Modified Spatio-Temporal Matched Filtering for Brain Responses Classification

Marian P. Kotas, Michal Piela, and Sonia H. Contreras-Ortiz

***Abstract*—In this article, we apply the method of spatio-temporal filtering (STF) to electroencephalographic (EEG) data processing for brain responses classification. The method operates similarly to linear discriminant analysis (LDA) but contrary to most applied classifiers, it uses the whole recorded EEG signal as a source of information instead of the precisely selected brain responses, only. This way it avoids the limitations of LDA and improves the classification accuracy. We emphasize the significance of the STF learning phase. To preclude the negative influence of super–Gaussian artifacts on accomplishment of this phase, we apply the discrete cosine transform (DCT) based method for their rejection. Later, we estimate the noise covariance matrix using all data available, and we improve the STF template construction. The further modifications are related with the constructed filters operation and consist in the changes of the STF interpretation rules. Consequently, a new tool for evoked potentials (EPs) classification has been developed. Applied to the analysis of signals stored in a publicly available database, prepared for the assessment of modern algorithms aimed in EPs detection (in the frames of the 2019 IFMBE Scientific Challenge), it allowed to achieve the second best result, very close to the best one, and significantly better than the ones achieved by other contestants of the challenge.**

***Index Terms*—Brain–computer interfaces (BCI), discrete cosine transform (DCT), generalized matched filtering (GMF), spatio–temporal filtering (STF), visual evoked potentials (EPs).**

I. Introduction

CAPABILITY for quick and reliable classification of brain responses plays a key role in designing modern and efficient brain-computer interfaces (BCI). Responses of the brain in BCI are triggered by external (visual, auditory) or internal (mental simulation of a physical action) stimuli, therefore, are called as evoked potentials (EP). Among a group of brain's responses to a visual stimulus, P300 is characterized by the greatest amplitude, therefore, considered as a suitable measure of the brain activity during visual tasks [1]–[8]. The idea to employ this potential as a control signal in BCI was first introduced in 1988 by Farwell and Donchin [1]. Since then, the major challenge remains the same and involves very low energy of the desired EP in reference to the spontaneous activity of the brain. Consequently, its proper classification requires multiple presentation of the same stimulus and averaging of the consecutive responses of the brain. This in turn prolongs the time of an experiment and thereby a period of high concentration a user needs to maintain. It is inconvenient, particularly for people whose medical condition affects their ability to stay focused long on the mental or visual task. Therefore, an effort is made to shorten this time by reducing a number of averaging steps necessary to recognize an EP.

Development of new, more effective classifiers is one of the ways to achieve this objective. The aim of the classifiers is to discriminate between two types of brain responses: target which are triggered by the stimuli, the user is focused on, and nontarget, i.e., caused by other stimuli, the user is trying to neglect. Popular algorithms that assure good performance are based on support vector machines [9], [10], ensemble of weighted support vector machines [11], [12], linear discriminant analysis (LDA) [13], stepwise LDA [14], shrinkage LDA [15], Fisher's and Bayesian LDA (FLDA, BLDA) [16], [17], sparse Bayesian LDA [18], and regularized group sparse LDA [19].

While the above listed methods perform an outright classification of the brain responses stored in signal matrices, there are also methods, which precede the classification by prior signal enhancement and feature extraction using spatial (the xDAWN algorithm [20]) or spatial and temporal projections [21]. In [22], a two-step projection using spatial LDA and temporal canonical correlation analysis was proposed to form feature vectors for later classification; a further extension of this approach was described in [23].

Although the linear methods appeared to be advantageous in application to EP detection [24], there are attempts to utilize also the nonlinear ones, like e.g., convolutional neural networks [25]–[27], self-organizing fuzzy neural networks [28], or extreme learning machines [29], [30]. Particularly fast development can be observed in applications of convolutional neural networks. In [25], the first convolutional layer performs spatial filtering and the second, the temporal one; later, the fully connected multilayer perceptron is employed for final classification. The authors achieved good results using an ensemble of such classifiers (called as a multiclassifier system). In [31], the order of spatial and temporal filtering was reversed. This architecture

Manuscript received July 9, 2021; revised January 29, 2022 and March 18, 2022; accepted March 31, 2022. Date of publication May 5, 2022; date of current version July 14, 2022. This work was supported in part by statutory funds of the Department of Cybernetics, Nanotechnology and Data Processing, Silesian University of Technology, BK-2022 and co-financed by European Union through the European Social Fund under Grant POWR.03.02.00-00-I029. This work was performed using the infrastructure supported by POIG.02.03.01-24-099/13 Grant: GeCONiI-Upper Silesian Center for Computational Science and Engineering. This article was recommended by Associate Editor T. H Falk. *(Corresponding author: Marian P. Kotas.)*

Marian P. Kotas and Michal Piela are with the Department of Cybernetics, Nanotechnology and Data Processing, Silesian University of Technology, 44-101 Gliwice, Poland (e-mail: mkotas@polsl.pl; michal.piela@polsl.pl).

Sonia H. Contreras-Ortiz is with the Universidad Tecnológica de Bolívar, Biomedical Engineering Program, 150003 Cartagena, Colombia (e-mail: scontreras@utb.edu.co).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/THMS.2022.3168421.

Digital Object Identifier 10.1109/THMS.2022.3168421

was also adopted in [32]. Similarly, in [33] and [34], where the convolutional neural network was preceded by additional preprocessing steps to improve the classification results.

Spatio-temporal filtering (STF) is an intrinsically linear method, applied, e.g.: to detection of low amplitude signals of repeatable morphology embedded in high energy noise [35], [36] or to motor imagery (MI) tasks classification [37]–[40], or to related operations like selection of the proper EEG channels [41], or extraction of movement related potentials [42] for later MI classification. The alternating spatial–temporal optimization and projection was applied to event related potentials classification in [43].

We have applied STF to tackle the fundamental problem of LDA based approaches to EP classification, i.e., the singularity of the within-class covariance matrix which has to be inverted [16]. This singularity results usually from a large dimension of a feature vector (it is equal to the product of a number of time samples covering a brain response and a number of channels recorded) and a limited number of the brain responses used for the classifiers learning. Since the within-class dispersion results mostly from the presence of noise in the feature vectors, this problem can be seen as an issue of the noise properties estimation. Accomplishment of this operation on the basis of target and nontarget responses, only, seems to be unjustified. We propose to reject this restriction and perform the noise covariance matrix estimation using different time segments of EEG signals, not necessarily those corresponding to the precisely selected brain responses. Moreover, taking into account that STF is most effective in the colored Gaussian noise (CGN) [44], we propose to reject super–Gaussian artifacts for better estimation of the CGN properties. This improves the EP classifiers learning without the necessity of inversion regularization [16], or feature vectors dimension reduction [20], [21].

A further improvement can be achieved if we face the fact that brain responses can have different time delays with respect to the stimuli (flashes) and it is not always justified to apply the classifiers to the same time segments after the stimuli.

The rest of this article is organized as follows. Section II explains the operation of a BCI system. Section III provides detailed description of the original STF method; its modifications are presented in Section IV. Results of experiments and their discussion are presented in Section V. Finally, Section VI concludes this article.

II. Evoked Potentials (EPs) Based BCI

During EPs based BCI experiments, a subject faces a screen filled with different objects (letters, signs or pictures, etc.) and is focused on the selected one. Fig. 1 shows an example where the user interface contains six household appliances, and the user focuses on the lamp. The six objects are flashed one at a time in a random order. The flash of the item the user is focused on (lamp) is considered as the target event; flashes of the other items are the nontarget ones. We expect the target to elicit higher responses of the brain than the nontargets. Once all the objects have been flashed (one target and five nontargets), we can compare the elicited brain responses. It is

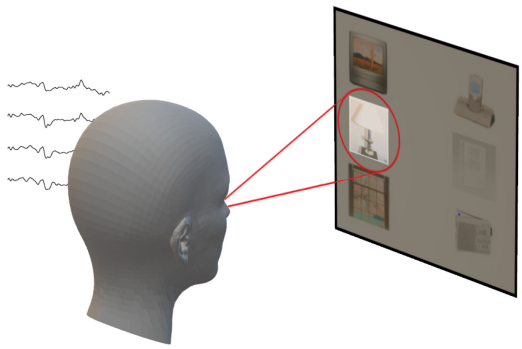

Fig. 1. Illustration of BCI experiment; a moment when visual stimulus (target) is captured.

accomplished by comparing the outputs of the classifier applied to these six events. The one with the highest score indicates the picture classified as a target. However, because of a very low amplitude of the brain responses to the visual stimuli, it is very rare to achieve proper discrimination at this step of the experiment.

Therefore, this part of the experiment, so called run (of six flashes), when each visual stimulus occurs only once, is insufficient. Multiple repetition of such runs of flashes is required. In each consecutive run, the order of pictures flashing is random, independent from the other runs. This way of presentation involves an element of surprise, necessary to elicit an EP. For every individual picture, the outputs obtained in the consecutive runs are averaged. Typically, taking into account more subsequent runs, the averaged response to a target stimulus is increased when compared to the averaged responses to the nontarget ones. Consequently, increasing the number of runs analyzed should increase the classification accuracy.

The described series of runs, related with one particular target object, is called as a block. During the experiments, this process (block) is usually repeated for every object (on the screen) regarded as a target. All blocks executed form a session. In the considered example, a session contains six blocks.

In the experimental section, we will analyze the dataset BCIAUT-P300, recorded during similar experiments. It was created for a purpose of the 2019 IFMBE Scientific Challenge [32]. The dataset contains recordings of 15 subjects diagnosed with autism spectrum disorder. The structure of the data is as follows: there are seven learning and seven testing sessions recorded by every subject. Each learning session contains 20 blocks. During each of the blocks, the user is focused on one picture (the one selected as a target, out of eight available on the user interface). Every block is built of 10 runs, each of which gathers single brain responses to all possible events (one target and seven nontargets). The testing subset is organized in a similar manner; however, the number of runs per block varies between 4 and 10. The EEG signals recorded are sampled with the frequency of 250 Hz, notch-filtered at 50 Hz, and bandpass-filtered between 2 and 30 Hz. Signals were acquired using eight-channel configuration (electrodes were positioned at C3, Cz, C4, CPz, P3, Pz, P4, POz).

III. Spatio-Temporal Filtering

Lets denote the m channel signal vector as $\mathbf{x}(k) = [x_1(k)x_2(k), \ldots, x_m(k)]^T$. To exploit the spatial and temporal relationships between the signal components, we have defined, similarly as in [35], the following vector representation of the multichannel signals:

$$\mathbf{x}^{(k)} = \begin{bmatrix} x_1(k - J\cdot\tau) \\ x_1(k-(J-1)\tau) \\ \cdots \\ x_1(k+(J-1)\tau) \\ x_1(k+J\cdot\tau) \\ \cdots \\ x_m(k-J\cdot\tau) \\ x_m(k-(J-1)\tau) \\ \cdots \\ x_m(k+(J-1)\tau) \\ x_m(k+J\cdot\tau) \end{bmatrix} \tag{1}$$

containing $2J+1$ time samples from m channels available (the $\mathbf{x}^{(k)}$ vector length is $p = (2J+1)m$).

Parameter τ is introduced to perform a kind of signal resampling (from the original frequency f_s to f_s/τ, applied in spatio-temporal vectors). We assume that this new frequency should not be smaller than 40 Hz, thus parameter τ can be calculated according to: $\tau = \lfloor \frac{f_s[Hz]}{40} \rfloor$, where $\lfloor\cdot\rfloor$ denotes the largest integer not greater than the argument.

The filtering operation can be expressed as

$$y(k) = \mathbf{h}^T\mathbf{x}^{(k)} \tag{2}$$

and if we want to perform detection of some signal pattern of a finite length T_p, the time interval $2\cdot J\cdot\tau$ should not be shorter. Thus, parameter J can be calculated according to: $J = \left\lceil \frac{f_s\cdot T_p}{2\cdot\tau} \right\rceil$, where $\lceil\cdot\rceil$ denotes the smallest integer not smaller than the argument.

A. STF Construction

Let's denote the moments when the target object is flashed as $k_i, i = 1, 2, \ldots, I_t$ (I_t denotes the total number of these flashes). We assume that the brain response to the ith flash begins at k_i and ends at $k_i + 2\cdot J\cdot\tau$. This time segment corresponds to spatio-temporal vector $\mathbf{x}^{(k_i+J\cdot\tau)}$. For this vector, the filter should produce the maximal output value. Since the slope of the output signal is limited, more of its neighboring time samples will be magnified ($\cdots, y(k_i+J\cdot\tau-1), y(k_i+J\cdot\tau), y(k_i+J\cdot\tau+1), \cdots$). Thus, to construct the filter we first specify the following set indicating the spatio–temporal vectors that should be maximized

$$\Psi_t(\delta) = \{k|\ \ |k-(k_i+J\cdot\tau)| \le \delta\cdot f_s,\ i = 1, 2, \ldots, I_t\} \tag{3}$$

where δ denotes the assumed time range.

We want the filter to be very sensitive to the corresponding set of vectors ($\mathbf{x}^{(k)}, k \in \Psi_t(\delta)$). On the other hand, we want it to produce as low output as possible in other moments of time. Therefore, we create an auxiliary set $\Psi_t(\Delta)$ with argument $\Delta > \delta$. On this basis, we create the set corresponding to spatio–temporal vectors that should be suppressed

$$\Psi_{\text{supp}}(\Delta) = \Psi - \Psi_t(\Delta) \tag{4}$$

where Ψ denotes the set of all time indices for which vectors $\mathbf{x}^{(k)}$ have been constructed. Thus $\Psi_{\text{supp}}(\Delta)$ indicates all vectors $\mathbf{x}^{(k)}$ whose time location is rather distant from the vectors maximized.

Finally, we create the following objective function:

$$Q(\mathbf{h}) = \frac{\frac{1}{|\Psi_t(\delta)|}\sum_{k\in\Psi_t(\delta)}(\mathbf{h}^T\mathbf{x}^{(k)})^2}{\frac{1}{|\Psi_{\text{supp}}(\Delta)|}\sum_{k\in\Psi_{\text{supp}}(\Delta)}(\mathbf{h}^T\mathbf{x}^{(k)})^2} \tag{5}$$

where $|\cdot|$ denotes the set cardinality.

We gather the spatio–temporal vectors that are maximized in matrix $\mathbf{A}_{p\times n_a} = [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_{n_a}]$, and those which we want to suppress in matrix $\mathbf{B}_{p\times n_b} = [\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_{n_b}]$, where $n_a = |\Psi_t(\delta)|$ and $n_b = |\Psi_{\text{supp}}(\Delta)|$. This way we obtain the following function:

$$Q(\mathbf{h}) = \frac{\frac{1}{n_a}\|\mathbf{h}^T\mathbf{A}\|^2}{\frac{1}{n_b}\|\mathbf{h}^T\mathbf{B}\|^2} = \frac{\mathbf{h}^T\mathbf{C}_a\mathbf{h}}{\mathbf{h}^T\mathbf{C}_b\mathbf{h}} \tag{6}$$

where $\mathbf{C}_a = \frac{1}{n_a}\mathbf{A}\mathbf{A}^T$ and $\mathbf{C}_b = \frac{1}{n_b}\mathbf{B}\mathbf{B}^T$.

Function $Q(\mathbf{h})$ has the form of a Rayleigh quotient whose maximization can be achieved using the generalized eigendecomposition

$$\mathbf{C}_a\mathbf{h} = \lambda\mathbf{C}_b\mathbf{h}. \tag{7}$$

The solution, we are interested in, is the generalized eigenvector $\mathbf{e}_1$ that corresponds to the greatest eigenvalue. Here, we face the problem of the eigenvector sign ambiguity [45]. We solve it by using either $\mathbf{h} = \mathbf{e}_1$ or $\mathbf{h} = -\mathbf{e}_1$, so as the average response of the filter to the learning vectors stored in matrix $\mathbf{A}$ was positive.

This original version of the STF method [based on eigendecomposition (7)] will further simply be denoted as OSTF.

B. STF Interpretation

Lets denote the onset of the oth object ith flash as ${}^ok_i$. The STF filter response to this flash is expected at ${}^ok_i + J\cdot\tau$ (see the explanation in Section III-A). It is therefore sufficient to calculate and store for further analysis the values

$$q_{i,o} = y({}^ok_i + J\cdot\tau);\ \ o = 1, 2, \ldots, I_o;\ i = 1, 2, \ldots, I_r \tag{8}$$

where I_o denotes the number of objects on the screen, and I_r is the number of analyzed runs of flashes.

Thus, the oth column of the formed $\mathbf{Q}_{I_r\times I_o}$ matrix contains the scores obtained for successive flashes of the oth object. If we want to select the target object on the basis of one run of flashes only ($I_r = 1$), we simply choose o for which $q_{1,o}$ has the greatest value. However, we usually take into account more runs ($I_r > 1$), and we select the column for which the mean of the corresponding I_r values is highest.

The described interpretation rules share an important feature with most classifiers applied to EP discrimination. They select the same located segments of the EEG signal following the

stimuli (contained in vectors $\mathbf{x}^{(k_i+J\cdot\tau)}, i = 1, 2, \cdots$) to perform classification. Therefore, they will further be referred to as the classical decision rules (CDR).

C. Averaging of the Maximized Spatio–Temporal Vectors

In [36], we propose to calculate the average of the vectors stored in $\mathbf{A}$

$$\bar{\mathbf{a}} = \frac{1}{n_a}\sum_{i=1}^{n_a}\mathbf{a}_i. \tag{9}$$

Objective function (6) can now be replaced by

$$Q(\mathbf{h}) = \frac{\mathbf{h}^T(\bar{\mathbf{a}}\bar{\mathbf{a}}^T)\mathbf{h}}{\mathbf{h}^T\mathbf{C}_b\mathbf{h}} \tag{10}$$

and consequently, the eigendecomposition (7) based formula is substituted by the product

$$\mathbf{h} = \mathbf{C}_b^{-1}\bar{\mathbf{a}} \tag{11}$$

which is computationally identical to the one used for construction of the highly acknowledged generalized matched filters [44]. Therefore, this solution will further be referred to as the generalized spatio–temporal matched filtering (GSTMF). GSTMF and OSTF will be regarded as reference methods for the assessment of the modified spatio–temporal matched filter (MSTMF), which we propose in this article.

IV. Modified Spatio–Temporal Matched Filtering

Although the brain responses to the nontarget flashes are very tiny, in [16], it was shown that taking into account these responses during classifiers learning (and not only the target responses) improves discrimination between them. To make use of this finding, we construct set $\Psi_{nt}(\delta)$ using (3) with k_i replaced by $k_i', i = 1, 2, \ldots, I_{nt}$ (corresponding to the moments when the nontarget objects are flashed). On this basis, we form matrix $\mathbf{A}'$ that contains vectors beginning close to these nontarget flashes $\{\mathbf{x}^{(k+J\tau)}|\ k \in \Psi_{nt}(\delta)\}$. By averaging these vectors, we construct the average nontarget response $\bar{\mathbf{a}}'$.

Now, objective function $Q(\mathbf{h})$ takes the following form:

$$Q(\mathbf{h}) = \frac{\mathbf{h}^T(\bar{\mathbf{a}}-\bar{\mathbf{a}}')(\bar{\mathbf{a}}-\bar{\mathbf{a}}')^T\mathbf{h}}{\mathbf{h}^T\mathbf{C}_b\mathbf{h}} \tag{12}$$

and its solution is given by

$$\mathbf{h} = \mathbf{C}_b^{-1}(\bar{\mathbf{a}}-\bar{\mathbf{a}}'). \tag{13}$$

The EEG signals can contain outlying super–Gaussian artifacts, which can have a detrimental influence on the filter template construction. Because of this, we precede the learning phase of the proposed method with the procedure for outlying artifacts rejection.

A. Rejection of Outlying Artifacts (ROA)

The procedure consists of the operation of outlying artifacts detection and of the stage of the detected artifacts suppression.

1) Detection of Outlying Artifacts: The operation is performed in each channel separately. Its first step involves calculation of the 5th and 95th percentile of all values in the channel. Then, multiplied by d these percentiles are regarded as thresholds that limit the negative and positive values accepted, respectively. In the experimental section, we use $d = 3$ (during initial experiments, for $d \in [2, 5]$, we have observed similar performance of EP classification). In each place, where the thresholds are crossed, we search for the signal maximum (or minimum), and starting from this position, we search for the artifact onset and offset. The criterion is based on the signal derivative: in order to find the first and the last sample of the artifact, we search for the positions (to the left and to the right from the extremum, respectively), where the derivative polarity changes. All samples that belong to the found artifacts are marked as outlying; for signal $\mathbf{x}(k)$ recorded during a single block, we form matrix $\mathbf{Z}$ containing the markers, which classify the samples of $\mathbf{x}(k)$ as outlying (0) or proper (1)

$$z_{i,j} = \begin{cases} 0, & if \quad x_j(i) \ \texttt{is outlying} \\ 1, & if \quad x_j(i) \ \texttt{is proper} \end{cases}. \tag{14}$$

Simultaneously, the $x_j(i)$ values marked as outlying are considered as missing and are replaced with zeros.

2) Missing Values Interpolation: To this end, we apply the iterative interpolation procedure, developed in [47] and facilitated as MATLAB function *inpaintn* by Garcia. The procedure is based on the discrete cosine transform (DCT) based smoothing proposed in [48]. The values of signal $\mathbf{x}(k)$ are stored in matrix $\mathbf{X}$ (subsequent channels in subsequent columns of $\mathbf{X}$). Then, $\mathbf{Z}$ defined by (14) and $\mathbf{X}$ are submitted to the *inpaintn* procedure. On the basis of the output matrix $\hat{\mathbf{X}}$ ($\hat{y}_{\{k+1\}}$ defined in [47, eq. (20)], where $\{k+1\}$ denotes the number of iteration), we obtain the signal with outlying artifacts suppressed $\hat{\mathbf{x}}(k)$.

B. Modification of the Decision Rules

The CDR, described in Section III-B, are less effective if the brain responses to the flashes are of changing delays. To overcome this inconvenience, we propose to calculate the average MSTMF responses to the I_r flashes of each object

$$\bar{y}_o(k) = \mathbf{h}^T\sum_{i=1}^{i=I_r}\mathbf{x}^{({}^ok_i+J\cdot\tau+k)} = \sum_{i=1}^{i=I_r} y({}^ok_i + J\cdot\tau + k). \tag{15}$$

The maximum of this average response to the oth object is expected at $k = 0$. However, we search for its true position around this point

$$k_o = \text{argmax}_{|k|\le r_s}\bar{y}_o(k) \tag{16}$$

where variable r_s limiting the range of the search can be established using $r_s = T_s \cdot f_s$ (following the preliminary experiments, $T_s = 0.1$ s has been applied).

The value $\bar{y}_o(k_o)$ is taken as the classifier score for the oth object. The object that achieves the highest value is regarded as the target one. This approach to STF interpretation will be referred to as the modified decision rules (MDR).

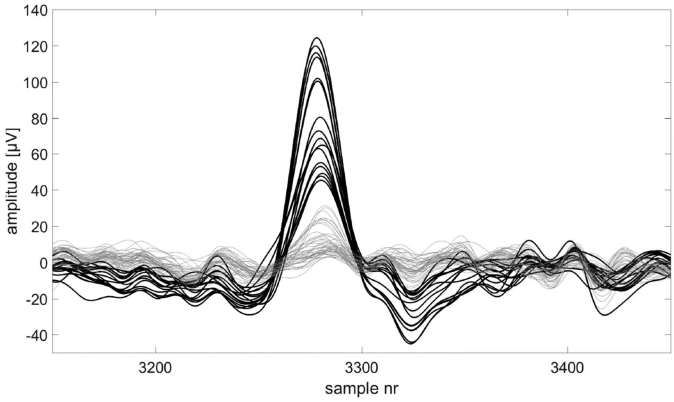

Fig. 2. Results of outlying artifacts suppression: original 64-channel EEG recording; in bold black: the channels with artifacts detected, in gray: the channels accepted.

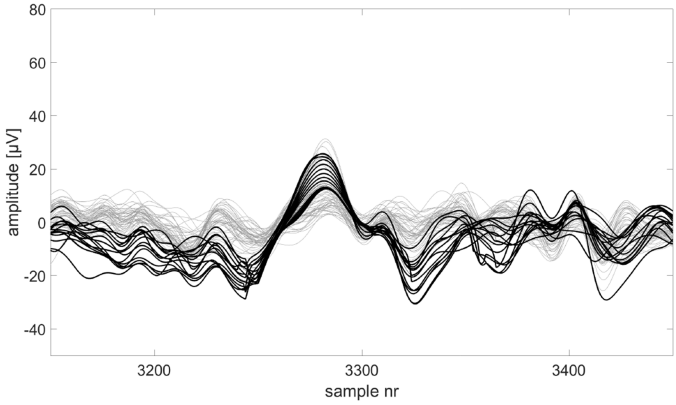

Fig. 3. Results of outlying artifacts suppression: the modified 64-channel EEG recording; in bold black: the modified channels, in gray: the unchanged channels. The amplitude scale has been changed with respect to Fig. 2.

C. MSTMF Responses as Feature Vectors

To exploit not only the height of the STF responses to the respective objects, but also their shape, we take the values: $\bar{y}_o(k_o+k), k=-r_s,-r_s+1,\ldots,r_s$, as the features of the oth object [r_s is the parameter used in (16)]. Those feature vectors can undergo classification for better discrimination between the target and nontarget brain responses. Since the amplitudes of the signals at the MSTMF output can vary substantially, we normalize them by dividing by their standard deviation (SD). This operation is accomplished during each individual classification task (for which the normalizing SD value is calculated on the basis of the signals recorded during all I_r flashes of the respective objects). In the experimental section, the SVM with linear kernel function will be applied for classification. The obtained combination of methods will simply be denoted as MSTMF+SVM (as opposed to MSTMF+CDR and MSTMF+MDR).

V. Results and Discussion

A. Signals Preprocessing

Both in the learning phase and during the STFs operation, the processed EEG signals undergo prior band-pass filtering [32]. ROA, which is rather time consuming, is applied during the learning phase, only. Figs. 2 and 3 show exemplary results of this operation. Fig. 2 contains an original 64-channel EEG recording with the noise artifacts; channels detected as noisy are marked in bold black. Exactly the same time frame of the signal is included in Fig. 3; here, the artifacts are significantly reduced (in bold black are the plots of the channels modified). In dataset BCIAUT-P300, used in this study, the signals are not stored in a continuous way (where all the channels and time samples are accessible). Instead, for each stimulus (a flash of a single object on the screen) a signal segment of 1.4 s length is written down: beginning 200 ms before the flash onset. Such length is insufficient for outlying artifacts rejection. To make this operation possible, all time segments corresponding to the flashes of 8 objects during 10 runs of flashes are first conflated. After this operation, we perform outlying artifacts rejection.

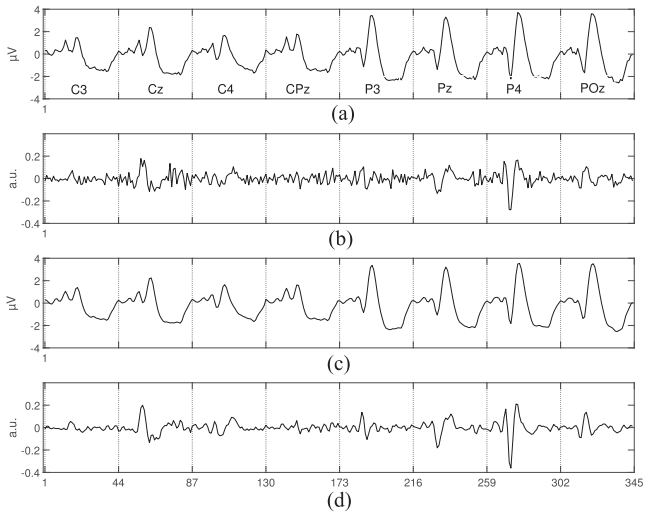

Fig. 4. LDA classifier (subplots A,B) versus MSTMF filter (C,D). Vectors are constructed according to (1); vertical dotted lines mark segments of $2J+1=43$ values, corresponding to individual EEG channels (whose names are given in subplot A). From the top: A) the difference between the centers of the target and nontarget classes, B) coefficients of the LDA classifier, C) the difference $\bar{\mathbf{a}}-\bar{\mathbf{a}}'$ (calculated to determine MSTMF) and D) the obtained MSTMF template; a.u. stands for arbitrary units. .

B. Construction of the Modified Spatio–Temporal Matched Filters

We have set parameters τ and J to 6 and 21, respectively, according to the formulas given below (1) and (2), so that the filter length ($2\cdot J\cdot\tau/f_s$) covered approximately 1 s period. Analysis of such a segment is oriented toward detection of the whole group of EPs, not only P300.

For each subject, we apply all learning data to calculate the mean target ($\bar{\mathbf{a}}$) and nontarget ($\bar{\mathbf{a}}'$) brain responses, and the covariance matrix $\mathbf{C}_b$. The exemplary difference $\bar{\mathbf{a}}-\bar{\mathbf{a}}'$ is presented in Fig. 4.C and the final MSTMF filter, in subplot D. For reference, we have plotted the results of using LDA (in A and B).

Subplots A and C are very similar, with the plot in C being slightly more smooth (as it is explained in Appendix A, for

TABLE I
ACCURACY (± STANDARD ERROR OF MEAN (SEM)) OF TARGET STIMULI DETECTION USING THE METHODS PROPOSED, THE BAYESIAN LDA, AND THE METHODS PARTICIPATING IN 2019 IFMBE SCIENTIFIC CHALLENGE (FOR THEIR PRECISE DESCRIPTION SEE [32])

Method	Accuracy[%]±SEM
OSTF+CDR $\delta = 44ms$	47.07± 8.01
GSTMF+CDR $\delta = 68ms$	82.90± 1.56
MSTMF+CDR $\delta = 60ms$	83.52± 1.41
MSTMF+MDR $\delta = 20ms$	90.50± 1.70
MSTMF+SVM $\delta = 20ms$	90.93± 1.67
BLDA	82.55± 1.26
CNN based on EEGNet	92.3± 1.8
CNN-BLSTM	84.0± 3.2
LDA, CWT, and PCA	82.0± 2.5
SVM	81.5± 2.6
400 LDA classifiers and FGDA	81.2± 2.1
VB-ARD	80.3± 2.2
BLDA, RUSBoost, and CNN	76.3± 2.9
MLP	70.0± 3.8
LDA	67.2± 3.3

$\delta > 0$ a kind of low-pass filtering is introduced). However, the final templates constructed are quite different; although we can see some similarity of the plots in B and D, the latter is of much better quality. It seems to be well matched to the local properties of the brain responses, whereas the former is embedded in wide band contaminations, and therefore can be expected to be more susceptible to noise. This substantial difference is primarily caused by more effective estimation of the noise covariance matrix $\mathbf{C}_b$ (if compared to the within class scatter matrix of the LDA classifier).

This can be explained in the following way: $\mathbf{C}_b$ is estimated on the basis of spatio–temporal vectors $\mathbf{x}^{(k)}$ indicated by set $\Psi_{\text{supp}}(\Delta)$. As it has been expressed below its definition (4), the set indicates vectors whose time location is rather distant from the vectors maximized (corresponding to target flashes). Thus, almost all vectors corresponding to nontarget flashes can be used for $\mathbf{C}_b$ estimation. As it has already been mentioned, the signals we use for experiments consist of 1.4 s long segments, corresponding to all individual flashes of the respective objects. With the sampling frequency $f_s = 250$ Hz and the assumed values of J and τ, applying (1) we construct 98 spatio-temporal vectors for each of such signal segments. All such vectors corresponding to nontarget flashes are used to estimate $\mathbf{C}_b$ and also a part of vectors corresponding to the target ones (their accepted number depends on the assumed value of parameter Δ). By contrast, only 1 vector corresponding to each individual flash (both target and nontarget) is used to estimate the within class scatter matrix in LDA.

C. Brain Responses Classification

We have applied all the methods described: OSTF, GSTMF, and MSTMF, and as a particular reference: the BLDA [16], i.e., the method that uses Bayesian regularization to overcome the LDA limitations (the software has been provided by the authors). The results are presented in the upper part of Table I.

For each of our methods, we had to select the values of parameters δ and Δ. It has, however, appeared that for the dataset used, Δ has hardly any influence on the classification results. Thus, a fixed value of $\Delta = 0.1$ s has been applied. For parameter δ, we performed a kind of cross validation. The filters were constructed on the basis of six learning sessions and applied to the seventh one. After rotation of this validation session and averaging of the results, an estimate of the accuracy was obtained for each value of δ. Following, we could select the one for which the best validation accuracy has been achieved. This value has been used in the final tests. BLDA has been applied using different decimation factors [16], and the best results achieved are presented. In the lower part of the table, for reference, the results of 2019 IFMBE Scientific Challenge are provided.

Watching the table, we can notice that using the generalized eigendecomposition (7) based OSTF with the CDR gives rather poor results. The large standard error of mean: SEM=8.01, shows that the method could have been fairly effective for some subjects and must have failed for some other ones. When formula (11) is applied in the learning phase, the filter (GSTMF) performs much better, competitively to BLDA and to most sophisticated methods competing in the challenge reported. The high accuracy of such a relatively simple method must be perceived as an outcome of the proper estimation of the noise covariance matrix $\mathbf{C}_b$: on the basis of the whole information available, contained in the recorded EEG signals. This has resulted in a fairly effective filter but the further modifications of a mean nontarget brain response ($\bar{\mathbf{a}}'$) subtraction, and DCT based outlying artifacts rejection, applied by MSTMF, have still raised its performance.

However, it is application of the MDR that has resulted in most significant improvement of the classification results. The reasons of such an outcome are illustrated in Fig. 5. Presented are the MSTMF responses to a block of nine runs of flashes (each run consisting of one target and seven nontarget flashes). Individual filter responses to the successive nine target flashes are presented in subplot A. Below in subplot B are the filter responses to nine flashes of a single nontarget object; we have selected the one (among seven) that has produced the highest average nontarget MSTMF response. Both in A and B, the plots seem to be dominated by high energy responses to the spontaneous EEG. A correct classification on the basis of such single responses does not seem possible. It is their averaging that is necessary to achieve the goal, and the average of plots from A is drawn with blue color in subplot C. We can observe a high amplitude peak, whose dominant component is the MSTMF response to the desired EP. What is surprising, however, is that the peak maximum is significantly shifted to the right from $k = 0$. Thus, application of CDR would cause the classification error: the height at $k = 0$ of the blue target response (marked by the red cross) is lower than the height of the corresponding nontarget one in D. Only the search for the maximum of the filter response leads to the correct classification because the height of the green point in C is greater than the corresponding one in D. Thus, in spite of the rather great height of the average target response, it is only application of MDR that allowed to avoid the classification error.

This observation is confirmed by the increased classification accuracy of MSTMF+MDR. It has risen up to above 90%. Using

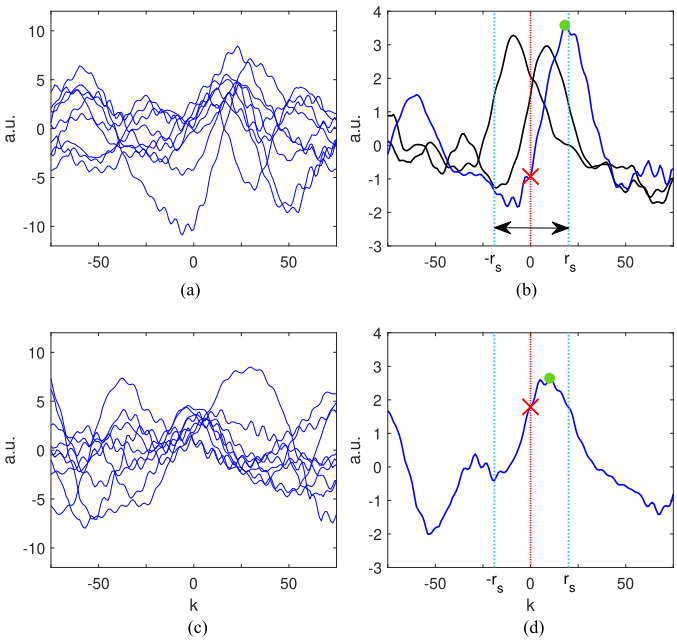

Fig. 5. Operation of the decision rules (MDR versus CDR): A) individual MSTMF responses to target flashes during a block of the successive $I_r = 9$ runs of flashes [variable k on the horizontal axis is introduced by (15)], B) the MSTMF responses to I_r flashes of a single nontarget object, C) the average target MSTMF responses (the one in blue is the mean of the plots in A; the black ones correspond to different blocks of flashes), D) the average nontarget MSTMF response (the mean of the plots in B). The double-sided arrow shows the region of the search for the maxima of the MSTMF responses. The red crosses mark the values of the scores obtained for $k = 0$ (according to CDR), and the green points, the ones obtained using MDR (the maxima in the specified range of the search). The additional comments are in the text.

TABLE II
STATISTICAL SIGNIFICANCE (S-SIGNIFICANT, N-NONSIGNIFICANT) OF THE DIFFERENCES BETWEEN THE METHODS TESTED: LETTER S MEANS THAT THE P-VALUE < 0.001

	OSTF+CDR	BLDA	GSTMF+CDR	MSTMF+CDR	MSTMF+MDR
BLDA	s	-	-	-	-
GSTMF+CDR	s	n	-	-	-
MSTMF+CDR	s	n	n	-	-
MSTMF+MDR	s	s	s	s	-
MSTMF+SVM	s	s	s	s	n

MSTMF responses as feature vectors and the SVM classifier has allowed to improve the results only slightly. Nevertheless, the increase of accuracy to nearly 91% is worth noticing.

To verify statistical significance of the differences between the methods tested, we have compared them pairwise, denoting as f_{12} the number of the cases classified correctly by the first, and incorrectly by the second method compared, and as f_{21} the number of opposite ones. On this basis, we have calculated the squared McNeymar statistic [49]

$$z = \frac{(f_{12} - f_{21})^2}{f_{12} + f_{21}} \tag{17}$$

which obeys the Chi-square distribution with one degree of freedom [50].

The results obtained are presented in Table II. For all methods tested, we have confirmed statistical significance of improvement with respect to the original OSTF method. This confirms

TABLE III
COMPUTATIONAL TIMES ($\pm$ STD)

Method	**learning time**$[s]\pm$STD	**testing time**[s]$\pm$STD
OSTF	39.28$\pm$ 0.26	7.19$\pm$ 0.85
GSTMF	23.26$\pm$ 0.12	7.19$\pm$ 0.83
MSTMF	99.03$\pm$ 3.66	7.17$\pm$ 0.85

inadequacy of using the eigendecomposition based approach for so noisy signals.

Comparing MSTMF (applied with the MDR or with the SVM classifier) with the reference method (BLDA), we can notice statistically significant improvement of the classification accuracy. This seems surprising because BLDA uses advanced Bayesian regularization for estimation of the classifier template. However, the need to apply regularization results from the abovementioned limitations of the LDA based approaches, and although the Bayesian regularization can effectively solve the problem of singularity of the scatter matrix to be inverted, it does not solve the problem of changing delays of the brain responses.

Thus, the most significant improvement of MSTMF accuracy is related with the modification of the decision rules. Averaging of the MSTMF responses to individual flashed objects and the search for maxima of such average responses, applied in MDR, have allowed to raise the classification accuracy to above 90%, and have made the MSTMF method not only significantly better than the reference BLDA, but even comparable to the winner of 2019 challenge, i.e., to the convolutional neural network based on the EEGnet software [31]. Moreover, compared to the second best competitor of the challenge, MSTMF appeared significantly more effective. This justifies using the modifications proposed, which are aimed to extract all the information from the whole EEG signal recorded.

The introduced modifications inevitably increase the computational costs of the methods. To illustrate this, in Table III, we have provided the times necessary to accomplish the learning stage and the operation stage of the selected versions of the STF method. Learning and testing times shown in Table III are the subject-averaged execution times. It should be noted that for each subject the total length of the learning signals is approximately equal to 37.3 min, and the length of the test signals is around 56 min. As far as the learning stage is realized, we can notice rather great differences between the respective methods. However, the operation stage is much much faster, and should not preclude any of the studied versions of the STF method from practical applications. For every method mentioned in Table III, the analysis of a 1.4 s long segment of the signal (related to one event) is executed in around 3 ms time.

One of the fundamental questions related with the development of real working BCI systems is: how many EEG channels are required to achieve satisfactory results. To give some insight into the matter, we have performed the following experiment. For MSTMF, we decreased the number of leads available (using the same cross validation as for selection of parameter δ) and calculated the classification accuracy achievable. Fig. 6.A presents the results obtained. It reveals that it is advantageous to measure at least six channels, since for five of them the accuracy discernibly drops down. The another question is: how dense

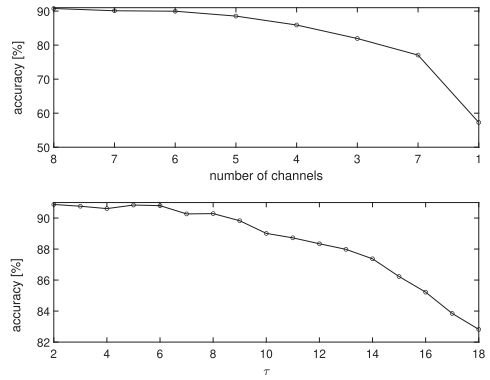

Fig. 6. Illustration of the factors that influence the classification accuracy of the MSTMF based BCI system: A) Accuracy as a function of the channels number, B) Accuracy as a function of τ.

should the brain responses be resampled, in the spatio–temporal vectors constructed, to achieve good classification results. This is related with the value of parameter τ. In Fig. 6.B, we can notice that for τ growing over the value of 10, the classification accuracy decreases, discernibly. The applied value of $\tau = 6$, calculated according to the formula given below (1), seems to be a good compromise between the computational costs of the method and its achievable classification accuracy.

The proper estimation of the noise covariance matrix is highly important. It not only allows to construct the filter with visually good properties, like the one presented in Fig. 4, but also assures effective suppression of stationary CGN of even extremely high energy. The problem can arise if the noise in test signals changes its properties with respect to this from the learning data. Such situation must be taken into account because we cannot assume the noise stationarity during different sessions, and even in long time intervals. A kind of remedy could be found if we had rather large learning database. This would allow us to construct many filters matched to different local noise properties. Later the weighted sum of the obtained ensemble of filters could be effective for the noise present in test signals. Development of such an ensemble of filters seems to be a promising direction for the future study.

However, we expect that it can be even more promising to apply the MDR in the already developed EP classifiers. Taking into account possible changes of the brain responses time delays (with respect to the stimuli) and looking for their correct localization, corresponding to the locally highest scores of the classifiers, should help to raise the accuracy of many existing algorithms.

VI. Conclusion

In this study, we have proposed application of spatio–temporal filtering to classification of visual EPs elicited by target and nontarget stimuli. We have shown that using the whole recorded EEG signal as a source of information instead of the same located brain responses to those stimuli, only, significantly improves the classification accuracy.

There are two major factors that contribute to this outcome. The first one consists in the appropriate construction of the STF, and the second, in the proper interpretation of the filtering results. To meet the first requirement, we have rejected the limitation of using only target and nontarget brain responses to estimate the noise covariance matrix. Instead, we have used all the data satisfying the necessary conditions, solving in this simple way, the problem of singularity of the matrix to be inverted. Moreover, we have proposed outlying artifacts rejection before the filter construction, what has still improved its performance. Applied with the CDR, the developed MSTMF filter has allowed to obtain relatively high accuracy of EP classification.

To meet the second requirement, we have addressed the problem of changing delays of the brain responses with respect to the visual stimuli. A delayed brain response at the MSTMF input inevitably results in a delayed filter response at its output. Therefore, it appears beneficial to look for the maxima of these responses. This search is preceded by averaging of the MSTMF responses to the flashes of the respective objects. Since averaging of the target responses causes their enhancement, it helps to find the correct maxima. These operations referred to as the MDR have significantly increased the EP classification accuracy.

With the very limited costs of MSTMF operation, we believe that our method can contribute to construction of more reliable and faster brain–computer interfaces.

Appendix A On the Noise Influence on STFs Construction

In the following, we will consider the filters defined by (11), related directly to the acknowledged generalized matched filtering (GMF). The GMF allows for detection of a known signal embedded in a CGN [44].

In the problem studied, the noise component is a complicated mixture produced mostly by the ongoing brain activity but also by the noncortical biologic sources (e.g., eye blinks and movements, or muscle and heart contractions) and the environmental sources (e.g., powerline, radio and electrical interference) [51]. Some of these components, like e.g., sinusoidal powerline interference are sub-Gaussian (with flatter distribution than Gaussian [52]) and some are super-Gaussian (sharper, with so called "heavy" tails, associated with relatively high probability of large amplitude peaks). This non-Gaussianity allows their separation using independent component analysis [52].

However, the recorded and processed signals are the mixtures of such components, and owing to the central limit theorem, we can make a simplifying assumption of their Gaussianity, unless one of these components has a predominant amplitude. Because of a strong correlation between the EEG channels and the successive signal samples, we can assume the noise to be colored.

The constructed average pattern $\bar{\mathbf{a}}$ can be expressed as

$$\bar{\mathbf{a}} = \frac{1}{n_a}\sum_{i=1}^{n_a}(\mathbf{s}+\mathbf{w}_i) = \mathbf{s}+\bar{\mathbf{w}} \tag{18}$$

where $\mathbf{s}$ is the fixed desired signal and $\mathbf{w}_i$, the additive noise. To achieve noise suppression, the noise vectors $\mathbf{w}_i$ should be [53]

1) of zero mean ($E\{w_{ki}\} = 0$, where w_{ki} is the kth entry of vector $\mathbf{w}_i$, E denotes expectation).

2) uncorrelated with the desired signal ($E\{s_n \cdot w_{ki}\} = 0$).
3) uncorrelated among themselves ($E\{w_{ki} \cdot w_{nj}\} = 0$, for $i \neq j$.

It should be noted that for Gaussian signals uncorrelatedness is equivalent to independence, with the latter being a stronger condition.

Since initially all signals are band-pass filtered, the zero mean assumption is naturally satisfied. Moreover, since the noise component is not related with the moments, the EPs are triggered, we can assume the noise vectors to be independent from the desired signal. When parameter δ of set Ψ_t [defined by (3)] equals 0, matrix $\mathbf{A}$ contains a single spatio–temporal vector for each target flash. Thus, for a relatively high distances between the flashes, we can also assume the noise vectors ($\mathbf{w}_i$) to be independent from each other (and consequently uncorrelated).

Following, for $\delta = 0$, we can expect effective suppression of noise as a result of averaging. However, for $\delta > 0$ matrix $\mathbf{A}$ contains many successive signal vectors related with the same target flash: $\{\ldots, \mathbf{x}^{(k+J\cdot\tau-1)}, \mathbf{x}^{(k+J\cdot\tau)}, \mathbf{x}^{(k+J\cdot\tau+1)}, \ldots\}$, thus we cannot expect the corresponding $\mathbf{w}_i$ vectors to be uncorrelated; nevertheless, their averaging will result in a kind of low-pass filtering, which can additionally decrease the level of noise in the constructed pattern $\bar{\mathbf{a}}$.

However, to avoid strong violation of the assumption concerning the fixed shape of the desired component, which is expressed by (18), the value of δ should not be large. Actually, it should be emphasized that the concept of using $\delta > 0$ can only be applied because of the very low frequency content of the visual EPs (which are additionally smoothed by the applied band-pass filtering).

REFERENCES

[1] L. A. Farwell and E. Donchin, "Talking off the top of your head: Toward a mental prosthesis utilizing event-related brain potentials," *Electroencephalogr. Clin. Neuriophysiol.*, vol. 70, no. 6, pp. 510–523, 1988.

[2] S. Sutton, M. Braren, J. Zubin, and E. R. John, "Evoked potential correlates of stimulus uncertainty," *Science*, vol. 150, pp. 1187–1188, 1965.

[3] Q. T. Obeidat, T. A. Campbell, and J. Kong, "Introducing the edges paradigm: A P300 brain–computer interface for spelling written words," *IEEE Trans. Human-Mach. Syst.*, vol. 45, no. 6, pp. 727–738, Dec. 2015.

[4] S. Noorzadeh, B. Rivet, and C. Jutten, "3-D interface for the P300 speller BCI," *IEEE Trans. Human-Mach. Syst.*, vol. 50, no. 6, pp. 604–612, Dec. 2020.

[5] J. Qu *et al.*, "A novel three-dimensional P300 speller based on stereo visual stimuli," *IEEE Trans. Human-Mach. Syst.*, vol. 48, no. 4, pp. 392–399, Aug. 2018.

[6] P. Stegman, C. S. Crawford, M. Andujar, A. Nijholt, and J. E. Gilbert, "Brain–computer interface software: A review and discussion," *IEEE Trans. Human-Mach. Syst.*, vol. 50, no. 2, pp. 101–115, Apr. 2020.

[7] A. Cruz, G. Pires, A. Lopes, C. Carona, and U. J. Nunes, "A self-paced BCI with a collaborative controller for highly reliable wheelchair driving: Experimental tests with physically disabled individuals," *IEEE Trans. Human-Mach. Syst.*, vol. 51, no. 2, pp. 109–119, Apr. 2021.

[8] A. Nourmohammadi, M. Jafari, and T. O. Zander, "A survey on unmanned aerial vehicle remote control using brain–computer interface," *IEEE Trans. Human-Mach. Syst.*, vol. 48, no. 4, pp. 337–348, Aug. 2018.

[9] A. Rakotomamonjy and V. Guigue, "BCI competition III: Dataset II – ensemble of SVMs for BCI P300 speller," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 3, pp. 1147–1154, Mar. 2008.

[10] M. Thulasidas, C. Guan, and J. Wu, "Robust classification of EEG signal for brain-computer interface," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 14, no. 1, pp. 24–29, Mar. 2006.

[11] S. Kundu and S. Ari, "P300 detection with brain–computer interface application using PCA and ensemble of weighted SVMs," *IETE J. Res.*, vol. 64, no. 3, pp. 406–416, 2018.

[12] Y-R. Lee and H-N. Kim, "A data partitioning method for increasing ensemble diversity of an eSVM-based P300 speller," *Biomed. Signal Process. Control*, vol. 39, pp. 53–63, 2018.

[13] C. Guger *et al.*, "How many people are able to control a P300-based brain-computer interface (BCI)?," *Neurosci. Lett.*, vol. 462, pp. 94–98, 2009.

[14] D. J. Krusienski *et al.*, "Toward enhanced P300 speller performance," *J. Neurosci. Methods*, vol. 167, no. 1, pp. 15–21, Jan. 2008.

[15] B. Blankertz *et al.*, "Single-trial analysis and classification of ERP components'a tutorial," *NeuroImage*, vol. 56, no. 2, pp. 814–825, May 2011.

[16] U. Hoffmann, J. Vesin, T. Ebrahimi, and K. Diserens, "An efficient P300-based brain computer interface for disabled subjects," *J. Neurosci.*, vol. 167, no. 1, pp. 115–125, 2008.

[17] Y. Miao *et al.*, "An ERP-based BCI with peripheral stimuli: Validation with ALS patients," *Cogn. Neurodynamics*, vol. 14, pp. 21–33, 2020.

[18] Y. Zhang, G. Zhou, J. Jin, Q. Zhao, X. Wang, and A. Cichocki, "Sparse Bayesian classification of EEG for brain–computer interface," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 11, pp. 2256–2267, Nov. 2016.

[19] Q. Wu, Y. Zhang, J. Liu, J. Sun, A. Cichocki, and F. Gao, "Regularized group sparse discriminant analysis for P300-based brain–computer interface," *Int. J. Neural Syst.*, vol. 29, no. 6, 2019, Art. no. 1950002.

[20] B. Rivet, A. Souloumiac, V. Attina, and G. Gibert, "xDAWN algorithm to enhance evoked potentials: Application to brain–computer interface," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 8, pp. 2035–2043, Aug. 2009.

[21] F. S. Rizi, V. Abootalebi, and M. T. Sadeghi, "Spatial and spatio-temporal filtering based on common spatial patterns and Max-SNR for detection of P300 component," *Biocybernetics Biomed. Eng.*, vol. 37, no. 3, pp. 365–372, 2017.

[22] X. Xiao, M. Xu, J. Jin, Y. Wang, T. -P. Jung, and D. Ming, "Discriminative canonical pattern matching for single-trial classification of ERP components," *IEEE Trans. Biomed. Eng.*, vol. 67, no. 8, pp. 2266–2275, Aug. 2020.

[23] X. Xiao *et al.*, "Enhancement for P300-speller classification using multi-window discriminative canonical pattern matching," *J. Neural Eng.*, vol. 18, no. 4, 2021, Art. no. 046079.

[24] D. J. Krusienski *et al.*, "A comparison of classification techniques for the P300 speller," *J. Neural Eng.*, vol. 3, no. 4, pp. 299–305, 2006.

[25] H. Cecotti and A. Graser, "Convolutional neural networks for P300 detection with application to brain-computer interfaces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 3, pp. 433–445, Mar. 2011.

[26] L. Vařeka, "Evaluation of convolutional neural networks using a large multi-subject P300 dataset," *Biomed. Signal Process. Control*, vol. 58, 2020, Art. no. 101837.

[27] O. Y. Kwon, M. H. Lee, C. Guan, and S. W. Lee, "Subject-independent brain–computer interfaces based on deep convolutional neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 10, pp. 3839–3852, Oct. 2020.

[28] D. Coyle, G. Prasad, and T. M. McGinnity, "Extracting features for a brain-computer interface by self-organising fuzzy neural network-based time series prediction," in *Proc. 26th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, San Francisco, USA, 2004, pp. 4371–4374, 2004.

[29] W. Kong *et al.*, "Weighted extreme learning machine for P300 detection with application to brain computer interface," *J. Ambient Intell. Humanized Comput.*, pp. 1–11, 2018.

[30] Z. Jin *et al.*, "EEG classification using sparse Bayesian extreme learning machine for brain–computer interface," *Neural Comput. Appl.*, vol. 32, pp. 6601–6609, 2020.

[31] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGNet: A compact convolutional neural network for EEG-based brain–computer interfaces," *J. Neural Eng.*, vol. 15, no. 5, 2018, Art. no. 056013.

[32] M. Simões *et al.*, "BCIAUT-P300: A multi-session and multi-subject benchmark dataset on autism for P300-based brain-computer-interfaces," *Front. Neurosci.*, vol. 14, 2020, Art. no. 568104.

[33] M. Liu, W. Wu, Z. Gu, Z. Yu, F. Qi, and Y. Li, "Deep learning based on batch normalization for P300 signal detection," *Neurocomputing*, vol. 275, pp. 288–297, 2018.

[34] F. Li, X. Li, F. Wang, D. Zhang, Y. Xia, and F. He, "A novel P300 classification algorithm based on a principal component analysis-convolutional neural network," *Appl. Sci.*, vol. 10, no. 4, 2020, Art. no. 1546.

[35] M. Kotas, J. Jeżewski, K. Horoba, and A. Matonia, "Application of spatio-temporal filtering to fetal electrocardiogram enhancement," *Comput. Methods Progr. Biomed.*, vol. 104, pp. 1–9, 2010.

[36] J. Giraldo-Guzmán, M. Kotas, F. Castells, S. H. Contreras-Ortiz, and M. Urina-Triana, "Estimation of PQ distance dispersion for atrial fibrillation detection," *Comput. Methods Progr. Biomed.*, vol. 208, 2021, Art. no. 106167.
[37] S. Lemm, B. Blankertz, G. Curio, and K-R. Müller, "Spatio-spectral filters for improving the classification of single trial EEG," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 9, pp. 1541–1548, Sep. 2005.
[38] G. Dornhege, B. Blankertz, M. Krauledat, F. Losch, G. Curio, and K. Muller, "Combined optimization of spatial and temporal filters for improving brain-computer interfacing," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 11, pp. 2274–2281, Nov. 2006.
[39] F. Qi, Y. Li, and W. Wu, "RSTFC: A novel algorithm for spatio-temporal filtering and classification of single-trial EEG," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 12, pp. 3070–3082, Dec. 2015.
[40] A. Jiang, J. Shang, X. Liu, Y. Tang, H. K. Kwan, and Y. Zhu, "Efficient CSP algorithm with spatio-temporal filtering for motor imagery classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 4, pp. 1006–1016, Apr. 2020.
[41] F. Qi *et al.*, "Spatiotemporal-filtering-based channel selection for single-trial EEG classification," *IEEE Trans. Cybern.*, vol. 51, no. 2, pp. 558–567, Feb. 2021.
[42] J. Lu, K. Xie, and D. J. McFarland, "Adaptive spatio-temporal filtering for movement related potentials in EEG-based brain–computer interfaces," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 22, no. 4, pp. 847–857, Jul. 2014.
[43] Y. Zhang, G. Zhou, Q. Zhao, J. Jin, X. Wang, and A. Cichocki, "Spatial-temporal discriminant analysis for ERP-based brain-computer interface," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 21, no. 2, pp. 233–243, Mar. 2013.
[44] S. M. Kay, "Fundamentals of statistical signal processing. detection theory," in *Signal Process. Ser.*. Hoboken, NJ, USA: Prentice-Hall, vol. 2, 1998.
[45] R. Bro, E. Acar, and Tamara G. Kolda, "Resolving the sign ambiguity in the singular value decomposition," *J. Chemometrics*, vol. 22, no. 2, pp. 135–140, 2008.
[46] G. D. Dawson, "A summation technique for detecting small signals in a large irregular background," *J. Physiol.*, vol. 115, no. 1, 1951.
[47] D. Garcia, "Robust smoothing of gridded data in one and higher dimensions with missing values," *Comput. Statist. Data Anal.*, vol. 54, no. 4, pp. 1167–1178, 2010.
[48] M. J. Buckley, "Fast computation of a discretized thin-plate smoothing spline for image data," *Biometrika*, vol. 81, no. 2, pp. 247–258, 1994.
[49] G. M. Foody, "Thematic map comparison: Evaluating the statistical significance of differences in classification accuracy," *Photogrammetric Eng. Remote Sens.*, vol. 70, no. 5, pp. 627–633, 2004.
[50] A. Agresti, *An Introduction to Categorical Data Analysis*. Wiley, 2018.
[51] M. N. Anastasiadou, M. Christodoulakis, E. S. Papathanasiou, S. S. Papacostas, and G. D. Mitsis, "Unsupervised detection and removal of muscle artifacts from scalp EEG recordings using canonical correlation analysis, wavelets and random forests," *Clin. Neuriophysiol.*, vol. 128, no. 9, pp. 1755–1769, 2017.
[52] R. Vigario, J. Sarela, V. Jousmiki, M. Hamalainen, and E. Oja, "Independent component approach to the analysis of EEG and MEG recordings," *IEEE Trans. Biomed. Eng.*, vol. 47, no. 5, pp. 589–593, May 2000.
[53] E. J. Berbari, S. M. Collins, and R. Arzbaecher, "Evaluation of esophageal electrodes for recording his-purkinje activity based upon signal variance," *IEEE Trans. Biomed. Eng.*, vol. BME-33, no. 10, pp. 922–928, Oct. 1986.

Marian P. Kotas received the M.Sc., Ph.D., and D.Sc. degrees in biomedical electronics from the Silesian University of Technology, Gliwice, Poland, in 1991, 1996, and 2012, respectively.

He is currently an Associate Professor with the Department of Cybernetics, Nanotechnology and Data Processing, Silesian University of Technology. His current research interests include linear and nonlinear filtering of biomedical signals, multivariate data processing and pattern recognition.

Michal Piela received the B.S. and M.Sc. degrees in biomedical engineering from the University of Science and Technology, Kraków, Poland, in 2010 and 2011, respectively. He is currently working toward the Ph.D. degree with the Department of Cybernetics, Nanotechnology and Data Processing, Silesian University of Technology, Gliwice, Poland.

His current research interests include biomedical signal processing, brain-computer interfaces, and evoked potentials detection.

Sonia H. Contreras-Ortiz received the B.S. and M.Sc. degrees in electronic engineering from Universidad Industrial de Santander, Bucaramanga, Colombia, in 2001, M.Sc. in 2004 respectively, and the Ph.D. degree in biomedical engineering from University of Connecticut, Storrs, CT, USA, in 2011.

She is currently the Director of the Biomedical Engineering program with Universidad Tecnológica de Bolívar, in Cartagena de Indias, Colombia. Her research interests include biosignal and medical image processing and analysis.