Using Multiple Decomposition Methods and Cluster Analysis to Find and Categorize Typical Patterns of EEG Activity in Motor Imagery Brain–Computer Interface Experiments

Alexander Frolov[1,2], Pavel Bobrov[1,2*], Elena Biryukova[1,2], Mikhail Isaev[1,2], Yaroslav Kerechanin[1,2], Dmitry Bobrov[1] and Alexander Lekin[1]

[1] Research Institute of Translational Medicine, Pirogov Russian National Research Medical University, Moscow, Russia,
[2] Institute of Higher Nervous Activity and Neurophysiology, Russian Academy of Science, Moscow, Russia

In this study, the sources of EEG activity in motor imagery brain–computer interface (BCI) control experiments were investigated. Sixteen linear decomposition methods for EEG source separation were compared according to different criteria. The criteria were mutual information reduction between the source activities and physiological plausibility. The latter was tested by estimating the dipolarity of the source topographic maps, i.e., the accuracy of approximating the map by potential distribution from a single current dipole, as well as by the specificity of the source activity for different motor imagery tasks. The decomposition methods were also compared according to the number of shared components found. The results indicate that most of the dipolar components are found by the Independent Component Analysis Methods AMICA and PWCICA, which also provided the highest information reduction. These two methods also found the most task-specific EEG patterns of the blind source separation algorithms used. They are outperformed only by non-blind Common Spatial Pattern methods in terms of pattern specificity. The components found by all of the methods were clustered using the Attractor Neural Network with Increasing Activity. The results of the cluster analysis revealed the most frequent patterns of electrical activity occurring in the experiments. The patterns reflect blinking, eye movements, sensorimotor rhythm suppression during the motor imagery, and activations in the precuneus, supplementary motor area, and premotor areas of both hemispheres. Overall, multi-method decomposition with subsequent clustering and task-specificity estimation is a viable and informative procedure for processing the recordings of electrophysiological experiments.

Keywords: brain–computer interface, motor imagery, blind source separation, independent component analysis, common spatial patterns, cluster analysis, EEG pattern extraction

INTRODUCTION

A brain–computer interface (BCI) is a system for controlling a device using registered brain activity. To control the device, the BCI operator is usually required to perform different mental tasks and does not need to use any muscles. In practice, BCI systems are intended to create new means of communication between the operator and his/her environment, which is essential for the paralyzed. Recently, BCI technology has been used not to create a new communication channel but rather to enhance the resting one lost due to stroke or trauma. BCI rehabilitation applications are a growing field, with more clinical data reported every year (Ang et al., 2011, 2015; Ramos-Murguialday et al., 2013; Ono et al., 2014; Frolov et al., 2017b; Mane et al., 2019; Bai et al., 2020). Aside from in the clinic, these systems can be used in fundamental research. The feedback provided by BCI system operation makes its user highly concentrated on performing the tasks required to control it. The user has to produce stable and repeatable patterns of the brain activity recorded, greatly facilitating the investigation of the task-specific patterns of brain activity (Frolov et al., 2012).

Although there are numerous BCI classifiers, i.e., algorithms for brain activity pattern classification, these algorithms typically do not provide pattern features that have clear physiological interpretation. In this work, we have decided to investigate the EEG recordings of our BCI experiments with different source separation methods. Methods such as Principal and Independent Component Analysis (ICA, Hyvärinen et al., 2004), as well as other Blind Source Separation (BSS, Cichocki and Amari, 2002) techniques and non-blind Common Spatial Patterns (CSP, Bashashati et al., 2007), are widely used in EEG research both in the BCI and non-BCI fields. As will be described later, all of the methods considered decompose the signal into the sum of sources, or components, each characterized by its topographic map of weights and temporal activity.

Our experiments are mainly focused on the investigation of the performance of different motor imagery tasks while controlling the BCI (Frolov et al., 2012, 2017a,b). Earlier, we showed that ICA can be effectively used to find the sources of task-related sensorimotor mu-rhythm desynchronization (Frolov et al., 2012) as well as to extract the patterns of activity of three other brain areas involved in motor imagery (Frolov et al., 2017a). However, these results were obtained using a single ICA method. In this paper, we investigate the results of the application of several source separation methods for processing our records. The methods are compared in terms of how many shared components they provide as well as the task specificity of EEG patterns they find. We are mainly focused on studying sources that have clear physiological interpretations.

In Delorme et al. (2012), it was shown that physiologically meaningful components are dipolar. Results were obtained by using several ICA methods to decompose EEG recordings corresponding to visual memory tasks performed. A component being dipolar means that its topographic weight map, which can be looked upon as a scalp potential distribution, is adequately approximated by the distribution resulting from a single current dipole. Note that EEG being a result of the superposition of multiple dipolar source potentials at a time is an accepted model of the EEG origin (Niedermeyer and Da Silva, 2005). The fact that a component is dipolar implies that it results from well-localized changes of electric activity in the brain, allowing it to be attributed to a certain brain area activation or deactivation.

Although the areas the activity of which results in the appearance of dipolar components are typically rather small (about several millimeters in diameter), they significantly impact the EEG signal recorded by all of the electrodes. That is why the main task of the decomposition method is to unmix the signal components. The unmixing done by the ICA and other BSS methods is often based on reducing the mutual information between the components (Hyvärinen et al., 2004). That is why mutual information reduction (MIR) is another criterion of method comparison.

Another indication of a component's relevance is the repeatability of its occurrence among all the subjects and records. The components reflecting common or task-specific EEG phenomena are expected to be found in most of the recordings. That is why a cluster analysis was applied to the topographic maps and power spectral densities of the components found. The results allow the evaluation of the rate at which a certain phenomenon can be discovered by the methods tested.

The paper is organized as follows. The next section provides information on the dataset and participants, as well as details of the 16 blind and non-blind signal decomposition methods used in this work. The measures of both similarity between the methods and similarity of the extracted components are described. The algorithms for estimating the task-specificity of the components and for cluster analysis are given. The results of the method comparison are then represented, and clusters of the most frequently encountered components are shown together with the cluster statistics. The results obtained are discussed, and possible physiological interpretation of the extracted components is given.

METHODS

We used the data from the BCI experiments carried out at the Institute for the Higher Nervous Activity and Neurophysiology of the Russian Academy of Science. The study was approved by the Institute's ethical committee.

Participants

Twenty-three volunteers (7 females and 16 males) aged from 21 to 36 (26 $\pm$ 4) participated in the experiments. The volunteers were right-handed according to the Edinburgh Handedness Inventory and had no reported neurological or other disorders. All of them gave written informed consent.

None of the participants had an experience of controlling a BCI. The ability of the participants to perform motor imagery was not assessed prior to the experiments.

Experimental Procedure

The original EEG dataset for the work was obtained using the same procedure as described in earlier works (Frolov et al., 2012, 2017a). Each subject had to sit relaxed, looking at the center of

the screen or had to imagine flexion of either his left or right hand according to visual cue presented on the screen. Each task had to be performed for 10 s, with a 2 s preparation period. There were 20 cue presentations for each hand motor image, which were split into blocks containing 2 cues for each hand. In each block, the cues were given in a random order. Each motor imagery task was preceded by the relaxation period.

There were from 10 to 20 experimental days (14 ± 3) for each subject, with 1–3 day gaps possible between the sessions.

The EEG data were recorded using the NVX52 acquisition device (Medical Computer Systems, Russia) with 32 electrodes placed according to the 10–10 system. The data were band-pass filtered with a 5–30 Hz Butterworth filter as well as a 50 Hz notch filter to suppress power line interference.

EEG Classification

The EEG data were classified with a Bayesian classifier under the assumption that the signal has a multivariate Gaussian distribution with a zero mean and a different covariance matrix for each task classified:

$$P(\mathbf{x}|i) = \frac{1}{\sqrt{2\pi \det \mathbf{C}_i}} \exp\left(-\frac{\mathbf{x}^T \mathbf{C}_i^{-1} \mathbf{x}}{2}\right), \qquad (1)$$

where $\mathbf{C}_i$ is a covariance matrix of the EEG corresponding to the ith task, and $\mathbf{x}$ is an EEG sample. According to the Bayes rule, the probability that a given sample will correspond to the ith class is proportional to $P(\mathbf{x}|i) P(i)$, where $P(i)$ is the probability of the ith task being cued. Thus, for a signal epoch logarithm of the probability of all its samples to correspond to the ith class is proportional to $\frac{1}{n_t}\left(-\Sigma_t \mathbf{x}_t^T \mathbf{C}_i^{-1} \mathbf{x}_t\right) - \ln \det \mathbf{C}_i + \ln P(i) + const$, where summation is performed over all the epoch samples, n_t is the epoch length in samples, and the constant term is independent of both the signal and the class number. Since $\frac{1}{n_t}\left(-\Sigma_t \mathbf{x}_t^T \mathbf{C}_i^{-1} \mathbf{x}_t\right)$ can be substituted with $\text{trace}\left(\text{cov}(\mathbf{X})\mathbf{C}_i^{-1}\right)$, where $\text{cov}(\mathbf{X})$ is the epoch covariance estimate under the assumption of a zero mean, the class number is

$$i_{class}(\mathbf{X}) = \underset{i}{\text{argmax}}\left(\ln P(i) - \text{trace}\left(\text{cov}(\mathbf{X})\,\mathbf{C}_i^{-1}\right) - \ln \det \mathbf{C}_i\right) \qquad (2)$$

We used a window of 1 s length sliding with a 100 ms shift to present the feedback during the sessions. The probabilities $P(i)$ were estimated based on the cue durations. The classifier was updated after each block of cues, with no feedback during the first block.

Signal Decomposition

All of the blind and non-blind source separation methods used can be, in general, represented in the form of multivariate linear decomposition with an optional noise term.

$$\mathbf{X} = \mathbf{a}_1 \boldsymbol{\xi}_1 + \ldots + \mathbf{a}_n \boldsymbol{\xi}_n + \mathbf{N} = \mathbf{A}\boldsymbol{\Xi} + \mathbf{N}, \qquad (3)$$

where $\mathbf{X}$ is the decomposed signal in matrix form, $\mathbf{a}_i$, $i = 1, \ldots, n$, are columns of $\mathbf{A}$ that define the source (component) weights representing the source topographic maps, $\boldsymbol{\xi}_i$, $i = 1, \ldots, n$, are

row vectors of the source activities, n is the number of EEG channels, and is the optional noise term. Matrix $\mathbf{A}$ is called the mixing matrix, and its inverse, $\mathbf{W}$, is called the unmixing matrix. The invertibility of $\mathbf{A}$ is based on the assumptions that there are as many components as there are signal channels and that the signal correlation matrix has full rank. These assumptions are in general not required but are common for linear source separation methods (Cichocki and Amari, 2002; Hyvärinen et al., 2004), and all of the decomposition methods used in the paper were derived under these assumptions.

Usually, the signal is whitened before searching for decomposition (3) so that the covariance matrix of $\mathbf{X}$ becomes an identity matrix. Also, it is well-known that the component activity variances are not defined, thus requiring the application of some constraints when the decomposition is computed. That is why we shall further consider the decomposed signal to be whitened, i.e.

$$\mathbf{X} = \mathbf{V}\mathbf{Z}, \qquad (4)$$

where $\mathbf{V}$ is a matrix such that $\mathbf{V}\text{cov}(\mathbf{X})\mathbf{V}^T = \mathbf{I}$. If we also suppose that all $\boldsymbol{\xi}_i$, $i = 1, \ldots, n$, have unit variance, then the mixing matrix for $\mathbf{Z}$ is orthogonal, $\mathbf{Z} = \mathbf{U}\boldsymbol{\Xi}$, $\mathbf{U}\mathbf{U}^T = \mathbf{I}$, and the EEG mixing matrix is $\mathbf{A}$ equals $\mathbf{V}\mathbf{U}$.

We used 16 source separation methods to obtain the decomposition (3).

PCA, or Principle Component Analysis, is a classic example of a technique for obtaining the decomposition (3). It uses Singular Value Decomposition to diagonalize the signal covariance matrix. Usually, the components with too much or too little variance are discarded. If one rather chooses to keep all of the components under the constraints that their variances are equal to 1, this will become signal whitening. In this case, $\mathbf{U} = \mathbf{I}$, and the unmixing matrix is not orthonormal.

KURT is an ICA method based on maximizing the difference of the component distributions from normal. The difference is measured by the absolute value of the component excess kurtosis, which serves as a cost function for iterative component search (Delfosse and Loubaton, 1995; Hyvärinen et al., 2004), i.e., the columns of unmixing matrix $\mathbf{W}$ are estimated sequentially by maximizing the excess kurtosis for $\mathbf{w}_i^T \mathbf{Z}$ under the assumption that $\mathbf{W}$ is orthogonal.

CUMUL is an ICA method using the signal non-stationarity as the criterion of statistical independence (Hyvarinen, 2001). The non-stationarity is measured by the fourth-order cumulant:

$$\begin{aligned}
\text{cum}_4(x) = {} & E\left\{x^2(t)\,x^2(t-\tau)\right\} - E\left\{x^2(t)\right\}E\left\{x^2(t-\tau)\right\} \\
& - 2E\left\{x(t)x(t-\tau)\right\}^2.
\end{aligned} \qquad (5)$$

The decomposition seeks to maximize the sum of the source cumulants. The proof of non-stationarity maximization being a criterion of component independence can be found in Hyvärinen et al. (2004), chapter 18. The time lag τ serves as the method parameter. In our experiment, it was set to 100 ms to extract the components with dominant alpha-band frequencies.

FastICA is an ICA method in which mutual information reduction between the components serves as the statistical

independence criterion. The mutual information between the components $\boldsymbol{\xi}_i$, $i = 1, \ldots, n$, is defined as

$$I\left(\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_n\right) = \sum_i \mathrm{H}(\boldsymbol{\xi}_i) - H(\boldsymbol{\Xi}), \tag{6}$$

where $H(\cdot)$ denotes entropy. The entropy of linear transformation is $H(\mathbf{WZ}) = H(\mathbf{Z}) + |\det(\mathbf{W})|$. Under the assumption of $\mathbf{Z}$ being whitened and the components having unit variance, $\mathbf{W}$ is orthogonal, which implies that $\det(\mathbf{W}) = 1$ and $H(\mathbf{WZ})$ does not depend on $\det(\mathbf{W})$. As a result, $H\left(\mathbf{w}_i^{\mathrm{T}}\mathbf{Z}\right)$ can be used as a cost function for sequential search of the unmixing matrix columns $\mathbf{w}_i$.

Since computing the entropy requires knowledge on the signal distribution, different parametric approximations are used. FastICA uses approximation in the form

$$H(x) \propto const - \left[E\left\{g(x)\right\} - E\left\{g(\nu)\right\}\right]^2, \tag{7}$$

where $g(x)$ is a non-linear function and ν is a random normally distributed variable with unit variance. FastICA can be computed using different non-linearities (Bingham and Hyvärinen, 2000; Hyvärinen et al., 2004). In this work, we used two functions, namely, $\tanh(x)$ and $x\exp\left(-x^2/2\right)$. The corresponding methods are denoted as FastICAT and FastICAG.

RunICA, or extended infomax, is an ICA method that is similar to FastICA. The method takes into account both super- and sub-Gaussian distributions by applying two different non-linearities, $g_+(x) = -2\log\cosh(x)$ and $g_-(x) = \log\cosh(x) - x^2/2$, and switching between them. Originally, the infomax algorithm was introduced by Bell and Sejnowski (1995) based on a neural network approach to maximize the entropy of the neural network output. It was later reformulated in terms of parametric non-linearities for source entropy estimation (Lee et al., 1999).

AMICA, which stands for Adaptive Mixture Independent Component Analysis, is an ICA method based on the assumption that all EEG components have generalized Gaussian mixture distributions

$$p(\xi_i) = \sum_{j=1, \ldots, k} \alpha_{ij} q(\xi_i, \rho_{ij}, \mu_{ij}, \beta_{ij}), \tag{8}$$

$$q(\xi, \rho, \mu, \beta) = \frac{\rho}{2\beta\Gamma(1/\rho)}\exp\left(-\left|\frac{\xi - \mu}{\beta}\right|^\rho\right). \tag{9}$$

AMICA can also segment the signal into several parts corresponding to different models (8) when the number of the models is set to higher than 1. All distribution parameters, unmixing matrix coefficients, and model scores (in the case of several models) are estimated using the expectation-maximization algorithm (Palmer et al., 2006, 2008, 2012). In our work, we compared AMICA with the default settings (2 models, 2 distributions in mixture) to AMICA with two models and Gaussian distributions with means fixed to zero (2 models, 2 distributions in mixture, $\rho_{i,j}^l = 2$, $\mu_{i,j}^l = 0$, where l indicates the model, $l = 1, 2$, and $i = 1, \ldots, n$, $j = 1, 2$), and to AMICA with a single model and Gaussian distributions with means fixed

to zero (1 model, 2 distributions in mixture, $\rho_{ij} = 2$, $\mu_{ij} = 0$, $i = 1, \ldots, n$, $j = 1, 2$). The two latter cases will be denoted as AMICA1 and AMICA2, respectively.

PWCICA is an ICA method mapping the signal into a complex domain using the first derivative (rate) of the signal as its imaginary part. The algorithm is described in Ball et al. (2016). It uses the complex variant of FastICA in order to find decomposition (3) in the complex domain (Bingham and Hyvärinen, 2000). After that, the real-valued unmixing matrix is computed using the algorithm provided in Ball et al. (2016).

SOBI, or Second Order Blind Identification, is a blind source separation method that uses joint diagonalization to remove signal cross-covariance for a set of defined time lags (Belouchrani et al., 1997). Particularly, a set of time lags δ_k is specified, and corresponding cross-covariance matrices are estimated:

$$\mathbf{C}_k = \mathrm{E}\left(\mathbf{X}(t), \mathbf{X}(t - \delta_k)\right), \tag{10}$$

Next, the unmixing matrix $\mathbf{W}$ is found by joint diagonalization of the estimated cross-covariance matrices:

$$\mathbf{W} = \underset{\mathbf{M}}{\arg\min}\left(\Sigma_k \left\|\mathbf{M}\mathbf{C}_k\mathbf{M}^{\mathrm{T}}\right\|_{\mathbf{off}}\right) \tag{11}$$

where $\|\cdot\|_{off}$ denotes the Euclidean norm of the off-diagonal part of a matrix. The norm of the unmixing matrix column is constrained to be equal to 1. There exist several algorithms for solving the problem (11). We used one proposed in Ziehe et al. (2004) to obtain the solution.

CSP, or Common Spatial Patterns (Ramoser et al., 2000; Bashashati et al., 2007, 2015), is a non-blind decomposition method, in contrast to the source separation techniques described above. This method utilizes the information on signal segmentation with respect to the experimental tasks in order to find the decomposition (3). The method is used for feature extraction in two-class separation problems. The decomposition is sought that maximizes the component variance ratio for the states classified under the assumption that total component variance is fixed. This is equivalent to finding matrix $\mathbf{W}$ such that

$$\left\{\begin{array}{l}\mathbf{WC}_1\mathbf{W}^T = \mathbf{D}_1 \\ \mathbf{W}(\mathbf{C}_1 + \mathbf{C}_2)\mathbf{W}^T = \mathbf{I}\end{array}\right\}, \tag{12}$$

where $\mathbf{C}_i$ is the i-th class covariance matrix and $\mathbf{D}_1$ is a diagonal matrix. The equations (12) have an explicit solution where $\mathbf{W}$ and $\mathbf{D}_1$ result from singular-value decomposition (SVD) of the matrix $\mathbf{D}^{-1/2}\mathbf{U}^{\mathrm{T}}\mathbf{C}_1\mathbf{U}\mathbf{D}^{-1/2}$ with orthogonal matrix $\mathbf{U}$ and diagonal matrix $\mathbf{D}$ given by SVD of $\mathbf{C}_1 + \mathbf{C}_2$. We used several CSP decompositions in this work, namely, CSP12 comparing relaxation and left hand motor imagery, CSP13 comparing relaxation and right hand motor imagery, CSP23 comparing left with right hand motor imagery, and CSP1X comparing relaxation and both left and right hand motor imagery. Although equations (12) are solved using signal epochs corresponding to the selected tasks only, the unmixing matrix was used to estimate the activity of the components at all other time moments.

MCSP is a generalization of the CSP method for multiple-class problems. There are several methods of generalization

(Lee et al., 2005; Bashashati et al., 2015). The most frequent approaches are to either compute all pairwise CSP and combine the corresponding components into a new signal to be classified or to use classifier voting or a tree of some sort when only two classes are compared at each step (e.g., relaxation vs. motor imagery with subsequent discrimination between the hands if the motor imagery was recognized). However, these techniques would not give us any additional information in terms of sources, as we have already considered all pairwise CSP decompositions for our experimental tasks, as well as the CSP1X comparing relaxation and motor imagery. That is why we have decided to use a different approach. Problem (12) was generalized so as to find $\mathbf{W}$ such that

$$
\begin{cases}
\mathbf{W}\mathbf{C_1}\mathbf{W}^{\mathrm{T}} = \mathbf{D_1} \\
\mathbf{W}\mathbf{C_2}\mathbf{W}^{\mathrm{T}} = \mathbf{D_2} \\
\mathbf{W}(\mathbf{C_1} + \mathbf{C_2} + \mathbf{C_3})\mathbf{W}^{\mathrm{T}} = \mathbf{I}
\end{cases}
$$

The equations are written for the case of three tasks, but the generalization for the case of an arbitrary number of classes is straightforward. The problem in general has no explicit solution and can be solved numerically using any joint diagonalization technique. The joint diagonalization algorithm was the same as was used to perform the SOBI decomposition (Ziehe et al., 2004). It should be noted that unlike the two-class CSP, component variance ratio maximization (or minimization) for different classes is not guaranteed, but if any component has high (small) variance for a certain class, it will have small (high) variance for other classes.

Mutual Information Reduction

The mutual information reduction (MIR) provided by the decomposition methods is defined as

$$
\mathrm{MIR} = \mathrm{I}\,(x_1, \ldots x_n) - \mathrm{I}(\xi_1, \ldots, \xi_n), \tag{13}
$$

where the mutual signal information I is given by equation (6). Using (6), we get

$$
\mathrm{MIR} = \Sigma_i \mathrm{H}\,(x_i) - \mathrm{H}\,(\mathbf{X}) - \Sigma_i \mathrm{H}\,(\xi_i) + H(\Xi) \tag{14}
$$

Since $\mathbf{X} = \mathbf{A}\Xi$,

$$
\mathrm{H}\,(\mathbf{X}) = \mathrm{H}\,(\Xi) + \log\left|\det \mathbf{A}\right| \tag{15}
$$

and thus

$$
\mathrm{MIR} = \Sigma_i \mathrm{H}\,(x_i) - \Sigma_i \mathrm{H}\,(\xi_i) + \log\left|\det \mathbf{A}\right| \tag{16}
$$

The entropy of each individual channel and component was estimated and bias-corrected as in Delorme et al. (2012).

Dipolar Component Selection

The decomposition (3) shows that the contribution of each component to the signal registered by the EEG electrodes is given by the weight vector, which can be treated as potential distribution over the scalp surface. This allows the source of electrical activity represented by the component to be localized.

As has been shown for the case of visual memory tasks, the physiologically meaningful components tend to have dipolar distributions of their weights, i.e., the distributions that could be adequately approximated by a single current dipole inside a head model (Delorme et al., 2012). We had neither anatomical magnetic resonance imagery (MRI) scans nor digitized electrode positions for most of our subjects. This is why a standardized head model based on the ICBM MNI atlas (Fonov et al., 2009, 2011) was used for estimating the dipolarity of the component weights. The dipolarity was computed as a residual variance of the best single dipole fit for the corresponding distribution. The fit was considered acceptable if the residual variance did not exceed 10%.

Shared Components

The number of the same components found by different decomposition methods can be used as a measure of similarity between the methods. On the other hand, the more methods reveal the component, the less likely it is that the component is an artifact of a certain mathematical procedure underlying the decomposition algorithm. That is why we focused our attention on the components found by different methods.

Two component similarity measures were used to determine which components were shared between the methods. The first measure was the absolute value of the cosine between normalized component weight vectors given by the corresponding mixing matrix columns. The second one was the absolute value of the Pearson correlation between the component activities. Two thresholds were chosen for the measures. The components were considered the same when both measures exceeded the corresponding thresholds. We have chosen 0.9 as the threshold for the topographic map similarity and 0.8 as the threshold for activity correlation.

When the number of shared components is obtained, the similarity between the decomposition methods can be measured as

$$
\mathrm{sim_{ij}} = \frac{n_s}{n_i + n_j - n_s}, \tag{17}
$$

where n_s is the number of shared components, and n_i and n_j are the numbers of components obtained by the i-th and j-th methods, respectively. The similarity measure varies from 0, when there are no shared components, to 1, when the methods have provided identical decomposition. Also, only the dipolar components can be considered when the similarity (17) is calculated, allowing non-dipolar components that are unlikely to have physiological meaning to be discarded. The number of methods that found a component minus one will be called the component rank. The component rank equals 0 when it is found by only one method and 15 when it is found by all of the 16 methods considered.

Classification Accuracy and the Component Specificity

Given a set of the EEG components, their activity can be used for estimating the accuracy of discriminating between the experimental tasks. This estimate can serve as a measure of

components' specificity for the performed tasks, allowing the task-specific EEG patterns to be extracted and both artifacts and irrelevant EEG activity to be removed.

We used a greedy algorithm to find the task-specific components as follows. Only the dipolar components were considered. At the first step, classification accuracy was estimated for each set containing one or two components. The best set was then selected. The remaining components were added to the set one by one so that the new set would provide the highest classification accuracy. The components from the set for which the classification accuracy was maximal (over all of the steps) were considered as the task-specific components. The classification was tested by 120 trials of cross-validation where 7 blocks were chosen for the training set and the 3 remaining blocks were used as the testing set.

The accuracy of task classification was measured by Cohen's kappa index, κ (Vieira et al., 2010), which was calculated from the elements of the confusion matrix (g_{ij}) resulting from the cross-validation. The element g_{ij} counts how many times the i-th task was recognized by the classifier under the condition that the classified EEG epoch corresponded to the j-th cue. Given g_{ij}, the κ index is

$$\kappa = \frac{g_0 \sum_i g_{ii} - \sum_j g_j^2}{g_0^2 - \sum_j g_j^2}, \tag{18}$$

where g_0 is the sum of all of the elements of the confusion matrix, $\sum_i g_{ii}$ is the sum of all diagonal elements of the confusion matrix, and g_j is the sum of all of the elements of the j-th column of the confusion matrix. The index varies from -1 to 1 (perfect classification) and equals 0 in case of random classification.

Component Clustering

The decomposition (3) gives at least as many components as there are electrodes for each of the methods used. Consequently, the number of components computed for all experimental recordings is too high to check the components manually. In order to investigate EEG patterns most typical for our experimental tasks, we used cluster analysis to group the components according to their similarity. The clustering was performed using the Attractor Neural Network with Increasing Activity (ANNIA), which was originally proposed for Boolean factor analysis (Frolov et al., 2007). Adaptation of the technique for cluster analysis is described in Bobrov et al. (2014). When the ANNIA is used for clustering, the strength of synaptic connections between the neurons of the network is determined by the similarity between the corresponding elements rather than by Hebbian learning. The stopping criterion is based on a threshold that specifies the minimal average similarity between the elements of a cluster. In Bobrov et al. (2014), the ANNIA was used to group the components with respect to their topographic map similarity. In this paper, we have also accounted for the component activity similarity using the correlation coefficient between the components' activity power spectral densities estimated for each of the experimental tasks. Spectral analysis was used since the components obtained from different recordings were compared, in contrast with the

TABLE 1 | Percentage of dipolar components found and the method MIR.

Method	% of dipolar components	MIR, bits/(sec·chan)
PCA	4.35 [2.17; 7.50]	63.56 [51.89; 77.34]
KURT	15.39 [6.98; 27.50]	70.15 [58.97; 83.69]
CUMUL	12.77 [6.38; 22.61]	70.96 [59.21; 84.81]
FastICAT	15.63 [8.51; 27.50]	68.24 [57.28; 82.96]
FastICAG	15.22 [8.33; 27.33]	67.58 [56.60; 83.75]
RunICA	10.87 [3.54; 25.00]	74.92 [65.21; 86.47]
AMICA	25.53 [11.11; 38.96]	68.98 [57.47; 83.99]
AMICA1	21.74 [12.50; 32.55]	69.23 [58.11; 83.95]
AMICA2	27.50 [15.38; 40.00]	66.74 [55.88; 80.30]
PWCICA	19.57 [6.38; 31.25]	67.42 [56.00; 81.38]
SOBI	15.22 [6.67; 26.67]	71.52 [58.80; 84.88]
CSP12	4.26 [1.37; 13.68]	62.99 [50.11; 78.13]
CSP13	4.17 [2.08; 12.74]	63.24 [48.94; 77.74]
CSP1X	4.26 [1.91; 14.97]	63.72 [50.34; 77.87]
CSP23	2.22 [1.23; 10.00]	63.18 [48.82; 78.13]
MCSP	6.90 [2.17; 17.95]	63.89 [51.40; 77.78]

shared components search described earlier. Both the component similarity measures considered vary from 0 to 1, and their mean was taken as the similarity measure for ANNIA.

RESULTS

Comparison of the Decomposition Methods

The methods were compared according to the criteria described in the previous section. The percentage of dipolar components found by each method is presented in **Table 1** (median values and quartiles are presented). The table also contains average values of mutual information reduction values for each of the methods. The p-values resulting from pairwise comparison of the percentage of the dipolar components found are presented in the upper diagonal part of **Table 2**. The values were obtained using a Wilcoxon test of the values pooled from all of the participants and sessions with subsequent Benjamini-Hochberg correction (initial significance level: 0.05; resulting critical value: 0.0429). The results indicate that the PWCICA and AMICA methods provide more dipolar components than the other methods. The difference is significant for the AMICA methods and insignificant when PWCICA is compared to other BSS methods. The PCA and CSP methods extracted significantly fewer dipolar components than the other methods.

Pairwise Wilcoxon comparison of the MIR values with Benjamini-Hochberg correction (initial significance level: 0.05; resulting critical value: 0.0275) shows that the ICA methods provide significantly higher mutual information reduction than the PCA and CSP methods. The difference between ICA and other methods in terms of MIR was insignificant, as was the difference between the PCA and CSP methods.

The method similarity was estimated by the fraction of shared components according to (17), providing a method similarity

TABLE 2 | The *p*-values of pairwise Wilcoxon testing of the percentage of the dipolar components found (upper triangle part) and the pattern classification accuracy (lower triangle part).

	MCSP	CSP23	CSP12	CSP1X	CSP13	AMICA1	AMICA	AMICA2	PWCICA	SOBI	FastICAT	KURT	FastICAG	CUMUL	RunICA	PCA
MCSP		<1e-4	<1e-4	<1e-4	<1e-4	<1e-4	<1e-4	<1e-4	<1e-4	<1e-4	<1e-4	<1e-4	<1e-4	<1e-4	0.0053	<1e-4
CSP23	0.0463		0.0312	0.0053	0.0045	<1e-4	<1e-4	<1e-4	<1e-4	<1e-4	<1e-4	<1e-4	<1e-4	<1e-4	<1e-4	0.0037
CSP12	**0.1150**	**0.6045**		**0.5743**	**0.5525**	<1e-4	<1e-4	<1e-4	<1e-4	<1e-4	<1e-4	<1e-4	<1e-4	<1e-4	<1e-4	**0.8583**
CSP1X	0.0352	**0.9696**	**0.5822**		**0.9963**	<1e-4	<1e-4	<1e-4	<1e-4	<1e-4	<1e-4	<1e-4	<1e-4	<1e-4	<1e-4	**0.3943**
CSP13	0.0094	**0.6366**	**0.3041**	**0.6029**		<1e-4	<1e-4	<1e-4	<1e-4	<1e-4	<1e-4	<1e-4	<1e-4	<1e-4	<1e-4	**0.5076**
AMICA1	<1e-4	0.0231	0.0056	0.0175	**0.0628**		0.0323	<1e-4	0.0004	<1e-4	<1e-4	<1e-4	<1e-4	<1e-4	<1e-4	<1e-4
AMICA	<1e-4	0.0056	0.0015	0.0046	0.0181	**0.6061**		0.0673	<1e-4	<1e-4	<1e-4	<1e-4	<1e-4	<1e-4	<1e-4	<1e-4
AMICA2	**0.0682**	**0.7721**	**0.5554**	**0.8452**	**0.8395**	**0.0940**	0.0457		<1e-4	<1e-4	<1e-4	<1e-4	<1e-4	<1e-4	<1e-4	<1e-4
PWCICA	<1e-4	0.0007	<1e-4	0.0005	0.0027	**0.2870**	**0.6421**	0.0076		**0.0937**	**0.5838**	**0.1890**	**0.2345**	0.0008	<1e-4	<1e-4
SOBI	<1e-4	0.0006	<1e-4	0.0003	0.0016	**0.2197**	**0.5266**	0.0055	**0.8586**		**0.2654**	**0.7172**	**0.6611**	0.0350	0.0003	<1e-4
FastICAT	<1e-4	<1e-4	<1e-4	<1e-4	0.0002	**0.0825**	**0.2646**	0.0012	**0.4937**	**0.6218**		**0.4582**	**0.5342**	0.0013	<1e-4	<1e-4
KURT	<1e-4	<1e-4	<1e-4	<1e-4	<1e-4	**0.0758**	**0.2594**	0.0009	**0.4703**	**0.5728**	**0.9670**		**0.9049**	0.0158	<1e-4	<1e-4
FastICAG	<1e-4	<1e-4	<1e-4	<1e-4	<1e-4	**0.0515**	**0.2027**	0.0007	**0.3798**	**0.5108**	**0.8375**	**0.9007**		0.0094	<1e-4	<1e-4
CUMUL	<1e-4	<1e-4	<1e-4	<1e-4	<1e-4	0.0145	**0.0839**	<1e-4	**0.1834**	**0.2308**	**0.5076**	**0.5392**	**0.6238**		**0.0526**	<1e-4
RunICA	<1e-4	<1e-4	<1e-4	<1e-4	<1e-4	0.0024	0.0202	<1e-4	**0.0391**	**0.0659**	**0.1537**	**0.1792**	**0.2098**	**0.4488**		<1e-4
PCA	<1e-4	<1e-4	<1e-4	<1e-4	<1e-4	<1e-4	<1e-4	<1e-4	<1e-4	<1e-4	<1e-4	<1e-4	<1e-4	0.0002	0.0100	
On-line	<1e-4	<1e-4	<1e-4	<1e-4	<1e-4	<1e-4	<1e-4	<1e-4	<1e-4	<1e-4	<1e-4	<1e-4	<1e-4	<1e-4	<1e-4	<1e-4

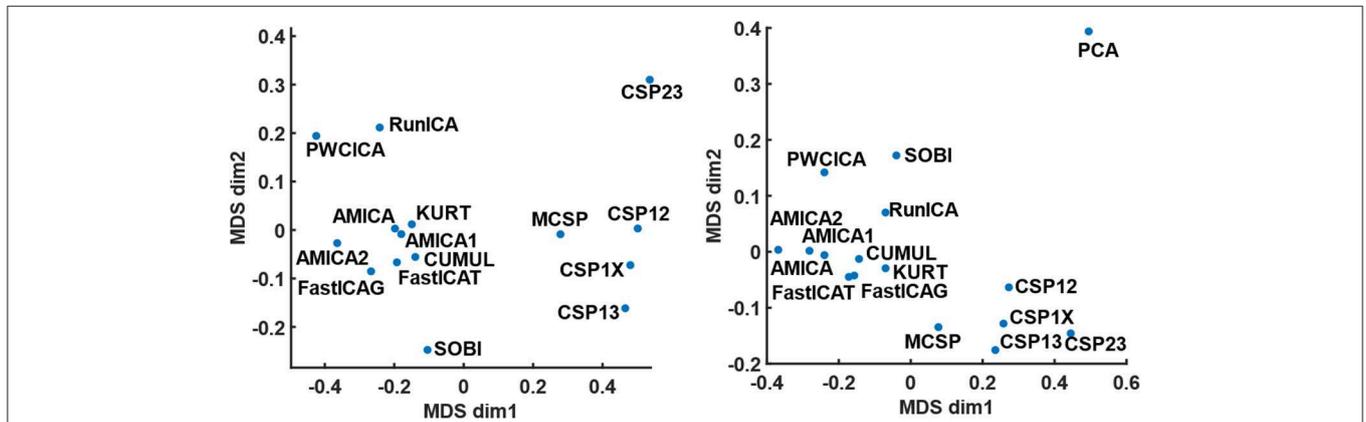*The values indicating insignificant difference are in bold.*

FIGURE 1 | The results of multidimensional scaling according to method similarity. The mapping on the left panel is obtained for the case where all the components were considered. The mapping on the right panel is obtained for the case where only the dipolar components were considered. The PCA point is omitted from the left panel due to its remoteness from all other points.

matrix. The similarity matrices for the cases when either all or only dipolar components were considered were used for mapping the method onto a 2D plane with a multidimensional scaling technique. The results of the mapping are shown in **Figure 1**. Apparently, the methods tend to group according to the decomposition criteria, one group containing the ICA methods except PWCICA, and another group containing the CSP methods. SOBI and PWCICA are distant from both groups, while the PCA is as an outlier.

The results for classification accuracy, obtained after searching for the task-specific components, are shown in **Figure 2**. Mean κ values are shown for each subject and each method. The methods are sorted according to the κ value averaged over

all the subjects. The subjects are sorted according to the κ value averaged over all the methods. The lower diagonal part of **Table 2** contains *p*-values resulting from the pairwise comparison of the κ values for different methods. The values were obtained using Wilcoxon test of the values pooled from all of the participants and sessions with Benjamini-Hochberg correction (initial significance level: 0.05; resulting critical value: 0.320). The CSP methods provide patterns that are significantly better classified than those found by the other methods. The PCA patterns are significantly worse-classified than the patterns found by the other methods. The task-specific component search yielded accuracy values significantly exceeding those obtained during on-line BCI operation,
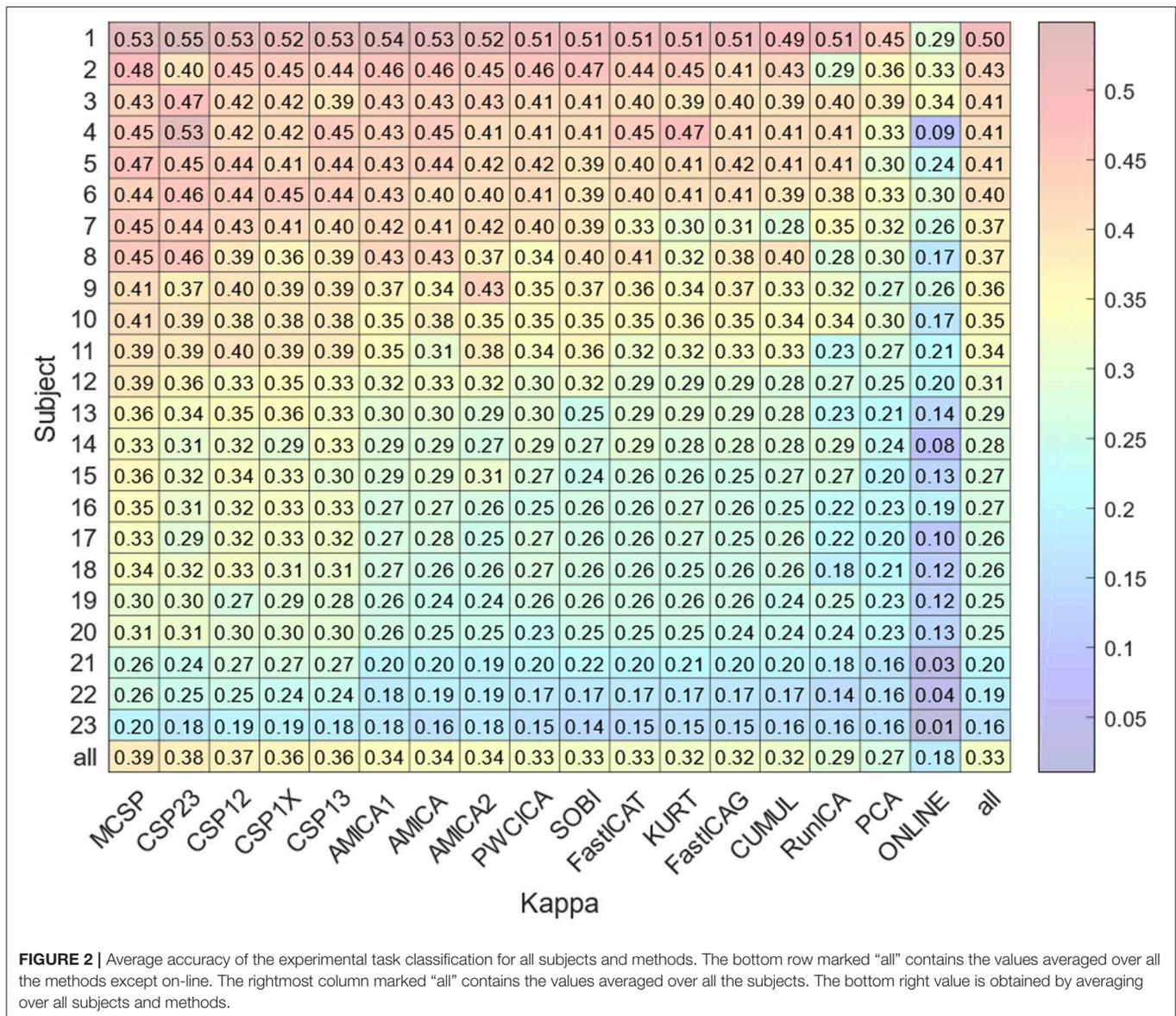
FIGURE 2 | Average accuracy of the experimental task classification for all subjects and methods. The bottom row marked "all" contains the values averaged over all the methods except on-line. The rightmost column marked "all" contains the values averaged over all the subjects. The bottom right value is obtained by averaging over all subjects and methods.

which is likely due to discarding irrelevant, artifact, and noisy components.

Component Clustering

The clusters obtained using the ANNIA were sorted according to their component occurrence, i.e., the percentage of the experimental session in which the components of the cluster were found. **Figure 3** shows topographic maps and activity power spectral densities (PSD) for the components of the first 12 clusters. The maps and PSDs were obtained by averaging over all of the cluster components regardless of the subject or session. **Table 3** presents the occurrence, average dipolarity, average rank, and specificity of the components of each cluster. As stated above, the component dipolarity was measured by the residual variance of the topographic map fit with a potential distribution resulting from a single current dipole, and the component rank is the

number of methods that found the component minus one. The specificity was calculated as the percentage of cases when the component was included in the best set of components (the set search is described in the Methods section) among the cases when the component was found.

It should be mentioned that AMICA restricted to only Gaussian distributions in mixtures performed as well as AMICA with the generalized super-Gaussian mixtures, suggesting that dipolar components obtained from band-pass filtered EEG signal have activity the distribution of which can be adequately approximated with a mixture of two Gaussians.

DISCUSSION

In the EEG\MEG field, and, particularly, in BCI research, ICA methods are mostly used for artifact removal during signal
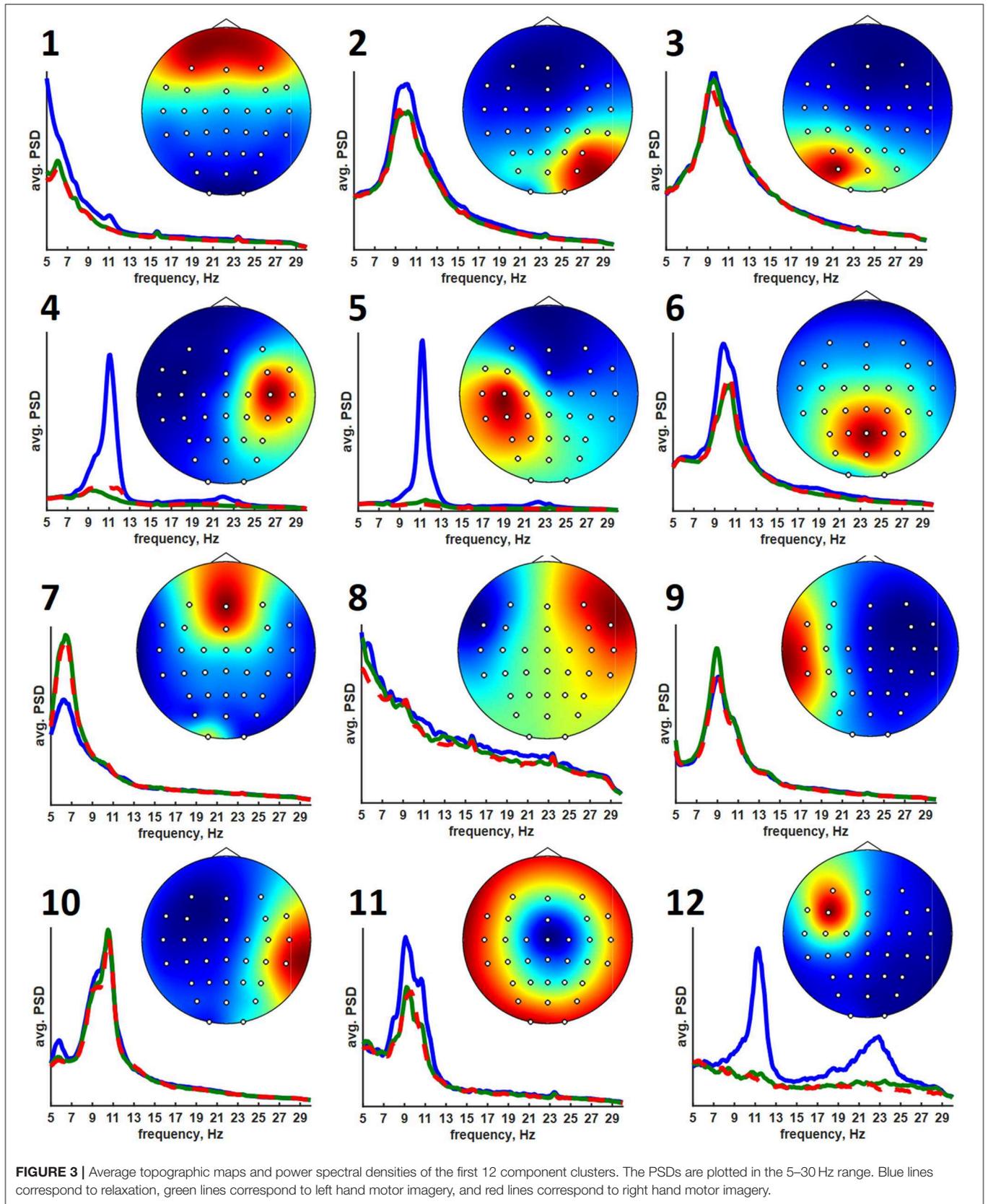
FIGURE 3 | Average topographic maps and power spectral densities of the first 12 component clusters. The PSDs are plotted in the 5–30 Hz range. Blue lines correspond to relaxation, green lines correspond to left hand motor imagery, and red lines correspond to right hand motor imagery.

TABLE 3 | The cluster statistics.

Cluster	1	2	3	4	5	6	7	8	9	10	11	12
Occurrence, %	71	70	61	52	52	48	39	38	35	30	26	22
Dipolarity, %	2	9	8	8	4	3	2	3	9	8	3	4
Rank	3.74	2.24	1.78	1.62	3.53	2.83	2.66	1.48	1.06	1.78	1.92	2.23
Specificity, %	21	10	8	24	59	61	34	10	16	18	26	41

preprocessing or epoch classification. They are applied for eliminating ocular (Höller et al., 2013; Dharmaprani et al., 2016; Sarin et al., 2020), motion (Zhou et al., 2016; Kobler et al., 2019), and muscle (Höller et al., 2013; Dharmaprani et al., 2016) artifacts. The most popular method used for artifact suppression is RunICA (extended infomax), likely due to its incorporation into EEGLAB (Delorme and Makeig, 2004) and BrainStorm (Tadel et al., 2011) software. The ICA methods are also believed to provide effective spatial filters for BCI feature selection (Kachenoura et al., 2007; Lotte et al., 2018; Xiaopei et al., 2019). The main limitation to using the methods is that they require a substantial amount of data to converge to meaningful decomposition, and many of the algorithms have high computational costs to implement on-line. These issues, as well as the nature of the algorithm used to automatically select the most specific components, limit the on-line implementation of the procedure proposed in this paper. The procedure is rather intended for assessing the overall task-specificity of the EEG signal and adequately arranging experimental recordings and subjects with respect to the BCI control efficiency. Performing the cluster analysis allows the most frequent and typical EEG patterns in the processed recordings to be extracted with no need to choose a single decomposition method. The best method may differ when different criteria are applied, such as the number of the dipolar components found, the mutual information reduction, or the accuracy of classification of the extracted patterns.

The method comparison shows that ICA methods provide results similar to each other despite the difference in the independence criteria underlying the algorithms. These methods find significantly more dipolar components and provide significantly higher MIR than CSP or PCA, with the best results being for AMICA and PWCICA. PCA also has the lowest number of components shared with the other methods, as indicated by the results of multidimensional scaling. The maximal similarity between the PCA and other methods, estimated according to (17), was five times less than the minimal similarity between the other methods when all of the components were considered and two times less when only the dipolar components were considered. On the other hand, our results indicate that non-blind CSP methods find more of the separable activity patterns, much like in Xiaopei et al. (2019), where three ICA methods were compared to CSP in terms of providing the best spatial filters for a BCI classifier. However, unlike our case, no significant difference between RunICA and CSP was observed in Xiaopei et al. (2019). CSP outperformed the ICA methods, likely due to accounting for the signal segmentation with respect to the experimental tasks.

Thus, CSP decomposition is likely to find the dipolar components with task-specific activity, while it is poor at extracting other task-irrelevant dipolar sources. However, finding irrelevant but physiologically plausible sources may be useful for proving that BCI works based on the activity of the targeted brain areas. E.g., it may be important to establish that motor imagery BCI used for post-stroke rehabilitation worked because of the activity of the motor areas and not just due to eye movements or differences in concentration level.

The results of cluster analysis agree with those obtained earlier for fewer healthy subjects and patients with subcortical lesions (Frolov et al., 2017a). In previous work, the components relevant to controlling a hand exoskeleton via BCI were found to correspond to the sources of sensorimotor mu-rhythm located at the bottom of the central sulci of the left (SIL) and right (SIR) hemispheres, alpha-rhythmic activity in the precuneus (Prc), and alpha- and beta- activity in both the supplementary motor area (SMA) and in the premotor cortex of the left hemisphere (PrmL). These sources correspond to clusters 5 (SIL), 6 (SIR), 3 (Prc), 11 (SMA), and 8 (PrmL). We have also found clusters corresponding to blinking (1) and eye movements (7), which sometimes exhibited task-relevant activity, e.g., the subjects often blinked more frequently during relaxation. Clusters 2 and 3 correspond to the occipital alpha rhythm, which has low task-specificity. Interestingly, the source symmetrical to PrmR was more frequent compared to the earlier results (Frolov et al., 2017a). Source 12 exhibited activity similar to the mu-rhythm, but unfortunately, the subjects for whom it was observed had no individual anatomical MRI scans, so the source could not be reliably localized. However, the SIL source was found for more than half of the recordings for which source 12 was extracted. Also, sources SIL and 12 were found together by the same method in 35% cases. These findings suggest that source 12 might have localization different from that of the SIL sources.

The average PSDs of the clustered component activities reflect the specificity presented in **Table 2**. The most specific components of the 5-th, 6-th, and 12-th clusters exhibit prominent rhythm desynchronization during the motor imagery. Notably, there is no evident lateralization in the average level of desynchronization between the hemispheres. There are different results on the mu-rhythm behavior in the ipsilateral hemisphere during the hand motor imagery. Our results agree with those obtained in Vasilyev et al. (2016), where ipsilateral rhythm suppression was observed. However, ipsilateral rhythm synchronization was observed in Nam et al. (2011). We suppose that the reason for the result disagreement lies in the BCI protocol. It seems that active motor imagery, when maintained

for a long time (from 10 s in these experiments to 20 s in our earlier NIRS study (Bobrov et al., 2016), involves activation of both hemispheres, either due to high concentration and the complexity of the task or due to the person's intention to "check" whether the requested hand movement has been imagined. The first explanation is supported by the results of Nam et al. (2011), where the ipsilateral rhythm synchronization was observed mainly in the end phase of brief motor imagery (1 s) trials, suggesting that brief and prolonged motor imagery may require different strategies. The latter explanation comes from some of the participants reporting that they had sometimes switched attention to the other hand to "make sure it isn't moving". Note, that both phenomena (different strategy and attention switch) may be present during prolonged motor imagery and underlie the absence of the asymmetry of rhythm suppression.

Unlike the SIL, SIR, and 12th sources, other sources do not exhibit such prominent rhythm desynchronization or synchronization during the motor imagery. However, the Prm, SMA, and Prc sources are often marked as being task-specific. These areas are often reported to be relevant to motor planning, execution, and imagery. The Prm sources are likely to reflect mirror neuron system activation when the image of the motion to be executed or imagined is generated to be compared to the incoming sensory information (Rizzolatti et al., 2014). The SMA source is likely to correspond to the supplementary motor area activation. This area is considered to be involved in the motion timing, motion sequence generation, or motion suppression (Rizzolatti et al., 2014). It was reported to be more active during the motor imagery than during real motion execution, suggesting that its main role is to suppress the actual movement when motor imagery is performed (Guillot et al., 2014). The occurrence of the Prm and SMA sources is not as high as that of the Prc and SI sources. Nevertheless, we believe that these areas are activated in all of the participants but that the activations are not often prominent in the electroencephalogram and thus are harder to find, even when using advanced techniques. This supposition

is supported by the results of numerous fMRI studies (Hétu et al., 2013) that report up to 34 areas being active during motor imagery.

In conclusion, we suggest the use of multi-method decomposition with subsequent component specificity estimation and clustering, focusing on the shared components. The primary methods to be used are PWCICA, AMICA, and MCSP, but utilizing other algorithms to support the findings seems to be reasonable.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available upon request by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics committee of Institute of Higher Nervous Activity and Neurophysiology of Russian Academy of Sciences. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

MI, YK, DB, and AL performed the experiments and collected the data. DB and AL designed the experimental software. PB, MI, and YK analyzed EEG data. AF and PB contributed to the interpretation of the data. AF, PB, and EB contributed to writing the manuscript. All authors approved the final version of the manuscript for submission.

FUNDING

REFERENCES

Ang, K. K., Chua, K. S. G., Phua, K. S., Wang, C., Chin, Z. Y., Kuah, C. W. K., et al. (2015). A randomized controlled trial of EEG-based motor imagery brain-computer interface robotic rehabilitation for stroke. *Clin. EEG Neurosci.* 46, 310–320. doi: 10.1177/1550059414522229

Ang, K. K., Guan, C., Chua, K. S. G., Ang, B. T., Kuah, C. W. K., Wang, C., et al. (2011). A large clinical study on the ability of stroke patients to use an EEG-based motor imagery brain-computer interface. *Clin. EEG Neurosci.* 42, 253–258. doi: 10.1177/155005941104200411

Bai, Z., Fong, K. N., Zhang, J. J., Chan, J., and Ting, K. (2020). Immediate and long-term effects of BCI-based rehabilitation of the upper extremity after stroke: a systematic review and meta-analysis. *J. Neuroeng. Rehabil.* 17, 1–20. doi: 10.1186/s12984-020-00686-2

Ball, K., Bigdely-Shamlo, N., Mullen, T., and Robbins, K. (2016). PWC-ICA: a method for stationary ordered blind source separation with application to EEG. *Comput. Intell. Neurosci.* 2016:20. doi: 10.1155/2016/9754813

Bashashati, A., Fatourechi, M., Ward, R. K., and Birch, G. E. (2007). A survey of signal processing algorithms in brain–computer interfaces based on

electrical brain signals. *J. Neural Eng.* 4, R32–R57. doi: 10.1088/1741-2560/4/2/R03

Bashashati, H., Ward, R. K., Birch, G. E., and Bashashati, A. (2015). Comparing different classifiers in sensory motor brain computer interfaces. *PLoS ONE* 10:e0129435. doi: 10.1371/journal.pone.0129435

Bell, A. J., and Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Comput.* 7, 1129–1159. doi: 10.1162/neco.1995.7.6.1129

Belouchrani, A., Abed-Meraim, K., Cardoso, J.-F., and Moulines, E. (1997). A blind source separation technique using second-order statistics. *IEEE Transac. Signal Proc.* 45, 434–444. doi: 10.1109/78.554307

Bingham, E., and Hyvärinen, A. (2000). A fast fixed-point algorithm for independent component analysis of complex valued signals. *Int. J. Neural Syst.* 10, 1–8. doi: 10.1142/S0129065700000028

Bobrov, P., Frolov, A., Husek, D., and Snasel, V. (2014). Clustering the sources of EEG activity during motor imagery by Attractor Neural Network with Increasing Activity (ANNIA). *Proc. Fifth Int. Conf. Innov. Bio Inspir. Comput. Appl.* 303, 183–191. doi: 10.1007/978-3-319-08156-4_19

Bobrov, P., Isaev, M., Korshakov, A., Oganesyan, V., Kerechanin, J., Popodko, A., et al. (2016). Sources of electrophysiological and foci of hemodynamic

brain activity most relevant for controlling a hybrid brain–computer interface based on classification of EEG patterns and near-infrared spectrography signals during motor imagery. *Hum. Physiol.* 42, 241–251. doi: 10.1134/S036211971603004X

Cichocki, A., and Amari, S.-I. (2002). *Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications.* Chichester: John Wiley and Sons.

Delfosse, N., and Loubaton, P. (1995). Adaptive blind separation of independent sources: a deflation approach. *Signal Proc.* 45, 59–83. doi: 10.1016/0165-1684(95)00042-C

Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21. doi: 10.1016/j.jneumeth.2003.10.009

Delorme, A., Palmer, J., Onton, J., Oostenveld, R., and Makeig, S. (2012). Independent EEG sources are dipolar. *PLoS ONE* 7:e30135. doi: 10.1371/journal.pone.0030135

Dharmaprani, D., Nguyen, H. K., Lewis, T. W., Delosangeles, D., Willoughby, J. O., and Pope, K. J. (2016). "A comparison of independent component analysis algorithms and measures to discriminate between EEG and artifact components," in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (Orlando, FL: IEEE), 825–828.

Fonov, V., Evans, A. C., Botteron, K., Almli, C. R., McKinstry, R. C., Collins, D. L., et al. (2011). Unbiased average age-appropriate atlases for pediatric studies. *Neuroimage* 54, 313–327. doi: 10.1016/j.neuroimage.2010.07.033

Fonov, V. S., Evans, A. C., McKinstry, R. C., Almli, C. R., and Collins, D. L. (2009). Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *Neuroimage* 47:S102. doi: 10.1016/S1053-8119(09)70884-5

Frolov, A., Aziatskaya, G., Bobrov, P., Luykmanov, R. K., Fedotova, I., Húsek, D., et al. (2017a). Electrophysiological brain activity during the control of a motor imagery-based brain–computer interface. *Hum. Physiol.* 43, 501–511. doi: 10.1134/S036211971705005X

Frolov, A., Husek, D., Bobrov, P., Korshakov, A., Chernikova, L., Konovalov, R., et al. (2012). Sources of EEG activity most relevant to performance of brain-computer interface based on motor imagery. *Neural Netw. World* 22, 21–37. doi: 10.14311/NNW.2012.22.002

Frolov, A.A., Husek, D., Muraviev, I.P., and Polyakov, P.Y. (2007). Boolean factor analysis by attractor neural network. *IEEE Trans. Neural Netw.* 18, 698–707. doi: 10.1109/TNN.2007.891664

Frolov, A.A., Mokienko, O., Lyukmanov, R., Biryukova, E., Kotov, S., Turbina, L., et al. (2017b). Post-stroke rehabilitation training with a motor-imagery-based brain-computer interface (BCI)-controlled hand exoskeleton: a randomized controlled multicenter trial. *Front. Neurosci.* 11:400. doi: 10.3389/fnins.2017.00400

Guillot, A., Di Rienzo, F., and Collet, C. (2014). "The neurofunctional architecture of motor imagery," in *Advanced Brain Neuroimaging Topics in Health and Disease-Methods and Applications*, eds T. D. Papageorgiou, G. I. Christopoulos, and S. M. Smirnakis (Rijeka: Intech), 433–456.

Hétu, S., Grégoire, M., Saimpont, A., Coll, M.-P., Eugène, F., Michon, P.-E., et al. (2013). The neural network of motor imagery: an ALE meta-analysis. *Neurosci. Biobehav. Rev.* 37, 930–949. doi: 10.1016/j.neubiorev.2013.03.017

Höller, Y., Bergmann, J., Thomschewski, A., Kronbichler, M., Höller, P., Crone, J. S., et al. (2013). Comparison of EEG-features and classification methods for motor imagery in patients with disorders of consciousness. *PloS ONE* 8:e80479. doi: 10.1371/journal.pone.0080479

Hyvarinen, A. (2001). Blind source separation by nonstationarity of variance: a cumulant-based approach. *IEEE Transac. Neural Netw.* 12, 1471–1474. doi: 10.1109/72.963782

Hyvärinen, A., Karhunen, J., and Oja, E. (2004). *Independent Component Analysis.* New York, NY: John Wiley and Sons.

Kachenoura, A., Albera, L., Senhadji, L., and Comon, P. (2007). ICA: a potential tool for BCI systems. *IEEE Signal Process. Mag.* 25, 57–68. doi: 10.1109/MSP.2008.4408442

Kobler, R., Hirata, M., Hashimoto, H., Dowaki, R., Sburlea, A. I., and Müller-Putz, G. (2019). "Simultaneous decoding of velocity and speed during executed and observed tracking movements: an MEG study," in *Proceedings of the 8th Graz Brain-Computer Interface Conference 2019: Bridging Science and Application.* Graz: Verlag der Technischen Universität Graz. 100–105.

Lee, F., Scherer, R., Leeb, R., Neuper, C., Bischof, H., and Pfurtscheller, G. (2005). "A comparative analysis of multi-class EEG classification for brain computer interface," in *Proceedings of the 10th Computer Vision Winter Workshop* (Graz), 195–204.

Lee, T.-W., Girolami, M., and Sejnowski, T. J. (1999). Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian sources. *Neural Comput.* 11, 417–441. doi: 10.1162/089976699300016719

Lotte, F., Bougrain, L., Cichocki, A., Clerc, M., Congedo, M., Rakotomamonjy, A., et al. (2018). A review of classification algorithms for EEG-based brain–computer interfaces: a 10 year update. *J. Neural Eng.* 15:031005. doi: 10.1088/1741-2552/aab2f2

Mane, R., Chew, E., Phua, K. S., Ang, K. K., Robinson, N., Vinod, A., et al. (2019). Prognostic and monitory EEG-biomarkers for BCI upper-limb stroke rehabilitation. *IEEE Transac. Neural Syst. Rehabil. Eng.* 27, 1654–1664. doi: 10.1109/TNSRE.2019.2924742

Nam, C. S., Jeon, Y., Kim, Y.-J., Lee, I., and Park, K. (2011). Movement imagery-related lateralization of event-related (de) synchronization (ERD/ERS): motor-imagery duration effects. *Clin. Neurophysiol.* 122, 567–577. doi: 10.1016/j.clinph.2010.08.002

Niedermeyer, E., and Da Silva, F. L. (2005). *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields.* Philadelphia, PA: Lippincott Williams and Wilkins.

Ono, T., Shindo, K., Kawashima, K., Ota, N., Ito, M., Ota, T., et al. (2014). Brain-computer interface with somatosensory feedback improves functional recovery from severe hemiplegia due to chronic stroke. *Front. Neuroeng.* 7:19. doi: 10.3389/fneng.2014.00019

Palmer, J.A., Kreutz-Delgado, K., and Makeig, S. (2006). "Super-Gaussian mixture source model for ICA," in *International Conference on Independent Component Analysis and Signal Separation* (Charleston, SC: Springer), 854–861.

Palmer, J.A., Kreutz-Delgado, K., and Makeig, S. (2012). *AMICA: An Adaptive Mixture of Independent Component Analyzers With Shared Components.* Swartz Center for Computatonal Neursoscience, University of California San Diego, Technical Report.

Palmer, J. A., Makeig, S., Kreutz-Delgado, K., and Rao, B. D. (2008). "Newton method for the ICA mixture model," in *Acoustics, Speech and Signal Processing. 2008. ICASSP 2008 (IEEE International Conference on: IEEE)* (Las Vegas, NV), 1805–1808.

Ramoser, H., Muller-Gerking, J., and Pfurtscheller, G. (2000). Optimal spatial filtering of single trial EEG during imagined hand movement. *IEEE Transa. Rehabil. Eng.* 8, 441–446. doi: 10.1109/86.895946

Ramos-Murguialday, A., Broetz, D., Rea, M., Laer, L., Yilmaz, O., Brasil, F.L., et al. (2013). Brain-machine interface in chronic stroke rehabilitation: a controlled study. *Ann. Neurol.* 74, 100–108. doi: 10.1002/ana.23879

Rizzolatti, G., Cattaneo, L., Fabbri-Destro, M., and Rozzi, S. (2014). Cortical mechanisms underlying the organization of goal-directed actions and mirror neuron-based action understanding. *Physiol. Rev.* 94, 655–706. doi: 10.1152/physrev.00009.2013

Sarin, M., Verma, A., Mehta, D. H., Shukla, P. K., and Verma, S. (2020). "Automated ocular artifacts identification and removal from eeg data using hybrid machine learning methods," in *2020 7th International Conference on Signal Processing and Integrated Networks* (SPIN) (Noida: IEEE), 1054–1059.

Tadel, F., Baillet, S., Mosher, J. C., Pantazis, D., and Leahy, R. M. (2011). Brainstorm: a user-friendly application for MEG/EEG analysis. *Comput. Intell. Neurosci.* 2011:879716. doi: 10.1155/2011/879716

Vasilyev, A., Liburkina, S., and Kaplan, A. Y. (2016). Lateralization of EEG patterns in humans during motor imagery of arm movements in the brain-computer interface. *Zhurnal Vysshei Nervnoi Deiatelnosti Imeni I.P. Pavlova.* 66, 302–312.

Vieira, S. M., Kaymak, U., and Sousa, J. M. (2010). "Cohen's kappa coefficient as a performance measure for feature selection," in *International Conference on Fuzzy Systems* (Barcelona: IEEE), 1–8.

Xiaopei, W., Zhou, B., Lv, Z., and Zhang, C. (2019). To explore the potentials of independent component analysis in brain-computer interface of motor imagery. *IEEE J. Biomed. Health Inform.* 24, 775–787. doi: 10.1109/JBHI.2019.2922976

Zhou, B., Wu, X., Lv, Z., Zhang, L., and Guo, X. (2016). A fully automated trial selection method for optimization of motor imagery based brain-computer interface. *PloS ONE* 11:e0162657. doi: 10.1371/journal.pone.0162657

Ziehe, A., Laskov, P., Nolte, G., and MãŽller, K.-R. (2004). A fast algorithm for joint diagonalization with non-orthogonal transformations and its application to blind source separation. *J. Mach. Learn. Res.* 5, 777–800. Available online at: http://www.jmlr.org/papers/v5/ziehe04a.html