Data-Driven Clustering of P300 EEG Data Using Coupled Tensor Decompositions

Rob Zink [1,2*], Borbála Hunyadi [1,2], Maarten De Vos [3] & Sabine Van Huffel [1,2]
[1] KU Leuven, Department of Electrical Engineering (ESAT), STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics, Kasteelpark Arenberg 10, 3001 Leuven, Belgium.
[2] imec, Leuven, Belgium.
[3] Engineering Department, Oxford University, Oxford, United Kingdom
[*] rob.zink@esat.kuleuven.be

Abstract **-** Recently, it was shown that auditory brain-computer-interface (BCI) classifications can be performed in real life environments. However, the need for initial training of existing supervised classifiers on large parts of that data discourages practical application. Reducing calibration time of BCI analyses would benefit short-term interactions. The current paper presents a carefully simulated study of P300 Event-Related-Potential (ERP) data to illustrate the performance of tensor decompositions for data-driven classification of the P300 effect. The aim of this study was to investigate whether coupling of a high- and low-noise dataset can enhance data-driven clustering of the P300. Imposing structure and linking the decompositions of higher dimensional data arrays called tensors was hypothesized to increase the classification accuracy. For the highest noise dataset (SNR=0.60), we demonstrated that imposing a coupling to datasets with a lower noise level can significantly improve the extracted clusters to classify target from non-target trials to achieve equal accuracy to the widely used supervised regularized Linear-Discriminant-Analysis (rLDA). We evaluated the performance of Canonical Polyadic Decomposition (CPD) and decomposition in multilinear rank Lr,Lr,1 terms (LL1). These structured models do not need a training phase or label information, although they require additional data. Finally, we illustrated the potential of the tensor approach for the analysis of simultaneous EEG recordings in which the trial mode is shared between subjects. Without a priori knowledge of the signal of interest, the tensor-based models successfully separated the two stimuli classes in the highest noise scenario up to 100% for coupling of five simulated subjects. These results highlight the benefits of exploiting structure in the stimuli and experimental setup (e.g., conditions or subjects).

Keywords - Tensor Decompositions, EEG, P300, Data-Driven Classification.

Introduction

Electroencephalography (EEG) is a measurement of electrical activities in the brain which are caused by large neuron group activity. Various electrodes positioned on the scalp record electrical activity, providing excellent resolution of time and enabling the observation and identification of specific areas of the brain that are active at any given point, even on sub-second timescales. Research frequently concentrates on distinguishing between states of the brain relative to events, e.g., the presentation of images (visual) or sounds (auditory). These different states can be utilized to control a device through a brain-computer interface (BCI). The most frequently studied feature in BCIs with EEG is the P300 Event-Related Potential (ERP) which is a positive deflection around 300-500ms in the EEG, produced as a response to both task-relevant and rare stimuli [1]. BCIs were originally conceived to function as access to computers for patients that are locked-in. In recent times they have also been recognized as having potential applications for users that are healthy–certainly if they are not restricted by the bounds of conventional laboratory scenarios [2]. The majority of studies on BCI rely on supervised machine-learning techniques to distinguish task-relevant stimuli [3]. 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. The use of calibration-free classifiers could increase the application potential significantly by removing this specific training phase.

As it is possible to naturally represent ERP data as a third-order tensor (i.e., channels × time × trials), it may be beneficial to exploit this multidimensional structure in the analysis. The existence of a specific spatiotemporal pattern which underlies the target trials, and is not present in the non-targets, can be elegantly exposed by means of tensor decompositions [4]. These optimization-based methods enable a priori knowledge of the data (e.g., noise levels, Expected ERP) and stimulus protocol to be incorporated in the analysis via constraints [5, 6].

Canonical Polyadic Decomposition (CPD) and decomposition in multilinear rank Lr,Lr,1 terms (LL1) are known to have the ability to classify the data of single-trial ERP in ways that are completely data driven [4]. In the current study these models are evaluated to derive meaningful P300 ERP-related components from coupling of various datasets with different levels of signal-to-noise ratio (SNR). Coupling between the datasets is expressed by using one or more common factors in the different factorizations of the tensor models. The coupling of different factors and the influence of noise on the derived clustering of the P300 is evaluated on simulated EEG data (i.e., 15 datasets of 200 trials) of four different noise levels (SNR-range: 3.73 – 0.60) which mimic realistic settings; for example, the visual P300 in lab recordings [7] as well as auditory P300 in recent mobile EEG recordings [8]. The tensor decomposition results were

compared to that of the widely utilized supervised rLDA classification [e.g., in 4,8].

The single CPD and LL1 models are shown to be able to separate ERP subcomponents such as the N100 and P300 and provide reliable differentiation of target and non-target stimuli for low noise levels. For the highest noise levels, coupling of a high- and low-noise dataset enhances the data-driven clustering of the stimuli significantly. Moreover, we illustrate different ways of coupling the various factors in the tensor decompositions to tailor the model to specific usage scenarios (e.g., multi-user recordings [9,10]).

In the next section we explain the simulated data generation and characteristics of the signal and noise. Consecutively the preprocessing steps performed are elaborated, followed by a description of the (coupled) tensor models that are considered in the analysis. Finally, the results of the tensor models and reference method are presented and discussed.

I. Data Generation and Preprocessing

The simulated data were generated using the BESA simulator (www.besa.de/products/besa-simulator/). This simulator utilizes a spherical four-shell head model to generate EEG data based on predefined dipoles [11]. Five dipoles were used for the generation of P300 ERPs as target trials, which are depicted in Fig. 1. Two dipoles with a frontal focus correspond primarily to the N100 effect, a central dipole to the N200 and P3a effects and finally two posterior dipoles that represent a P300-like deflection around 450 ms after stimulus presentation. Consecutively, the two dipoles corresponding to the P300 (i.e., dipole 4 and 5 in Fig. 1) were diminished in amplitude to obtain a second set of EEG data without explicit P300 effects, hereafter referred to as the non-target Trials. Differentiating between the target and non-target trials is the primary aim in BCI approaches, and the simulated data resembles an auditory oddball paradigm [e.g., 8, 12].

The EEG data were simulated at 24 channels corresponding to standard 10-20 locations: FP1, FP2, Fz, F7, F8, FC1, FC2, Cz, C3, C4, T7, T8, CPz, CP1, CP2, CP5, CP6, TP9, TP10, Pz, P3, P4, O1 and O2. The data were generated with a sampling rate of 500Hz. The topography and sampling rate are identical to that of recently published mobile EEG studies ([12]). To obtain realistic data, several types of noise were added to the model. First, the amplitude of the sources is modulated randomly to differ up to 50% in between trials. Second, EEG-like broadband noise (i.e., as implemented in the BESA simulator) was added to the dataset to simulate background activity unrelated to the simulated ERP events. The noise had a relative high correlation between signal amplitudes from electrodes that were close together. noise-free data were generated for reference purposes, and in addition four different proportions of noise were added to the ERP effects. The noise amplitude is scaled by its root mean square (RMS) value to match noise values of 0.5, 1.0, 1.5 and 2.0 uV, and alpha activity at 8-12Hz was generated and added to the signals with an amplitude that was 50% of the overall noise RMS value per trial. The four noise levels corresponded to SNR values of 3.73, 1.75, 0.97, 0.60 from high to low, respectively. The signal-to-noise ratio (SNR) was calculated following previously described procedures in the literature as the ratio of the maximum P300 peak and the pre-stimulus RMS of the EEG.

We simulated 15 datasets of 100 target and 100 non-target Trials for each of the four noise levels and the noise-free case. One trial lasted 1000ms in which the stimulus onset was simulated at 200ms. The data were preprocessed offline using EEGLAB and MATLAB (Mathworks Inc., Natick, MA). The EEG data were 0.5-20 Hz band-pass filtered and baseline-corrected (−200–0 ms) after re-referencing to the mean of TP9 and TP10. To eliminate the effect of outliers in the data, we performed a Z-score normalization per channel and trial. This way, the overall variance is more uniform before analyzing the data with the tensor models. The average ERPs at channel Pz

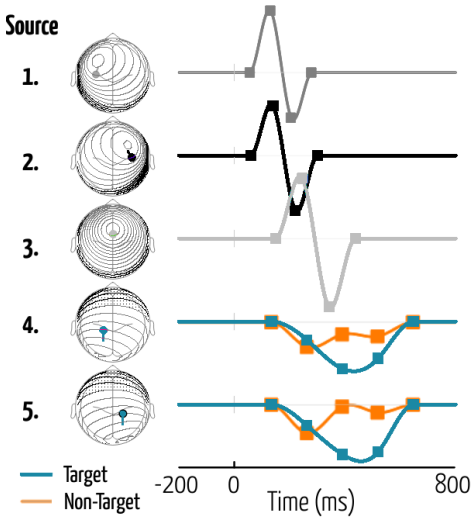


Fig. 1. Overview of the dipole spatial location on a top-down view (left) and temporal characteristics of each dipole on the right. The blue lines depict the temporal progression in the Target trials and the orange for the Non-Target trials. Image obtained from the BESA simulator (available online at www.besa.de/products/besa-simulator) and modified to highlight the P300 source.

for the target and non-target stimuli are presented in Fig. 2 for each of the noise levels. The corresponding SNR values illustrate that the last two noise levels portray situations in which the noise is equal to or higher than the signal of interest (i.e., P300), mimicking realistic mobile EEG P300 scenarios [8].

II. Tensor Based Models

CPD

Multidimensional signals can be decomposed by the CPD as a sum of rank-1 terms [13]. For the three-dimensional case,

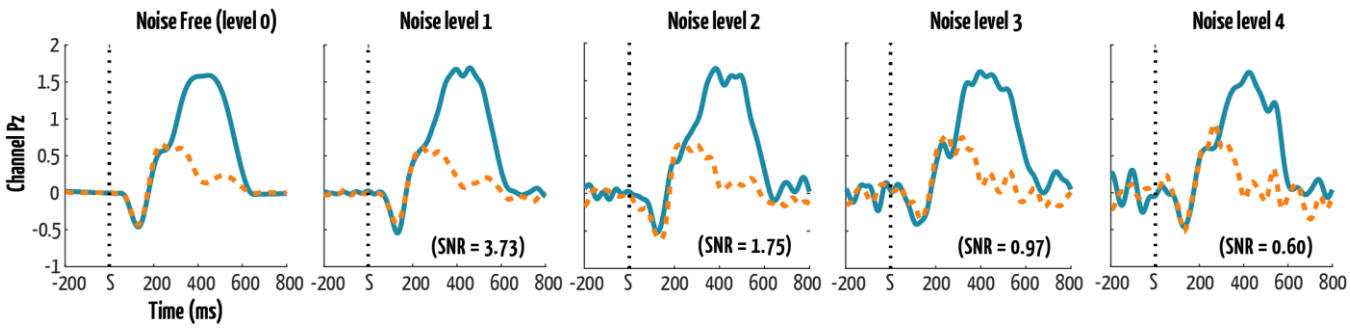
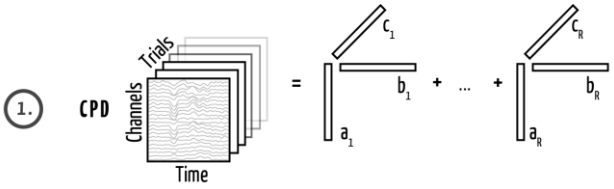


Fig. 2. Average ERPs at Channel Pz for the noise free case (most left) and the four increasing levels of noise from left to right respectively.

the CPD will decompose a tensor *X* as follows:

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



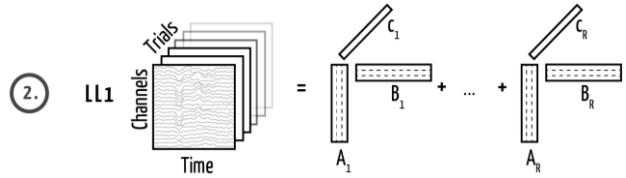
with *R* representing the number of components, a_r, b_r, and c_r, the signatures of every atom in each of the modes, and ε, the model error. Each mode has a specific signature which characterizes the extracted component; in the three-dimensional tensor representing the ERP as a channel × time × trials 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. An example of a decomposition with CPD is illustrated in Fig. 3. This represents the decomposition of a dataset of intermediate noise level (SNR = 1.75). The first component extracts characteristics that relate to the N100 ERP with a frontal central focus, and the second to the P300 effect with a posterior topography. The last component reflects an alpha noise source. From the third mode (i.e., trial mode), it can be noted that the second component clearly differentiates between the two classes in the data. In this dataset the trials could be separated with 92% accuracy.

The CPD 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 [14]. The size of the data tensor for the single CPD and single LL1 models is 24 × 500 × 200 for the channels time and trial dimension, respectively.

LL1 decompositions

Although CPD provides interpretable components, the model can be too restrictive for some applications, as it does not model all variability in the data [4, 15]. LL1 allows the modeling of more variation in two factors (denoted by Lr) [16, 17]. While for the CPD model the time course of a certain component is the same on all channels, LL1 will allow some variation of the time course on the different channels. The LL1 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. A three-dimensional data tensor *X* can be decomposed by a LL1 as:

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



The tensor *X* is the sum of the outer products of a rank L_r matrix (the product of matrices A_r and Br-transposed) and the component vector c_r, with R representing the number of components and ε, again the model error. Similar to the rank, L_r should be set a priori. In our example we allow the spatial and temporal mode to be of higher rank as compared to the CPD models. This is expected to capture time and waveform variability more accurately between the target and non-target effects as constituted in the trials. The interpretation of the trial dimension c is similar as to the one from CPD model.

An overview of tensor decompositions used in signal processing applications is presented in [18] and [19].

Coupled Decompositions

Instead of decomposing a single dataset, CPD and LL1 can be used to express coupling between datasets. A so-called coupled decomposition of two datasets can be achieved in a framework known as Structured-Data Fusion (SDF) which is a specific syntax to impose structure on factors in the tensor decomposition [5, 6]. For example, we can impose equality between the spatial factors of two decomposed datasets of different SNR level. This is illustrated for the CPD case in the schematic below, indicating equality between both the spatial and temporal mode in the two datasets. In the current work, we studied the effectiveness on the P300 separation when the

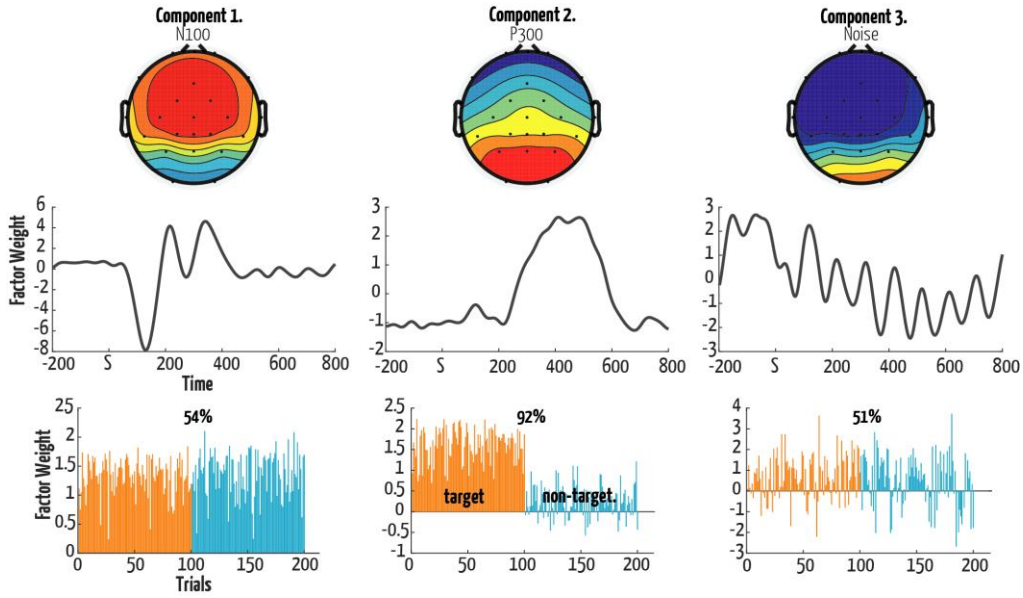
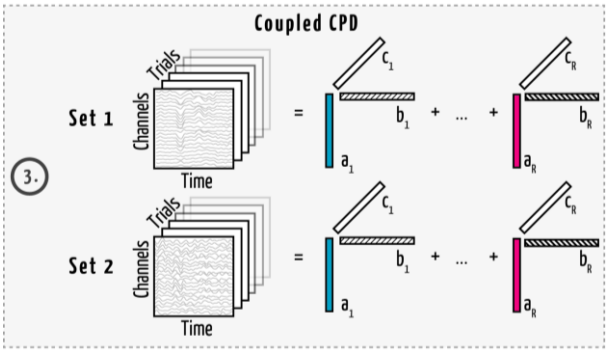


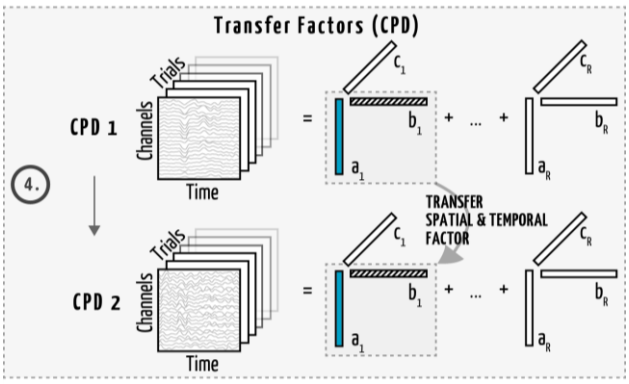
Fig. 3. Example CPD outcome of a dataset from noise-level 2, illustrating the decomposition with the corresponding spatial, temporal and trial modes for a model or Rank = 3. The components depict the N100, P300 and noise signature from left to right, respectively. The first half of the trials corresponds to the target Trials, the latter to non-targets. Component 2 is able to separate the classes by 92%, based on the factor weights of the third mode.

highest noise condition was coupled with a lower-noise-level dataset. This is based on the assumption that the underlying source model for the P300 in both datasets is (roughly) similar and only the variability over trials (over the course of the experiment) is different. For both the CPD and LL1, we evaluated the clustering on the lowest SNR dataset when jointly decomposed with higher SNR datasets. The decomposition of both datasets happens simultaneously and without any additional stimulus information. The size of the coupled model is comprised of 2 data tensors of dimensions 24 × 500 × 200 for the channel, time and trial dimension, respectively.



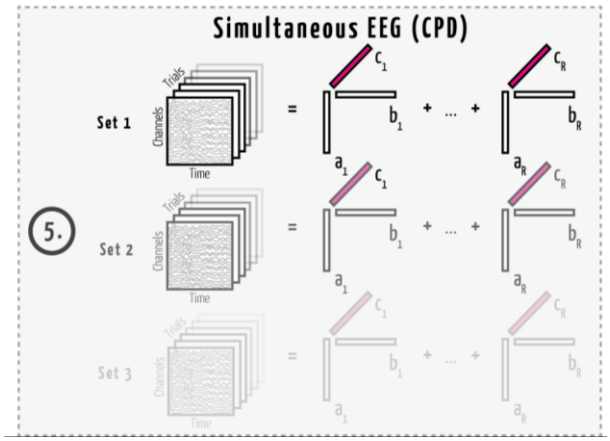
Fixed-Factor Decompositions

Tensor decompositions can be applied in a supervised way by transferring factors between models. For example, a spatiotemporal signature from one CPD or LL1 model from a low noise dataset can be chosen and imposed on the decomposition of a high noise dataset. This way, we aim to achieve a better clustering of the target trials in contrast to the non-targets in the high noise conditions. This so-called fixed-factor decomposition requires the datasets to be decomposed consecutively and relies on identifying the most useful component in the first decomposition that is to be transferred. This concept is illustrated in the schematic below. Note that the top decomposition is evaluated first and the best temporal-spatial pattern is transferred to the second model. The size of each data-tensor is 24 × 500 × 200 (channels, time and trial), and each is decomposed independently of the other.



Simultaneous Recordings (multi-user BCI)

Besides coupling or fixing factors of the spatial or temporal modes, one could opt for coupling the trial mode of two or more datasets. This could be useful for a specific type of paradigm that is gaining momentum in the past years in the field of EEG research and is comprised of simultaneous recordings of multiple subjects. One example of such paradigm is the control of a BCI Space Invaders game by two subjects [9] and another, the simultaneous analysis of EEG in response to video in a classroom environment [10]. Since these subjects may have different spatiotemporal patterns, the spatial and temporal modes are left unconstrained. However, the trial (i.e., time) dimension is assumed to be identical, as the subjects are doing the same task at the same time. The schematic below illustrates the constrained CPD model in which the third (trial) mode is shared between datasets. This way the CPD or LL1 model derives the shared information between two or more of such datasets on the trial factor. The additional structure in the trial factor is employed to increase the overall SNR. The performance of this coupling is evaluated in the current study by combining up to five datasets of the highest noise scenario (i.e., SNR = 0.60) to increase the overall distinction between target and non-target trials. The size of data that is used in the decompositions is 1 to 5 times a single subject dataset of dimensions 24 × 500 × 200 that represent the channel, time and trial dimension, respectively.



Decomposition Parameters

In this study, all CPD and LL1 models were computed with the nonlinear least squares (NLS) algorithm in the publicly available Tensorlab 3.0 toolbox [5]. These methods are dependent on several parameters of which the following are evaluated: the initialization of the model, the number of iterations of the NLS algorithm, the number of components (i.e., Rank *R*) and, specifically for LL1, the value of L. Unless stated otherwise, all models were initialized randomly. By default, all computations were allowed up to 1000 iterations for the NLS. Post hoc evaluation of the number of iterations did not improve the results and these results were therefore omitted because of space limitations. In the case of LL1, values of L>3 showed no clear differences on an extra simulated dataset and therefore a value of L=3 was chosen for all LL1 models in the current analysis. The Rank of the models was evaluated explicitly for Ranks of 1 to 10.

Clustering and Classification

The factor loading on the trial dimension can be used to naturally obtain two clusters (e.g., Fig. 3, Component 2).

Accuracies were obtained by taking the median value of the trial factor as threshold for each component in the decomposition to obtain two clusters. The resulting classes are compared to the true labels, and this results in a clustering percentage for each component. This method only allows for separation of the trials into two classes; it does not identify which class corresponds to the target or distractor stimuli. Nevertheless, identifying the stimulus may be derived from the corresponding spatiotemporal signature in the other modes. Evaluation of the Tensor-based clustering is achieved by a comparison to those of rLDA, which is one of the widely used P300 classification algorithms [e.g., 8, 12]. This requires a separate training phase to calibrate the classifier function. 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 [3]. The basic LDA feature set comprised seventeen 47ms data bins on all 12 electrodes between 0–800 ms, which is reported repeatedly as an efficient feature set for classifying the P300 [8, 12]. Shrinkage regularization as implemented in BCILAB is used for rLDA classification. Per subject, the classifiers are trained based on five-fold cross-validation procedure.

III. RESULTS

Single Dataset Models

The grand average classification accuracies are summarized in Table 1. For the conditions 'without noise' and 'lowest noise level' (SNR = 3.73), all models performed near 100% for separating the target from non-target trials. In the case of an SNR of 1.75, the clustering was on average correct for 92% of the trials, for all methods. The two highest noise cases (noise 3 and 4) were significantly better clustered by the supervised rLDA, as opposed to the LL1 models, ($t14= 5.93$, $p<0.0001$) and ($t14= 14.33$, $p<0.0001$) respectively. The CPD results were on par with the rLDA for the noise 3 case, even though they were lower in the highest noise data ($t14= 4.53$, $p<0.001$).

Table 1. Average accuracies of the rLDA, CPD and LL1 method for discriminating between target and non-target trials for each of the five different levels of noise. Values with an asterix were significantly lower as compared to rLDA ($p < 0.001$).

Method	**Average Clustering % (± SD)**				
	Noise-Free	***Noise 1 SNR = 3.73***	***Noise 2 SNR = 1.75***	***Noise 3 SNR = 0.97***	***Noise 4 SNR = 0.60***
Supervised rLDA	100.0 (±0)	99.0 (±0.9)	91.8 (±1.3)	85.6 (±3.0)	78.9 (±2.9)
uncoupled CPD	100.0 (±0)	99.6 (±0.5)	92.6 (±2.0)	85.0 (±3.0)	**71.0 (±7.1)***
uncoupled LL1	100.0 (±0)	99.3 (±0.6)	91.9 (±1.9)	**73.3 (±7.5)***	**62.9 (±3.2)***

Fig. 4 illustrates the influence of the Rank of the tensor models on the clustering percentage. The number of components was found to be optimal for 3-4 components over all noise levels. This is similar to previously described results on clustering mobile EEG data with CPD [20]. LL1 outperforms the CPD results for models of only one component for all noise levels. Similarly, the LL1 achieves better clustering for higher component numbers (i.e., > 5) for the lowest noise level. A remarkable finding is that the LL1 models performed significantly lower on the two highest noise levels, as compared to CPD if the Rank of the models is larger than 1.

Transferring Spatial and Temporal factors

Transferring the most discriminative spatial and temporal factor from a low-noise dataset to the decomposition of the highest-noise dataset improves the clustering for both the CPD and LL1 up to 83% and 81%, respectively. This procedure can be best compared to transferring knowledge of the same subject in between different conditions or recordings. Fig. 5 illustrates the clustering results on the highest noise level and the origin of the fixed component: the noise-free case, or one of the other three noise levels. Surprisingly, the transferred spatial and temporal factor from the noise-free (i.e., noise level 0) component is not very efficient in the highest noise dataset, as is evident from the blue line in Fig. 5. Note that these results are not dependent on the number of components. This is to be expected, as the first component is always fixed and already contains a good estimate of the P300 signal of interest, and therefore the other components are mostly filled by noise. For almost all presented fixed factor decompositions, the fixed component was the most discriminative. This removes the need for component selection but requires label information of the data tensor that the factors were taken from, rendering the method supervised.

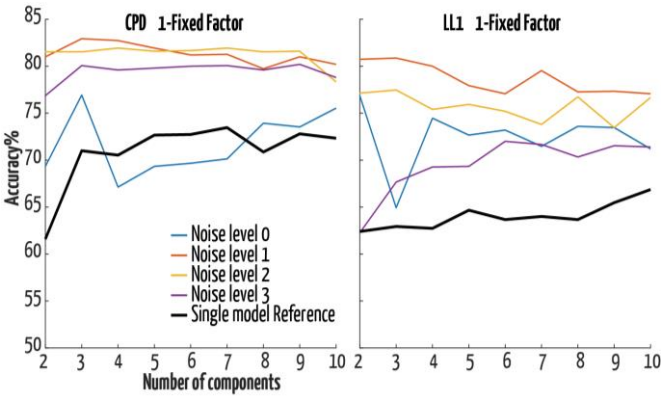


Fig. 5. Accuracies of CPD and LL1 with a fixed factor for the highest noise condition. Each line represents a different noise level from which the best spatial and temporal factors were identified to be fixated in the highest noise dataset decomposition.

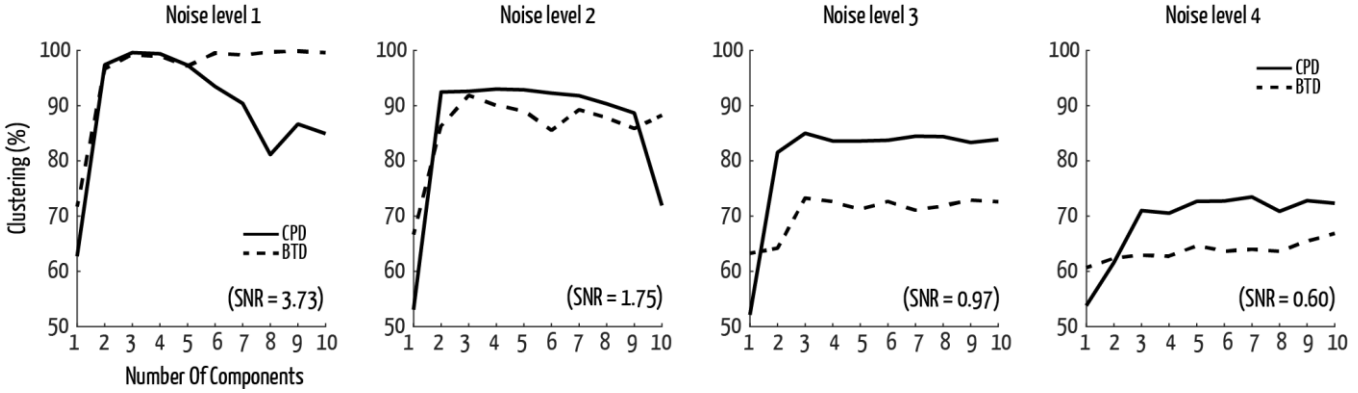


Fig. 4. Average clustering percentage of the CPD and LL1 method for each of the four noise levels, dependent on the number of components considered in the models.

Coupled models of low and high noise

With the coupled models, we aim to increase the clustering of the highest noise level, as here, the single-model results fall short of supervised alternatives such as the rLDA presented here. Clustering the highest noise condition (SNR = 0.60) through a structured coupled CPD with a dataset from one of the lower noise-level datasets resulted in a clustering percentage of 81.9% (± 2.0), 80.3% (±2.7), 79.4% (±2.5) and 78.3% (±3.0) for the lowest to highest noise dataset. These values are significantly higher compared to the single CPD model (i.e., 71.0% (±7.1)). Similarly, the single model LL1 estimate (62.9% (±3.2)) could be improved substantially for the noisiest dataset when coupled to one of the four lower-noise-level sets, 82.9% (± 2.4), 82.1% (±3.3), 75.7% (±3.7) and 64.2% (±3.4) from low to high noise, respectively. Fig. 6 depicts the influence of the noise level of the dataset that was jointly decomposed and the dependency on the Rank of the models for CPD and LL1. Specifying the number of components appears more crucial for LL1 as compared to CPD: for the former, higher ranks are required when coupled to noisier level datasets (c.q. noise level 0-1 to 2-3). In summary, these results are similar to the fixed-factor results presented in the previous paragraph without the need of specific label information.

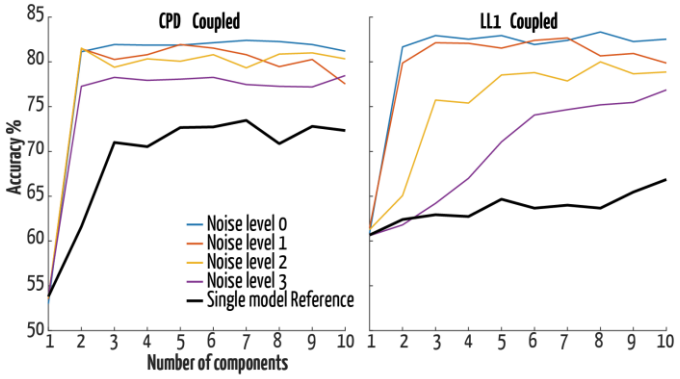


Fig. 6. The CPD and LL1 clustering accuracy of the highest-noise-level dataset if it was coupled to one of the lowest four noise levels (indicated as level 0-3) in contrast to the number of components in the decomposition.

Coupling in the Trial Dimension

The coupling of several datasets of the highest noise level with a shared trial mode results in improved clustering of the stimuli in the highest noise condition. Fig. 7 illustrates the clustering percentage in relation to the number of coupled datasets. Evidently there is a linear increase in the number of datasets that are combined and provide near 100% separation in the case of five simultaneously decomposed datasets. Moreover, these results are not dependent on the number of components as long as R >1.

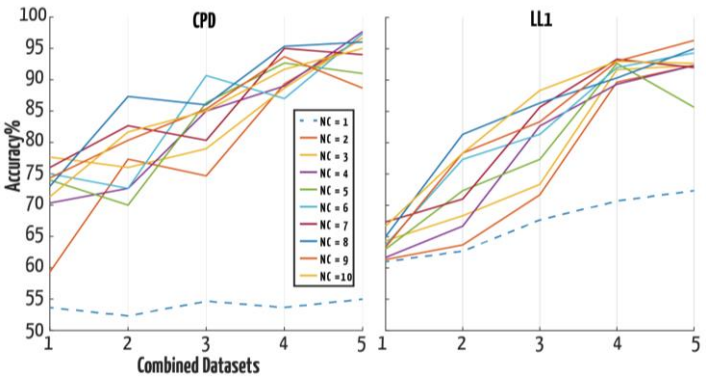


Fig. 7. Clustering percentage in relation to the number of coupled datasets in the trial-mode for the highest-noise-level dataset (SNR = 0.60). Each line corresponds to a different Rank.

IV. Discussion and Conclusion

Recently, it was shown that auditory BCI classifications can be performed in real life environments [8, 12]. However, the need for initial training of supervised classifiers on large parts of that data discourages practical application. Faster interaction with a BCI would likely increase user interest and engagement. Here, we explored the incorporation of structural information into the analysis in several ways to increase data-driven clustering of target and non-target trials. The current study carefully simulated P300 ERP data from 15 subjects at 5 different noise levels, mimicking realistic settings. We showed that a simple single CPD or LL1 model is able, without supervision, to separate signal and noise in simulated single-trial P300 ERP data (e.g. the components that represent the N100, P300 and noise in Fig. 3). At low noise levels, these models perform equally to that of supervised rLDA, which requires specific label and training phases (table 1.). In contrast, the CPD and LL1 models lack the ability to achieve high performance for low SNR datasets. For the highest-noise dataset (SNR=0.60), we illustrated that imposing a coupling to datasets with a lower noise level can significantly improve the extracted clusters to classify target from non-target trials to achieve equal accuracy to rLDA for both the CPD and LL1 (Fig 6.). These structured models function without a training phase or label information, although they require additional data. Transferring a derived spatiotemporal component from a low- to high-noise dataset resulted in a similar improvement of the clustering, albeit in a supervised way; the best component on the low noise dataset has to be determined beforehand. Finally, we illustrated the potential of coupled models for the analysis of simultaneous EEG recordings in which the trial mode is shared between subjects. Without a priori knowledge of the signal of interest, the tensor-based models successfully separated the two stimuli classes in the highest noise scenario up to 100% for coupling of five simulated subjects.

The finding that CPD and LL1 models did not surpass the rLDA classification suggests that the limiting factor in this case for the separation is the lack of task-related P300 ERPs. Further extension of the presented approaches could be to identify which of the extracted clusters represents the target stimuli, based on the spatiotemporal modes of the component. Adding additional constraints (e.g. independence or sparsity) on the factors in the tensors models might improve the clustering performance [21]. The LL1 models estimate a higher number of parameters, compared to the CPD (at equal Rank). Whether this is the cause of the lower results at higher noise levels compared to CPD remains an unanswered question. One potential reason could be that the LL1 models over-fit to the noise, because, at these SNR levels, the level of noise is substantially higher than the P300 signal.

In summary, these results highlight the benefits of exploiting structure in the stimuli and experimental setup for classification of single-trial EEG data.

Acknowledgment

The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/20072013)/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.

REFERENCES

[1] Polich J 2007 Updating P300: an integrative theory of P3a and P3b Clin. Neurophysiol. 118 2128–48

[2] Zander T O and Kothe C 2011 Towards passive brain–computer interfaces: applying brain–computer interface technology to human-machine systems in general J. Neural Eng. 8 025005

[3] Lotte F, Congedo M, Lécuyer A and Lamarche F 2007 A review of classification algorithms for EEG-based brain–computer interfaces J. Neural Eng. 4 R1

[4] Zink R, Hunyadi B, Huffel S V and De Vos M 2016 Tensor-based classification of auditory mobile BCI without subject-specific calibration phase J. Neural Eng. 13 026005

[5] Vervliet, N., Debals, O., Sorber, L., Van Barel, M., & De Lathauwer, L. (2016). Tensorlab 3.0, 2016. Available online, Mar.

[6] Sorber, L., Van Barel, M., & De Lathauwer, L. (2015). Structured data fusion. IEEE Journal of Selected Topics in Signal Processing, 9(4), 586-600.

[7] Heinrich, S. P., & Bach, M. (2008). Signal and noise in P300 recordings to visual stimuli. *Documenta Ophthalmologica*, *117*(1), 73-83.

[8] 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.

[9] Korczowski, L., Barachant, A., Andreev, A., Jutten, C., & Congedo, M. (2016, May). " Brain Invaders 2": an open source Plug & Play multi-user BCI videogame. In *6th International Brain-Computer Interface Meeting* (pp. 10-3217).

[10] Poulsen, A. T., Kamronn, S., Dmochowski, J., Parra, L. C., & Hansen, L. K. (2017). EEG in the classroom: Synchronised neural recordings during video presentation. *Scientific Reports*, *7*.

[11] Berg, P., & Scherg, M. (1994). A fast method for forward computation of multiple-shell spherical head models. *Electroencephalography and clinical neurophysiology*, *90*(1), 58-64.

[12] Zink, R., Hunyadi, B., Van Huffel, S., & De Vos, M. (2016). Mobile EEG on the bike: disentangling attentional and physical contributions to auditory attention tasks. *Journal of Neural Engineering*, *13*(4), 046017.

[13] Harshman R A 1970 Foundations of the parafac procedure: models and conditions for an ‘explanatory’ multimodal factor analysis. *UCLA Working Pap. Phonetics* 16 1–18

[14] Kolda T G and Bader B W 2009 Tensor decompositions and applications *SIAM Rev*. 51 455–500

[15] Hunyadi B, Camps D, Sorber L, Paesschen W, De Vos M, Huffel S and Lathauwer L 2014 Block term decomposition for modelling epileptic seizures *EURASIP J. Adv. Signal Process*. 2014 139

[16] Sorber, L., Van Barel, M., & De Lathauwer, L. (2013). 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.

[17] De Lathauwer L 2008 Decompositions of a higher-order tensor in block terms-part II: definitions and uniqueness *SIAM. J. Matrix Anal. Appl*. 30 1033–66

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

[19] Cong, F., Lin, Q. H., Kuang, L. D., Gong, X. F., Astikainen, P., & Ristaniemi, T. (2015). Tensor decomposition of EEG signals: a brief review. *Journal of neuroscience methods*, *248*, 59-69.

[20] Zink, R., Hunyadi, B., Van Huffel, S., & De Vos, M. (2015, April). Exploring CPD based unsupervised classification for auditory BCI with mobile EEG. In *Neural Engineering (NER), 2015 7th International IEEE/EMBS Conference on* (pp. 53-56). IEEE.

[21] De Vos, M., Nion, D., Van Huffel, S., & De Lathauwer, L. (2012). A combination of parallel factor and independent component analysis. *Signal Processing*, *92(12)*, 2990-2999.