Institute of Physics Publishing — Journal of Neural Engineering

J. Neural Eng. **3** (2006) 208–216 doi:10.1088/1741-2560/3/3/003

Seperability of four-class motor imagery data using independent components analysis

M Naeem, C Brunner, R Leeb, B Graimann and G Pfurtscheller

Laboratory of Brain-Computer Interfaces (BCI-Lab), Graz University of Technology, Krenngasse 37, 8010 Graz, Austria

E-mail: pfurtscheller@tugraz.at

Received 27 January 2006
Accepted for publication 19 May 2006
Published 27 June 2006
Online at stacks.iop.org/JNE/3/208

Abstract

This paper compares different ICA preprocessing algorithms on cross-validated training data as well as on unseen test data. The EEG data were recorded from 22 electrodes placed over the whole scalp during motor imagery tasks consisting of four different classes, namely the imagination of right hand, left hand, foot and tongue movements. Two sessions on different days were recorded for eight subjects. Three different independent components analysis (ICA) algorithms (Infomax, FastICA and SOBI) were studied and compared to common spatial patterns (CSP), Laplacian derivations and standard bipolar derivations, which are other well-known preprocessing methods. Among the ICA algorithms, the best performance was achieved by Infomax when using all 22 components as well as for the selected 6 components. However, the performance of Laplacian derivations was comparable with Infomax for both cross-validated and unseen data. The overall best four-class classification accuracies (between 33% and 84%) were obtained with CSP. For the cross-validated training data, CSP performed slightly better than Infomax, whereas for unseen test data, CSP yielded significantly better classification results than Infomax in one of the sessions.

(Some figures in this article are in colour only in the electronic version)

1. Introduction

Motor imagery is a frequently used mental strategy to modify brain oscillations and to operate a brain-computer interface (BCI) (Wolpaw *et al* 2002, Pfurtscheller *et al* 2005). It is well known that execution or imagination of limb movement result in similar somatotopically organized activation patterns (Lotze *et al* 1999). For example, one-sided hand motor imagery reveals an event-related desynchronization (ERD) of mu and central beta rhythms focused over the contralateral hand representation area (Pfurtscheller and Neuper 1997).

Although mu rhythms of similar form and frequency are found in both hemispheres, they demonstrate no obvious bilateral coherence (Storm Van Leeuwen *et al* 1978, Andrew and Pfurtscheller 1996) or phase coupling (Spiegler *et al* 2004) and thus give evidence that mu generation systems are relatively independent of each other and exist in both hemispheres. The mu rhythm plays a dominant role in mental cursor control (Wolpaw *et al* 2000, Cheng *et al* 2004) and is therefore of special interest in BCI research.

Independent components analysis (ICA) separates multichannel EEG data into statistically independent components. It has been shown that ICA is especially suitable for removing a wide variety of artifacts in EEG recordings (Jung *et al* 2000) and separating mu rhythms generated in both hemispheres (Makeig *et al* 2004). ICA is therefore a useful method for constructing spatial filters for preprocessing raw multichannel EEG data in BCI research. Since its conception in the late 1990s and successful implementation in various fields of data mining, a surge of interest in ICA has resulted in the broad availability of various ICA algorithms. In this paper, the performance of three well-known ICA-based algorithms was evaluated. The specific algorithms studied were Infomax, FastICA, and SOBI (second order blind identification). The choice of these algorithms was made somewhat heuristically on the basis of past and more recent applications to biomedical data and the results achieved (Makeig *et al* 2005, Tang *et al* 2005b). The availability of algorithms also played a major factor in the choice. Moreover, the diversity in the methods of source extraction was a significant aspect of the choice made.

Another well-known preprocessing method is based on the decomposition of multichannel EEG data into spatial patterns which are calculated from two populations of brain patterns, known as common spatial patterns (CSP) (Koles *et al* 1995). In contrast to the ICA method, which is unsupervised, CSP is a supervised method where class information must be available *a priori*. The usefulness of the CSP method has already been proven in a number of BCI studies (Guger *et al* 2000, Ramoser *et al* 2000, Lemm *et al* 2005).

The main goal of this study is to address the following question: Does EEG preprocessing with different ICA algorithms improve the classification accuracy in the context of BCIs? To this end, we analyzed and classified 22-channel EEG recordings during four motor imagery tasks. For preprocessing we used, on the one hand, three different ICA algorithms and either all or a reduced number of components for feature extraction and classification, and on the other hand CSP. The data analysis was performed without removing any artifacts (EMG or EOG) in order to determine the robustness of the methods.

2. Methods

2.1. Subjects and experimental paradigm

The datasets were recorded from eight able-bodied subjects, three females and five males, with a mean age of 23.8 years (standard deviation of 2.5 years). The subjects were naive in BCI training. They were seated in a comfortable armchair in front of an LCD monitor.

The training paradigm consisted of a repetition of cue-based (synchronous) trials of four different motor imagery tasks, namely the imagination of left hand, right hand, foot and tongue movement. The beginning of each trial (at $t = 0$ s) was indicated by a short beep along with the display of a fixation cross in middle of the screen. After 2 s (at $t = 2$ s), a visual cue (an arrow pointing either left, right, up or down) appeared for 1.25 s on the screen. Each position of the arrow required the subject to perform the corresponding imaginary movement, i.e., either left, right, tongue or foot movement, respectively. Specifically, they were asked to keep up the imagination of the movement between 3 and 6 s. After 6 s (at $t = 6$ s), the fixation cross disappeared, allowing the subject to relax. The next trial started after 1.5–2.5 s resting period. The exact timing scheme is displayed in figure 1. The datasets of each subject consist of two sessions that were recorded on different days, each session comprising six runs. Each of the four types of cue were displayed 12 times within each run (which yielded a total of 72 trials per session) in a randomized order. For the subsequent data analysis in this paper, all the runs of each session were concatenated. Overall, there were 288 trials in each session for every subject.

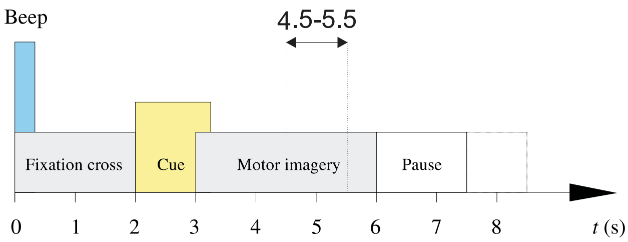

Figure 1. Timing of a trial of the training paradigm. The time slice between seconds 4.5 and 5.5 was used to train the classifiers. In the case of CSP the same time slice was also employed to calculate the spatial filters. However, for calculating ICA spatial filters the entire time period (0–7.5 s) was used.

2.2. Spatial filtering

Spatial smearing of EEG signals due to volume conduction through the scalp, skull and other layers of the brain is a well-known fact. To address this issue, various techniques of spatial filtering are used in EEG data analysis. The goal of spatial filtering is to determine, under different assumptions, an unmixing matrix $\mathbf{W} = [\mathbf{w}_1, \ldots, \mathbf{w}_n]$ so that new time series $\mathbf{y}(t) = [y_1(t), \ldots, y_n(t)]$ with a reduced smearing effect can be obtained. These components can be calculated by multiplying the unmixing matrix $\mathbf{W}$ to the raw EEG signals $\mathbf{x}(t)$:

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

Here, the EEG signals $\mathbf{x}(t)$ consist of n channels; therefore the unmixing matrix $\mathbf{W}$ is a square $n \times n$ matrix.

2.2.1. ICA-based spatial filtering. The fundamental assumption of ICA is statistical independence. There are a number of measures that estimate statistical independence between two time series (components), each leading to different optimization processes for the determination of $\mathbf{W}$. In turn, each of these optimality conditions gives rise to different ICA algorithms. All the ICA algorithms seek to unmix signals produced by the linear combination of multiple independent sources.

The ICA algorithms chosen for this study are Infomax, FastICA and SOBI. The first two are linear instantaneous versions of ICA, whereas SOBI exploits the time structure of the data for source extraction. The main idea of Infomax (Bell and Sejnowski 1995) is the minimization of the mutual information among the output components. FastICA (Hyvarinen 1997, Hyvarinen and Oja 1997) works on the principle of separately maximizing the negentropy of each mixture. In contrast, SOBI (Belouchrani *et al* 1993, 1997) relies only on stationary second-order statistics that are based on a joint diagonalization of a set of covariance matrices.

One salient feature of FastICA is its deflationary approach where independent components are estimated one by one like in projection pursuit (Hyvarinen and Oja 2000). Infomax, on the other hand, employs a symmetric approach where all the sources are extracted in parallel. The estimation of the sources in SOBI is accomplished by the process of simultaneous diagonalization.

Both FastICA and Infomax algorithms offer a variety of parameters that have to be tuned to obtain good results. It should be noted that the contrast function used in these algorithms is approximated by choosing a suitable nonlinearity, g. The choice of g is important in optimizing the performance of the algorithm. The default non-linearity used in FastICA is a third-order polynomial, which is recommended when there are no outliers. In our case, since we did not remove artifacts, a tangent hyperbolic was used instead. The choice of tanh was taken because it is a good general-purpose contrast function. In Infomax, all the default parameters were used—for example, the logistic Infomax algorithm with natural gradient features was employed (the RunICA algorithm) to search for super-Gaussian distribution of activity. It should be mentioned that for the detection of sub-Gaussian distributions such as slow cortical potentials, the optional extended ICA algorithm can also be used. In our case, since we were looking for ERD/ERS brain patterns, we did not choose the extended version of Infomax. By contrast, SOBI offers only one parameter to adjust, namely the temporal delays. In our preliminary analysis we experimented with delays 2 to 250 with a step size of 1. Although the best result was obtained with 120 delays, it was only marginally better than the default value of 50 delays. Therefore, the default value of 50 delays was used for this study.

Since ICA is essentially blind to order and scaling, the importance of the components cannot be determined on this basis. However, for the purpose of comparison, the components were sorted according to their mean projected variances. That is, the first component contributes the most to the EEG signal, the second contributes the second most and so on. This does not imply that the first component will contain the most significant ERD/ERS pattern, but it does suggest that components at the end are less likely to contain essential information. FastICA and SOBI originally do not sort components according to mean projected variances by default, but this feature was incorporated in those two algorithms in order to make them comparable with Infomax.

2.2.2. Common spatial patterns. The method of CSP designs spatial filters in such a way that the variances of the filtered time series are maximized and optimally discriminable with respect to the different motor imagery classes through joint diagonalization of the two corresponding covariance matrices. The resulting projection matrix is sorted in descending order of the eigenvalues, meaning that the most important filter pair is the first and the last entry in the matrix, the second best pair is the second and second last and so on.

The original method of CSP can discriminate only between two different classes or states (e.g., left-versus-right). Therefore, in order to extend the method to four-class motor imagery, four spatial filters were estimated on the principle of one-versus-rest, effectively reducing the multiclass problem to several binary problems (e.g., left-versus-rest etc). In the presented experiments, four spatial filters (one for each motor imagery type) have been created by calculating the covariance matrices over a 1 s long segment (4.5–5.5 s). From each filter matrix the projections corresponding to the three largest and to the three smallest eigenvalues were selected. For details about the extension of the CSP algorithm to multiclass problems see, e.g., Dornhege *et al* (2004).

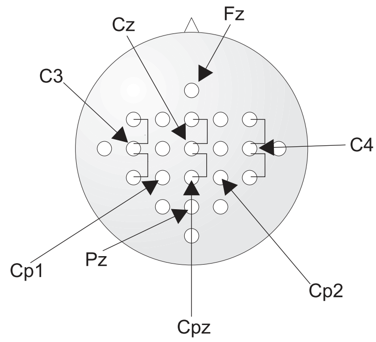

Figure 2. EEG electrode setup, some labels corresponding to positions in the international 10–20 system as well as the channels used to compile six bipolar channels are marked. Similarly, channels used to calculate six Laplacian derivations are also marked.

Table 1. Number of artifact-contaminated trials for each subject and session (the total number of trials per session was 288).

Subjects	s1	s2	s3	s4	s5	s6	s7	s8
Session 1	15	18	18	24	51	26	26	17
Session 2	7	5	15	17	24	60	12	11

2.3. Data analysis

2.3.1. Data recording. 22 Ag/AgCl electrodes (with inter-electrode distances of 3.5 cm), placed according to the scheme in figure 2, were used to record the EEG. Monopolar derivations were used throughout all recordings (the left mastoid served as reference and the right mastoid as ground). The signals were sampled with 250 Hz and filtered between 0.5 and 100 Hz. A 50 Hz notch filter was enabled to suppress line noise.

A visual inspection of the datasets was carried out by an expert and trials contaminated with artifacts (e.g., EMG, EOG) were marked (see table 1). However, since ICA has proven to be robust to outliers and artifacts (Jung *et al* 2000, James and Gibson 2003, Iriarte *et al* 2004), none of these trials were precluded from further analysis. In fact, this ability of ICA might also increase the statistical significance of the results. For the purpose of calculating ICA-based spatial filters, all the datasets were first triggered (i.e., each trial was cut out from continuous EEG data and then joined in the same order) and then detrended by a second-order polynomial (i.e., each individual 7.5 s trial was detrended).

2.3.2. Feature extraction and classification. In order to find optimal feature sets, components extracted by the ICA algorithms were visually examined with the help of time-frequency maps (Graimann *et al* 2002) and topographic maps (Delorme and Makeig 2004) with the goal of finding important and unnecessary components. The analysis by visual inspection led us to conclude that 6 out of the

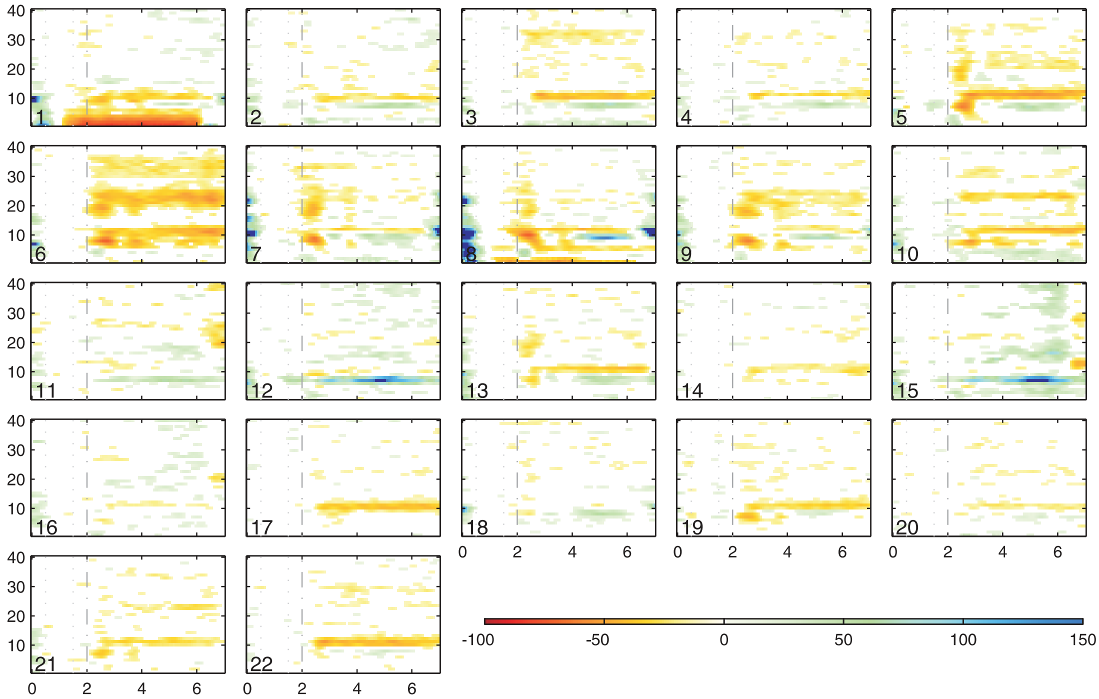

Figure 3. Time-frequency maps (Infomax): subject s3 session 1 (all 22 components). The six components selected are: 3, 5, 6, 7, 8, 10.

22 components were enough to represent the task-related activities.

More specifically, *a priori* knowledge of the physiological processes underlying motor imagery helped us in selecting the six most important components. In the case of hand imagery, the most important components were those focused on contralateral regions over the motor cortex area containing mu or beta ERD. The ipsilateral components containing ERS activity were also important. For foot or tongue imagery, midcentral or parietal components containing localized and prominent activity were considered. The components chosen to depict tongue imagery contained dominant ERS activity, whereas for foot imagery both ERD and ERS patterns were more significant (Pfurtscheller *et al* 2006). In addition, the components that showed scattered activity over the whole surface on a topographic map (which is merely a projection of the components $[\mathbf{w}_1, \ldots, \mathbf{w}_n]$ on a two-dimensional head surface) were not chosen. As an example time-frequency maps and topographic maps of subject s3 session 1 for Infomax are shown in figures 3 and 4

It should be mentioned that for building ICA-based spatial filters the entire triggered data (i.e., a whole session) for each subject were used. This ICA spatial filter (i.e., unmixing matrix) was then multiplied to the entire data. Then, six components were selected using the procedure described above. As a next step, for each of these six components, two logarithmic bandpower features (which correspond to the logarithm of the power in certain frequency bands) were calculated. The frequency bands selected were 10–12 Hz and 16–24 Hz for all subjects, covering the alpha and beta bands. Overall, there were 12 bandpower features from 6 components.

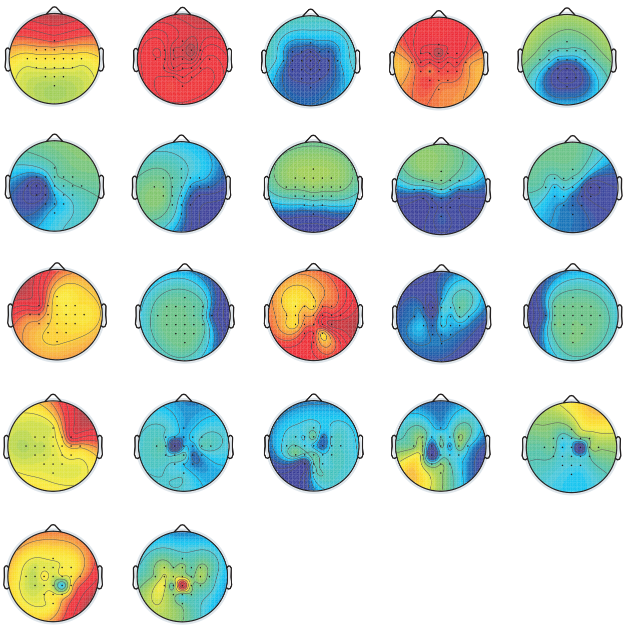

Figure 4. Topographic maps (Infomax): subject s3 session 1 (all 22 components). The six components (from left to right) selected are: 3, 5, 6, 7, 8, 10.

For the sake of comparison, all the 44 features (computed from all 22 ICA components) were also calculated.

In the next step, samples were partitioned into ten parts and each part was used as test set once in the following way. Nine parts (90%) of the data were used to train four linear statistical classifiers (fisher linear discriminant analysis, FDA) (Duda *et al* 2001) combined in a one-versus-the-rest classification scheme (i.e., one class was compared to all the remaining classes). These classifiers were then applied to the remaining 10% of the data and the classification accuracy was calculated by choosing the class corresponding to the maximum value of the four FDAs. The whole procedure was repeated ten times, i.e., a 10 × 10-fold cross-validation procedure (Duda *et al* 2001) was performed in order to avoid overfitting. Within each trial, samples between 4.5 and 5.5 s were used to train the classifiers. This time slice was chosen because it showed a high discriminability of the data for all the subjects.

In order to compare the classification accuracies of the ICA-filtered datasets with the results from other spatial preprocessing methods, CSP-preprocessed data (with the best six components) are also presented in this paper. It should be mentioned here that special care was taken to carry out the cross-validation correctly in the case of CSP, since this is a supervised method and the class labels are known when calculating the spatial filter matrices. More specifically, signals were first band-pass filtered in the range of 8–30 Hz and then samples were partitioned into ten parts before building four CSP spatial filters. Each part was used as test set only once in the following way. The (four) spatial filters were calculated on the basis of the 90% portion (nine parts) and were then multiplied to these data. In the next step, six components (the first and last three) were chosen and log-transformed normalized variances were calculated for each of the components. Next, these features were forwarded to four linear statistical classifiers (again using a one-versus-the-rest scheme). It should be noted that each classifier gets the same 24 features. The classifier weights were calculated and these classifiers plus the four spatial filters were then applied to the remaining 10% of the data. The whole procedure was repeated ten times, i.e., a 10 × 10-fold cross-validation procedure (Duda *et al* 2001) was performed and classification accuracies were determined. The same time slice (between 4.5 and 5.5 s) was used to train the classifiers, as in the case of ICA. It should be mentioned that in the case of CSP this time slice was also used to calculate the spatial filters.

Since our main goal was to find out whether ICA-based preprocessing improves classification results, we also present the classification accuracies of simple bipolar derivations (six components) and Laplacian derivations (six components) in this paper. The electrodes chosen for the bipolar derivations, in addition to C3, C4 and Cz, are one electrode anterior and posterior over these recording sites (see figure 2), whereas in the case of Laplacian derivations the electrodes chosen were C3, C4, Cz, CP1, CP2 and CPz. The same bandpower features as in the case of ICA were calculated. Similarly, the same procedure as in the case of ICA for feature extraction and classification was carried out.

In addition to the cross-validated analysis described above, the performance of ICA-based spatial filters was evaluated in session-to-session transfer. To this end, the unmixing matrices and classifiers from one session were used to calculate the classification accuracies of the remaining unseen session. The same procedure was carried out for the CSP components, bipolar derivations and the Laplacian derivations.

Table 2. Mean overall accuracies (in %) for all subjects and both sessions for the CSP methods using different numbers of components. The results were calculated on the basis of running classifier and without prefiltering for this analysis.

	Two components	Four components	Six components
With artifacts	63.3	67.2	68.9
Without artifacts	64.0	67.9	69.2

3. Results

3.1. CSP: preliminary analysis

In our preliminary analysis we experimented with two, four and six CSP components to optimize the results. Table 2 shows six components gives better results (higher overall accuracies) in comparison with two and four components. Moreover, the results using all trials were comparable with CSP spatial filters built with artifact-free trials only (see table 2); therefore we decided to keep all available trials.

It should be mentioned that for this particular analysis (table 2) a running classifier instead of the static classifier was used to calculate the results. Since our purpose was to show the comparative performance of the number of components chosen for CSP, and a running classifier calculates the results faster, it was preferred for this preliminary analysis (table 2). For the running classifier, the same segment of the data was used for training and testing purposes, whereas for a static classifier one particular segment (in our case between 4.5 and 5.5 s) was used for training and then the entire data set (0–7.5 s) for the purpose of testing. It should be noted that in all the subsequent analyses static classifiers were used.

3.2. Cross-validated data

The accuracies for features extracted from six visually selected ICA components for all the subjects are shown in table 3. In addition, the corresponding results of the CSP method and the Laplacian derivations are presented.

Surprisingly, the mean accuracies for the six visually selected cross-validated ICA components (table 3) are slightly lower (in the range of 1–2%) than the corresponding values of all 22 components (results not shown). The time points of highest classification accuracies were also slightly earlier especially for Infomax and FastICA.

The pairwise *t*-test, in this comparison, was performed on the data for each session separately at a 5% significance level. The results revealed that CSP showed significantly better performance in session 2, with respect to accuracies in comparison with FastICA, SOBI and the Laplacian derivations. Similarly, Infomax was found to be significantly better than FastICA and SOBI in session 2. On the other hand, in session 1 the *t*-test only found CSP to be significantly better than FastICA, SOBI and the Laplacian derivations. Moreover, when compared to Infomax, CSP did not show a significant

Table 3. Accuracies (cross-validated and unseen data) of six selected components—Overall accuracies of ICA algorithms, CSP and Laplacian derivation. The abbreviation STD denotes standard deviation.

	Cross-validated data					Session-to-session transfer				
Subject	Info	Fast	SOBI	CSP	Lap	Info	Fast	SOBI	CSP	Lap
s1/1	57.9	56.6	57.8	73.1	60.4	59.7	56.9	60.1	67.7	58.0
s2/1	37.0	52.7	52.6	54.4	50.8	35.4	35.4	35.7	45.5	51.7
s3/1	72.1	59.5	68.7	80.2	74.4	72.2	66.7	63.9	69.8	70.8
s4/1	76.0	67.1	62.6	74.5	68.7	64.9	55.2	55.2	69.1	65.6
s5/1	79.8	74.0	75.9	76.8	76.4	71.5	62.5	66.3	75.0	75.0
s6/1	46.0	47.2	42.9	50.4	37.7	51.3	49.0	45.8	49.3	39.6
s7/1	38.6	37.1	34.8	38.1	36.7	38.2	38.2	35.4	36.5	35.7
s8/1	70.9	70.0	56.1	69.2	71.4	59.7	61.5	55.6	67.0	58.7
Mean/1	59.8	58.0	56.4	64.6	59.6	56.6	53.2	52.3	60.0	56.9
STD /1	17.3	12.4	13.3	15.1	16.1	14.0	11.4	12.0	14.1	14.0
s1/2	70.3	69.7	65.7	73.1	60.8	64.9	64.9	65.7	74.3	57.6
s2/2	35.6	35.3	33.0	48.9	48.2	37.2	50.4	49.7	51.4	50.0
s3/2	73.0	66.8	68.7	78.2	75.2	71.2	52.8	63.9	76.0	75.7
s4/2	70.6	58.6	53.8	75.6	65.8	71.5	60.1	56.6	72.2	65.6
s5/2	77.8	72.1	73.2	83.4	73.2	69.8	59.6	64.9	71.2	72.6
s6/2	61.1	54.4	56.3	57.6	50.1	50.4	47.6	46.9	53.1	41.3
s7/2	43.3	41.4	38.3	41.1	33.6	35.4	35.8	32.6	37.5	38.5
s8/2	74.7	73.7	67.6	78.1	74.9	66.7	65.6	51.7	72.9	67.0
Mean/2	63.3	59.0	57.1	67.0	60.2	58.4	54.6	54.0	63.6	58.5
STD/2	15.6	14.4	14.8	15.7	15.1	15.2	10.0	11.3	14.3	14.1

difference in either session. This is interesting because in each session the mean accuracy values for CSP are higher in the range of 3–5 % as compared to the values obtained with Infomax.

Therefore, this comparison would place CSP as the best performing algorithm followed by Infomax, Laplacian, FastICA and finally SOBI. It should be mentioned that bipolar derivation performed significantly worse than Laplacian derivation for this comparison.

Contrary to the accuracy, pairwise *t*-tests with respect to the time points revealed nothing significant for any pair of the algorithms.

The accuracy plots of CSP (examples of all methods are presented in figure 5) show that the time courses in general are not similar to ICA algorithms or to Laplacian derivations. For example, the curve stays at a high level around the maximum value for a longer time period, which is unlike in the case of ICA algorithms. For online applications, such a behavior would be preferable, but that depends also on the paradigms one wants to use. On the other hand, the accuracy curves of ICA algorithms look quite alike.

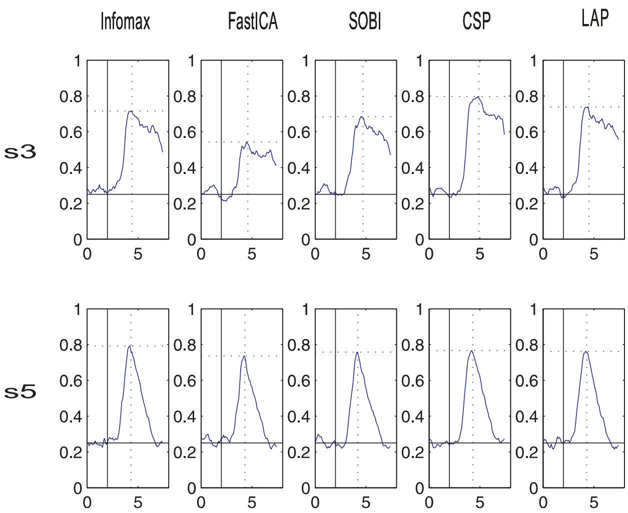

Figure 5. Comparison of different ICA algorithms, CSP and Laplacian derivations for two subjects and six components (cross-validated) showing the classification accuracy for the four-class problem (from 0 to 1) versus the time (in seconds). The cue onset is marked by a solid vertical line. The location of the maximum in each curve is marked by two dotted lines.

3.3. Session-to-session transfer

In order to simulate an online BCI application, the spatial filters and classifiers built with data from session 1 were applied to session 2 and vice versa. Therefore, it makes sense that the accuracy of each session in cross-validated data is compared with the accuracy of the corresponding session (in this case unseen data) in session-to-session transfer for all the algorithms.

The results are shown in table 3. One immediate conclusion that can be drawn from these results is that CSP has the best accuracy values in comparison with all the other algorithms. The performance of Laplacian derivations and Infomax is comparable followed by FastICA and SOBI.

Although the best classification result is obtained with CSP, comparing the results with cross-validated data shows a decrease in mean accuracy values of 4.6% and 3.4% for session 1 and session 2, respectively (see table 3 for more details). For the purpose of comparison, accuracy curves of s3 and s5 are presented (see figure 6).

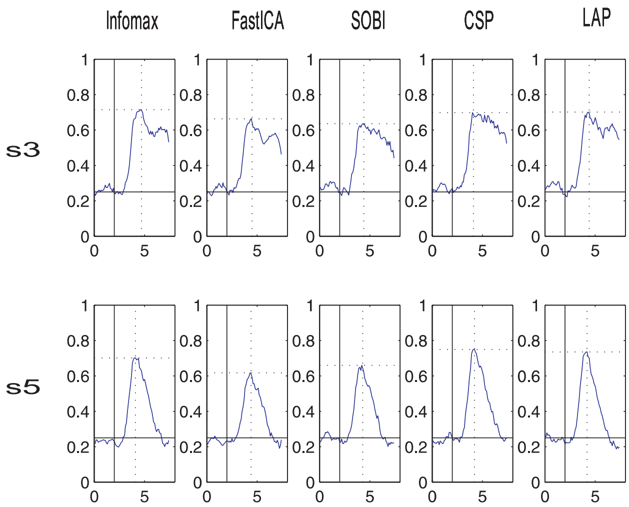

Figure 6. Comparison of different ICA algorithms, CSP and Laplacian derivations for two subjects and six components (unseen data) showing the classification accuracy for the four-class problem (from 0 to 1) versus the time (in seconds). The cue onset is marked by a solid vertical line. The location of the maximum in each curve is marked by two dotted lines.

The pairwise *t*-test at a 5% significance level showed that CSP has a higher and significantly different mean accuracy values in session 2 in comparison to all other algorithms. However, in session 1, CSP only performed significantly better than FastICA and SOBI. Similarly, Infomax was found to be significantly better than FastICA and SOBI in session 1. The differences between the remaining pairs, in both sessions, were not found to be statistically significant. It should be mentioned that Bipolar derivations performed slightly worse than Laplacian derivations for this comparison. Another important fact to note is that the performance of Infomax with 22 components in session-to-session transfer is comparable with the corresponding analysis with 6 selected components. In contrast, FastICA and SOBI performed worse for six visually selected components.

The pairwise *t*-test was also conducted with respect to the classification timing and no significant differences were found.

4. Discussion and conclusion

It is of importance to note that the EEG data analyzed were recorded from untrained subjects without feedback. They were only instructed to perform the indicated motor imagery tasks over a time period of 3 s (see figure 1). Two sessions were recorded on different days. On the one hand, the intrasubject stability of motor imagery related brain states could be studied and on the other hand, separated training and test sets could be generated for the classification. Due to the untrained subjects, the four-class classification accuracy demonstrated a great variability between about 33% and 84%.

4.1. Comparison of the different ICA algorithms

The best performance was achieved by Infomax when using all 22 components as well as for the selected 6 components in comparison with FastICA and SOBI. Therefore, among the ICA algorithms considered Infomax can be recommended for preprocessing EEG data in BCI research. In contrast, SOBI's overall performance is poor, worse than Laplacian or even bipolar derivations. One of the plausible reasons for this failure could be that time delay models of ICA may not be suitable to represent ERD/ERS related brain dynamics. Another reason may be the sensitive dependence of SOBI (Tang *et al* 2005a) to a proper choice of time delays (number and range). Although in our preliminary analysis the default value of 50 delays showed near optimum results, a direction of further research, as already pointed out in (Tang *et al* 2005a), is to investigate the optimum number and also the range of time delays for oscillatory activities.

On the other hand, robustness is an issue in the case of algorithms such as Infomax and FastICA. That is, two runs of FastICA, for example, will yield slightly different results. SOBI does not have this drawback. The lack of robustness can be directly traced to the optimization process of these algorithms. The search for a global minimum is sensitively dependent on the choice of initial values, which in turn are arbitrary and random. To address this issue, constrained ICA is an area of active research these days. The basic idea of constrained ICA is to incorporate prior information in ICA and other blind source separation (BSS) algorithms, thereby making them semi-blind. Recent publications (James and Gibson 2003, Lu and Rajapakse 2005) suggest various ways to do this. One of the approaches works by creating reference signals (James and Gibson 2003) of different classes and artifacts based on some distinct representative attributes and employing them as constraints in ICA algorithms. The other way is to create probability distribution functions of reference signals (Lu and Rajapakse 2005) and incorporate them directly as models in ICA algorithms.

It should be noted that, in addition to robustness, SOBI distinguishes itself in terms of speed of convergence and least number of tunable parameters requiring adjustment. On average, for this study, it took SOBI about 1.2 min for source resolution, while the average time taken by FastICA and Infomax was 12.1 min and 9.2 min, respectively. This could make SOBI a better candidate for a possible online adaptive implementation. The only adjustable parameter that SOBI has is the time delay. In contrast, Infomax and FastICA require many parameters to be adjusted for different situations.

Since there is an inherent indeterminacy with respect to scaling and permutation, in this study for the purpose of comparison, the components were sorted according to their mean projected variances. In our experience of the datasets used in this study, more often than not components at the end can be associated with EMG artifacts or in some cases electrode artifacts. But this does not imply that the components at the end are always useless and can never contain important information. Ideally, ICA can be expected to resolve one unique component for each movement imagery class. However, independence is a stringent criterion and different ICA algorithms would only seek an approximation to independence; therefore in practice often more than one extracted component is needed to represent a single task-related activity. In the case of our datasets, six components were enough to represent task-related activities.

4.2. Comparison of ICA with CSP and Laplacian derivations

Both in the case of cross-validated data as well as unseen test data the CSP method yielded better results in comparison with all the other algorithms considered. The overall better performance of CSP is an expected result as this is a supervised method where class information is known, whereas all the ICA algorithms employ blind source separation techniques in source extraction. Another important thing to note is that CSP performed significantly better than Infomax in the case of session-to-session transfer as determined by a *t*-test. However, for cross-validated data Infomax performed only slightly worse in comparison with CSP. In this context, it should be noted that for ICA algorithms we were using physiological knowledge by selecting corresponding components (related to motor imagery type) and using suitable bandpower features. In contrast, physiological information was not used for selecting the components in the case of CSP. Therefore, the performance of CSP might be further enhanced by utilizing physiological information in component selection.

Similarly high classification results were obtained with unseen testing data without sophisticated preprocessing algorithms by just calculating six Laplacian and bipolar derivations. Interestingly, the performance of Laplacian derivations was found to be comparable with Infomax preprocessed data and better than FastICA and SOBI. But perhaps by using sophisticated methods of automatic component selection, the performance of ICA-based algorithms can be improved further. This is important for the reason that visual selection of components has a subjective bias that may culminate in a cumulative error, whereas methods of automatic component selection use objective criteria which might result in improved performance.

ICA has two important advantages. First, ICA (in theory) is able to separate EEG data into physiologically and functionally discrete sources (Makeig *et al* 2004) and second, ICA separates the contamination of biological artifacts, e.g., eye movement and muscle activity, as well as technical artifacts such as line noise. The first property is especially interesting because it underlines the importance of selecting a small number of components (in our case six components were selected) based on time-frequency (ERD/ERS) maps and topographic maps of the most reactive EEG components.

Acknowledgments

This work has been funded by the Higher Education Commission (HEC), Pakistan, and the European Presencia project (IST-2001-37927).

References

Andrew C and Pfurtscheller G 1996 Event-related coherence as a tool for studying dynamic interaction of brain regions *Electroenceph. Clin. Neurophysiol.* **98** 144–8

Bell A J and Sejnowski T J 1995 An information-maximization approach to blind separation and blind deconvolution *Neural Comput.* **7** 1129–59

Belouchrani A, Meraim K A, Cardoso J F and Moulines E 1993 Second-order blind separation of correlated sources *Int. Conf. on Digital Sig. Proc.* pp 346–51

Belouchrani A, Meraim K A, Cardoso J F and Moulines E 1997 A blind source separation technique using second-order statistics *IEEE Trans. Sig. Proc.* **45** 434–44

Cheng M, Jia W, Gao X, Gao S and Yang F 2004 Mu rhythm-based cursor control: an offline analysis *Clin. Neurophysiol.* **115** 745–51

Delorme A and Makeig S 2004 Eeglab: an open source toolbox for analysis of single trial EEG dynamics including independent component analysis *J. Neurosci. Methods* **134** 9–21

Dornhege G, Blankertz B, Curio G and 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 and Stork D G 2001 *Pattern Classification* (New York: Wiley)

Graimann B, Huggins J E, Levine S P and Pfurtscheller G 2002 Visualization of significant ERD/ERS patterns in multichannel EEG and ECOG data *Clin. Neurophysiol.* **113** 43–7

Guger C, Ramoser H and Pfurtscheller G 2000 Real-time EEG analysis with subject-specific spatial patterns for a brain-computer interface (BCI) *IEEE Trans. Rehabil. Eng.* **8** 447–56

Hyvarinen A 1997 A family of fixed-point algorithms for independent component analysis *IEEE Int. Conference on Acoustics, Speech, and Signal Processing* vol 5 pp 3917–20

Hyvarinen A and Oja E 1997 A fast fixed-point algorithm for independent component analysis *Neural Comput.* **9** 1482–92

Hyvarinen A and Oja E 2000 Independent component analysis: algorithms and application *Neural Netw.* **13** 411–30

Iriarte J, Urrestarazu E, Valencia M, Alegre M, Malanda A, Viteri C and Artieda J 2004 Independent component analysis as a tool to eliminate artifacts in EEG: a quantitative study *J. Clin. Neurophysiol.* **20** 249–57

James C J and Gibson O J 2003 Temporally constrained ICA: an application to artifact rejection in electromagnetic brain signal analysis *IEEE Trans. Biomed. Eng.* **50** 1108–16

Jung T P, Makeig S, Westerfield M, Townsend J, Courchesne E and Sejnowski T J 2000 Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects *Clin. Neurophysiol.* **111** 1745–58

Koles Z J, Lind J C and Soong A C 1995 Spatio-temporal decomposition of the EEG: a general approach to the isolation and localization of sources *Electroenceph. Clin. Neurophysiol.* **95** 219–30

Lemm S, Blankertz B, Curio G and Müller K R 2005 Spatio-spectral filters for improved classification of single trial EEG *IEEE Trans. Biomed. Eng.* **52** 1541–8

Lotze M, Montoya P, Erb M, Hulsmann E, Flor H, Klose U, Birbaumer N and Grodd W 1999 Activation of cortical and cerebellar motor areas during executed and imagined hand movements: an FMRI study *J. Cogn. Neurosci.* **11** 491–501

Lu W and Rajapakse J 2005 Approach and applications of constrained ICA *IEEE Trans. Neural Netw.* **16** 203–12

Makeig S, Debener S, Onton J and Delorme A 2004 Mining event-related brain dynamics *Trends Cogn. Sci.* **8** 204–10

Makeig S, Hammon P and Onton J 2005 Modeling mechanisms of biological control of EEG rhythms *Brain-Computer Interface Technology: 3rd International Meeting (Rensselaerville, NY)*

Pfurtscheller G, Brunner C, Schlögl A and Lopes Da Silva F H 2006 Mu rhythm (de)synchronization and EEG single-trial classification of different motor imagery tasks *Neuroimage* **31** 153–9

Pfurtscheller G and Neuper C 1997 Motor imagery activates primary sensorimotor area in humans *Neurosci. Lett.* **239** 65–8

Pfurtscheller G, Neuper C and Birbaumer N 2005 Human brain-computer interface *Motor Cortex in Voluntary Movement* (Boca Raton, FL: CRC) pp 367–401

Ramoser H, Müller-Gerking J and Pfurtscheller G 2000 Optimal spatial filtering of single trial EEG during imagined hand movement *IEEE Trans. Rehabil. Eng.* **8** 441–6

Spiegler A, Graimann B and Pfurtscheller G 2004 Phase coupling between different motor areas during tongue-movement imagery *Neurosci. Lett.* **369** 50–4

Storm Van Leeuwen W, Wieneke G, Spoelstra P and Versteeg H 1978 Lack of bilateral coherence of mu rhythm *Electroenceph. Clin. Neurophysiol.* **44** 140–6

Tang A C, Liu J and Sutherland M T 2005a Recovery of correlated neuronal sources from EEG: The good and bad ways of using SOBI *Neuroimage* **28** 507–19

Tang A C, Sutherland M T and McKinney C J 2005b Validation of SOBI components from high-density EEG *Neuroimage* **25** 539–53

Wolpaw J R, Birbaumer N, McFarland D J, Pfurtscheller G and Vaughan T 2002 Brain-computer interfaces for communication and control *Clin. Neurophysiol.* **113** 767–91

Wolpaw J R, McFarland D J and Vaughan T 2000 Brain-computer interface research at the Wadsworth center *IEEE Trans. Rehabil. Eng.* **8** 222–6