial EEG classification accuracy by using ICA

* Corresponding author. Tel.: +43 316 873 5315; fax: +43 316 873 5349.
E-mail address: clemens.brunner@tugraz.at (C. Brunner).

0167-8655/$ - see front matter
doi:10.1016/j.patrec.2007.01.002

for spatial preprocessing and a subsequent feature selection algorithm for selecting the most relevant components. Here, bandpower features in a number of different bands between 8 Hz and 30 Hz were calculated from the preprocessed data. For comparison, the well-known spatial filtering method called common spatial patterns (CSP) (Koles, 1991; Müller-Gerking et al., 2000) was applied to the same data. In contrast to ICA, CSP is a supervised method that requires additional a priori information about the class of the data. As another reference method, bipolar derivations (which are simply differences between two monopolar EEG channels) were calculated. All preprocessing methods were compared with the results obtained from the original (monopolar) raw EEG data.

2. Subjects and experimental paradigm

In this study, the EEG data of eight subjects (three females and five males with a mean age of 23.8 years and a standard deviation of 2.5 years), recorded during a cue-based four class motor imagery task, was analyzed. Two sessions on different days were recorded for each subject, each session consisting of six runs separated by short (a couple of minutes) breaks. One run consisted of 48 trials (12 for each of the four possible classes), yielding a total of 288 trials per session.

The subjects were sitting in a comfortable armchair in front of a computer screen. As mentioned above, the paradigm consisted of four different tasks, namely the imagination of movement (motor imagery) of the left hand, right hand, foot, and tongue, respectively. At the beginning of each trial ($t = 0$ s), a fixation cross-appeared on the black screen. In addition, a short acoustic warning tone was presented at this time instant. After two seconds (at $t = 2$ s), a cue in the form of an arrow pointing either to the left, right, down or up (corresponding to one of the four classes left hand, right hand, foot or tongue) appeared for 1.25 s, prompting the subjects to perform the desired motor imagery task. No feedback (neither visual nor acoustic) was provided. The subjects were asked to carry out the mental imagination until the fixation cross-disappeared from the screen at $t = 6$ s. A short break followed, lasting at least 1.5 s. After that, the next trial started. The paradigm is illustrated in Fig. 1 (left).

Twenty two Ag/AgCl electrodes (with inter-electrode distances of 3.5 cm) were used to record the EEG, the setup is depicted in Fig. 1 (right). Monopolar derivations were used throughout all recordings, where the left mastoid served as reference and the right mastoid as ground. The signals were sampled at 250 Hz and bandpass-filtered between 0.5 Hz and 100 Hz. An additional 50 Hz notch filter was enabled to suppress line noise.

Although a visual inspection of the raw EEG data was performed by an expert, no trials were removed from the subsequent analysis in this study in order to evaluate the robustness and sensitivity to outliers and artifacts of each method. The fraction of artifacteous trials over all subjects was rather low anyway, namely 7.5% on average (median value of 6.1%).

3. Methods

3.1. Preprocessing

3.1.1. Spatial filters

The contamination of EEG signals with a variety of different artifacts such as EOG or EMG is an important issue in EEG data analysis. Appropriate precautions have to be taken in order to deal with this problem. Furthermore, the spatial resolution of EEG signals is compromised due to volume conduction through the scalp, skull and other layers of the brain. In the field of BCI research, these factors influence the classification accuracy of task-related activity. To address these problems, various spatial filtering techniques, for example common average reference (CAR), orthogonal source derivations, common spatial patterns (CSP), principle components analysis (PCA) and independent components analysis (ICA), can be utilized.

All these spatial filtering methods seek to solve the problems mentioned above by creating new components from the original data channels. In general, a spatial filter tries to estimate a so-called unmixing matrix $\boldsymbol{W} = [\boldsymbol{w}_1, \ldots, \boldsymbol{w}_n]$ such that the obtained components $\boldsymbol{y}(t) = [y_1(t), \ldots, y_n(t)]$ are as representative of the underlying sources as possible.

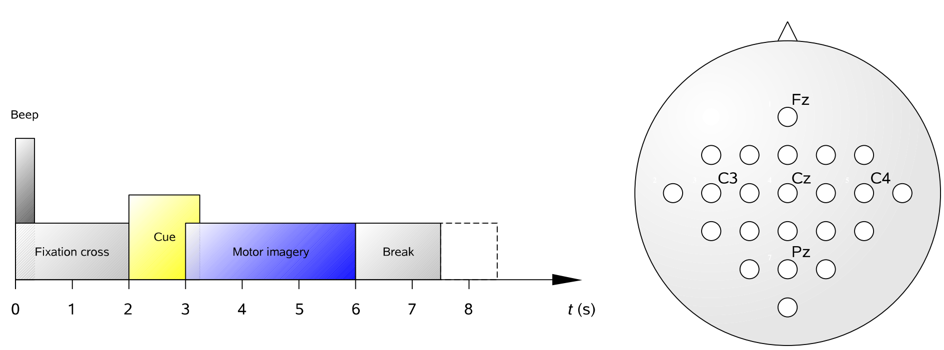

Fig. 1. Timing scheme of the BCI paradigm (left) and electrode setup of the 22 channels with inter-electrode distances of 3.5 cm. Some locations corresponding to the international 10–20 system are labeled (right).

These components can be calculated by multiplying the unmixing matrix $\boldsymbol{W}$ to the raw EEG signals $\boldsymbol{x}(t)$:

$$\boldsymbol{y}(t) = \boldsymbol{W}^{\mathrm{T}}\boldsymbol{x}(t).$$

Here, the vector of EEG signals $\boldsymbol{x}(t)$ consists of n channels, therefore the unmixing matrix $\boldsymbol{W}$ is, in general, a square $n \times n$ matrix. However, if the number of components is reduced to k by applying PCA as a first step, the unmixing matrix is only $k \times n$-dimensional.

3.1.2. Independent components analysis

Independent components analysis (ICA), like the loosely related principal components analysis (PCA), is a blind source separation technique (Vigário et al., 2000). The goal of these techniques is to recover sources from mixtures (observations) of signals with no information about the mixing matrix under certain assumptions and conditions. In other words, ICA is an unsupervised method, which means that there is no a priori information available – in the field of BCIs this implies that the class labels of the motor imagery data are not known to the method. The fundamental assumption employed by ICA for source extraction is statistical mutual independence (in contrast to PCA, which merely assumes uncorrelatedness). However, independence is only a concept, represented by different ICA algorithms in the form of a suitable contrast function that provides a measure of the degree of independence. The optimization of these contrast functions through iterative processes estimates the unmixing matrix $\boldsymbol{W}$ and ultimately yields the distinct independent components. Another assumption ICA makes is the non-gaussian distribution of the components.

In this paper, the specific ICA algorithms employed were Infomax, FastICA and SOBI (Second Order Blind Identification). One of the reasons for this choice was the large number of successful applications in various fields of data mining, particularly medical signal processing (Makeig et al., 2004; Tang et al., 2005a). The other factors were the broad availability of the algorithms and the diversity in the way the unmixing matrix is estimated.

The information maximization algorithm, or short Infomax (Bell and Sejnowski, 1995), maximizes the joint entropy $H(\boldsymbol{y})$ of the outputs, thereby minimizing the mutual information $I(\boldsymbol{y})$ among the output components. The minimization of the mutual information therefore resolves output variables into independent components. The FastICA algorithm (Hyvarinen and Oja, 1997) extracts independent components by separately maximizing the negentropy $J(\boldsymbol{y})$ of each mixture. In contrast to Infomax and FastICA, which are linear instantaneous versions of ICA, SOBI (Belouchrani et al., 1997) exploits the time structure of the data for source extraction. Furthermore, SOBI relies only on stationary second order statistics that are based on a simultaneous diagonalization of a set of covariance matrices for the computation of an estimate of the unmixing matrix.

FastICA has two algorithmic approaches available for the estimation of the unmixing matrix: one is the symmetric approach and the other is the deflationary approach where independent components are estimated one by one like in projection pursuit (Hyvarinen and Oja, 2000). Infomax, on the other hand, has only a symmetric approach available where all the sources are extracted in parallel whereas source estimation in SOBI is accomplished by the process of joint diagonalization.

There are a number of parameters available that can be tuned in order to optimize the performances of both FastICA and Infomax. For this study, FastICA was using a tangent hyperbolic as its non-linearity instead of the default third order polynomial, which was not found to be a robust parameter, because it is only recommended when there are no outliers. This was not the case for the data analyzed in this paper, because all trials were retained and no artifacts were rejected. Moreover, the deflationary approach was used in favor to the symmetric approach. In the case of Infomax, all default parameters were used – for example, instead of the optional extended ICA algorithm, the logistic ICA algorithm with natural gradient features was used. On the other hand, SOBI has only the number of temporal delays as an adjustable parameter (Tang et al., 2005b). For this study, the default value of 50 delays ranging from 4 ms to 200 ms was used, which proved to be a good choice based on preliminary variations of the parameter between 2 and 250 delays (Naeem et al., 2006).

For both Infomax and FastICA, the number of components was also reduced by applying PCA and retaining only the 10 components containing most of the signal variance; these two additional variants were also analyzed in this paper.

Before calculating the spatial filters with all three ICA algorithms, all data sets were triggered (meaning that data epochs between the trials were thrown away) and detrended by a second order polynomial as a first step.

3.1.3. Common spatial patterns

The method of common spatial patterns is another spatial filtering technique that calculates new signals in such a way that the variances of these components contain the most discriminative information with respect to the different motor imagery classes (Müller-Gerking et al., 1999). This is accomplished by jointly diagonalizing the two corresponding covariance matrices, which means that in its original form this method can only be applied to binary (i.e., two classes) problems. However, extensions to multiclass problems have already been developed (Dornhege et al., 2004) by combining two or more spatial filters, thereby reducing the multiclass problem to several binary decisions. For the data analyzed in this study, four filters in a one-versus-the-rest scheme were necessary (one for each imagery type versus the remaining three classes). These four matrices were calculated within a time segment of 4.5–5.5 s inside a trial because a screening of different one-second time windows within a trial revealed this epoch to yield the best discriminability of the data.

As a first step, the raw data was bandpass filtered between 8 Hz and 30 Hz, which is a good general choice for EEG data (Ramoser et al., 2000). In order to be comparable with the other methods, the variance was calculated within a one second time window. Three different feature subsets were created by taking only the first and the last column of the filtered data for the first subset (i.e., the projections corresponding to the largest and smallest eigenvalues), while the second set contained the second and second last columns in addition, and the third set comprised also the third and third last columns. In contrast to the ICA methods, the data filtered with CSP already constitutes features that can be directly used as input for the following classifiers.

3.1.4. Bipolar derivations

A very simple spatial filter can be constructed by calculating bipolar derivations, which means subtracting a signal originating from one electrode site from a signal stemming from a different place. In order to compare the two sophisticated methods described above (ICA and CSP) with this commonly used derivation, three bipolar channels were created from the original 22 channels. Electrode positions C3, C*z*, and C4 were approximated by subtracting the signal anterior from the signal posterior to the corresponding site, because from physiological knowledge these sites over the motor cortex are known to contribute highly distinguishable information in this four class motor imagery task (Pfurtscheller et al., 2006).

3.1.5. Spatial filter variants

Summing up, the following spatial filter variants were analyzed and compared in this paper:

- Infomax.
- PCA and Infomax.
- FastICA.
- PCA and FastICA.
- SOBI.
- Bipolar derivations.
- CSP.
- Monopolar (unfiltered) data.

3.2. Feature extraction

After multiplying the spatial filters (calculated by the various ICA algorithms and the bipolar derivations) with the raw EEG signals, logarithmic bandpower features were calculated. The power was computed in a number of frequency bands within the range of 8–30 Hz, namely 21 bands with width 2 Hz, 19 bands with width 4 Hz, 15 bands with width 8 Hz and 7 bands with width 16 Hz, yielding a total of 62 overlapping bands. Finally, a one second moving average window was applied to the data.

As already mentioned above, EEG signals filtered with CSP matrices can readily be used as features, therefore no further processing was necessary here. However, by calculating the CSP features in a one second window, we made sure that all methods receive the same amount of information about the data.

3.3. Feature selection

Each method using bandpower features (i.e., all ICA variants, bipolar derivations and raw monopolar data) was subjected to an optimization process. As already mentioned above, 62 different frequency bands for each of the 22 ICA components were calculated. Out of these 1364 features (due to the number of channels, the bipolar data comprised only 186 features), a feature selection algorithm was used to find the best combination consisting of up to 10 different features. To this end, we used the so-called sequential floating forward selection (SFFS) algorithm, which was shown to yield very good results within a short amount of calculation time (Pudil et al., 1994). The algorithm is based on a bottom–up approach, which means that it starts with an empty feature set. In each step, one feature is added based on a certain performance criterion, which in our case was the maximum of the cross-validated classification accuracy c (number of correctly classified trials over the total number of trials). In the next step, the algorithm tries to remove one feature and keeps the reduced set if the performance improved. Then another feature is added and the whole procedure is repeated until the desired number of features is reached.

The feature selection algorithm was not applied to the CSP-filtered data, instead, three separate subsets were calculated (see Section 3.1.3) to optimize this method.

3.4. Classification

In order to classify the feature vectors, a set of Fisher's linear discriminant analyses (LDAs) was used to discriminate between the four motor imagery classes. Four of these statistical classifiers were combined into a one-versus-the-rest scheme without a majority voting strategy involved – the class was assigned to the classifier yielding the largest discriminant value (Devijver and Kittler, 1982; Duda et al., 2001).

Before doing that, the optimal time slice for training the classifiers was determined by a so-called running classifier (Müller-Gerking et al., 2000), which means that the trials were segmented into overlapping one-second slices and a classifier was trained and tested on the same data segment. That way it was possible to choose the segment yielding the highest classification accuracy for each subject separately.

In order to avoid overfitting, a 10×10 cross-validation strategy was used, meaning that the data was divided into 10 equally-sized portions. The training of the four LDAs was carried out only on nine segments (i.e., 90%) of this data, whereas the testing was done on the remaining unseen segment (10%). This procedure was repeated such that each segment was used as the testing set once. In addition, this random partitioning into 10 segments was repeated another 10 times, which yielded a total of 100 training and test sets, respectively. The estimated classification accuracy c was then computed by averaging over all results.

In this paper, two different calculations were performed, first the cross-validated classification results for each subject and session separately with the corresponding optimized feature sets and second, simulated online results. The latter is an important measure for actual BCI systems, since everything (the optimized features, the spatial filter matrix and the corresponding classifier) is set up using only training data and is then applied to the completely unseen test data, which was recorded on a different day. For this purpose, the complete first session of each subject was taken to train the system. The second session was then the unknown test data – this procedure was also repeated the other way round, i.e., the second session was the training set and the first the test set.

4. Results

4.1. Feature optimization

In most of the subjects, 10 different bandpower features were used because they showed the highest classification accuracy in the feature selection process. However, there were some exceptions where less than 10 features were necessary, but that was mostly the case for the subjects with poor performance. It is very difficult and not even very interesting to present the results in all its details here (showing the optimal feature sets for each subject), since the main purpose of this procedure was to automatically construct an optimal basis of features that can be used in the subsequent analysis. Fig. 2, for example, shows the number of selected features versus classification accuracy for all subjects and sessions in the case of Infomax. It can clearly be seen that the maxima are reached at the end of the scale, but the steepness decreases after three selected features. For CSP, the results of the three different feature sets are presented in the next section.

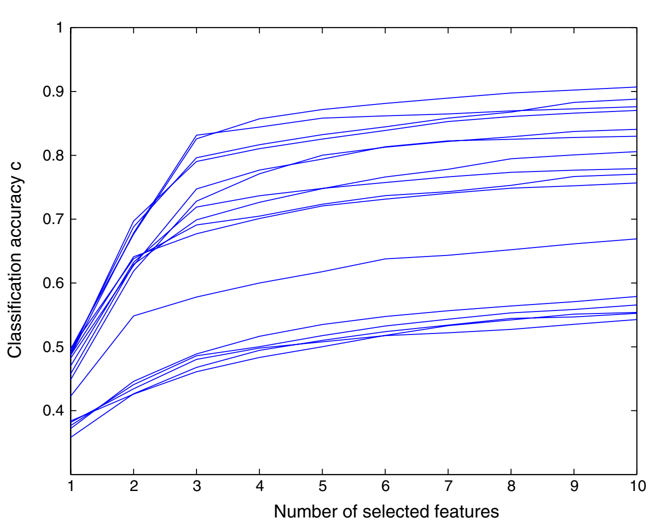

Fig. 2. Number of features (abscissa) versus classification accuracy (ordinate) for all subjects and sessions calculated for Infomax.

4.2. Cross-validated results

4.2.1. CSP

One preliminary goal was to find out how many columns of the CSP-filtered data are necessary to obtain good classification results. To this end, results were calculated with 2 columns (first and last), 4 columns and 6 columns (see Table 1). As can readily be observed, using only the columns corresponding to the largest and smallest eigenvalues yields significantly lower results as compared to four or six columns (in both cases a two-sided t-test was performed, which resulted in $p < 0.05$). There is no significant difference (the t-test yielded $p = 0.15$) between four and six components, but since there is still a higher mean classification accuracy for six components, this feature set was chosen as a representative and optimized CSP result. The time instants when the maxima occur within a trial is about the same for each variant (about second 4.9).

4.2.2. ICA algorithms

The results of the comparison of the three different ICA variants as well as bipolar derivations were twofold (see Fig. 3). First, Infomax (75.53%) greatly improves the mean classification accuracy over all subjects as opposed to the

Table 1
Maximum classification accuracies max(c) and corresponding time instant t (s) for each of the three different CSP-filtered data sets

Subject	CSP2		CSP4		CSP6	
	max(c)	t (s)	max(c)	t (s)	max(c)	t (s)
s1–1	0.7159	4.2	0.7176	4.2	**0.7309**	4.2
s1–2	0.7219	4.6	**0.7345**	5.1	0.7313	5.0
s2–1	**0.5625**	6.5	0.5564	6.5	0.5443	6.5
s2–2	**0.5013**	6.7	0.4977	6.8	0.4888	6.8
s3–1	0.7976	4.3	0.7917	4.7	**0.8015**	4.7
s3–2	0.7449	4.0	**0.7824**	4.5	0.7823	4.7
s4–1	0.4908	4.6	0.4968	4.4	**0.5039**	4.5
s4–2	0.5601	4.5	**0.5783**	4.1	0.5755	4.1
s5–1	0.3719	4.6	**0.3863**	4.6	0.3810	4.6
s5–2	0.3912	6.4	**0.4182**	6.4	0.4105	6.4
s6–1	0.6628	5.1	0.6850	4.8	**0.6924**	4.7
s6–2	**0.7851**	4.5	0.7728	4.6	0.7805	4.4
s7–1	0.7153	5.2	0.7369	4.6	**0.7448**	4.6
s7–2	0.6320	4.0	0.7267	4.6	**0.7564**	4.6
s8–1	0.6060	4.2	0.7174	4.2	**0.7675**	4.2
s8–2	0.8146	4.0	0.8328	4.1	**0.8340**	4.2
Mean	0.6296	4.84	0.6520	4.89	**0.6579**	4.89
STD	0.1398	0.91	0.1416	0.88	0.1490	0.87

CSP2 means that only the first and last column of the filtered data was used, CSP4 and CSP6 use additional 2 and 4 columns, respectively. The last two rows show the means and standard deviations (STD) of each CSP result. The best value for each subject and session is marked with a boldface font.

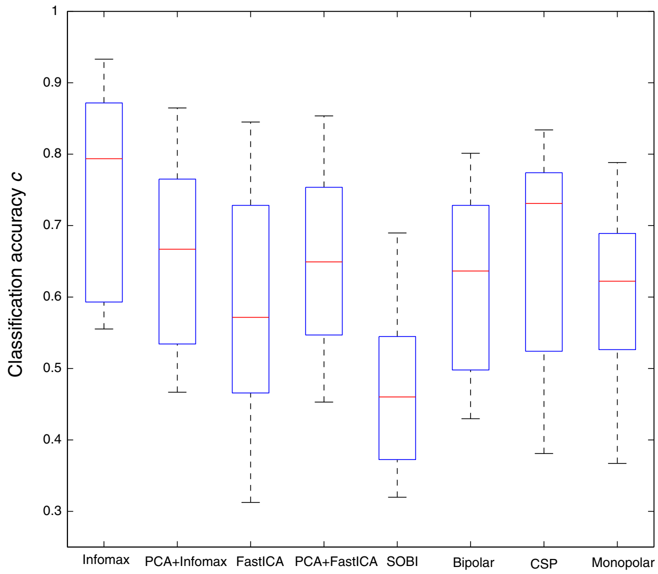

Fig. 3. Boxplots showing the cross-validated performance of each method. The upper and lower lines of a box show the upper and lower quartile, respectively, while the median corresponds to the line within the box. The whiskers are lines that indicate the range of the rest of the data.

monopolar data (59.62%) – the chance level is at 25% in a four-class problem. Bipolar derivations show only a slight improvement of the results (62.02%). More specifically, Infomax reaches much higher results for each single subject, whereas the accuracies for the bipolar derivations are sometimes better and sometimes worse, depending on the subject. Infomax performs significantly better than bipolar derivations and yields by far the highest accuracies of all algorithms. Performing Infomax on PCA-filtered data did not improve the classification accuracies – on the contrary, the results dropped to 66.12%.

Second, FastICA and especially SOBI yielded worse results as compared to unfiltered data with average classification accuracies of 58.71% and 47.06%, respectively. The difference is not extremely pronounced for FastICA – the fact that this ICA variant performed slightly worse on average is due to its complete failure for subject s4, where it reached accuracies of about 40%, whereas monopolar data was much better with 54% and 60%. On the other hand, the classification results obtained with SOBI were always worse than those obtained with monopolar data.

The performance of FastICA could be greatly improved by PCA. By only providing the first 10 PCA components, FastICA reached classification accuracies of 65.17% on average – this is significantly higher than the results without PCA, which was only 58.71%.

Taking also into consideration the cross-validated results from the CSP-filtered data, it is clear that Infomax outperformed all other algorithms by far. In second place comes CSP with a still competitive result (65.79%, about 10% lower than Infomax but still about 6% higher than monopolar data). It can be noted that for subject s5, CSP yielded a much lower classification accuracy than Infomax – for all other subjects, this difference is not that pronounced. Approximately the same good performance could be reached by combining PCA with Infomax and FastICA – in the first case, the results dropped to 66.12%, whereas in the latter case the accuracy greatly improved to 65.17%. The other algorithms performed significantly worse, with the bipolar derivations being better than monopolar data, FastICA, and lastly SOBI. In addition to the boxplot, the means and standard deviations of each method are summarized in Table 2.

The time instants where the maximum classification accuracy occurred are not presented, because they were all in a similar range of around 4.2–4.9 s. The only small outlier is SOBI, where the maximum occurred after 5 s on average.

4.3. Simulated online results

The same analysis was also performed for the simulation of an online experiment (Fig. 4). It is not surprising that the overall classification performance decreased as compared to the cross-validated results for all subjects. However,

Table 2
Mean and standard deviation of the maximum classification accuracies max(c) of the different preprocessing algorithms for cross-validated data

	Mean	Standard deviation
Infomax	0.7553	0.1361
PCA and Infomax	0.6612	0.1360
FastICA	0.5871	0.1677
PCA and FastICA	0.6517	0.1247
SOBI	0.4706	0.1075
Bipolar	0.6202	0.1219
CSP	0.6579	0.1490
Monopolar	0.5962	0.1305

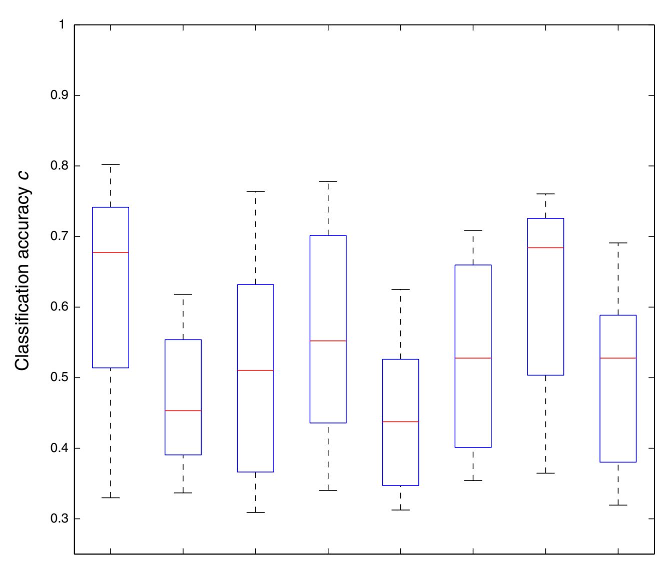

Fig. 4. Boxplots showing the simulated online performance of each method, for a description see Fig. 3.

Table 3
Mean and standard deviation of the maximum classification accuracies max(c) of the different preprocessing algorithms for simulated online data

	Mean	Standard deviation
Infomax	0.6196	0.1534
PCA and Infomax	0.4696	0.0956
FastICA	0.5148	0.1496
PCA and FastICA	0.5595	0.1515
SOBI	0.4460	0.1056
Bipolar	0.5295	0.1397
CSP	0.6179	0.1383
Monopolar	0.4985	0.1235

Infomax still performed best on average, but with CSP now very close behind. Actually, a two-sided t-test was not able to reject the null hypothesis that the means of those two methods differ, which means that it can be assumed that both methods perform equally well on completely unseen data. There are no outliers any more where Infomax performed much better than CSP, as was the case for the cross-validated data.

FastICA, bipolar derivations and monopolar data give approximately the same results (51.48%, 52.95% and 49.85%, respectively), whereas SOBI is again performing worst. Applying PCA before FastICA improved the performance of this algorithm to 55.95%. Infomax, on the other hand, does not benefit from reducing the dimensionality with PCA – the classification accuracy drops to 46.96%, which is worse than monopolar data and only slightly better than SOBI. In addition to the boxplots, the means and standard deviations of all methods are presented in Table 3.

5. Discussion

Most BCI experiments are based on the following strategy: First, the untrained (naive) subjects have to take part in a training session where they do not receive feedback in any form. The purpose of this procedure is to record data and afterwards calculate a classifier that is adapted to the specific subject's brain patterns. After that, this classifier can be used in a new online session, where feedback proportional to the output of the classifier can be provided. In order to simulate such a real-world situation, session-to-session transfer results were calculated in addition to cross-validated accuracies. When the two best performing methods, namely Infomax and CSP, are compared, notable differences between cross-validated and simulated online behavior can be observed.

First of all, it is quite surprising that Infomax performs significantly better than CSP in the cross-validated case, since CSP is a supervised method and gets a priori information about the data. In contrast, ICA does not need class label information since it is an unsupervised method. Still, there is another difference between those two spatial filtering techniques in our study, namely that ICA received the data of the whole trials (but no class labels), whereas CSP got only one second epochs per trial. This potential difference in the amount of information is not relevant, since the time slice between second 4.5 and 5.5 was found to contain the most discriminative information in the data. Moreover, ICA filter matrices from the same one-second epochs were also calculated, but the results did not change significantly.

Obviously, the potential advantage of CSP being a supervised method could not be exploited, at least with the data analyzed in this paper. A reason for this might be that this method could further benefit from additional optimizations, such as subject-specific bandpass filters (instead of 8–30 Hz) or time segments (instead of 4.5–5.5 s). Although taking into consideration only the columns corresponding to the largest eigenvalues (i.e., the first and the last column) would be sufficient, significantly better results can be obtained when choosing one or two additional components. On the other hand, as the online results of CSP became much better (or rather, the performance of Infomax decreased) such that both Infomax and CSP yielded similarly good results as opposed to cross-validated results, these predefined parameters did not seem to be bad choices at all. It rather seems that despite of having done a proper cross-validation, the results of Infomax seem to be slightly overfitted and components are extracted that are not present in the subjects' second sessions anymore.

Another way of improving the performance of CSP would be to train the filters on artifact-free data. However, a prerequisite in this paper was that all methods should receive the whole amount of data including artifacts in order to evaluate the robustness to outliers. In fact, CSP performed very well in comparison to the various ICA algorithms, especially in the case of simulated online data, which is a suitable measure for real-world BCI applications.

The results of the other two ICA algorithms (FastICA and SOBI) are a bit disappointing, especially those of SOBI. The latter performed even worse than unprocessed monopolar data, thus it must be concluded that this ICA variant exploiting the time structure of the data is not very suitable for improving the classification accuracy of multichannel EEG data. Another reason could be the sensitive dependence of SOBI to a proper choice of the time delays (number and range) (Tang et al., 2005b), which can be subject-specific. The difference in performance between Infomax and SOBI is also an interesting point, because they are very similar and are mathematically equivalent. However, when applied to real EEG data, the different optimization algorithms perform differently and Infomax seems to yield more robust results. That said, both Infomax and CSP can be recommended for online BCI applications where multiple channels are available.

Using PCA to reduce the dimensionality of the data yields different results for the ICA algorithms Infomax and FastICA, respectively. In the first case, the performance is reduced, whereas in the case of FastICA, PCA significantly improves the classification accuracy. This is true for both cross-validated results as well as simulated online data.

Finally, bipolar derivations yielded slightly better results in general as opposed to monopolar data, but they are nowhere near to Infomax or CSP. However, the amount of electrodes used is significantly lower, which might be a very practical advantage in BCI applications that involve patients. The importance of bipolar derivations over left and right sensorimotor cortical areas (close to electrode positions C3 and C4) has recently been demonstrated (Schlögl et al., 2005; Pfurtscheller et al., 2006). However, in addition to optimizing the frequency bands, also the exact bipolar locations could be optimized in order to further improve the performance of this method.

Acknowledgements

This work was supported by the European project Eye-To-IT (IST-2006-517590), the European project Presenccia (IST-2006-27731), EU COST B27 and the Higher Education Commission (HEC) of Pakistan.

References

Bell, A.J., Sejnowski, T.J., 1995. An information–maximization approach to blind separation and blind deconvolution. Neural Comput. 7, 1129–1159.

Belouchrani, A., Meraim, K.A., Cardoso, J.F., Moulines, E., 1997. A blind source separation technique using second-order statistics. IEEE Trans. Signal Proc. 45, 434–444.

Devijver, P.A., Kittler, J., 1982. Pattern recognition – A statistical approach. Prentice Hall.

Dornhege, G., Blankertz, B., Curio, G., Müller, K.-R., 2004. Boosting bit rates in noninvasive EEG single-trial classifications by feature combination and multiclass paradigms. IEEE Trans. Biomed. Eng. 51, 993–1002.

Duda, R.O., Hart, P.E., Stork, D.G., 2001. Pattern classification. Wiley.

Hyvarinen, A., Oja, E., 1997. A fast fixed-point algorithm for independent component analysis. Neural Comput. 9, 1482–1492.

Hyvarinen, A., Oja, E., 2000. Independent component analysis: Algorithms and application. Neural Networks 13, 411–430.

Jung, T.P., Makeig, S., Humphries, C., Lee, T.W., McKeown, M.J., Iragui, V., Sejnowski, T.J., 2000a. Removing electroencephalographic artifacts by blind source separation. Psychophysiol. 37, 163–178.

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

Koles, Z.J., 1991. The quantitative extraction and topographic mapping of the abnormal components in the clinical EEG. Electroenceph. Clin. Neurophysiol. 79, 440–447.

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

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

Müller-Gerking, J., Pfurtscheller, G., Flyvbjerg, H., 2000. Classification of movement-related EEG in a memorized delay task experiment. Clin. Neurophysiol. 111, 1353–1365.

Naeem, M., Brunner, C., Leeb, R., Graimann, B., Pfurtscheller, G., 2006. Separability of four-class motor imagery data using independent components analysis. J. Neural Eng. 3, 208–216.

Pfurtscheller, G., Brunner, C., Schlögl, A., Lopes da Silva, F.H., 2006. Mu rhythm (de)synchronization and EEG single-trial classification of different motor imagery tasks. Neuroimage 31, 153–159.

Pudil, P., Novovicova, J., Kittler, J., 1994. Floating search methods in feature selection. Pattern Recognition Lett. 15, 1119–1125.

Ramoser, H., Müller-Gerking, J., Pfurtscheller, G., 2000. Optimal spatial filtering of single trial EEG during imagined hand movement. IEEE Trans. Rehab. Eng. 8, 441–446.

Schlögl, A., Lee, F., Bischof, H., Pfurtscheller, G., 2005. Characterization of four-class motor imagery EEG data for the BCI-competition 2005. J. Neural Eng. 2, L1–L9.

Tang, A.C., Sutherland, M.T., McKinney, C.J., 2005a. Validation of SOBI components from high-density EEG. Neuroimage 25, 539–553.

Tang, A.C., Liu, J., Sutherland, M.T., 2005b. Recovery of correlated neuronal sources from EEG: The good and bad ways of using SOBI. Neuroimage 28, 507–519.

Vigário, R., Särelä, J., Jousmäki, V., Hämäläinen, M., Oja, E., 2000. Independent component approach to the analysis of EEG and MEG recordings. IEEE Trans. Biomed. Eng. 47, 589–593.