Tensor-Based Classification of Auditory Mobile BCI without Subject-Specific Calibration Phase

Rob Zink[1,2], **Borbála Hunyadi**[1,2], **Sabine Van Huffel**[1,2], **Maarten De Vos**[3]

[1]KU Leuven, Department of Electrical Engineering (ESAT), STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics, Kasteelpark Arenberg 10, 3001 Heverlee, Belgium and iMinds Medical IT.
[2]iMinds Medical IT, Leuven, Belgium.
[3]Engineering Department, Oxford University, Oxford, United Kingdom and Cluster of Excellence Hearing4all, University of Oldenburg, Germany.

E-mail: rob.zink@esat.kuleuven.be

Abstract. **Objective** One of the major drawbacks in EEG Brain Computer Interfaces (BCI) is the need for subject-specific training of the classifier. By removing the need for a supervised calibration phase, new users could potentially explore a BCI faster. In this work we aim to remove this subject-specific calibration phase and allow direct classification. **Approach** We explore Canonical Polyadic Decompositions (CPD) and Block Term Decompositions (BTD) of the EEG. These methods exploit structure in higher dimensional data arrays called tensors. The BCI tensors are constructed by concatenating ERP templates from other subjects to a target and non-target trial and the inherent structure guides a decomposition that allows accurate classification. We illustrate the new method on data from a three-class auditory oddball paradigm. **Main results** The presented approach leads to a fast and intuitive classification with accuracies competitive with a supervised and cross-validated LDA approach. **Significance** The described methods are a promising new way of classifying BCI data with a forthright link to the original P300 ERP signal over the conventional and widely used supervised approaches.

Keywords: Mobile EEG, Brain-Computer-Interface, Subject-Specific Calibration-free, Tensor Decompositions, auditory P300.

1. Introduction

Research in the field of Brain-Computer-Interfaces (BCI) has made significant progress in methodologies and applications. To date, the use of non-invasive EEG (electroencephalography) signals as primary input is most often used for such interfaces. EEG-based BCIs exploit often the P300, generated in response to rare and task-relevant stimuli (e.g. [1, 2, 3, 4]). Although BCIs were originally developed with the intent of providing accessibility to computers for locked-in patients [5], they have also potential in applications for healthy users [6, 7]. However, one of the biggest hurdles for wider application is the feasibility of fast practical application in real life situations. An increasing number of studies show that mobile EEG applications can be deployed in natural, real life situations [8, 9] with comparable accuracy as using traditional EEG systems [10]. These studies illustrate that in outdoor sitting and walking circumstances it is possible to use an auditory BCI without explicit control of environmental variables. This comes close to a practical application of auditory BCI in real-life as envisioned in the literature [11, 12].

Besides the limited validation in real-life scenarios, a second hurdle is that most BCIs exploit supervised classification methods. This requires a separate training phase to calibrate the classifier function [13]. Essentially a substantial part of the data is discarded for model training rather than interacting with the BCI, which comes at the cost of consuming users time and effort. It is not uncommon that half of the experimental time is devoted to recording training data (e.g. [9]). The use of calibration-free classifiers could increase the application potential significantly by removing this specific training phase. Few studies have focused on developing such out-of-the box classifiers. Incipiently, an unsupervised approach for the P300 speller was presented offline in [14] relying on Bayesian statistics and exploiting constraints imposed by the BCI stimulation setup. The utilization of prior distributions of Target and Non-Target stimuli and repetitions was shown to lead to a fast adapting calibration free classifier. The same authors reported high performance in [15] extending the Bayesian approach with transfer learning techniques and re-learning of the classifiers by reason of language models. This approach was also successfully translated to an online scenario [16]. A novel classification approach for motor imagery BCIs that has gained momentum in the past two years exploits the spatial covariance matrices of EEG signals and rely on Riemannian geometry to obtain accurate classification [17]. Successively, modifications of this approach were presented in [18, 19] to extend the framework for calibration free ERP (Event Related Potential) analysis. This allowed instantaneous P300 classification based on a special form of space-time covariance matrices and a Riemannian distance-to-class-average as classifier function. Another pertinent approach aimed to incorporate information transfer between BCI sessions in the EEG by transferring session-specific changes between experimental conditions to new subjects [20]. The use of session specific changes in the EEG was shown to convey decisive information for classifier performance. Ultimately the aforementioned techniques allow to track classifier performance across subjects and sessions and grant instantaneous classification. All these approaches are of particular interest for bringing BCI applications into practical use.

In the current work, we present a novel subject-specific calibration-free method based on tensor decompositions of the EEG applied to auditory P300 data (i.e. on data obtained from [9]). Since ERP data can be naturally represented as a channels $\times$ time $\times$ trials tensor, it can be advantageous to exploit such multidimensional structure in the analysis. Tensor-based methods such as the Canonical-Polyadic-Decomposition (CPD) have been used successfully in several biomedical applications for feature extraction or classification in a supervised approach (e.g.[21, 22, 23, 24, 25, 26]). However, the data-driven nature of these methods lends these methods also to be applied for unsupervised classification [27]. Block Term Decomposition (BTD) was recently demonstrated to be superior over CPD for certain EEG applications as it allows to model more variability [28]. We will explore both CPD and BTD for unsupervised BCI classification.

By constructing a tensor from the trials we want to classify, and relying on structural properties in the data, we demonstrate that CPD and BTD are able to provide accurate labels. The method elegantly exploits the presence of a particular spatiotemporal pattern underlying the target trials which is absent in the non-targets. We add non-specific ERP templates to the two deviant stimuli epochs to obtain a tensor and a subsequent decomposition identifies discriminative features between target and non-target trials. The decompositions create a uniquely adapted (i.e. from the templates) spatiotemporal pattern to separate target and non-target trials that can vary over time and sessions. In this way we achieve classification results without the need for a subject-specific calibration phase.

In order to demonstrate the potential of the new method, we compare the results to regularized LDA (rLDA), as this has been shown to be amongst the most effective algorithms for P300 classification [29]. As a cross-subject trained classifier would also

be able to derive instantaneous classification of a new subject (in contrast to the subject-specific LDA-trained classifiers), we also compare our results to a cross-subject trained rLDA classifier [19].

The paper is structured as follows: Section II summarizes the most important steps in the data acquisition procedure from [9], followed by a description of the data-driven classification methods with the CPD/BTD and how the different LDA classifiers are constructed. In Section III the results are presented, followed by a discussion and future perspective (Section IV).

2. Methods

2.1. Participants, Stimuli and Data Acquisition

The most important facts of the auditory attention dataset are summarized here, for more detail we refer to [9]. The used dataset consists of 20 healthy subjects (10 females) with a mean age of 24.6 years. The paradigm was a three-class auditory oddball task, where the participants were requested to count the target tones while ignoring the other two tones. The standard tone was 900Hz while the two deviant tones were 600Hz and 1200Hz respectively. Ten subjects were instructed to pay attention to the 1200 Hz tone and the other ten to the 600 Hz tone. The stimuli were presented randomly and binaurally through headphones. The mean inter stimulus interval was 1000ms and incorporated an evenly distributed jitter in the range of 0ms to 375ms. Per session the subjects were exposed to 94 target deviants, 94 non-target deviants and 504 standard tones that were randomly presented. Each subject had to undergo tests in two different outdoor settings. One test was conducted while walking along a planned route at the Oldenburg campus. The other test was conducted in a silent corner of the university in which the volunteer and examiner were seated on two chairs. It is important to note that a significant amount of unwarranted distraction could have occurred during each pair of recordings due to regular activities taking place both in and off the campus. The stimulus delivery and experimental control was carried out using OpenViBE software running on a laptop [30].

The data acquisition was conducted using an original Emotiv amplifier (emotiv.com) connected to a modern infracerebral electrodecap comprising of 14 channels (easycap.de). Further details regarding the modified Emotiv EEG system are described in [8]. The signals of the amplifier were bandpass-filtered between 0.16Hz and 45Hz. and were sampled at 128Hz. Fourteen Ag/AgCl sintered electrodes were fixed at 10-20 positions: F3, Fz, C3, FPz, Tp10, Cz, O1, O2, F4, C4, TP9, Pz, P4 and P3. The common mode suppression for the electrodes (online reference) was at AFz and they were grounded at FCz.

2.2. Preprocessing

The data were preprocessed offline using EEGLAB [31] and MATLAB (Mathworks Inc., Natick, MA). Eye-blink artifacts were semi-automatically attenuated by means of extended infomax independent component analysis [32, 33]. EEG data were 20Hz low-pass filtered, and epochs were extracted from -200ms to 800ms with respect to stimulus onset and baseline-corrected (-200ms to 0ms) after re-referencing to the mean of TP9 and TP10. In order to reduce the complexity of the data, we down-sampled the data from 128Hz to 30Hz for this approach. This decreases computation time of the tensor models and is yet still expected to capture most of the P300 waveforms [27, 34]. Prior to the downsampling a 15Hz low-pass filter was applied to avoid alliasing artefacts.

2.3. Binary Classification

The aim of the classification process is to classify single trial pairs consisting of one 600Hz trial and one 1200Hz trial. It is to be expected that when subjects attend to one tone, these attended tones elicit a P300 oddball ERP pattern which is absent for the unattended and baseline tones. The subject's binary choice of focussing his/her attention on one of the two stimuli makes it logical to approach the analysis as classifying these trial pairs rather than individual stimuli. An example of an application could be to pre-code the stimuli to different hearing aid processing modes, allowing the user to switch between background processing or foreground processing by attending to the proper tone. The attended and unattended tones are referred to as "target" and "non-target" trials respectively. The paradigm's stimulation codes can be used to know when these sounds are presented and the selective attention of the subject causes one tone to elicit a P300 response. Therefore we approach the problem as distinguishing target and non-target trials in a set of 2 (i.e. trial-pair) rather than making a decision for individual trials as is traditionally done. This available paradigm information was not incorporated in [9].

2.3.1. Construction of the Data Tensor The interpretability of the components extracted from a tensor decomposition depend on the construction of the data tensor prior to the decomposition. A window of 167ms $-$ 633ms after stimulus onset (SO) is determined as the most discriminative window for the P300 signal. The windowed data is normalized by converting the time course to z-scores for every channel.

Tensor decomposition techniques are able to identify the most structured signature in a multidimensional data tensor. In order to enhance the likelyhood of extracting a task-related signature, an average target and baseline ERP are added to the trial pair data tensor. These ERP templates are the average ERP of all other subjects for the baseline and target stimuli. Moreover, this inclusion of templates simplifies the identification of the class of each trial from the decomposed factors; which would be otherwise a non-trivial task [27]. Since some subjects might sometimes incorrectly focus to the non-targets, the baseline stimuli provide a cleaner non-target average template as compared to averaging all non-target trials. Taken together, for each trial pair we obtain a 12 × 15 × 4 data tensor in which the last mode consists of two trials (Target, non-Target of which we do not know the identity) and two templates (Target, Baseline). All analysis steps regarding the construction of the data tensor are conceptually visualized in Figure 1.

2.3.2. CPD Multidimensional signals can be decomposed by the CPD (Canonical Polyadic Decomposition) as a sum of rank-1 terms [35]. For the three-dimensional case the CPD will decompose a tensor X as follows:

$$X = \sum_{r=1}^{R} a_r \circ b_r \circ c_r + \varepsilon \tag{1}$$

with R representing the number of components, a_r, b_r, and c_r the signatures of every atom in each of the modes, and ε the model error. Each mode has a specific signature which characterizes the extracted component; in the three-dimensional tensor representing the ERP as a channel × time × trials structure, the spatial distribution of the different atoms would be contained in a_r, the time courses would be contained in b_r, and a strength of the space-time signature across trials would be given in c_r. The Canonical Polyadic Decomposition model is trilinear, which means that each mode's vectors are proportional to each other within a rank-1 component. Generally if the data follows a rank R structure, the decomposition is unique up to permutation and scaling of the extracted components [36]. In this study, CPD was computed with the nonlinear least squares (NLS) algorithm in the publicly available Tensorlab toolbox [37].

2.3.3. BTD Although CPD provides interpretable components, the model can be too restrictive for some applications as it does not model all variability in the data [28]. A BTD (Block Tensor Decomposition) allows to model more variation in two factors in a so called rank (L_r, L_r, 1) BTD. The rank (L_r, L_r, 1) BTD approximates a third-order tensor by a sum of R terms, each of which is an outer product of a rank-L_r matrix and a nonzero vector [38, 39]. A three-dimensional data tensor X can be decomposed by a (L_r, L_r, 1) BTD as:

$$X = \sum_{r=1}^{R} (A_r \cdot B_r^T) \circ c_r + \varepsilon \tag{2}$$

The tensor is the sum of the outer products of a rank L_r matrix (the product of matrices A_r and B_r-transposed) and the component vector c_r, with R representing the number of components and ε again the model error. Similarly to the rank, L_r should be set a priori. We aim to model additional variance with the BTD as opposed to the CPD. However, the exact mixture of the data tensor is unknown. The choice of L_r is therefore rather conceptual, depending on a priori knowledge about the task and preprocessing. Since our data-tensors are rather small (i.e. 12 × 15 × 4), values of 2 to 5 for L_r were explored. This is expected to capture time and waveform variability more accurately between the target and non-target effects as constituted in the trials and templates. Similar to the CPD we utilize the NLS algorithm within the Tensorlab toolbox for the BTD. Both the CPD and BTD models are extracted to retrieve a single component (R=1) and are initialized randomly. Figure 2 illustrates the CPD and BTD models in a single data tensor. In the case of a CPD decomposition it can be seen that the same waveform is linearly scaled over the channels (i.e rank-1), while in case of a BTD solution the rank-2 (or higher) spatiotemporal pattern allows more variation on the waveform on the different channels. For an larger overview of tensor decompositions used in signal processing applications we kindly refer the reader to [26].

2.3.4. Interpretation of the Decomposition Figure 3 illustrates a CPD and BTD decomposition for a single trial pair constructed from real data. The modes depict the extracted spatial and temporal waveform together with the extracted weights in the third dimension. To classify a single trial pair we can utilize these weights as follows. A target trial is expected to differ most from the baseline template and differ the least from the target template; vice versa for a non-target trial. For every single-trial-pair augmented with the two templates a CPD and BTD model is obtained. The third factor represents the trial weights accredited to the templates and the 2 unknown trials. In order to classify which trial is the Target and non-Target trial we calculate per trial in the third factor the absolute difference between the trial weight and the Baseline template weight and from this we subtract the absolute difference of the trial weight and the Target template

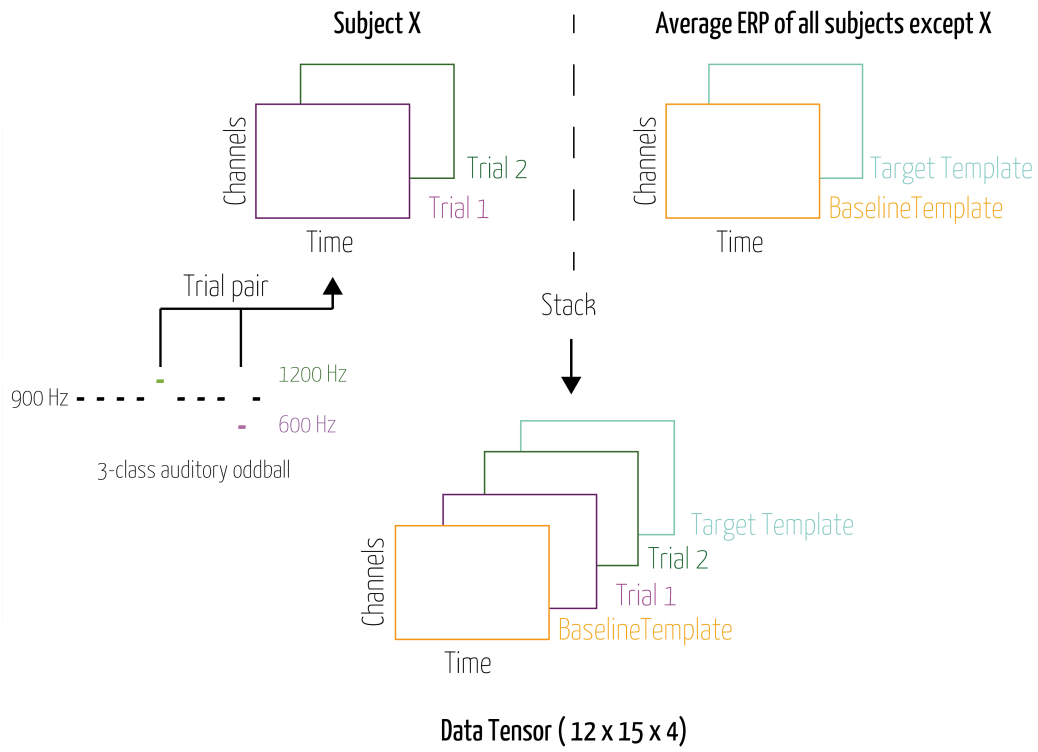

Figure 1. Overview of the tensor construction. Every target (attended tone) and non-target (unattended tone) trial are concatenated with average ERP templates from other subjects of the target and baseline stimuli. In the end per binary command we constructed a 12 channels x 15 time points x 4 trials/templates tensor.

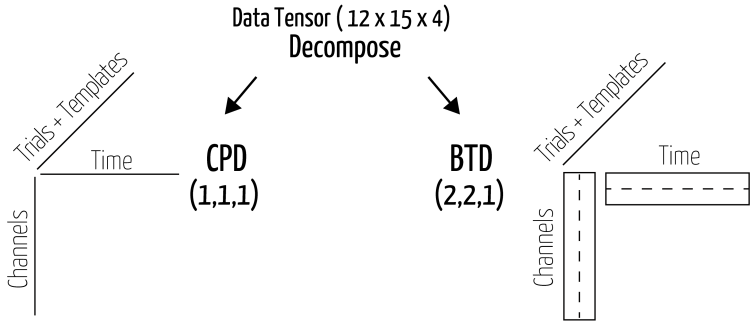

Figure 2. Overview of the tensor construction. Every target (attended tone) and non-target (unattended tone) trial are concatenated with average ERP templates from other subjects of the target and baseline stimuli. In the end per binary command we constructed a 12 channels x 15 time points x 4 trials/templates tensor.

weight. The trial with the largest obtained value is considered the Target trial, the smallest the non-Target.

Figure 3a illustrates the spatiotemporal pattern and trial mode of an extracted CPD component. In the trial mode the distance of the trials to the baseline and target template indicate that trial 2 is considered the target trial and 1 the non-target; the latter is closer to the baseline template (BT) weight and the former to the target template (TT) weight. However, the difference between Trial 2 and Trial 1 is marginal. Similarly we can decompose the data-tensor with BTD, for example a (2, 2, 1)-BTD to derive similar estimates and derive class labels from the third mode estimates. It should be noted that the individual spatial and temporal signatures in the BTD model are not unique in the extracted form [38, 39]; linear combinations of these solutions are equally plausible outcomes. In order to obtain the true unique spatiotemporal signature of the component we need to construct the combined spatiotemporal subspace (i.e. multiplying A with B^T in equation 1). The result is visualized in Figure 3b. In this matter the BTD models extract a more fine-tuned signature, namely a Parietal Occipital shift of activity. This results in a similar correct classification of the trial pair (Figure 3b), however the better model also enlarges the distinction between trial 1 and trial 2 (c.q. Figure 3a, b the third mode).

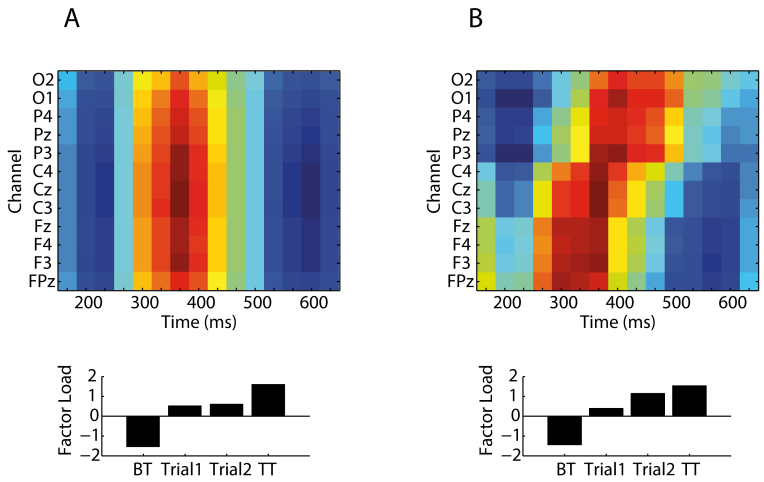

Figure 3. Examples of decomposing a single trial pair tensor with (a) CPD, Illustrating the factor loadings in the spatiotemporal matrix and trial/template modes from top to bottom respectively. (b) BTD with L=3,the obtained BTD spatiotemporal matrix is derived from multiplying the first two modes of the BTD estimates (not shown). The BTD model estimates 2 distinct sources whereas the CPD estimates an average target-non-target effect. The trial/template estimates in the last mode will lead to classifying Trial2 as the Target trial since its value is close to the Target Template (TT) and further away for the Baseline Template (BT) as opposed to Trial1. Note, in this particular example the subject focused on Trial2. The BTD model separates the target and non-target signature more accurately which is evident from a larger difference between Trial1 and Trial2.

2.3.5. Evaluating the CPD/BTD Classification accuracy is an important factor to evaluate the CPD/BTD classifiers. However, another important aspect is to understand subject-wise differences with the newly proposed method. Since the CPD/BTD classification process exploits the morphologic structure guided by the ERP templates and the P300 is known to vary in both amplitude and latency between subjects [40], we hypothesize that the results might depend on the subjects' P300 morphology. In order to investigate the influence of the added template ERPs, we define features that we correlate with the retrieved classification accuracies.

- Terp-Btemp = difference between: targetERP and the Baseline template.
- Terp-Ttemp = difference between: targetERP and the Target template.

The targetERP is defined as a matrix with the averages across all target trials per channel. The difference is calculated as the summed absolute difference between the targetERP and template matrix. It should be noted that these ERP features are derived offline after the classification process for evaluation purposes only.

2.4. Linear Discriminant Analysis

In order to compare the results from the newly proposed method, we also compute supervised classification accuracies with different LDA variants as described here. Similarly to the CPD/BTD method we classify single-trial-pairs allowing the classifier to discriminate between the target and non-target, rather than classify the individual responses. Traditionally this paradigm information is not included in the classifier function, therefore the LDA was adapted accordingly to allow for fair comparison between CPD/BTD and the LDA. The included 'pair' information does not incorporate knowledge about which class (Target-NonTarget) the trial belongs to and is therefore a valuable addition to the classification of three-class oddball data.

2.4.1. Offline Subject Dependent The basic LDA feature set comprised seventeen 47ms data bins on all twelve electrodes between 0−800ms. Shrinkage regularization as implemented in BCILAB [41, 42] is used for rLDA classification. Per subject the classifiers are trained based on five-fold cross-validation procedure.

2.4.2. Pseudo Online Subject Dependent One way to evaluate the performance of rLDA without long calibration is to (re)train the classifiers for every single trial pair, in which all past trials of that subject serve as training input for evaluating the next trial under investigation. The accuracies reported at every trial pair are calculated as the number of correctly classified trials from that moment in time, until the end (i.e. of the session) with a model that is trained on the history of trials up to that point. It is expected that the performance will increase in time as the number of training data increases. In contrast, the CPD and BTD are not expected to show such an increase as the method does not require an explicit subject-specific calibration phase.

2.4.3. LDA-based Cross-Subject Classification As it is also possible to construct an LDA classifier on data from other subjects, we compare the performance of the new method to rLDA trained on cross-subject data. The feature formation and selection steps are identical to the steps described in the previous paragraph, with the exception that now, the model is trained on data from other non-test subjects. We compute accuracies as a function of the potential number of training subjects available, from 1-19. For all divisions of these 19 different scenarios the LDAs were trained and tested on the remaining independent subject(s). In the case of training 1-18 subjects, the index of subjects chosen for training was selected at random from the 19 non-test subjects. For example, if we train on five subjects, many sets of five can be drawn out of the total 19 subjects that can be used for training. Therefore this process was repeated an arbitrarily chosen 25 times and the average accuracies were reported. As the CPD and BTD performance is also influenced by the morphology of the templates added to the model, these used template averages are in turn also varied from 1-19 subjects in exactly the same way as for the LDA cross-subject results.

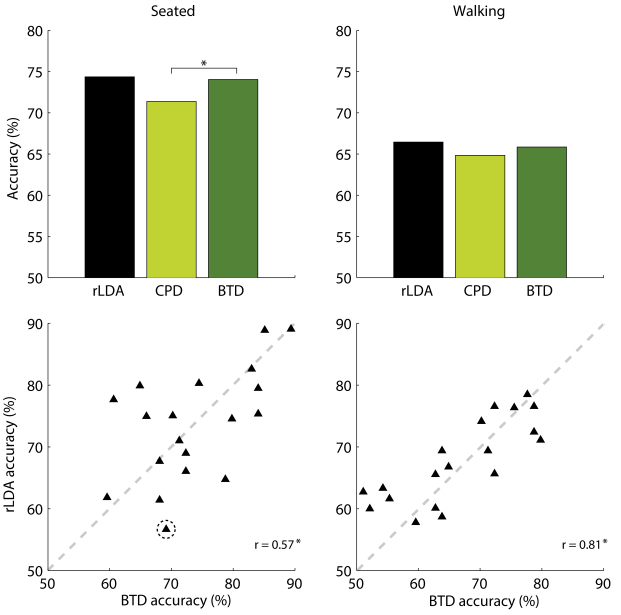

Figure 4. Top, grand average accuracies. Bottom, Relation between BTD and rLDA with Pearsons correlation coefficient. Respectively in the seated condition (left) and walking (right). Significant differences and correlations with p<0.05 are indicated by an asterisk.

3. Results

The presented (L_r,L_r,1) BTD results are obtained with L=2. However, for all considered values of L the BTD results did not differ significantly in any condition.

3.1. Offline Comparison

Average classification accuracies per subject are presented in Figure 4. The grand average accuracies (and SD) are74.4 (9.7), 71.7 (8.6) and 74.0% (8.5) in the seated condition for rLDA, CPD and BTD respectively. The BTD results were significantly higher compared to the CPD (t_{19} = 3.34, p<0.01) and the rLDA difference to CPD was not significant (t_{19} = 1.44, p=0.17). In the walking condition, grand average accuracies (and SD) were 66.4 (8.7) 64.8 (10.3) and 65.9% (9.9) for rLDA, CPD, BTD respectively. It should be noted that although the means are very similar across methods, differences in BTD and rLDA performance between subjects can be substantial (e.g. the subject indicated with a circle in Figure 4). When only including the Baseline Template, BTD accuracy dropped 1.1 and 1.6 percentage points for the seated and walking condition, respectively.

In both conditions the CPD and BTD estimates displayed a strong correlation in the seated (Pearsons r=0.93, p<0.001) and walking condition (r=0.97, p<0.001). The correlation between BTD and rLDA is moderate for the seated condition (Pearsons r=0.58, p<0.05) and strong in the walking condition (Pearsons r=0.81, p<0.001). In order to understand how basic ERP features could predict classification performance, we computed the correlation between some Target P300 ERP features (described in section 2.3.5) and BTD classification accuracies. The Terp-Btemp and Terp-Ttemp correlations with the BTD accuracies were (Pearsons r=0.45, p=0.09) and (Pearsons r=-0.51, p=<0.05) in the seated condition and (Pearsons r = 0.56, p=<0.05) and (Pearsons r =0.-67, p=<0.01) for the walking condition. Because of this dependence on morphologic features, we wondered if we could use these features to predict which of the two best performing methods, rLDA or BTD, would provide the best classification. The Terp-Ttemp feature is indeed mildly correlated with the difference of rLDA-BTD accuracy in both the seated and walking conditions (Pearsons r=0.18 and 0.53, respectively, p=0.43 and p<0.05). This implies that if the subject's average target ERPs are similar to the Target template, BTD would perform better than rLDA. Post-hoc analysis showed that an arbitrary threshold can be defined on the Terp-Ttemp in order to indicate which method (rLDA or BTD) would perform best per subject. For the seated condition we could define two subjects and in the walking condition nine subjects for which the Target ERP differed substantially from the Target Template resulting in a higher rLDA accuracy as compared to BTD. Substituting the BTD results of those subjects by their rLDA counterparts would lead to a mildly improved grand average accuracy of +1.2

and +1.7% for the seated and walking conditions respectively. These results support the notion that morphology is an important factor in the classification procedure.

3.2. Pseudo Online Scenario

The previously mentioned rLDA results are based on training the classifiers on 80% of the subject specific data. In order to use this method online, in the best case only the past trials (and label information) at moment T can be used for classifier training whereas information from other subjects can be included in the CPD/BTD approach without subject-specific data. First we present the results of the rLDA, CPD and BTD method in such an online scenario. Secondly we present the results of the cross-subject trained rLDA, meaning that classifier is trained on data from non-test-subjects.

3.2.1. Subject Dependent rLDA vs CPD/BTD Our CPD and BTD allow the classification process to start instantaneously. Evidently the rLDA needs sufficient training data in order to derive reliable estimates. Figure 5 displays this relation between training data for the LDAs and corresponding CPD/BTD accuracies at similar sections of the data. The BTD estimates outperform the rLDA estimates for the first 29% (shaded area) of the trials (p-values$<$0.05). In the walking condition similar results can be observed albeit more spread over time. For the first quarter of the trials the BTD accuracy is significantly ($p<0.05$) higher or marginally ($p<0.1$) higher compared to rLDA.

3.2.2. Cross-Subject LDAs vs CPD/BTD Consecutively we compare the results of the CPD/BTD to that of an rLDA classifier trained on data from other subjects. Figure 6 shows the grand average accuracies depending on the number of available subjects for model estimation. It can be seen that the CPD/BTD methods reach their optimal performance when 3 or 4 subjects are available, whereas the LDA based classifiers need data from more subjects to achieve similar accuracy to CPD and requires even more data when compared to BTD. The shaded area shows the significantly ($p<0.05$) higher BTD accuracy as compared to the rLDA in both conditions.

Finally, it is interesting to note that a cross-subject LDA trained on data from 10 subjects provides competitive performance to that of subject-specific LDA (c.q. Figure 6 and Figure 4).

4. Discussion

Recently it was shown that it is feasible to perform a binary auditory BCI classification in a real life environment [9]. However, the need for training supervised LDA classifiers on large parts of that data at the start of a session does not encourage applications outside experimental studies. In the current work we showed that it is possible to remove that explicit subject-dependent calibration phase with a tensor-based decomposition (CPD/BTD) augmented with non-subject-specific templates without sacrifying classification accuracy. This allows for instantaneous classification results that on average are similar to those of the subject-specific trained models. This allows faster interaction with the BCI and is likely to increase the user interest and engagement. Interestingly, we also demonstrate that specific subject-related ERP features are predictive of the BTD results. This direct relation between ERP characteristics and the accuracy of the BTD models makes the classifier more easy to interpret, compared to LDA methods that do not select structured features and rely on additional post-processing steps for interpretation (e.g. [43]). The BTD method presented here has a forthright link to the original P300 ERP signal.

We assumed that the differences between the target and non-target trials were fixed between subjects to some extent. Moreover, our method diminishes a certain amount of subject specific information by adding the average ERP templates to the trials prior to the low rank approximation by CPD/BTD. However, the obtained approximation is an efficient combination of the data and templates to classify the single trial pairs. Transferring additional information from subject to subject or session to session might increase the performance. For example, we noticed a small overall increase in accuracy in the walking condition if the data tensors were constructed with templates from the seated condition. In addition the CPD/BTD might improve with an ability to recognize uncommon ERP morphologies and subsequently update the templates. A similar approach in which the templates incorporate past subject specific data has been proposed in an Riemannian framework in [19]. Furthermore, a comparison to several prototypical responses instead of the 1-2 fixed templates as described currently might be a valuable line of future research.

We argue that the CPD/BTD methods might not lead to the highest accuracy possible, but do lead to interesting insights regarding the EEG BCI signals at hand. For example in a feedback paradigm where the aim is to train a robust P300 waveform by the users, traditional LDA analysis might derive features that are not inherent to the P300 morphology as the proposed method.

The LDAs feature selection step selects very

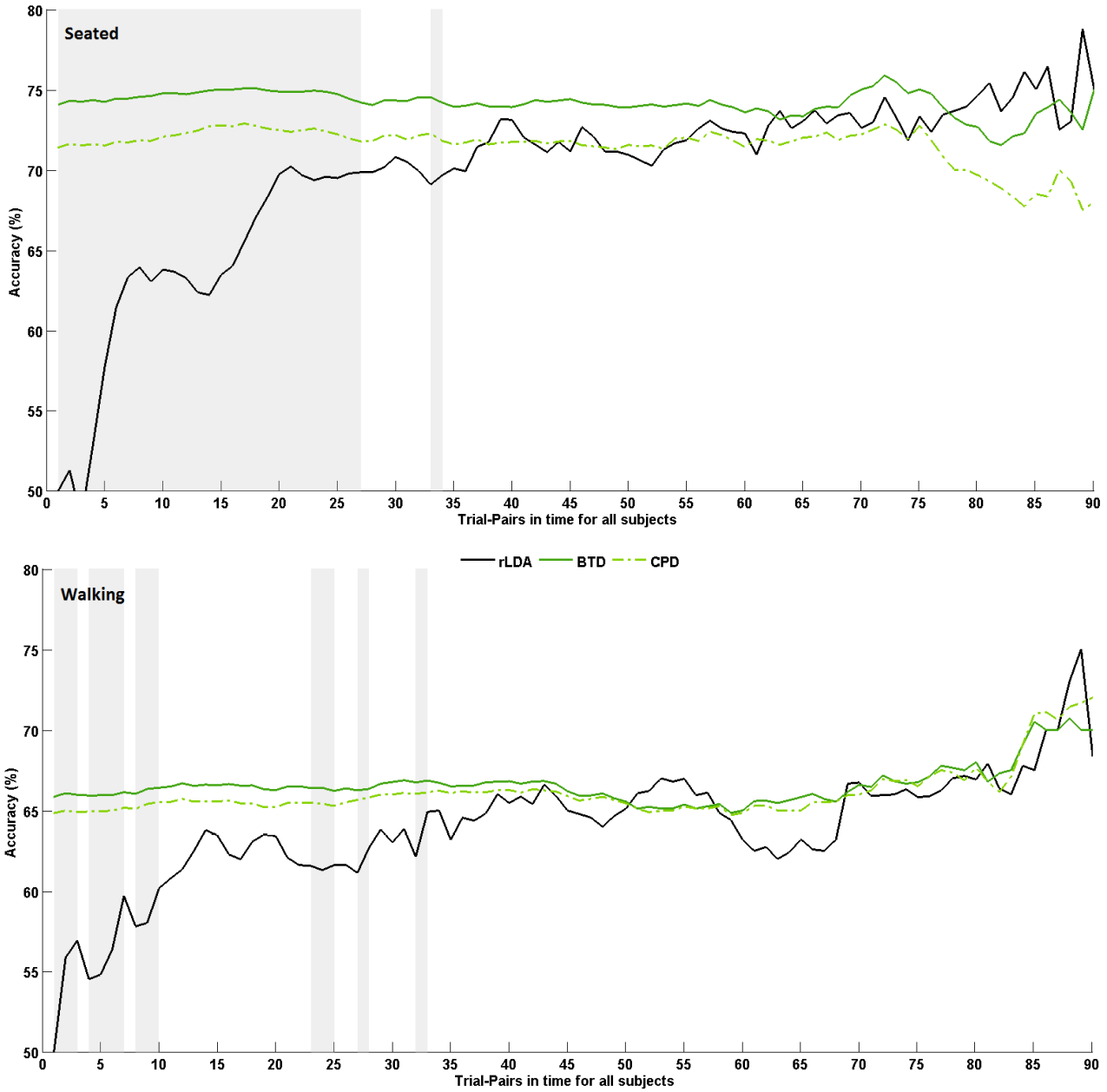

Figure 5. Average classification accuracy across subjects as a function of the available trials for training, when the classifier is evaluated on the remaining data. rLDAs and swLDA estimates are compared with the CPD and BTD classification accuracies. Top: Seated condition. Bottom: Walking condition. The shaded areas mark significant differences between BTD and rLDA at the 0.05 significance level.

distributed features in time and space. Indeed as one expects the most chosen features are at central posterior electrodes in the 250-500ms range, however a substantial number of features is selected towards the extremes of the 0-800ms time window (not shown here). To investigate the effect of these features we limited the LDA feature selection to the same time-window as the CPD/BTD and this did not lead to a significant difference in accuracies (results not shown here).

An important question is to what extent we can generalize our findings. First, the three class oddball paradigms allows for a clean non-target estimate based on the baseline stimuli; something that is not present in the other P300 paradigms (i.e. the frequently used P300-speller). Moreover, the results between the two conditions differ. The BTD and rLDA differences are larger in the seated condition as compared to the walking condition. However, all methods performed significantly lower in the walking

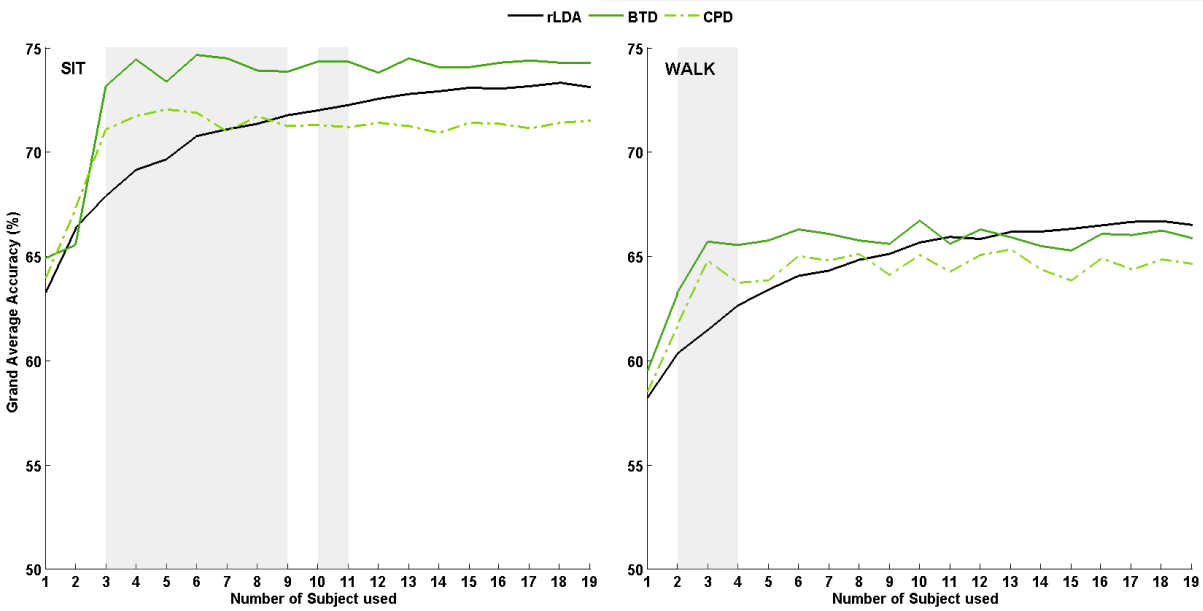

Figure 6. Grand average accuracy as function of the non-specific subject used for training (LDA) or estimating templates (CPD/BTD) for the seated and walking condition left and right respectively. The shaded areas mark significant differences between BTD and rLDA at the 0.05 significance level.

condition as compared to the seated one (Note: these significances are not indicated in figure 4). Consecutively, the CPD/BTD method distinctly needs a clear Target and non-Target template. Similar results with CPD were obtained with only the inclusion of a Baseline template in [44]. It is an open question if such reliable templates can be constructed for other -faster BCI experiments (e.g. speller) where the brain response to stimulation overlaps among stimulus presentations. Nevertheless, this is an important step towards constructing intuitive classification methods by exploiting data signatures from other subjects and structural paradigm information. Such approach has also proven useful and has been advocated by others [16, 18]. We encourage that BCI users really start using classifiers constructed on data from other subjects to maximize time where the BCI is actually used rather than spending significant amounts of time recording training data.

The CPD and BTD performances are shown to be closely related.This was also evident from their high correlation described in the previous section. In general the BTD values outperform the CPD results in the seated condition but not in the walking condition. The CPD model approximates the data tensors as a rank-1 component which is shown to be too restrictive to achieve a discriminative target/non-target signature. The (L_r,L_r,1) BTD is better at modeling the related differences as is evident in the obtained accuracies and interpretation of the factors (c.q. figure 3a, 3b). Similar results are obtained if L was increased to 5. This would suggest that the constructed tensors have prominent target and non-target differences that are already captured with a low rank model of L=2. It can be noted that the rLDA and BTD values correlate strongly in the walking condition but only moderately in the seated. Conversely, the correlations between the ERP features and the rLDA-BTD difference displayed the highest correlation (r>0.5) in the walking condition indicating that the BTD method performs better than rLDA in case of highly prototypical ERPs.

Finally, besides the achievement of instanteneous classification, the cross-subject LDAs comparison to BTD shows that fewer subjects are needed for BTD in order to reach adequate classification. Even though substantial differences are observed between the two methods between subjects, only mild overall improvement can be achieved. It is striking that two completely distinctive methods are not able to improve the results excessively in either condition. This strongly suggests that the limiting factor in this case for the BCI is the lack of task related ERPs. Therefore, future work should focus also on the understanding of the fluctuations in brain responses on single trials and improving BCI paradigms to elicit strong responses rather than merely improving classifiers (e.g. [45, 46, 47]). This can lead to better results with an adaptive BCI approach, especially in real-life scenarios in which the distraction levels are high [48].

5. Conclusion

A method that removes the subject-specifc calibration phase for classification has been shown to have significant benefits over traditional supervised methods such as rLDA on a three-class auditory mobile BCI dataset. With structured CPD and BTD decompositions of single trials and templates our estimates compared favourably to more complex model training in supervised ways. Future work should focus more on the understanding of the fluctuations in brain responses on single trials and incorporating structural information rather than merely improving classifier functioning.

Acknowledgments

We are grateful to Stefan Debener for providing the dataset used in the present study and the reviewers for providing useful comments on the manuscript.

Research supported by Research Council KUL: GOA/10/09 MaNet, CoE PFV/10/002 (OPTEC); Flemish Government, FWO projects: G.0427.10N; EU: ERC Advanced Grant: BIOTENSORS (nr. 339804). This paper reflects only the authors' views, and the Union is not liable for any use that may be made of the contained information.

6. References

[1] J. Polich. Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology*, 118(10):2128–2148, 2007.

[2] A. Furdea, S. Halder, D.J. Krusienski, D. Bross, F. Nijboer, N. Birbaumer, and A. Kübler. An auditory oddball (P300) spelling system for brain-computer interfaces. *Psychophysiology*, 46(3):617–625, 2009.

[3] S. Halder, M. Rea, R. Andreoni, F. Nijboer, E.M. Hammer, S.C. Kleih, N. Birbaumer, and A. Kübler. An auditory oddball brain-computer interface for binary choices. *Clinical Neurophysiology*, 121(4):516–523, 2010.

[4] S. Debener, C. Kranczioch, C.S. Herrmann, and A.K. Engel. Auditory novelty oddball allows reliable distinction of top-down and bottom-up processes of attention. *International Journal of Psychophysiology*, 46(1):77–84, 2002.

[5] N. Birbaumer, N. Ghanayim, T. Hinterberger, I. Iversen, B. Kotchoubey, A. Kübler, J. Perelmouter, E. Taub, and H. Flor. A spelling device for the paralysed. *Nature*, 398(6725):297–298, 1999.

[6] T.O. Zander and C. Kothe. Towards passive brain-computer interfaces: applying brain-computer interface technology to human-machine systems in general. *Journal of Neural Engineering*, 8(2):025005, 2011.

[7] J. van Erp, F. Lotte, and M. Tangermann. Brain-computer interfaces: Beyond medical applications. *Computer*, 45(4):26–34, 2012.

[8] S. Debener, F. Minow, R. Emkes, K. Gandras, and M. De Vos. How about taking a low-cost, small, and wireless EEG for a walk? *Psychophysiol*, 49(11):1617–1621, 2012.

[9] M. De Vos, K. Gandras, and S. Debener. Towards a truly mobile auditory brain-computer interface: Exploring the P300 to take away. *International Journal of Psychophysiology*, 91(1):46–53, 2014.

[10] M. De Vos, M. Kroesen, R. Emkes, and S. Debener. P300 speller BCI with a mobile EEG system: comparison to a traditional amplifier. *Journal of Neural Engineering*, 11(3):036008, 2014.

[11] V. Mihajlovic, B. Grundlehner, R. Vullers, and J. Penders. Wearable, wireless EEG solutions in daily life applications: What are we missing? *IEEE J. Biomed. Health Inform.*, 19(1):6–21, 2015.

[12] B. He, T. Coleman, G.M. Genin, G. Glover, X. Hu, N. Johnson, T. Liu, S. Makeig, P. Sajda, and K. Ye. Grand challenges in mapping the human brain: NSF workshop report. *IEEE Transactions on Biomedical Engineering*, 60(11):2983–2992, 2013.

[13] F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, and B. Arnaldi. A review of classification algorithms for EEG-based brain-computer interfaces. *Journal of Neural Engineering*, 4(2):R1–R13, 2007.

[14] P.J. Kindermans, D. Verstraeten, and B. Schrauwen. A Bayesian model for exploiting application constraints to enable unsupervised training of a P300-based bci. *PLoS ONE*, 7(4):e33758, 2012.

[15] P.J. Kindermans, M. Tangermann, K.R. Müller, and B. Schrauwen. Integrating dynamic stopping, transfer learning and language models in an adaptive zero-training ERP speller. *Journal of Neural Engineering*, 11(3):035005, 2014.

[16] P.J. Kindermans, M. Schreuder, B. Schrauwen, K.R. Müller, and M. Tangermann. True zero-training brain-computer interfacing - an online study. *PLoS ONE*, 9(7):e102504, 2014.

[17] A. Barachant, S. Bonnet, M. Congedo, and C. Jutten. Multiclass brain-computer interface classification by Riemannian geometry. *IEEE Transactions on Biomedical Engineering*, 59(4):920–928, 2012.

[18] A. Barachant and M. Congedo. A plug &play P300 BCI using information geometry. *arXiv:1409.0107*, 2014.

[19] Marco Congedo, Alexandre Barachant, and Anton Andreev. A new generation of brain-computer interface based on riemannian geometry. *arXiv preprint arXiv:1310.8115*, 2013.

[20] W. Samek, F.C. Meinecke, and K.R. Müller. Transferring subspaces between subjects in brain-computer interfacing. *IEEE Transactions on Biomedical Engineering*, 60(8):2289–2298, 2013.

[21] E. Acar, C. Aykut-Bingol, H. Bingol, R. Bro, and B. Yener. Multiway analysis of epilepsy tensors. *Bioinformatics*, 23(13):i10–i18, 2007.

[22] K. Vanderperren, B. Mijovic, N. Novitskiy, B. Vanrumste, P. Stiers, B.R.H. Van den Bergh, L. Lagae, S. Sunaert, J. Wagemans, and S. et al. Van Huffel. Single trial ERP reading based on parallel factor analysis. *Psychophysiol*, 50(1):97–110, 2012.

[23] A. Cichocki, Y. Washizawa, T. Rutkowski, H. Bakardjian, A.H. Phan, S. Choi, H. Lee, Q. Zhao, L. Zhang, and Y. Li. Noninvasive BCIs: Multiway signal-processing array decompositions. *Computer*, 41(10):34–42, 2008.

[24] F. Miwakeichi, E. Martinez-Montes, P.A. Valdés-Sosa, N. Nishiyama, H. Mizuhara, and Y. Yamaguchi. Decomposing EEG data into space-time-frequency components using parallel factor analysis. *NeuroImage*, 22(3):1035–1045, 2004.

[25] M. De Vos, A. Vergult, L. De Lathauwer, W. De Clercq, S. Van Huffel, P. Dupont, A. Palmini, and W. Van Paesschen. Canonical decomposition of ictal scalp EEG reliably detects the seizure onset zone. *NeuroImage*, 37(3):844–854, 2007.

[26] A. Cichocki, D. Mandic, L. De Lathauwer, G. Zhou, Q. Zhao, C. Caiafa, and H.A. Phan. Tensor decompositions for signal processing applications: From two-way to multiway component analysis. *Signal Processing Magazine, IEEE*, 32(2):145–163, 2015.

[27] R. Zink, B. Hunyadi, S. Van Huffel, and M. De Vos. Exploring cpd based unsupervised classification for auditory bci with mobile eeg. In *Neural Engineering (NER), 2015 7th International IEEE/EMBS Conference on*, pages 53–56, April 2015.

[28] B. Hunyadi, D. Camps, L. Sorber, W. Paesschen, M. De Vos, S. Huffel, and L. Lathauwer. Block term decomposition for modelling epileptic seizures. *EURASIP Journal on Advances in Signal Processing*, 2014(1):139, 2014.

[29] J. Farquhar and N.J. Hill. Interactions between pre-processing and classification methods for event-related-potential classification. *Neuroinformatics*, 11(2):175–192, 2012.

[30] Y. Renard, F. Lotte, G. Gibert, M. Congedo, E. Maby, V. Delannoy, O. Bertrand, and A. Lécuyer. OpenViBe: An open-source software platform to design, test, and use brain-computer interfaces in real and virtual environments. *Presence: Teleoperators and Virtual Environments*, 19(1):35–53, 2010.

[31] A. Delorme and S. Makeig. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134(1):9–21, 2004.

[32] M. De Vos, L. De Lathauwer, and S. Van Huffel. Spatially constrained ICA algorithm with an application in EEG processing. *Signal Processing*, 91(8):1963–1972, 2011.

[33] F. Campos Viola, J. Thorne, B. Edmonds, T. Schneider, T. Eichele, and S. Debener. Semi-automatic identification of independent components representing EEG artifact. *Clinical Neurophysiology*, 120(5):868–877, 2009.

[34] T. Demiralp, A. Ademoglu, Y. Istefanopulos, C. Basar-Eroglu, and E. Basar. Wavelet analysis of oddball P300. *International Journal of Psychophysiology*, 39(2-3):221–227, 2001.

[35] R. A. Harshman. *Foundations of the parafac procedure: models and conditions for an" explanatory" multimodal factor analysis*. 1970.

[36] T.G. Kolda and B.W. Bader. Tensor decompositions and applications. *SIAM Rev.*, 51(3):455–500, 2009.

[37] L Sorber, M Van Barel, and L De Lathauwer. Tensorlab v2.0, 2014.

[38] L. Sorber, M. Van Barel, and L. De Lathauwer. Optimization-based algorithms for tensor decompositions: Canonical polyadic decomposition, decomposition in rank-(l_r,l_r,1) terms, and a new generalization. *SIAM Journal on Optimization*, 23(2):695–720, 2013.

[39] L. De Lathauwer. Decompositions of a higher-order tensor in block terms-part II: Definitions and uniqueness. *SIAM. J. Matrix Anal. & Appl.*, 30(3):1033–1066, 2008.

[40] P. Anderer, H.V. Semlitsch, and B. Saletu. Multichannel auditory event-related brain potentials: effects of normal aging on the scalp distribution of N1, P2, N2 and P300 latencies and amplitudes. *Electroencephalography and Clinical Neurophysiology*, 99(5):458–472, 1996.

[41] A. Delorme, T. Mullen, C. Kothe, Z. Akalin Acar, N. Bigdely-Shamlo, A. Vankov, and S. Makeig. EEGLAB, SIFT, NFT, BCILAB, and ERICA: New Tools for Advanced EEG Processing. *Computational Intelligence and Neuroscience*, 2011:1–12, 2011.

[42] B. Blankertz, S. Lemm, M. Treder, S. Haufe, and K.R. Mller. Single-trial analysis and classification of ERP components - a tutorial. *NeuroImage*, 56(2):814–825, 2011.

[43] Stefan Haufe, Frank Meinecke, Kai Görgen, Sven Dähne, John-Dylan Haynes, Benjamin Blankertz, and Felix Bießmann. On the interpretation of weight vectors of linear models in multivariate neuroimaging. *Neuroimage*, 87:96–110, 2014.

[44] R. Zink, B. Hunyadi, S. Van Huffel, and M. De Vos. Classifying the auditory P300 using mobile EEG recordings without calibration phase. *37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2015.

[45] M. De Vos, J.D. Thorne, G. Yovel, and S. Debener. Let's face it, from trial to trial: Comparing procedures for N170 single-trial estimation. *NeuroImage*, 63(3):1196–1202, 2012.

[46] A. Myrden and T. Chau. Effects of user mental state on EEG-BCI performance. *Front. Hum. Neurosci*, 9(308), 2015.

[47] J. Spinnato, M.C. Roubaud, B. Burle, and B. Torresani. Detecting single-trial EEG evoked potential using a wavelet domain linear mixed model: application to error potentials classification. *Journal of Neural Engineering*, 12(3):036013, 2015.

[48] A. Kübler, E.M. Holz, E.W. Sellers, and T.M. Vaughan. Toward independent home use of brain-computer interfaces: a decision algorithm for selection of potential end-users. *Archives of physical medicine and rehabilitation*, 96(3):S27–S32, 2015.