Neurocomputing xxx (xxxx) xxx

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Oscillatory Source Tensor Discriminant Analysis (OSTDA): A regularized tensor pipeline for SSVEP-based BCI systems

Tania Jorajuría [a,b,1,2], Mina Jamshidi Idaji [b,c,d,2], Zafer İşcan [e], Marisol Gómez [a], Vadim V. Nikulin [b,f,*,3], Carmen Vidaurre [a,*,3]

[a] Department of Statistics, Computer Science and Mathematics, Universidad Pública de Navarra, Pamplona, Spain
[b] Department of Neurology, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany
[c] International Max Planck Research School NeuroCom, Leipzig, Germany
[d] Machine Learning Group, Technical University of Berlin, Berlin, Germany
[e] Department of Electrical & Electronics Engineering, Faculty of Engineering and Natural Sciences, Bahcesehir University, Istanbul, Turkey
[f] Centre for Cognition and Decision Making, Institute for Cognitive Neuroscience, National Research University Higher School of Economics, Russian Federation

ARTICLE INFO

Article history:
Received 26 February 2021
Revised 28 May 2021
Accepted 25 July 2021
Available online xxxx
Communicated by Zidong Wang

Keywords:
Brain-computer interface
Steady-state visual evoked potential
Spatio-spectral decomposition
Higher order discriminant analysis
Analytical regularization
Tensor-based feature reduction

ABSTRACT

Periodic signals called Steady-State Visual Evoked Potentials (SSVEP) are elicited in the brain by flickering stimuli. They are usually detected by means of regression techniques that need relatively long trial lengths to provide feedback and/or sufficient number of calibration trials to be reliably estimated in the context of brain-computer interface (BCI). Thus, for BCI systems designed to operate with SSVEP signals, reliability is achieved at the expense of speed or extra recording time. Furthermore, regardless of the trial length, *calibration free* regression-based methods have been shown to suffer from significant performance drops when cognitive perturbations are present affecting the attention to the flickering stimuli. In this study we present a novel technique called Oscillatory Source Tensor Discriminant Analysis (OSTDA) that extracts oscillatory sources and classifies them using the newly developed tensor-based discriminant analysis with shrinkage. The proposed approach is robust for small sample size settings where only a few calibration trials are available. Besides, it works well with both low- and high-number-of-channel settings, using trials as short as one second. OSTDA performs similarly or significantly better than other three benchmarked state-of-the-art techniques under different experimental settings, including those with cognitive disturbances (i.e. four datasets with control, listening, speaking and thinking conditions). Overall, in this paper we show that OSTDA is the only pipeline among all the studied ones that can achieve optimal results in all analyzed conditions.

1. Introduction

A Brain-Computer Interface (BCI) is a system that uses brain signals to deliver commands according to their decoding [1]. Its applications range from neurological rehabilitation [2–6] to communication for disabled people [1,7], controlling external devices [8,9], marketing [10] and entertainment [11]. Over the past years, several BCI systems have been developed using different electroencephalogram (EEG) brain responses, including sensorimotor rhythms [12–18], event-related potentials [19,20] and visual-evoked potentials [21–25]. Among them, Steady-State Visual Evoked Potentials (SSVEP) have attracted attention due to their high signal-to-noise ratio (SNR) [26], robustness [27] and greater information transfer rate [28,29], when compared to other brain responses.

SSVEP signals are generated when a subject is looking at a flickering stimulus. As a response, periodic signals with fundamental frequency equal to the stimulus frequency, together with its harmonics, can be observed in the occipito-parietal regions [30]. Based on this phenomenon, the SSVEP-based BCIs can differentiate the brain response to stimuli with different flickering frequencies (classes) and therefore generate a specific output for each of those classes. In order to analyze SSVEP signals, several methodologies have been used in the recent years. Some of them include fast Fourier transform [31,32], power spectral density-based analysis [33–36], common spatial patterns [37], minimum energy combination [27,38,39], multivariate linear regression [40] and linear dis-

* Corresponding authors.
E-mail addresses: nikulin@cbs.mpg.de (V.V. Nikulin), carmen.vidaurre@unavarra.es (C. Vidaurre).
[1] Present address.
[2] Both authors contributed equally.
[3] Both senior authors contributed equally.

https://doi.org/10.1016/j.neucom.2021.07.103
0925-2312/

criminant analysis [41], among others. However, the most popular method in the literature is the Canonical Correlation Analysis (CCA) [42,43,39,44,45] and its extensions [46–48] or combinations with other methods [49]. CCA-based methods aim at finding spatial filters that maximize the correlation between SSVEP signals and sine–cosine reference templates at the fundamental and harmonic frequencies of the stimuli. Many CCA-based methods have the advantage of not requiring calibration (training) data, thus reducing or even eliminating the need of performing extra-recordings. Nevertheless, they do not consider the possibly high overlap between different classes and are very sensitive to cognitive perturbations that can be present in out-of-the-lab environments [50]. Recently, a new method called Task-Related Component Analysis (TRCA) [51,52] has been developed in order to reduce trial length in the design of SSVEP-based BCIs and operate at higher speeds. This method is based on the maximization of the inter-trial covariance matrix, and it was shown to be significantly more efficient than CCA under time restricting conditions such as short trial lengths. Another state-of-the-art method developed for high-speed SSVEP-based BCIs, which we will refer to as High-Speed BCI (HSBCI), was presented in [28]. HSBCI decomposes SSVEPs and template signals into sub-band components by applying a filter bank analysis, and employs CCA-based spatial filters to extract correlation features. Both TRCA and HSBCI use calibration data to train their corresponding feature extraction procedures. In fact, feature training has proven to critically increase SSVEP-based BCI performance [53]. TRCA and HSBCI methods were originally introduced for a set of pre-selected channels over parietal and occipital areas. Even though they both achieve high accuracy results with short trial lengths, their ability to accurately classify SSVEP data in the small sample size and in settings with higher number of channels has to be established yet. Besides, in this manuscript we also study how cognitive perturbations affect the aforementioned state-of-the art methods to understand how they could work in out-of-the-lab environments.

In this paper we propose a novel analysis pipeline, namely Oscillatory Source Tensor Discriminant Analysis (OSTDA), to extract SSVEP features from EEG data. Within this pipeline, we introduce a method called shrinkage higher order discriminant analysis (sHODA). sHODA is a modification of higher order discriminant analysis (HODA) [54] that includes an analytic shrinkage inspired by [55] as a tool for overcoming the small sample size problem. sHODA is then used to extract features from oscillatory brain sources, resulting in high accuracy results with trial lengths as short as one second. In this manuscript, we compare OSTDA to the aforementioned TRCA, HSBCI and CCA methods, and show that it has the benefits of all of them, but none of their disadvantages. In particular, we demonstrate that OSTDA is more robust in the small sample size setting, and less affected by perturbations. Moreover, our pipeline performs well with few channels and also in settings with a high number of channels, which is a required condition in certain applications. Thus, in this manuscript we propose a practical and fast SSVEP-based BCI system.

The paper is organized as follows: we introduce our proposed pipeline, OSTDA, in Section 2. There, we also present sHODA and show the details of the comparison between OSTDA and state-of-the-art methods. In Section 3 we describe the experimental data used in this study as well as the evaluation settings. In Section 4, we explain the details about OSTDA parameter selection, and in Section 5 we detail the statistical analysis done. Section 6 is dedicated to analyze results, which are discussed in Section 7. Finally, Section 8 includes a brief conclusion of the paper.

2. Methods

2.1. Notation

In this manuscript, tensors are denoted by calligraphic letters (e.g. $\mathcal{X}$), matrices are denoted by boldface capital letters (e.g. $\mathbf{X}$), whereas boldface lower-case letters are used to denote vectors (e.g. $\mathbf{x}$), and regular letters for scalars (e.g. x).

We denote the data matrix of the k-th training trial as $\mathbf{X}^{(k)} \in \mathbb{R}^{C\times T}$, where C is the number of channels and T is the number of time samples. Therefore, the training data with K trials is a three-way tensor $\mathcal{X} \in \mathbb{R}^{C\times T\times K}$.

2.2. Oscillatory Source Tensor Discriminant Analysis (OSTDA)

Fig. 1 shows the block diagram of our proposed feature extraction and classification pipeline, named Oscillatory Source Tensor Discriminant Analysis (OSTDA). In the following sections, we elaborate upon the details necessary to employ OSTDA.

2.2.1. Feature extraction

Spatio-Spectral Decomposition (SSD). SSD [56] is a multivariate and unsupervised method that aims at separating oscillatory sources from brain signals recorded with multiple electrodes by maximizing their signal-to-noise ratio (SNR) at the frequency band of interest. It computes spatial filters that are basis vectors of a subspace where the ratio of the signal power in a specific frequency band to the power of signals at its marginal bands is maximized.

Assume that the measured signal is $\mathbf{X} = \mathbf{S} + \mathbf{N}$, with $\mathbf{S}$ corresponding to the signal-of-interest in a specific frequency band and $\mathbf{N}$ the noise component in a broad band. While filtering $\mathbf{X}$ in the frequency band of interest (resulting in $\mathbf{X}_s$) gives an approximation of the signal-of-interest plus noise, filtering in two narrow bands (1 to 2 Hz) around the frequency band of interest (resulting in $\mathbf{X}_n$) approximates the noise alone. Using generalized eigenvalue

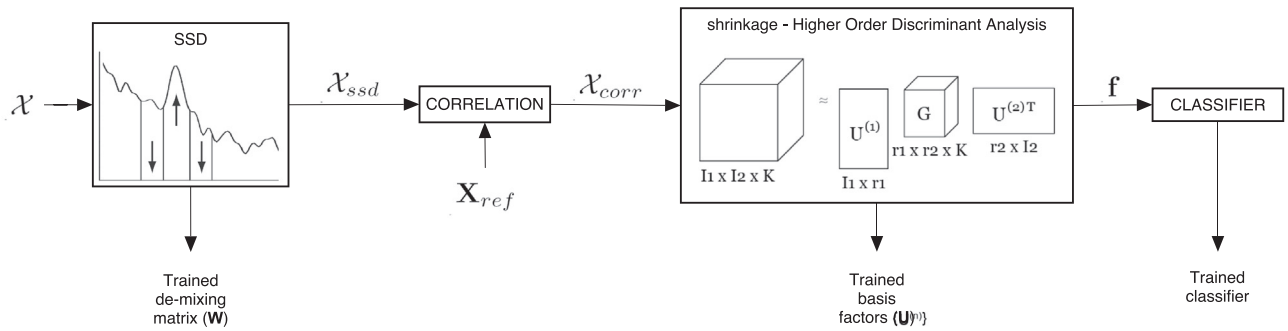

Fig. 1. The block diagram of our proposed pipeline, OSTDA.

decomposition, SSD finds the optimum spatial filter $\mathbf{w} \in \mathbb{R}^{C\times 1}$ that maximizes the ratio $\frac{\mathbf{w}^T\mathbf{C}_s\mathbf{w}}{\mathbf{w}^T\mathbf{C}_n\mathbf{w}}$, where $\mathbf{C}_s = \frac{1}{T}\mathbf{X}_s\mathbf{X}_s^T$ and $\mathbf{C}_n = \frac{1}{T}\mathbf{X}_n\mathbf{X}_n^T$.

In order to apply SSD to our data, we concatenated all the trials of the training tensor $\mathcal{X} \in \mathbb{R}^{C\times T\times K}$ to have a $\mathbf{X} \in \mathbb{R}^{C\times TK}$ data matrix. Then, we applied SSD on a broad-band range between $[5, 32]$ Hz. This band was selected because it covers all the fundamental and second harmonic frequencies of the flickering stimuli. Afterwards, we selected n_{SSD} sources ($n_{SSD} \leqslant C$) from the SSD components. We obtained the SSD spatial filters separately for each subject using the training trials. We denote the SSD spatial filters by $\mathbf{W} \in \mathbb{R}^{C\times n_{SSD}}$. The projection of $\mathcal{X}$ onto the SSD subspace is denoted by $\mathcal{X}_{ssd} \in \mathbb{R}^{n_{SSD}\times T\times K}$, that corresponds to brain sources with improved SNR within the frequency band $[5, 32]$ Hz.

Correlation with reference signals. As discussed before, a flickering stimulus elicits brain activities oscillating with the fundamental frequency equal to the flickering frequency. Therefore, one of the most informative SSVEP features is the similarity of the EEG with single-tone sines and cosines at each flickering frequency and its corresponding higher harmonics. In the case of this study, there are four classes and we use the information of first and second harmonics (similar to [50]). Thus, a total of $M = 16$ reference signals are necessary. In order to extract these similarities and to reduce the dimensionality of the features, we computed the Pearson's correlation of each SSD source signal with these reference signals grouped in $\mathbf{X}_{ref} \in \mathbb{R}^{M\times T}$. This resulted in the training data tensor $\mathcal{X}_{corr} \in \mathbb{R}^{n_{SSD}\times M\times K}$, which is a concatenation of feature matrices of all trials $\{\mathbf{X}_{corr}^{(k)}\}_{k=1}^{K} \subset \mathbb{R}^{n_{SSD}\times M}$. In the following, we refer to these feature matrices as *correlation feature matrices*.

Shrinkage Higher Order Discriminant Analysis (sHODA). Higher Order Discriminant Analysis (HODA) [54], initially introduced as DATER [57], is a generalization of linear discriminant analysis (LDA) [58] for tensor data $\mathcal{X}_k \subset \mathbb{R}^{I_1\times I_2\times\cdots\times I_N}$. In the special case of matrix data, which is also the case of this manuscript, assume $\{\mathbf{X}_k\}_{k=1}^{K} \subset \mathbb{R}^{I_1\times I_2}$ is the set of the training data points categorized in multiple classes. In order to find the discriminant subspace for this data, a simultaneous matrix factorization should be performed as the following [59]:

$$\mathbf{X}_k \approx \mathbf{U}^{(1)}\mathbf{G}_k\mathbf{U}^{(2)^T}, k = 1, \cdots, K \tag{1}$$

The orthogonal basis factors $\{\mathbf{U}^{(n)} \in \mathbb{R}^{I_n\times r_n}\}_{n=1}^{2}$ are computed, with the same idea as in LDA, by maximizing between-class covariance of the projections and minimizing the within-class covariance. The projections of data points (or the features) are computed as:

$$\mathbf{G}_k = \mathbf{U}^{(1)^T}\mathbf{X}_k\mathbf{U}^{(2)} \tag{2}$$

which can be vectorized and used for training a classifier. Test data can be projected onto this discriminant subspace by using Eq. 2. Note that $\mathbf{G}_k \in \mathbb{R}^{r_1\times r_2}$. Since $r_n < I_n, n = 1, 2$, the discriminant subspace has a lower dimension than the original subspace of the data.

HODA employs an alternating least squares (ALS) algorithmic scheme that iterates over the dimensions of the data tensor (here the first two dimensions). In each iteration scatter matrices are computed and the basis factors are estimated using an eigenvalue decomposition. In other words, each iteration of the ALS algorithm is an LDA problem as the following:

$$\boldsymbol{U}^{(n)} = \underset{\boldsymbol{U}^{(n)}}{argmax}\ \frac{tr\left[\boldsymbol{U}^{(n)^T}\boldsymbol{S}_b^{-n}\boldsymbol{U}^{(n)}\right]}{tr\left[\boldsymbol{U}^{(n)^T}\boldsymbol{S}_t^{-n}\boldsymbol{U}^{(n)}\right]}, \quad s.t. \quad \boldsymbol{U}^{(n)T}\boldsymbol{U}^{(n)} = \boldsymbol{I} \tag{3}$$

where $\boldsymbol{S}_b^{-n}$ and $\boldsymbol{S}_t^{-n}$ are the between-class and total scatter matrices of the n-th mode of the data tensor. In our special case of matrix data, $n = 1, 2$. More details about the computation of the scatter matrices are presented in appendix A.

Tensor-based feature extraction methods are known to cope well with small sample size scenarios [60,54]. However, even these methods suffer from ill-conditioned scatter matrices. Therefore, similar to conventional LDA, it is suggested that the within-class covariance matrix should be regularized [54]. Previous works [54,61] have implemented this regularization by means of a constant regularization parameter. However, as we know from the literature on EEG classification, LDA can benefit from shrinkage, where the shrinkage parameter is computed analytically in a data-driven manner [55,62,63]. Inspired by [55], we implemented an analytic shrinkage in HODA, and called it shrinkage-HODA (sHODA), where the total scatter matrix of Eq. 3 is regularized in each of the iterations of the ALS algorithm. The optimization problem of sHODA is formulated as the following:

$$\boldsymbol{U}^{(n)} = \underset{\boldsymbol{U}^{(n)}}{argmax}\ \frac{tr[\boldsymbol{U}^{(n)^T}\boldsymbol{S}_b^{-n}\boldsymbol{U}^{(n)}]}{tr[\boldsymbol{U}^{(n)^T}(\boldsymbol{S}_t^{-n}+\gamma^{(n)}\mathbf{I})\boldsymbol{U}^{(n)}]}, \quad s.t. \quad \boldsymbol{U}^{(n)T}\boldsymbol{U}^{(n)} = \boldsymbol{I} \tag{4}$$

where $\gamma^{(n)}$ is the regularization parameter of mode n. sHODA computes this regularization parameter of each mode in a data-driven approach. As mentioned before, [54,61] have introduced a constant regularization parameter to the total scatter matrix, i.e. $\gamma^{(n)} = \gamma_0 = const, \forall n$. The constant regularization adds a challenge of selecting this parameter as a hyper-parameter of the pipeline, which makes the calibration phase computationally more expensive. Additionally, a constant regularization parameter may not be the optimum value for regularizing the LDA problem in all iterations of the ALS algorithm. sHODA overcomes both of these drawbacks and can improve the performance of the feature extraction in small sample size scenarios. Additionally, data-driven computation of the shrinkage parameter eliminates one parameter from the cross-validation parameter tuning. For the details about the formulation and algorithm of sHODA, we refer the reader to appendix A of this manuscript.

After applying sHODA on the correlation features of SSVEP data (i.e. $\{\mathbf{X}_{corr}^{(k)}\}_{k=1}^{K} \subset \mathbb{R}^{n_{SSD}\times M}$), the feature matrix $\mathbf{f} \in \mathbb{R}^{K\times r_1 r_2}$ is obtained and used for classification. Note that the only parameters of sHODA are the new dimensions of the feature space, i.e. $\mathbf{R} = [r_1, r_2]$.

2.2.2. Classification

As in the original paper describing and analysing the datasets used here [50], we applied the K-Nearest Neighbor (KNN) classifier with five neighbors for all the analyses performed in this paper with the OSTDA pipeline. The KNN implementation of the Statistics and Machine Learning Toolbox of MATLAB® was used.

2.3. Benchmarking

We compared our proposed pipeline, OSTDA, with the following state-of-the art methods.

Canonical Correlation Analysis (CCA). CCA is probably the most popular method for feature extraction in SSVEP-based BCIs [42,43,39,44,45]. CCA solves an optimization problem to find linear mappings that maximize the correlation of two matrices. Here, these two matrices are the EEG data from multiple electrodes and the reference signals. In [50], for each SSVEP flickering frequency, two sine–cosine reference signal matrices were used to calculate canonical correlations; one of these matrices corresponded to the stimulation frequency and the other one to its second harmonic. Therefore, four canonical correlation values (CCA features) were computed for each class, resulting in sixteen CCA features.

To implement the above-mentioned CCA-based feature extraction in this paper, each trial was first detrended and band-pass fil-

tered between $[0.53, 40]$ Hz. This step removes DC, high frequency artifacts and the 50 Hz power line noise [50]. Afterwards, the sixteen CCA features were computed for each trial and classified with a KNN classifier with five neighbors. This classifier achieved the best classification results in the testing dataset in [50].

Task-Related Component Analysis (TRCA). TRCA [51] is a recently published algorithm applied to high-speed SSVEP-based BCIs. This method finds spatial filters that maximize the inter-trial covariance matrix of each class separately by means of a generalized eigenvalue decomposition (more details can be found in [52]).

In this paper, we followed the procedure explained in [51] to implement TRCA, where the SSVEP signals are filtered into sub-band components by using a filter bank. In order to fit the parameters of that work to the data used in this manuscript, the lowest cut-off frequency was set to 5 Hz, thus we could include the smallest flickering frequency. We computed the ensemble spatial filters and the averaged templates for each stimuli and sub-band with the training trials.

High-Speed BCI (HSBCI). A new high-speed SSVEP-based BCI was introduced in [28]. HSBCI consists on a filter bank analysis that decomposes SSVEPs and template signals into sub-band components to extract correlation features using CCA-based spatial filters. We calculated the SSVEP template signals with the training trials.

In order to implement HSBCI, we followed the procedure explained in [28] with some exceptions to adapt the method to the data analyzed here: the reference signals matrix to compute CCA included the first and second harmonics, so the total frequency range was $[5, 32]$ Hz. As for TRCA, we also set the lowest cut-off frequency to 5 Hz to include the smallest flickering frequency.

3. Experimental data

3.1. Data description

In this paper, we analyzed a public EEG dataset of a SSVEP-based BCI recorded from 24 subjects from [50]. All subjects participated in the study after giving a written informed consent in accordance with the declaration of Helsinki and the experiments were approved by the local ethics committee of National Research University Higher School of Economics, Moscow [50]. The visual stimuli consisted of four circles placed in different locations of a screen, flickering at distinct frequencies of 5.45 Hz, 8.57 Hz, 12 Hz, and 15 Hz. Frequencies multiples of each other were avoided in order to prevent coincidences between frequencies of flickering stimuli and their higher harmonics. For more details about the experimental setup, we refer the reader to [50].

Participants underwent offline and online (with feedback) recording sessions. In this study, we used the offline data as the training dataset and the online data as the testing dataset.

The offline session consisted of four runs of 25 trials each, with a total of 100 trials (25 trials per class). In each offline trial, the subjects were asked to focus on a randomly selected flickering circle surrounded by a red oval for three seconds, followed by a one-second resting period. In the online session, the subjects freely chose a flickering circle to focus on for three seconds. After each trial, the classification result was presented to the subject as a feedback and they had to confirm or reject the result using the keyboard. There were four conditions (with and without perturbation) in the online session (100 trials each, per subject). During each condition, the subjects were asked to perform randomly one of the following tasks, while looking at the stimulus: control (no perturbation), speaking (counting loudly and repeatedly from one to ten), thinking (counting silently from one to ten repeatedly), and listening (paying attention to a pre-recorded audio file where the participant had counted from one to ten repeatedly).

EEG data were recorded from 60 channels based on the international 10–20 system, as in Fig. 2, with the reference electrode on the left mastoid and 1 kHz sampling rate. The data were segmented using the stimuli markers specifying the onset and end of the flickering. As aforementioned, offline and online trials of the studied datasets were three seconds long. However, long trials decrease the speed of BCI systems, thus an effort must be done to obtain reliable results with short trial lengths. Consequently, in this study we focused on short trials of only one second, with the aim of increasing the information transfer rate of our system.

3.2. Evaluation settings

We evaluated the performance of two different settings with our proposed pipeline and with the selected state-of-the-art methods. In particular, we analyzed how the training sample size and the number of channels affect the performance of the studied methods. The factor "sample size" addresses the amount of data that is necessary to obtain a reliable performance. Reducing the need of training (also called calibration) data, increases the ease of use of BCIs. Typically, recording offline data is challenging because participants become tired before starting the actual feedback BCI session. This problem is even greater for patients, whose attention span might be compromised. Thus, BCI researchers investigate how short the training sessions can be to still obtain reliable online performances [64–66], as we do in this work. On the other hand, the factor "number of channels" is related to the trade-off between a fast and easy setup with a low number of dimensions and the need of flexible paradigms, specially in the case of participants with special needs, for example due to pathologies of the nervous system. This means that, specially for the case of patients, it is not always possible to select only a few optimal channels, and hence flexible systems with a higher number of dimensions might be necessary. Thus, in order to design BCI systems that can be used in out-of-the-lab conditions, it is necessary to study how reliable they are under low sample size settings and with higher number of dimensions, as we do here. That is, it is necessary to analyze the trade-off between number of channels and the amount of available training data.

Sample size. Small sample size (SSS) problems in classification arise when there are not enough training data points in comparison to the number of unkowns. This condition results in overfitting of the machine learning methods and therefore, a significant drop in the classification results of test data is observed. In order to assess the performance of our pipeline and other state-of-the-art methods in SSS settings, we computed the classification results for two numbers of training trials values. In particular, we used only the first 5 or 25 trials of each class from the training dataset for training. We refer to the case of using all the available training data (i.e. 25 trials per class) as the full sample size (FSS) setting.

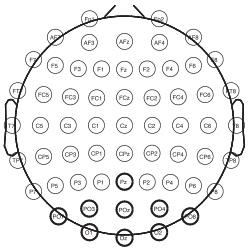

Fig. 2. Scalp plot of the electrodes configurations used in this study. 60-channel setting: all shown electrodes; 9-channel setting: bold-circled electrodes.

Note that there are four classes and therefore we have in total 100 trials to train in FSS setting and 20 in the SSS setting.

Number of channels. We evaluated two different settings regarding the number of channels used. The first setting included all 60 EEG channels (all electrodes shown in Fig. 2). For the second setting, only nine channels located in parietal and occipital areas (bold-circled electrodes in Fig. 2) were selected according to [28,51]. These channels are known to provide very good performance for steady-state *visual* evoked potential paradigms [67].

4. Parameter selection

OSTDA pipeline has two parameters that were subject-specifically selected: n_{SSD} (the number of SSD components that are retained) and $\mathbf{R} = [r_1, r_2]$ (the number of components retained by sHODA in each mode).

The range of n_{SSD} was determined based on the number of channels used [68]. For the 9-channel setting $5 \leqslant n_{SSD} \leqslant 9$, while for the 60-channel setting $10 \leqslant n_{SSD} \leqslant 35$.

$\mathbf{R}$ was restricted to a specific set of possible values based on the resulting total number of features (equal to the product of the two elements of vector $\mathbf{R}$, i.e. $r_1 r_2$) according to the following criteria: as we have four SSVEP flickering frequencies (classes), the lower limit was set to three; additionally, taking into account that the total number of trials in the FSS is 100, the upper bound was limited to about a 10% of this value [69], in this case to 12. Thus, we had $3 \leqslant r_1 r_2 \leqslant 12$. Therefore, the resulting possible values for $\mathbf{R}$ were: $[1,3], [1,4], [1,5], [1,6], [1,7], [1,8], [1,9], [1,10], [1,11], [1,12], [2,2], [2,3], [2,4], [2,5], [2,6], [3,1], [3,2], [3,3], [3,4], [4,1], [4,2], [4,3], [5,1], [5,2], [6,1], [6,2], [7,1], [8,1], [9,1], [10,1], [11,1], [12,1]$.
Among these values, the ones that satisfied $r_1 \leqslant n_{SSD}$ were used for each n_{SSD}.

For each evaluation setting, we selected n_{SSD} and $\mathbf{R}$ for each subject based on the eigenvalues obtained in sHODA, taking the combination of these two parameters that maximized the explained percentage of variance. This was done with a chronological hold-out cross-validation using the corresponding training dataset in each sample size setting, taking the last 40% trials of each class as the validation set. With these subject-specific parameters, OSTDA was trained with all available trials in each sample size setting.

5. Statistical Analysis

Two 3-way repeated measures ANOVA were applied to the classification results obtained with the testing dataset to see the influence of the analysis *approach* (OSTDA, TRCA, HSBCI, CCA), *number of training trials* per class $(5, 25)$ and *perturbation* (Control, Listening, Speaking, Thinking) factors. They were performed separately for individual accuracy values obtained with the 9- and 60-channel settings, respectively.

For each ANOVA, we analyzed the sphericity of the response variables in the repeated measures model, using a Mauchly's test. In case of rejection, we report the Greenhouse-Geisser adjusted p-values. Furthermore, after each ANOVA we performed post hoc tests using Tuckey's honestly significant difference (HSD), when the interaction of main factors was significant.

6. Results

6.1. Parameter selection (training dataset)

As mentioned in Section 4, we performed a chronological hold-out to tune the parameters of OSTDA (n_{SSD} and $\mathbf{R}$) for each subject. The last 40% trials of each class of the corresponding training data were used as the validation set. Thus, subject-dependent parameters were selected in each of the evaluation settings, i.e. 9- and 60-channel as well as small and full sample size (refer to Section 3.2).

Fig. 3 shows the joint distribution of the selected parameters, $\{n_{SSD}, \mathbf{R}\}$, over all the subjects in each setting. In the 9-channel setting the best parameters for most of the subjects are $\{5, [2,6]\}$ in the SSS setting, selected for 75% of subjects, and $\{5, [1,3]\}$ and $\{5, [2,6]\}$ in the FSS setting, selected for 50% and 45.83% of subjects, respectively. In the 60-channel setting the predominant parameters are $\{10, [1,3]\}$, which were selected for 70.83% and 75% of subjects in the SSS and the FSS settings, respectively.

6.2. Classification (testing dataset)

We benchmarked OSTDA against three state-of-the art methods, namely CCA, TRCA and HSBCI. Fig. 4 depicts test classification mean accuracy values (%) and standard errors in different evaluation settings and under different perturbations. Besides, in Fig. 5 these mean accuracy values obtained with analyzed approaches are ranked for each evaluation setting and perturbation condition.

Two 3-way repeated measures ANOVA were separately conducted (see Tables B.1 and B.2 in appendix B) for the individual accuracies obtained in 9-channel and 60-channel settings. Each ANOVA was designed with factors *approach* (OSTDA, TRCA, HSBCI, CCA), *number of training trials* per class $(5, 25)$ and *perturbation* (Control, Listening, Speaking, Thinking).

9-channel setting. The analysis of the 9-channel setting delivered significant results for all main effects ($p < 0.001$ in each case) and also for the 2-way interaction between *approach* and *perturbation* ($p < 0.01$) factors. Post-hoc tests showed that OSTDA performed significantly better than CCA in Control, Speaking, Thinking ($p < 0.001$ in each case) and Listening ($p < 0.01$) conditions. There were no significant differences between OSTDA and HSBCI nor between OSTDA and TRCA in any perturbation. Quantitatively, however, mean accuracy values achieved with OSTDA were higher than for TRCA and HSBCI in all sample sizes and perturbations. Specifically, they were 4.7% and 0.7% higher than for

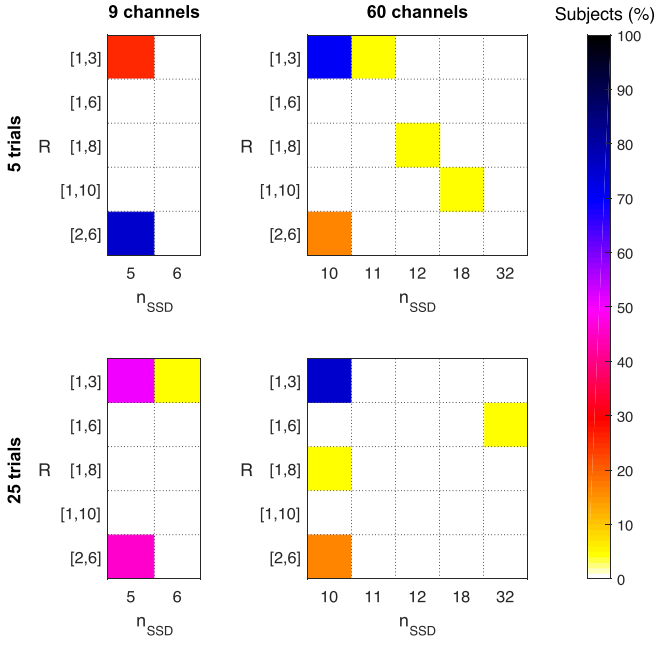

Fig. 3. The joint distribution of selected OSTDA parameters ($\{n_{SSD}, \mathbf{R}\}$) over $N = 24$ subjects in 9-channel and 60-channel settings for the SSS (5 training trials per class) and the FSS (25 training trials per class) settings.

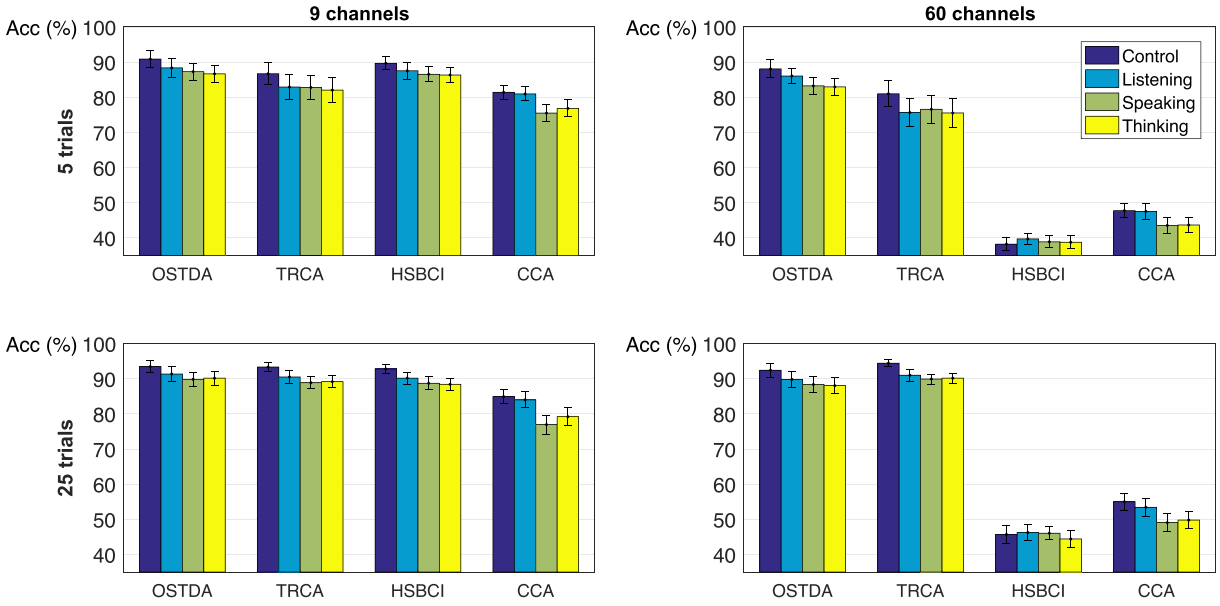

Fig. 4. Test classification mean accuracy values (%) and standard errors under different perturbations for different evaluation settings (N = 24 subjects). Different approaches are represented by different bar groups.

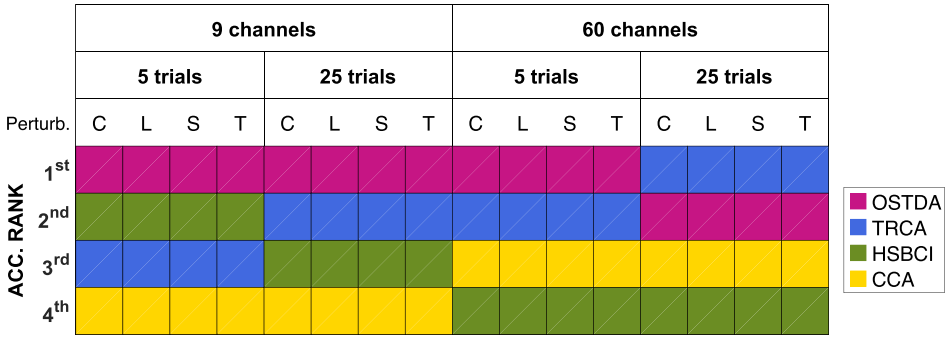

Fig. 5. Ranking of test classification mean accuracy values in each evaluation setting and perturbation (C: control, L: listening, S: speaking, T: thinking). First rank means highest mean accuracy value. Each approach is represented with a different color.

TRCA in SSS and FSS settings, respectively, and 1% higher than for HSBCI in average for both sample size settings.

When comparing mean accuracies between perturbations for each approach, post hoc tests showed that Control had significantly better accuracy than Speaking and Thinking for CCA, HSBCI, TRCA and OSTDA ($p < 0.05$ in each case). Moreover, mean accuracies in Control were also higher than in Listening ($p < 0.05$) for HSBCI. Besides, in the case of CCA Listening condition had better accuracy than Speaking ($p < 0.01$) and Thinking ($p < 0.05$).

60-channel setting. The analysis of 60-channel setting showed a significant 2-way interaction between *approach* and *number of training trials* ($p < 0.01$) factors, between *approach* and *perturbation* ($p < 0.05$), and significant main effects for *approach*, *number of training trials* ($p < 0.001$ in each case) and *perturbation* ($p < 0.01$) factors. Post-hoc comparisons revealed that OSTDA and TRCA obtained significantly better results than CCA and HSBCI in all evaluated sample size settings and perturbations ($p < 0.001$ in each case). In relation to OSTDA and TRCA, there were no significant differences between these two approaches in the FSS setting, but the SSS setting showed that OSTDA was significantly better than TRCA ($p < 0.05$).

Regarding perturbations, OSTDA performed significantly better than TRCA in Listening ($p < 0.05$) condition and, when averaging results of both sample size settings, quantitatively better in Control (2.5%), Speaking (2.6%) and Thinking (2.6%). Comparing mean accuracy values between perturbations for each of these two approaches, post hoc results showed that Control condition had better accuracy than Speaking and Thinking ($p < 0.05$ and $p < 0.01$, respectively) for both OSTDA and TRCA. Moreover, classification results in Control condition were also better than in Listening ($p < 0.05$) for TRCA.

Finally, in order to depict how perturbations are handled by OSTDA, we studied the distribution of power on the scalp before and after applying SSD. In particular, we computed the mean power of each channel over all trials in the 60-channel and full sample size setting. In Fig. 6, scalp plots of the power distribution of an illustrative subject are depicted before and after SSD, in different conditions. We used EEGLAB [70] Toolbox of MATLAB® to plot this figure. In order to plot the power distribution after SSD, we projected subject-specific n_{SSD} sources back to the sensor space with SSD trained patterns, and then computed the sensor powers.

As shown in Fig. 6, before SSD, the brain activity is located at different brain regions depending on the perturbation. In particular, before SSD the Listening condition activity is mainly located at temporal areas and during the Thinking condition predominant areas are frontal and temporal regions. Regarding the Speaking perturbation, the power is distributed through the whole scalp, including frontal and sensorimotor areas. After SSD, however, the

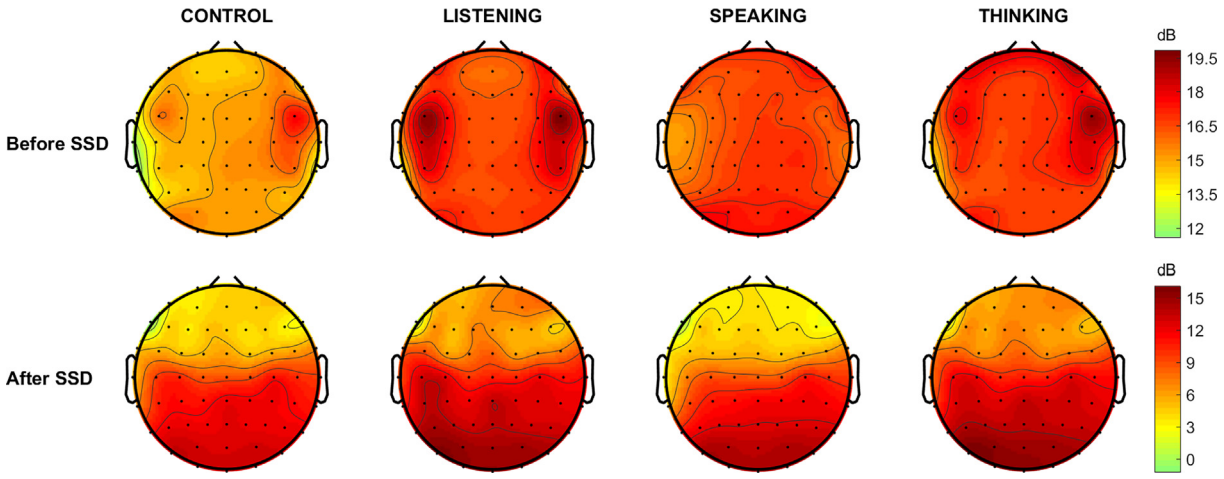

Fig. 6. Power distribution on scalp (dB) of an illustrative subject under perturbation conditions, before and after SSD (testing dataset).

brain activity is mainly located at occipital and parietal areas in all conditions.

7. Discussion

7.1. OSTDA versus state-of-the-art methods

In Section 6.2, it was shown that OSTDA is the only pipeline that can achieve optimal results in all analyzed settings (number of channels and number of training trials). OSTDA performed similarly well to HSBCI and TRCA in the 9-channel setting, although it was quantitatively better in both sample sizes and all analyzed perturbations, and significantly better than CCA. In the 60-channel setting, HSBCI and CCA failed in comparison to OSTDA and TRCA in all sample size settings and perturbations. In this last case, HSBCI attained quantitatively the worst performance of all pipelines, as shown in Fig. 5. Regarding OSTDA and TRCA in the 60-channel setting, OSTDA achieved significantly better classification results than TRCA in the SSS setting, and similar in the FSS. Moreover, OSTDA performed significantly better than TRCA in Listening condition, and similar in the other three, showing however quantitatively higher mean accuracies when averaging both sample sizes.

OSTDA is more robust in the small sample size setting than other analyzed approaches. In Section 6.2 it was shown that there was a significant two-way interaction between *approach* and *number of training trials* factors when using 60 channels. In this channel setting, CCA and HSBCI did not even reach 60% mean accuracy in any sample size (see Fig. 4), and both performed significantly worse than OSTDA. Furthermore, OSTDA achieved significantly better results than TRCA for the SSS setting, and there were no significant differences in the FSS setting.

Thus, OSTDA is less sensitive than the other studied approaches to the small sample size problem. There are two reasons for this: on the one hand, SSD not only increases the SNR of the data, but also allows dimensionality reduction based on their oscillatory properties. On the other hand, tensor based techniques are known to be more robust than their vectorized counterparts when not enough training data is available [60,71,57]. The reason of this robustness in small sample size settings is that in tensor-based feature extraction methods the dimension of each mode is smaller than the whole dimension of the vectorized data constructed from concatenating all the data modes together, alleviating the curse of dimensionality problem. This, in turn, offers a desirable alternative to design robust BCI-based systems under small sample size conditions.

OSTDA is robust with short trial lengths. As shown in [50], CCA is limited by the long trial lengths necessary to obtain stable results. This is because *calibration free* regression-based techniques need long epoch-lengths that contain a sufficient number of cycles in order to achieve reliable estimates. As demonstrated in this study, this is not the case for OSTDA, which operates well with trial lengths as short as one second. Therefore, OSTDA is a suitable candidate for implementing fast SSVEP-based BCIs, which is a desired feature for these systems.

OSTDA is optimal using a low or a large number of electrodes. As discussed before, OSTDA was the only pipeline that did not fail in any analyzed setting. Its performance in the 9-channel setting was quantitatively the best one among all approaches (see Fig. 5), being significantly better than CCA and similar to HSBCI and TRCA. Those 9 channels are in fact optimal for visual paradigms [67]. However, it is not always possible to know beforehand which channels are the best. For instance, when a new experimental condition should be tested or a paradigm mixes different types of stimuli (for example visual and auditory), it might be necessary to increase the number of channels to be able to efficiently perform experiments. This might also be the case when a paradigm is designed to use different visual stimuli in left and right visual hemifields. It might also be necessary to increase the number of channels when there is a need to discriminate between retinotopically close stimuli. These stimuli are processed in neighboring brain regions, and thus require a high spatial EEG sampling. In all these situations, OSTDA presents a clear advantage over the studied approaches.

In particular, our results showed that CCA and HSBCI fail in the 60-channel setting. In [72], an electrode channel optimization method is proposed for SSVEP-based BCIs, where CCA and filter bank canonical correlation analysis methods (in which HSBCI is based) are applied as target identification algorithms. Their results show that, when increasing the number of electrode channels, the performance also increases until it reaches a maximum value, and then starts decreasing. The reason for this low performance with high number of channels is probably the collinearity problem that arises in multivariate regression algorithms [73].

Additionally, TRCA performed similarly to OSTDA in the FSS setting with 60 channels, but significantly worse under SSS. Thus, OSTDA offers an optimal option for both channel settings, specially for SSS problem in 60-channel setting, where the other three analyzed approaches performed significantly worse.

7.2. Effect of perturbations in SSVEP-based BCI

As shown in Section 6.2, in 9-channel setting OSTDA, HSBCI and TRCA performed similarly well under the perturbations. Quantitatively, however, OSTDA was better than both of them, particularly almost 5% better than TRCA in the SSS setting. On the other hand, CCA was worse than OSTDA in all perturbations, in the two sample size settings. Besides, all approaches showed significant differences between Control-Speaking and Control-Thinking conditions, but HSBCI exhibited differences between Control-Listening conditions as well, and CCA also between Listening-Speaking and Listening-Thinking conditions, thus being less robust to perturbations. In the 60-channel setting, CCA and HSBCI failed under all perturbations. OSTDA performed significantly better than TRCA in Listening condition, and similar in the other three, but obtained quantitatively better results than TRCA when averaging results of both sample size settings. Moreover, OSTDA and TRCA showed significant differences between Control-Speaking and Control-Thinking conditions, but TRCA also revealed differences between Control-Listening conditions, proving to be less robust to the Listening perturbation than OSTDA.

In summary, OSTDA is the least affected approach under perturbations. This implies that OSTDA might be able to facilitate the design of practical systems based on SSVEP to be used in out of the laboratory conditions. Such benefits are due to the combination of SSD and our tensor-based approach. SSD subspace, computed from the training data, extracts the oscillatory sources and improves their SNR in the desired frequency band. At the same time, disturbances, as the ones studied here, are often not of oscillatory nature. Thus, SSD not only reduces dimensions to those containing relevant information, but simultaneously it also reduces the effect of non-oscillatory activity. Fig. 6 of Section 6.2 shows that after applying SSD, the power of the brain activity corresponding to the perturbations decreases. This, in turn, allows the occipital and parietal neuronal activities containing SSVEP information emerge. Paralelly, in the SSVEP setting, reducing dimensions to those with higher SNR means that the power and hence amplitude of the signals improves, simultaneously reducing noise contributions. These two effects can significantly improve the estimation of correlations. Furthermore, and as our results suggest, by projecting contaminated test data on only relevant informative dimensions it is possible to reduce the effect of perturbations.

8. Conclusion

In this manuscript we presented a new pipeline called Oscillatory Source Tensor Discriminant Analysis (OSTDA) to robustly and efficiently classify SSVEP data. We showed that it is possible to use short trial lengths and a very short calibration recording of just a few samples per class to train the OSTDA pipeline. We also analyzed its performance in different perturbing states of the test data resembling out-of-the-lab conditions, and in settings that had low or high number of channels. We were able to show that, unlike the rest of studied pipelines, OSTDA does not fail in any setting; it is always better or similar to the other studied approaches, thus offering an optimal alternative. Specifically, for small sample size problems with high number of channels, a difficult setting since there are many features and few training trials, OSTDA is significantly better than all the other three state-of-the-art pipelines.

CRediT authorship contribution statement

Tania Jorajuría: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing - original draft, Writing - review & editing, Visualization. **Mina Jamshidi Idaji:** Conceptualization, Methodology, Software, Validation, Formal analysis, Writing - original draft, Writing - review & editing. **Zafer İşcan:** Investigation, Data curation, Writing - review & editing. **Marisol Gómez:** Formal analysis, Writing - review & editing. **Vadim V. Nikulin:** Conceptualization, Methodology, Formal analysis, Resources, Writing - review & editing, Supervision, Project administration. **Carmen Vidaurre:** Conceptualization, Methodology, Software, Formal analysis, Investigation, Software, Writing - review & editing, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

TJ was partly supported by the European Erasmus + Program for international mobility within Campus Iberus. VVN was supported in part by the Basic Research Program at the National Research University Higher School of Economics (HSE University). CV was supported by MINECO-RyC-2014–15671.

Appendix A. sHODA

Notation. We use the notations of [60,54] for tensor algebra. The mode-n matricization (unfolding) of a tensor $\mathscr{X} \in \mathbb{R}^{I_1 \times \cdots \times I_N}$ is defined as rearranging its mode-n fibers in a matrix denoted by $\boldsymbol{X}_{(n)} \in \mathbb{R}^{I_n \times (\prod_{i=1, i\neq n}^{N} I_i)}$. The product of the mode-n unfolding $\mathscr{X}$ and a matrix $\boldsymbol{A} \in \mathbb{R}^{J_n \times I_n}$ defines the mode-n product of a tensor and matrix denoted by $\mathscr{Z} = \mathscr{X} \times_n \boldsymbol{A} \in \mathbb{R}^{I_1 \times \cdots \times J_n \times \cdots \times I_N}$, where $\boldsymbol{Z}_{(n)} = \boldsymbol{A}\boldsymbol{X}_{(n)}$. Finally, the two products $\mathscr{Z} = \mathscr{X} \times_1 \boldsymbol{A}^{(1)} \times_2 \cdots \times_N \boldsymbol{A}^{(N)}$ and $\mathscr{Z} = \mathscr{X} \times_1 \boldsymbol{A}^{(1)} \times_2 \cdots \times_{n-1} \boldsymbol{A}^{(n-1)} \times_{n+1} \boldsymbol{A}^{(n+1)} \times_{n+2} \cdots \times_N \boldsymbol{A}^{(N)}$ are written in summarized form as $\mathscr{Z} = \mathscr{X} \times \{\boldsymbol{A}\}$ and $\mathscr{Z} = \mathscr{X} \times_{-n} \{\boldsymbol{A}\}$ respectively. Finally $\langle \mathscr{X}, \mathscr{X} \rangle_{-n} := \boldsymbol{X}_{(n)} \boldsymbol{X}_{(n)}^T \in \mathbb{R}^{I_n \times I_n}$.

Problem Formulation. Let $\{\mathscr{X}_k \in \mathbb{R}^{I_1 \times \cdots \times I_N}\}_{k=1}^{K}$ be the training tensor data categorized in C classes with K_c members in each class. The multi-way feature extraction methods aim at computing orthonormal basis factors $\{\mathbf{U}^{(n)} \in \mathbb{R}^{I_n \times r_n}\}_{n=1}^{N}$ that define a *feature subspace* where the separability of the classes is maximized. The projection of a tensor data $\mathscr{X}_k$ onto this subspace is computed by $\mathscr{G}_k = \mathscr{X}_k \times \{\mathbf{U}^T\}$.

Higher Order Discriminant Analysis (HODA). HODA is a multi-way extension of linear discriminant analysis (LDA), where the basis of the feature subspace is computed by simultaneous maximization of the between-class covariance and minimization of the within-class covariance of the features on the feature subspace. Based on a similar idea, HODA maximizes the Fisher ratio of the projections of the training data on the feature subspace as the following[4]:

$$\{\mathbf{U}\} = \underset{\boldsymbol{U}^{(1)},\cdots,\boldsymbol{U}^{(N)}}{argmax} \frac{\sum_{c=1}^{C} K_c ||\bar{\mathscr{G}}^{(c)} - \bar{\mathscr{G}}||_F^2}{\sum_{k=1}^{K} ||\mathscr{G}^{(k)} - \bar{\mathscr{G}}^{(c_k)}||_F^2} \quad s.t. \quad \boldsymbol{U}^{(n)T}\boldsymbol{U}^{(n)} = \boldsymbol{I}, \quad n = 1, \cdots, N \tag{A.1}$$

[4] The notation and almost all the equations of this part are taken from [60].

In the above equation, $\bar{\bar{\mathscr{G}}}$ is the mean of all $\{\mathscr{G}_k\}_{k=1}^{K}$. While $\bar{\mathscr{G}}^{(c)}$ is the mean of the members of class c, $\bar{\mathscr{G}}^{(c_k)}$ is the mean of the members of the class to which the k-th data tensor belongs. The numerator and denominator of Eq. A.1 reflect the between-class and within-class variance, respectively.

The above equation is solved using the alternating least square (ALS) algorithm. In the n-th iteration of the ALS algorithm, $\mathbf{U}^{(n)}$ is optimized while all other basis matrices are used from the estimates of the previous iterations. Therefore, in iteration n, Eq. A.1 can be rewritten as the following:

$$\boldsymbol{U}^{(n)} = \underset{\boldsymbol{U}^{(n)}}{argmax} \frac{tr[\boldsymbol{U}^{(n)^T}\boldsymbol{S}_b^{-n}\boldsymbol{U}^{(n)}]}{tr[\boldsymbol{U}^{(n)^T}\boldsymbol{S}_t^{-n}\boldsymbol{U}^{(n)}]}, \quad s.t. \quad \boldsymbol{U}^{(n)T}\boldsymbol{U}^{(n)} = \boldsymbol{I} \tag{A.2}$$

where $\boldsymbol{S}_b^{-n}$ and $\boldsymbol{S}_t^{-n}$ are the between-class and total scatter matrices defined as the following:

$$\boldsymbol{S}_b^{-n} = \sum_{c=1}^{C} \left\langle \check{\mathscr{X}}^{-n}_{(c)}, \check{\mathscr{X}}^{-n}_{(c)} \right\rangle_{-n} = \left\langle \check{\mathscr{X}}^{-n}, \check{\mathscr{X}}^{-n} \right\rangle_{-n} \tag{A.3}$$

where

$$\check{\mathscr{X}}^{(c)} = \sqrt{K_c}\left(\bar{\mathscr{X}}^{(c)} - \bar{\bar{\mathscr{X}}}\right), \quad \check{\mathscr{X}} = cat\left(N+1, \check{\mathscr{X}}^{(1)}, \cdots, \check{\mathscr{X}}^{(C)}\right) \tag{A.4}$$

$$\check{\mathscr{X}}^{-n}_{(c)} = \check{\mathscr{X}}^{(c)} \times_{-n} \{\boldsymbol{U}^T\}, \quad \check{\mathscr{X}}^{-n} = \check{\mathscr{X}} \times_{-(n,N+1)} \{\boldsymbol{U}^T\} \tag{A.5}$$

and

$$\bar{\bar{\mathscr{X}}} = \frac{1}{K}\sum_{k=1}^{K} \mathscr{X}^{(k)} \tag{A.6}$$

$$\bar{\mathscr{X}}^{(c)} = \frac{1}{K_c}\sum_{k\in\Omega_c} \mathscr{X}^{(k)}, \quad c = 1, \cdots, C \tag{A.7}$$

Additionally:

$$\boldsymbol{S}_t^{-n} = \sum_{k=1}^{K} \left\langle \tilde{\mathscr{X}}^{-n}_{(k)}, \tilde{\mathscr{X}}^{-n}_{(k)} \right\rangle_{-n} = \left\langle \tilde{\mathscr{X}}^{-n}, \tilde{\mathscr{X}}^{-n} \right\rangle_{-n} \tag{A.8}$$

where

$$\tilde{\mathscr{X}}^{(k)} = \mathscr{X}^{(k)} - \bar{\bar{\mathscr{X}}}, \quad \tilde{\mathscr{X}} = cat\left(N+1, \tilde{\mathscr{X}}^{(1)}, \cdots, \tilde{\mathscr{X}}^{(K)}\right) \tag{A.9}$$

$$\tilde{\mathscr{X}}^{-n}_{(k)} = \tilde{\mathscr{X}}^{(k)} \times_{-n} \{\boldsymbol{U}^T\}, \quad \tilde{\mathscr{X}}^{-n} = \tilde{\mathscr{X}} \times_{-(n,N+1)} \{\boldsymbol{U}^T\} \tag{A.10}$$

Shrinkage Higher Order Discriminant Analysis (sHODA). In order to prevent the ill-conditioned behaviour of the within-class scatter matrix (or here total scatter matrix) in the generalized eigenvalue decomposition, Phan et. al. [54] suggested that it should be regularized with a *constant* regularization parameter, i.e. in each iteration replace $\mathbf{S}_t^{-n}$ by $\mathbf{S}_t^{-n} + \gamma\mathbf{I}$, where $\mathbf{I}$ is the identity matrix. This regularization parameter should be optimized in a cross-validation scheme. In many analysis pipelines, like OSTDA in this paper, there are other parameters and the regularization parameter of HODA adds up to them, making the parameter tuning harder.

To overcome this problem, we propose to compute the shrinkage data-driven and automatically in each iteration of the ALS algorithm. We do so using the idea of shrinkage in [55] and implement it based on the scatter matrices of n-th iteration of the ALS algorithm as follows:

$$\gamma^{(n)} = f\left(\frac{M}{(M-1)^2}\frac{\sum_{i,j} var(\tilde{\mathbf{Z}}^{-n}_{(n)}(i,:) \odot \tilde{\mathbf{Z}}^{-n}_{(n)}(j,:))}{||\mathbf{S}_t^{-n} - v\mathbf{I}||^2}\right) \tag{A.11}$$

where $\odot$ is the Hadamard product, $v = \frac{1}{I_n} trace(\mathbf{S}_t^{-n})$, and f is defined as the following function:

$$f(x) = \begin{cases} 0, & x \leqslant 0 \\ x, & 0 < x < 1 \\ 1, & x \geqslant 1 \end{cases}$$

Therefore, the optimization problem of each iteration of the ALS algorithm of sHODA changes from Eq. (A.2) to the following:

$$\boldsymbol{U}^{(n)} = \underset{\boldsymbol{U}^{(n)}}{argmax} \frac{tr[\boldsymbol{U}^{(n)^T}\boldsymbol{S}_b^{-n}\boldsymbol{U}^{(n)}]}{tr[\boldsymbol{U}^{(n)^T}(\boldsymbol{S}_t^{-n}+\gamma^{(n)}\mathbf{I})\boldsymbol{U}^{(n)}]}, \quad s.t. \quad \boldsymbol{U}^{(n)T}\boldsymbol{U}^{(n)} = \boldsymbol{I} \tag{A.12}$$

The main advantage of sHODA over regularized HODA is that regularization is performed analytically, thus eliminating the need of performing cross-validation to optimize the regularization parameter.

Appendix B. ANOVA Tables

Tables B.1 and B.2.

Table B.1
Three-way repeated measures ANOVA for individual accuracies in testing dataset with 9-channel setting, with *approach*, *number of training trials* per class and *perturbation* factors.

Source	SS	df	MS	F	p
Approach	11241.2	3	3747.08	17.140	<0.001
Error(Approach)	15084.5	69	218.62	–	–
Number trials	2644.0	1	2644.04	22.914	<0.001
Error(Number trials)	2654.0	23	115.39	–	–
Perturbation	2617.7	3	872.58	7.121	<0.001
Error(Perturbation)	8455.5	69	122.54	–	–
Approach x Number trials	636.4	3	212.14	2.931	0.0887
Error(Approach x Number trials)	4994.2	69	72.38	–	–
Approach x Perturbation	510.5	9	56.72	3.254	<0.01
Error(Approach x Perturbation)	3608.2	207	17.43	–	–
Number trials x Perturbation	28.9	3	9.63	1.329	0.2724
Error(Number trials x Perturbation)	499.7	69	7.24	–	–
Approach x Number trials x Perturbation	29.8	9	3.32	0.489	0.8204
Error(Approach x Number trials x Perturbation)	1403.4	207	6.78	–	–

Table B.2
Three-way repeated measures ANOVA for individual accuracies in testing dataset with 60-channel setting, with *approach*, *number of training trials* per class and *perturbation* factors.

Source	SS	df	MS	F	p
Approach	317445.4	3	105815.13	289.756	<0.001
Error(Approach)	25197.9	69	365.19	–	–
Number trials	12033.3	1	12033.33	82.834	<0.001
Error(Number trials)	3341.2	23	145.27	–	–
Perturbation	1655.0	3	551.65	5.866	<0.01
Error(Perturbation)	6489.1	69	94.04	–	–
Approach x Number trials	2628.6	3	876.20	6.394	<0.01
Error(Approach x Number trials)	9455.6	69	137.04	–	–
Approach x Perturbation	805.9	9	89.54	2.683	<0.05
Error(Approach x Perturbation)	6908.8	207	33.38	–	–
Number trials x Perturbation	3.0	3	1.00	0.089	0.9563
Error(Number trials x Perturbation)	776.4	69	11.25	–	–
Approach x Number trials x Perturbation	89.7	9	9.97	0.691	0.6512
Error(Approach x Number trials x Perturbation)	2983.1	207	14.41	–	–

References

[1] J.R. Wolpaw, N. Birbaumer, D.J. McFarland, G. Pfurtscheller, T.M. Vaughan, Brain–computer interfaces for communication and control, Clin. Neurophysiol. 113 (6) (2002) 767–791.
[2] J.J. Daly, J.R. Wolpaw, Brain–computer interfaces in neurological rehabilitation, Lancet Neurol. 7 (11) (2008) 1032–1043.
[3] B.H. Dobkin, Brain–computer interface technology as a tool to augment plasticity and outcomes for neurological rehabilitation, J. Physiol. 579 (3) (2007) 637–642.
[4] C. Vidaurre, J. Pascual, A. Ramos-Murguialday, R. Lorenz, B. Blankertz, N. Birbaumer, K.-R. Müller, Neuromuscular electrical stimulation induced brain patterns to decode motor imagery, Clin. Neurophysiol. 124 (9) (2013) 1824–1834.
[5] C. Vidaurre, C. Klauer, T. Schauer, A. Ramos-Murguialday, K.-R. Mueller, Eeg-based bci for the linear control of an upper-limb neuroprosthesis, Med. Eng. Phys. 38 (2016) 1195–1204.
[6] C. Vidaurre, A.R. Murguialday, S. Haufe, M. Gómez, K.-R. Müller, V. Nikulin, Enhancing sensorimotor bci performance with assistive afferent activity: An online evaluation, NeuroImage 199 (2019) 375–386.
[7] N. Birbaumer, Breaking the silence: brain–computer interfaces (bci) for communication and motor control, Psychophysiology 43 (6) (2006) 517–532.
[8] X. Gao, D. Xu, M. Cheng, S. Gao, A bci-based environmental controller for the motion-disabled, IEEE Trans. Neural Syst. Rehab. Eng. 11 (2) (2003) 137–140.
[9] R. Ortner, B.Z. Allison, G. Korisek, H. Gaggl, G. Pfurtscheller, An ssvep bci to control a hand orthosis for persons with tetraplegia, IEEE Trans. Neural Syst. Rehab. Eng. 19 (1) (2010) 1–5.
[10] G. Vecchiato, L. Astolfi, F.D.V. Fallani, S. Salinari, F. Cincotti, F. Aloise, D. Mattia, M.G. Marciani, L. Bianchi, R. Soranzo, et al., The study of brain activity during the observation of commercial advertsing by using high resolution eeg techniques, 2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE 2009 (2009) 57–60.
[11] M. Ahn, M. Lee, J. Choi, S. Jun, A review of brain-computer interface games and an opinion survey from researchers, developers and users, Sensors 14 (8) (2014) 14601–14633.
[12] G. Pfurtscheller, C. Brunner, A. Schlögl, F.L. Da Silva, Mu rhythm (de) synchronization and eeg single-trial classification of different motor imagery tasks, NeuroImage 31 (1) (2006) 153–159.
[13] A. Schlögl, F. Lee, H. Bischof, G. Pfurtscheller, Characterization of four-class motor imagery eeg data for the bci-competition 2005, J. Neural Eng. 2 (4) (2005) L14.
[14] H.-J. Hwang, K. Kwon, C.-H. Im, Neurofeedback-based motor imagery training for brain–computer interface (bci), J. Neurosci. Methods 179 (1) (2009) 150–156.
[15] C. Park, D. Looney, N. ur Rehman, A. Ahrabian, D.P. Mandic, Classification of motor imagery bci using multivariate empirical mode decomposition, IEEE Trans. Neural Syst. Rehab. Eng. 21(1) (2012) 10–22..
[16] R. Ortner, D.-C. Irimia, J. Scharinger, C. Guger, A motor imagery based brain-computer interface for stroke rehabilitation, Annu. Rev. Cybertherapy Telemedicine 181 (2012) 319–323.
[17] C. Sannelli, C. Vidaurre, K. Müller, B. Blankertz, A large scale screening study with a smr-based bci: Categorization of bci users and differences in their smr activity, PLOS ONE 14 (2019) e0207351.
[18] T. Nierhaus, C. Vidaurre, C. Sannelli, K.-R. Mueller, A. Villringer, Immediate brain plasticity after one hour of brain-computer interface (bci), J. Physiol. doi:10.1113/jp278118..
[19] Y. Zhang, Q. Zhao, J. Jin, X. Wang, A. Cichocki, A novel bci based on erp components sensitive to configural processing of human faces, J. Neural Eng. 9 (2) (2012) 026018.
[20] J. Jin, I. Daly, Y. Zhang, X. Wang, A. Cichocki, An optimized erp brain–computer interface based on facial expression changes, J. Neural Eng. 11 (3) (2014) 036004.
[21] M. Cheng, X. Gao, S. Gao, D. Xu, Design and implementation of a brain-computer interface with high transfer rates, IEEE Trans. Biomed. Eng. 49 (10) (2002) 1181–1186.
[22] M.M. Movahedi, A. Mehdizadeh, A. Alipour, Development of a brain computer interface (bci) speller system based on ssvep signals, J. Biomed. Phys. Eng. 3 (3 Sep)..
[23] Y. Wang, X. Gao, B. Hong, C. Jia, S. Gao, Brain-computer interfaces based on visual evoked potentials, IEEE Eng. Med. Biol. Mag. 27 (5) (2008) 64–71.
[24] C. Jia, H. Xu, B. Hong, X. Gao, Z. Zhang, S. Gao, A human computer interface using ssvep-based bci technology, in, International Conference on Foundations of Augmented Cognition, Springer (2007) 113–119.
[25] Y. Wang, R. Wang, X. Gao, B. Hong, S. Gao, A practical vep-based brain-computer interface, IEEE Trans. Neural Syst. Rehab. Eng. 14 (2) (2006) 234–240.
[26] R. Srinivasan, F.A. Bibi, P.L. Nunez, Steady-state visual evoked potentials: distributed local sources and wave-like dynamics are sensitive to flicker frequency, Brain topography 18 (3) (2006) 167–187.
[27] O. Friman, I. Volosyak, A. Graser, Multiple channel detection of steady-state visual evoked potentials for brain-computer interfaces, IEEE Trans. Biomed. Eng. 54 (4) (2007) 742–750.
[28] X. Chen, Y. Wang, M. Nakanishi, X. Gao, T.-P. Jung, S. Gao, High-speed spelling with a noninvasive brain–computer interface, Proc. Natl. Acad. Sci. 112 (44) (2015) E6058–E6067.
[29] M. Nakanishi, Y. Wang, Y.-T. Wang, Y. Mitsukura, T.-P. Jung, A high-speed brain speller using steady-state visual evoked potentials, Int. J. Neural Syst. 24 (06) (2014) 1450019.
[30] G.R. Müller-Putz, R. Scherer, C. Brauneis, G. Pfurtscheller, Steady-state visual evoked potential (ssvep)-based communication: impact of harmonic frequency components, J. Neural Eng. 2 (4) (2005) 123.
[31] P.F. Diez, V.A. Mut, E.M.A. Perona, E.L. Leber, Asynchronous bci control using high-frequency ssvep, J. Neuroeng. Rehab. 8 (1) (2011) 39.
[32] P.F. Diez, S.M.T. Müller, V.A. Mut, E. Laciar, E. Avila, T.F. Bastos-Filho, M. Sarcinelli-Filho, Commanding a robotic wheelchair with a high-frequency steady-state visual evoked potential based brain–computer interface, Med. Eng. Phys. 35 (8) (2013) 1155–1164.
[33] A.P. Liavas, G.V. Moustakides, G. Henning, E.Z. Psarakis, P. Husar, A periodogram-based method for the detection of steady-state visually evoked potentials, IEEE Trans. Biomed. Eng. 45 (2) (1998) 242–248.
[34] M. Middendorf, G. McMillan, G. Calhoun, K.S. Jones, Brain-computer interfaces based on the steady-state visual-evoked response, IEEE Trans. Rehab. Eng. 8 (2) (2000) 211–214.
[35] E.C. Lalor, S.P. Kelly, C. Finucane, R. Burke, R. Smith, R.B. Reilly, G. Mcdarby, Steady-state vep-based brain-computer interface control in an immersive 3d gaming environment, EURASIP J. Adv. Signal Process. 2005 (19) (2005) 706906.
[36] J. Castillo-Garcia, S. Müller, E. Caicedo, A. Cotrina, T. Bastos, Comparison among feature extraction techniques based on power spectrum for a ssvep-bci, in: 2014 12th IEEE International Conference on Industrial Informatics (INDIN) IEEE, 2014, pp. 284–288.
[37] L. Acqualagna, S. Bosse, A.K. Porbadnigk, G. Curio, K.-R. Müller, T. Wiegand, B. Blankertz, Eeg-based classification of video quality perception using steady state visual evoked potentials (ssveps), J. Neural Eng. 12 (2) (2015) 026012.
[38] I. Volosyak, D. Valbuena, T. Luth, A. Gräser, Towards an ssvep based bci with high itr, IEEE Trans. Biomed. Eng..
[39] W. Nan, C.M. Wong, B. Wang, F. Wan, P.U. Mak, P.I. Mak, M.I. Vai, A comparison of minimum energy combination and canonical correlation analysis for ssvep detection, in: 2011 5th International IEEE/EMBS Conference on Neural Engineering IEEE, 2011, pp. 469–472.
[40] H. Wang, Y. Zhang, N.R. Waytowich, D.J. Krusienski, G. Zhou, J. Jin, X. Wang, A. Cichocki, Discriminative feature extraction via multivariate linear regression for ssvep-based bci, IEEE Trans. Neural Syst. Rehabil. Eng. 24 (5) (2016) 532–541.
[41] T. Jorajuría, M. Gómez, C. Vidaurre, A fast ssvep-based brain-computer interface, in: E.A. de la Cal, J.R. Villar Flecha, H. Quintián, E. Corchado (Eds.),

Hybrid Artificial Intelligent Systems, Springer International Publishing, Cham, 2020, pp. 49–60.
[42] Z. Lin, C. Zhang, W. Wu, X. Gao, Frequency recognition based on canonical correlation analysis for ssvep-based bcis, IEEE Trans. Biomed. Eng. 53 (12) (2006) 2610–2614.
[43] G. Bin, X. Gao, Z. Yan, B. Hong, S. Gao, An online multi-channel ssvep-based brain–computer interface using a canonical correlation analysis method, J. Neural Eng. 6 (4) (2009), 046002.
[44] E. Yin, Z. Zhou, J. Jiang, Y. Yu, D. Hu, A dynamically optimized ssvep brain–computer interface (bci) speller, IEEE Trans. Biomed. Eng. 62 (6) (2014) 1447–1456.
[45] L. Cao, Z. Ju, J. Li, R. Jian, C. Jiang, Sequence detection analysis based on canonical correlation for steady-state visual evoked potential brain computer interfaces, J. Neurosci. Methods 253 (2015) 10–17.
[46] Y. Zhang, G. Zhou, Q. Zhao, A. Onishi, J. Jin, X. Wang, A. Cichocki, Multiway canonical correlation analysis for frequency components recognition in ssvep-based bcis, in: International Conference on Neural Information Processing, Springer, 2011, pp. 287–295..
[47] Y. Zhang, G. Zhou, J. Jin, X. Wang, A. Cichocki, Frequency recognition in ssvep-based bci using multiset canonical correlation analysis, Int. J. Neural Syst. 24 (04) (2014) 1450013.
[48] Y. Zhang, G. Zhou, J. Jin, M. Wang, X. Wang, A. Cichocki, L1-regularized multiway canonical correlation analysis for ssvep-based bci, IEEE Trans. Neural Syst. Rehab. Eng. 21 (6) (2013) 887–896.
[49] M. Farooq, O. Dehzangi, High accuracy wearable ssvep detection using feature profiling and dimensionality reduction, in: 2017 IEEE 14th International Conference on Wearable and Implantable Body Sensor Networks (BSN) IEEE, 2017, pp. 161–164.
[50] Z. İşcan, V.V. Nikulin, Steady state visual evoked potential (ssvep) based brain-computer interface (bci) performance under different perturbations, PloS one 13 (1) (2018) e0191673.
[51] X. Xing, Y. Wang, W. Pei, X. Guo, Z. Liu, F. Wang, G. Ming, H. Zhao, Q. Gui, H. Chen, A high-speed ssvep-based bci using dry eeg electrodes, Sci. Rep. 8 (1) (2018) 14708.
[52] H. Tanaka, T. Katura, H. Sato, Task-related component analysis for functional neuroimaging and application to near-infrared spectroscopy data, NeuroImage 64 (2013) 308–327.
[53] Y. Wang, M. Nakanishi, T. Wang, T. Jung, Enhancing detection of steady-state visual evoked potentials using individual training data, Conf. Proc. IEEE Eng. Med. Biol. Soc. (2014, 2014,) 3037–3040.
[54] A. Phan, A. Cichocki, Tensor decompositions for feature extraction and classification of high dimensional datasets, Nonlinear Theory Appl., IEICE 1 (1) (2010) 37–68.
[55] B. Blankertz, S. Lemm, M. Treder, S. Haufe, K.-R. Müller, Single-trial analysis and classification of erp components–a tutorial, NeuroImage 56 (2) (2011) 814–825.
[56] V.V. Nikulin, G. Nolte, G. Curio, A novel method for reliable and fast extraction of neuronal eeg/meg oscillations on the basis of spatio-spectral decomposition, NeuroImage 55 (4) (2011) 1528–1535.
[57] S. Yan, D. Xu, Q. Yang, L. Zhang, X. Tang, H.-J. Zhang, Discriminant analysis with tensor representation, in: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), Vol. 1, IEEE, 2005, pp. 526–532..
[58] K.-R. Muller, C.W. Anderson, G.E. Birch, Linear and nonlinear methods for brain-computer interfaces, IEEE Trans. Neural Syst. Rehab. Eng. 11 (2) (2003) 165–169.
[59] A. Cichocki, Tensor decompositions: a new concept in brain data analysis?, arXiv preprint arXiv:1305.0395..
[60] M.J. Idaji, M.B. Shamsollahi, S.H. Sardouie, Higher order spectral regression discriminant analysis (hosrda): A tensor feature reduction method for erp detection, Pattern Recogn. 70 (2017) 152–162.
[61] A.H. Phan, Nfea: Tensor toolbox for feature extraction and applications, Lab for Advanced Brain Signal Processing, BSI, RIKEN, Tech. Rep..
[62] O. Ledoit, M. Wolf, Honey, i shrunk the sample covariance matrix, J. Portfolio Manage. 30 (4) (2004) 110–119.
[63] C. Vidaurre, N. Krämer, B. Blankertz, A. Schlögl, Time domain parameters as a feature for eeg-based brain–computer interfaces, Neural Networks 22 (9) (2009) 1313–1319.
[64] C. Sannelli, C. Vidaurre, K.-R. Mller, B. Blankertz, CSP patches: an ensemble of optimized spatial filters. an evaluation study, J. Neural Eng. 8 (2) (2011) 025012.
[65] C. Sannelli, C. Vidaurre, K.-R. Mller, B. Blankertz, Ensembles of adaptive spatial filters increase BCI performance: an online evaluation, J. Neural Eng. 13 (4) (2016) 046003.
[66] T. Verhoeven, D. Hbner, M. Tangermann, K.R. Mller, J. Dambre, P.J. Kindermans, Improving zero-training brain-computer interfaces by mixing model estimators, J. Neural Eng. 14 (3) (2017) 036021.
[67] X. Chen, Y. Wang, S. Gao, T.-P. Jung, X. Gao, Filter bank canonical correlation analysis for implementing a high-speed ssvep-based brain–computer interface, J. Neural Eng. 12 (4) (2015) 046008.
[68] S. Haufe, S. Dähne, V.V. Nikulin, Dimensionality reduction for the analysis of brain oscillations, NeuroImage 101 (2014) 583–597.
[69] C. Beleites, U. Neugebauer, T. Bocklitz, C. Krafft, J. Popp, Sample size planning for classification models, Analytica chimica acta 760 (2013) 25–33.
[70] A. Delorme, S. Makeig, Eeglab: an open source toolbox for analysis of single-trial eeg dynamics including independent component analysis, J. Neurosci. Methods 134 (1) (2004) 9–21.
[71] Y. Zhang, G. Zhou, Q. Zhao, J. Jin, X. Wang, A. Cichocki, Spatial-temporal discriminant analysis for erp-based brain-computer interface, IEEE Trans. Neural Syst. Rehabil. Eng. 21 (2) (2013) 233–243, https://doi.org/10.1109/TNSRE.2013.2243471.
[72] K. Ma, S. Wang, S. Zhang, Y. Sun, D. Zheng, Electrode channel optimisation method for steady-state visual evoked potentials, J. Eng. 2019 (23) (2019) 8632–8636.
[73] S. Wold, A. Ruhe, H. Wold, W. Dunn, Iii, The collinearity problem in linear regression. the partial least squares (pls) approach to generalized inverses, SIAM J. Sci. Stat. Comput. 5 (3) (1984) 735–743.

Tania Jorajuría received her MSc. in Telecommunication Engineering from Public University of Navarre (Upna), Spain, in 2015. Since then, she has worked as a software developer in different companies, being nowadays employed at the Industry Association of Navarre as data scientist. Since 2017 she is also pursuing her doctoral thesis in brain-computer interfacing based on electroencephalographic signals, at Upna. In 2019 she was guest researcher at the Max Planck Institute for Human Cognitive and Brain Sciences, in Leipzig, Germany. Her research interests include neuroscience, brain-computer interfaces, data processing and analysis, and biomedical engineering.

Mina Jamshidi Idaji received her B.Sc. in electrical engineering and pure mathematics from Isfahan University of Technology, Isfahan, Iran in 2014, and her M.Sc. degree in biomedical engineering from Sharif University of Technology, Tehran, Iran in 2016. She is currently a doctoral researcher at the Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany. Her research interests include signal and image processing, particularly with biomedical applications, brain-computer interface, multilinear (tensor) algebra, and network science, especially network neuroscience.

Zafer İşcan received his B.Sc. in Electronics & Communication Engineering, M.Sc. in Biomedical Engineering, and Ph.D. in Electronics Engineering from Istanbul Technical University in 2002, 2005, and 2012, respectively. Afterwards, he worked as a postdoctoral researcher in the U.S. (Stony Brook University), Russia (Higher School of Economics), and France (Inserm). Since 2018, he has been working as an assistant professor in the Department of Electrical and Electronics Engineering at Bahçesehir University. His main research areas are neuroscience, brain computer interfaces, pattern recognition, machine learning, artificial neural networks, and image processing.

Marisol Gómez holds a bachelor's degree in Mathematics from Salamanca University, and a PhD in Mathematics and an expert title in Big Data from the Public University of Navarre. She is associate professor in the Algebra area at this university. Professor Gómez is co-author of more than 60 publications in magazines and international conferences. She has participated in 20 research projects funded by public bodies. She has co-directed 4 doctoral theses. She is also a founding partner of Movalsys S.L., a technology-based company of Upna. Her research interests include the applications of Algebra techniques in the analysis of biomedical signals.

Vadim Nikulin (PhD) is a principal investigator in Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany. He is also a visiting professor in the Institute for Cognitive Neuroscience, HSE, Moscow, Russia. His interests include investigation of complex neuronal dynamics unfolding at different temporal and spatial scales as well as investigation of interactions between local and global neuronal processes. He is particularly interested in investigation of non-invasive neuronal signatures of neurological (stroke, Parkinson's disease) and psychiatric disorders (Schizophrenia, Depression,...). He also develops novel techniques for the investigation of multidimensional neuronal data with applications for brain neuromodulation and BCI.

Carmen Vidaurre studied Telecommunication Engineering and received a PhD in Engineering at the Public University of Navarre. She was then awarded a Marie Curie IE Fellowship for the Fraunhofer Institute (Berlin). Afterwards, she held a position as post doc researcher at the Technical University of Berlin. She is currently a Ramon y Cajal fellow at the Public University of Navarre. Her current research is about the development and application of methods to understand the central and peripheral nervous system.