COMPARISON OF FOUR CLASSIFICATION METHODS FOR BRAIN-COMPUTER INTERFACE

Alexander Frolov[*], *Dušan Húsek*[†], *Pavel Bobrov*[‡]

Abstract: This paper examines the performance of four classifiers for Brain Computer Interface (BCI) systems based on multichannel EEG recordings. The classifiers are designed to distinguish EEG patterns corresponding to performance of several mental tasks. The first one is the basic Bayesian classifier (BC) which exploits only interchannel covariance matrices corresponding to different mental tasks. The second classifier is also based on Bayesian approach but it takes into account EEG frequency structure by exploiting interchannel covariance matrices estimated separately for several frequency bands (Multiband Bayesian Classifier, MBBC). The third one is based on the method of Multiclass Common Spatial Patterns (MSCP) exploiting only interchannel covariance matrices as BC. The fourth one is based on the Common Tensor Discriminant Analysis (CTDA), which is a generalization of MCSP, taking EEG frequency structure into account. The MBBC and CTDA classifiers are shown to perform significantly better than the two other methods. Computational complexity of the four methods is estimated. It is shown that for all classifiers the increase in the classifying quality is always accompanied by a significant increase of computational complexity.

Key words: *Brain computer interface, motor imagery, visual imagery, EEG pattern classification, Bayesian classification, Common Spatial Patterns, Common Tensor Discriminant Analysis*

Received: March 1, 2011
Revised and accepted: April 5, 2011

[*]Alexander Frolov
Institute of Higher Nervous Activity and Neurophysiology, RAS, Butlerova 5a, Moscow, Russia, E-mail: aafrolov@mail.ru

[†]Dušan Húsek
Institute of Computer Science, Academy of Sciences of the Czech Republic, Pod Vodárenskou věží 2, Prague 8, Czech Republic, E-mail: dusan@cs.cas.cz

[‡]Pavel Bobrov
Institute of Higher Nervous Activity and Neurophysiology, RAS, Butlerova 5a, Moscow, Russia, Department of Computer Science, VŠB-Technical University of Ostrava, 17. listopadu 15, Ostrava – Poruba, Czech Republic, E-mail: p-bobrov@yandex.ru

1. Introduction

A brain-computer interface (BCI) provides a direct functional interaction between a human brain and an external device. BCI consists of a brain signal acquisition system, data processing software for feature extraction and pattern classification, and a system to transfer commands to an external device and, thus, providing feedback to an operator. Many kinds of signals (from electromagnetic to metabolic [21, 34, 33, 14]) could be used in BCI. However, the most prevalent BCI systems are based on the discrimination of EEG patterns related to execution of different mental tasks [22, 18, 20]. This approach is justified by the presence of correlation between brain signal features and tasks performed, revealed by basic research [37, 22, 25, 27, 28]. Implementation of such systems is encouraged by emergence of low-cost commercial EEG devices (e.g. Emotiv EPOC EEG headset [13]). The proliferation of these devices on the consumer market has been accelerated by the ability to utilize BCI for rehabilitation of patients with different motor disabilities (see [27, 23]) and by a growing interest in using BCI for gaming and other consumer applications [24, 15, 10].

One crucial part of a BCI system is the EEG pattern classifier. The classifier is trained to distinguish between EEG patterns related to performance of different mental tasks. Execution of each task causes a certain command being sent to an external device, allowing the operator to control it by voluntarily switching between different tasks. If commands sent to the external device trigger different movements, then psychologically compatible mental states are imaginary movements of different extremities. For example, when a subject controls a vehicle or a wheelchair, he can easily associate right hand movement with a right turn of the device. Moreover, mental states related to imaginary movements of extremities are clearly identified by corresponding EEG patterns (synchronization and desynchronization reactions of the mu rhythm, [1, 29]), as demonstrated in successful BCI projects, such as Graz [27, 28, 26] and Berlin [6] BCI. However, potential applications of BCI extend beyond motion control, including controlling home appliances, selecting contacts in a phone address book, or web search engine manipulation. Such tasks are more naturally accomplished by controlling the BCI with voluntary generation of corresponding visual images. Recently we found that BCI classifier is able to distinguish not only EEG patterns corresponding to motor imagery but also to visual imagery [8].

There are many approaches to designing a BCI classifier [3]. One of the most efficient classifiers is based on the method of Common Spatial Patters (CSP) [30], allowing classification of states of two classes. In a variety of combinations with other methods it is widely used in BCI systems [4]. Later, CSP method was generalized for classifying more than two classes of mental states(Multiple-class Common Spatial Patterns, MCSP) [12, 17]. Both CSP and MCSP are based on multichannel EEG covariance matrices for different mental states ignoring the frequency signal structure. Recently MCSP was generalized to take it into account (Common Tensor Discriminant Analysis, CTDA) [38]. The method CTDA demonstrates very high efficacy in EEG pattern classification but it is computationally expensive. We showed [8] that the simplest Bayesian classifier (BC) based on EEG covariation matrices (as CSP and MCSP) also provides classification accuracy comparable with CSP and MCSP. Since accounting of EEG frequency structure in CTDA significantly increased classification accuracy, when compared with MCSP, it was a challenge for us to modify BC to account this structure. We called the modified classifier Multiband Bayesian classifier (MBBC). The goal of the present study is to compare classification accuracy and computational cost of four classifiers: BC, MCSP, MBBC and CTDA. The first two classifiers ignore EEG frequency structure while the latter two take it into account.

To estimate classification accuracy, we used three EEG data sets. The first two were obtained in our experiments with motor (MTR) and visual imagery (VIS) [7, 8]. The third EEG data set (BCIC) contains EEG recordings corresponding to motor imagery. It was provided by the organizers of BCI Competition IV (data set 2A, [4]). Different data sets were used to check whether the experimental paradigm and data acquisition technique have influence on results of classifier comparison.

2. Description of the Data Sets

2.1 BCIC data set

The data set was provided by the Institute for Knowledge Discovery (Laboratory of Brain Computer Interfaces), Graz University of Technology for BCI Competition IV in 2008 [4]. The data correspond to imagining of 4 different movements: of left hand, right hand, foot, and tongue. BCIC contains 22 electrode EEG recordings of two sessions for 9 subjects. Each session is comprised of 6 runs of movement mental imagery with 48 trials in each run (12 trials for each mental task). During every trial subjects had to perform a particular mental task for 4 sec. Totally each mental task had to be performed for 288 sec. A detailed description of experiment paradigm is available [4].

2.2 MTR data set

MTR data set contains EEG recordings for 7 male subjects, 23-30 years of age. EEG was recorded by 24 ActiCap (Munich, Germany) electrodes (Fz, F3, F4, Fcz, Fc3, Fc4, Fc7, Fc8, Cz, C3, C4, Cpz, Cp3, Cp4, P3, P4, Po3, Po4, Po7, Po8, Oz, O1, O2), Afz was used as reference. The data were digitized by 16 bit ADC NBL640 (NeuroBioLab, Russia) with sampling frequency 200 Hz and filtered within 1-30 Hz passband.

Subject was sitting in a comfortable chair, one meter from a 17” monitor, and was instructed to fix his gaze on a motionless circle (1 cm in diameter) in the middle of the screen. Four gray markers were placed around the circle to indicate the mental task to be performed. The change of the marker color into green signaled the subject to perform the corresponding mental task. Left and right markers corresponded to left and right hand movement imagining respectively. The bottom marker corresponded to leg movement imagining and the upper one corresponded to relaxation. Each command to imagine a movement was displayed for 15 seconds and was preceded by a relaxation period of 5 seconds. Each clue was preceded by a 2-second warning. Three such pairs “relaxation - motor imagination” presented in random order constituted a block, two blocks constituted a session. There were two sessions for each subject, training and testing. During the training session classifier was switched off and recording was used only for its learning. During the following testing session classifier was switched on and the result of classification was presented to a subject by color of the central circle. The circle became green if the result coincided with the instruction, and red in the opposite case. During the instruction to relax the presentation of classification result was switched off not to attract the subject‘s attention. The second, testing, session was designed to provide subjects with the output of the BCI classifier in real time to enhance their efforts to imagine movements.

On the whole, each instruction was presented for 30 sec during each session.

2.3 VIS data set

VIS data set contains EEG recordings for 7 male subjects, 24-31 years of age. The data correspond to performance of three mental tasks: to imagine a house, a face and to relax. At the beginning of the study each subject was presented with two types of pictures: faces (10 pictures from the Yale Face Database B [16] and houses (10 pictures from the Microsoft Research Cambridge Object Recognition Data Base, version 1 [32], adjusted to black-and-white). Subject selected one face and one house as samples to imagine during the experiment.

The experimental setting for each session of VIS data was very similar to that used for MTR data set with 4 exceptions. First, the experiment with each subject was conducted on 4 consecutive days, two sessions (training and testing) on each day. Second, there were not four but three instructions. Third, during the first three days of the study, EEG was recorded using the Emotiv Systems Inc. (San Francisco, USA) EPOC 16-electrode cap. Fourth, each series contained 3 blocks, each block contained two instruction for each visual image, and each instruction to relax lasted 7 sec. Thus, in each training and testing series, each instruction to imagine a picture was presented for 90 sec and to relax 84 sec. On the 4th day of the experiment, EEG was recorded using the same system as for MTR data set. Only data of the last, fourth, day were used for classifier comparing.

3. EEG Pattern Classification

3.1 Bayesian approach

Suppose that there are L different mental tasks to be distinguished and probabilities of each task to be performed are equal to $1/L$. Let also distribution of each mental task EEG signal recorded by N_c electrodes be Gaussian with zero mean. Also, let $\mathbf{C}_i$, a covariance matrix of the signal corresponding to execution of the i-th mental task ($i = 1, \ldots, L$), be non-singular. Then, probability to obtain EEG signal $\mathbf{X}(t)$ under the condition that the signal corresponds to performing the i-th mental task is $P(\mathbf{X}(t) \mid i) \propto e^{-\frac{V_i(t)}{2}}$, where $V_i(t) = \mathbf{X}^{\mathrm{T}}(t)\mathbf{C}_i^{-1}\mathbf{X}(t) + \ln(\det(\mathbf{C}_i))$, and $\mathbf{X}(t)$ is a column vector of dimensionality N_c. Following the Bayesian approach, the maximum value of $P(\mathbf{X}(t) \mid i)$ over all i determines the class to which $\mathbf{X}(t)$ belongs. Hence, the signal $\mathbf{X}(t)$ is considered to correspond to execution of the k-th mental task as soon as $k = \operatorname{argmin}_i \{V_i(t)\}$. The equality $\mathbf{X}^{\mathrm{T}}(t)\mathbf{C}_i^{-1}\mathbf{X}(t) =$

$\operatorname{trace}\left(\mathbf{C}_i^{-1}\mathbf{X}(t)\mathbf{X}^{\mathrm{T}}(t)\right)$ implies that

$$V_i(t) = \operatorname{trace}\left(\mathbf{C}_i^{-1}\mathbf{X}(t)\mathbf{X}^{\mathrm{T}}(t)\right) + \ln\left(\det(\mathbf{C}_i)\right). \tag{1}$$

In practice, all $V_i(t)$ are rather variable so it is more beneficial to split signal into epochs and compute average value $\langle V_i \rangle$ for each EEG epoch to be classified, using equation (1)

$$\langle V_i \rangle = \operatorname{trace}\left(\mathbf{C}\mathbf{C}_i^{-1}\right) + \ln\left(\det(\mathbf{C}_i)\right), \tag{2}$$

where $\mathbf{C}$ denotes an epoch data covariance matrix estimated as $\left\langle \mathbf{X}\mathbf{X}^{\mathrm{T}} \right\rangle$. Therefore, to perform the classifier training it is sufficient to compute the covariance matrices corresponding to each mental task. It makes BC computationally inexpensive. In our experience reasonable classification accuracy is achieved when epochs of 1 second length are used for $\langle V_i \rangle$ computation. This is a tradeoff between time resolution and classification accuracy.

3.2 Multi-band Bayesian approach

A natural way to take EEG frequency structure into consideration is to filter EEG within several non-overlapping passbands, perform separate classification of the filtered signals, and combine the results obtained for differed bands. This technique can be rather effective, as suggested by results of the BCI Competition IV [4], see also [2] for description of the classifier which was the most effective on the dataset 2a. The Bayesian approach can also be generalized in such a way so that Bayesian classification is performed for signals filtered in different passbands. More detailed description is as follows.

Let $\mathbf{X}_b$ be an epoch of EEG signal filtered within the b-th frequency band. Then, as for BC (see (2)), probability of $\mathbf{X}_b$ to belong to the i-th class is determined by

$$\langle V_{b,i} \rangle = \operatorname{trace}\left(\mathbf{C}_b\mathbf{C}_{b,i}^{-1}\right) + \ln\left(\det(\mathbf{C}_{b,i})\right),$$

where $\mathbf{C}_{b,i}$ is the i-th class covariance matrix of EEG filtered within the b-th passband and $\mathbf{C}_b = \left\langle \mathbf{X}_b\mathbf{X}_b^{\mathrm{T}} \right\rangle$. Under the assumption that signals $\mathbf{X}_b$ are mutually independent for different bands, the probability of signal $\mathbf{X}$ to correspond to the i-th mental task is

$$P\left(\mathbf{X} \mid i\right) = \prod_b P\left(\mathbf{X}_b \mid i\right) \propto \exp\left(-\frac{1}{2}\sum_b \langle V_{b,i} \rangle\right).$$

Hence, the epoch can be attributed to the class with the number $k = \operatorname{argmin}_i \left\{\left\langle \widetilde{V}_i \right\rangle\right\}$, where $\left\langle \widetilde{V}_i \right\rangle$, $i = 1, \ldots, L$, are computed as

$$\left\langle \widetilde{V}_i \right\rangle = \sum_b \left[\operatorname{trace}\left(\mathbf{C}_{b,i}^{-1}\mathbf{C}_b\right) + \ln\left(\det(\mathbf{C}_{b,i})\right)\right].$$

3.3 Multi-class Common Spatial Patterns

Classification method based on MCSP can be described as follows. First, covariance matrices $\mathbf{C}_i$ are estimated over multichannel EEG data recorded during the classifier training. Then, matrices $\mathbf{W}_i$ are sought to meet the following requirements

$$\begin{aligned} \mathbf{W}_i \mathbf{C}_i \mathbf{W}_i^{\mathrm{T}} &= \mathbf{D}_i \\ \mathbf{W}_i \widetilde{\mathbf{C}} \mathbf{W}_i^{\mathrm{T}} &= \mathbf{I}, \end{aligned} \tag{3}$$

where $\widetilde{\mathbf{C}} = \mathbf{C}_1 + \cdots + \mathbf{C}_L$, $\mathbf{I}$ is the identity matrix, and $\mathbf{D}_i$ are diagonal matrices. The problem of obtaining the matrices $\mathbf{W}_i$ has an explicit solution. Indeed, if $\widetilde{\mathbf{C}} = \mathbf{U}\mathbf{D}\mathbf{U}^{\mathrm{T}}$ is the singular value decomposition (SVD) of the matrix $\widetilde{\mathbf{C}}$, with $\mathbf{U}$ being a unitary matrix and $\mathbf{D}$ being a diagonal matrix, then it is easy to prove that $\mathbf{W}_i = \mathbf{U}_i \mathbf{D}^{-\frac{1}{2}} \mathbf{U}$ with $\mathbf{U}_i$ being unitary matrix from SVD decomposition $\mathbf{D}^{-\frac{1}{2}} \mathbf{U} \mathbf{C}_i \mathbf{U}^{\mathrm{T}} \mathbf{D}^{-\frac{1}{2}} = \mathbf{U}_i \mathbf{D}_i \mathbf{U}_i^{\mathrm{T}}$.

After the matrices $\mathbf{W}_i$ are found, signal corresponding to a certain state is segmented into epochs and for each epoch $\mathbf{X}$ vectors $\nu_i, i = 1, \ldots, L$, are computed by estimating variances of all components of vectors $\xi_i = \mathbf{W}_i \mathbf{X}$, i.e. $\nu_i = \operatorname{diag}\left(\mathbf{W}_i \left\langle \mathbf{X}\mathbf{X}^{\mathrm{T}} \right\rangle \mathbf{W}_i^{\mathrm{T}}\right)$. Then, $\mathbf{X}$ is mapped onto a feature vector $\ln(\nu)$, where ν is concatenation of all vectors ν_i and $\ln(\cdot)$ means component-wise log transform.

Classification of the feature vectors can be performed using any multiclass classifier. In our experiments, we used multiclass SVM algorithm [11] with general radial basis kernel $\gamma = 0.5$ and Euclidian metric for feature vectors classification.

When only two mental states are classified, MCSP is reduced to CSP. In this case, $\mathbf{W}_1 = \mathbf{W}_2 = \mathbf{W}$, $\xi_1 = \xi_2 = \xi$ and $\left\langle \xi\xi^{\mathrm{T}} \right\rangle$ is a diagonal matrix in both states ($\mathbf{D}_1$ in the 1st state and $\mathbf{D}_2$ in the 2nd state). If a particular component of ξ has low variance for a certain state, then this component has high variance for the other state and vice versa, since $\mathbf{D}_1 + \mathbf{D}_2 = \mathbf{I}$. It implies large difference in feature vector components for different states, which underlines the possibility of state discrimination. Similarly, when MCSP is applied to classify more than two states, for each state i there usually exist components of ξ_i with low variance in the state. This can explain the efficacy of MCSP.

3.4 Common Tensor Discriminant Analysis

MCSP method can be generalized in order to account for EEG frequency structure, as proposed in [38]. EEG is represented as a tensor $\mathcal{X} \in \mathbb{R}^{N_c \times N_f \times N_t}$, where N_c is the number of EEG channels, N_f is the number of frequency steps, and N_t is the number of samples. An appropriate way to obtain such representation is to use wavelet transform, in which case N_f equals the number of wavelet scales.

In tensor presentation covariance matrix $\mathbf{C}$ becomes a covariance tensor $\mathcal{R} \in \mathbb{R}^{N_c \times N_f \times N_f \times N_c}$ which is defined as contraction of outer product $(1/N_t)\, \mathcal{X} \circ \mathcal{X}$ on the last index which corresponds to time.

Let $\mathcal{X}_i$, $i = 1, \ldots, L$ be EEG tensors of L classes. The goal of CTDA is to find matrices $\mathbf{W}_{i,1}$ projecting $\mathcal{X}_i$ on the 1st mode and $\mathbf{W}_{i,2}$ projecting $\mathcal{X}_i$ on the 2nd mode, so that transformed tensors $\mathcal{Z}_i = \mathcal{X}_i \times_1 W_1^{\mathrm{T}} \times_2 W_2^{\mathrm{T}}$ would have diagonal covariance tensors and sum of the covariance tensors would also be diagonal with all non-zero elements equal to 1:

$$\begin{aligned} \mathcal{R}_i \times_1 W_{i,1}^{\mathrm{T}} \times_2 W_{i,2}^{\mathrm{T}} \times_3 W_{i,2}^{\mathrm{T}} \times_4 W_{i,1}^{\mathrm{T}} &= \mathcal{D}_i \\ \left(\sum_{i=1}^{L} \mathcal{R}_i\right) \times_1 W_{i,1}^{\mathrm{T}} \times_2 W_{i,2}^{\mathrm{T}} \times_3 W_{i,2}^{\mathrm{T}} \times_4 W_{i,1}^{\mathrm{T}} &= \mathcal{I}, \end{aligned} \tag{4}$$

where $\mathcal{R}_i$ is covariance tensors of $\mathcal{X}_i$, $\mathcal{D}_i = \mathbf{D}_{i,1} \circ \mathbf{D}_{i,2}$, and $\mathcal{I} = \mathbf{I}_{i,1} \circ \mathbf{I}_{i,2}$ with $\mathbf{D}_{i,k}$ being diagonal and $\mathbf{I}_{i,k}$ being identity matrices. In matrix representation of EEG data, i.e. when EEG recording is an $N_c \times N_t$ matrix, (4) is reduced to (3) and the problem of finding projection matrices becomes equivalent to that described in the previous subsection.

After the projection matrices are computed, based on training data EEG epoch tensor $\mathcal{X}$ to be classified is transformed as $\mathcal{Z} = \mathcal{X} \times_1 W_1^{\mathrm{T}} \times_2 W_2^{\mathrm{T}}$ and feature vector $\nu = \ln\left(\operatorname{diag}\left(\operatorname{mat}_3\left(\mathcal{Z}\right)\operatorname{mat}_3(\mathcal{Z})^{\mathrm{T}}\right)\right)$ is obtained. The feature vectors are classified in the same way as when MCSP method is used.

Note that in contrast to MBBC the CTDA accounts for correlations between EEG rhythms of different frequencies which are reported to be present in EEG signals [31].

4. Classification Quality Evaluation

To compare the classifiers, off-line analysis of both training and testing session data was performed using MATLAB (the Mathworks Inc., Natick, MA, USA). Additional filtering within 5-30 Hz passband was performed prior to BC and MCSP classification, and within multiple passbands (4-8, 8-12, 12-16, 16-20, 20-24, and 24-28 Hz) prior to MBBC classification. The data were converted into tensor of order 3 using continuous wavelet transform (CWT) prior to CTDA classification. Order 5 Tchebychev Type 2 filters (MATLAB Filter Design Toolbox) were used for filtering. Continuous wavelet transform was performed using MATLAB Wavelet Toolbox `cwt` function with Morlet wavelet and 26 scales corresponding to frequencies from 5 to 30 Hz. CWT with Morlet wavelet is reported to be an effective technique of band-power extraction in BCI technology [9].

The preprocessed EEG recordings corresponding to execution of mental tasks were then split into epochs of 1-second length. Then, 70% of epochs corresponding to each state were chosen randomly for classifier training, and the remaining 30% of epochs were used to test classifier. 100 such classification trials were made. Averaging over all classification trials resulted in $L \times L$ confusion matrix $\mathbf{P} = (p_{ij})$ for each classifier. Here p_{ij} is an estimate of probability to recognize the i-th mental task in case the instruction is to perform the j-th mental task.

We chose the mean probability of correct classification p, mutual information g between states recognized and instructions presented, and Cohen's κ as indices of classification efficacy. Given the confusion matrix $\mathbf{P}$, these indices can be calculated as follows:

$$
\begin{aligned}
p &= \frac{1}{L}\sum_i p_{ii}, \\
g &= -\sum_{i,j} p_{ij}p_{0j}\log_2\left(p_{ij}/p_{i0}\right), \\
\kappa &= \frac{\sum_i p_{ii}p_{0i} - \sum_i\left[p_{0i}p_{i0}\right]}{1-\sum_i\left[p_{0i}p_{i0}\right]},
\end{aligned}
\tag{5}
$$

where p_{0j} is probability of the j-th instruction to be presented and $p_{i0} = \sum_j p_{ij}p_{0j}$ is probability of the i-th mental state to be recognized. The probabilities p_{0j} were estimated by dividing the number of epochs corresponding to the j-th state by the number of all epochs. For all datasets considered p_{0j} were equal or very close to $1/L$. This is in agreement with our a priori supposition that all BCI commands are equally needed.

The better classifier performs the more confusion matrix is close to identity matrix. In case L states are classified perfectly $p = 1$, $g = \log_2 L$, and $\kappa = 1$. If classification is random, i.e. $p_{ij} = p_{i0}$ for all j, then $p = 1/L$, $g = 0$, and $\kappa = 0$.

Index p has an advantage of being evidently interpreted as the percentage of correct classification. Its disadvantage is that it does not depend on non-diagonal elements of confusion matrix and hence does not account for distribution of errors. That is why information index g was used together with p. Cohen's κ was chosen to be presented because this index is widely used to estimate classification quality [4].

When all probabilities of correct classification are equal, i.e. $p_{ii} = p$ for all i, and all probabilities of incorrect classification are equal, i.e. $p_{ij} = (1-p)/(L-1)$ for all $i \neq j$, the mutual information between instructions presented and the states classified can be obtained as:

$$
g = \log_2 L + p\log_2 p + (1-p)\log_2\left(\frac{1-p}{L-1}\right). \tag{6}
$$

Based on [35], (6) is often used to estimate BCI efficacy ([5], [19] [36]). But if the corresponding assumptions do not hold true, the value of g, calculated according to (6), is lower than the actual mutual information. In this study, we used the general formula (5).

5. Results

MTR data set. Tab. I exemplifies confusion matrix obtained as a result of Bayesian classification of one subject testing session data. The matrix is diagonally dominant, which means prevalence of correct classifier responses. For the presented confusion matrix $p = 0.79$, $g = 0.99$, and $\kappa = 0.71$. Note that $p = 0.25$ for random classification of 4 states.

Fig. 1 shows p, g, and κ values computed after classification of all subjects‘ training (bar graphs A, B, C) and testing (bar graphs D, E, F) session data using each method. Maximum values of indices presented over all subjects and sessions

		Instructions presented			
		1	2	3	4
States recognized	1	0.72	0.00	0.00	0.02
	2	0.09	0.77	0.04	0.07
	3	0.09	0.08	0.84	0.06
	4	0.11	0.15	0.12	0.85

Tab. I *Confusion matrices obtained as a result of BC classification of one subject testing session data.*

are $p = 0.79$, $g = 0.99$, and $\kappa = 0.71$ for BC, $p = 0.78$, $g = 0.92$, and $\kappa = 0.71$ for MCSP, $p = 0.87$, $g = 1.3$, and $\kappa = 0.83$ for MBBC, $p = 0.77$, $g = 0.92$, and $\kappa = 0.69$ for CTDA.

Although Fig. 1 demonstrates large difference in subject performance, all indices exceed values characterizing random classification. Classification accuracy of BC and MCSP methods, both ignoring EEG frequency structure, differs just slightly. Although on average MCSP occurred to perform 15% better in terms of κ, the difference is insignificant in terms of p and g indices (one sample t-test, $P > 0.06$ for both indices). Accounting for EEG frequency structure has increased classification quality significantly. MBBC performs much better than BC (one sample t-test, $P < 10^{-5}$ for all indices) and CTDA is better than MCSP (one sample t-test, $P < 10^{-5}$), moreover, CTDA has shown better performance than a MBBC (one sample t-test, $P < 0.02$ for all quality indices). Despite maximum values of quality indices are lower for CTDA, in average its classification accuracy is much higher than for other methods, as shown in Tabs. II, III, and IV.

VIS data set. Relations between methods, revealed by VIS data set analysis, are similar to those observed for MTR data set, as suggested by Fig. 2 and Tabs. II, III, IV. Again, MCSP performed slightly better than BC. But for VIS data the difference in performance is significant in terms of all indices (one sample t-test, $P < 0.02$ for all indices). MBBC performed better than BC (one sample t-test, $P < 10^{-5}$ for all indices) and CTDA performed better than MCSP (one sample t-test, $P < 10^{-4}$ for all indices). CTDA classification accuracy was also the best in average.

Maximum values of indices presented over all subjects and sessions are $p = 0.62$, $g = 0.31$, and $\kappa = 0.44$ for BC, $p = 0.66$, $g = 0.41$, and $\kappa = 0.51$ for MCSP, $p = 0.67$, $g = 0.41$, and $\kappa = 0.52$ for MBBC, $p = 0.79$, $g = 0.63$, and $\kappa = 0.68$ for CTDA. Note that $p = 0.33$ for random classification of three states.

BCIC data set. The results of BCIC data analysis are shown in Fig. 3 and Tabs. II, III, IV. Similarly to MTR and VIS data sets MCSP has shown better performance than BC (one sample t-test, $P < 0.001$ for all indices). MBBC performed better than BC (one sample t-test, $P < 10^{-4}$ for all indices) and CTDA performed better than MCSP (one sample t-test, $P < 0.002$ for all indices).

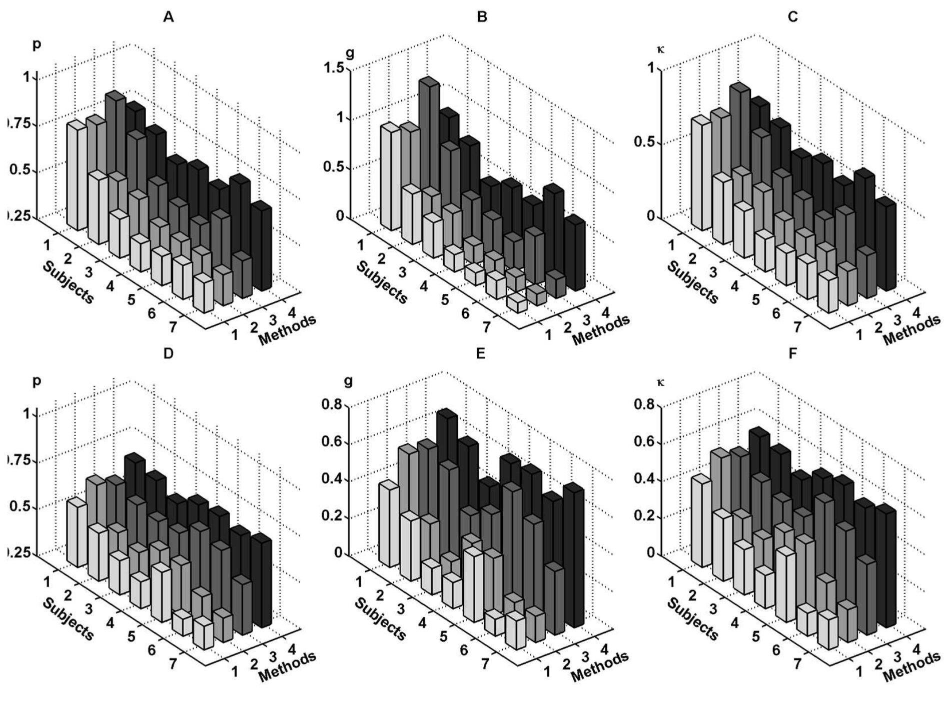

Fig. 1 *Indices p, g and κ obtained as a result of MTR training and testing sessions classification for all subjects, using all methods (1 — BC, 2 — MCSP, 3 — MBBC, 4 — CTDA). Note that graphs A and D represent exceeding of p over level of random classification.*

	BC	MCSP	MBBC	CTDA
1a	0.50 ± 0.05	0.52 ± 0.05	0.61 ± 0.06	0.70 ± 0.02
1b	0.45 ± 0.03	0.49 ± 0.03	0.59 ± 0.02	0.68 ± 0.01
2a	0.50 ± 0.01	0.52 ± 0.02	0.56 ± 0.02	0.64 ± 0.02
2b	0.52 ± 0.02	0.57 ± 0.02	0.60 ± 0.02	0.71 ± 0.02
3a	0.47 ± 0.04	0.51 ± 0.04	0.54 ± 0.04	0.63 ± 0.02
3b	0.46 ± 0.04	0.49 ± 0.04	0.55 ± 0.04	0.64 ± 0.03

Tab. II *Average values of p index computed after processing of different data sets: MTR training session (1a), MTR testing session (1b), VIS training session (2a), VIS testing session (2b), two BCIC sessions (3a and 3b).*

6. Discussion

Figs. 1, 2, and 3 show that three quality indices p, g, and κ are in good agreement, particularly, high value of any index for some subject implies that values of other two indices are also high for this subject. Generally, these indices do not duplicate

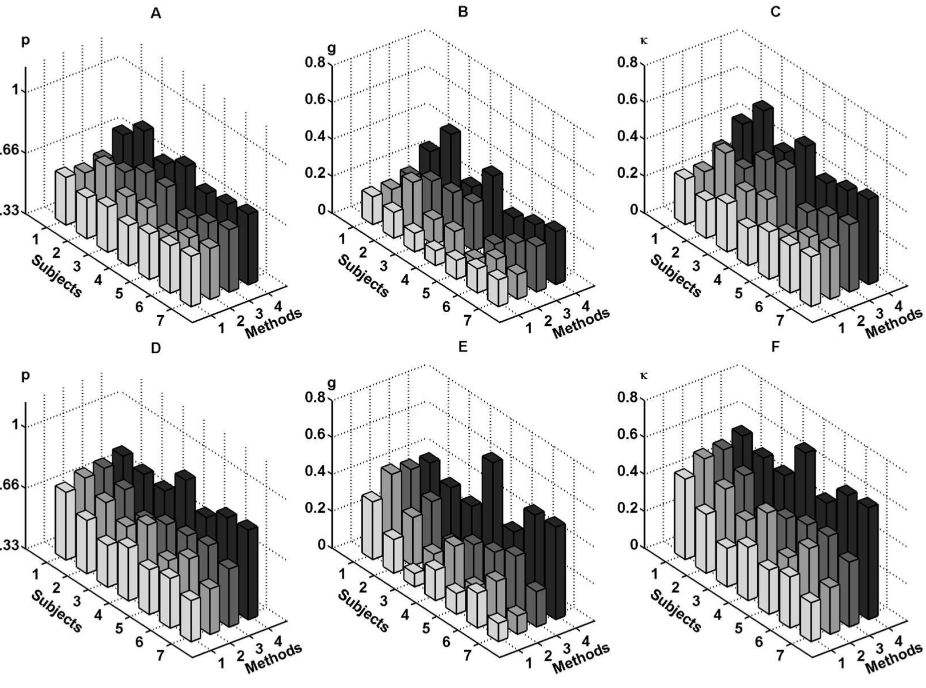

Fig. 2 *Indices p, g and κ obtained as a result of VIS training and testing sessions classification for all subjects, using all methods (1 — BC, 2 — MCSP, 3 — MBBC, 4 — CTDA). Note that graphs A and D represent exceeding of p over level of random classification.*

	BC	MCSP	MBBC	CTDA
1a	0.35 ± 0.12	0.33 ± 0.11	0.55 ± 0.14	0.70 ± 0.06
1b	0.23 ± 0.05	0.27 ± 0.06	0.48 ± 0.04	0.63 ± 0.03
2a	0.12 ± 0.01	0.15 ± 0.02	0.21 ± 0.02	0.29 ± 0.03
2b	0.16 ± 0.03	0.21 ± 0.04	0.26 ± 0.03	0.43 ± 0.04
3a	0.30 ± 0.05	0.35 ± 0.06	0.38 ± 0.05	0.53 ± 0.07
3b	0.28 ± 0.06	0.32 ± 0.06	0.40 ± 0.05	0.50 ± 0.05

Tab. III *Average values of g index computed after processing of different data sets: MTR training session (1a), MTR testing session (1b), VIS training session (2a), VIS testing session (2b), two BCIC sessions (3a and 3b).*

but complement each other. Index p has the most evident interpretation but it ignores classification error distribution. In contrast, indices g and κ depend on error distribution. Index κ evaluates only classification quality reaching 1 in case of perfect classification independently of number of states classified while information index g evaluates information gain provided by BCI reaching $\log_2 L$ in case L states

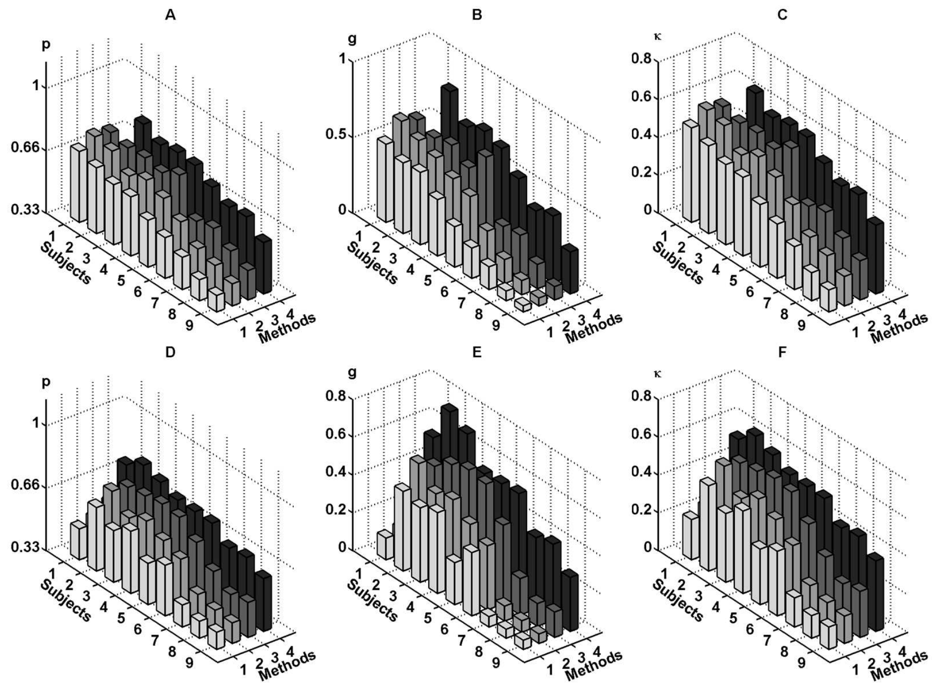

Fig. 3 *Indices p, g and κ obtained as a result of BCIC training and testing sessions classification for all subjects, using all methods (1 — BC, 2 — MCSP, 3 — MBBC, 4 — CTDA). Note that graphs A and D represent exceeding of p over level of random classification.*

	BC	MCSP	MBBC	CTDA
1a	0.33 ± 0.07	0.37 ± 0.06	0.48 ± 0.07	0.60 ± 0.02
1b	0.26 ± 0.05	0.33 ± 0.05	0.46 ± 0.02	0.58 ± 0.02
2a	0.24 ± 0.01	0.29 ± 0.03	0.35 ± 0.02	0.46 ± 0.02
2b	0.28 ± 0.03	0.37 ± 0.04	0.40 ± 0.03	0.56 ± 0.03
3a	0.25 ± 0.07	0.31 ± 0.08	0.35 ± 0.07	0.50 ± 0.04
3b	0.25 ± 0.07	0.28 ± 0.08	0.36 ± 0.08	0.51 ± 0.04

Tab. IV *Average values of κ index computed after processing of different data sets: MTR training session (1a), MTR testing session (1b), VIS training session (2a), VIS testing session (2b), two BCIC sessions (3a and 3b).*

are classified perfectly. Therefore, information index g is an appropriate measure for comparing performance of brain-computer interfaces based on different principles (e.g. BCI based on voluntary switching of different mental states and BCI based on P300).

Classifier comparison yielded quite similar results for all data sets. BC and MCSP classifiers based solely on interchannel covariance are shown to be comparable in performance while loosing to MBBC and CTDA classifiers which additionally account for EEG frequency structure. Although for MTR data set maximal quality of performance is provided by MBBC, CTDA based classifier is shown to provide the best classification accuracy on average over subjects for all data set. One explanation for CTDA superiority might be that it accounts for cross-band correlations which are reported to be present in EEG [31].

In conclusion, it is interesting to compare computational costs of the classifiers. Computational cost was evaluated by time spent on EEG preprocessing, classifier training, and classifier testing. Evaluation was performed using one subject's MTR training session data. Preprocessing was done as described in Section 4. A single classification trial was made with all data used both to train and to test the classifiers. Results of evaluation obtained on PC, Intel Core 2 Duo T9300 CPU 2.50 GHz, are shown in Tab. V along with kappa averaged over all subjects and datasets.

	κ	$\mathbf{t_{pre}, sec}$	$\mathbf{t_{trn}, sec}$	$\mathbf{t_{tst}, sec}$
BC	0.29	0.10	0.04	0.07
MCSP	0.34	0.10	0.48	0.06
MBBC	0.45	0.57	0.17	0.24
CTDA	0.55	0.61	50.43	21.39

Tab. V *Comparison of execution time for different classification methods and different classification steps: preprocessing(t_{pre}), classifier training(t_{trn}), and classifier testing (t_{tst}).*

In most cases higher classification accuracy demands more execution time as one could expect. Surprising is that computationally inexpensive basic Bayesian classifier showed accuracy comparable with that of much more sophisticated classifiers.

Acknowledgement

The work was supported in part by program of the Presidium of RAS "Basic research for medicine" and by the projects MSMT 1M0567, GA CR 201/05/0079, GA CR P202/10/02 and by the IRP ICS v. v. i. AS CR AV0Z10300504.

References

[1] Allison B., Brunner C., Kaiser V., Muller-Putz G., Neuper C., et al.: Toward a hybrid brain-computer interface based on imagined movement and visual attention. Journal of Neural Engineering, 7, 2010, 026007.

[2] Ang K. K., Chin Z. Y., Zhang H., Guan C.: Filter Bank Common Spatial Patterns (FBCSP) in Brain-Computer Interface. Proceedings of the IEEE International Joint Conference on Neural Networks, 2008, pp. 2391-2398.

[3] Bashashati A., Fatourechi M., Ward R., Birch G.: A survey of signal processing algorithms in brain-computer interfaces based on electrical brain signals. Journal of Neural Engineering, 4, 2007, R32.

[4] BCI competition IV. http://www.bbci.de/competition/

[5] Besserve M., Jerbi K., Laurent F., Baillet S., Martinerie J., et al.: Classification methods for ongoing EEG and MEG signals. Biological Research, 40, 2007, pp. 415-437.

[6] Blankertz B., Dornhege G., Krauledat M., Muller K., Curio G.: The non-invasive Berlin Brain-Computer Interface: Fast acquisition of effective performance in untrained subjects. NeuroImage, 37 2007, pp. 539-550.

[7] Bobrov P. D., Korshakov A. V., Roschin V. Y., Frolov A. A.: Investigation of Bayesian Approach in Motor-Base BCI Implementation. 10-th International Conference on Pattern Recognition and Image Analysis: New Information Technologies, 2010.

[8] Bobrov P., Frolov A., Cantor C., Fedulova I., Bakhnyan M., Zhavoronkov A.: Brain-computer interface based on generation of visual images Plos-Biology (in press).

[9] Brodu N., Lotte F., Lecuyer A.: Comparative Study of Band-Power Extracyion Techniques for Motor Imagery Classification, IEEE Symposium on Computational Intellignce, Codnitive Algorithms, Mind and Brain, 2011.

[10] Campbell A., Choudhury T., Hu S., Lu H., Mukerjee M., et al.: NeuroPhone: brain-mobile phone interface using a wireless EEG headset, ACM, 2010, pp. 3-8.

[11] Crammer K., Singer Y.: On the Algorithmic Implementation of Multiclass Kernel-based Vector Machines, Journal of Machine Learning Research, 2, 2004, pp. 265-292.

[12] Dornhege G., Blankertz B., Curio G., Muller K.: Increase information transfer rates in BCI by CSP extension to multi-class. Advances in Neural Information Processing Systems, 16, 2004, pp. 733-740.

[13] Emotiv Systems. Emotiv – brain computer interface technology. http://www.emotiv.com

[14] Faber J., Pekny J., Pieknik R., Tichy T., Faber V., Bouchner P., Novak M.: Simultaneous recording of electric and methabolic brain activity. Neural Network World, 20, 4, 2010, pp. 539-557.

[15] Finke A., Lenhardt A., Ritter H.: The MindGame: A P300-based brain-computer interface game. Neural Networks, 22, 2009, pp. 1329-1333.

[16] Georghiades A., Belhumeur P., Kriegman D.: From few to many: Illumination cone models for face recognition under variable lighting and pose. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 23, 2002, pp. 643-660.

[17] Grosse-Wentrup M., Buss M.: Multiclass common spatial patterns and information theoretic feature extraction. Biomedical Engineering, IEEE Transactions on, 55, 2008, pp. 1991-2000.

[18] Haynes J., Rees G.: Decoding mental states from brain activity in humans. Nature Reviews Neuroscience, 7, 2006, pp. 523-534.

[19] Krepki R., Curio G., Blankertz B., Muller K.: Berlin Brain-Computer Interface – The HCI communication channel for discovery. International Journal of Human-Computer Studies, 65, 2007, pp. 460-477.

[20] Leuthardt E., Schalk G., Roland J., Rouse A., Moran D.: Evolution of brain-computer interfaces: going beyond classic motor physiology. Neurosurgical Focus, 27, 2009, E4.

[21] Mellinger J., Schalk G., Braun C., Preissl H., Rosenstiel W., Birbaumer N., Kubler A.: An MEG-based brain-computer interface (BCI). Neuroimage, 36, 2007, pp. 581-593.

[22] Millan J., Mourino J., Marciani M., Babiloni F., Topani F., et al.: Adaptive brain interfaces for physically-disabled people, Citeseer, 1998, pp. 2008-2011.

[23] Millan J., Rupp R., Muller-Putz G., Murray-Smith R., Giugliemma C., et al.: Combining Brain-Computer Interfaces and Assistive Technologies: State-of-the-Art and Challenges. Frontiers in Neuroscience, 4, 2010.

[24] Nijholt A.: BCI for games: A 'state of the art' survey. Entertainment Computing-ICEC, 2008, pp. 225-228.

[25] Nikolaev A., Ivanitskii G., Ivanitskii A.: Reproducible EEG alpha-patterns in psychological task solving. Human Physiology, 24, 1998, pp. 261-268.

[26] Neuper C., Scherer R., Wriessnegger S., Pfurtscheller G.: Motor imagery and action observation: modulation of sensorimotor brain rhythms during mental control of a brain-computer interface. Clinical Neurophysiology, 120, 2009, pp. 239-247.

[27] Pfurtscheller G., Flotzinger D., Kalcher J.: Brain-computer Interface – a new communication device for handicapped persons. Journal of Microcomputer Applications, 16, 1993, pp. 293-299.

[28] Pfurtscheller G., Neuper C., Flotzinger D., Pregenzer M.: EEG-based discrimination between imagination of right and left hand movement. Electroencephalography and Clinical Neurophysiology, 103, 1997, pp. 642-651.

[29] Pfurtscheller G., Brunner C., Schlogl A., Lopes da Silva F.: Mu rhythm (de) synchronization and EEG single-trial classification of different motor imagery tasks. NeuroImage, 31, 2006, pp. 153-159.

[30] Ramoser H., Muller-Gerking J., Pfurtscheller G.: Optimal spatial filtering of single trial EEG during imagined hand movement. IEEE Transactions on Rehabilitation Engineering, 8, 2000, pp. 441-446.

[31] Schanze T., Eckhorn R.: Phase correlation among rhythms present at different frequencies: spectral methods, application to microelectrode recordings from visual cortex, and functional implications. Int. J. Neurophysiol., 26, 1997, pp. 171-189.

[32] Shotton J., Winn J. M., Rother C., Criminisi A.: Texton-Boost: Joint appearance, shape and context modeling for multiclass object recognition and segmentation. In Proceedings of the European conference on computer vision, 2006, pp. 1-15.

[33] Sitaram R., Zhang H., Guan C., Thulasidas M., Hoshi Y., Ishikawa A., Shimizu K., Birbaumer N.: Temporal classification of multichannel near-infrared spectroscopy signals of motor imagery for developing a brain-computer interface. NeuroImage **34**, 4, 2007, pp. 1416-1427.

[34] Weiskopf N., Veit R., Erb M., Mathiak K., Grodd W., Goebel R., Birbaumer N.: Physiological self-regulation of regional brain activity using real-time functional magnetic resonance imaging (fMRI): methodology and exemplary data. Neuroimage, 19, 2003, pp. 577-586.

[35] Wolpaw J., Birbaumer N., Heetderks W., McFarland D., Peckham P., et al.: Brain-computer interface technology: a review of the first international meeting. IEEE Transactions on Rehabilitation Engineering, 8, 2000, pp. 164-173.

[36] Wolpaw J., Birbaumer N., McFarland D., Pfurtscheller G., Vaughan T.: Brain-computer interfaces for communication and control. Clinical Neurophysiology, 113, 2002, pp. 767-791.

[37] Wolpaw J., McFarland D.: Control of a two-dimensional movement signal by a noninvasive brain-computer interface in humans. Proceedings of the National Academy of Sciences of the United States of America, 101, 2004, pp. 17849-17854.

[38] Zhao Q., Zhang L., Cichocki A.: Multilinear generalization of Common Spatial Pattern, IEEE, 2009, pp. 525-528.