



INDEPENDENT EEG COMPONENTS ARE MEANINGFUL (FOR BCI BASED ON MOTOR IMAGERY)

Y.V. Kerechanin*, P.D. Bobrov*, A.A. Frolov*, D. Húsek†

Abstract: Eight methods of decomposition of a multichannel EEG signal are compared in terms of their ability to identify the most physiologically significant components. The criterion for the meaningfulness of a method is its ability to reduce mutual information between components; to create components that can be attributed to the activity of dipoles located in the cerebral cortex; find components that are provided by other methods and for this case; and, at the same time, these components should most contribute to the accuracy of the BCI based on imaginary movement. Independent component analysis methods AMICA, RUNICA and FASTICA outperform others in the first three criteria and are second only to the common spatial patterns method in the fourth criterion. The components created by all methods for 386 experimental sessions of 27 subjects were combined into more than 100 clusters containing more than 10 elements. Additionally, the components of the 12 largest clusters were analyzed. They have proven to be of great importance in controlling BCI, their origins can be modeled using dipoles in the brain, and they have been detected by several degradation methods. Five of the 12 selected components have been identified and described in our previous articles. Even if the physiological and functional origins of the rest of identified components are to be the subject of further research, we have shown that their physiological nature is at least highly probable.

Key words: *EEG analysis, independent component analysis, ICA, common spatial patterns, CSP, principal component analysis, PCA, brain computer interface, BCI, features selection*

Received: September 11, 2020

DOI: 10.14311/NNW.2021.31.020

Revised and accepted: October 30, 2021

1. Introduction

Brain-computer interface (BCI) is a soft-and hard-ware complex that allows to control external technical devices directly by signals of the brain without commonly used muscle activity. In recent years, motor imagery based BCI systems have

*Yaroslav Kerechanin – Corresponding author, Pavel D. Bobrov, Alexander A. Frolov; The Institute of Higher Nervous Activity of RAS, 5A Butlerova st., Moscow RU-117485, Russia Moscow, Russian Federation, E-mail: leeleeekelee@gmail.com, p-bobrov@yandex.ru

†Dušan Húsek; Institute of Computer Science of the Czech Academy of Sciences, Pod Vodárenskou věží 271/2, CZ-182 07 Praha 8, Czech Republic, E-mail: Dusan.Husek@cs.cas.cz

become widespread in the motor rehabilitation of patients after a stroke or a trauma [1, 10, 21, 22]. However, as shown [16], BCI technology yields great benefits not only for practical application, but it can also make a major contribution to brain research. Via biological feedback the subjects train to stabilize and contrast brain activity patterns corresponding to specific mental tasks in BCI sessions [33]. This provides more reliable recognition and extraction of such patterns and facilitates better understanding of brain functionality while performing relevant mental tasks. The two main tools that we use in our study to research brain activity during motor imagery BCI performance are the blind source separation (BSS) methods, mainly the independent component analysis (ICA) methods, and solution of the inverse EEG problem, based on individual geometry of brain and its covers [16].

The ICA method, as all BSS methods understands multichannel EEG signal measured on the scalp as a superposition of brain activity signals (components), each of them having a specific placement on gray matter, and its activity is time-dependent.

In recent years, the ICA methods have become widely used in EEG processing, especially in BCI studies [26]. Leading researchers in this field – S. Makeig and his group, argue that the ICA methods, combined with the solution of the inverse EEG problem, “should bring EEG once again to the forefront of brain imaging, merging its high time and frequency resolution with enhanced cm-scale spatial resolution of its cortical sources” [31].

However, there is still the question of whether the estimated components reflect the specific physiological properties of brain activity or rather the mathematical properties of the ICA method used, where the criterion of statistical independence (SI) is an important parameter.

This problem is relevant due to the obvious noisy nature of many independent components, thus early ICA applications to EEG processing were only used to eliminate noise [9].

In EEG analysis it is supposed that every physiologically meaningful component corresponds to a singular current dipole [11] on the brain. Current dipoles are a common tool for modeling sources of brain electrical activity recorded from the surface of the scalp.

The dipoles have orientation and location on the brain, while component activity corresponds to a dipole moment varying in time.

The suggestion that each independent component corresponds to single current dipole is based on the anatomical knowledge that short-range connections between cortex neurons are much more dense than long-range connections. Thus, locally synchronized activity in cortex areas of a few square millimeters (which is, for example, the estimated size of the cortex area responsible for sensorimotor mu-rhythm generation [25]) can be considered as independent of activities of distant areas of the same size.

Based on the foregoing, in our previous works [16, 18] we classified the components of ICA as physiologically significant based on two main criteria: 1) A component can be generated by a current dipole model, 2) Component is significant for BCI control efficiency. Moreover, we checked on a sample of subjects using functional magnetic resonance imaging (fMRI) that in the area where the dipole should take place, an increase in hemodynamic activity appeared.

The main goals of the present study are, first, to evaluate the ability of the eight methods of the multi-channel EEG signal decomposition methods to extract physiologically significant components from multichannel EEG signal, and second to estimate the physiological significance of components in terms of BCI efficiency.

The ability, of a total of 22 linear decomposition algorithms (20 ICA or blind sources separation (BSS) algorithms plus Principal Component Analysis (PCA) plus sphering decomposition), to perform physiologically plausible EEG decomposition was evaluated by A. Delorme et al. in [11] by three criteria: 1) mutual information reduction (MIR), 2) pairwise mutual information reduction and 3) the method's ability to extract from EEG signal the largest amount of dipolar components.

In the present study we evaluate the ability of 7 BSS methods (5 ICA, second-order blind identification (SOBI) and PCA) plus a common spatial patterns (CSP) method to perform meaningful EEG decomposition.

We evaluate these methods in accordance with the degree to which they provide meaningful components in terms of the four criteria 1) Dipolarity i.e. the ability create components that can be assigned to the activities of dipoles, 2) MIR i.e. the ability reduce mutual information between components; 3) Significance i.e. the component is significant for BCI control efficiency. 4) Physiological significance – component repeatedly appears in different subjects and in different relations of one subject and is detected simultaneously by several different ICA methods.

In order to be able to quantify the above criteria, we used the indicators and other tools that will be introduced in the Section 2.3 “Indicators”.

Next, the results of the analysis we used for the selection of the most physiologically significant components, which, according to the analysis, are also the optimal features for the BCI classifier, see Section 3.2 “The components most suitable to be treated as physio-logically meaningful”

We also compare our results with the relevant results of A. Delorme [11]. All the BSS methods considered here were included in the analysis carried out in this work. In contrast to this, we extended our analysis by adding the CSP method, because its application is the most common [2] in BCI classifiers. Note that both criteria MIR and dipolarity are common in both of these evaluations. However, let us not forget, that the main difference between the two analyses is that the experimental data analyzed in this study correspond to completely different mental tasks, as compared to [11]. In our study, motor imagery data are analyzed, while in [11] subjects performed a modified Sternberg visual working memory task [30].

2. Methods

2.1 Experimental procedure

Twenty-seven subjects (20 male and 7 female), aged between 21 and 36 without visible physical and neurological impairment, participated in the study. All the subjects were acquainted with the experimental protocol and gave their written consent to participate in the experiment. The experimental protocol was approved by the board of ethics at the Institute for the Higher Nervous Activity and Neurophysiology RAS.

There were between 10 and 20 experimental sessions of BCI training for each subject, 1-2 sessions a day. The gap between experimental days was from 1 to 4 days. The subjects had to execute one of 3 instructions presented on the computer screen: to relax or to imagine kinesthetically left or right fist extension. The subjects had to perform one of the 3 instructions presented on the computer screen: to relax or to imagine opening their left or right fist kinesthetically.

Each instruction was presented for 10 seconds. Two instructions for right arm and 2 for left arm were presented in a random order and formed a block. Each session contained 10 such blocks. Each MI instruction was preceded by instruction to relax. More detailed description of the experimental protocol is given in [5].

The Bayesian classifier described in [7], which has much lower computational costs, but the classification accuracy comparable to other more complex classifiers [15], was used in our BCI implementation.

EEG recording was carried out by electroencephalograph gUSBamp (g-tec, Austria) with 48 active electrodes and 256 Hz sampling frequency. Before EEG signal decomposition recordings were filtered by 5-30 Hz pass-band filter.

2.2 EEG signal decomposition

All the evaluated methods represent multichannel EEG signal \mathbf{X} (where components of \mathbf{X} are electric potentials on N individual electrodes on the scalp/head surface) as a weighted superposition of N component activities ξ_i

$$\mathbf{X} = \mathbf{W}\Xi = \mathbf{w}_1\xi_1 + \mathbf{w}_2\xi_2 + \dots + \mathbf{w}_N\xi_N. \quad (1)$$

In Eq. (1) \mathbf{X} is N by M matrix, its each row corresponds to a single electrode time series and each column is an EEG signal at some time moment. Thus M , is equal to the whole number of EEG times samples. Column vector \mathbf{w}_i (N by 1) represents i -th component potential distribution over EEG electrodes, let us call it topographic map (TM), and row vector ξ_i (M by 1) gives i -th component activity at all moments of time.

2.2.1 ICA BSS methods

Before matrix \mathbf{X} decomposition the whitening procedure is usually used, which transforms \mathbf{X} by the equality $\mathbf{Z} = \mathbf{V}\mathbf{X}$ so that $\mathbf{E}\{\mathbf{Z}\mathbf{Z}^T\} = \mathbf{I}$, where \mathbf{V} is the whitening matrix and $\mathbf{E}\{\cdot\}$ denotes signal expectation. Thus, due to whitening \mathbf{X} becomes decorrelated. It can be shown [24] that \mathbf{V} can be found from the covariation matrix $\mathbf{C} = \mathbf{E}\{\mathbf{X}\mathbf{X}^T\}$: $\mathbf{V} = \mathbf{C}^{-1/2}$. Due to whitening, Eq. (1) transforms to:

$$\mathbf{Z} = \mathbf{U}\xi, \quad (2)$$

where $\mathbf{U} = \mathbf{V}\mathbf{W}$ and ξ_k are uncorrelated and of a unit variance. In contrast to matrix \mathbf{W} , \mathbf{U} is orthonormal.

Component activities Ξ can be found by inversion of Eq. (2):

$$\Xi = \mathbf{A}\mathbf{Z}, \quad \mathbf{A} = \mathbf{U}^T. \quad (3)$$

thus

$$\mathbf{W} = (\mathbf{A}\mathbf{V})^{-1}. \quad (4)$$

Since \mathbf{U} is orthonormal then \mathbf{A} is also orthonormal, which greatly simplifies the solution of the ICA problem.

Five ICA methods were used for EEG decomposition.

KURT [24] is based on maximizing deflection of component kurtosis from the kurtosis of the normal distribution which is equal to 3

$$\mathbf{a}_i = \arg \max_{\mathbf{v}, \|\mathbf{v}\|=1} \left| \frac{E\{(\mathbf{v}\mathbf{Z})^4\}}{[E\{(\mathbf{v}\mathbf{Z})^2\}]^2} - 3 \right|,$$

where the row vector \mathbf{a}_i is supposed to be orthogonal to all previously found vectors \mathbf{a}_j ($j < i$). When all vectors \mathbf{a}_i are found, \mathbf{W} can be found from Eq. (4).

Since distribution of a superposition of independent signals is closer to normal than distributions of signals themselves then deflection of component kurtosis from 3 is a good index that components are independent. This method evidently fails if independent signals are normally distributed. However, this is a common property of all the ICA methods that do not work with a normal distribution of independent components.

FASTICA [24] is based on minimization of mutual information between component activities. Mutual information I between activities $\xi_1, \xi_2, \dots, \xi_k$ is defined as:

$$I(\xi_1, \xi_2, \dots, \xi_k) = \sum_l H(\xi_l) - H(\Xi), \tag{5}$$

where H is the entropy of signals and $\Xi = \mathbf{AZ}$. For any linear transformation of \mathbf{Z} , $H(\Xi) = H(\mathbf{Z}) + \det|\mathbf{A}|$. Since

$$\det(E\{\Xi\Xi^T\}) = \det(E\{\mathbf{AZZ}^T\mathbf{A}^T\}) = \det(\mathbf{A})^2 \det(E\{\mathbf{ZZ}^T\})$$

and matrices $E\{\Xi\Xi^T\}$ and $E\{\mathbf{ZZ}^T\}$ are orthonormal, then $|\det(\mathbf{A})| = 1$ and, consequently, $H(\Xi)$ does not depend on \mathbf{A} and the minimization of mutual information is equivalent to finding \mathbf{a}_i which minimizes the entropy $H(\xi_i)$:

$$\mathbf{a}_i = \arg \min_{\mathbf{v}, \|\mathbf{v}\|=1} \{H(\mathbf{v}\mathbf{Z})\}.$$

Exact calculation of the entropy is practically unavailable, and different methods of its estimation are used. In this particular case, the method approximates the entropies of component activities as follows:

$$H(\xi_i) \propto \text{const} - [E\{g(\xi_i)\} - E\{g(\nu)\}]^2, \tag{6}$$

where $g(\cdot) = \log \cosh(\cdot)$ and ν is Gaussian variable of zero mean and unit variance.

Thus, each row vector of matrix \mathbf{A} is found by the equation:

$$\mathbf{a}_i = \arg \max_{\mathbf{v}, \|\mathbf{v}\|=1} \{E\{g(\mathbf{v}\mathbf{Z})\}\},$$

where, as in the KURT method, \mathbf{a}_i is supposed to be orthogonal to all previously found vectors \mathbf{a}_j ($j < i$).

RUNICA [3] is a modification of FASTICA taking into account that two different functions for estimation of entropy are required for super-Gaussian (kurtosis > 3) and sub-Gaussian (kurtosis < 3) components. For super-Gaussian components, $g_+(\cdot) = -2 \log \cosh(\cdot)$, for sub-Gaussian, $g_-(\cdot) = \log \cosh(\cdot) - (\cdot)^2/2$.

AMICA [32] represents each component ξ_i as a mixture ($j = 1, \dots, K$) of supergaussian signals. The component activity distribution $p(\xi_i(t))$ is given by following formulae:

$$p(\xi_i(t)) = \sum_{j=1}^K \alpha_{ij} \cdot q(\xi_i(t); \rho_{ij}, \mu_{ij}, \beta_{ij}),$$

$$q(\xi_i(t); \rho, \mu, \beta) = \frac{\rho}{2\beta\Gamma(1/\rho)} \exp\left(-\left|\frac{\xi_i(t) - \mu}{\beta}\right|^\rho\right).$$

Where $q(\xi_i(t))$ is super-Gaussian probability density function and Γ is gamma function. In this regard, for each component activity ξ_i several parameters must be estimated in general. First, these are: mean μ_{ij} , scale β_{ij} , and shape ρ_{ij} of super-Gaussian probability density function q_{ij} . The next one are α_{ij} , which indicates the relative contribution of each distribution q_{ij} to the component activity distribution. For α_{ij} , the following equality must hold: $\sum_{j=1}^K \alpha_{ij} = 1$.

When $\mu_{ij} = 0$, $\beta_{ij} = 1$ and $\rho_{ij} = 2$, q_{ij} represents zero-mean Gaussian distribution with unit variance. In our calculations each component activity distribution $p(\xi_i)$ was approximated by a mixture of $K = 2$ super-Gaussian ($\rho > 2$) distributions. Likelihood maximization was used to estimate parameters of these mixtures.

CUMUL [24] is similar to the KURT method, but instead of kurtosis the maximization of the 4-th order cross-cumulant is used:

$$\text{cum}_4(y, \tau) = E\{(y(t)^2 y(t - \tau)^2)\} - E\{y(t)^2\}E\{y(t - \tau)^2\} - 2E\{y(t)y(t - \tau)\}^2. \quad (7)$$

In the present paper, the value of time shift τ was set to 80 ms, which is equal to the period of 12 Hz signal. Thus, the method is intended to extract components that have alpha-rhythm frequency.

2.2.2 Other BSS methods

In addition to the ICA methods we used two other BSS methods: second-order blind identification (SOBI) and PCA.

SOBI [4] is based on diagonalization of a set of time-shifted cross-covariance matrices. The function estimates the unmixing matrix in a second order stationary source separation model by jointly diagonalizing the covariance matrix and several autocovariance matrices at different lags.

$$E\{\xi(t)\xi(t - \tau)\} = \mathbf{A}\mathbf{C}_k\mathbf{A}^T,$$

where $\mathbf{C}_k = E\{\mathbf{X}(t)\mathbf{X}^T(t - \tau)\}$, $\tau = k\delta$, $k = 0, \dots, K$ and δ is a sampling step. Thus, SOBI minimizes correlation between component activities in a given range of time shifts:

$$\mathbf{A} = \arg \min \sum_k f(\mathbf{A}\mathbf{C}_k\mathbf{A}^T),$$

where for arbitrary $N \times N$ matrix \mathbf{Y} , $f(\mathbf{Y})$ is defined as

$$f(\mathbf{Y}) = \sum_{1 \leq i, j \leq N, i \neq j} |Y_{ij}|^2.$$

PCA implies that in Eq. (1) columns of \mathbf{W} are eigenvectors of covariance matrix $\text{cov}(\mathbf{X})$, thus matrix \mathbf{W} is orthonormal, component activities ξ_i are not correlated, and their variances are equal to eigenvalues of covariance matrix.

2.2.3 Non BSS method recognized in BCI

CSP does not belong to BSS methods family, we, as mentioned in Section 1 incorporated this method in our study, as it is the most frequently used for BCI control [2]. We used original CSP, which is intended for separation of EEG signals only into two classes. To eliminate this drawback, we propose procedure given below. According to CSP, matrix $\mathbf{A} = \mathbf{W}^{-1}$ has to satisfy two equations:

$$\mathbf{A}\mathbf{C}_1\mathbf{A}^T = \mathbf{D}^{(1)} \quad \mathbf{A}(\mathbf{C}_1 + \mathbf{C}_2)\mathbf{A}^T = \mathbf{I}, \quad (8)$$

where \mathbf{C}_i are covariance matrices for two mental states ($i = 1, 2$), $\mathbf{D}^{(1)}$ is a diagonal and \mathbf{I} is a unit matrix. Then in both mental states component activities are not correlated and their variances are given by $\mathbf{D}^{(1)}$ for the first and $\mathbf{D}^{(2)} = \mathbf{I} - \mathbf{D}^{(1)}$ for the second mental state. Therefore, if variance is small for one state it is large for another. Thus, each row vector \mathbf{a}_i of matrix \mathbf{A} defines the direction \mathbf{a}_i^T in the signal space so that projection of \mathbf{X} on this direction maximizes or minimizes the ratio of projection variances $d_{ii}^{(1)}/d_{ii}^{(2)}$ in two mental states. Maximal difference in these variances determines the high ability of CSP for state recognition during BCI control.

Matrix \mathbf{A} can be found explicitly. It is easy to prove that both equations (8) satisfy if $\mathbf{A} = \mathbf{U}_1^T \mathbf{D}^{-\frac{1}{2}} \mathbf{U}^T$, where \mathbf{U} is a unitary matrix and \mathbf{D} is a diagonal matrix given by singular value decomposition (SVD) of

$$\mathbf{C}_1 + \mathbf{C}_2 = \mathbf{U}\mathbf{D}\mathbf{U}^T$$

and \mathbf{U}_1 is a unitary matrix given by SVD of

$$\mathbf{D}^{-\frac{1}{2}} \mathbf{U}^T \mathbf{C}_1 \mathbf{U} \mathbf{D}^{-\frac{1}{2}} = \mathbf{U}_1 \mathbf{D}^{(1)} \mathbf{U}_1^T.$$

The definition of CSP is made under the assumption that there are two classes (states). Since there were three tasks (classes) in our experiments, the CSP decomposition was performed for each pair of the tasks, resulting in three sets of components for each of the BCI-control sessions.

2.3 Indicators

2.3.1 Mutual information reduction

Mutual information reduction (MIR), i.e. the amount of mutual information removed from set of channels, is [11] given by:

$$\text{MIR} = I(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N) - I(\xi_1, \xi_2, \dots, \xi_N),$$

where mutual information I is given by (5). Then

$$\text{MIR} = \sum_i H(x_i) - H(\mathbf{X}) - \sum_i H(\xi_i) + H(\Xi).$$

Since $\mathbf{X} = \mathbf{W}\Xi$, then

$$H(\mathbf{X}) = H(\Xi) + \log \det |\mathbf{W}|.$$

Hence

$$\text{MIR} = \sum_i H(x_i) - \sum_i H(\xi_i) - \log \det |\mathbf{W}|. \quad (9)$$

Just reminder: Zero mutual information between two random variables means that the variables are independent, so as a result, MIR is an indicator of a method's efficiency.

2.3.2 Dipolarity

The majority of extracted components are noisy components, originated from eye movements, blinking, myogram, head movements and loss of contact between electrodes and skin. These components were treated as noisy if it was impossible to approximate their TMs with single current dipole potential distribution. Each TM was matched to a single current dipole located in brain cortex. The best-fitting dipole approximation for the component TM \mathbf{w} was found using finite element head model created using the standardized MRI head image provided by the Montreal Neurological Institute (MNI). The model allows to evaluate the potential distribution at the electrodes, which results from a single dipolar current source located at the center of each of the model elements. Thus, for each of the elements it is possible to find the best dipolar fit TMD_i for \mathbf{w} . The best-fitting distribution can then be found as:

$$\text{TMD} = \arg \min_{i=1, \dots, N_{el}} (\text{TMD}_i),$$

where N_{el} denotes the total number of elements in the model. Goodness of fit can be measured by residual variance:

$$\varepsilon_{\text{dip}} = \frac{\|\mathbf{w} - \text{TMD}\|^2}{\|\mathbf{w}\|^2}$$

Components were treated as noisy if residual variance of dipole approximation was higher than 10%. It should be noted that eye movement and blink components can be fitted by a single dipole with high precision but they were also removed as noisy because their dipole location was not in the brain cortex. As a result of such filtration, only a third of the whole components were treated as “dipolar”.

To evaluate the ability of the method to provide dipolar components, we introduce indicator dipolar components ratio:

$$ND = \frac{N_{\text{dip}}}{N_{\text{all}}}, \quad (10)$$

where N_{dip} is the number of dipole components and N_{all} is the number of components identified during all the sessions of all subjects.

2.3.3 Methods similarity indicator

All the methods used to decompose the EEG signal are clearly based on different approaches, as described above in Section 2.2, they may be implicitly similar. Thus, it is reasonable to evaluate their internal ability to give similar results. Therefore we introduce similarity indicator MS:

$$MS = n_s / (n_1 + n_2 - n_s) \quad (11)$$

where n_s is a number of identical components, n_i is the total number of components revealed by the i -th method. If there are no identical components $MS = 0$, if all components are identical $MS = 1$.

2.3.4 Component rank indicator

Different methods are based on different assumptions, so it is reasonable to suppose that a component that is extracted by several methods contains information about the EEG signal and it is not an artefact of the method. Thus, in addition to the dipolarity criterion, only components that were extracted in each experimental session at least by two decomposition methods were considered physiologically significant and used for analysis. As a measure of this, we introduce component rank (R) indicator:

$$R = N_{m, \xi_i} \quad (12)$$

where R is component rank and N_{m, ξ_i} number of methods that extracted the same component.

2.3.5 Components identity criterion

As mentioned earlier, the resulting components are never exactly the same, so we must introduce the criterion of component identity. It is supposed that components extracted by different methods are identical if their TM, i.e. \mathbf{w}_1 and \mathbf{w}_2 and activities ξ_1 and ξ_2 respectively, obtained by two methods, 1 and 2 are close. The closeness of TMs activities is given by the formulae:

$$S_{1,2} = |\mathbf{w}_1^T \mathbf{w}_2| / (|\mathbf{w}_1| |\mathbf{w}_2|), \quad (13)$$

and closeness ξ_1 and ξ_2 by the formulae:

$$C_{1,2} = |\xi_1 \xi_2^T| / \sqrt{(\xi_1 \xi_1^T)(\xi_2 \xi_2^T)}. \quad (14)$$

It was supposed that components are identical if $S_{1,2} > 0.9$ and $C_{1,2} > 0.8$. According to this two criteria approximately half of the total components had rank $R > 1$.

Components that were recognized as identical were combined into a single component with mean TM, and mean activity.

2.3.6 The selection of components most relevant for BCI control

To evaluate the relevance of the components for BCI control we used the approach described in [16, 18]. According to this approach, the components belonging to the component combination providing the highest efficiency of BCI control were treated as the most relevant. For all the methods except CSP the components are activities $\xi_i, i = 1, \dots, N$ directly obtained by EEG decomposition Eq. (1).

For CSP, the components are obtained by combining the components obtained by 3 decompositions applied to each pair of mental states. For each pairwise decomposition 6 components with highest and 6 with smallest eigenvalues were taken and then combined. The optimal combination of components was selected from these 36 combined components.

The efficiency of BCI control was evaluated off-line by cross-validation of each experimental session. Ten blocks of a session were randomly split into sets of seven and three blocks. The first set was used for BCI classifier training and the second one for its testing. Thus, non-overlapping data sets were used for training and testing of the classifier.

We used a Bayesian classifier that compare covariance matrices for three mental tasks performed by the subject in accordance with the instructions [7, 15]. The BCI training actually consists in calculating the three EEG covariance matrices for each of the mental tasks on data from the selected seven blocks. Missing features are therefore taken into account automatically compared to [29].

The classifier was tested by extracting epochs lasting one second from the three remaining blocks and calculating the covariance matrix for each epoch. The classifier selected the task for which a covariance matrix was closest to that calculated for the epoch [7, 15]. As a result of averaging over 50 such splits, the confusion matrix $\mathbf{P} = p_{ij}$ was obtained, where p_{ij} is the probability estimate for recognizing the i -th mental task if the j -th mental task is to be performed. The efficiency of BCI control was evaluated by Cohen's kappa [28] indicator. Given the confusion matrix \mathbf{P} , this indicator is calculated as

$$\kappa = \frac{\sum_i p_{ii} p_{i0} - \sum_i p_{0i} p_{i0}}{1 - \sum_i p_{0i} p_{i0}}, \quad (15)$$

where p_{0j} is a probability of the j -th instruction to be presented and $p_{i0} = \sum_j p_{ij} p_{0j}$ is a 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. Thus, p_{0j} were equal or very close to $1/3$. The better the classifier performs the more the confusion matrix \mathbf{P} is close to identity matrix \mathbf{I} . If classification is perfect ($\mathbf{P} = \mathbf{I}$) and the $\kappa = 1$. If classification is random, i.e. $p_{ij} = p_{i0}$ for all j , then $\kappa = 0$.

For each experimental session and each method of signal decomposition, an optimal combination of components was found, where the value of the index κ was considered as a criterion of optimality. The higher the κ , the better.

Because the total number of possible combinations of 2^N components is extremely large, we have found an optimal combination of only 3 components using an exhaustive search first.

Then, in finding the optimal set of components, we used a “greedy” algorithm that added more components one by one, starting with these 3 components. In each step, a new component was added that provided the maximum increment to the coefficient κ . The significance of each component was evaluated as the frequency of cases where such a component was included in the optimal combination.

2.3.7 Topographic map clustering

Functionally identical components, when they are identified on different subjects, never have identical TM, and this fact is true even if these components are identified at different experimental sessions, where the subject participating in the experiment remains the same. This is due to anatomical differences in the structure of the subjects’ brains, different electrode settings during the sessions, and specific noise in each session. However, it can be assumed that these TMs maps should be at least somehow similar. This task does not have an unambiguous solution because the maps smoothly transit into each other, so boundaries of clusters depend on the clustering method. Due to this peculiarity we used attractor neural network with increasing activity (ANNIA) as a clustering method [13, 14, 17]. The method was elaborated for the Boolean factor analysis. According to this method, each TM which satisfied dipolarity and rank criteria was matched with a single binary neuron whose states were 1 (active) or 0 (not active). Neurons were connected into the neural network by “synaptic” connections equal to similarities between corresponding TMs. The similarity S_{ij} between TMs i and j (that is the synaptic connection between neurons i and j) was specified by Eq. (13) as $S_{i,j}$ [8]. To reveal each cluster we used the two-run procedure. The network dynamics was initialized by the presentation of a random initial binary pattern of the network activity $\mathbf{X}(0)$ at the moment $t = 0$. The number of active neurons k_{init} in the initial pattern was taken to be much smaller than the supposed size of the cluster. On the presentation of $\mathbf{X}(0)$, network activity evolves to an attractor. This evolution is determined by the synchronous discrete time dynamics. At each time step

$$X_i(t+1) = \Theta(h_i(t) - T(t)), \quad i = 1, \dots, N_n,$$

where h_i are components of the vector of synaptic excitations $\mathbf{h}(t) = \mathbf{S}\mathbf{X}(t)$, Θ is a step function, T is activation threshold, and N_n the number of neurons in the network.

At each time step of the searching process the threshold $T(t)$ was taken to provide constant number k_{init} of active neurons in $\mathbf{X}(t+1)$. As shown in [12], this choice of the activation threshold provides the permanent increase of the Lyapunov function

$$\Lambda = \mathbf{X}^T \mathbf{S} \mathbf{X}$$

until the network activity stabilizes in point or cyclic attractor of length two. The Lyapunov function actually corresponds to the total sum of connections between active neurons. Thus, the searching process with a fixed number of active neurons provides the choice of neurons with maximal connections, i.e. TMs with maximal similarities.

When activity stabilizes at the initial level k_{init} a new neuron with maximal synaptic excitation is added to the neurons active at the previous step and network activity evolves to attractor at the new level of activity and so on until the number of active neurons reaches some value k_{fin} . Thus, to find one cluster, $k_{\text{fin}} - k_{\text{init}}$ external steps and several steps inside each external step required to reach an attractor for a given level of activity.

Usually we begun the recall process starting from 5 randomly chosen active neurons and ended it when the mean value of synaptic connections between neurons in the cluster decreases to 95% of its maximal value over the neurodynamic trajectory. After one cluster was found, all its neurons were deleted from the network and new searching procedure was initiated, while the network was not depleted.

2.3.8 Rate of appearance indicator

In the introduction, however, we set a tougher criterion of physiological significance – the component has to appear repeatedly in different subjects and in different relations of one subject. In addition, this pattern should be detected simultaneously by several different ICA methods simultaneously.

So we introduced indicator rate of appearance indicator:

$$\text{RA} = N_{\text{clust}}/N_{\text{sessions}} \cdot 100 \quad [\%], \quad (16)$$

where N_{clust} is the total number of component TM in the cluster associated with this component, and N_{sessions} is the total number of all experimental sessions.

3. Results

Experimental data include 386 sessions for 27 subjects, approximately 10 – 20 sessions for each subject. Each session was processed by 6 ICA, 2 other BSS methods and CSP.

3.1 Methods comparison

The signal decomposition methods were compared according to the 6 main indicators described above: mutual information reduction MIR – Eq. 9, dipolar component extraction coefficient ND – Eq. 10, methods similarity MS – Eq. 11, component rank R see Eq. 12, then component significance hereupon κ – Eq. 15 is used as a BCI control quality criterion, and rate of appearance RA Eq. 16.

To evaluate the ability of the method to provide dipolar components (dipolarity) the ND indicator Eq. 10 was calculated.

Its value ranged from 0.36 for AMICA to 0.08 for PCA method – see Fig. 1 – Dipolar components extraction ratio. Therefore, dipolarity is highest for the ICA methods. The PCA method is the worst. This result agrees with those obtained

in [11] for the task related to visual memory. It is, therefore, reasonable to assume that they really reflect the property of the method, and do not depend on the mental task. In addition, we obtained three values for pairwise analysis of CSPs: CSP12, CSP13, and CSP23, which correspond to pairs of mental states: 1 for left-hand movement, 2 for right-hand, and 3 for relaxation. As these values are relatively close, instead of these three pairs of values, we give only their average value.

Mutual information reduction was also calculated for each method. The final numbers were obtained as an average over all the sessions of all the subjects. The results are shown in Fig. 1–MIR. Values ranged from 918 bps for AMICA to 897 bps for CUMUL, CSP and PCA. And once again, the highest MIR values are provided by ICA methods. And, again, the results are in good agreement with those presented in [11].

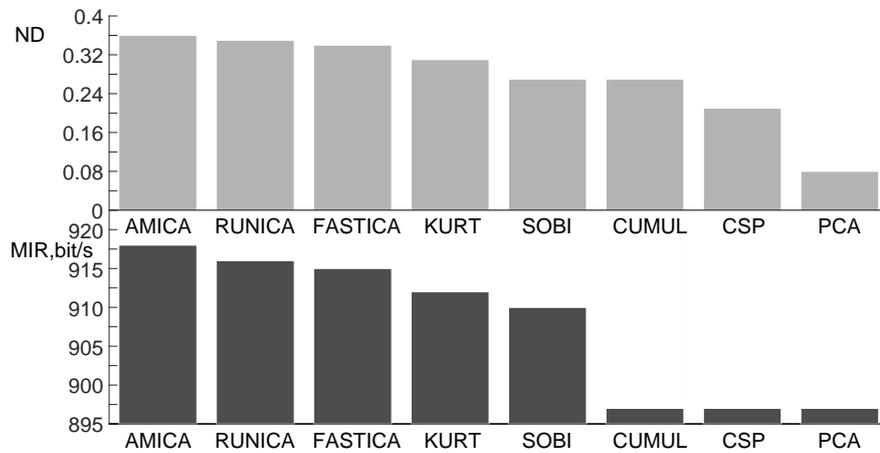


Fig. 1 Dipolar components extraction ratio ND and mutual information reduction MIR averaged over all the method.

Although all the methods used to decompose the EEG signal are clearly based on different approaches, as described in Section 2, they may be implicitly similar. Thus, before assessing the ability of the methods to create components with a high rank, it is reasonable to evaluate their internal ability to give similar results. The results of such assessment are given in Tab. I. Methods similarity was estimated by Eq. 11.

The MS values below the diagonal were obtained for all the components found. In this case, $n_1 = n_2 = N$, where N is the number of EEG electrodes. The MS values above the diagonal were obtained only for dipole components. In this case n_1 and n_2 are generally not equal and are smaller than N . MS were obtained by averaging over all the sessions of all the subjects.

The highest MS was obtained between AMICA and RUNICA in spite of the fact that these methods are based on completely different approaches as described

	AMICA	RUNICA	FASTICA	KURT	SOBI	CUMUL	CSP	PCA
AMICA		0.56	0.44	0.24	0.10	0.11	0.03	0.006
RUNICA	0.42		0.48	0.22	0.08	0.11	0.02	0.004
FASTICA	0.33	0.36		0.20	0.09	0.09	0.03	0.007
KURT	0.20	0.20	0.17		0.06	0.06	0.02	0.004
SOBI	0.08	0.07	0.07	0.05		0.02	0.02	0.01
CUMUL	0.08	0.08	0.07	0.05	0.02		0.007	0.002
CSP	0.02	0.02	0.02	0.01	0.01	0.003		0.002
PCA	0.004	0.004	0.004	0.003	0.005	0.001	0.001	

Tab. I Similarity of the MS methods in producing identical components when all the components (below diagonal) and only dipole components (above diagonal) are taken into account.

in Section 2 “Methods”. This is explained by their ability to reveal the largest amount of dipole components (see above) which are supposed to have physiological significance. This is also shown in Tab. I: for all components MS between AMICA and RUNICA amounts to 0.42, and for only dipole components – 0.56. Thus, the great similarity of the results of these methods reflects not their common internal property to reveal identical components but their ability to find the proper components corresponding to the signal nature. Since MS is higher for dipole components than for all the components also for other methods, this is true for all the methods. Hence, it is possible to calculate the rank of the components when ignoring the methods internal ability to produce identical components due to implicit similarity of the MIR criteria.

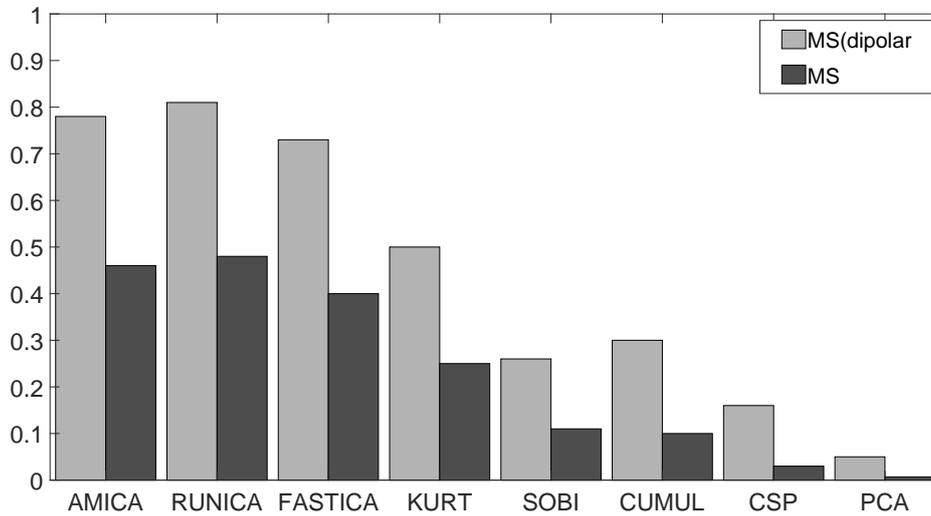


Fig. 2 index MS for all (light bars) and dipolar only (dark bars) components.

We estimated the global ability of the methods to produce components with a high rank by calculating the index MS by the Eq. (11) but where n_s is the number of identical components found by one of the methods and by all other methods. For all the components found (dipole and not dipole), the MS ranged from 0.48 for RUNICA to 0.007 for PCA. For only dipole components MS ranged from 0.81 to 0.05, respectively. Thus, in terms of dipolarity and MIR, the three methods AMICA, RUNICA and FASTICA are superior to the other methods in their ability to produce components with a high rank. For dipole components, each of them reveals almost the same number of components with the rank exceeding 2 as all the other components together.

The ability of the methods to produce components most relevant to control BCI is depicted in Fig. 3. The mean values of κ for each subject evaluated by the original multi-channel EEG signal (light bars) and optimal (darker bars) components are shown. The subjects in Fig. 3 are sorted over κ obtained by the original EEG signal. The figure demonstrates the wide subject distribution over BCI control quality that was noted earlier in [20] and is consistent with the results of other studies [6].

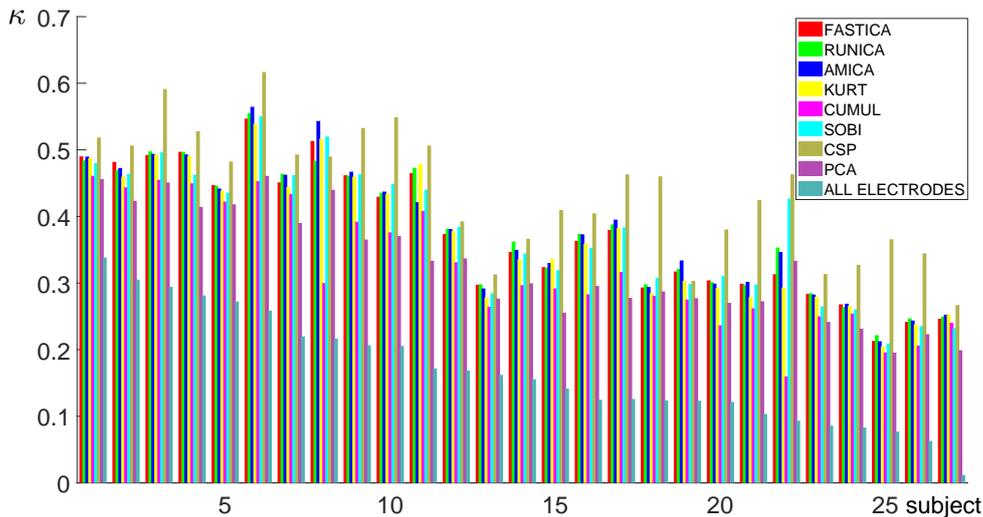


Fig. 3 Over all the sessions the mean BCI control quality index κ achieved by the subjects in depending on the signal decomposition method. The dark green bars indicate the quality of control achieved by the subjects based on the original EEG signal. The other colored bars are quality indicators for the optimal component sets obtained by different methods for a given subject. The correspondence of the decomposition method to the color of the bar is shown in the upper right part of the figure.

Generally BCI, control quality significantly increases when removing noise components. On average for all the subjects, κ increased from value 0.17 obtained using

the original EEG signal to the following values 0.38 for AMICA, 0.38 – RUNICA, 0.37 – FASTICA, 0.37 – KURT, 0.37 – SOBI, 0.33 – CUMUL, 0.4374 – CSP, 0.32 – PCA. The improvement compared to the value obtained for the original EEG signal amounted to 0.20 for AMICA, 0.20 – RUNICA, 20 – FAST ICA, 19 – KURT, 0.20 – SOBI, 0.16 – CUMUL, 0.264 – CSP, 0.15 – PCA 0.37 – for optimal components. ANOVA analysis did not reveal any statistically significant differences between κ values obtained by different ICA methods however, pairwise comparisons by Mann-Whitney test showed that the CUMUL method was inferior to the majority of the other ICA methods. The closeness of κ values obtained by different ICA method shows that wide distribution of subjects in BCI control quality most likely reflects their physiological differences rather than parameters of classification and EEG analysis methods.

The increase in the quality of control by choosing the optimal components compared to those obtained using the original EEG signal was the highest for CSP. This is not surprising since this method is the most effective in BCI for recognizing two mental states. Its modification to recognize the three states described in Section 2 “Methods” may be effective for a multiclass BCI, but it must be studied in real-time experiments.

3.2 The components most suitable to be treated as physiologically meaningful

After removing the components which do not satisfy the criteria of dipolarity and rank, approximately 10,000 components were left for further analysis. ANNIA clustering divided these TMs into 200 clusters larger than 10 in size.

Fig. 4 depict topoplots and spectral activity densities of the relative activities of components in the first 12 largest clusters, i.e. the components which most regularly appear in all experimental sessions of all the subjects.

Fig. 5 depict the main characteristics for the selected clusters: (A) – rate of appearance, i.e. the percentage of sessions where cluster components were extracted, (B) – the mean rank of cluster components, (C) – the percent of sessions where the cluster components were identified as significant, and D – the mean dipolarity of the cluster components. As shown, the components of the first 5 clusters appeared in more than 50 % of sessions and the 12-th cluster components – in 5 %. For almost all clusters, the mean value of component rank was higher then 3, ie on average these components were extracted by more than three decomposition methods. The components were identified as significant in more than 20 % sessions and dipolarity was less than 5 %. Thus, all components shown in Fig. 4 should be treated as physiologically significant.

4. Conclusions

A comparison was made of eight methods for decomposition of a multichannel EEG signal by the main indicators that determine their ability to extract physiologically significant components from the signal; mainly: 1) mutual information reduction, 2) dipolarity of the selected component, 3) the average number of methods retriev-

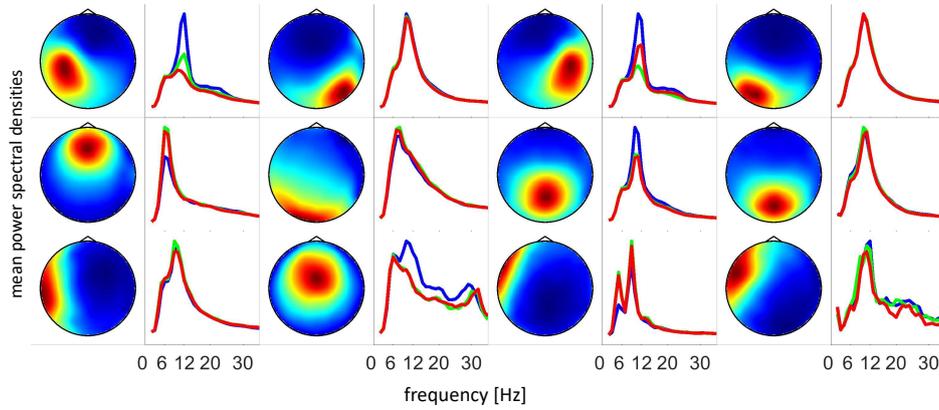


Fig. 4 Topoplots and spectral activity densities of components for the 12 largest clusters ordered by cluster size. Topoplots and relative spectral densities were averaged over all the components in each cluster. Spectral densities are shown in arbitrary units in the frequency range from 5 to 30 Hz for three tasks: blue lines – relax, red and green lines – imagination of right-hand and left-hand movement.

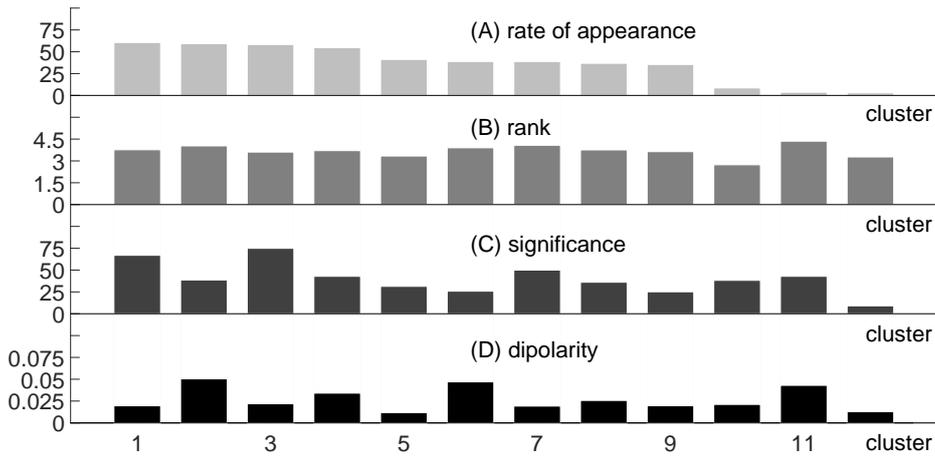


Fig. 5 Indicators of component meaningfulness for the first 12 components. *A* – rate of appearance RA. *B* – average rank of component R. *C* – the frequency of cluster components to be optimal for BCI control in percent of the total number of cluster components. *D* – average ε_{dip} of component.

ing the same component; and 4) the ability to produce components most suitable for BCI control.

The three ICA methods – AMICA, RUNICA and FASTICA – are superior to the other methods in the first three indicators. This result is consistent with the

results obtained in [11], where 22 EEG decomposition methods were compared for MIR and dipolarity for a task related to visual memory. The similarity of the results obtained for completely different mental tasks [11] and the results obtained in this study corroborate the conclusion that the distribution of methods according to these indicators reflects properties of the methods themselves and is independent of the task.

It is worth noting that FASTICA is slightly behind AMICA and RUNICA in terms of the first three indicators, but significantly surpasses them in computational cost. This makes it reasonable to use FASTICA even in real time experiments. CSP demonstrates greatest ability to create components most suitable for BCI control (Fig. 3). This is not surprising since namely this method provides the highest accuracy for recognizing mental states by EEG and is most frequently used for BCI control. However, it is originally assigned to recognize only pair of mental states. The CSP modification proposed in the present paper for recognizing three states in off-line analysis may be efficient for multiclass BCI performing in real time.

After removing the components that do not satisfy the criteria of dipolarity and rank, approximately 10,000 components were left for further analysis. The components were combined in clusters according to the similarity of their TMs. We analyzed the properties of components in 12 largest clusters. The components in these clusters had high rank, dipolarity and significance for BCI control. As a result, they may be treated as physiologically significant. The five components (1, 3, 5, 7 and 9) shown in Fig. 4 were described earlier in [19] but they were obtained by only one ICA method AMICA and for the small group of 7 healthy subjects and 4 post-stroke patients. The result obtained here for large group of 27 healthy subjects by 8 methods of EEG decomposition confirms the correct choice of these components in [19] as the most physiologically significant. The result obtained in [19] also confirms the high ability of AMICA to reveal physiologically meaningful components obtained in the present study.

The sources of the five components were localized in [19] by solving the inverse EEG problem taking into account individual geometry of the brain and its covers obtained by magnetic resonance imaging (MRI) [16]. The sources of the components 1 and 3 were localized in primary somatosensory cortex of the left and right hemispheres, 5 in the supplementary motor area, 7 in the precuneus and 9 in the lateral pre-motor-cortex. The detail description of motor functions for these brain areas is given in [27]. The functions of 7 others have to be additionally studied.

Acknowledgement

The work was supported by the Russian Foundation for Basic Research, grant 20-015-00370. D. Húsek was supported by the long-term strategic development financing of the Institute of Computer Science (RVO:67985807).

References

- [1] ANG K.K., CHUA K.S., PHUA K.S., WANG C., CHIN Z. Y., KUAH C.W., LOW W., GUAN C.A. Randomized Controlled Trial of EEG-Based Motor Imagery Brain-Computer

- Interface Robotic Rehabilitation for Stroke. *Clinical EEG and Neuroscience*. 2015, 46(4), pp. 310–320, doi: [10.1177/1550059414522229](https://doi.org/10.1177/1550059414522229).
- [2] BASHASHATI H., WARD R.K., BIRCH G.E., BASHASHATI A. Comparing different classifiers in sensory motor brain computer interfaces. *PLoS ONE*. 2015, doi: [10.1371/journal.pone.0129435](https://doi.org/10.1371/journal.pone.0129435).
- [3] BELL A.J., SEJNOWSKI T.J. An information – maximization approach to blind separation and blind deconvolution. *Neural Computation*. 1995, 7(6), pp. 1129–1159, doi: [10.1162/neco.1995.7.6.1129](https://doi.org/10.1162/neco.1995.7.6.1129).
- [4] BELOUCHRANI A., ABED-MERAIM K., CARDOSO J.F., MOULINES E. A blind source separation technique using second-order statistics. *IEEE Transactions on signal processing*. 1997, 45(2), pp. 434–444, doi: [10.1109/78.554307](https://doi.org/10.1109/78.554307).
- [5] BIRYUKOVA E.V., PAVLOVA O.G., KURGANSKAYA M.E., BOBROV P.D., TURBINA L.G., FROLOV A.A., DAVYDOV V.I., SILCHENKO A.V., MOKIENKO O.A. Recovery of the motor function of the arm with the aid of a hand exoskeleton controlled by a brain–computer interface in a patient with an extensive brain lesion. *Human Physiology* 2016, 42(1), pp. 13–23, doi: [10.1134/s0362119716010035](https://doi.org/10.1134/s0362119716010035).
- [6] BLANKERTZ B., DORNHEGE G., KRAULEDAT M., MÜLLER K.R., CURIO G. The non-invasive Berlin brain–computer interface: fast acquisition of effective performance in untrained subjects. *NeuroImage*. 2007, 37(2), pp. 539–550, doi: [10.1016/j.neuroimage.2007.01.051](https://doi.org/10.1016/j.neuroimage.2007.01.051).
- [7] BOBROV P.D., KORSHAKOV A.V., ROSHCIN V.I., FROLOV A.A. Bayesian classifier for brain-computer interface based on mental representation of movements. *Zhurnal vysshei nervnoi deiatelnosti imeni I P Pavlova*. 2012, 62(1), pp. 89–99, PMID: 22567990. (in Russian)
- [8] BOBROV P., FROLOV A., HÚSEK D., SNÁŠEL V. Clustering the Sources of EEG Activity during Motor Imagery by Attractor Neural Network with Increasing Activity (ANNIA). In: *Proceedings of the Fifth International Conference on Innovations. Bio-Inspired Computing and Applications IBICA*. Cham: Springer, 2014, pp. 183–191, doi: [10.1007/978-3-319-08156-4_19](https://doi.org/10.1007/978-3-319-08156-4_19).
- [9] BOBROV P., FROLOV A., CANTOR C., BAKHNYAN M., ZHAVORONKOV A. Brain-computer interface based on generation of visual images. *PloS one*. 2011, 6(6), e20674, doi: [10.1371/journal.pone.0020674](https://doi.org/10.1371/journal.pone.0020674).
- [10] CERVERA M.A., SOEKADAR S.R., USHIBA J., MILLÁN J.D.R., LIU M., BIRBAUMER N., GARIPPELLI G. Brain-computer interfaces for post-stroke motor rehabilitation: a meta-analysis. *Annals of clinical and translational neurology*. 2018, 5(5), pp. 651–663, doi: [10.1002/acn3.544](https://doi.org/10.1002/acn3.544).
- [11] DELORME A., PALMER J., ONTON J., OOSTENVELD R., MAKEIG S. Independent EEG sources are dipolar. *PloS one*. 2012, 7(2), p. e30135, doi: [10.1371/journal.pone.0030135](https://doi.org/10.1371/journal.pone.0030135).
- [12] FROLOV A.A., SIROTA A.M., HÚSEK D., MURAVIEV I.P., POLYAKOV P.Y. Binary factorization in Hopfield-like neural networks: single-step approximation and computer simulations. *Neural Network World*. 2004, 14(2), pp. 139–152.
- [13] FROLOV A.A., HÚSEK D., MURAVIEV I.P., POLYAKOV P.Y. Boolean factor analysis by attractor neural network. *IEEE Transactions on Neural Networks*. 2007, 18(3), pp. 698–707, doi: [10.1109/TNN.2007.891664](https://doi.org/10.1109/TNN.2007.891664).
- [14] FROLOV A.A., HÚSEK D., POLYAKOV P.Y. Recurrent-neural-network-based Boolean factor analysis and its application to word clustering. *IEEE transactions on neural networks*. 2009, 20(7), pp. 1073–1086, doi: [10.1109/TNN.2009.2016090](https://doi.org/10.1109/TNN.2009.2016090).
- [15] FROLOV A., HÚSEK D., BOBROV P. Comparison of four classification methods for brain–computer interface. *Neural Network World*. 2011, 21(2), pp. 101–115.
- [16] FROLOV A., HÚSEK D., BOBROV P., KORSHAKOV A., CHERNIKOVA L., KONVALOV R., MOKIENKO O. Sources of EEG activity most relevant to performance of brain-computer interface based on motor imagery. *Neural Network World*. 2012, 22(1), pp. 21–37.

- [17] FROLOV A.A., HÚSEK D., POLYAKOV P.Y. Comparison of seven methods for Boolean factor analysis and their evaluation by information gain. *IEEE transactions on neural networks and learning systems*, 2015, 27(3), pp. 538–550, doi: [10.1109/TNNLS.2015.2412686](https://doi.org/10.1109/TNNLS.2015.2412686).
- [18] FROLOV A., HÚSEK D., BOBROV P., MOKIENKO O., TINTERA J. Sources of electrical brain activity most relevant to performance of brain-computer interface based on motor imagery. In: *Brain-Computer Interface Systems-Recent Progress and Future Prospects*, 2012, pp. 175–193, doi: [10.5772/55166](https://doi.org/10.5772/55166).
- [19] FROLOV A.A., AZIATSKAYA G.A., BOBROV P.D., LUYKMANOV R.K., FEDOTOVA I.R., HÚSEK D., SNAŠEL V. Electrophysiological brain activity during the control of a motor imagery-based brain-computer interface. *Human Physiology*. 2017, 43, pp. 501–511, doi: [10.1134/S036211971705005X](https://doi.org/10.1134/S036211971705005X).
- [20] FROLOV A.A., FEDOTOVA I.R., HÚSEK D., BOBROV P.D. Rhythmic brain activity and brain-computer interface based on motor imagination. *Advances in physiological sciences (in Russian)*. 2017a, 48(3), pp. 72–91.
- [21] FROLOV A.A., MOKIENKO O., LYUKMANOV R., BIRYUKOVA E., KOTOV S., TURBINA L., NADAREYSHVILY G., BUSHKOVA Y. Post-stroke rehabilitation training with a motor-imagery-based brain-computer interface (BCI)-controlled hand exoskeleton: a randomized controlled multicenter trial. *Frontiers in neuroscience*. 2017, 11, p. 400, doi: [10.3389/fnins.2017.00400](https://doi.org/10.3389/fnins.2017.00400).
- [22] FROLOV A.A., BOBROV P.D., BIRYUKOVA E.V., SILCHENKO A.V., KONDUR A.A., DZHALAGONIYA I.Z., MASSION J. Electrical, hemodynamic, and motor activity in BCI post-stroke rehabilitation: clinical case study. *Frontiers in neurology*. 2018, 9, p. 1135, doi: [10.3389/fneur.2018.01135](https://doi.org/10.3389/fneur.2018.01135).
- [23] GRECH R., CASSAR T., MUSCAT J., CAMILLERI K.P., FABRI S.G., ZERVAKIS M., XANTHOPOULOS P., SAKKALIS V., VANRUMSTE B. Review on solving the inverse problem in EEG source analysis. *Journal of Neuroengineering and Rehabilitation*. 2008, 5(25), doi: [10.1186/1743-0003-5-25](https://doi.org/10.1186/1743-0003-5-25).
- [24] HYVARINEN A., KARHUNEN J., OJA E. *Independent component analysis*. New-York: Wiley, 2001. p. 480.
- [25] JONES S.R., PRITCHETT D.L., SIKORA M.A., STUFFLEBEAM S.M., HÄMÄLÄINEN M., MOORE C.I. Quantitative analysis and biophysically realistic neural modeling of the MEG mu rhythm: rhythmogenesis and modulation of sensory-evoked responses. *Journal of neurophysiology*. 2009, 102(6), pp. 3554–3572, doi: [10.1152/jn.00535.2009](https://doi.org/10.1152/jn.00535.2009).
- [26] KACHENOURA A., ALBERA L., SENHADJI L., COMON P. ICA: a potential tool for BCI systems. *IEEE Signal Processing Magazine*. 2008, 25(1), pp. 57–68, doi: [10.1109/MSP.2008.4408442](https://doi.org/10.1109/MSP.2008.4408442).
- [27] KERECHANIN Y.V., HÚSEK D., BOBROV P.D., FEDOTOVA I.R., FROLOV A.A. Sources of electrical brain activity of brain areas involved in motor imagery. *J. Zhurnal Vysshei Nervnoi Deyatelnosti Imeni I.P. Pavlova*. 2019, 69(6) pp. 711–725, (in Russian) doi: [10.1134/S0044467719060066](https://doi.org/10.1134/S0044467719060066).
- [28] KOHAVI R., PROVOST F. Glossary of terms. *Machine Learning (Special Issue on Applications of Machine Learning and the Knowledge Discovery Process)*. 1998, 30, pp. 271–274, doi: [10.1023/A:1017181826899](https://doi.org/10.1023/A:1017181826899).
- [29] MILOŠEVIĆ N., RACKOVIĆ M. Classification based on missing features in deep Convolutional Neural Networks, *Neural Network World*. 2019, 29(6), pp. 221–234, doi: [10.14311/NNW.2019.29.015](https://doi.org/10.14311/NNW.2019.29.015).
- [30] ONTON J., DELORME A., MAKEIG S. Frontal midline EEG dynamics during working memory. *Neuroimage* 2005, 27(2), pp. 341–356, doi: [10.1016/j.neuroimage.2005.04.014](https://doi.org/10.1016/j.neuroimage.2005.04.014).
- [31] ONTON J., WESTERFIELD M., TOWNSEND J., MAKEIG S. Imaging human EEG dynamics using independent component analysis. *Neuroscience and biobehavioral reviews*. 2006, 30(6), pp. 808–822, doi: [10.1016/j.neubiorev.2006.06.007](https://doi.org/10.1016/j.neubiorev.2006.06.007).

- [32] PALMER J.A., KREUTZ-DELGADO K., MAKEIG S. *AMICA: An Adaptive Mixture of Independent Component Analyzers with Shared Components*. San Diego, CA, Technical report, Swartz Center for Computational Neuroscience. 2011, [Viewed 21.12.2021], https://sccn.ucsd.edu/~jason/amica_a.pdf.
- [33] SITARAM R., ROS T., STOECKEL L., HALLER S., SCHARNOWSKI F., LEWIS-PEACOCK J., WEISKOPF N., BLEFARI M., RANA M., OBLAK E., BIRBAUMER N., SULZER J. Closed-loop brain training: the science of neurofeedback. *Nature Reviews Neuroscience*. 2017, 18(2), pp. 86–100, doi: [10.1038/nrn.2016.164](https://doi.org/10.1038/nrn.2016.164).