PAPER • **OPEN ACCESS**

ETucker: a constrained tensor decomposition for single trial ERP extraction

To cite this article: Behrad TaghiBeyglou and Mohammad Bagher Shamsollahi 2023 *Physiol. Meas.* **44** 075005

IOP Publishing Physiol. Meas. 44 (2023) 075005 https://doi.org/10.1088/1361-6579/ace510

Physiological Measurement

IPEM Institute of Physics and Engineering in Medicine

OPEN ACCESS

RECEIVED
2 March 2023

REVISED
29 June 2023

ACCEPTED FOR PUBLICATION
6 July 2023

PUBLISHED
21 July 2023

PAPER

ETucker: a constrained tensor decomposition for single trial ERP extraction

Behrad TaghiBeyglou[1] **and Mohammad Bagher Shamsollahi**[2]

[1] Institute of Biomedical Engineering, University of Toronto, Ontario, Canada
[2] School of Electrical Engineering, Sharif University of Technology, Tehran, Iran

E-mail: behrad.taghibeyglou@mail.utoronto.ca

Keywords: electroencephalogram, event-related potentials, ETucker, single trial, tensor, Tucker decomposition

Abstract

Objective. In this paper, we propose a new tensor decomposition to extract event-related potentials (ERP) by adding a physiologically meaningful constraint to the Tucker decomposition. *Approach.* We analyze the performance of the proposed model and compare it with Tucker decomposition by synthesizing a dataset. The simulated dataset is generated using a 12th-order autoregressive model in combination with independent component analysis (ICA) on real no-task electroencephalogram (EEG) recordings. The dataset is manipulated to contain the P300 ERP component and to cover different SNR conditions, ranging from 0 to −30 dB, to simulate the presence of the P300 component in extremely noisy recordings. Furthermore, in order to assess the practicality of the proposed methodology in real-world scenarios, we utilized the brain-computer interface (BCI) competition III-dataset II. *Main results.* Our primary results demonstrate the superior performance of our approach compared to conventional methods commonly employed for single-trial estimation. Additionally, our method outperformed both Tucker decomposition and non-negative Tucker decomposition in the synthesized dataset. Furthermore, the results obtained from real-world data exhibited meaningful performance and provided insightful interpretations for the extracted P300 component. *Significance.* The findings suggest that the proposed decomposition is eminently capable of extracting the target P300 component's waveform, including latency and amplitude as well as its spatial location, using single-trial EEG recordings.

1. Introduction

Event-related potentials (ERPs) are the responses of brain to an internal/external stimulus and are mostly observable in electroencephalogram (EEG) signals. Depending upon the specifications of the stimulus and the characteristics of the person responding to it, amplitude, time of occurrence (after the onset), polarity, and the spatial position of the ERPs may vary (Schomer and Da Silva 2012, Luck 2014, Ting *et al* 2014, Luck *et al* 2000). As an example, P300 ERP component, which is provoked as a result of an oddball task during EEG recording, mostly occurs after 250–650 ms of the onset, has positive amplitude, and is located in thesuperior and inferior frontal and parietal lobe (Kaplan *et al* 2013, Sabeti *et al* 2016).

Cross-trial temporal averaging, also known as grand averaging method, which averages trials over time, has been widely used to extract ERP components from EEG recordings, particularly in brain-computer interface applications (Xiao *et al* 2019, Ma *et al* 2021). The underlying assumption of this method is that the ERP component is deterministic and the background EEG is a white noise (Luck 2014). This assumption, though practical and handy, has been questioned in several studies in presence of the jitter effect, which may happen due to mental fatigue or attention deficit (Käthner *et al* 2014). Moreover, the amplitude, latency, and spatial location of the ERP components have been prone to variations in different recording trials (between-trial-variability) (Jarchi *et al* 2011). In this respect, several studies have attempted to propose methods for extracting ERP (sub) component(s) from single-trial EEG recordings. The approaches can be classified into five groups; blind source

separation (BSS), e.g. independent component analysis (ICA) (Lee *et al* 2016) or principal component analysis (PCA) (Dien 2012), adaptive filter, e.g. Wiener filter (Cerutti *et al* 1987) or Kalman filter (Georgiadis *et al* 2005), denoising algorithms such as wavelet transform (Quiroga and Garcia 2003, Aniyan *et al* 2014), iterative estimation algorithm such as RIDE (Ouyang *et al* 2011), and spatio-temporal filtering methods (Li *et al* 2009, Jarchi *et al* 2010, Ranjbar *et al* 2018). While some methods suffer from lack of interpretable spatial estimation for the targeted ERP component, the others require several trials to estimate the component in each trial or fail to extract the dynamics of the component in bad SNR conditions (lower than −15 dB).

Tensor decompositions have been widely used in EEG analysis in various applications, including but not limited to multi-modal EEG and functional magnetic resonance imaging (fMRI) analysis (Mørup *et al* 2008, Eliseyev and Aksenova 2013), brain-computer interfaces (BCIs) (Cichocki *et al* 2008), seizure detection and localization (De Vos *et al* 2007, Deburchgraeve *et al* 2009), and ERP analysis (Vanderperren *et al* 2013, Idaji *et al* 2017, Maki *et al* 2018, Wang *et al* 2018, Zhang *et al* 2020). Different tensor decompositions, such as canonical polyadic (CP), also known as parallel factor analysis (PARAFAC) (Mørup *et al* 2006, 2008, Eliseyev and Aksenova 2013), Tucker decomposition (Latchoumane *et al* 2012, Cong *et al* 2013), Non-negative Tucker decomposition (NTD) (Dao *et al* 2020, Rošťáková *et al* 2020, Wang *et al* 2020) , and their variations (Zhao *et al* 2012, Cichocki *et al* 2015, Idaji *et al* 2017), have been explored in these studies. While much research in the field of ERP analysis has focused on extracting features for pattern recognition models, to the best of our knowledge, no previous work has proposed a new tensor decomposition specifically designed for robust single-trial ERP estimation from raw EEG recordings in low signal-to-noise ratio (SNR) conditions. The flexibility offered by tensors in handling multi-dimensional data may be beneficial in addressing the current limitations in single-trial ERP estimation. A new approach, which can benefit from multi-dimensional flexibility of tensors may be useful to address the current gaps in single-trial ERP estimation.

The contribution of the paper is to propose a customized tensor decomposition by including a constraint, which is inspired by spatio-temporal filtering models, to the conventional Tucker decomposition. In this regard, the notations, original Tucker decomposition, Non-negative Tucker decomposition, and our proposed method (called ETucker) are discussed in section 2, then the performance of the proposed method on a physiologically meaningful synthesized EEG dataset as well as on a real-world dataset is validated in section 3. And some conclusions are drawn in section 4.

2. Methods

2.1. Notations

Before bringing up the details of the proposed decomposition, we need to demonstrate the assumptions and notations. We use the same notations and definitions used by Kolda and Bader (Kolda and Bader 2009). Therefore, a tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ is characterized as an Nth dimensional array, also called an N-way tensor. Boldface calligraphic letters, e.g. $\mathcal{X}$ symbolizes tensors of dimension 3 or greater; bold capital letters like $\mathbf{X}$ symbolize matrices; bold letters in lower-case, e.g. $\mathbf{x}$ symbolize vectors; and lowercase letters, e.g. x denote scalars. '$\otimes$' and '$\circ$' denote the Kronecker and outer products, respectively, and the n-mode tensor-times-matrix product of $\mathcal{X}$ and the matrix $\mathbf{M} \in \mathbb{R}^{J \times I_n}$ is indicated by $\mathcal{X} \times_n \mathbf{M} \in \mathbb{R}^{I_1 \times \cdots \times I_{n-1} \times J \times I_{n+1} \times I_N}$. Also, $\mathbf{X}_{(n)} \in \mathbb{R}^{I_n \times M}$ denotes the mode-n unfolding of the tensor $\mathcal{X}$, where $M = \prod_{i=1, i \neq n}^{N} I_i$ and is defined by mapping from element $(i_1, i_2, \cdots, i_N)$ to (i_n, j) where $j = \prod_{k=1, k \neq n}^{N} i_k \times \prod_{m=k+1, k \neq n}^{N} I_m$. The transpose of matrix $\mathbf{X} \in \mathbb{R}^{I_1 \times I_2}$ is denoted as $\mathbf{X}^\top \in \mathbb{R}^{I_2 \times I_1}$, and the pseudo inverse of matrix $\mathbf{X}$ is represented as $\mathbf{X}^\dagger \in \mathbb{R}^{I_2 \times I_1}$ and can be calculated either as $(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top$ or $\mathbf{X}^\top (\mathbf{X}\mathbf{X}^\top)^{-1}$ according to the rank of matrix $\mathbf{X}$.

2.2. Tucker decomposition

The Tucker decomposition of a tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ is

$$\begin{aligned} \mathcal{X} &= \mathcal{G} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \times_3 \mathbf{U}^{(3)} =: [[\mathcal{G}; \mathbf{U}^{(1)}, \mathbf{U}^{(2)}, \mathbf{U}^{(3)}]] \\ &= \sum_{p=1}^{R_1} \sum_{q=1}^{R_2} \sum_{r=1}^{R_3} g_{pqr} \mathbf{u}_p^{(1)} \circ \mathbf{u}_q^{(2)} \circ \mathbf{u}_r^{(3)}, \end{aligned} \tag{1}$$

where $\mathcal{G} \in \mathbb{R}^{R_1 \times R_2 \times R_3}$ is called the core tensor, and each $\mathbf{U}^{(i)} \in \mathbb{R}^{I_i \times R_i}$ is called factor matrix. The factor matrices are assumed to have orthonormal columns (Malik and Becker 2018). Also, the tensor in (1) is a rank-(R_1, R_2, R_3) tensor. Tucker decomposition can also be transformed to an optimization problem using the cost function as follows:

$$\min_{\mathcal{G}, \mathbf{U}^{(1)}, \mathbf{U}^{(2)}, \mathbf{U}^{(3)}} \left\| \mathcal{X} - \mathcal{G} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \times_3 \mathbf{U}^{(3)} \right\|_F^2, \tag{2}$$

where $\|.\|_F^2$ indicates the Frobenius norm of a tensor. Throughout this paper, it is assumed that the tensor $\mathcal{X} \in \mathrm{IR}^{C\times T\times 2}$ corresponds to a single trial EEG epoch, and accordingly $\mathbf{U}^{(1)}$ and $\mathbf{U}^{(2)}$ and $\mathbf{U}^{(3)}$ contain information regarding spatial dispersal, temporal dynamics, and contribution of the estimated sources in reconstructing each slice of the tensor, respectively.

2.3. Non-negative Tucker decomposition

The NTD is a variant of Tucker decomposition where the core tensor ($\mathcal{G}$) and factor matrices ($\mathbf{U}^{(i)}$) are constrained to have only non-negative elements (Kim and Choi 2007). This non-negativity constraint ensures that all elements in the core tensor and factor matrices are non-negative, promoting sparsity and interpretability of the factors.

The non-negativity constraint ($\mathcal{G},\ \mathbf{U}^{(i)} \geqslant 0$ for $i = 1, 2, 3$) is incorporated in (2), and the problem is solved through an optimization process, which are commonly projected gradient methods or alternating non-negative least squares algorithms (Kim and Choi 2007). In this paper, we utilize NTD to compare the performance of our model in cases with positive peaks in the synthesized dataset. In contrast to all the algorithms implemented in MATLAB for this study, NTD was implemented using the Tensorly library in Python (Kossaifi *et al* 2018).

2.4. Constrained Tucker decomposition for ERP extraction

Though capable of source separation, Tucker decomposition fails to extract sources properly in low SNR conditions (Cichocki *et al* 2009, Zhou and Cichocki 2012). In such cases, adding *a priori* information regarding the sources can enhance not only the convergence possibility of the algorithm but also the accuracy of estimated sources (Fonał and Zdunek 2019, Chen *et al* 2021).

In spatio-temporal filtering-based papers (Li *et al* 2009, Jarchi *et al* 2010, Ranjbar *et al* 2018), it is assumed that the ERP component can be extracted by applying a spatial filter ($\mathbf{w} \in \mathrm{IR}^{C}$) to the observation matrix ($\mathbf{X} \in \mathrm{IR}^{C\times T}$) and comparing the result with a pre-assumed waveform with different delays ($\mathbf{g}_0(\tau)$), which can be formulated as the following problem:

$$\min_{\mathbf{w}} \ \|\mathbf{w}^\top\mathbf{X} - \mathbf{g}_0(\tau)\|_2^2, \tag{3}$$

where C is the number of channels, T is the number of temporal samples, and $\mathbf{g}_0(\tau)$ is the τ-sample shifted gamma function. The original gamma function is defined as:

$$g(t) = ct^{K-1}\exp\left(\frac{-t}{\theta}\right), \tag{4}$$

where c and K and θ are amplitude, bandwidth, and peak of the waveform, respectively. Similar assumptions can likewise be made on the tensor's factor matrices. In this respect, the cost function in (5) is defined so that the estimated sources in the second factor matrix of the tensor form an analogous waveform to the gamma function ($\mathbf{r}(t)$ is the vectorized notion of $g(t)$).

$$\min_{\mathbf{w},\mathbf{U}^{(2)},\tau} \ \|\mathbf{w}^\top\mathbf{U}^{(2)} - \mathbf{r}(\tau)\|_2^2. \tag{5}$$

Since (2) and (5) should be optimized concurrently, the second equation is appended in (2) as a constraint using Lagrange multiplier (λ). As a result, the Tucker decomposition in (2) is transformed to the following problem:

$$\begin{aligned}\min_{\mathcal{G},\mathbf{U}^{(1)},\mathbf{U}^{(2)},\mathbf{U}^{(3)},\mathbf{w},\tau} \ &\|\mathcal{X} - \mathcal{G}\times_1\mathbf{U}^{(1)}\times_2\mathbf{U}^{(2)}\times_3\mathbf{U}^{(3)}\|_F^2 + \\ &\lambda\,\|\mathbf{w}^\top\mathbf{U}^{(2)} - r(\tau)\|_2^2,\end{aligned} \tag{6}$$

where core tensor, factor matrices, filter coefficients (w_i), and delay (τ) are target parameters and should be estimated through an optimization process. To this end, the objective function in (6) is split into three objective functions as follows:

$$\begin{aligned}J^{(1)} &= \|\mathbf{X}_{(1)} - \mathbf{U}^{(1)}\mathbf{C}^{(1)}\|_F^2 \\ J^{(2)} &= \|\mathbf{X}_{(2)} - \mathbf{U}^{(2)}\mathbf{C}^{(2)}\|_F^2 + \lambda\,\|\mathbf{w}^\top\mathbf{U}^{(2)} - \mathbf{r}(\tau)\|_2^2 \\ J^{(3)} &= \|\mathbf{X}_{(3)} - \mathbf{U}^{(3)}\mathbf{C}^{(3)}\|_F^2,\end{aligned} \tag{7}$$

where $\|.\|_F^2$ denotes Frobenius norm of the matrix, and $\mathbf{C}^{(i)}$s are defined as follows:

$$\begin{aligned}\mathbf{C}^{(1)} &= \mathbf{G}_{(1)}(\mathbf{U}^{(3)}\otimes\mathbf{U}^{(2)})^\top \\ \mathbf{C}^{(2)} &= \mathbf{G}_{(2)}(\mathbf{U}^{(3)}\otimes\mathbf{U}^{(1)})^\top \\ \mathbf{C}^{(3)} &= \mathbf{G}_{(3)}(\mathbf{U}^{(2)}\otimes\mathbf{U}^{(1)})^\top.\end{aligned} \tag{8}$$

To reduce the risk of divergence, gradient descent method was employed to estimate the target parameters. The first-order derivatives with respect to the factor matrices and the filter are demonstrated below (Akbari *et al* 2015):

$$\begin{aligned}
\frac{\partial J^{(1)}}{\partial \mathbf{U}^{(1)}} &= -2\mathbf{X}_{(1)}\mathbf{C}^{(1)\top} + 2\mathbf{U}^{(1)}\mathbf{C}^{(1)}\mathbf{C}^{(1)\top} \\
\frac{\partial J^{(2)}}{\partial \mathbf{U}^{(2)}} &= -2\mathbf{X}_{(2)}\mathbf{C}^{(2)\top} + 2\mathbf{U}^{(2)}\mathbf{C}^{(2)}\mathbf{C}^{(2)\top} + \\
&\quad \lambda(2\mathbf{U}^{(2)}\mathbf{w}^{\top}\mathbf{w} - 2\mathbf{r}(\tau)^{\top}\mathbf{w}) \\
\frac{\partial J^{(3)}}{\partial \mathbf{U}^{(3)}} &= -2\mathbf{X}_{(3)}\mathbf{C}^{(3)\top} + 2\mathbf{U}^{(3)}\mathbf{C}^{(3)}\mathbf{C}^{(3)\top} \\
\frac{\partial J^{(2)}}{\partial \mathbf{w}} &= \lambda(2\mathbf{U}^{(2)\top}\mathbf{U}^{(2)}\mathbf{w}^{\top} - 2\mathbf{U}^{(2)\top}\mathbf{r}(\tau)^{\top}).
\end{aligned} \tag{9}$$

The parameters are initialized using algorithm 1, where R_i specifies the expected rank of the tensor and $f(.)$ denotes any transformation or algorithm that can be used to extract ERP components in high SNR conditions, e.g. spatio-temporal, ICA, or PCA-based methods (MacDonald and Barry 2014, Lee *et al* 2016, Ranjbar *et al* 2020, Zang *et al* 2022). Subsequently, the factor matrices and filter are updated recursively using (10) until the algorithm converges (the difference between the norm of two successive estimated filtering vectors, namely **w**, is below 0.0001) or reaches the maximum number of iterations (1000).

Algorithm 1. Initialization algorithm for factor matrices and the core tensor

Require: $\mathcal{X}, R_1, R_2, R_3$
Ensure: $\mathcal{G}$, $\mathbf{U}^{(1)}$, $\mathbf{U}^{(2)}$, $\mathbf{U}^{(3)}$, $\mathbf{w}$
1: $\mathbf{U}^{(1)} \leftarrow R_1$ leading left singular vector of $\mathbf{X}_{(1)}$
2: $\mathbf{U}^{(2)} \leftarrow R_2$ leading ERP components of $f(\mathbf{X}_{(2)})$
3: $\mathbf{U}^{(3)} \leftarrow R_3$ leading left singular vector of $\mathbf{X}_{(3)}$
4: $\mathbf{w} \leftarrow \mathbf{0.5} \in \mathrm{IR}^{R_2}$
5: $\mathcal{G} = \mathcal{X} \times_1 \mathbf{U}^{(1)^{\dagger}} \times_2 \mathbf{U}^{(2)^{\dagger}} \times_3 \mathbf{U}^{(3)^{\dagger}}$
6: **return** $\mathcal{G}$, $\mathbf{U}^{(1)}$, $\mathbf{U}^{(2)}$, $\mathbf{U}^{(3)}$, $\mathbf{w}$

$$\begin{aligned}
\mathbf{U}^{(n)}_{k+1} &= \mathbf{U}^{(n)}_{k} - \gamma \frac{\partial J^{(n)}}{\partial \mathbf{U}^{(n)}}\Bigg|_{\mathbf{U}^{(n)}=\mathbf{U}^{(n)}_{k}} \\
\mathbf{w}_{k+1} &= \mathbf{w}_{k} - \gamma \frac{\partial J^{(2)}}{\partial \mathbf{w}}\Bigg|_{\mathbf{w}=\mathbf{w}_{k}},
\end{aligned} \tag{10}$$

where γ is the learning rate (0.001 in this paper). The proposed method is summarized in algorithm 2, where T_s is the suspected range of the target ERP component (0.25s–0.65s in this work), **s** is the extracted ERP waveform, and **a** is the spatial vector of the extracted component.

Algorithm 2. ETucker

Require: $\mathcal{X}, R_1, R_2, R_3, \lambda, r(t), T_s$
Ensure: $\mathcal{G}$, $\mathbf{U}^{(1)}$, $\mathbf{U}^{(2)}$, $\mathbf{U}^{(3)}$, $\mathbf{w}$, $\mathbf{s}$, $\mathbf{a}$.
1: Initialize $\mathbf{U}^{(i)}$ and $\mathcal{G}$ using algorithm 1.
2: $z \leftarrow 1$
3: **for** Every τ in range of T_s **do**
4: $k \leftarrow 1$
5: **while:** *not* converged OR *not* reached maximum number of iterations **do**
6: Update $\mathbf{U}^{(i)}$ for $i = 1, 2, 3$ using (10)
7: Update $\mathbf{w}$ using (10)
8: $\mathcal{G} = \mathcal{X} \times_1 \mathbf{U}^{(1)^{\dagger}} \times_2 \mathbf{U}^{(2)^{\dagger}} \times_3 \mathbf{U}^{(3)^{\dagger}}$
9: $k \leftarrow k + 1$
10: **end while**
11: $J(z) :=$ value of cost function in (6)
12: $z \leftarrow z + 1$
13: **end for**
14: $L = \arg\min_{\tau} J(\tau) \quad \tau \in [0, T_s]$
15: $\mathbf{r} = \mathbf{r}(L)$ and repeat 6 to 11
16: $\mathbf{s} = \mathbf{w}^{\top}\mathbf{U}^{(2)}$
17: $j = \arg\max_{\alpha} w_{\alpha} \quad \alpha = 1, 2, \cdots, R_2$
18: $i = \arg\max_{\beta} \mathcal{G}_{\beta,j,1} \quad \beta = 1, 2, \cdots, R_1$
19: $\mathbf{a} = \mathbf{U}^{(1)}_{:,i}$
20: **return** $\mathcal{G}$, $\mathbf{U}^{(1)}$, $\mathbf{U}^{(2)}$, $\mathbf{U}^{(3)}$, $\mathbf{w}$, $\mathbf{s}$, $\mathbf{a}$

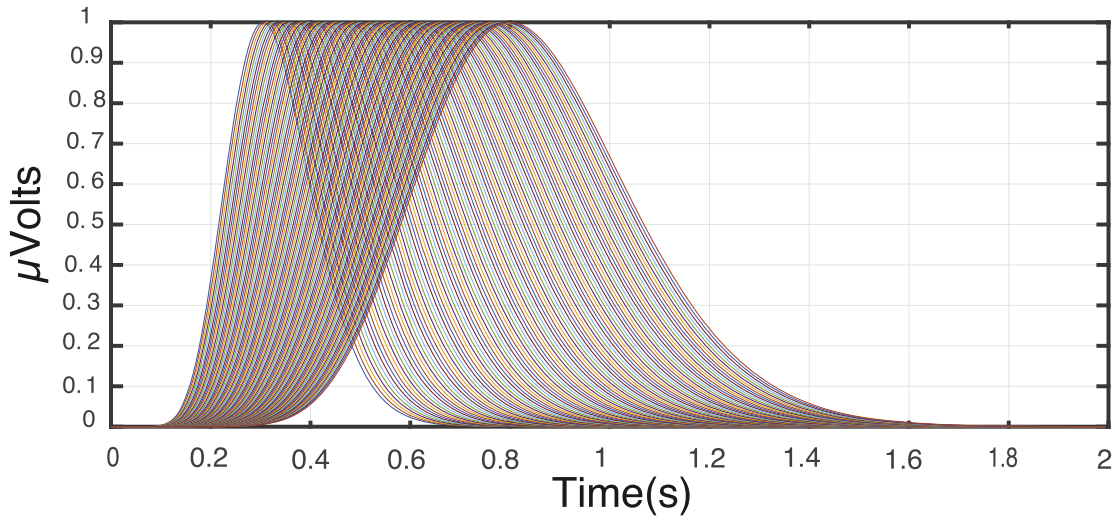

Figure 1. Temporal distribution of generated P300 components through all the trials.

3. Experiments

3.1. Synthesizing EEG recordings

Since there is no gold standard to compare the estimated ERP component and the original one directly, previous research has used synthesizing methods to generate EEG trials containing target ERP (sub)component(s) (Li *et al* 2009, Jarchi *et al* 2010, Lee *et al* 2016, Ranjbar *et al* 2020). Such studies have assumed that the background EEG activity is similar to the white noise, and accordingly used the Gaussian probability density function to generate the temporal signals. Also, uniform distribution has been used to generate the spatial vector of the ERP component. However, this approach does not fully act following the frequency-domain characteristics of the EEG signals such as hyperbolic $(1/f)$ morphology of the power spectrum distribution. Therefore, in this work, an approach was developed to utilize real EEG recordings and simulate the synthetic recordings for further performance validation of the proposed decomposition. Accordingly, ICA was used to extract sources of a dataset recording no-task EEG. The dataset had been originally recorded with closed-eyes using 19 channels and a sampling frequency of 200 Hz to simulate and identify the effect of electrooculogram artifacts on EEG signals (Klados and Bamidis 2016). In this research, an autoregressive (AR) model was used to learn the behaviour of estimated sources from ICA. The model is characterized as follows:

$$x(n) = -\sum_{k=1}^{q} w_k x(n-k) + u(n), \tag{11}$$

where $u(n)$ is the input of the system and is usually considered as zero-mean Gaussian white noise (GWN), $x(n)$ is the observation sequence, and w_k, $1 \leqslant k \leqslant q$ is the AR parameter with q being the order of the AR model (Li *et al* 2015). For training the proposed AR model, q was set to 12, and the Yule-Walker algorithm was used to find the AR parameters (w_k). Using the AR model and white Gaussian noise as the input, multiple meaningful EEG sources can be generated. Regarding the target ERP component, which is P300 in this article, the parameters of the gamma function are set as $c = 1$, $K = 14$, and $\theta = \frac{0.3}{K-1}$. Furthermore, to simulate the jitter effect, the values of K and θ are incremented by 0.02 and 0.005 in each iteration over 100 trials, respectively. The sampling frequency is 200 Hz and the duration of each trial is 2 s. Simulated ERP components are illustrated in figure 1. To project the sources in a meaningful manner, the Brainstorm toolbox was used, and the sources' positions were manually located. The scalp and skull and brain tissue conductivities were set to 0.33, 0.004, and 0.33 μs cm^{-1}, respectively. The position of the corresponding source was also located around C_z channel location (Linden 2005). The projection of sources from the source space to the observation space was done using $\mathbf{X} = \sum_{i=1}^{N_s} \mathbf{a}_i^T \mathbf{s}_i$, where $\mathbf{X}$ is the observation matrix, N_s is the number of sources or 19 (in this article), $\mathbf{a}_i$ is the spatial vector corresponding to the ith source, and $\mathbf{s}_i$ is the ith source. In figure 2, the power spectral density (PSD) of the generated recordings using the proposed synthesis method and white noise are shown, and it can be

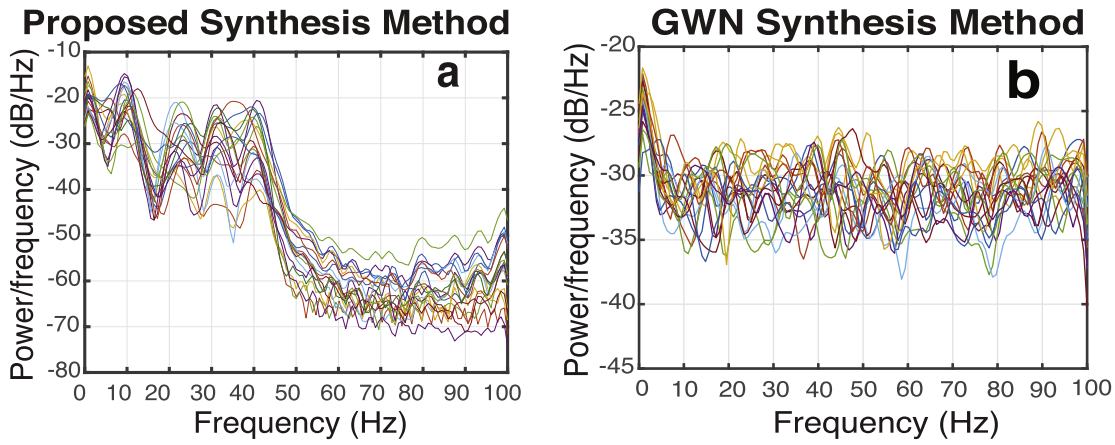

Figure 2. PSD comparison (over channels) between two different methods for generating ERP-included EEG recordings; (a) the proposed synthesis method and (b) GWN-based synthesis method.

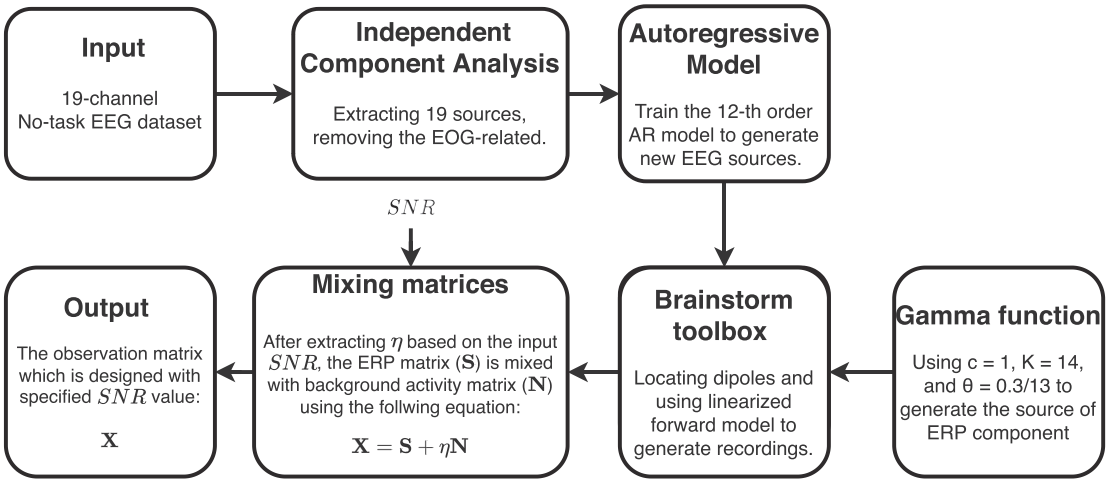

Figure 3. The proposed EEG synthesis flowchart.

conceivable that the proposed method could retrieve more information in all frequency bands; moreover, it was able to preserve the hyperbolic shape of the EEG PSD.

As the performance of the methods should be evaluated in different SNR conditions, the matrix corresponding to the reconstructed ERP component (**S**) and the matrix of noise signals or background EEG recording (**N**) should be merged using different coefficients, $\mathbf{X} = \mathbf{S} + \eta\mathbf{N}$, where η is calculated as follows:

$$\eta = \frac{\|\mathbf{S}\|_F^2}{10^{\frac{SNR}{10}} * \|\mathbf{N}\|_F^2}. \tag{12}$$

The EEG synthesizing algorithm is illustrated in figure 3.

3.2. Real data

To evaluate the effectiveness of the proposed algorithm in real-world scenarios, we utilized EEG signals from BCI Competition III (2004)-Dataset II (Blankertz *et al* 2004, 2006). The dataset consisted of participants engaging in five sessions of several runs, where they were tasked with spelling a series of characters by observing a 6×6 speller matrix. During each character presentation, the matrix was displayed for a duration of 2.5 s with uniform intensity (i.e. blank matrix). Subsequently, individual rows and columns of the matrix were randomly intensified for 100 ms, resulting in 12 different stimuli. After each intensification, there was a 75 ms blank period. The order of row/column intensifications was randomized within blocks of 12, and each set of 12 intensifications was repeated 15 times for each character, resulting in a total of 180 intensifications per character. Following each sequence of 15 sets of intensifications, there was another 2.5 s blank period.

Since the target character was repeated in two out of the 12 rows/columns, the experimental paradigm was considered as an odd-ball paradigm, with the intensification of the target row/column expected to elicit a P300 component. EEG signals were recorded using a 64-channel recording scheme at a sampling frequency of 240 Hz.

For this study, we utilized the training set of 'Subject A,' which consisted of 85 epochs (characters). To focus specifically on the time range where P300 components are predominantly observed, we extracted the signals from the onset of stimuli to 650 ms after the onset. Subsequently, an 8th-order Chebyshev bandpass filter with a passband of 0.1–10 Hz was applied to remove baseline drift and high-frequency artifacts from the signals. To estimate the P300 component, we only considered the first repetition (trial) out of the 15 trials available for each character. The estimated component was then correlated with the average of the 15 trials for each character. To demonstrate the significance of our work, we applied the proposed algorithm to both target trials (containing P300 components) and non-target trials (without P300 components), allowing for a comprehensive evaluation of its performance.

3.3. Metrics

To compare the estimated and the original ERP component in the synthesized EEG dataset directly, four different metrics have been proposed; (1) delay, which is the difference between the index of the maximum amplitude in estimated and the original waveform, (2) relative amplitude, which is the ratio of maximum amplitude of estimated component to the original one, (3) temporal correlation, which is the correlation between the estimated waveform and the original one, and (4) spatial correlation, which is the correlation between spatial vectors.

3.4. Tensorization

Making tensors from EEG recording matrix can be done in several ways, e.g. concatenating recording of different epochs in the third mode of tensor, using time-frequency representation such as wavelet, or using transforms that have been developed for ERP extraction and adding the reconstructed component as the second slice of the tensor. In this work, though exploring all approaches, the third family has shown better performance, and accordingly were chosen for presentation and comparison of the results. Temporal PCA (TPCA), and ICA were used as the blind source separation methods in which sources are extracted from the observation matrix (Jung *et al* 1999, Kayser and Tenke 2003, Bernat *et al* 2005). Then, the source that is more correlated with the standard gamma function is selected as the candidate source, and projected to the observation space using the corresponding spatial vector. Thereafter, the reconstructed matrix is placed in the second slice of the third mode of tensor. The implementation of the mentioned methods has been done using ERP PCA Toolkit and ICALAB (Cichocki 2002, Dien 2010). The other method is based on spatio-temporal filtering, called STF, which estimates the source and the spatial vector utilizing several trials (Li *et al* 2009). Afterwards, the source is projected to observation space using the forward model ($\mathbf{a}^T\mathbf{s}$) and the reconstructed matrix is placed in the third mode of the tensor.

Regarding the rank estimation, we employed the modified eigenvalues estimator for Tucker rank determination (MEET) algorithm (Yokota *et al* 2016), to estimate the rank of the first mode (R_1). After careful consideration and for the sake of consistency among all SNRs, we set the first mode rank to be 8 and the third mode rank (R_3) to be 2. Additionally, we intuitively set the temporal rank, R_2, as 3 due to the presence of one ERP source, one physiologically meaningful EEG component, and one non-physiological source (e.g. noise).

3.5. Results

3.5.1. Synthesized EEG

The comparison of the performance of the proposed method using different tensorization approaches is shown in figure 4. It can be inferred that integrating the spatio-temporal filtering method (STF) with the ETucker decomposition outperformed the other approaches. What can be clearly seen in this figure is the robustness of the algorithm in low SNR (high noise) conditions. However, almost every approach failed to estimate the spatial vector as the SNR gets worse (less than −20 dB).

As a tensor decomposition, Tucker has been shown as a source separation algorithm in which each vector of the factor matrix can be considered as the source basis. In table 1, the performance of Tucker decomposition and ETucker, which was exclusively extended for single-trial ERP extraction application, were compared with the reference algorithms that were used initially for generating tensors. The table also shows that there has been a sharp drop in the performance of Tucker in low SNR conditions as if the Tucker is unable to distinguish source and background noise.

As table 1 illustrates, utilizing ICA alone can accurately estimate the temporal occurrence and morphology of the target component across all SNR conditions. However, it is found to be inadequate in determining the peak amplitude and spatial vector of the component. The proposed method, ETucker, in conjunction with ICA,

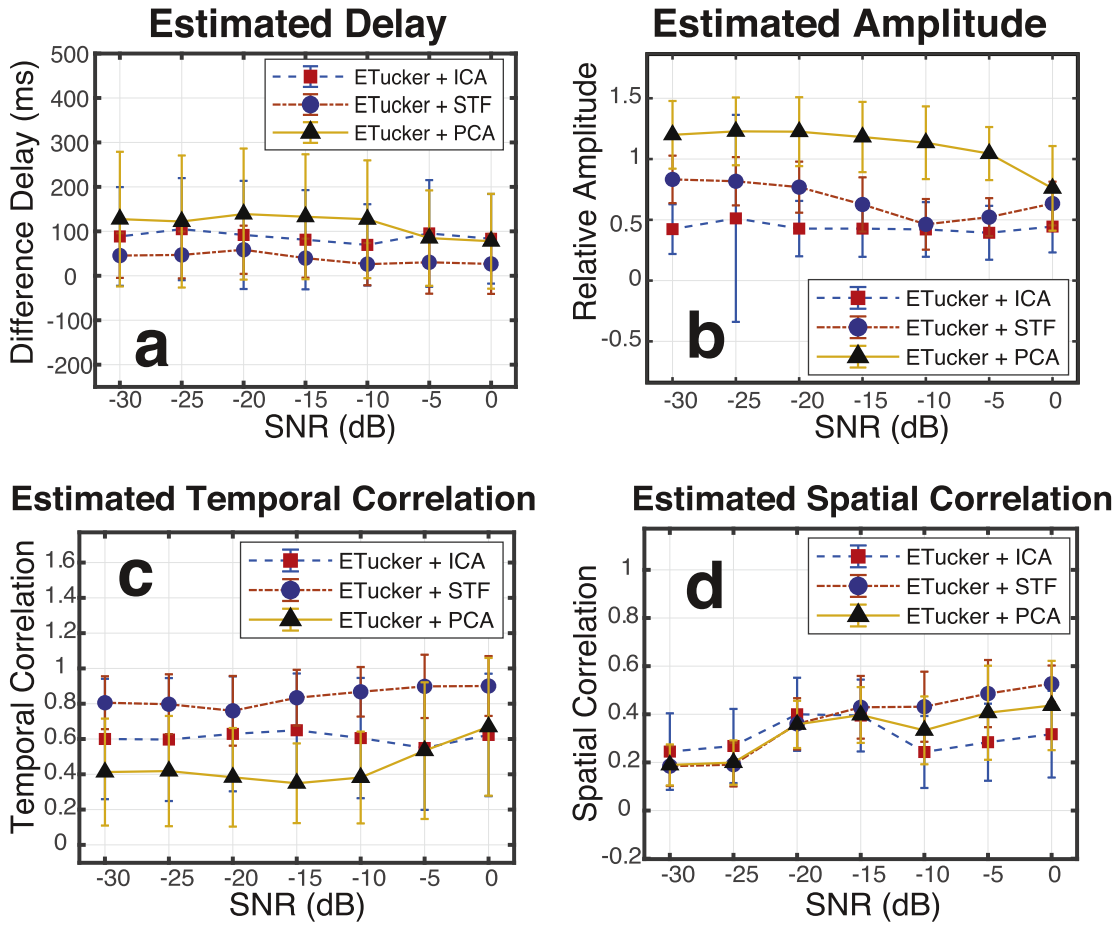

Figure 4. The comparison of different tensorization approaches (where PCA represents TPCA) using different metrics; (a) shows difference in estimated delay, (b) demonstrates the estimated relative amplitude, (c) and (d) show the temporal and spatial correlation between estimated and original ERP component.

significantly improves the peak detection and peak onset estimation, while the spatial distribution estimation remains inferior. A comparison of the proposed method with the Tucker decomposition and ICA alone, reveals that ETucker demonstrates superior performance in almost every SNR condition, particularly in low SNR or extremely noisy conditions. The results of applying STF on our synthesized dataset suggest that the method demonstrates reliable performance in high SNR conditions. However, it is found to be inadequate in the estimation of amplitude and peak occurrence in lower SNR conditions. These findings align with the results of the original paper (Li *et al* 2008, 2009). Furthermore, the results indicate that the use of the proposed ETucker method, in conjunction with STF, not only improves the performance of amplitude estimation, but also enhances component occurrence detection and temporal correlation. It is important to note, however, that the spatial correlation remains inadequate in high SNR conditions for all methods. Additionally, the findings of our study indicate that the Tucker decomposition built upon the STF method is ineffective in extracting the target component, particularly in high SNR levels where the component is almost diminished.

Furthermore, table 1 suggests that TPCA method performs relatively acceptable in estimating the target component at zero SNR level. However, as the SNR decreases, the temporal occurrence and correlation of the estimation deteriorate significantly. To address this limitation, we propose the use of ETucker in conjunction with TPCA. The outcomes imply that this approach improves the delay, temporal correlation, and amplitude estimation of the target component with a high level of significance. Despite this, it is important to note that all methods, including TPCA, exhibit limitations in accurately estimating the spatial vectors of the target component in noisy conditions. This is a common limitation among all methods, regardless of the use of Tucker or ETucker decomposition.

In addition to Tucker decomposition, we also compared our proposed method with NTD, specifically targeting the peak corresponding to our intended ERP component, which is the P300 characterized by a positive peak. Table 2 presents the results of this comparison. The temporal estimation performance of NTD was found to be superior to Tucker decomposition, indicating its ability to capture the temporal dynamics of the P300 component more accurately. However, it is important to note that the spatial precision of NTD was poor. This

Table 1. The comparison between the proposed tensor decomposition and the Tucker decomposition using different tensorization methods with the methods without tensorizations in different SNR conditions. The numbers are reported in the form of mean ± standard deviation over 100 trials.

Tensor	SNR (dB)	Proposed method (ETucker)				Tucker				No Tensorization			
		Delay	Peak	Temporal	Spatial	Delay	Peak	Temporal	Spatial	Delay	Peak	Temporal	Spatial
ICA (Jung *et al* 1999)	0	83.20 ± 100.79	0.44 ± 0.21	0.62 ± 0.35	0.27 ± 0.18	76.85 ± 96.06	0.15 ± 0.02	0.67 ± 0.29	0.34 ± 0.17	85.25 ± 110.50	3.09 ± 0.51	0.60 ± 0.32	0.22 ± 0.14
	−5	95.25 ± 120.02	0.39 ± 0.22	0.55 ± 0.35	0.23 ± 0.16	106.95 ± 114.49	0.14 ± 0.02	0.49 ± 0.32	0.26 ± 0.15	90.95 ± 111.25	3.20 ± 0.51	0.58 ± 0.33	0.21 ± 0.14
	−10	69.50 ± 91.46	0.42 ± 0.22	0.61 ± 0.34	0.19 ± 0.15	138.35 ± 145.12	0.14 ± 0.02	0.44 ± 0.31	0.24 ± 0.16	92.05 ± 105.60	3.15 ± 0.53	0.60 ± 0.33	0.20 ± 0.14
	−15	81.10 ± 111.62	0.43 ± 0.23	0.65 ± 0.32	0.19 ± 0.15	107.80 ± 115.72	0.14 ± 0.02	0.48 ± 0.34	0.19 ± 0.13	102.40 ± 126.04	3.10 ± 0.55	0.55 ± 0.32	0.24 ± 0.16
	−20	91.90 ± 121.68	0.43 ± 0.23	0.63 ± 0.33	0.20 ± 0.15	113.90 ± 129.51	0.14 ± 0.02	0.45 ± 0.32	0.18 ± 0.14	90.95 ± 113.77	3.12 ± 0.52	0.59 ± 0.30	0.21 ± 0.15
	−25	104.90 ± 114.84	0.51 ± 0.85	0.60 ± 0.35	0.22 ± 0.15	107.25 ± 116.38	0.14 ± 0.02	0.45 ± 0.29	0.21 ± 0.15	112.80 ± 128.21	3.15 ± 0.58	0.55 ± 0.33	0.22 ± 0.16
	−30	88.50 ± 111.02	0.42 ± 0.20	0.60 ± 0.34	0.20 ± 0.16	131.55 ± 138.32	0.14 ± 0.02	0.46 ± 0.32	0.18 ± 0.13	119.40 ± 130.37	3.15 ± 0.51	0.55 ± 0.36	0.20 ± 0.15
STF (Li *et al* 2008, 2009)	0	26.65 ± 67.35	0.63 ± 0.18	0.90 ± 0.17	0.48 ± 0.08	121.30 ± 148.54	0.18 ± 0.03	0.58 ± 0.41	0.30 ± 0.21	87.10 ± 254.12	0.74 ± 0.18	0.84 ± 0.24	0.45 ± 0.08
	−5	30.35 ± 70.60	0.52 ± 0.16	0.90 ± 0.18	0.44 ± 0.14	70.90 ± 98.31	0.17 ± 0.02	0.71 ± 0.36	0.39 ± 0.16	91.95 ± 258.96	0.55 ± 0.17	0.85 ± 0.25	0.40 ± 0.04
	−10	26.20 ± 47.50	0.46 ± 0.21	0.87 ± 0.14	0.38 ± 0.15	79.75 ± 111.84	0.16 ± 0.02	0.54 ± 0.28	0.34 ± 0.14	96.18 ± 263.20	0.36 ± 0.17	0.75 ± 0.15	0.39 ± 0.07
	−15	39.70 ± 43.89	0.63 ± 0.22	0.83 ± 0.16	0.23 ± 0.13	193.00 ± 165.52	0.15 ± 0.02	0.19 ± 0.19	0.18 ± 0.12	99.95 ± 266.97	0.23 ± 0.16	0.79 ± 0.19	0.37 ± 0.13
	−20	58.55 ± 54.20	0.77 ± 0.21	0.76 ± 0.20	0.16 ± 0.11	207.90 ± 160.76	0.15 ± 0.02	0.13 ± 0.11	0.14 ± 0.11	103.35 ± 270.37	0.17 ± 0.14	0.77 ± 0.18	0.28 ± 0.16
	−25	46.95 ± 52.80	0.82 ± 0.20	0.80 ± 0.17	0.14 ± 0.09	231.85 ± 158.19	0.15 ± 0.02	0.11 ± 0.11	0.12 ± 0.09	106.45 ± 273.47	0.14 ± 0.13	0.70 ± 0.10	0.16 ± 0.14
	−30	45.50 ± 50.25	0.83 ± 0.20	0.81 ± 0.15	0.13 ± 0.08	218.80 ± 163.39	0.15 ± 0.02	0.12 ± 0.10	0.12 ± 0.09	109.30 ± 276.31	0.12 ± 0.12	0.71 ± 0.12	0.09 ± 0.15
TPCA (Bernat *et al* 2005)	0	77.75 ± 106.71	0.76 ± 0.35	0.67 ± 0.39	0.39 ± 0.19	76.75 ± 113.44	0.15 ± 0.02	0.77 ± 0.37	0.39 ± 0.18	52.75 ± 69.7	0.85 ± 0.27	0.40 ± 0.03	0.37 ± 0.15
	−5	84.80 ± 107.28	1.05 ± 0.22	0.53 ± 0.39	0.36 ± 0.20	88.25 ± 110.28	0.15 ± 0.02	0.68 ± 0.34	0.39 ± 0.17	105.20 ± 175.3	0.77 ± 0.26	0.36 ± 0.04	0.36 ± 0.14
	−10	127.10 ± 132.55	1.13 ± 0.30	0.38 ± 0.26	0.28 ± 0.14	110.45 ± 114.05	0.15 ± 0.02	0.35 ± 0.22	0.27 ± 0.14	331.80 ± 395.9	0.60 ± 0.32	0.26 ± 0.05	0.35 ± 0.16
	−15	132.80 ± 140.52	1.18 ± 0.29	0.35 ± 0.23	0.20 ± 0.12	193.70 ± 164.24	0.15 ± 0.02	0.13 ± 0.12	0.17 ± 0.12	732.50 ± 389.8	0.40 ± 0.35	0.11 ± 0.07	0.21 ± 0.14
	−20	138.85 ± 147.45	1.23 ± 0.28	0.38 ± 0.28	0.16 ± 0.10	212.35 ± 163.11	0.15 ± 0.02	0.11 ± 0.08	0.14 ± 0.10	822.35 ± 371.8	0.42 ± 0.36	0.06 ± 0.05	0.20 ± 0.13
	−25	122.10 ± 148.53	1.23 ± 0.28	0.42 ± 0.31	0.15 ± 0.09	221.55 ± 169.94	0.15 ± 0.02	0.10 ± 0.08	0.13 ± 0.10	848.60 ± 376.5	0.42 ± 0.36	0.05 ± 0.04	0.20 ± 0.13
	−30	127.45 ± 151.52	1.20 ± 0.28	0.41 ± 0.30	0.14 ± 0.09	225.95 ± 169.84	0.15 ± 0.02	0.10 ± 0.08	0.13 ± 0.10	867.05 ± 363.8	0.42 ± 0.36	0.04 ± 0.04	0.20 ± 0.13

Table 2. The comparison between ETucker, Tucker decomposition, and NTD using STF tensorization method in different SNR conditions. The numbers are reported in the form of mean ± standard deviation over 100 trials.

Tensor	SNR (dB)	Proposed method (ETucker)				Tucker				Non-negative Tucker (Kim and Choi 2007)			
		Delay	Peak	Temporal	Spatial	Delay	Peak	Temporal	Spatial	Delay	Peak	Temporal	Spatial
STF (Li *et al* 2008, 2009)	0	26.65 ± 67.35	0.63 ± 0.18	0.90 ± 0.17	0.48 ± 0.08	121.30 ± 148.54	0.18 ± 0.03	0.58 ± 0.41	0.30 ± 0.21	23.75 ± 20.3	1.38 ± 0.21	0.47 ± 0.28	0.08 ± 0.01
	−5	30.35 ± 70.60	0.52 ± 0.16	0.90 ± 0.18	0.44 ± 0.14	70.90 ± 98.31	0.17 ± 0.02	0.71 ± 0.36	0.39 ± 0.16	33.65 ± 27.8	0.3 ± 0.23	0.44 ± 0.18	0.08 ± 0.01
	−10	26.20 ± 47.50	0.46 ± 0.21	0.87 ± 0.14	0.38 ± 0.15	79.75 ± 111.84	0.16 ± 0.02	0.54 ± 0.28	0.34 ± 0.14	112.75 ± 82	0.27 ± 0.12	0.24 ± 0.12	0.07 ± 0.01
	−15	39.70 ± 43.89	0.63 ± 0.22	0.83 ± 0.16	0.23 ± 0.13	193.00 ± 165.52	0.15 ± 0.02	0.19 ± 0.19	0.18 ± 0.12	110.8 ± 87.02	0.25 ± 0.1	0.25 ± 0.13	0.06 ± 0.01
	−20	58.55 ± 54.20	0.77 ± 0.21	0.76 ± 0.20	0.16 ± 0.11	207.90 ± 160.76	0.15 ± 0.02	0.13 ± 0.11	0.14 ± 0.11	113.12 ± 89.25	0.24 ± 0.12	0.26 ± 0.14	0.07 ± 0.02
	−25	46.95 ± 52.80	0.82 ± 0.20	0.80 ± 0.17	0.14 ± 0.09	231.85 ± 158.19	0.15 ± 0.02	0.11 ± 0.11	0.12 ± 0.09	115.42 ± 89.1	0.22 ± 0.13	0.26 ± 0.15	0.08 ± 0.02
	−30	45.50 ± 50.25	0.83 ± 0.20	0.81 ± 0.15	0.13 ± 0.08	218.80 ± 163.39	0.15 ± 0.02	0.12 ± 0.10	0.12 ± 0.09	113.2 ± 91.2	0.22 ± 0.14	0.26 ± 0.16	0.08 ± 0.03

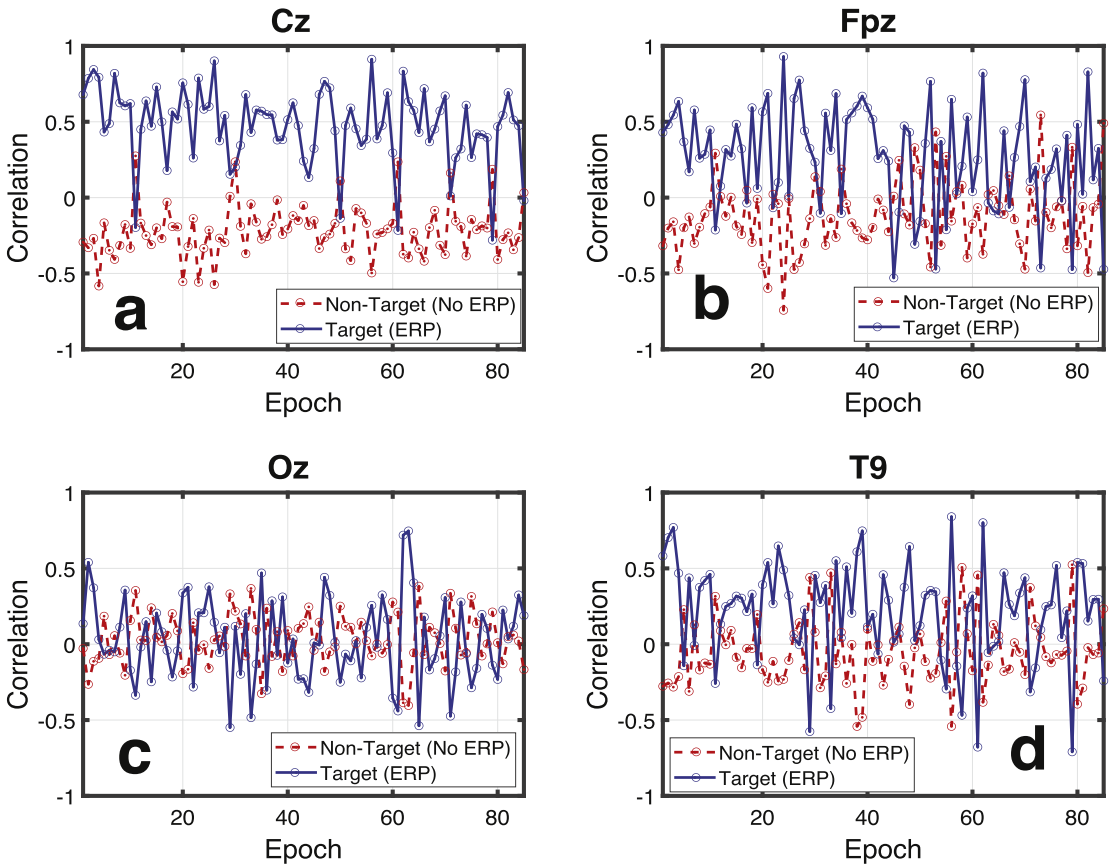

Figure 5. Correlation between estimated ERP component and grand-averaged ERP in four channels: (a) Cz channel, (b) Fpz channel, (c) Oz channel, (d) T9 channel. Blue lines and circles represent correlations of the target ERP, obtained by averaging the correlation of the target row ERP with the grand average and the correlation of the target column ERP with the grand average. Red lines and dots correspond to the non-target stimuli.

can be attributed to the specific characteristics of the synthesized data used in our study, where different weights were assigned to project sources onto the EEG subspace, resulting in a mixture of positive and negative values. The optimization procedure of NTD, which enforces non-negativity, ignores the presence of negative values, leading to suboptimal spatial estimation. Furthermore, when comparing our proposed method with NTD, our model demonstrated superior performance. This can be attributed to the fact that our approach considers the condition as a physiologically meaningful constraint, rather than solely relying on non-negativity.

3.5.2. Real data

The STF tensorization approach, using the specified hyperparameters that were used to simulate synthesized EEG data, was employed to generate a tensor from the multi-channel recordings. Temporal correlation was calculated between the estimated component in both target and non-target epochs and the grand average of four specific channels: Cz, Fpz, Oz, and T9. The selection of Cz and Fpz channels was based on their proximity to the source of the P300 component, while Oz and T9 were chosen as they are spatially distant and not expected to contain substantial P300-related information. The reported results present the average correlation of the target row and column for target trials (2 out of 12), as well as the average correlation of all rows and columns for non-target trials (10 out of 12). These correlations provide insights into the performance of the algorithm in distinguishing between target and non-target trials.

The findings indicate that the Cz channel (figure 5(a)) and Fpz channel (figure 5(b)) exhibit the highest mean correlation across different epochs, suggesting their sensitivity to the underlying ERP characteristics. Furthermore, these channels demonstrate a significant distinction between target and non-target trials. Conversely, in figures 5(c) and (d), the mean correlation between target and non-target ERPs is not significant. This indicates that the algorithm successfully detects the ERP component in target trials, while being unable to identify a significant component in channels that lack sufficient information about the underlying ERP component.

3.6. Limitations and future directions

The proposed algorithm in this study has been shown to perform superiorly compared to Tucker, NTD, and other conventional matrix-based transformations for the task of source separation. However, the convergence of the algorithm is not fully guaranteed due to the complexity of the loss function. This complexity leads to some unexpected trends in the results, as seen in figures 4 and 5 and tables 1 and 2. In future directions of this study, more complex and non-linear optimization algorithms should be investigated to address this challenge. One possible approach is to use second-order methods, such as Newton's method and quasi-Newton methods. These methods require the computation of the Hessian matrix, which can be computationally expensive, but they can provide faster convergence than first-order methods. Another approach is to use heuristic methods, such as genetic algorithms or simulated annealing, which do not rely on gradient information and can be less sensitive to the loss function's complexity.

In this study, we utilized higher-order singular value decomposition (HOSVD) for initializing the first and third modes of the tensor decompositions under investigation. While the uniqueness of HOSVD is guaranteed due to the uniqueness of the corresponding singular vector matrices of the unloading factors of the tensor, we adopted a different approach to initialize the second factor matrix, which in turn compromised the uniqueness of the proposed decomposition. However, it is important to note that the main focus of this study was to explore the Tucker decomposition in a way that allows for accurate prediction of the underlying ERP by appropriately combining different fibers of the second factor matrix. Additionally, the introduced constraint $(\lambda\ |\mathbf{w}^{\top}\mathbf{U}^{(2)} - \mathbf{r}(\tau)|_2^2)$ ensures that the amplitude of the estimated component falls within an acceptable range. Moreover, the presence of $\mathbf{w}$ helps alleviate the issue of column permutation within the second mode factor matrix. In future steps of this study, additional constraints and/or unique initialization methods can be incorporated into the decomposition to address the issue of non-uniqueness.

Additionally, it is important to investigate the robustness of the proposed algorithm to different types of noise and different EEG datasets. This will allow us to better understand the limitations and strengths of the proposed algorithm and to identify areas for further improvement. Furthermore, it would be also interesting to study the extension of the proposed algorithm to other source separation tasks such as audio and image source separation.

Overall, this study provides a promising approach for source separation, but there is still room for improvement. By investigating more complex and non-linear optimization algorithms and studying the robustness of the proposed algorithm, we can further improve its performance and make it more widely applicable to different source separation tasks.

4. Conclusion

In this paper, we proposed an extended tensor decomposition for single-trial ERP extraction. The proposed model is based on Tucker decomposition and adds meaningful constraints to the tensor factorizations to improve performance in source separation. Our experimental results in the synthesized EEG dataset show that the proposed model outperforms Tucker decomposition in terms of source separation accuracy. Additionally, the feasibility of leveraging the model in clinical applications is evaluated using a publicly available dataset that includes the P300 component.

This study has important implications for the field of EEG signal processing and source separation. Our findings suggest that modified tensor decompositions may be a promising approach for extracting single-trial ERPs from EEG data. Furthermore, the proposed method may also be applied to other source separation problems, such as audio and image source separation.

In future work, we plan to investigate the effectiveness of the proposed method on different EEG datasets and with different types of noise. Additionally, we will explore the possibility of incorporating additional constraints into the tensor decomposition to further improve the performance of the proposed method. Overall, this study has the potential to pave the way for new and more accurate methods for extracting single-trial ERPs and other source separation applications.

Data availability statement

The data are simulated and can be provided upon request. The data that support the findings of this study are available upon reasonable request from the authors. The real-world data are also available through BCI competition website (https://www.bbci.de/competition/iii/.)

ORCID iDs

Behrad TaghiBeyglou https://orcid.org/0000-0003-4031-9537

References

Akbari H, Shamsollahi M B and Phlypo R 2015 Fetal ECG extraction using πTucker decomposition *2015 Int. Conf. on Systems, Signals and Image Processing (IWSSIP)* (IEEE) pp 174–8

Aniyan A K, Philip N S, Samar V J, Desjardins J A and Segalowitz S J 2014 A wavelet based algorithm for the identification of oscillatory event-related potential components *J. Neurosci. Methods* **233** 63–72

Bernat E M, Williams W J and Gehring W J 2005 Decomposing ERP time-frequency energy using PCA *Clin. Neurophysiol.* **116** 1314–34

Blankertz B, Muller K-R, Krusienski D J, Schalk G, Wolpaw J R, Schlogl A, Pfurtscheller G, Millan J R, Schroder M and Birbaumer N 2006 The BCI competition III: validating alternative approaches to actual BCI problems *IEEE Trans. Neural Syst. Rehabil. Eng.* **14** 153–9

Blankertz B *et al* 2004 The BCI competition 2003: progress and perspectives in detection and discrimination of EEG single trials *IEEE Trans. Biomed. Eng.* **51** 1044–51

Cerutti S, Bersani V, Carrara A and Liberati D 1987 Analysis of visual evoked potentials through Wiener filtering applied to a small number of sweeps *J. Biomed. Eng.* **9** 3–12

Chen Z, Xu Z and Wang D 2021 Deep transfer tensor decomposition with orthogonal constraint for recommender systems *Proc. of the AAAI Conf. on Artificial Intelligence* 35 pp 4010–8

Cichocki A 2002 ICALAB toolboxes (http://bsp.brain.riken.jp/ICALAB)

Cichocki A, Mandic D, De Lathauwer L, Zhou G, Zhao Q, Caiafa C and Phan H A 2015 Tensor decompositions for signal processing applications: from two-way to multiway component analysis *IEEE Signal Process Mag.* **32** 145–63

Cichocki A, Washizawa Y, Rutkowski T, Bakardjian H, Phan A-H, Choi S, Lee H, Zhao Q, Zhang L and Li Y 2008 Noninvasive BCIs: multiway signal-processing array decompositions *Computer* **41** 34–42

Cichocki A, Zdunek R, Phan A H and Amari S I 2009 *Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-way Data Analysis and Blind Source Separation* (New York: Wiley)

Cong F, Phan A-H, Astikainen P, Zhao Q, Wu Q, Hietanen J K, Ristaniemi T and Cichocki A 2013 Multi-domain feature extraction for small event-related potentials through nonnegative multi-way array decomposition from low dense array EEG *Int. J. Neural Syst.* **23** 1350006

Dao N T A *et al* 2020 Multi-channel EEG epileptic spike detection by a new method of tensor decomposition *J. Neural Eng.* **17** 016023

De Vos M, Vergult A, De Lathauwer L, De Clercq W, Van Huffel S, Dupont P, Palmini A and Van Paesschen W 2007 Canonical decomposition of ictal scalp EEG reliably detects the seizure onset zone *NeuroImage* **37** 844–54

Deburchgraeve W, Cherian P J, De Vos M, Swarte R M, Blok J H, Visser G H, Govaert P and Van Huffel S 2009 Neonatal seizure localization using PARAFAC decomposition *Clin. Neurophysiol.* **120** 1787–96

Dien J 2010 The ERP PCA Toolkit: an open source program for advanced statistical analysis of event-related potential data *J. Neurosci. Methods* **187** 138–45

Dien J 2012 Applying principal components analysis to event-related potentials: a tutorial *Developmental Neuropsychology* **37** 497–517

Eliseyev A and Aksenova T 2013 Recursive N-way partial least squares for brain-computer interface *PLoS One* **8** e69962

Fonał K and Zdunek R 2019 Fast recursive nonnegative standard and hierarchical Tucker decomposition *IEEE Signal Process Lett.* **26** 1265–9

Georgiadis S D, Ranta-aho P O, Tarvainen M P and Karjalainen P A 2005 Single-trial dynamical estimation of event-related potentials: a Kalman filter-based approach *IEEE Trans. Biomed. Eng.* **52** 1397–406

Idaji M J, Shamsollahi M B and Sardouie S H 2017 Higher order spectral regression discriminant analysis (HOSRDA): A tensor feature reduction method for ERP detection *Pattern Recognit.* **70** 152–62

Jarchi D, Sanei S, Mohseni H R and Lorist M M 2011 Coupled particle filtering: A new approach for P300-based analysis of mental fatigue *Biomed. Signal Process. Control* **6** 175–85

Jarchi D, Sanei S, Principe J C and Makkiabadi B 2010 A new spatiotemporal filtering method for single-trial estimation of correlated ERP subcomponents *IEEE Trans. Biomed. Eng.* **58** 132–43

Jung T, Makeig S, Westerfield M, Townsend J, Courchesne E and Sejnowski T 1999 Independent component analysis of single-trial event-related potentials *Proc. ICA* **99** 173–9

Kaplan A Y, Shishkin S L, Ganin I P, Basyul I A and Zhigalov A Y 2013 Adapting the P300-based brain-computer interface for gaming: a review," IEEE Transactions on *Computat. Intell. AI Games* **5** 141–9

Käthner I, Wriessnegger S C, Müller-Putz G R, Kübler A and Halder S 2014 Effects of mental workload and fatigue on the P300, alpha and theta band power during operation of an ERP (P300) brain-computer interface *Biol. Psychol.* **102** 118–29

Kayser J and Tenke C E 2003 Optimizing PCA methodology for ERP component identification and measurement: theoretical rationale and empirical evaluation *Clin. Neurophysiol.* **114** 2307–25

Kim Y-D and Choi S 2007 Nonnegative Tucker decomposition *2007 IEEE Conf. on Computer vision and Pattern Recognition (IEEE)* pp 1–8

Klados M A and Bamidis P D 2016 A semi-simulated EEG/EOG dataset for the comparison of EOG artifact rejection techniques *Data Brief* **8** 1004–6

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

Kossaifi J, Panagakis Y, Anandkumar A and Pantic M 2018 Tensorly: Tensor learning in python arXiv:1610.09555

Latchoumane C-F V, Vialatte F-B, Solé-Casals J, Maurice M, Wimalaratna S R, Hudson N, Jeong J and Cichocki A 2012 Multiway array decomposition analysis of EEGs in alzheimer's disease *J. Neurosci. Methods* **207** 41–50

Lee W L, Tan T, Falkmer T and Leung Y H 2016 Single-trial event-related potential extraction through one-unit ICA-with-reference *J. Neural Eng.* **13** 066010

Li P *et al* 2015 Autoregressive model in the Lp norm space for EEG analysis *J. Neurosci. Methods* **240** 170–8

Li R, Keil A and Principe J C 2009 Single-trial P300 estimation with a spatiotemporal filtering method *J. Neurosci. Methods* **177** 488–96

Li R, Principe J C, Bradley M and Ferrari V 2008 A spatiotemporal filtering methodology for single-trial ERP component estimation *IEEE Trans. Biomed. Eng.* **56** 83–92

Linden D E 2005 The P300: where in the brain is it produced and what does it tell us? *Neuroscientist* **11** 563–76

Luck S J 2014 *An Introduction to the Event-related Potential Technique* (Cambridge, MA: MIT Press)

Luck S J, Woodman G F and Vogel E K 2000 Event-related potential studies of attention *Trends Cogn Sci* **4** 432–40

Ma R, Yu T, Zhong X, Yu Z L, Li Y and Gu Z 2021 Capsule network for ERP detection in brain-computer interface *IEEE Trans. Neural Syst. Rehabil. Eng.* **29** 718–30
MacDonald B and Barry R J 2014 Trial effects in single-trial ERP components and autonomic responses at very long ISIs *Int. J. Psychophysiol.* **92** 99–112
Maki H, Tanaka H, Sakti S and Nakamura S 2018 Graph regularized tensor factorization for single-trial EEG analysis *2018 IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)* (IEEE) pp 846–50
Malik O A and Becker S 2018 Low-rank Tucker decomposition of large tensors using tensorsketch *Advances in Neural Information Processing Systems*
Mørup M, Hansen L K, Arnfred S M, Lim L-H and Madsen K H 2008 Shift-invariant multilinear decomposition of neuroimaging data *NeuroImage* **42** 1439–50
Mørup M, Hansen L K, Herrmann C S, Parnas J and Arnfred S M 2006 Parallel factor analysis as an exploratory tool for wavelet transformed event-related EEG *NeuroImage* **29** 938–47
Ouyang G, Herzmann G, Zhou C and Sommer W 2011 Residue iteration decomposition (RIDE): a new method to separate ERP components on the basis of latency variability in single trials, *Psychophysiology* **48** 1631–47
Quiroga R Q and Garcia H 2003 Single-trial event-related potentials with wavelet denoising *Clin. Neurophysiology* **114** 376–90
Ranjbar M, Mikaeili M and Banaraki A K 2018 Single trial estimation of peak latency and amplitude of multiple correlated ERP components *J. Med. Biol. Eng.* **38** 161–72
Ranjbar M, Mikaeili M and Banaraki A K 2020 Single trial estimation of event-related potential components using spatiotemporal filtering and artificial bee colony optimized Gaussian kernel mixture model *Int. J. Adapt Control Signal Process.* **34** 1135–47
Rošťáková Z, Rosipal R, Seifpour S and Trejo L J 2020 A comparison of non-negative Tucker decomposition and parallel factor analysis for identification and measurement of human EEG rhythms *Meas. Sci. Rev.* **20** 126–38
Sabeti M, Katebi S, Rastgar K and Azimifar Z 2016 A multi-resolution approach to localize neural sources of P300 event-related brain potential *Comput. Methods Programs Biomed.* **133** 155–68
Schomer D L and Da Silva F L 2012 *Niedermeyer's Electroencephalography: Basic Principles, Clinical Applications, and Related Fields* (Philadelphia, PA: Lippincott Williams & Wilkins)
Ting C-M, Salleh S-H, Zainuddin Z M and Bahar A 2014 Artifact removal from single-trial ERPs using non-Gaussian stochastic volatility models and particle filter *IEEE Signal Process Lett.* **21** 923–7
Vanderperren K *et al* 2013 Single trial ERP reading based on parallel factor analysis *Psychophysiology* **50** 97–110
Wang D, Zhu Y, Ristaniemi T and Cong F 2018 Extracting multi-mode ERP features using fifth-order nonnegative tensor decomposition *J. Neurosci. Methods* **308** 240–7
Wang X, Liu W, Toiviainen P, Ristaniemi T and Cong F 2020 Group analysis of ongoing EEG data based on fast double-coupled nonnegative tensor decomposition *J. Neurosci. Methods* **330** 108502
Xiao X, Xu M, Jin J, Wang Y, Jung T-P and Ming D 2019 Discriminative canonical pattern matching for single-trial classification of ERP components *IEEE Trans. Biomed. Eng.* **67** 2266–75
Yokota T, Lee N and Cichocki A 2016 Robust multilinear tensor rank estimation using higher order singular value decomposition and information criteria *IEEE Trans. Signal Process.* **65** 1196–206
Zang S, Ding X, Wu M and Zhou C 2022 An EEG classification-based method for single-trial N170 latency detection and estimation *Comput. Math. Methods Med.* **2022** 6331956
Zhang G *et al* 2020 Multi-domain features of the non-phase-locked component of interest extracted from ERP data by tensor decomposition, *Brain Topography* **33** 37–47
Zhao Q, Caiafa C F, Mandic D P, Chao Z C, Nagasaka Y, Fujii N, Zhang L and Cichocki A 2012 Higher order partial least squares (HOPLS): a generalized multilinear regression method *IEEE Trans. Pattern Anal. Mach. Intell.* **35** 1660–73
Zhou G and Cichocki A 2012 Fast and unique Tucker decompositions via multiway blind source separation *Bull. Polish Acad. Sci. Tech. Sci.* **60** 389–405