KU LEUVEN

STADIUS
Center for Dynamical Systems, Signal Processing and Data Analytics

Citation/Reference	Zink, R., Hunyadi, B. ; Van Huffel, S. ; De Vos, M. 2015 **Exploring CPD based unsupervised classification for auditory BCI with mobile EEG** Neural Engineering (NER), 2015 7th International IEEE/EMBS Conference on, 53 - 56.
Archived version	Author manuscript: the content is identical to the content of the published paper, but without the final typesetting by the publisher
Published version	http://dx.doi.org/10.1109/NER.2015.7146558
Journal homepage	http://neuro.embs.org/2015/
Author contact	rob.zink@esat.kuleuven.be +32 16 32 76 83
IR	Klik hier als u tekst wilt invoeren.

Exploring CPD based unsupervised classification for auditory BCI with mobile EEG.*

R. Zink, B. Hunyadi, S. Van Huffel, and M. De Vos.

Abstract— **The analysis of mobile EEG Brain Computer Interface (BCI) recordings can benefit from unsupervised learning methods. Removing the calibration phase allows for faster and shorter interactions with a BCI and could potentially deal with non-stationarity issues in the signal quality. Here we present a data-driven approach based on a trilinear decomposition, Canonical Polyadic Decomposition (CPD), applied to an auditory BCI dataset. Different ways to construct a data-tensor for this purpose and how the results can be interpreted are explained. We also discuss current limitations in terms of trial identification and model initialization. The results of the new analysis are shown to be comparable to those of the traditional supervised stepwise LDA approach.**

I. Introduction

To date, the majority of studies exploring brain dynamics using EEG are still conducted in artificial settings like hospitals and research labs [1]. These recordings lack the ability to investigate the full dynamics of our brains in real-life. In particular, the field of Brain Computer Interfaces (BCIs) would tremendously benefit from fully mobile recordings [2].

Currently most BCI systems involve 2 stages: training and testing. During the training stage, a supervised classifier is trained on the recorded data to discriminate between different stimulus classes. However, supervised classification methods require a reasonable amount of training data to optimize classifier performance. Acquiring separate train and test data increases the overall setup time. Moreover, non-stationarities due to varying signal quality and brain adaptations over time decrease the classifier's performance when train and test data are further apart in time. Especially in augmented reality scenarios in which only sparse interaction with a BCI is needed, supervised techniques still carry significant drawbacks.

Unsupervised classification methods can potentially address these issues and be advantageous for use in mobile BCI applications [3,4]. This paper proposes a new approach for unsupervised classification of single trial ERP data based on Canonical Polyadic Decomposition (CPD). CPD is a data-driven method that exploits multidimensional structure in data naturally present in single trial BCI datasets. It allows automatic derivation of multidimensional fingerprints, given that appropriate model parameters are set. Tensor based methods allow for intuitive interpretation of a structured dataset [5]. CPD has been applied in supervised ways to classify single trial ERP data before [6]. We extend this work and propose an unsupervised CPD approach to automatically cluster auditory BCI data. Label information from prior trials is still utilized to evaluate the accuracy of the CPD models; however the actual CPD estimates are independent of any labeled data.

We demonstrate the potential of the new approach on a three-class auditory oddball dataset recorded with a recently developed mobile EEG device. We hypothesize that CPD is able to extract signal sources that allow discriminating target and distractor stimuli. Exploiting the spatial, temporal, and frequency information present in the EEG, discrimination between different stimulus types is obtained. To illustrate our methodology, we first explore the relationship between the CPD parameters and performance on large datasets of 188 trials. An accurate CPD on smaller time-windows would allow for applications involving short-term interactions with a BCI (e.g. operating assistive technologies). To this end we present a method to classify single trials in a 10-trial window. The results are compared to those of stepwise LDA (swLDA), which is considered a prominent classifier for the P300 ERP. We also define future lines of research tackling the current limitations of CPD based classification for online unsupervised BCI.

II. Data & Methods

A. Data acquisition and stimuli

Eight healthy subjects (mean age 24.6 years) performed a three-class oddball auditory task. These subjects are part of the data described in [2]. A standard tone (900 Hz) and two deviant tones (600, 1200 Hz) of 62-ms duration (10 ms rise/fall time) were presented in random order (ISI 1000 ms, Jitter 0-375 ms). The participants' task was to silently count the target tones (i.e. for 4 subjects the 600Hz tone and other 4 subjects the 1200Hz tone) and ignore the two other tones. 504 standards, 94 non-target deviants and 94 target deviants were presented in randomized order while the subject was sitting on a chair outside. OpenViBE software running on a notebook was used for stimulus delivery and experiment control. Data acquisition was performed with a modified Emotiv (www.emotiv.com) EEG system as described in [7,8]. Sintered Ag/AgCl electrodes were placed at the 10–20 positions FPz, F3, Fz, F4, C3, Cz, C4, TP9, TP10, P3, Pz, P4, O1, and O2 (with common mode suppression ("online reference") at AFz and driven right leg ("ground") at FCz).

* Research supported by Research Council KUL: GOA/10/09 MaNet, CoE PFV/10/002 (OPTEC); PhD/Postdoc grants; Flemish Government, FWO projects: G.0427.10N; Belgian Federal Science Policy Office: IUAP P7/19/ (DYSCO, `Dynamical systems, control and optimization', 2012-2017); EU: ERC Advanced Grant: BIOTENSORS (n 339804).

R. Z., B. H and S. V. H. are with 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, (e-mail: rob.zink@esat.kuleuven.be, bori.hunyadi@esat.kuleuven.be and Sabine.VanHuffel@esat.kuleuven.be).

M. D. V. is with the Engineering Department, Oxford University, Oxford United Kingdom (e-mail: maarten.devos@eng.ox.ac.uk) and Cluster of Excellence Hearing4all, University of Oldenburg, Germany.

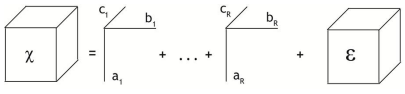
B. Pre-processing

The data were analyzed offline using EEGLAB [9] and MATLAB. Extended infomax independent component analysis (ICA) was used to semi-automatically attenuate contributions from eye blink artifact [10]. EEG data were 20 Hz low-pass filtered, re-referenced to the average of Tp9 and Tp10, and epochs were extracted from −200 to 800 ms and baseline corrected (−200 to 0 ms). In order to speed up the CPD, the data were down-sampled from 128 Hz to 30 Hz. As input for the CPD we concurrently looked at the temporal signal as well as two derived signals: the normalized time signal and the frequency power. For the former this was achieved by normalizing every epoch to z scores and the latter by computing a fast-fourier transform of every epoch. CPD decompositions were computed on a post stimulus interval of 100-800ms. For the latter we selected the amplitude and phase information in a 2-14 Hz interval, as this is expected to capture most of the P300 waveforms [11].

C. CPD

Canonical Polyadic Decomposition (CPD) decomposes multidimensional signals into a sum of rank-1 tensors [12]. Every extracted component is characterized by a certain signature in each of the modes; in the usual ERP analysis' *[channel x time x trials]* tensor, 'a' would be the spatial distribution, 'b' would be the time course, and 'c' would be the trial loadings of a given source. For the three-dimensional case the CPD will decompose a tensor χ as follows:

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



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 $\mathcal{E}$ the model error. CPD is a trilinear model, meaning the vectors along each mode are proportional to each other. Hence, the CPD model assumes that the source maintains the same P300-waveform and topography within the observed trials.

The data is structured in three different tensors following the last three preprocessing steps. Firstly, the data is structured as *[channels x time x trials],* which is expected to capture the P300 waveform present in the data; this is referred to as *CPDtime*. Similarly we construct a tensor with the normalized time signals which is referred to as *CPDtime-Norm*. Normalizing the time course is expected to improve the CPD model's accuracy in detecting small amplitude changes on some channels in contrast to high amplitudes due to artifacts on others. For both described time models the fluctuation in magnitude of the P300 can be taken care of by CPD; it is reflected in the trial mode signature. However, shifts in the P300 latency between trials cannot be resolved thusly. In order to deal with these latency changes we apply a third CPD model on the frequency estimates from every trial in a *[channels x frequency x trials]* tensor, referred to as *CPDfreq*. In this study, the CPD using the alternating least squares algorithm as implemented in Tensorlab 2.0 Toolbox was used [13].

D. Model Parameters

The success of CPD depends on the appropriate setup of certain parameters in the analysis. For the full dataset analysis we discuss the most influential parameters: initialization and rank determination. Being a non-convex optimization method, CPD is not guaranteed to always find the optimal solution. Therefore we report our results based on 100 random initializations. This way we derive reliability estimates and present results based on the best initialization as evaluated by classification accuracy. The rank (R) is expected to differ with the complexity of the dataset. Although semi-automatic rank determination algorithms exist, for noisy EEG data they tend to greatly overestimate the rank. Without defining a specific rank beforehand, we evaluate the results for different ranks ranging from 1 to 12. Finally, the 10 trial windows are decomposed into rank-1 components, as this is the most basic setup.

E. Discriminating trials

In order to discriminate between target and distractor trials with CPD we focus on the factor estimates in the trial-mode. If the rank equals 1, we obtain a single vector with trial weights. Two classes can be distinguished with opposite signs, allowing for a simple, yet effective separation technique (e.g. Figure 3). In the case of a higher rank we consider a k-means cluster algorithm (k-means++) to find 2 distinct clusters. All trial modes of the R-components are given as input for k-means. Being dependent on semi-random initializations, the k-means procedure is repeated 999 times and the most frequent prediction vector is used as final prediction.

First we explore the relationship between the CPD parameters and performance on large datasets of 188 trials. Secondly, we investigate its performance on 10-trial subsets (with equal class distribution). Consequently, the 10-trial overall performance is based on 178 models per subject of 10 trials which are time-shifted by 1 trial for the consecutive model. Single trials are analyzed together with the previous 9 trials in time. The number of correctly classified trials (max = 178) of the 10-trial models is reported as performance measure for each subject. Figure 1 summarizes the most important steps in the CPD analysis. Both the sign and k-means clustering only allow for separation of trials into two classes; it does not identify which class corresponds to the target or distractor stimuli. Label information from prior trials is utilized to evaluate the accuracy of the derived CPD estimates in the 10-trial decompositions. This was achieved by comparing the estimate of the newly added trial to that of the other trials with known labels in that window. This way we can identify which sign is associated with the target class. Future extensions for an adaptive way of determining the identity of the trials are presented in the discussion.

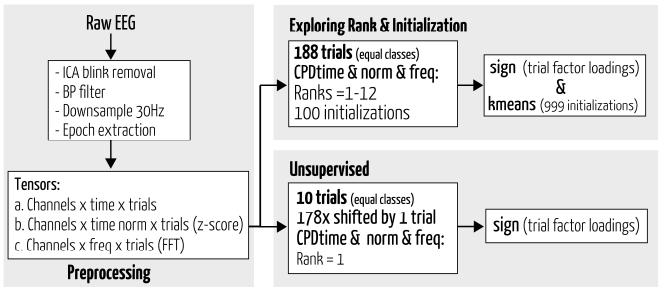


Figure 1: Overview of the most important CPD analysis steps.

Evaluation of the CPD based accuracies is achieved by a comparison to those of swLDA, which is one of the widely used P300 classification algorithms. The original feature set consisted of seventeen 47 ms data bins between 0 and 800 ms on all 12 electrodes. The swLDA method adds relevant features sequentially. A new feature is added to the final feature set if it statistically improves class discrimination (p_{in} < 0.1). After adding a new feature, reanalysis of the current features could lead to removal of a redundant feature (p_{out} > 0.15). Although the stepwise feature selection reduces the number of features used, shrinkage regularization as implemented in BCILAB [9] was applied to further reduce the risk of overfitting. Half of the trials are used for training and the other half for testing and vice-versa in the full dataset. Classification results for the small 10-trial subsets are obtained with a 5 fold cross-validation procedure.

III. Results

A. Electrophysiological results

The subject average subtraction of target-distractor ERPs are presented in Figure 2. These averages illustrate different waveforms of the P300 and peak amplitude and latency between the eight subjects. For all subjects, the P300 of the target stimuli was significantly (at a 5% chance level) larger compared to the distractor stimuli at the P300 peak latency on the Pz electrode. Trial type classification for target and distractor resulted on average in 74.7% accuracy (Standard Deviation (SD) = 6.7) for swLDA across subjects.

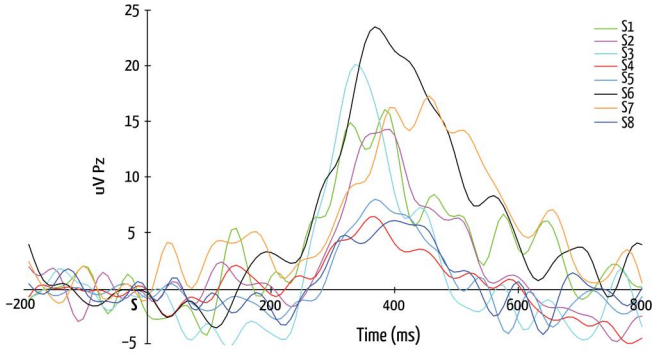


Figure 2: Average difference ERPs per subject at electrode Pz.

B. CPD parameters on all trials

In order to illustrate the underlying single subject decomposition process, components from both a *CPDtime* and *CPDfreq* for a representative subject are shown in figure 3. The modes correspond to the spatial, temporal and trial dimensions from top to bottom, respectively. The *CPDtime* component follows the average target ERP (red dots). Similarly, the *CPDfreq* component is congruent with the target stimulus' frequency spectrum. This leads to similar trial distributions in the third mode. Based on these trial factor loadings, 74.7 and 78.2% correct identification of the trials was obtained for the time and frequency model, respectively. These CPD components are uniquely reconstructed up to arbitrary scaling and permutation of the modes which explains the opposite signs for class-1 between the two CPD models. Finally, equivalent spatial maps can be observed with a strong focus on electrodes Cz and Pz.

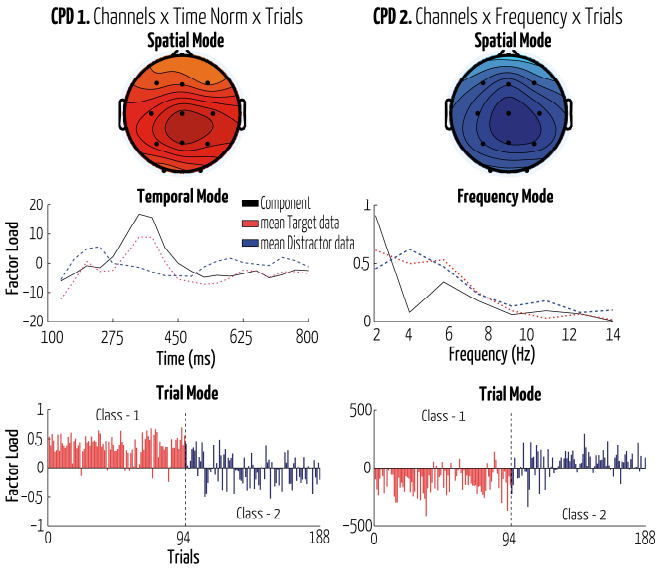


Figure 3: Single subject (S3, rank 1) example of a CPDtime and CPDfreq component on the large dataset. The modes correspond to the spatial, temporal and trial dimensions from top to bottom, respectively. The first half of the trials corresponds to the target stimuli, the latter to the distractor. This is also reflected in the trial factor loadings by CPD (75% and 78% accuracy, respectively).

Figure 4 illustrates the average discrimination across subjects for rank 1 to 12 based on optimal initializations for CPD. On average across subjects, for all three CPD models (i.e. time, time-norm and frequency) we obtained up to 75.5% (SD = 6.8), 75.0% (SD = 5.7) and 77.5% (SD = 5.4) discrimination respectively. No significant differences were observed between swLDA and neither the *CPDtime* for ranks > 3 nor the *CPDfreq* for ranks > 1.

The *CPDtime-norm* and *CPDfreq* models outperform the time based models for lower ranks on this large dataset. However, the fluctuations between models with different initializations differ substantially. The *CPDfreq* models are considerably influenced by the initialization compared to the time models at lower ranks (<6). Further analysis showed that for low rank it often gets stuck at a local minimum which is not discriminative.

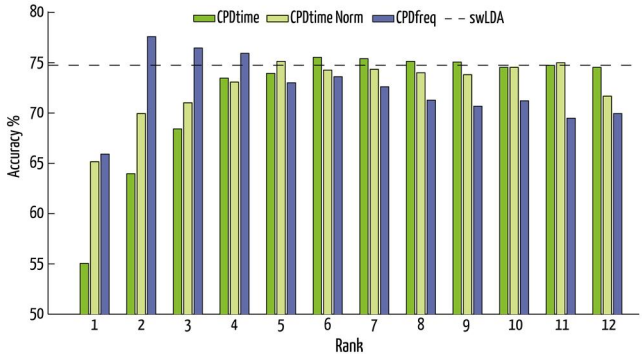


Figure 4: Grand-average accuracy for different ranks of the CPD models. The swLDA accuracy is indicated as reference.

C. 10 trial CPD models

Classification accuracies on the 10-trial subsets for the swLDA are on average 62.1% (SD = 4.7) across subjects.

For the CPD models based on the time, normalized time and frequency tensor we obtained 56.4 (SD = 2.1), 63.0 (SD = 5.8) and 62.9% (SD = 4.7) accuracy respectively. The normalized time and frequency CPD results are comparable to those of swLDA. Figure 5 illustrates the average accuracies for each subject. Further analysis showed that the CPD*time* outcomes were only mildly influenced by the changes between random initializations. The frequency CPD models varied substantially, similar to the full dataset results. Finally, we observed that components that correlated positively with the target ERP as opposed to the distractor ERP (or vice-versa) scored higher in discriminating power.

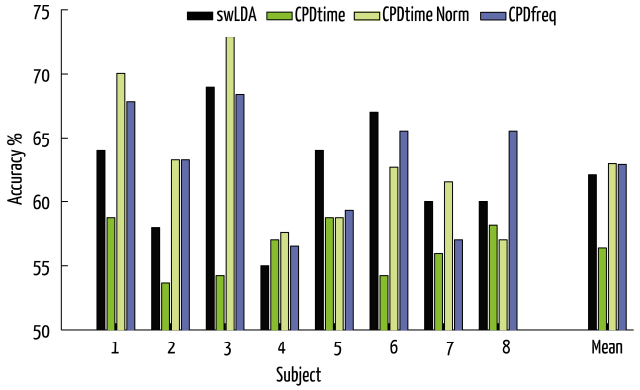


Figure 5: Average accuracy per subject with a rank-1 CPD model.

IV. Discussion & Conclusion

The current study presented a novel way to analyze auditory BCI data based on CPD. The benefits of CPD based discrimination of single trial ERPs in a 3-class auditory oddball task are shown. Even in small datasets, CPD is able to produce meaningful estimates. Applying CPD in an adaptive way so that it could update only the necessary structures is expected to further increase the results. A model initialized on only a few trials that can be updated over time so that it allows exclusion of old data when non-stationarity is occurring has potential for a successful zero-training algorithm. The principal drawback of the presented approach is the inability to identify extracted clusters without prior label information of some trials. Our CPD approach finds distinct clusters that could rely on either non-target or target responses. Future work should focus on the automatic detection of which cluster represents the target stimuli. Addition of the known baseline trials in the CPD might solve this issue as the non-target response should be closely related to the baseline trials. Another solution might arise from the analysis of the 10-trial models which showed that the most discriminative models (80-100%) are strongly correlated to either the target or distractor stimulus and oppositely to the other. Incorporating the results of the second mode in the CPD decomposition (i.e. time or frequency estimates) for identifying the class-stimuli relation might lead to correct labels.

Our results are moderately influenced by the initialization of the CPD models. Finding an objective measure to handle this instability would increase the application potential in noisy datasets. We considered using the residual error in the CPDs as a measure for model fitness to improve the stability. However, this measure proved unsuccessful in identifying well-classifying models.

Improvement of the classification results by combining the trial dimensions of the *CPDtime* and *CPDfreq* models in a joint classifier did not lead to an overall improved classification. This suggests that both time and frequency decompositions provide signal components that represent approximately similar underlying source activation. For the 10-trial analysis, the normalized tensor seems to be well decomposed by CPD. The normalization step decreases the amplitude differences across trials, making the optimization algorithm less prone to outliers due to artifacts. Apart from the CPD, our results can also be influenced by the clustering algorithm. Given the nature of k-means it is not always optimal in finding discriminative clusters. Classifying only on CPD components that represent the signal of interest can increase the results for ranks >1. However, this would require an additional (automatic) component selection step.

In summary, CPD seems to be able to, with a fairly simple model, separate signal and noise in single trial ERP data. Even in small near real-time datasets meaningful estimates are derived without the need for specific training data. Future work should focus on the identity of the derived clusters to facilitate online BCI applications.

References

[1] Mihajlovic, V., Grundlehner, B., Vullers, R., & Penders, J. (2014). Wearable, Wireless EEG Solutions in Daily Life Applications: What are we missing? Biomedical and Health Informatics, IEEE Journal of , vol.PP, no.99, pp.1,1

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

[3] Kindermans, P. J., Schreuder, M., Schrauwen, B., Müller, K. R., & Tangermann, M. (2014). True Zero-Training Brain-Computer Interfacing–An Online Study. PloS one, 9(7), e102504.

[4] Grizou, J., Iturrate, I., Montesano, L., & Oudeyer, P. Y. (2014). Calibration-Free BCI Based Control. In Twenty-Eighth AAAI Conference on Artificial Intelligence (No. EPFL-CONF-198763).

[5] Kolda, T. G., & Bader, B. W. (2009). Tensor decompositions and applications. SIAM review, 51(3), 455-500.

[6] Vanderperren, K., Mijović, B., Novitskiy, N., Vanrumste, B., Stiers, P., Van den Bergh, B. R., ... & De Vos, M. (2013). Single trial ERP reading based on parallel factor analysis. Psychophysiology, 50(1), 97-110.

[7] Debener, S., Minow, F., Emkes, R., Gandras, K., & Vos, M. (2012). How about taking a low-cost, small, and wireless EEG for a walk?. Psychophysiology, 49(11), 1617-1621.

[8] De Vos, M., Kroesen, M., Emkes, R., & Debener, S. (2014). P300 speller BCI with a mobile EEG system: comparison to a traditional amplifier. Journal of neural engineering, 11(3), 036008.

[9] Delorme, A., Mullen, T., Kothe, C., Acar, Z. A., Bigdely-Shamlo, N., Vankov, A., & Makeig, S. (2011). EEGLAB, SIFT, NFT, BCILAB, and ERICA: new tools for advanced EEG processing. Computational intelligence and neuroscience, 2011, 10.

[10] De Vos, M., De Lathauwer, L., & Van Huffel, S. (2011). Spatially constrained ICA algorithm with an application in EEG processing. Signal Processing, 91(8), 1963-1972.

[11] Demiralp, T., Ademoglu, A., Istefanopulos, Y., Başar-Eroglu, C., & Başar, E. (2001). Wavelet analysis of oddball P300. International journal of psychophysiology, 39(2), 221-227.

[12] Harshman, R. A. (1970). Foundations of the parafac procedure: models and conditions for an" explanatory" multimodal factor analysis. Technical report, UCLA Working Papers in Phonetics, 1-84.

[13] Sorber, L., Van Barel, M., & De Lathauwer, L. T. (2014).Tensorlab v2. 0. Available online, URL: http://www. tensorlab. net.