Asynchronous BCI Based on Motor Imagery With Automated Calibration and Neurofeedback Training

Rafał Kuś, Diana Valbuena, Jarosław Żygierewicz, Tatsiana Malechka, Axel Gräser, *Member, IEEE*, and Piotr Durka

***Abstract*—A new multiclass Brain–computer Interface (BCI) based on the modulation of sensorimotor oscillations by imagining movements is described. By the application of advanced signal processing tools, statistics and machine learning, this BCI system offers: (a) asynchronous mode of operation (b) automatic selection of user-dependent parameters based on an initial calibration (c) incremental update of the classifier parameters from feedback data. The signal classification uses spatially filtered signals and is based on spectral power estimation computed in individualized frequency bands, which are automatically identified by a specially tailored AR-based model. Relevant features are chosen by a criterion based on Mutual Information. Final recognition of motor imagery is effectuated by a multinomial logistic regression classifier. This BCI system was evaluated in two studies. In the first study, five participants trained the ability to imagine movements of the right hand, left hand and feet in response to visual cues. The accuracy of the classifier was evaluated across four training sessions with feedback. The second study assessed the information transfer rate (ITR) of the BCI in an asynchronous application. The subjects' task was to navigate a cursor along a computer rendered 2D maze. A peak information transfer rate of 8.0 bit/min was achieved. Five subjects performed with a mean ITR of 4.5 bit/min and an accuracy of 74.84%. These results demonstrate that the use of automated interfaces to reduce complexity for the intended operator (outside the laboratory) is indeed possible. The signal processing and classifier source code embedded in BCI2000 is available from https://www.brain-project.org/downloads.html.**

***Index Terms*—Brain–computer interface (BCI), electroencephalography (EEG), event-related synchronization and desynchronization (ERD/ERS), motor imagery, neurofeedback.**

I. Introduction

BRAIN-COMPUTER Interfaces (BCIs) measure human brain activity to detect and discriminate the occurrence of specific phenomena in the brain allowing users to communicate or control external devices [1]. Although BCIs may measure brain activity through magnetoencephalography (MEG), functional magnetic resonance imaging (fMRI), functional near infrared imaging (fNIR), positron emission tomography (PET), electrocorticography (ECoG) or microelectrode recordings [2], most BCIs measure electrical potentials on the scalp (electroencephalography (EEG)) [3]. The main application of a BCI has been and is to control assistive devices and provide communication for severely disabled users. BCIs can be used for basic communication [4], control of external devices, such as wheelchairs [5], robot arms [6] or prosthetic devices [7]; and entertainment applications [8]–[10]. Further, recent progress in BCI research has broadened the field of applications, in which the principal goal of the BCI is not the communication but rehabilitation [11] and neuro-physiological regulation, which is known as neurofeedback [12]. A BCI system can be considered to be the most advanced neurofeedback system available.

One type of EEG-based BCI exploits the fact that during execution of different tasks the brain rhythms are specifically modulated. The specificity is in time-course, frequency ranges, and spatial organization of the modulation. The relative decrease of power of the oscillations is called event-related desynchronization (ERD); the opposite, increase of rhythmic activity is known as event-related synchronization (ERS) [13]. A robust effect is that both movement and the imagination of movement (motor imagery (MI)) of a limb are accompanied by decrease of power in μ (7–13 Hz) and β (13–30 Hz) frequency bands relative to the baseline level followed by a rebound of power in the beta band [14]. Moreover, due to somatotopic organization of the motor cortex the spatial patterns of ERD/ERS are specific for each limb. Because of its characteristics, motor imagery constitutes the most natural paradigm for building a BCI, but achieving an efficient MI-BCI is demanding for the user and challenging for the constructor. It is demanding for a user because, it usually requires repeated training to master the control of the MI for it to be precise enough to be useful (cf. [15]). It is challenging for the constructors since it requires more advanced signal processing tools. The most difficult issue in this paradigm is the selection of the individual set of features that provide the most efficient BCI. These features are most often some functions of the energy of EEG signal in a set of frequency-bands, spatial and time ranges.

Generally, there are two possible modes of a MI-BCI system [16]. The synchronous or cue-paced interface allows interactions only in fixed time windows determined by the BCI. This kind of interface is easier to implement, but less convenient to use. Knowledge of the onset of the MI simplifies the signal processing and classification, but forces the user to follow the rate of cues. The other mode, known as asynchronous MI-BCI, assumes that the user generates the MI in

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7) under grant agreement n° 224156 and from the German Federal Ministry of Economics and Technology (BMWi) under grant agreement n° 16136BG.

R. Kuś, J. Żygierewicz and P. Durka are with the Faculty of Physics, University of Warsaw, ul. Hoża 69, 00-681 Warsaw, Poland.

D. Valbuena was with the Institute of Automation, University of Bremen, Otto-Hahn-Allee 1, 28359 Bremen, Germany.

T. Malechka is with the Institute of Automation, University of Bremen and the Friedrich Wilhelm Bessel Institute m.b.H., Otto-Hahn-Allee 1, 28359 Bremen, Germany.

A. Gräser are with the Institute of Automation, University of Bremen, Otto-Hahn-Allee 1, 28359 Bremen, Germany.

a self-paced manner. In a system implementing this approach, the signal is analyzed continuously. One of the first approaches to develop asynchronous systems was presented in late 90s [17]. That system performed quasi-continuous classification of the user motor imagination and fed it back in a visual form. However, there were some limitations. The user was presented with the cue of the desired imagination. Next, he/she had to perform the motor imagination continuously until the end of the trial. The system was not designed to classify periods in which the user did not perform any motor imagination. Moreover, the Kalman filters used in that approach were evolving between the cue and the end of imagination period. Thus, the system can not be considered asynchronous and in reality does not allow a self paced usage. A system with similar behavior is implemented in the standard application of BCI2000 [18].

The present work presents advanced signal processing tools used to detect and discriminate different motor imagery classes in ongoing EEG. Our approach includes calibration runs where data are collected to set-up a subject-specific classifier (no feedback is provided), training runs where the classifier is used to provide feedback to the user while imagination is requested in a trial-based paradigm, and application runs where the user controls a cursor on a computer rendered 2D maze or any other available application in a self-paced manner. The interface used for user training is similar to the paradigm described in [17] with some essential differences. Each classification is performed independently of the preceding ones, and the system is designed to classify the non-control state of the subject. Moreover, the temporal information since the beginning of the trial, or the cue is not used by the classifier. The following aspects of the proposed MI-BCI system can be highlighted:

- the calibration function to fit the classifier parameters to the user is completely automated;
- the online classification of non-control and three motor imagery classes is asynchronous;
- the classifier can be updated with data recorded during training runs;
- the graphical user interface (GUI) provides online feedback to the user.

In the following section, the signal processing approach used in this work and the proposed methodology to evaluate the system are presented in detail.

II. Methods

A. BCI Signal Processing

This section describes the signal processing procedures used for calibration, i.e., obtaining individualized parameters for classification, updating these parameters after the subsequent user training, and online signal processing and classification for BCI operation. These issues are described in detail in subsections 1–3.

1) Calibration Algorithm: The aim of the calibration is to derive individualized parameterization of EEG signals. In fact, to construct the classifier, a separate parameterization is sought for discriminating each intentional control (IC) state, related to different MI, from the non-control (NC) state, and for discriminating among various IC states. In this work, three IC states related to imagination of movement of: left hand (IC$_{LH}$), right hand (IC$_{RH}$), and feet (IC$_F$), are considered. The parameterized data is subsequently used to train a hierarchy of logistic regression classifiers [19]. The calibration algorithm consists of the following main steps:

(a) Estimation of spatial filters.
(b) Identification of individualized μ and β frequency bands.
(c) Parameterization of EEG.
(d) Selection of features relevant for discrimination between NC and each of the IC states.
(e) Training of the logistic regression classifiers for discrimination between NC and each of IC states.
(f) Selection of features relevant for discrimination between the IC states.
(g) Training of the multinomial logistic regression classifier for discrimination between the IC states.

Each of these steps is described in detail in paragraphs *a)–g)* below.

a) Estimation of spatial filters: Differences of ERD/ERS spatial patterns for different limbs can be made more evident if the multichannel data is subjected to a suitable transformation [13]. Let us assume that there are m simultaneously active, statistically independent sources $\mathbf{s}[t] = \{s_1[t], \ldots, s_m[t]\}$. Then, we can consider the signal $x_i[t]$, recorded by the i^{th} electrode, as an instantaneous linear mixture of signals originating from the sources $\mathbf{s}[t]$. This is due to the linear superposition of electromagnetic fields and almost instantaneous propagation in the spatial scale of a head. This assumption leads to the model:

$$\mathbf{x}[t] = \mathbf{A}\mathbf{s}[t], \tag{1}$$

where $\mathbf{A}$ is the mixing matrix, which element a_{ij} describes the contribution of j^{th} source to the signal recorded by i^{th} electrode. The transformation[1] $\hat{\mathbf{P}} = \hat{\mathbf{A}}^{-1}$ can be used to find the unmixed signals that approximate the signals of independent sources:

$$\hat{\mathbf{s}}[t] = \hat{\mathbf{P}}\mathbf{x}[t]. \tag{2}$$

In real settings, we know neither the source signals $\mathbf{s}[t]$ nor the matrix $\mathbf{A}$. Only the signals $\mathbf{x}[t]$ are observed. The goal of estimating the approximated matrix $\hat{\mathbf{A}}$ and signal $\hat{\mathbf{s}}[t]$ from $\mathbf{x}[t]$ is known as the blind source separation (BSS) [20] (see also Appendix A). The individual unmixed signals $\hat{\mathbf{s}}$ are related to some abstract statistically independent sources.[2] Signals transformed by spatial filter $\hat{\mathbf{P}}$ approximate the statistically independent components which are more specific for different classes of motor imagery than the original signals. In the system presented in this work two transformations were implemented: one, $\hat{\mathbf{P}}_\mu$, for μ (7–13 Hz) and the other, $\hat{\mathbf{P}}_\beta$, for β (13–30 Hz) band. Thus in order to estimate the transformations, EEG data are band-pass filtered in the corresponding frequency ranges. Then, the IC epochs are extracted and split into partially overlapping, and quasi stationary time intervals.

[1] The ˆ symbol denotes estimated quantities.

[2] There is no direct correspondence between the physiological sources and the unmixed signals estimated by BSS since e.g., a number of correlated physiological sources can contribute to a single unmixed signal.

We found that 2 s intervals can be considered quasi-stationary. For each IC class and each time interval the covariance matrix is estimated. The transformation that jointly approximately diagonalizes the set of covariance matrices is the sought spatial filter $\hat{\mathbf{P}}$. It can be computed using the FFDiag [21] or JADiag [22] algorithm.

b) Identification of individual μ and β frequency bands: It is known that characteristic motor imagery causes modulations of EEG power spectra. In most BCI systems, the selection of features from the spectra requires interaction with an operator. A workaround of this problem was proposed in [23]. It relied on application of Adaptive Autoregressive Model (AAR) to analyze the ERD/ERS effect. The authors showed that power spectrum changes are reflected in changes of the autoregressive coefficients. However, ERD/ERS changes occur mainly in μ and β frequency bands. Analyses of the whole spectrum is not optimal and the exact range of these frequency bands vary between subjects. Thus, it is desirable to determine the individual limits of these bands. In this work, the individualized bands, parameterized by mean frequency and band-width, are identified automatically using the following procedure:

- From the raw calibration signals extract the NC epochs.
- Band-pass filter the data in the frequency range 5–35 Hz.
- Fit an autoregressive model (AR) for each channel separately. The order of the model is estimated by means of Schwarz criterion [24]:
$$SC(p) = \log(E) + \frac{2p\log(N)}{N} \tag{3}$$
where E is the residual variance, N is the number of samples in a single epoch. The selected order corresponds to the minimum of the SC curve for odd $p \in \{5, \dots, 15\}$.
- Decompose the transfer function of each AR model by means of frequency-amplitude-damping (FAD) method (see Appendix B). This decomposition yields a parameterization of spectral peaks in terms of their mean frequency and span.
- For each channel, select the spectral peaks localized within the μ and β frequency ranges.
- Combine the peaks in the μ band to obtain one μ_{ind} band common to all channels. The combination is done using the k-means method with number of cluster $k = 1$. Similarly the combination of the β band peaks into two β_{ind} bands is achieved by k-means analysis with $k = 2$.

c) Parameterization of EEG: At this stage of calibration the following computation is undertaken: spatial filters $\hat{\mathbf{P}}_\mu$ and $\hat{\mathbf{P}}_\beta$ for general μ (7–13 Hz) and β bands (13–30 Hz), and the ranges of individualized μ_{ind} and β_{ind} bands. The features are obtained for the intervals during the MI and NC epochs for each of the general frequency bands $b \in \{\mu, \beta\}$ using the following procedure:

- The raw calibration signals $\mathbf{x}[t]$ are first spatially filtered giving components $\hat{\mathbf{s}}_b[t]$ associated with the given frequency band b:
$$\hat{\mathbf{s}}_b[t] = \hat{\mathbf{P}}_b\mathbf{x}[t] \tag{4}$$
- For each i^{th} component $\hat{s}_{i,b}[t]$, compute the power spectrum estimate, using the multitaper method (MTM) (see Appendix C):
$$S_{i,b}(f) = MTM(\hat{s}_{i,b}[t]). \tag{5}$$
- Integrate the power spectrum in these individualized frequency ranges $R = \{\mu_{ind}, \beta_{ind}\}$, which are contained in the band b:
$$S_{i,R} = \sum_{f \in R} S_{i,b}(f). \tag{6}$$
- A feature is defined as:
$$X(i, R) = \log_{10} S_{i,R}. \tag{7}$$

As a result of the above procedure an interval of the EEG signal is parameterized by N = (number of channels)$\times$ (number of individualized frequency bands) features. The vector of features, evaluated for the epoch of EEG, gives a representation of this signal in the N dimensional vector space spanned by the features.

d) Selection of features relevant for discrimination between NC and each of the IC states: Since the classifier is meant to work in an asynchronous mode, the time intervals for which each type of motor imagery is most distinct from the NC state is determined. Features for discriminating between NC and each of the IC state are derived. However, due to the limited number of realizations of IC, the construction of a reliable classifier requires reduction of the dimensionality of the feature space. This reduction is reasonable also since some of the features are irrelevant to the classification of motor imagery, others may be redundant. A common approach to solve the problem of selection of an optimal subset of features is to use the mutual information between the feature and the class label. This method was also applied in this work and is described in detail in Appendix D. Finally, for each time interval, and for each IC_k ($k \in \{RH, LH, F\}$) separately, the following steps are undertaken:

- Extract data from all trials representing given time interval within the given IC_k.
- Extract data from all trials representing NC state.
- Evaluate features for the extracted data according to (4)–(7).
- Determine the reduced feature space (see Appendix D in (21)–(24)).
- Form two sets of reduced feature vectors: one for the IC_k, and the other for NC data.
- Compute the Mahalanobis distance (MD) (see Appendix E) between the two sets of vectors.

The best time interval is defined as the one which gives the greatest MD. As a result, for each IC class a reduced feature space optimally suited for discriminating this IC class from the resting state is determined. The sets of reduced feature vectors for the best time interval for the IC, and for the NC data are saved in a file for use in the on-line mode.

e) Training of the logistic regression classifiers for discrimination between NC and each of IC states: The set of features computed for the best time intervals for discrimination between the NC and IC state are used as training data to estimate parameters of the first stage of multinomial logistic regression classifiers.

f) Selection of features relevant for discrimination between the IC states: The data from the best time intervals of each IC class are subsequently used to select the reduced feature space for second stage classifier. The process of reduction of number of these features is as follows:

- Extract data from all trials representing best time intervals selected within the given IC_k.
- Evaluate features for the extracted data according to (4)–(7).
- Determine the reduced feature space (see Appendix D in (21)–(24)).

The sets of reduced feature vectors for discrimination of the IC data are saved in a file for use in the on-line classification.

g) Training of the multinomial logistic regression classifier for discrimination between the three IC states: The set of features identified for discriminating between various IC classes is used as training data to estimate parameters of the second stage multinomial logistic regression classifier.

2) Updating the classifier parameters: The EEG data recorded during the training runs can be used to update subject-specific parameters of the classifier, as also proposed in [25]. In our work, the re-estimation of the classifier parameters is based on all epochs of data from the initial calibration, and those epochs from previous and current training runs which were classified in accordance with the initial cue. Note, that during training runs the classifier does not use the information about the cue or the time since the cue onset. That is, the EEG data is continuously monitored, and the classifier first decides whether the user is currently trying to generate a control signal, then decide exactly which control signal is being generated. This information is passed to the training interface for providing feedback to the user and is stored for later offline use. The update of the classifier takes place offline. If the classifier output corresponds to desired MI-task, the previous 2 s EEG-data (length of the time window) was added to database of correct trials. The algorithm for the update of spatial filter, feature evaluation and selection, and the training of the logistic regression classifier are analogous to the ones used for calibration. The approach of estimating the classifier parameters based on calibration runs followed by training runs and subsequent re-estimation of the parameters encompasses both: development and strengthening of the features identified in the calibration run and also adjusting of the classifier parameters as the motor imagery patterns of the user evolve. This approach is meant also to help the subject in transition from synchronous calibration mode to truly asynchronous operation.

3) Online Classification Algorithm: The classification procedure operates in asynchronous mode. The BCI2000 system delivers a chunk of data in regular time intervals. These data are stored in a first-in-first-out (FIFO) buffer of the same length as used for time intervals by joint diagonalization procedure (in our case 2 s). The classification is based on the content of this buffer. The flowchart of the algorithm is presented in Fig. 1. The classification procedure consists of three stages:

a) Preprocessing: The data from the current FIFO buffer is transformed into features according to (4)–(7). As a result a full set of features is available.

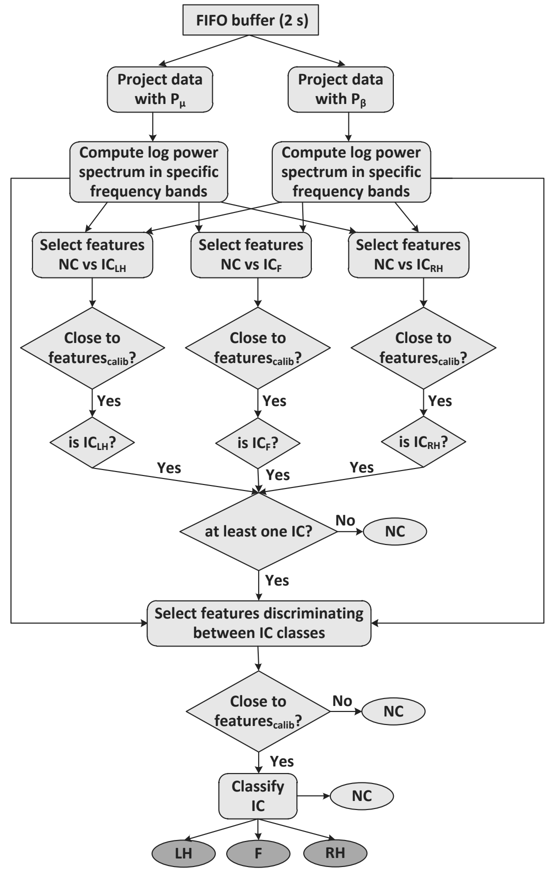


Fig. 1. Flowchart representing the decisions taken in the classification algorithm.

b) NC vs IC classification: At this point a decision as to whether the current FIFO buffer represents NC or some IC state is taken. For every IC_k class the following procedure is repeated:

- From the full feature set select the features identified to discriminate given IC_k state from the NC state.
- Check if the reduced feature vector of the data in the current FIFO buffer is close to the corresponding vectors obtained in calibration run. To accomplish this: Compute the Mahalanobis distances between the reduced feature vector and the corresponding vectors from the calibration run from the IC_k and from the NC state. If both MDs are larger than 97.5 quantile of the χ^2_ν distribution with ν degrees of freedom, where the number of degrees of freedom equals the dimension of reduced feature space, label the data as "unknown." Otherwise proceed to the next step.
- Apply the corresponding two-class logistic regression classifier. As a result the data are labeled as belonging to one of the classes: currently analyzed IC_k class, NC or "unknown." The "unknown" label is assigned in case when none of the probabilities of the data belonging to one of the classes passes the threshold defined as 0.5 plus the error of probability estimation.

After all IC classes have been checked the current FIFO buffer may have two possible types of label sets:

- all labels “unknown” or NC
- at least one label indicating IC state

For “unknown” or NC, the classification finishes at this stage and the current FIFO buffer is assigned the NC class. Otherwise, proceed with the second stage of classification.

c) Motor imagery classification: In this step, a decision is taken about the MI class that represents the current FIFO buffer. A procedure analogous to the one in previous stage is followed:

- From the full feature set select the features identified to discriminate between IC classes.
- Check if the features of the data in the current FIFO buffer are in the correct range. The Mahalanobis distance between the reduced feature vector and the corresponding set of reduced feature vectors from the calibration run representing the IC intervals should not be larger than the 97.5 quantile of the χ^2_ν distribution with ν degrees of freedom, where the number of degrees of freedom equals the dimension of reduced feature space. If the MD is too large the data is labeled as “unknown.” Otherwise proceed to the next step.
- Apply the logistic regression classifier. As a result, the data is labeled as belonging to one of the IC classes: LH, RH or F, or as “unknown.” The “unknown” label is assigned in case when none of the probabilities of the data belonging to one of the classes, i.e., does not pass the threshold defined as 0.5 plus the error of probability estimation. The “unknown” label is treated as NC.

B. Neurofeedback Interfaces

The neurofeedback interface is the component of the BCI system that provides user interaction. This interface is observed by the user during interaction to perform mental tasks and obtain performance feedback about the correct or incorrect response of BCI system. In the proposed system, the two neurofeedback interfaces are available: a training interface and a virtual maze. The training interface has elements typical for synchronous BCI. The system indicates the movement to be imagined, but the classification procedures don’t use the information about the cue or time since the beginning of the trial. The virtual maze works completely asynchronous, the user decides when to initiate the MI. Both interfaces are described in detail in the following subsections.

1) Training interface: This interface supplies cues for performing mental tasks and translates the logical control signals produced by the classifier into a graphical signal representation. The aim of this interface is to support users to improve control over the modulation of their neural activity, and teach them the brain-signal control, which is needed later to interact with a device (e.g. maze) via BCI. Learning to operate a BCI based on the modulation of sensorimotor rhythms requires repeated practice with feedback and reward [12]. Fig. 2 presents the training interface used to train users. The display contains three targets: left, up and right, each one corresponding to one of the following mental tasks: left hand, both feet and right hand imagery of movement, respectively. This training interface operates in a trial-based paradigm, i.e., it supplies blocks of randomized sequences of target codes (cues) for performing mental tasks. These sequences proceed in phases; each trial consists of a PreFeedback, Feedback, and Postfeedback phase, which is followed by an inter-trial interval (ITI). PreFeedback indicates the user’s task, i.e., to select the indicated target icon by imagining the corresponding movement. During Feedback, the result of the classification of the different subject intentions is visually fed back to the user. The classification of user actions and feedback are performed continuously (the incoming data block is analyzed continuously). The indicators (stripes) on each of the icons represent a result of single classification of the subject state as one of the IC classes. It means that the system has detected that the subject is not in an NC state and was able to assign an IC class to the current data buffer. Only when a desired number of stripes are collected, the selection of the icon is performed. The selection of a target occurs during PostFeedbak when an arrow reaches the desired level, then the chosen target changes color to confirm selection. For the feedback phase there is maximum time to perform the imagination. If no selection is performed within the maximum feedback duration, the selection is aborted. Positive or negative feedback (visual and acoustic) is presented to the user depending on the result of the selection. When the selected icon is the one suggested by the cue, the icon is colored green and the message “You got it” is displayed on it. This situation is called a hit. For a miss, i.e., when the selected icon differs from the suggested one, the selected icon color turns red and the message “Try again” appears on it. During the ITI the subject is requested to rest. A training session consists of several trials.

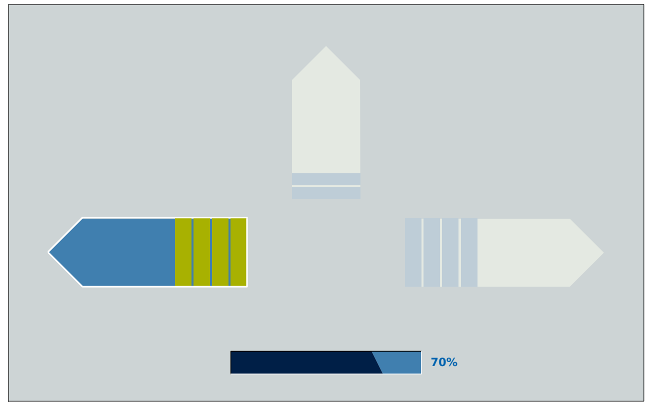


Fig. 2. The display used for user training of three motor imagery tasks. This interface provides feedback about the classifier result every 125 ms. Each stripe represents a classification.

2) Virtual maze: The virtual maze is an asynchronous interface. The user performs the imagination of movements in a self-paced manner and has the opportunity to correct or modify his/her behavior based on the continuous feedback. At the beginning of each run, the cursor is located at the “Start” position. The subject’s task is to navigate the cursor through the maze until the “Finish” position is reached (see Fig. 3). The “move forward” (MI feet) command moves the cursor to the next position in the maze. Commands “turn left” (MI left hand) and “turn right” (MI right hand) rotate the cursor

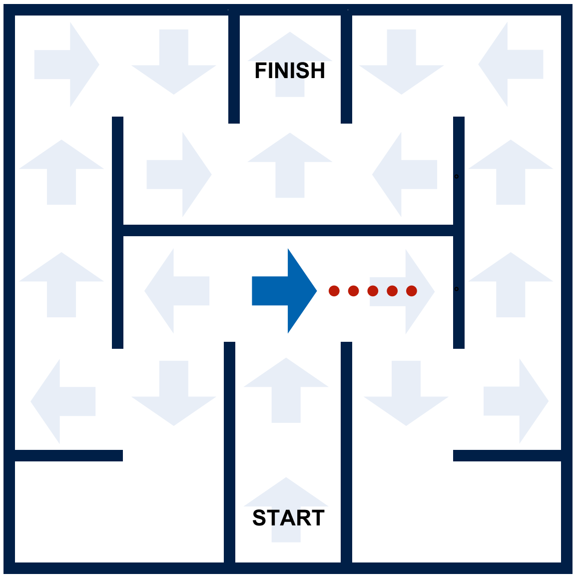


Fig. 3. Virtual maze application. At the beginning of each run, the cursor is located at the "start" position. With "move forward" command the cursor moves to the next position in the maze. With commands "turn left" and "turn right" the cursor rotates by 90 degrees to the left and to the right, respectively.

by 90 degrees to the left and to the right, respectively. A run starts with "move forward" command. From this position, the minimum of 19 commands is needed in order to reach the final position. The run ends automatically after the cursor has reached the "Finish" position.

C. Subjects

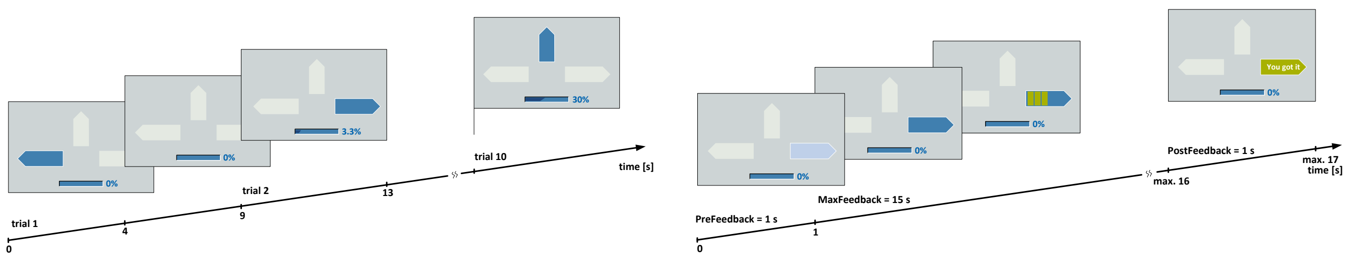
A total of eight able-bodied subjects were recruited for the evaluation of the MI-BCI system. First, five participants (aged 24.2 ± 4.5 years; 4 women and 1 male) were recruited for the training study, each knowing that they would be required to undertake several phases of recording which would include calibration and training. In total four sessions were recorded on four different days. Two of the subjects who participated in the training study and three new subjects agreed to participate in the application study (aged 25.8 ± 3.8 years; 3 women and 2 male). They undertook calibration runs, followed by training runs and finally they controlled the virtual maze via BCI. All was done in one single session. All subjects were right-handed, had normal or corrected to normal vision. According to self-reports, Subject H suffered from acusticus neurinom. Six of the subjects had no prior experience with BCI. Two subjects had experience with SSVEP-BCI (Subjects A and E) and one with MI-BCI (Subject A).

D. Data Collection

Data were recorded from the surface of the scalp via 16 sintered Ag/Ag-Cl EEG electrodes. They were placed on AF_Z for ground and $F_Z, FC_3, FC_z, FC_4, C_5, C_3, C_1, C_z, C_2, C_4, C_6, CP_3, CP_z, CP_4, P_z$ as the input electrodes on the 10–10 international system of EEG measurement [26]. All impedances were kept below 5 KΩ. An EEG amplifier Porti32 (Twente Medical Systems International, Oldenzaal, Netherlands) was used for these experiments. The Porti32 amplifier records unipolar inputs configured as the reference amplifier, i.e., all channels are amplified against the average of all connected inputs. The signals were digitized with a sampling rate of 256 Hz and transmitted to the computer by means of a bi-directional glass fibre. The Porti32 was then connected to the USB port of a desktop quad core computer running the BCI2000 general software framework [18]. BCI2000 consists of four modules: source, signal processing, application and operator. The operator used was the release provided by the BCI2000 developers; all other modules were reimplemented by the authors. The source module acquired signals from the Porti32 amplifier in blocks of 32 samples and applied a high-pass filter at 0.1 Hz to reduce signal offset and a notch filter at 50 Hz to reduce the power line noise. The acquired signals were then transmitted to the signal processing module, which collected the EEG signals in segments of 2 seconds, the data buffer for data processing. The buffer's overlap was 125 ms. The signal processing module implements the methodology presented in section II-A3 using the Matlab interface of BCI2000. The training interface and virtual maze were implemented as external modules in C++. They received control signals from BCI2000 via the User Datagram Protocol (UDP).

E. Experimental Protocol

1) Training study: Subjects participated in four training sessions on different recording days. The first session started for all subjects with four calibration runs consisting of 30 trials each (120 trials in total). The number of trials per task was the same for all tasks in a run (10 trials). The mental tasks were the mental imagination of the right hand, left hand and feet movement. The subject were asked to avoid the limb movements during imagination task. After each calibration run, a break of five minutes followed. During a calibration run, the display initially showed three icons in form of arrows colored light gray, and pointing left, up and right. Each trial started when one of the icons changed its color to blue indicating the mental task to perform. The target cue was displayed on the screen for four seconds. The time between the trials (ITI) was five seconds. Subjects were instructed to imagine left/right hand or feet movements according to the cue presented on the screen as shown in Fig. 4a. Specifically, they were given the instruction to imagine the continuous opening and closing of the left/right hand (left/right arrow), and to imagine to grip an object with both feet (up arrow). The data recorded from the four calibration runs were concatenated in a single file, which was given as input argument to the procedure described in II-A1. This step was necessary to adapt the classifier parameters to the individual user. Subject-specific parameters, like μ and β frequency bands, transformation matrix $\hat{\mathbf{P}}$ and significant features were used to train the logistic regression classifier. The created classifier was applied to the calibration data to estimate the quality of classification. All the correct estimated trials were stored in a database. After calibration, subjects proceeded with at least three training runs with feedback (30 trials per run). The intent of the training was to improve the user's skill level. A training run proceeded in



(a) Calibration run

(b) Training run

Fig. 4. Trial-based paradigm for calibration and training. In both, cue stimulus in form of arrows pointing left/right/up indicated the subject the kind of imagination to perform left hand/right hand/feet, respectively. (a) For calibration, subjects performed the imagination continuously for four seconds, no feedback was provided. (b) For training, subjects were instructed to perform the imagination until the corresponding arrow was completely filled (six indicators). A maximum of 15 seconds was given to the user to complete the task.

phases as explained in section II-B1. The PreFeedback time of 1 second indicated the subject to be ready for the trial. In the feedback phase, subjects were instructed to perform the desired motor imagery task until the corresponding arrow was filled with six indicators (counted as a hit), another arrow was filled (counted as a miss), or the maximum feedback duration of 15 seconds expired (counted as an invalid trial). The PostFeedback phase showed the result of the selection for 1 second. The interval between the trials was 5 seconds. Fig. 4b shows the timing of the training run. The subject-dependent classifier was recalibrated whenever was possible, i.e., if enough correct classified trials were available from the training runs. The recalibration was done using the procedure described in II-A2. All correct trials were stored again in the database for each subject. The next three training sessions consisted of six training runs unless a new calibration was required. That depended on the performance of the subject. If a calibration was conducted, then three training runs were available.

2) Application study: The objective of this study was to demonstrate direct brain control, i.e., the normal activity of a user performing tasks in an (unconstrained) environment. The interface paradigm was object positioning through virtual manipulation and the temporal control paradigm was self-paced (continuously available–idling supported). Before the run with the virtual maze could begin, the BCI system was first fitted to the specific individual. Subjects had to perform four calibration runs and three training runs. The paradigm, timing and the number of trials for calibration and training runs were configured as in the training study. The subject-specific parameters of the classifier were estimated based on data recorded during calibration runs. The data from two training runs was used to update the classifier. This update compensated the changes in the signals between runs from the calibration and training. The third training run and application run were performed using the adjusted classifier. For the application run, the participants used the asynchronous MI-BCI to navigate the cursor through the maze. The virtual maze provided the user with a continuous visual feedback. The classifier output in the form of red dots were used as feedback: dots ahead of the cursor for feet MI, on the left/right side of cursor for left/right hand MI (see Fig. 3). To move the cursor, the user had to perform imagination until eight dots were collected. The run ended when the cursor was located at the “Finish” position in the maze. All runs were performed on the same day with an intermediate break of a few minutes. The entire procedure took about one hour on average per subject.

F. Measures of BCI Performance

The performance of a BCI system is usually assessed by the calculation of the classification accuracy (ACC) and information transfer rate (ITR). To measure the performance of classification in training runs, we calculated the specific accuracy for each class i, as:

$$ACC_i = 100 \times \frac{hits_i}{trials_i}. \tag{8}$$

The ITR measures the rate at which information is successfully transmitted from the user, through the BCI channel, to the application [27]. For application runs with the virtual maze, the ITR was calculated based on the following formula:

$$B = \log_2 N + P \log_2 P + (1-P) \log_2 \left(\frac{1-P}{N-1}\right). \tag{9}$$

In this formula, B represents the number of bits per trial, P represents the probability of correct classification and N is the number of choices. ITR was calculated on the basis of the stages necessary to reach the goal, i.e., the low level commands sent to the transducer. This lead to an N of 3, based on the 3 movement commands (“turn left,” “move forward” and “turn right”). P was calculated as the number of all correct commands (hits) divided by the total number of executed commands (trials). To obtain ITR in bits per minute, B was multiplied by the speed, i.e., the number of executed commands divided by the total duration of the run (time). The three variables hits, trials and time required to calculate ITR were calculated online directly in the maze software. The total time was measured as the elapsed time since the cursor left the “Start” position (indicated by the first command “move forward”) until reaching the “Finish” position. Each executed command was rated as correct or incorrect by comparing the current cursor position and direction of the cursor, with the next target position. For example, after leaving the “Start” position, only the “move forward” command, is correct, all other commands are treated as incorrect for the calculation of accuracy. If the cursor was facing a wall, and “move forward” command was executed, the command was also counted (as incorrect), but the cursor did not move forward. The user could choose one of the two paths to reach the goal. Both paths were valid. Furthermore, for the application study we evaluated the mean time needed to issue a correct movement command denoted as $\bar{t}_{hit}$. The time periods used for this calculation were counted from the previous to the subsequent cursor movement.

III. Results

A. Training study

Table I summarizes the results obtained in the training study with five subjects. This table presents the specific accuracy for each motor imagery class and the overall mean accuracy achieved in each training session. A session is constituted by several training runs with 30 trials each. Additionally, the best training run for each subject with the maximum accuracy is highlighted. The accuracy was defined as the percentage of correct trials divided by the total number of trials (valid and invalid trials). An invalid trial was considered when the subject did not reach the desired level of indicators for any of the classes during the maximum feedback duration. Only the periods of time in which the user was engaged with the imagination of movements were evaluated, i.e., during the feedback phase. During the ITI period, the subject did not received feedback and was instructed to rest and to prepare for the next trial. These results of accuracy for each class helped the operator to decide when to recalibrate the system. This procedure was not straightforward, it depended on the subject's performance and therefore different for most of the participants. Subjects A and B only required a calibration of the system in the first session. For the next sessions, the original classifier settings were only updated always after each training run using the function described in II-A2. Subject C required a recalibration in session S4 as the accuracy for right hand imagination dropped to 3.3% in session S3. Similarly, Subject D required a recalibration in session S3 as the classifier was not detecting feet imagination in session S2. Subject E required a new calibration of the system at the beginning of session S1 to S3. Only for session S4, an update was sufficient BCI operation. Fig. 5a presents the BCI performance for a representative subject (B) in the course of the training sessions. Subject B completed three feedback runs (90 trials) in the first session and six feedback runs (180 trials) in each of the next training sessions. This figure shows clear variations across the training runs. Subject B could increased his accuracy from 64.4% to 74.4%. Fig. 5b shows the learning curves for each subject during four training sessions. These results show considerable intersubject differences, as widely reported in other BCI studies [28].

TABLE I
Training study: Results of five subjects during four training sessions

Subject	Session	ACC_i [%] LH	F	RH	Mean accuracy [%]
A	S1*	55.0	63.3	83.3	67.2
	S2	50.0	70.0	85.0	68.3
	S3	46.7	70.0	85.0	67.2
	S4	60.0	66.7	91.7	72.8
	Mean	52.9	67.5	86.3	68.9
	Max	**70.0**	**70.0**	**100.0**	**80.0**
B	S1*	40.0	90.0	63.3	64.4
	S2	56.7	53.3	83.3	64.4
	S3	65.0	53.3	85.0	67.8
	S4	78.3	75.0	70.0	74.4
	Mean	60.0	67.9	75.4	67.8
	Max	**80.0**	**80.0**	**100.0**	**86.7**
C	S1*	40.0	90.0	55.0	61.7
	S2	78.3	93.3	6.7	59.4
	S3	100.0	70.0	3.3	57.8
	S4*	80.0	66.7	33.3	60.0
	Mean	74.6	80.0	24.6	59.7
	Max	**70.0**	**90.0**	**80.0**	**80.0**
D	S1*	12.5	37.5	87.5	45.8
	S2	71.7	0.0	43.3	38.3
	S3*	56.0	48.0	78.0	60.7
	S4	68.3	26.7	78.3	57.8
	Mean	52.1	28.0	71.8	50.7
	Max	**70.0**	**80.0**	**60.0**	**70.0**
E	S1*	20.0	30.0	5.0	18.3
	S2*	16.7	53.3	20.0	30.0
	S3*	53.3	30.0	45.0	42.8
	S4	66.7	71.7	18.3	54.2
	Mean	39.2	46.3	22.1	35.8
	Max	**100.0**	**80.0**	**0.0**	**60.0**

* This session started with a calibration of the system

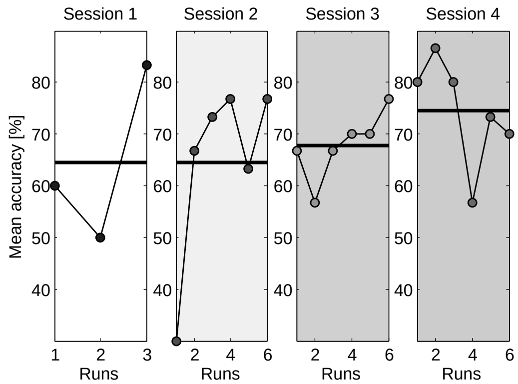
B. Application study

The results of the application study are presented in Table II. This table shows the results of the training runs and the application run with the virtual maze. The results of the training runs are presented in the same form as in the training study. For the run with the virtual maze, four measures of performance were available: accuracy *(hits/trials)*, overall time for the maze task, mean time needed to issue a correct movement command $(\bar{t}_{hits})$ and ITR. The calculation of information transfer rate was specified by using the formula described in II-F. Subjects achieved a mean information transfer rate of 4.51 bit/min and a mean accuracy of 74.84%. The time needed to complete the task was 5.27 minutes on average. The subjects spent on each correct movement command 8.81 s on average. The mean number of commands required to navigate the cursor to the final position was 35.6, what significantly differs from best case scenario with only 19 commands. This can be explained by the maze layout: A false command at some places in the maze must be corrected by three following commands.

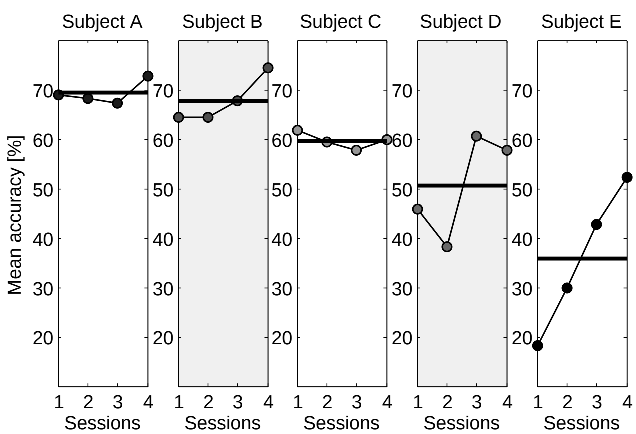
Subject A achieved the maximum ITR of 8.00 bit/min and accuracy of 81.8% and could finish the task in only 2.97 minutes. From the table, it can be observed that Subject H could achieve an overall accuracy of 70% although one of the conditions was very week. This was possible by avoiding turning left. Results from subjects F, G and H are remarkable as they used the system for the first time and had no prior experience with BCIs.

IV. Discussion

The main contribution of this work is the development of a multiclass asynchronous BCI based on motor imagery, in which the selection of individual characteristics of each user and update of the classifier settings using data from the training



(a) Subject B



(b) All subjects

Fig. 5. Classification accuracy during training sessions. (a) Performance of a representative subject in the course of four training sessions. The training runs in one session were acquired the same recorded day. The thick line is the mean accuracy in %. (b) Mean accuracy for each subject in the course of four training sessions.

TABLE II
APPLICATION STUDY: RESULTS OF THE TRAINING AND APPLICATION RUNS FOR FIVE SUBJECTS

Subject	Training runs ACC_i [%]			Application run ACC [%] *(hits/trials)*	Application run Time [min] ($\bar{t}_{hits}$ [s])	Application run ITR [bit/min]
	LH	F	RH			
A	60.0	90.0	50.0	81.8 *(27/33)*	2.97 *(5.57)*	8.00
E	13.3	73.3	66.7	77.4 *(24/31)*	4.78 *(9.29)*	3.82
F	73.3	56.7	86.7	76.2 *(32/42)*	3.97 *(5.29)*	5.87
G	30.0	56.7	10.0	68.8 *(22/32)*	10.28 *(17.54)*	1.17
H	6.7	96.7	60.0	70.0 *(28/40)*	4.35 *(6.42)*	3.71
Mean	**36.7**	**74.7**	**54.7**	**74.84** ***(26.6/35.6)***	**5.27** ***(8.81)***	**4.51**

sessions is done fully automatic by the software. This may reduce the complexity of this paradigm for non BCI experts who intend using the system outside the laboratory. The performance data show that this MI-BCI system could provide effective control for all five subjects participating in the application study (trained and untrained). In this section, two main characteristics of the BCI are evaluated, the automatic selection of subject-specific parameters and the multiclass detection, which makes an important contribution to the BCI state-of-the-art. Finally, the current trends in MI-BCIs are assessed.

A. Automatic selection of subject-specific parameters

The brain oscillations related to the imagination of limb movement varies from one subject to another. As a result, the BCI system must be adapted individually to each user. To our knowledge, most MI-based BCI systems the features were selected manually based on the visual analysis of time-frequency-spatial properties of the signals by experts. This procedure is extremely time-consuming and requires an expert with technical and scientific knowledge of BCIs. Today, there exists only few papers ([29], [30]), which address the automatic or semi-automatic feature selection. In this paper a solution which automates most of these tasks is proposed:

- adjustment of the individual frequency-bands,
- selection of optimal linear combination of available channels,
- selection of relevant and not redundant set of features for motor-imagery classification,
- search for the best time interval for training of two stage logistic regression classifier.

The results showed that with the automatic calibration, the subjects in this study were able to achieve a good accuracy already in the first session ($> 70\%$).

B. Multiclass MI-BCI

The mainstream of research has been focused on the development of signal processing methods that improve the feature extraction and/or classification [31]–[34]. The research presented in these papers has been done with offline data analysis using EEG data that were previously recorded. A comparison of performance of many different algorithms on the same data sets can be found at http://www.bbci.de/competition/iv/results/index.html. Today, there exists only few studies, which have investigated the performance of three or more MI classes. Scherer *et al.*, [35] used the smiley paradigm as a cue-based training interface for three-class MI-BCI. Three subjects achieved the mean classification accuracies of 75%, 80% and 60% during feedback training. Galán *et al.*, [5] and Milán *et al.*, [36] combined motor imagery BCI with other mental tasks like words association, cube rotation and subtraction. In our work, we demonstrate that three class control is possible. Although two subjects (E and H) had problems with a third MI class. This results also suggest that the BCI application should be adapted to the user according the number of MI classes.

C. Current trends in MI-BCI research

Recently, there is an increase interest in systems working asynchronously. Basically, two directions of development within this paradigm can be identified. One is the minimalistic approach with as few electrodes as possible and simple signal processing. This line of development is aimed at construction of robust "brain switch" [37], [38]. The other direction is aimed at the construction of a system that can be operated in an asynchronous way and is able to distinguish multiple classes of MI. This direction needs more sophisticated signal processing and additional electrodes. It is common for this approach to divide the classification process into two stages: the classification of the NC vs IC state, and in case IC state was detected the second stage classifier determined the actual IC state. Of course, the increase of classes impairs the classification accuracy. Despite this using more classes has the potential to increase information transfer rate.

In general, there are two main concepts of the cooperation of the user with the computer. The first idea, "let the machines learn" promotes the approach in which the machine learning algorithms are meant to adjust classification to the patterns generated by user in the calibration session. The fact that users patterns may evolve during the usage of the system is neglected. In the opposite idea, "let the user learn," the subject has to train in neurofeedback sessions the generation of patterns hardwired in the system. The system proposed in this paper utilizes both concepts. The features and the classifier parameters are set with the help of machine learning. The evolution of the pattern generated by the user and a reestimation of the parameters after each training session are utilized. The reestimation uses all the data that was correctly classified in all performed sessions. The fact that only the correctly classified data is used to update of the parameters has two positive consequences. First, it encourages the user to strengthen the classifiable patterns. Second, if the patterns evolve during the training sessions, the reestimation of the classification parameters allows system to capture the changes. The learning curves of most of the subjects show a positive effect on the performance. This could be explained by the incremental update of the classifier parameters and the development of successful strategies by the subject.

V. Conclusion

An asynchronous BCI system tailored for detecting ERD/ERS changes in online EEG data has been presented. The signal processing methodology automatically selects user-dependent parameters from calibration data and allows the incremental update of those parameters based on data recorded from synchronous training sessions prior to operation. When using the system to navigate a cursor along a virtual maze, a good performance in terms of accuracy (74.84%) and information transfer rate (4.51 bit/min) was achieved. This result is comparable with systems that instead require experts to select the optimal user parameters. This research facilitates BCIs that can adapt to each user with little or no expert help.

Appendix A
Blind source separation by approximated joint diagonalization

The problem of finding the demixing matrix $\hat{\mathbf{P}}$ in (2) can be formulated with the formalism of general joint diagonalization. Let us consider the spatial covariance matrix of the mixed signals $\mathbf{x}[t]$:

$$\mathbf{C}_x = E\left\{\mathbf{x}[t]\mathbf{x}[t]^T\right\} \tag{10}$$

where the expectation is taken over time t. Substituting (1) to (10) we obtain:

$$\mathbf{C}_x = E\left\{\mathbf{A}\mathbf{s}[t]\left(\mathbf{A}\mathbf{s}[t]\right)^T\right\} = \mathbf{A}E\left\{\mathbf{s}[t]\mathbf{s}[t]^T\right\}\mathbf{A}^T = \mathbf{A}\mathbf{C}_s\mathbf{A}^T \tag{11}$$

For independent sources the off-diagonal elements of matrix $\mathbf{C}_s$ should be zero. Thus the matrix $\mathbf{P}$ has the property that it diagonalizes the signal covariance matrix $\mathbf{C}_x$:

$$\hat{\mathbf{A}}^{-1}\mathbf{C}_x\left(\hat{\mathbf{A}}^T\right)^{-1} = \hat{\mathbf{P}}\mathbf{C}_x\hat{\mathbf{P}}^T = \mathbf{C}_s \tag{12}$$

This property can be used to estimate the matrix $\mathbf{P}$ from the data. The above derivation is often applied in the construction of spatial filters for BCIs. Let's assume that:

- there are K different MI classes—corresponding to imaginations of movement of different limbs,
- the spatio-temporal pattern of band power modulation for each of the classes is related to the same set of sources related to motor imagery, but the modulations of activity of different subsets of these sources are specific for each motor imagery.
- the time course of the power modulation can be split into intervals in which the data are quasi-stationary.

Then the transformation $\hat{\mathbf{P}}$, which jointly diagonalizes the covariance matrices and which is estimated for all quasi-stationary intervals of the motor imagery epochs, is an approximation of the demixing matrix $\mathbf{P}$ [31].

Appendix B
Frequency-Amplitude-Damping decomposition of an autoregressive model

Let's consider an autoregressive model of the form:

$$x[n] = \sum_{i=1}^{p} a[i]x[n-i] + e[n] \tag{13}$$

where $x[i]$ is the value of the i-th sample, $a[i]$—the i-th coefficient of AR model, $e[n]$— random component, p—the order of the AR model. After applying the Z-transform to (13) we get:

$$x[z] = \sum_{i=1}^{p} a[i]x[z]z^{-i} + e[z] \tag{14}$$

This equation can be solved for $x[z]$:

$$x[z] = \frac{e[z]}{1 - \sum_{i=1}^{p} a[i]z^{-i}} = h[z]e[z] \tag{15}$$

where

$$h[z] = \frac{1}{1 - \sum_{i=1}^{p} a[i]z^{-i}} \tag{16}$$

is called the transfer function of the autoregressive model. The poles of the transfer function are connected with the maxima in the power spectrum [39], thus they correspond to the relevant rhythms present in the signal. Under the assumption that $h[z]$ has only single poles[3], (16) may be written in the form:

$$h[z] = \sum_{j=1}^{p} C_j \frac{z}{z - z_j} \tag{17}$$

[3]Multiple poles might happen in the EEG signal, when order of model is overestimated

where z_j is the j-th pole of $h[z]$,

$$C_j = \lim_{z \to z_j} \frac{(z - z_j)h[z]}{z}$$

By means of the z_j and C_j coefficients, the spectral peak parameters: mean frequency, bandwidth, amplitude and phase, may be estimated. The frequency of the j-th oscillation is:

$$\omega_j = \text{imag}(\alpha_j)\frac{f_s}{2\pi} \tag{18}$$

where f_s—sampling frequency, $\alpha_j = \ln(z_j)$

The full width at half maximum (FWHM) of the corresponding spectral peak is proportional to $2\beta_j$, where β_j is:

$$\beta_j = -\text{real}(\alpha_j)\frac{f_s}{2\pi} \tag{19}$$

Appendix C
Multitaper method

The multitaper method (MTM) described in [40], uses a sequence of windows that are orthogonal to each other (discrete prolate spheroidal sequences). Each window is used to compute the windowed periodogram of the signal. Subsequently the periodograms are averaged.

In the current paper we use a number of windows in the sequence, N_w, which is related to the length of the time window Δt and to the width, Δf, of a spectral peak estimated by means of FAD:

$$N_w = 2\Delta t \Delta f - 1 \tag{20}$$

The width of the spectral peak Δf is constrained not to be lower than the spectral resolution of the periodogram ($\frac{1}{\Delta f}$).

Appendix D
Reduction of feature space dimensionality

Mutual information $I(X;Y)$ measures the amount of information shared by random variables X and Y. Formally, for discrete random variables it is defined as:

$$I(X;Y) = \sum_{y \in Y} \sum_{x \in X} p(x,y) \log \frac{p(x,y)}{p_1(x)p_2(y)} \tag{21}$$

where $p(x,y)$ is the joint probability distribution of X and Y, and $p_1(x)$ and $p_2(y)$ are the marginal probability distribution functions of X and Y respectively. We will designate features as X and class label as Y. Let the classifier function $g(X)$ assign a class label to feature. The minimal achievable classification error is bounded by two inequalities, lower bound is the Fano's inequality [41], and the upper bound is half of the conditional entropy [42]:

$$\frac{H(Y) - I(X;Y) - 1}{\log |Y|} \leq P(g(X) \neq Y) \leq \frac{1}{2} H(Y|X) \tag{22}$$

where H is the information entropy and $|Y|$ is the number of elements in Y. The left inequality indicates that the maximization of mutual information minimizes the bound. Whether the minimum can be reached depends on the classification function $g(X)$. For a set of features $\mathcal{X} = \{X_1, \ldots, X_n\}$ the selection of features should aim at maximization of the joint mutual information $I(X_{1:n};Y)$. Unfortunately, the evaluation of joint mutual information involves high dimensional distributions, which in practical cases cannot be estimated reliably. The Shannon mutual information can be expanded into a sum of interaction information terms over all possible subsets T drawn from S [42]:

$$I(X_{1:n};Y) = \sum_{T \subseteq \mathcal{X}} I(\{T \cup Y\}), \quad |T| \geq 1 \tag{23}$$

Keeping only terms for $|T| \leq 2$ is equivalent to assume that there are only conditional and unconditional pairwise relations, and no higher order relations. Such a truncated expansion can be used to form a ranking measure:

$$J_n = I(X_n;Y) - \beta \sum_{k=1}^{n-1} I(X_n;X_k) \tag{24}$$

The above formula includes the objective term $I(X_n;Y)$ to ensure feature relevance and the penalty term $\sum_{k=1}^{n-1} I(X_n;X_k)$ to enforce low feature redundancy. We used the formula (24) with $\beta = \frac{1}{n-1}$. Such a formula was proposed by Peng et al. [43] as Maximum Relevance Minimum Redundancy (MRMR). It can be understood as taking the mean of the redundancy terms. In the iterative way consecutive features with the highest ranking J are selected. The iterations are stopped at k for which $J_k < 0$. To improve the generalization of the classifier on a new data set, the list of selected features is truncated at the number corresponding to 10% of the number of available calibration trials. Our numerical experiments confirm in this respect results obtained by [44].

To make use of criterion (24), we need to estimate the probabilities in formula (21). Due to the limited number of MI realizations, the values of features need to be quantized. Based on our numerical experiments we decided that the maximal number of quantization levels should not exceed 5% of available observations.

Appendix E
Mahalanobis distance

Mahalanobis distance [45] is a distance measure in a multidimentional vector space. It can be used to determine similarity of a vector $\mathbf{x} = \{x_1, x_2, \ldots, x_n\}$ to a given set of vectors A. It differs from Euclidean distance in that it takes into account the correlation structure of the vector set A and is scale-invariant. The Mahalanobis distance is defined in the following way:

$$d = \sqrt{(\mathbf{x} - \mu_A)^T C^{-1} (\mathbf{x} - \mu_A)} \tag{25}$$

where $\mu_A = \{\mu_1, \mu_2, \ldots, \mu_n\}$ is the mean vector averaged over the group of the vectors in A, C — covariance matrix of vectors in A.

Mahalanobis distance can be defined also to measure the distance between two sets of vectors A and B:

$$d = \sqrt{(\mu_A - \mu_B)^T C^{-1} (\mu_A - \mu_B)} \tag{26}$$

where: $\mu_A = \{\mu_1^A, \mu_2^A, \ldots, \mu_n^A\}$ and $\mu_B = \{\mu_1^B, \mu_2^B, \ldots, \mu_n^B\}$ represents the means of the vectors from set A and B respectively. C is the joint covariance matrix:

$$C = \frac{n_A C_A + n_B C_B}{n_A + n_B - 2} \tag{27}$$

where C_A, C_B the covariance matrixes of the distributions of the vectors belonging to set A and B respectively.

REFERENCES

[1] J. Mak and J. Wolpaw, "Clinical applications of brain-computer interfaces: Current state and future prospects," *IEEE Rev. Biomed. Eng.*, vol. 2, pp. 187–199, 2009.

[2] G. Schalk, "Sensor modalities for brain-computer interfacing," *Human-Computer Interaction. Novel Interaction Methods and Techniques*, vol. 5611, pp. 616–622, 2009.

[3] S. Mason, A. Bashashati, M. Fatourechi, K. Navarro, and G. Birch, "A comprehensive survey of brain interface technology designs," *Ann Biomed Eng*, vol. 35, pp. 137–169, 2007.

[4] N. Birbaumer, N. Ghanayim, T. Hinterberger, I. Iversen, B. Kotchoubey, A. Kübler, J. Perelmouter, E. Taub, and H. Flor, "A spelling device for the paralysed," *Nature*, vol. 398, pp. 297–298, 1999.

[5] F. Galán, M. Nuttin, E. Lew, P. Ferrez, G. Vanacker, J. Philips, and J. del R. Millán, "A brain-actuated wheelchair: Asynchronous and non-invasive braincomputer interfaces for continuous control of robots," *Clin. Neurophysiol.*, vol. 119, no. 9, pp. 2159–2169, 2008.

[6] S. M. Grigorescu, T. Lüth, C. Fragkopoulos, M. Cyriacks, and A. Gräser, "A BCI-controlled robotic assistant for quadriplegic people in domestic and professional life," *Robotica*, vol. 1, pp. 1–13, 2011.

[7] G. Müller-Putz and G. Pfurtscheller, "Control of an electrical prosthesis with an SSVEP-based BCI," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 1, pp. 361–364, 2008.

[8] M. Tangermann, M. Krauledat, K. Grzeska, M. Sagebaum, B. Blankertz, C. Vidaurre, and K.-R. Müller, "Playing pinball with non-invasive BCI," in *Advances in Neural Information Processing Systems*, vol. 21, 2009, pp. 1641–1648.

[9] R. Krepki, B. Blankertz, G. Curio, and K.-R. Müller, "The Berlin Brain-Computer Interface (BBCI) towards a new communication channel for online control in gaming applications," *Multimed Tools Appl*, vol. 33, pp. 73–90, 2007.

[10] A. A. Karim, T. Hinterberger, J. Richter, J. Mellinger, N. Neumann, H. Flor, A. Kübler, and N. Birbaumer, "Neural internet: Web surfing with brain potentials for the completely paralyzed," *Neurorehab Neural Repair*, vol. 20, pp. 508–515, 2006.

[11] N. Birbaumer and P. Sauseng, "Brain-computer interface in neurorehabilitation," *Brain-Computer Interfaces*, pp. 155–169, 2010.

[12] C. Neuper and G. Pfurtscheller, "Neurofeedback Training for BCI Control," *Brain-Computer Interfaces*, pp. 65–78, 2010.

[13] G. Pfurtscheller and F. H. Lopes da Silva, "Event-related EEG/MEG synchronization and desynchronization: basic principles," *Clin. Neurophysiol.*, vol. 110, pp. 1842–1857, 1999.

[14] G. Pfurtscheller and C. Neuper, "Motor imagery and direct brain-computer communication," *Proc. IEEE*, vol. 89, pp. 1123–1134, 2001.

[15] R. Leeb, D. Friedman, G. R. Müller-Putz, R. Scherer, M. Slater, and G. Pfurtscheller, "Self-paced (asynchronous) BCI control of a wheelchair in virtual environments: A case study with a tetraplegic," *Comput. Intell. Neurosci.*, vol. 2007, pp. 1–8, 2007 (ID 79642).

[16] G. Townsend, B. Graimann, and G. Pfurtscheller, "Continuous EEG classification during motor imagery-simulation of an asynchronous BCI," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 12, pp. 258–265, 2004.

[17] A. Schloegl, C. Neuper, and G. Pfurtscheller, "Subject specific EEG patterns during motor imaginary," in *Proc. 19th Ann. Int. Conf. IEEE/EMBS*, 1997, pp. 1530–1532.

[18] G. Schalk, D. McFarland, T. Hinterberger, N. Birbaumer, and J. Wolpaw, "BCI2000: A general-purpose brain-computer interface (BCI) system," *IEEE Trans. Biomed. Eng.*, vol. 51, pp. 1034–1043, 2004.

[19] P. McCullagh and J. A. Nelder, *Generalized Linear Models*, 2nd ed. Chapman & Hall, 1989.

[20] C. Jutten and J. Herault, "Blind separation of sources, part 1: An adaptive algorithm based on neuromimetic architecture," *Signal Processing*, vol. 24, pp. 1–10, 1991.

[21] A. Ziehe, P. Laskov, G. Nolte, and K.-R. Müller, "A fast algorithm for joint diagonalization with non-orthogonal transformations and its application to blind source separation," *J. Mach. Learn. Res.*, vol. 5, pp. 777–800, 2004.

[22] D.-T. Pham and J.-F. Cardoso, "Blind separation of instantaneous mixtures of nonstationary sources," *IEEE Trans. Signal Process.*, vol. 49, pp. 1837–1848, 2001.

[23] A. Schloegl, K. Lugger, and G. Pfurtscheller, "Using adaptive autoregressive parameters for a brain-computer-interface experiment," in *Proc. 19th Ann. Int. Conf. IEEE/EMBS*, 1997, pp. 1533–1535.

[24] G. Schwarz, "Estimating the dimension of a model," *Ann. Statistics*, vol. 6, pp. 461–464, 1978.

[25] H. Zhang, S. R. Liyanage, C. Wang, and C. Guan, "Learning from feedback training data at a self-paced brain-computer interface," *J. Neural Eng.*, vol. 8, 2011 (ID 046035).

[26] M. R. Nuwer, "Recording Electrode Site Nomenclature," *J. Clin. Neurophysiol.*, vol. 4, pp. 121–133, 1987.

[27] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, 1948.

[28] B. Z. Allison, E. W. Wolpaw, and J. R. Wolpaw, "Brain-computer interface systems: Progress and prospects," *Expert Rev. Med. Devices*, vol. 4, pp. 463–474, 2007.

[29] M. Fatourechi, A. Bashashati, G. Birch, and R. Ward, "Automatic user customization for improving the performance of a self-paced brain interface system," *Med. Bio. Eng. Comput.*, vol. 44, pp. 1093–1104, 2006.

[30] S. Fazli, M. Danóczy, F. Popescu, B. Blankertz, and K.-R. Müller, "Using rest class and control paradigms for brain computer interfacing," in *Proc. 10th Int. Work-Conf. Artificial Neural Networks*, 2009, pp. 651–665.

[31] C. Gouy-Pailler, M. Congedo, C. Brunner, C. Jutten, and G. Pfurtscheller, "Nonstationary brain source separation for multiclass motor imagery," *IEEE Trans. Biomed. Eng.*, vol. 57, pp. 469–478, 2010.

[32] M. Grosse-Wentrup and M. Buss, "Multiclass common spatial patterns and information theoretic feature extraction," *IEEE Trans. Biomed. Eng.*, vol. 55, pp. 1991–2000, 2008.

[33] G. Dornhege, B. Blankertz, G. Curio, and K.-R. Muller, "Boosting bit rates in noninvasive EEG single-trial classifications by feature combination and multiclass paradigms," *IEEE Trans. Biomed. Eng.*, vol. 51, pp. 993–1002, 2004.

[34] C. Brunner, M. Naeem, R. Leeb, B. Graimann, and G. Pfurtscheller, "Spatial filtering and selection of optimized components in four class motor imagery EEG data using independent components analysis," *Pattern Recognit. Lett.*, vol. 28, pp. 957–964, 2007.

[35] R. Scherer, F. Lee, A. Schlogl, R. Leeb, H. Bischof, and G. Pfurtscheller, "Toward self-paced brain-computer communication: navigation through virtual worlds," *IEEE Trans. Biomed. Eng.*, vol. 55, pp. 675–682, 2008.

[36] J. Millán, F. Renkens, J. Mouriño, and W. Gerstner, "Brain-actuated interaction," *Artif. Intell.*, vol. 159, pp. 241–259, 2004.

[37] G. Pfurtscheller and T. Solis-Escalante, "Could the beta rebound in the EEG be suitable to realize a "brain switch"?" *Clin. Neurophysiol.*, vol. 120, pp. 24–29, 2009.

[38] G. Müller-Putz, V. Kaiser, T. Solis-Escalante, and G. Pfurtscheller, "Fast set-up asynchronous brain-switch based on detection of foot motor imagery in 1-channel EEG," *Med. Bio. Eng. Comput.*, vol. 48, pp. 229–233, 2010.

[39] P. J. Franaszczuk and K. J. Blinowska, "Linear model of brain electrical activity–EEG as a superposition of damped oscillatory modes," *Biol. Cybern.*, vol. 53, pp. 19–25, 1985.

[40] D. J. Thomson, "Spectrum estimation and harmonic analysis," *Proc. IEEE*, vol. 70, pp. 1055–1096, 1982.

[41] R. Fano, *Transmission of Information: Statistical Theory of Communications.* Wiley, 1961.

[42] G. Brown, "An information theoretic perspective on multiple classifier systems," in *Proc. 8th Intl. Work. Multiple Classifier Systems*, 2009, pp. 344–353.

[43] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, pp. 1226–1238, 2005.

[44] P. Peduzzi, J. Concato, E. Kemper, T. R. Holford, and A. R. Feinstein, "A simulation study of the number of events per variable in logistic regression analysis," *J. Clin. Epidemiol.*, vol. 49, pp. 1373–1379, 1996.

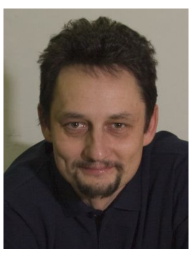
[45] P. C. Mahalanobis, "On the generalised distance in statistics," in *Proc. Nat. Inst. Science*, 1936, pp. 49–55.

Rafał Kuś received the M.Sc. degree in medical physics, in 2002 and the Ph.D. degree in physics, in 2007, from the University of Warsaw, Warsaw, Poland. He is currently an Assistant Professor at the Biomedical Physics Division of the University of Warsaw.

His main research interests concern investigation of biophysics, bioelectrical signals and application of advanced signal analysis methods to brain-computer interfaces.

Diana Valbuena received the Dipl.Ing. degree in electrical engineering from the University of Ibague, Ibague, Colombia, in 2005 and the Ph.D. degree in electrical engineering from the University of Bremen, Bremen, Germany, in 2011.

Her research focus is the improvement of brain–computer interfaces through neurofeedback training. Her research interests include digital signal processing, software design and human–computer interaction.



Jarosław Żygierewicz received the M.Sc. (1995) and Ph.D. (2000) and Ph.D. D.Sc (2012) degrees in Biomedical Physics from the University of Warsaw, Warsaw, Poland. He is currently an Assistant Professor at the Department of Biomedical Physics, Faculty of Physics of the University of Warsaw.

His research interests concern time-frequency analysis of EEG and MEG signals, especially event related power changes of those signals, and realistic neuronal network models that provide insight into the mechanisms underlying the effects observed in EEG and MEG signals.

Tatsiana Malechka received the M.Sc. degree in medical engineering from the University of Applied Sciences, Gelsenkirchen, Germany, in 2009. She is currently working toward the Ph.D. degree at the Institute of Automation of the University of Bremen, Bremen, Germany.

Her research interests include digital and statistical signal processing with applications to brain-computer interfaces, and software design.

Axel Gräser received the diploma in electrical engineering from the University of Karlsruhe, Karlsruhe, Germany, in 1976 and the Ph.D. degree in control theory from the Technical University of Darmstadt, Darmstadt, Germany, in 1982.

From 1982 to 1990, he was the Head of the Control and Software Department of Lippke GmbH, Germany. From 1990 to 1994, he was Professor of Control Systems, Process Automation and Real-time Systems at the University of Applied Sciences, Koblenz, Germany. Since 1994, he has been the Head of the Institute of Automation, University of Bremen, Bremen, Germany. He was the Coordinator of the European Union projects BRAINROBOT and BRAIN. His research interests include service robotics, brain–computer interfaces, visual servoing, digital image processing, and augmented reality.

Piotr Durka received his Ph.D. degree and habilitation in physics from the University of Warsaw, Warsaw, Poland. He is currently the Head of the Division of Biomedical Physics, world's first Neuroinformatics BSc curriculum, and several projects, including http://OpenBCI.pl. He is also President of BrainTech Ltd.

His views on the mission of a physicist are coherent with Richard Feynman's speech on Cargo Cult Science and Jon Clearbout's notion of Reproducible Research. Author of 5 books and dozens scientific papers, details at http://durka.info.