

## Group analysis of ongoing EEG data based on fast double-coupled nonnegative tensor decomposition

Xiulin Wang<sup>a,b</sup>, Wenya Liu<sup>b</sup>, Petri Toivainen<sup>c</sup>, Tapani Ristaniemi<sup>b</sup>, Fengyu Cong<sup>a,b,\*</sup>

<sup>a</sup> School of Biomedical Engineering, Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology, Dalian, China

<sup>b</sup> Faculty of Information Technology, University of Jyväskylä, Jyväskylä, Finland

<sup>c</sup> Finnish Centre of Excellence in Interdisciplinary Music Research, Department of Music, University of Jyväskylä, Jyväskylä, Finland

### ARTICLE INFO

**Keywords:**  
Coupled  
Music  
Nonnegative  
Tensor decomposition  
Ongoing EEG

### ABSTRACT

**Background:** Ongoing EEG data are recorded as mixtures of stimulus-elicited EEG, spontaneous EEG and noises, which require advanced signal processing techniques for separation and analysis. Existing methods cannot simultaneously consider common and individual characteristics among/within subjects when extracting stimulus-elicited brain activities from ongoing EEG elicited by 512-s long modern tango music.

**New method:** Aiming to discover the commonly music-elicited brain activities among subjects, we provide a comprehensive framework based on fast double-coupled nonnegative tensor decomposition (FDC-NTD) algorithm. The proposed algorithm with a generalized model is capable of simultaneously decomposing EEG tensors into common and individual components.

**Results:** With the proposed framework, the brain activities can be effectively extracted and sorted into the clusters of interest. The proposed algorithm based on the generalized model achieved higher fittings and stronger robustness. In addition to the distribution of centro-parietal and occipito-parietal regions with theta and alpha oscillations, the music-elicited brain activities were also located in the frontal region and distributed in the 4~11 Hz band.

**Comparison with existing method(s):** The present study, by providing a solution of how to separate common stimulus-elicited brain activities using coupled tensor decomposition, has shed new light on the processing and analysis of ongoing EEG data in multi-subject level. It can also reveal more links between brain responses and the continuous musical stimulus.

**Conclusions:** The proposed framework based on coupled tensor decomposition can be successfully applied to group analysis of ongoing EEG data, as it can be reliably inferred that those brain activities we obtained are associated with musical stimulus.

### 1. Introduction

Listening to music has proven to be an effective strategy to improve and rehabilitate the human health (Koelsch, 2012; MacDonald et al., 2013), especially for people with insomnia, depression, schizophrenia or similar illnesses (Maratos et al., 2008; Mössler et al., 2011; Jespersen et al., 2015). Therefore, revealing brain activities during listening to music has drawn an increasing amount of research interest in recent decades (Cong et al., 2012b, 2013a; Wang et al., 2016; Li et al., 2016; Zhu et al., 2019). The advent of brain imaging techniques has provided

researchers with the opportunity and insight to probe the brain functions elicited by listening to music. For example, Electroencephalography (EEG) is a collection of potentials along the scalp that reflect electrical activities of the brain. Since Hans Berger first introduced EEG to the world in 1929 (Berger, 1929), it has been widely used in the study of brain functions (Cong et al., 2013a; Huber et al., 2004; Herrmann, 2001) and diagnosis of neurological diseases/disorders (Jeong, 2004; Adeli and Ghosh-Dastidar, 2010; Siuly et al., 2016). Unlike spontaneous EEG recorded in resting state (Berger, 1929) or event-related potentials (ERP) acquired through repeated

\* Corresponding author at: School of Biomedical Engineering, Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology, Dalian, China.

E-mail addresses: [xiulin.wang@foxmail.com](mailto:xiulin.wang@foxmail.com) (X. Wang), [wenyaliu0912@foxmail.com](mailto:wenyaliu0912@foxmail.com) (W. Liu), [petri.toivainen@jyu.fi](mailto:petri.toivainen@jyu.fi) (P. Toivainen), [tapani.e.ristaniemi@jyu.fi](mailto:tapani.e.ristaniemi@jyu.fi) (T. Ristaniemi), [cong@dlut.edu.cn](mailto:cong@dlut.edu.cn) (F. Cong).

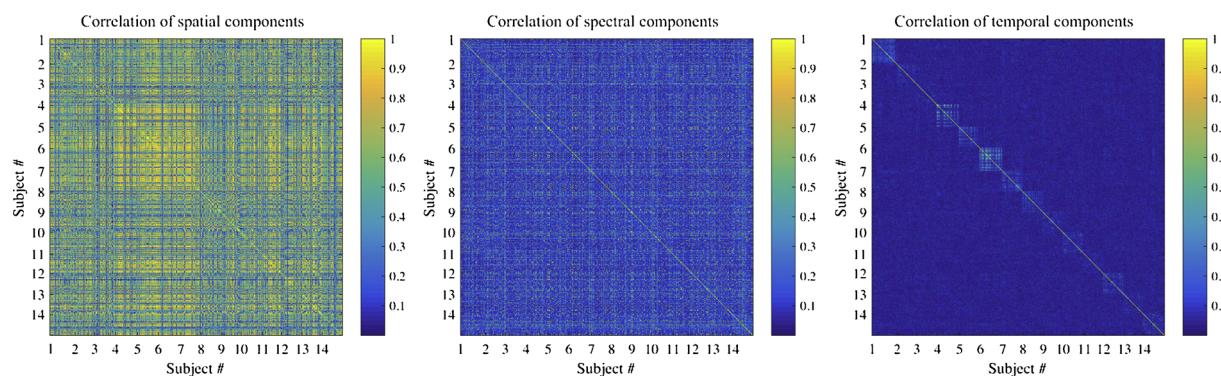
URLs: <http://www.xiulin.wang> (X. Wang), <http://www.users.jyu.fi/~ptoivai/> (P. Toivainen), <http://www.users.jyu.fi/~riesta> (T. Ristaniemi), <http://www.escience.cn/people/cong> (F. Cong).

presentation of stimuli (Luck, 2014), ongoing EEG is a direct response to brain activities in naturalistic and continuous context (e.g. listening to music or watching movies) (Busch et al., 2009; Cong et al., 2012b, 2013a), which makes it possible to study brain functions during real-world experiences. In ongoing EEG experiment, the recorded data can be viewed as mixtures of stimulus-elicited EEG, spontaneous EEG and noises, but how to separate the stimulus-elicited brain activities from ongoing EEG data still remains open for research (Cong et al., 2013a, 2015b; Zhu et al., 2019). Therefore, this study is devoted to the separation and analysis of ongoing EEG data elicited by a 512-s long piece of modern tango music.

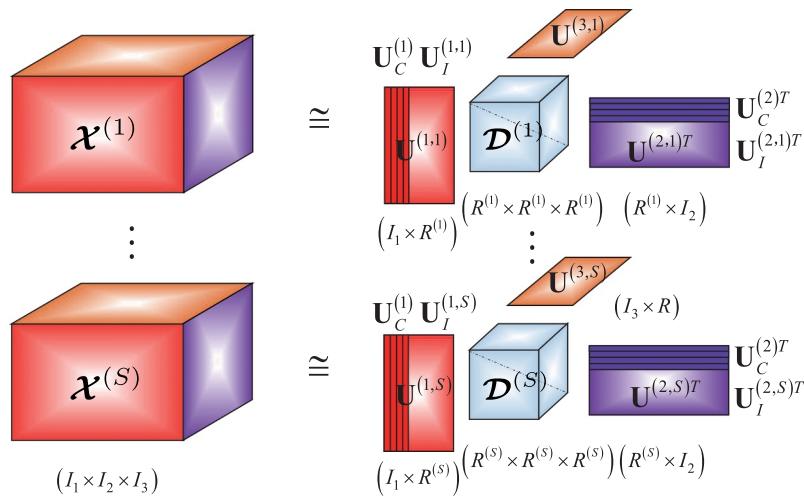
In recent years, for the data used in this paper, several studies have been tried to extract the music-elicited brain activities. Cong et al. constructed a fourth-order EEG tensor of channel  $\times$  frequency  $\times$  time  $\times$  subject and the tensor was decomposed using nonnegative tensor factorization (NTF) (Cong et al., 2012b). It should be noted that the analysis of high-order tensors is based on the assumption that the underlying information in temporal, spatial and spectral modes are consistent among subjects (Wang et al., 2018b). However, we found that there was almost no consistent temporal information among subjects as shown in Fig. 1. In Cong et al. (2013a) and Zhu et al. (2019), the authors first adopted independent component analysis (ICA) or spatial ICA to decompose two-way ongoing EEG data represented by each subject and then applied time-frequency analysis and K-means clustering to find spatial, spectral and temporal information of interest. Apparently, such ICA-based approach did not take into account the high-correlation information in space and frequency modes among subjects as shown in Fig. 1, and did not fully utilize the inherent structural information of the tensors represented spatial, temporal and spectral modes (Cong et al., 2015a). In Wang et al. (2016), multilinear partial least squares (PLS) was performed on the tensor (represented by ongoing EEG data) and matrix (represented by musical features), however, it did not consider the nonnegative nature of EEG tensor brought by the time-frequency analysis. With the consideration of phase characteristics, Li et al. applied the ordered Parallel Factors (PARAFAC) algorithm to the ongoing EEG data elicited by the same tango music (Li et al., 2016), but no coupled information among subjects was utilized in the data processing. Tensors, also termed as multi-way arrays, are the higher-order extension of matrices (Kolda and Bader, 2009). Ongoing EEG data can be naturally represented as tensors in which the structural information of inherent interactions between different modes can be fully utilized (Cong et al., 2015a). For example, considering the time-frequency representation of EEG data in each channel, a third-order tensor of

channel  $\times$  time  $\times$  frequency can be formed (Acar et al., 2007). Tensor decomposition allows for simultaneous consideration of spatial, temporal and spectral information, which provides convincing solutions with physiological or pathological interpretations (Cichocki, 2013). However, when it comes to the analysis of an ensemble of ongoing EEG data (e.g., the data collected from different subjects under the same tango music), it is unreasonable to represent them as a high-order tensor of channel  $\times$  frequency  $\times$  time  $\times$  subject and apply high-order tensor decomposition owing to the incomplete consistency in channel, time and frequency patterns across subjects. Moreover, when analyzing the data through two-way matrix or individual tensor decomposition methods, potential interactions among the multi-way structure of tensors or the coupling information among tensors will inevitably be lost (Cong et al., 2015a; Mørup, 2011).

Coupled tensor decomposition, the extension of tensor decomposition to multi-block tensors, provides a natural framework for the simultaneous analysis of heterogeneous tensors with coupling information (Zhou et al., 2016; Sørensen et al., 2015; Gong et al., 2016). The crucial difference between them is that tensor decomposition processes tensors of  $\times$  frequency  $\times$  time individually or a higher-order tensor of channel  $\times$  frequency  $\times$  time  $\times$  subject (generated by stacking tensors from different subjects with the consistent assumption of spatial, temporal and spectral information among the third-order tensors) (Wang et al., 2018b; Mørup et al., 2006; Cong et al., 2012a), while coupled tensor decomposition generalizes tensor decomposition to cover the sharing information across multiple tensor blocks (Sørensen et al., 2015; Gong et al., 2016; Ermiş et al., 2015; Yokota et al., 2012). Compared with its matrix counterparts (Chen et al., 2016; Gong et al., 2015; Calhoun et al., 2009), coupled tensor decomposition can achieve unique solutions and interpretable components, while circumventing the independence constraint (Hunyadi et al., 2017; Mørup, 2011). Given the data collected under the same stimulus, it is reasonable to expect identical or highly correlated stimulus-elicited information among subjects, which can be regarded as a prerequisite for applying coupled tensor decomposition. However, the inter-component similarity among subjects has rarely been considered in previous methods (Cong et al., 2013a; Wang et al., 2016; Li et al., 2016; Zhu et al., 2019). Due to individual differences, individual characteristics in each subject may lead to inconsistent number of components among tensors (Zhou et al., 2016; Ermiş et al., 2015). This inconsistency is not considered in the realization of linked canonical polyadic tensor decomposition (LCPTD) model in Yokota et al. (2012) and Wang et al. (2019). In addition, the time consumption load will be extremely heavy due to the



**Fig. 1.** Inter- and intra-subject correlations of spatial, spectral and temporal components. The spatial (spectral or temporal) components decomposed from ongoing EEG data of 14 subjects by tensor decomposition individually (here we use the fast hierarchical alternative least squares (fast-HALS) algorithm Cichocki and Phan (2009)) are concatenated together, and then the correlation coefficients are calculated.



**Fig. 2.** Conceptual illustration of generalized LCPTD model with double-coupled constraint (adapted from Cichocki (2013)), in which the factor matrices of mode-1 and mode-2 among tensors are partially linked, as they share the same components  $\mathbf{U}_C^{(1)}$  and  $\mathbf{U}_C^{(2)}$ , respectively.

high-dimensional and nonnegative nature of ongoing EEG data (Wang et al., 2019; Zhou et al., 2012). Our preliminary exploration on the LCPTD model in the application of ongoing EEG data processing has been reported in Wang et al. (2019). Based on the above issues, the specific contributions of this paper can be listed as follows:

First, regarding the individual differences in ongoing EEG data, this study developed a more versatile and flexible model with inconsistent component number in each tensor for coupled tensor decomposition. This model enables the simultaneous decomposition of common components and individual components among tensors.

Second, based on the model mentioned above, this study proposed an efficient data-driven coupled tensor decomposition method termed as fast double-coupled nonnegative tensor decomposition (FDC-NTD) algorithm.

Third, in order to discover the reliable links between brain responses and musical stimulus, this study proposed a general framework based on coupled tensor decomposition for ongoing EEG data processing and analysis. To the best of our knowledge, this is the first attempt to apply coupled tensor decomposition to the group analysis of ongoing EEG data.

### 1.1. Why nonnegative and double-coupled constraints?

From the perspective of data analysis, imposing specific constraints on different modes or components during the decomposition process would contribute to obtaining more meaningful solutions (Cichocki, 2013; Wang et al., 2018b). After performing TFR, nonnegative constraint is naturally brought into the EEG data. Correspondingly, the temporal, spectral and spatial components of EEG tensor are all non-negative, representing the specific physical meanings of time envelope, spectrum and topography, respectively (Wang et al., 2018b).

Given the ongoing EEG data collected under the same stimulus, it is reasonable to expect coupled (identical or highly correlated) components among subjects. Fig. 1 shows the inter- and intra-subject correlations of spatial, spectral and temporal components respectively, which are extracted from the ongoing EEG data of 14 subjects by conventional tensor decomposition individually (i.e., regardless of any coupled information). The detailed information of these data is

described in Section 3. We can see that the correlations of some components among subjects in spatial and spectral modes are very significant, while the correlations of temporal components are almost nonexistent. Due to the sparse nature in spectral mode, of course, the correlations of spectral components are not as pronounced as correlations of the spatial components. Therefore, in this study, we consider imposing double coupled constraints in spatial and spectral modes.

## 2. Fast double-coupled nonnegative tensor decomposition

### 2.1. Basic notations and mathematical operations

Generally, scalars, vectors, matrices and tensors are respectively denoted by lowercase, boldface lowercase, boldface uppercase and calligraphic boldface uppercase letters, e.g.  $x$ ,  $\mathbf{x}$ ,  $X$ ,  $\mathcal{X}$ .  $\mathbb{R}$  and  $\mathbb{R}_+$  denote real number and nonnegative real number. Operators  $(\cdot)^T$  and  $\|\cdot\|_F$  denote transpose and Frobenius norm, respectively. Outer product, Khatri-Rao product, Hadamard product and element-wise division are denoted by ‘ $\odot$ ’, ‘ $\odot^*$ ’, ‘ $\circledast$ ’ and ‘ $\oslash$ ’, respectively. Moreover,  $\mathbf{U}^{(1)} \odot \mathbf{U}^{(2)} \dots \odot \mathbf{U}^{(N)}$ ,  $\mathbf{U}^{(1)} \odot \dots \odot \mathbf{U}^{(n-1)} \odot \mathbf{U}^{(n+1)} \dots \odot \mathbf{U}^{(N)}$ ,  $\mathbf{U}^{(1)} \circledast \mathbf{U}^{(2)} \dots \circledast \mathbf{U}^{(N)}$  and  $\mathbf{U}^{(1)} \circledast \dots \circledast \mathbf{U}^{(n-1)} \circledast \mathbf{U}^{(n+1)} \dots \circledast \mathbf{U}^{(N)}$  are defined as  $\{\mathbf{U}\}^\odot$ ,  $\{\mathbf{U}\}^{\odot-n}$ ,  $\{\mathbf{U}\}^\circledast$ ,  $\{\mathbf{U}\}^{\circledast-n}$ , respectively. The mode- $n$  matricization of a tensor  $\mathcal{X} \in \mathbb{R}_{I_1 \times I_2 \times \dots \times I_N}$  is termed as  $\mathcal{X}_{(n)}$  with the size of  $I_n \times (I_1 \dots I_{n-1} I_{n+1} \dots I_N)$ . Please refer to Kolda and Bader (2009) for a more detailed description of standard notations and basic tensor operations.

### 2.2. Model generalization

Aiming to process multi-block tensors with coupled information, Yokota et al. proposed the LCPTD model (Yokota et al., 2012), which can enable the simultaneous extraction of common components, individual components and core tensors. This model assumes that those tensors are linked together for sharing some common components. However, even if the tensors are generated under the same conditions, individual differences between them will present as individual characteristics, which may result in inconsistent number of components in each tensor. This inconsistency was not considered in the LCPTD model.

Therefore, in this section, we first extend a generalized LCPTD model of inconsistent component number  $R^{(s)}$ . Given a set of  $N$ th-order non-negative tensors  $X^{(s)} \in \mathbb{R}_+^{I_1 \times I_2 \times \dots \times I_N}$ ,  $s = 1, 2, \dots, S$ , the generalized nonnegative LCPTD model can be expressed as:

$$\begin{aligned} X^{(s)} \approx \hat{X}^{(s)} &= \sum_{r=1}^{R^{(s)}} \lambda_r^{(s)} \mathbf{u}_r^{(1,s)} \circ \mathbf{u}_r^{(2,s)} \circ \dots \circ \mathbf{u}_r^{(N,s)} \\ &= \mathcal{D}^{(s)} \times_1 \mathbf{U}^{(1,s)} \times_2 \mathbf{U}^{(2,s)} \dots \times_N \mathbf{U}^{(N,s)}, \end{aligned} \quad (1)$$

where the tensor  $\hat{X}^{(s)} \in \mathbb{R}_+^{I_1 \times I_2 \times \dots \times I_N}$  denotes the estimated item of tensor  $X^{(s)}$ .  $\mathbf{u}_r^{(n,s)}$  denotes the  $r$ th column of  $n$ -mode factor matrix  $\mathbf{U}^{(n,s)} \in \mathbb{R}_+^{I_n \times R^{(s)}}$  of  $s$ th tensor ( $s = 1, 2, \dots, S$ ,  $n = 1, 2, \dots, N$ ), and  $\mathbf{U}^{(n,s)} = [\mathbf{u}_1^{(n,s)}, \mathbf{u}_2^{(n,s)}, \dots, \mathbf{u}_R^{(n,s)}]$ .  $\mathcal{D}^{(s)} \in \mathbb{R}_+^{R^{(s)} \times R^{(s)} \dots \times R^{(s)}}$  represents the  $s$ th core tensor with non-zero entries  $\lambda_r^{(s)}$  only on the super-diagonal elements ( $r, r, \dots, r$ ),  $r = 1, 2, \dots, R^{(s)}$ . Most importantly, in generalized LCPTD model, each factor matrix  $\mathbf{U}^{(n,s)}$  includes two parts:  $\mathbf{U}_C^{(n)} \in \mathbb{R}_+^{I_n \times L_n}$ ,  $0 \leq L_n \leq R^{(s)}$  and  $\mathbf{U}_I^{(n,s)} \in \mathbb{R}_+^{I_n \times (R^{(s)} - L_n)}$ .  $\mathbf{U}_C^{(n)}$  shared by all tensors represents the coupling information among them, while  $\mathbf{U}_I^{(n,s)}$  corresponds to the individual characteristics of each tensor. Fig. 2 gives the conceptual illustration of generalized double-coupled tensor decomposition model.

### 2.3. Model realization

In this section, aiming to extract the constrained factor matrices and core tensors  $\mathcal{D}^{(s)}$ , a solution based on fast-HALS algorithm for the generalized LCPTD model is provided. Note that the scalar factor  $\lambda_r^{(s)}$  of core tensor  $\mathcal{D}^{(s)}$  can be absorbed into one denormalized component, such as  $\mathbf{u}_r^{(N,s)}$ , so the cost function using squared Euclidean distance minimization can be represented in a simplified form as follows:

$$\min \sum_{s=1}^S \left\| X^{(s)} - \sum_{r=1}^{R^{(s)}} \mathbf{u}_r^{(1,s)} \circ \mathbf{u}_r^{(2,s)} \circ \dots \circ \mathbf{u}_r^{(N,s)} \right\|_F^2 \quad (2)$$

$$\text{s. t. } \mathbf{u}_r^{(n,1)} = \dots = \mathbf{u}_r^{(n,S)}, \quad r \leq L_n,$$

$$\|\mathbf{u}_r^{(n,s)}\| = 1, \quad n = 1 \dots N - 1, \quad r = 1 \dots R^{(s)}, \quad s = 1 \dots S.$$

The above optimization problem can be converted into  $\max(R^{(s)})$  suboptimization problems via HALS algorithm (Cichocki et al., 2007), in which  $\mathbf{u}_r^{(n,s)}$  can be calculated sequentially and iteratively. To address the issue of high computation cost, we further introduce the fast-HALS algorithm (Cichocki and Phan, 2009) to the proposed model. Therefore, the updating rule of  $\mathbf{u}_r^{(n,s)}$  can be defined as follows:

$$\mathbf{u}_r^{(n,s)} = \begin{cases} \left[ \sum_s \zeta_r^{(n,s)} \right] / \sum_s \gamma_r^{(n,s)}, & r \leq L_n, \\ \zeta_r^{(n,s)} / \gamma_r^{(n,s)}, & r > L_n, \end{cases} \quad (3)$$

where  $[\cdot]_+$  means “half-rectifying” nonlinear projection to obtain non-negative components and  $\zeta_r^{(n,s)}$  is defined as:

$$\zeta_r^{(n,s)} = [X_{(n)}^{(s)} \circ [U^{(s)}]_{\odot-n}]_r - U^{(n,s)} [\xi_{(n)}^{(s)}]_r + \gamma_r^{(n,s)} \mathbf{u}_r^{(n,s)} \quad (4)$$

with  $\xi_{(n)}^{(s)} = (U^{(s)T} U^{(s)})^{\otimes} \otimes (U^{(n,s)T} U^{(n,s)})$  and  $X_{(n)}^{(s)}$  is the mode- $n$  matricization of tensor  $X^{(s)}$ . The scaling coefficients  $\gamma_r^{(n,s)}$  can be formulated as:

$$\gamma_r^{(n,s)} = \begin{cases} \mathbf{u}_r^{(N,s)T} \mathbf{u}_r^{(N,s)}, & n \neq N, \\ 1, & n = N. \end{cases} \quad (5)$$

In each iteration, we perform the updates of  $\mathbf{u}_r^{(n,s)}$  with the indexes  $n = 1, 2, \dots, N$ ,  $s = 1, 2, \dots, S$  and  $r = 1, 2, \dots, R^{(s)}$ .

successively, while normalizing it to unit variance by  $\mathbf{u}_r^{(n,s)} \leftarrow \mathbf{u}_r^{(n,s)} / \|\mathbf{u}_r^{(n,s)}\|_2$  except  $n \neq N$ . As illustrated in (3), the calculation of common component  $\mathbf{u}_r^{(n,s)}$  depends on all tensors and the individual component needs to be calculated separately. These components are alternatively updated one after another until convergence. In this study, considering that those nonnegative tensors represented by ongoing EEG data are assumed to be coupled in spatial and spectral modes (i.d.,  $L_n = 0$ ,  $n > 2$ ), the proposed algorithm is termed as fast double-coupled nonnegative tensor decomposition (FDC-NTD) algorithm. We summarize the proposed FDC-NTD algorithm in Algorithm 1, and its detailed derivation is given in the appendix.

### Algorithm 1. FDC-NTD algorithm

---

**Input:**  $\mathcal{X}^{(s)}$ ,  $L_1$ ,  $L_2$  and  $R^{(s)}$ ,  $s = 1, \dots, S$   
**Initialization:**  $\mathbf{U}^{(n,s)}$ ,  $n = 1, \dots, N$ ,  $s = 1, \dots, S$   
 $\mathbf{u}_r^{(n,s)} \leftarrow \mathbf{u}_r^{(n,s)} / \|\mathbf{u}_r^{(n,s)}\|_2$ ,  $n \neq N$ ,  
 $\mathcal{X}_{(n)}^{(s)}$ ,  $n = 1, \dots, N$ ,  $s = 1, \dots, S$   
 $\mathbf{E}^{(s)} = (\mathbf{U}^{(s)T} \mathbf{U}^{(s)})^{\otimes}$ ,  $s = 1, \dots, S$

**while** not convergence **do**

- for**  $n = 1, 2, \dots, N$  **do**
- $\mathbf{E}^{(s)} = \mathbf{E}^{(s)} \odot (\mathbf{U}^{(n,s)T} \mathbf{U}^{(n,s)})$ ,  $s = 1, \dots, S$
- for**  $r = 1, 2, \dots, \max(R^{(s)})$  **do**
- update  $\mathbf{u}_r^{(n,s)}$ ,  $r \leq R^{(s)}$ ,  $s = 1, 2, \dots, S$
- via equation (3)
- end**
- $\mathbf{E}^{(s)} = \mathbf{E}^{(s)} \circledast (\mathbf{U}^{(n,s)T} \mathbf{U}^{(n,s)})$ ,  $s = 1, \dots, S$
- end**

**Output:**  $\mathbf{U}^{(n,s)}$ ,  $n = 1, \dots, N$ ,  $s = 1, \dots, S$

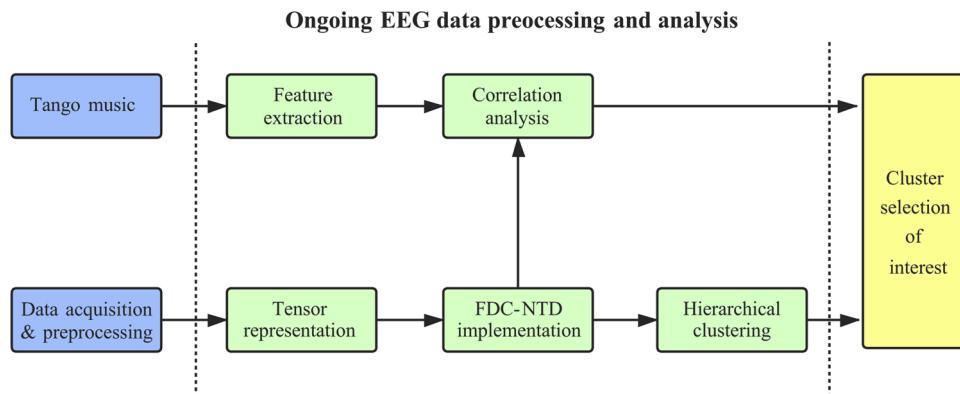
---

## 3. Experiments and methods

In this section, we provide a comprehensive framework for ongoing EEG data processing and analysis based on the proposed FDC-NTD algorithm, aiming to find commonly appearing brain activities elicited by naturalistic and continuous musical stimulus. Undoubtedly, such common information shared by the majority of subjects is more reliable than individual information from particular subject (Cong et al., 2013a; Lee and Choi, 2009). Through TFR and FDC-NTD algorithm, tensors with dimensions of channel  $\times$  time  $\times$  frequency can be constructed and decomposed into common and individual components in the spatial, spectral and temporal modes. Meanwhile, five long-term musical features can be extracted from the music. Correlation analysis and hierarchical clustering are performed together to determine the cluster of interest. Fig. 3 illustrates the overall flow diagram of ongoing EEG data processing and analysis.

### 3.1. Data acquisition & preprocessing

The ongoing EEG data collected from 14 participants aged from 20 to 46 years old were used in this study. All participants were right-handed and healthy, without musical expertise and any problem of hearing loss or history of neurological disorders. The musical stimulus adopted an 8.5-min piece of modern tango by Astor Piazzolla (Alluri et al., 2012). The data were recorded using BioSemi bioactive electrode caps with the sampling rate of 2048 Hz, according to the international 10–20 system. The collected EEG data were preprocessed off-line using EEGLAB toolbox (Delorme and Makeig, 2004) and MATLAB R2016b,



**Fig. 3.** FDC-NTD-based ongoing EEG data analysis includes the following steps: (1) data acquisition & preprocessing; (2) musical feature extraction; (3) tensor representation; (4) FDC-NTD implementation; (5) correlation analysis; (6) hierarchical clustering; (7) cluster selection of interest.

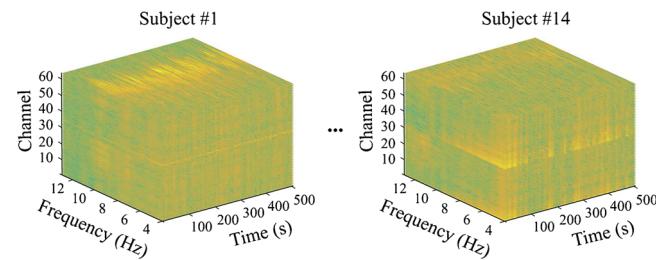
down-sampled to 256 Hz, and filtered by high-pass and low-pass filters with 1 Hz and 30 Hz cutoff frequencies. Detailed information about data acquisition and preprocessing can be found in our previous work (Cong et al., 2013a).

### 3.2. Tensor representation

TFR of the preprocessed EEG data was obtained by short-time Fourier transform (STFT). The Hamming window was adopted as the window function, with the window length of 3 s and 66.7% overlap ratio between windows. The number of Fourier points in each window was 1024, which was four times of the sampling rate. Power spectrum of EEG data are often evaluated in several frequency bands, such as delta (0.5~4 Hz), theta (4~8 Hz), alpha (8~13 Hz) and beta (13~30 Hz) (Siuly et al., 2016). According to previous work (Sammel et al., 2007; Lin et al., 2008, 2010; Shahin et al., 2009; Schaefer et al., 2011; Cong et al., 2012b, 2013a; Li et al., 2016), frequency fluctuations of brain activities elicited by musical stimulus are generally distributed in theta and alpha bands, hence in this study, frequency ranging from 4 to 13 Hz was used for further analysis. Therefore, third-order tensors including the spectrograms of EEG data with dimensions of 46 (frequency bins)  $\times$  510 (time samples)  $\times$  64 (space channels) were generated for 14 participants, as shown in Fig. 4.

### 3.3. Musical feature extraction

In this study, five long-term musical features (tonal and rhythmic, Fig. 5) were extracted by a frame-by-frame analysis method, providing a bridge for analyzing the connections between musical stimulus and ongoing EEG (Alluri et al., 2012; Cong et al., 2013a; Zhu et al., 2019). The duration of each frame was 3 s and overlap ratio between frames was 66.7%, which was consistent with the window settings in the STFT of EEG data. Therefore, the corresponding temporal courses with 510 samples were generated for those features. Furthermore, for the tonal features, Mode denotes the strength of major or minor mode, and Key Clarity is the measure of the tonal clarity (Alluri et al., 2012). For the rhythmic features, Fluctuation Centroid is defined as the geometric mean of the fluctuation spectrum, and it represents the global repartition of rhythm periodicities within the range of 0~10 Hz, indicating the average frequency of these periodicities (Alluri et al., 2012). Fluctuation entropy is the Shannon entropy of the fluctuation spectrum, and it represents the global repartition of rhythm periodicities. Pulse Clarity is regarded as an estimate of clarity of the pulse. The details of musical features and extraction method can be found in Latrillot and Toivainen (2007), Alluri et al. (2012), and Cong et al. (2013a).



**Fig. 4.** Third-order tensors for 14 participants, and each tensor includes three dimensions of 64 channels, 46 frequency bins (4~13 Hz) and 510 time samples.

### 3.4. FDC-NTD implementation

#### 3.4.1. Parameter initialization

The input factor matrices of spatial, spectral and temporal modes were initialized with uniformly distributed pseudorandom numbers generated by MATLAB function `rand`.

#### 3.4.2. Termination criteria

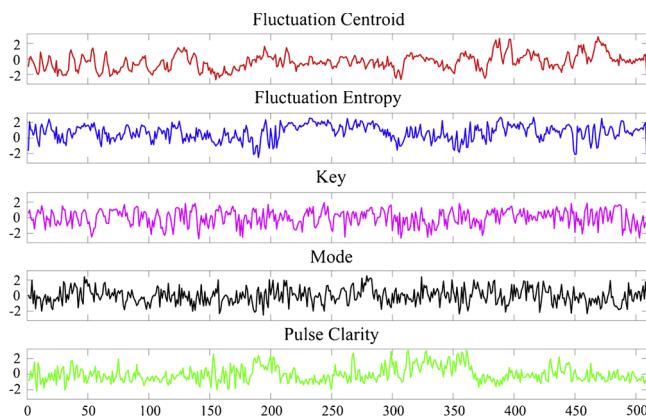
In this study, two iteration termination criteria of FDC-NTD algorithm were adopted. (a)  $|Fit_{new} - Fit_{old}| < \varepsilon$ , it means that the Fit change between the adjacent iterations should be smaller than the predefined threshold (e.g.,  $\varepsilon = 1e-6$ ). Tensor fitting is defined as  $Fit \triangleq \frac{1}{S} \sum_{s=1}^S [1 - \|\mathcal{X}^{(s)} - \hat{\mathcal{X}}^{(s)}\|_F / \|\mathcal{X}^{(s)}\|_F]$ , where  $\mathcal{X}^{(s)}$  and  $\hat{\mathcal{X}}^{(s)}$  are original and recovered tensors respectively. Furthermore, the relative error is defined as  $RelErr \triangleq \sum_{s=1}^S [\|\mathcal{X}^{(s)} - \hat{\mathcal{X}}^{(s)}\|_F / \|\mathcal{X}^{(s)}\|_F]$ . (b) The maximum number of iterations is no more than 1000.

#### 3.4.3. Component number selection

To determine the number of components, a multi-dimensional model order selection technique termed as *R*-dimensional minimum description length (*R*-D MDL, da Costa et al., 2011) was adopted in this study. The *R*-D MDL method based on information theoretic criterium extended 1-D MDL (modified MDL) method to the multi-dimensional case by using the global eigenvalues, providing low computational complexity and maintaining good performance even for lower SNR scenarios (da Costa, 2010). Its optimization problem is given as follows:

$$\hat{d} = \operatorname{argmin}_P I(\alpha^{(G)} - P) \log \left( \frac{g^{(G)}(P)}{\alpha^{(G)}(P)} \right) + p(P, I, \alpha^{(G)}) \quad (6)$$

where the penalty function  $p(P, I, \alpha^{(G)})$  is chosen as  $\frac{1}{2}P(2\alpha^{(G)} - P)\log(I)$ .  $\hat{d}$  denotes an estimated of the true model order  $d$ ,  $g^{(G)}(P)$  and  $\alpha^{(G)}(P)$  represent the geometric and arithmetic means of



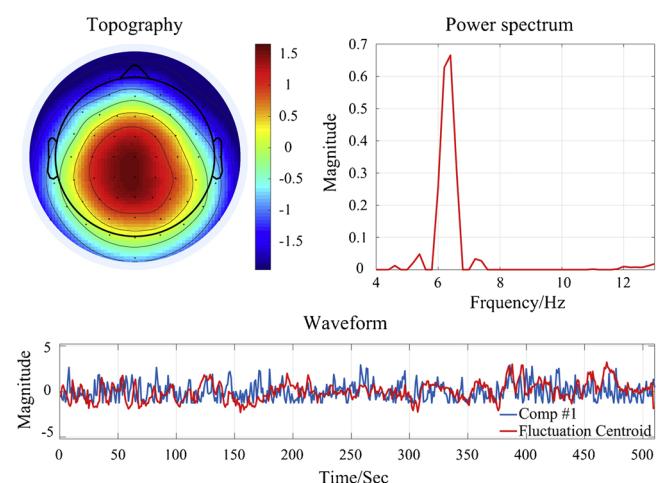
**Fig. 5.** Temporal courses of five musical features, including Fluctuation Centroid, Fluctuation Entropy, Key Clarity, Mode and Pulse Clarity.

the  $P$  smallest global eigenvalues, respectively.  $I$  is set as  $I = \max(I_1, I_2, \dots, I_N)$ , and  $\alpha^{(G)}$  is the total number of adaptively defined global eigenvalues. Therefore, for the EEG tensors of 14 subjects, the number of components were respectively selected as {44, 34, 36, 38, 36, 39, 35, 35, 34, 37, 33, 36, 34, 35} via R-D MDL algorithm adapted from IPM software.<sup>1</sup>

Regarding the number of common components, we first ran 10 times of individual fast-HALS decompositions on each EEG tensor and then performed correlation analysis on the spatial/spectral modes between any two subjects successively. We obtained the averaged correlation coefficients  $r = 0.8714$  and  $r = 0.9031$  at level  $p < 0.001$  on the two modes respectively. There was an average of 26 high-correlated spatial/spectral components between any two subjects (the correlation coefficients of 0.7 to 1 are considered to represent high or very high correlations; Asuero et al., 2006). Therefore, considering the hypothesis of double-coupled constraint, we set  $L_1 = L_2 = 26$  and  $L_3 = 0$ .

### 3.5. Correlation analysis

After extracting the components using FDC-NTD algorithm, it is necessary to determine which ones are relevant to musical stimulus. According to our previous work Cong et al. (2012b, 2013a), correlation analysis was conducted between temporal courses of extracted temporal components and temporal courses of musical features, aiming to find the brain activities elicited by musical stimulus. Pearson correlation analysis was applied to calculate the correlation coefficient, and then Monte Carlo method and permutation test were used to determine significant thresholds of correlation and correct for multiple comparisons (Alluri et al., 2012; Groppe et al., 2011). Moreover, a threshold (at level  $p < 0.05$ ) of correlation coefficient was calculated by a musical feature and  $R^{(s)}$  temporal components from each participant. Those temporal components which are significantly correlated with temporal courses of musical features were considered to be relevant to musical stimulus, and will be of interest and further analyzed. Fig. 6 shows an example of spatial, spectral and temporal components of EEG data, represented as topography, power spectrum and waveform, respectively. The temporal component was significantly correlated with the musical feature of 'Fluctuation Centroid' (i.e., ). In addition, we can see that the occipital region of subject #11 is activated with theta oscillation.



**Fig. 6.** The topography, power spectrum and waveform of the 1st EEG components from subject #11 of Run #1. The temporal course of Comp #1 is significantly correlated with the temporal course of 'Fluctuation Centroid' with a correlation coefficient of 0.1128 and a significant correlation threshold of 0.1064 (at level  $p < 0.05$ ).

### 3.6. Hierarchical clustering

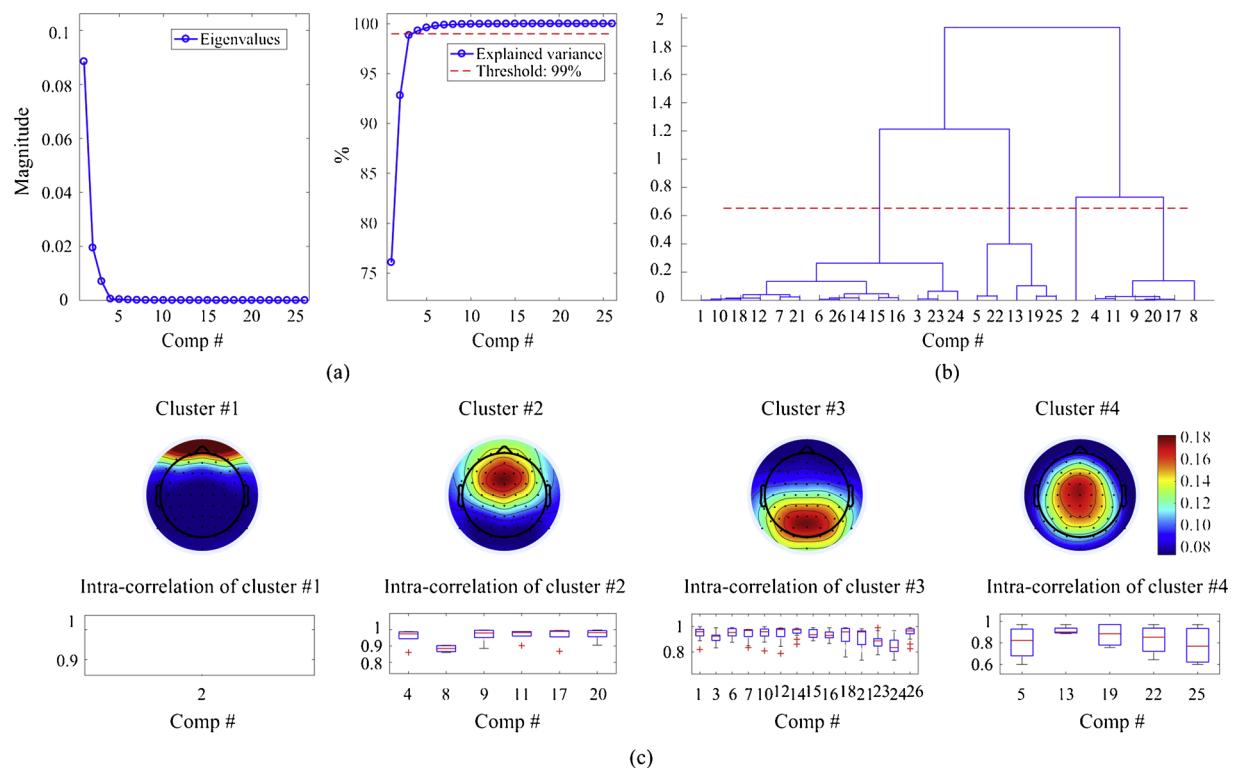
It should be noted that the correlations in Fig. 1 include two parts: auto-correlation (intra-subject) and cross-correlation (inter-subject). Therefore, in addition to the high spatial cross-correlation of inter-subject, we also found the high correlations between spatial components within the subject. Different from imposing coupled constraints to address inter-subject correlations, in this study, we adopted hierarchical clustering to merge the highly correlated spatial components within the subject. Through the FDC-NTD algorithm,  $L_1$  common spatial and  $L_2$  spectral components from the ongoing EEG can be extracted. By virtue of the coupled constraints across subjects, we only need to cluster the  $L_1$  common spatial components.

Moreover, clustering  $L_1$  spatial components is simpler than clustering all of the spatial components extracted from 14 subjects by independent component analysis (ICA) individually (Cong et al., 2013a). For stable clustering, we adopted hierarchical agglomerative clustering algorithm, in which complete linkage was used to calculate the furthest distance (here we used correlation) between pairs of clusters and the pairs of clusters with the nearest distance were merged. We applied PCA to  $L_1$  spatial components, and the number of principal components with 99% explained variance was selected as the number of clusters. Figs. 7 and 8 give some illustrations and results relevant to hierarchical clustering analysis about 26 spatial components. As shown in Fig. 7(a), when the cumulative explained variance exceeds the threshold of 99% (red dash line), 4 is selected as the number of clusters. Therefore, the hierarchical tree in the hierarchical clustering is split into 4 clusters by cutting branches (red dash line). Fig. 7(c) shows the averaged spatial maps of each cluster. In cluster #1, there is only one component. For clusters #2, #3 and #4, the mean correlation coefficients between spatial components within each cluster are 0.9518, 0.9268 and 0.8397, and the corresponding standard deviations (SDs) are 0.0502, 0.0596 and 0.1314, respectively. This indicates that the components are highly correlated with each other in each cluster. The low correlations of averaged spatial components between clusters, as shown in Fig. 8, also demonstrate the accuracy of clustering results.

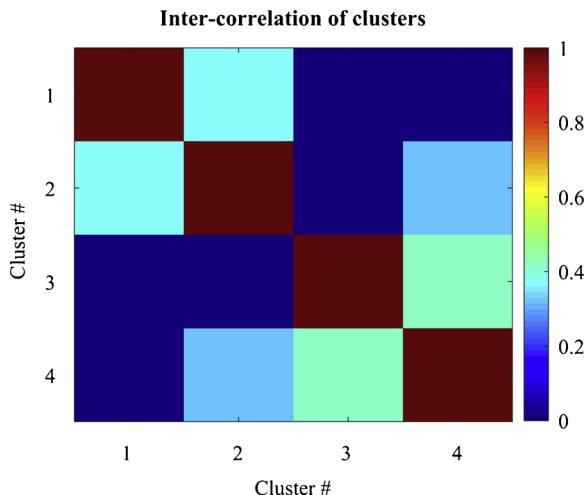
### 3.7. Cluster selection of interest

In group analysis, phenomena commonly appearing in most subjects are more attractive than the individual ones of a particular subject.

<sup>1</sup> <https://lasp.unb.br/index.php/publications/softwares/>



**Fig. 7.** Hierarchical clustering results of 26 spatial components of Run #7. (a) Selection of the number of clusters; (b) dendrogram output of hierarchical clustering; (c) averaged topographies of clusters and correlations between components within the clusters.



**Fig. 8.** Correlations of spatial maps (the averaged spatial component in each cluster) between clusters.

Therefore, in this section, our object is to determine the brain activities shared by the majority of subjects from the ongoing EEG data. Through comprehensively analyzing the results from correlation analysis (temporal components and musical features) and hierarchical clustering (common spatial components), we can obtain  $q$  clusters of spatial components whose parallel temporal components satisfy the threshold of significant correlation coefficients (with any musical feature). If the number of subjects contributing to a cluster exceeds half of the total number of subjects, the cluster will be selected as the cluster of interest and kept for further analysis in this study (Cong et al., 2013a). For each cluster of interest, the corresponding brain responses in most subjects are considered to be related to the musical stimulus. In Table 1, the subjects contributing to the 4 clusters in Fig. 7 are listed separately, and clusters #2, #3 and #4 are selected as the cluster of interest based on the predefined criterion. For the sake of simplicity, here we integrate the clustered spatial components and their corresponding temporal and spectral components into category of the cluster of interest.

#### 4. Results

The uniqueness of the decomposition is critical to the interpretation of extracted components (Hunyadi et al., 2017). For the ongoing EEG data contaminated with noise, it is difficult to verify that the recovered

**Table 1**

Subject distribution of 4 clusters after comprehensively analyzing the results of correlation analysis and hierarchical clustering of Run #1. ‘1’ and ‘-’ indicate that the subject contributes or does not contribute to the cluster. The number of subjects contributing to cluster #2, #3 and #4 exceed half of the total number of subject.

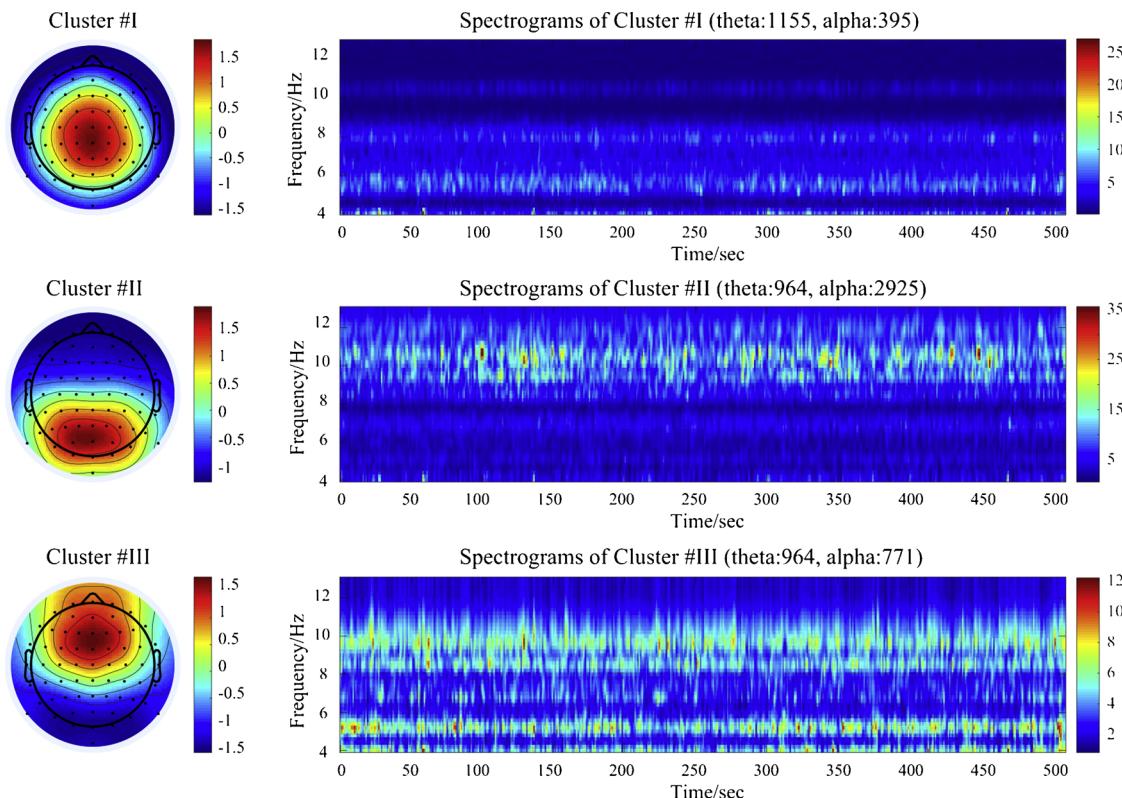
Subject	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12	#13	#14	Total
Cluster #1	-	-	-	-	-	-	1	-	-	1	-	-	-	1	3
Cluster #2	1	-	1	-	1	-	-	-	-	1	1	1	1	1	8
Cluster #3	1	1	1	-	1	1	1	1	1	1	1	1	1	-	12
Cluster #4	1	1	1	-	1	1	1	-	-	1	-	1	-	-	8

**Table 2**

Performance comparison between FDC-NTD and LCPTD algorithms. Evaluation indices include Obj, RelErr, Fit and Time averaged from 100 runs, and the number of occurrence of Clusters #I, #II and #III in 100 runs.

	Obj	RelErr	Fit	Time/s	Cluster #I	Cluster #II	Cluster #III
FDC-NTD	<b>1.1263e + 11</b>	<b>3.9202</b>	<b>0.7200</b>	164.06	83/100	100/100	96/100
LCPTD	1.1669e + 11	3.9634	0.7169	<b>153.00</b>	82/100	100/100	93/100

Obj: objective function value; RelErr: relative error; Fit: tensor fitting; Time: running time.

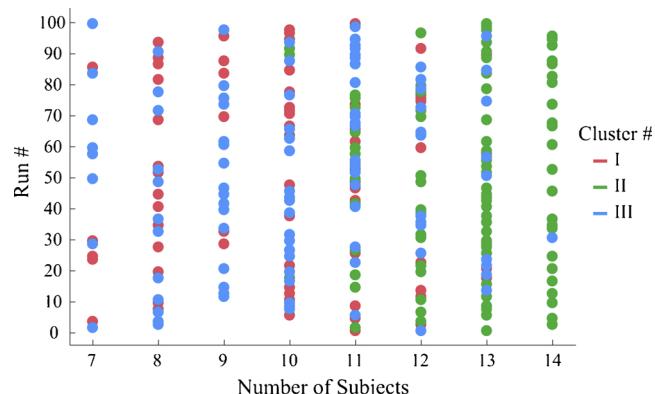


**Fig. 9.** Illustrations of averaged clusters of interest #I, #II and #III obtained from 100 runs. Spatial information, presented by the averaged topographies (left column), indicate the activations of centro-parietal, occipito-parietal and frontal regions of the brain elicited by musical stimulus, respectively. Overall spectrograms of clusters #I, #II and #III (right column) from 100 runs illustrate the frequency oscillations over the entire period. For cluster #I, the numbers of theta and alpha components are 1155 and 395. For cluster #II, the numbers of theta and alpha components are 964 and 2925. For cluster #III, the numbers of theta and alpha components are 964 and 771.

components are the true versions of the observed tensors. Therefore, in order to validate the reliability and stability of the solutions, we performed FDC-NTD algorithm 100 times on the generalized LCPTD model in this experiment. At the same time, to prove the validity of the generalized model, we also ran 100 times of Fast-HALS algorithm on the LCPTD model (for simplicity, we named it the LCPTD algorithm). According to Cong et al. (2012b), we chose 35 as the number of components in the LCPTD algorithm.

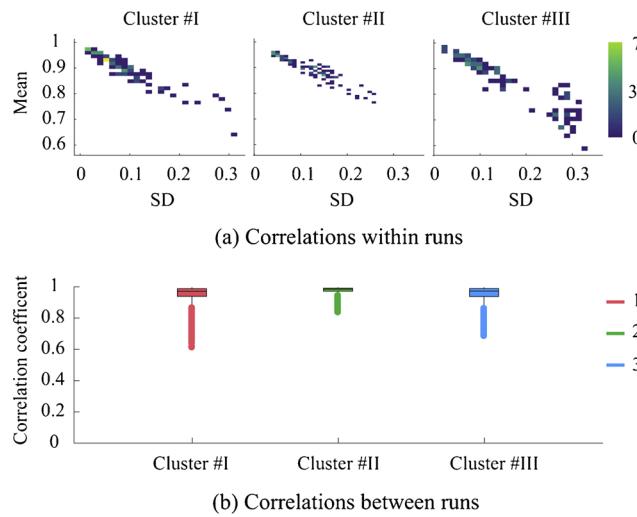
The experiments were carried out with the following computer configurations: CPU: Intel Core i5-7500 @ 3.40Hz 3.41Hz; Memory: 16.00 Gb; System: 64-bit Windows 10; Matlab R2016b. For data visualization, Figs. 10 and 11 were plotted using the graphics toolbox gramm<sup>2</sup> (Morel, 2018).

As shown in Table 2, we compare FDC-NTD and LCPTD algorithms in five aspects including objective function value (Obj), relative error



**Fig. 10.** Distribution of the number of subjects contributing to the clusters #I, #II and #III in 100 runs. Cluster #1, #II and #III appeared 83, 100 and 96 times in 100 runs, respectively.

<sup>2</sup> <https://github.com/piermorel/gramm>



**Fig. 11.** Spatial correlations within and between runs for clusters #I, #II and #III in 100 runs. (a) Distribution of means and SDs of correlation coefficients calculated by the internal spatial components in each run for clusters #I, #II and #III. (b) Illustration of correlation coefficients calculated by the averaged spatial components between runs for clusters #I, #II and #III.

(RelErr), tensor fitting (Fit), running time (Time) and occurrence probabilities of clusters #I, #II and #III. For the averaged Obj, RelErr and Fit value, the FDC-NTD algorithm performs slightly better than LCPTD algorithm, but requires more consuming time. Through correlation analysis, hierarchical clustering and cluster selection of interest, three kinds of clusters of interest are obtained from 100 decomposition results. For example, the corresponding averaged topographies obtained by FDC-NTD algorithm are plotted in Fig. 9. For the FDC-NTD algorithm, the probabilities of clusters #I, #II, and #III occurring in 100 runs reach 83% (83/100), 100% (100/100) and 96% (96/100), while the occurrence probabilities of clusters #I, #II, and #III obtained by LCPTD algorithm are 82% (82/100), 100% (100/100) and 93% (93/100). It should be noted that the FDC-NTD algorithm is more stable than the LCPTD algorithm.

Fig. 9 also illustrates the overall spectrograms of clusters #I, #II and #III obtained in FDC-NTD algorithm. The spectrogram can be generated by back-projection of spectral and temporal components, presenting a qualitative and quantitative evaluation of frequency oscillations over the entire period. Regarding cluster #I, the topography reveals that the centro-parietal region of the brain was activated with quite a lot theta oscillations (4~7 Hz, 74.52%, 1155/1550) but little alpha oscillations (around 10 Hz, 25.48%, 395/1550). Conversely, the occipito-parietal region of the brain is activated with significant alpha oscillations (8~13 Hz, 75.21%, 2925/3889), accompanied by a small amount of theta oscillations (4~8 Hz, 24.79%, 964/3889) in cluster #II. In addition, we also obtained the topography representing the activation of

frontal region of the brain, as shown in cluster #III. The frequency oscillations of cluster #III are distributed in the range of 4~11 Hz (theta-55.56%, 964/1735, alpha-44.44%, 771/1735).

The number of subjects contributing to the clusters #I, #II and #III in each run is visualized in Fig. 10. Regarding cluster #II, the number of subjects in each run is concentrated at 12, 13 and 14 (green circles). The number of subjects contributing to clusters #I and #III in 100 runs is mostly distributed in 7, 8, 9 and 10 (pink and blue circles). Furthermore, the gramm plot of spatial correlations within and between runs for clusters #I, #II and #III in 100 runs is illustrated in Fig. 11. From Fig. 11(a), we can see that the distribution of cluster #II is more compressed than that of clusters #I and #III. Regarding cluster #II, the averaged mean and SD of correlations in 100 runs are 0.8949 and 0.0949. For cluster #I, the averaged mean and SD of correlations in 83 runs are 0.8959 and 0.0985. For cluster #III, the averaged mean and SD of correlations in 96 runs are 0.8459 and 0.1402, which are inferior to the ones of clusters #I and #II. Fig. 11(b) illustrates the correlation coefficients between runs for clusters #I, #II and #III. We can find that the cluster #II is more stable with less outliers. Regarding cluster #II, the mean and SD of correlation coefficient are 0.9813 and 0.0182. For cluster #I, the mean and SD are 0.9570 and 0.0492. For cluster #III, the mean and SD are 0.9591 and 0.0449. From the results in Fig. 11, we can conclude that spatial components within and between runs for each cluster are highly correlated with each other. It strongly demonstrates the accuracy and stability of clustering results obtained by FDC-NTD algorithm.

## 5. Discussion

Tensor decomposition and group-level ICA methods have been generally used to extract stimulus-elicited components from a higher-order EEG tensor or concatenated EEG matrix of different subjects for group-level analysis in the cognitive research (Eichele et al., 2011; Cong et al., 2012b, 2015a). Only if the number of sources or the hidden information in EEG data of different subjects is consistent, the above methods will make sense to stack/concatenate the data to the tensor/matrix for analysis (Cong and He, 2013b; Wang et al., 2018b). Individual ICA method is not naturally suited to explore group inferences since the result fusion across individuals is sometimes a non-trial problem (Eichele et al., 2011). Coupled tensor decomposition, an extension of tensor decomposition to multi-block tensors, has been widely utilized to explore the potential common phenomenon across tensors (Sørensen et al., 2015; Ermiş et al., 2015; Gong et al., 2016). Therefore, in this study, a comprehensive framework based on coupled tensor decomposition applying to group-level analysis of ongoing EEG data during free listening to a 8.5-min long tango music was investigated.

The theoretical principle of coupled tensor decomposition is that the different tensors share some of the same or partially identical factor matrices (Zhou et al., 2016; Yokota et al., 2012). For the ongoing EEG data, we indeed found that there were highly correlated information in both spatial and spectral modes among the tensors represented from 14 subjects (see it in Fig. 1), which can be regarded as a prerequisite for applying coupled tensor decomposition in this study. Meanwhile, there is also individual information for each tensor, which may lead to inconsistent number of components among tensors. Considering the

common and individual information with the inconsistent number of components, we extended the LCPTD model to a general case. For the validation of the generalized model, we ran 100 times of two algorithms to compare their performance as shown in Table 2. It indicates that the our generalized LCPTD model has higher model fit and stronger robustness than the LCPTD model, and the performance development depends to a large extent on the actual number of components we use. For clusters whose occurrence rate are not 100%, this is mainly due to the inconsistency of local optimal solutions caused by random initialization in 100 algorithm implementations. Wang et al. demonstrated that sparsity regularization can improve the extraction stability of EEG components (Wang et al., 2018a), which provides a good perspective for our future work.

From the results of 100 FCD-NTD algorithm implementations, we found that the brain activities of selected clusters #I, #II and #III relevant to the musical stimulus were mainly distributed in the bands of 4~8 Hz (theta), 8~13 Hz (alpha) and 4~11 Hz, and located in the centro-parietal, occipito-parietal and frontal regions respectively. For the same data in this study, such theta and alpha activities in the central and occipital regions were reported in Cong et al. (2013a), but no such activities in the frontal region. For the ICA-based method in Cong et al. (2013a), posterior K-means clustering was adopted to cluster all of the spatial components extracted from 14 subjects individually. However, the prior coupling information present in spatial and spectral modes was not employed when extract the hidden information, which may result in the failure of information extraction. In addition, clustering  $L_{1,2}$  common spatial components is obviously much simpler and stable than clustering all of the spatial components in Cong et al. (2013a), where the mean correlation coefficients between spatial maps in clusters #I and #II were only 0.85 and 0.81 respectively. Regarding the clusters #I and #II, the significant theta and alpha oscillations were also reported in Li et al. (2016), where Li et al. utilized tensor decomposition imposing EEG phase characteristics to explore the brain responses to the naturalistic and continuous musical stimulus. Cong et al. extracted only alpha activity in the posterior region using fourth-order nonnegative tensor decomposition without considering the existence of individual information for each subject (Cong et al., 2012b). Compared to previous work, the proposed FDC-NTD algorithm can avoid strong constraint that imposes consistency on temporal, spatial and spectral modes between EEG tensors (Cong et al., 2012b; Li et al., 2016). In addition, it can utilize the multi-way structure of tensor-represented data and the coupled relationship across tensor blocks, and can decompose EEG tensors into common components and individual components in each mode. The extraction of common components among data makes it easier to discover the commonly appearing brain activities among majority of subjects. The high means and low SDs of correlation coefficients within/between 100 runs can demonstrate the

stability and practicability of coupled tensor decomposition applied to the group-level analysis of ongoing EEG data.

Besides the studies of ongoing EEG analysis elicited by naturalistic and continuous musical stimulus, previous work on the use of EEG activities to analyze emotion and musical stimuli can provide some solid references for the results in this study (Sammel et al., 2007; Lin et al., 2008, 2010; Schmidt and Trainor, 2001). During listening to the emotional music, the spectrum power asymmetry indexes located in the brain areas corresponding to RASM12 (namely, 12 symmetric electrode pairs including Fp1-Fp2, F7-F8, F3-F4, FT7-FT8, FC3-FC4, T7-T8, P7-P8, C3-C4, TP7-TP8, CP3-CP4, P3-P4, and O1-O2) are sensitive to the brain activations associated with emotion responses (Lin et al., 2008). Lin et al. found the frontal and parietal lobes across frequency bands including theta and alpha contributed a lot in the emotion recognition during music listening (Lin et al., 2010). According to Sammel et al. (2007), the increase of theta power over the frontal midline was associated with pleasant music, while the frontal alpha asymmetry of ongoing EEG activity was used to distinguish the emotional valence of musical stimuli (Schmidt and Trainor, 2001). The previous studies regarding EEG and musical stimuli verified the plausibility of our findings to some extent.

In conclusion, we proposed a comprehensive framework based on coupled tensor decomposition for the group analysis of ongoing EEG data, elicited by naturalistic and continuous musical stimulus. Specifically, the proposed framework includes the following seven steps: data acquisition & preprocessing, musical feature extraction, tensor representation, algorithm implementation, correlation analysis, hierarchical clustering and cluster selection of interest, aiming to discover commonly appearing brain activities among subjects. The results obtained in the proposed framework illustrate that our findings are in line with the results of previous studies, and it can be inferred that those brain activities we extracted are associated with musical stimulus. Furthermore, the proposed framework based on coupled tensor decomposition in this study provides a new perspective for the processing and analysis of multi- subject ongoing EEG data. Coupled tensor decomposition methods with different optimization strategies can be applied for comparison to find more convincing solutions when processing and analyzing ongoing EEG data, which will be one of our future work.

## Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant Nos. 91748105 and 81471742), the Fundamental Research Funds for the Central Universities [DUT2019] in Dalian University of Technology in China and the scholarships from China Scholarship Council (Nos. 201706060262 and 201706060263).

## Appendix A. Derivation of FDC-NTD algorithm

The minimized optimization problem in (2) can be converted into  $\max(R^{(s)})$  rank-1 tensor approximation problems via HALS algorithm (Cichocki et al., 2007), which can be solved sequentially and iteratively as follows:

$$\min D_F^{(r)}(\mathbf{u}_r^{(n,s)}) = \sum_{s=1}^S \|\mathcal{Y}_r^{(s)} - \mathbf{u}_r^{(1,s)} \circ \dots \circ \mathbf{u}_r^{(N,s)}\|_F^2 = \sum_{s=1}^S \|\mathcal{Y}_{r,(n)}^{(s)} - \mathbf{u}_r^{(n,s)} \{\mathbf{u}_r^{(s)}\} \odot_{-n}^T\|_F^2, \quad (7)$$

where  $\mathcal{Y}_r^{(s)} \doteq \mathcal{X}^{(s)} - \sum_{k \neq r}^{R^{(s)}} \mathbf{u}_k^{(1,s)} \circ \mathbf{u}_k^{(2,s)} \circ \dots \circ \mathbf{u}_k^{(N,s)}$  and  $\mathcal{Y}_{r,(n)}^{(s)}$  is the mode- $n$  matricization of  $\mathcal{Y}_r^{(s)}$ . Mathematically, using the trace property of matrix, (7)

$$\min D_F^{(r)}(\mathbf{u}_r^{(n,s)}) = \sum_{s=1}^S \text{tr}[(\mathcal{Y}_{r,(n)}^{(s)} - \mathbf{u}_r^{(n,s)} \{\mathbf{u}_r^{(s)}\}^{\odot T_n}) (\mathcal{Y}_{r,(n)}^{(s)} - \mathbf{u}_r^{(n,s)} \{\mathbf{u}_r^{(s)}\}^{\odot T_n})^T] \quad (8)$$

We can calculate the gradient of  $D_F^{(r)}(\mathbf{u}_r^{(n,s)})$  in (8) with respect to  $\mathbf{u}_r^{(n,s)}$  as:

$$\frac{\partial D_F^{(r)}(\mathbf{u}_r^{(n,s)})}{\partial \mathbf{u}_r^{(n,s)}} = \begin{cases} \sum_{s=1}^S [-2\mathcal{Y}_{r,(n)}^{(s)} \{\mathbf{u}_r^{(s)}\}^{\odot -n} + 2\mathbf{u}_r^{(n,s)} \{\mathbf{u}_r^{(s)}\}^{\odot T_n} \{\mathbf{u}_r^{(s)}\}^{\odot -n}], & r \leq L_n \\ -2\mathcal{Y}_{r,(n)}^{(s)} \{\mathbf{u}_r^{(s)}\}^{\odot -n} + 2\mathbf{u}_r^{(n,s)} \{\mathbf{u}_r^{(s)}\}^{\odot T_n} \{\mathbf{u}_r^{(s)}\}^{\odot -n}, & r > L_n \end{cases} \quad (9)$$

For the solution of  $\mathbf{u}_r^{(n,s)}$ , we only need to set the gradient in (9) to zero. Therefore, the learning rule of  $\mathbf{u}_r^{(n,s)}$  obtained via HALS strategy can be formulated as:

$$\mathbf{u}_r^{(n,s)} = \begin{cases} \sum_s [\mathcal{Y}_{r,(n)}^{(s)} \{\mathbf{u}_r^{(s)}\}^{\odot -n}] / \sum_s \gamma_r^{(n,s)}, & r \leq L_n \\ [\mathcal{Y}_{r,(n)}^{(s)} \{\mathbf{u}_r^{(s)}\}^{\odot -n}] / \gamma_r^{(n,s)}, & r > L_n, \end{cases} \quad (10)$$

where  $\{\mathbf{u}_r^{(s)}\}^{\odot T_n} \{\mathbf{u}_r^{(s)}\}^{\odot -n}$  is denoted by  $\gamma_r^{(n,s)}$  for simplicity, and it can be reformulated as  $\{\mathbf{u}_r^{(s)T} \mathbf{u}_r^{(s)}\}^{\odot -n}$ . As  $\|\mathbf{u}_r^{(n,s)}\| = 1$ ,  $n = 1 \dots N-1$ , the scaling coefficients  $\gamma_r^{(n,s)}$  can be abbreviated as (5). It is important to note that the calculation of  $\mathcal{Y}_{r,(n)}^{(s)} \{\mathbf{u}_r^{(s)}\}^{\odot -n}$  including the mode- $n$  matricization of  $\mathcal{Y}_r^{(s)}$  and the Khatri-Rao product of  $\mathbf{u}_r^{(m,s)}$ ,  $m = 1 \dots n-1, n+1, \dots, N$  needs to be performed in each iteration. It can result in rather high computational cost, especially for large-scale problems (Cichocki and Phan, 2009). Since the fast-HALS algorithm has been proven to be more efficient in Cichocki and Phan (2009) than HALS algorithm, in this study, we further extend it to the generalized LCPTD model. Inspired by fast-HALS algorithm,  $\mathcal{Y}_r^{(s)}$  can be represented as:

$$\mathcal{Y}_r^{(s)} = \mathcal{X}^{(s)} - \sum_{r=1}^{R^{(s)}} \mathbf{u}_r^{(1,s)} \circ \mathbf{u}_r^{(2,s)} \circ \dots \circ \mathbf{u}_r^{(N,s)} + \mathbf{u}_r^{(1,s)} \circ \mathbf{u}_r^{(2,s)} \circ \dots \circ \mathbf{u}_r^{(N,s)} = \mathcal{X}^{(s)} - \hat{\mathcal{X}}^{(s)} + \mathbf{u}_r^{(1,s)} \circ \mathbf{u}_r^{(2,s)} \circ \dots \circ \mathbf{u}_r^{(N,s)}. \quad (11)$$

By defining  $\zeta_r^{(n,s)} \doteq \mathcal{Y}_{r,(n)}^{(s)} \{\mathbf{u}_r^{(s)}\}^{\odot -n}$ , we substitute the mode- $n$  matricization of  $\mathcal{Y}_r^{(s)}$  in (11) into  $\zeta_r^{(n,s)}$  and it can be represented as:

$$\begin{aligned} \zeta_r^{(n,s)} &= [\mathcal{X}_n^{(s)} - \mathbf{U}^{(n,s)} \{\mathbf{U}^{(s)}\}^{\odot T_n} + \mathbf{u}_r^{(n,s)} \{\mathbf{u}_r^{(s)}\}^{\odot T_n} \{\mathbf{u}_r^{(s)}\}^{\odot -n} \\ &= \mathcal{X}_n^{(s)} \{\mathbf{u}_r^{(s)}\}^{\odot -n} - \mathbf{U}^{(n,s)} \{\mathbf{U}^{(s)}\}^{\odot T_n} \{\mathbf{u}_r^{(s)}\}^{\odot -n} + \mathbf{u}_r^{(n,s)} \{\mathbf{u}_r^{(s)}\}^{\odot T_n} \{\mathbf{u}_r^{(s)}\}^{\odot -n} \\ &= [\mathcal{X}_n^{(s)} \{\mathbf{U}^{(s)}\}^{\odot -n}]_r - \mathbf{U}^{(n,s)} [\{\mathbf{U}^{(s)}\}^{\odot T_n} \{\mathbf{U}^{(s)}\}^{\odot -n}]_r + \mathbf{u}_r^{(n,s)} \{\mathbf{u}_r^{(s)}\}^{\odot T_n} \{\mathbf{u}_r^{(s)}\}^{\odot -n} \\ &= [\mathcal{X}_n^{(s)} \{\mathbf{U}^{(s)}\}^{\odot -n}]_r - \mathbf{U}^{(n,s)} [\{\mathbf{U}^{(s)T} \mathbf{U}^{(s)}\}^{\odot -n}]_r + \gamma_r^{(n,s)} \mathbf{u}_r^{(n,s)} \end{aligned} \quad (12)$$

with

$$\{\mathbf{U}^{(s)T} \mathbf{U}^{(s)}\}^{\odot -n} = \{\mathbf{U}^{(s)T} \mathbf{U}^{(s)}\}^{\odot} \otimes (\mathbf{U}^{(n,s)T} \mathbf{U}^{(n,s)}). \quad (13)$$

At last, we obtain the learning rule of  $\mathbf{u}_r^{(n,s)}$  based on fast-HALS algorithm Cichocki and Phan (2009) in generalized LCPTD model (1) as follows:

$$\mathbf{u}_r^{(n,s)} = \begin{cases} \sum_s \zeta_r^{(n,s)} / \sum_s \gamma_r^{(n,s)}, & r \leq L_n \\ \zeta_r^{(n,s)} / \gamma_r^{(n,s)}, & r > L_n \end{cases} \quad (14)$$

with the definitions of  $\zeta_r^{(n,s)}$  in (12) and  $\gamma_r^{(n,s)}$  in (5). The mode- $n$  matricization  $\mathcal{X}_n^{(s)}$  in  $\zeta_r^{(n,s)}$  only needs to be performed once in initialization, which greatly improves the computation efficiency of the proposed FDC-NTD algorithm.

## References

- Acar, E., Aykut-Bingol, C., Bingol, H., Bro, R., Yener, B., 2007. Multiway analysis of epilepsy tensors. Bioinformatics 23, i10–i18. <https://doi.org/10.1093/bioinformatics/btm210>.
- Adeli, H., Ghosh-Dastidar, S., 2010. Automated EEG-Based Diagnosis of Neurological Disorders: Inventing the Future of Neurology. CRC press <https://doi.org/10.1201/9781439815328>.
- Alluri, V., Toivainen, P., Jääskeläinen, I.P., Glerean, E., Sams, M., Brattico, E., 2012. Large-scale brain networks emerge from dynamic processing of musical timbre, key and rhythm. Neuroimage 59, 3677–3689. <https://doi.org/10.1016/j.neuroimage.2011.11.019>.
- Asuero, A.G., Sayago, A., Gonzalez, A., 2006. The correlation coefficient: an overview.

- Crit. Rev. Anal. Chem. 36, 41–59. <https://doi.org/10.1080/10408340500526766>.
- Berger, H., 1929. Über das elektrenkephalogramm des menschen. Arch. Psychiatr. Nervenkrankh. 87, 527–570. <https://doi.org/10.1007/BF01797193>.
- Busch, N.A., Dubois, J., VanRullen, R., 2009. The phase of ongoing EEG oscillations predicts visual perception. *J. Neurosci.* 29, 7869–7876. <https://doi.org/10.1523/JNEUROSCI.0113-09.2009>.
- Calhoun, V.D., Liu, J., Adali, T., 2009. A review of group ICA for fMRI data and ICA for joint inference of imaging, genetic, and ERP data. *Neuroimage* 45, S163–S172. <https://doi.org/10.1016/j.neuroimage.2008.10.057>.
- Chen, X., Wang, Z.J., McKeown, M., 2016. Joint blind source separation for neurophysiological data analysis: multiset and multimodal methods. *IEEE Signal Process. Mag.* 33, 86–107. <https://doi.org/10.1109/MSP.2016.2521870>.
- Cichocki, A. 2013. Tensor decompositions: a new concept in brain data analysis? arXiv preprint arXiv:1305.0395, pp. 1–19.
- Cichocki, A., Phan, A.-H., 2009. Fast local algorithms for large scale nonnegative matrix and tensor factorizations. *IEICE Trans. Fundam. Electron. Commun. Comput. Sci.* 92, 708–721. <https://doi.org/10.1587/transfun.E92.A.708>.
- Cichocki, A., Zdunek, R., Amari, S.-i., 2007. Hierarchical ALS algorithms for nonnegative matrix and 3D tensor factorization. In: International Conference on Independent Component Analysis and Signal Separation. Springer. pp. 169–176. [https://doi.org/10.1007/978-3-540-74494-8\\_22](https://doi.org/10.1007/978-3-540-74494-8_22).
- Cong, F., Alluri, V., Nandi, A.K., Toivainen, P., Fa, R., Abu-Jamous, B., Gong, L., Craenen, B.G., Poikonen, H., Huotilainen, M., et al., 2013a. Linking brain responses to naturalistic music through analysis of ongoing EEG and stimulus features. *IEEE Trans. Multimed.* 15, 1060–1069. <https://doi.org/10.1109/TMM.2013.2253452>.
- Cong, F., He, Z., Hämäläinen, J., Leppänen, P.H., Lyytinen, H., Cichocki, A., Ristaniemi, T., 2013b. Validating rationale of group-level component analysis based on estimating number of sources in EEG through model order selection. *J. Neurosci. Methods* 212, 165–172. <https://doi.org/10.1016/j.jneumeth.2012.09.029>.
- Cong, F., Lin, Q.-H., Kuang, L.-D., Gong, X.-F., Astikainen, P., Ristaniemi, T., 2015a. Tensor decomposition of EEG signals: a brief review. *J. Neurosci. Methods* 248, 59–69. <https://doi.org/10.1016/j.jneumeth.2015.03.018>.
- Cong, F., Phan, A.H., Zhao, Q., Huttunen-Scott, T., Kaartinen, J., Ristaniemi, T., Lyytinen, H., Cichocki, A., 2012a. Benefits of multi-domain feature of mismatch negativity extracted by non-negative tensor factorization from EEG collected by low-density array. *Int. J. Neural Syst.* 22, 1250025. <https://doi.org/10.1142/S0129065712500256>.
- Cong, F., Phan, A.H., Zhao, Q., Nandi, A.K., Alluri, V., Toivainen, P., Poikonen, H., Huotilainen, M., Cichocki, A., Ristaniemi, T., 2012b. Analysis of ongoing EEG elicited by natural music stimuli using nonnegative tensor factorization. In: 2012 Proceedings of the 20th European Signal Processing Conference (EUSIPCO). IEEE. pp. 494–498.
- Cong, F., Ristaniemi, T., Lyytinen, H., 2015b. Advanced signal processing on brain event-related potentials: filtering ERPs in time. In: Frequency and Space Domains Sequentially and Simultaneously, vol. 13. World Scientific. <https://doi.org/10.1142/9306>.
- da Costa, J.P.C.L., 2010. Parameter Estimation Techniques for Multi-Dimensional Array Signal Processing. Shaker Verlag.
- da Costa, J.P.C.L., Roemer, F., Haardt, M., de Sousa, R.T., 2011. Multi-dimensional model order selection. *EURASIP J. Adv. Signal Process.* 2011, 26. <https://doi.org/10.1186/1687-6180-2011-26>.
- Delorme, A., Makeig, S., 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21. <https://doi.org/10.1016/j.jneumeth.2003.10.009>.
- Eichele, T., Rachakonda, S., Brakedal, B., Eikeland, R., Calhoun, V., 2011. EEGIFT: group independent component analysis for event-related EEG data. *Comput. Intell. Neurosci.* 2011, 9. <https://doi.org/10.1155/2011/129365>.
- Ermiş, B., Acar, E., Cemgil, A.T., 2015. Link prediction in heterogeneous data via generalized coupled tensor factorization. *Data Min. Knowl. Discov.* 29, 203–236. <https://doi.org/10.1007/s10618-013-0341-y>.
- Gong, X., Lin, Q., Cong, F., De Lathauwer, L., 2016. Double coupled canonical polyadic decomposition for joint blind source separation. *IEEE Trans. Signal Process.* 66, 3475–3490. <https://doi.org/10.1109/TSP.2018.2830317>.
- Gong, X., Wang, X., Lin, Q., et al., 2015. Generalized non-orthogonal joint diagonalization with LU decomposition and successive rotations. *IEEE Trans. Signal Processing* 63, 1322–1334. <https://doi.org/10.1109/TSP.2015.2391074>.
- Groppe, D.M., Urbach, T.P., Kutas, M., 2011. Mass univariate analysis of event-related brain potentials/fields I: a critical tutorial review. *Psychophysiology* 48, 1711–1725. <https://doi.org/10.1111/j.1469-8986.2011.01273.x>.
- Herrmann, C.S., 2001. Human EEG responses to 1–100 Hz flicker: resonance phenomena in visual cortex and their potential correlation to cognitive phenomena. *Exp. Brain Res.* 137, 346–353. <https://doi.org/10.1007/s002210100682>.
- Huber, R., Ghilardi, M.F., Massimini, M., Tononi, G., 2004. Local sleep and learning. *Nature* 430, 78. <https://doi.org/10.1038/nature02663>.
- Hunyadi, B., Dupont, P., Van Paesschen, W., Van Huffel, S., 2017. Tensor decompositions and data fusion in epileptic electroencephalography and functional magnetic resonance imaging data. *Wiley Interdiscip. Rev.: Data Min. Knowl. Discov.* 7, e1197. <https://doi.org/10.1002/widm.1197>.
- Jeong, J., 2004. EEG dynamics in patients with Alzheimer's disease. *Clin. Neurophysiol.* 115, 1490–1505. <https://doi.org/10.1016/j.clinph.2004.01.001>.
- Jespersen, K.V., Koenig, J., Jennum, P., Vuust, P., 2015. Music for insomnia in adults. *Cochrane Database Syst. Rev.* <https://doi.org/10.1002/14651858.CD010459.pub2>.
- Koelsch, S., 2012. *Brain and Music*. John Wiley & Sons.
- Kolda, T.G., Bader, B.W., 2009. Tensor decompositions and applications. *SIAM Rev.* 51, 455–500. <https://doi.org/10.1137/07070111X>.
- Latrillot, O., Toivainen, P., 2007. MIR in Matlab: a toolbox for musical feature extraction. *Proceedings of the International Conference on Music Information Retrieval*.
- Lee, H., Choi, S., 2009. Group nonnegative matrix factorization for EEG classification. *Artificial Intelligence and Statistics.* pp. 320–327.
- Li, J., Ji, H., Gu, R., Hou, L., Zhang, Z., Wu, Q., Lu, R., Li, M., 2016. Explore the brain response to naturalistic and continuous music using EEG phase characteristics. In: International Conference on Intelligent Computing. Springer. pp. 294–305. [https://doi.org/10.1007/978-3-319-42291-6\\_29](https://doi.org/10.1007/978-3-319-42291-6_29).
- Lin, Y.P., Wang, C.H., Jung, T.P., Wu, T.L., Jeng, S.K., Duann, J.R., Chen, J.H., 2010. EEG-based emotion recognition in music listening. *IEEE Trans. Biomed. Eng.* 57, 1798–1806. <https://doi.org/10.1109/TBME.2010.2048568>.
- Lin, Y.P., Wang, C.H., Wu, T.L., Jeng, S.-K., Chen, J.-H., 2008. Support vector machine for EEG signal classification during listening to emotional music. In: 2008 IEEE 10th Workshop on Multimedia Signal Processing. IEEE. pp. 127–130. <https://doi.org/10.1109/MMSP.2008.4665061>.
- Luck, S.J., 2014. *An Introduction to the Event-Related Potential Technique*. MIT Press.
- MacDonald, R., Kreutz, G., Mitchell, L., 2013. *Music, Health, and Wellbeing*. Oxford University Press.
- Maratos, A., Gold, C., Wang, X., Crawford, M., 2008. Music therapy for depression. *Cochrane Database Syst. Rev.* <https://doi.org/10.1002/14651858.CD004517.pub2>.
- Morel, P., 2018. Gramm: grammar of graphics plotting in Matlab. *J. Open Source Softw.* 3, 568.
- Mørup, M., 2011. Applications of tensor (multiway array) factorizations and decompositions in data mining. *Wiley Interdiscip. Rev.: Data Min. Knowl. Discov.* 1, 24–40. <https://doi.org/10.1002/widm.1>.
- Mørup, M., Hansen, L.K., Parnas, J., Arnfred, S.M., 2006. Decomposing the Time-Frequency Representation of EEG using Non-Negative Matrix and Multi-Way Factorization. Technical University of Denmark Technical Report. pp. 1–28.
- Mössler, K., Chen, X., Heldal, T.O., Gold, C., 2011. Music therapy for people with schizophrenia and schizophrenia-like disorders. *Cochrane Database Syst. Rev.* <https://doi.org/10.1002/14651858.CD004253.pub3>.
- Sammller, D., Grigutsch, M., Fritz, T., Koelsch, S., 2007. Music and emotion: electrophysiological correlates of the processing of pleasant and unpleasant music. *Psychophysiology* 44, 293–304. <https://doi.org/10.1111/j.1469-8986.2007.00497.x>.
- Schaefer, R.S., Vlek, R.J., Desain, P., 2011. Music perception and imagery in EEG: alpha band effects of task and stimulus. *Int. J. Psychophysiol.* 82, 254–259. <https://doi.org/10.1016/j.ijpsycho.2011.09.007>.
- Schmidt, L.A., Trainor, L.J., 2001. Frontal brain electrical activity (EEG) distinguishes valence and intensity of musical emotions. *Cogn. Emotion* 15, 487–500. <https://doi.org/10.1080/0269930126048>.
- Shahin, A.J., Trainor, L.J., Roberts, L.E., Backer, K.C., Miller, L.M., 2009. Development of auditory phase-locked activity for music sounds. *J. Neurophysiol.* 103, 218–229. <https://doi.org/10.1152/jn.00402.2009>.
- Siuly, S., Li, Y., Zhang, Y., 2016. *EEG Signal Analysis and Classification: Techniques and Applications*. Springer International Publishing. <https://doi.org/10.1007/978-3-319-47633-7>.
- Sørensen, M., Domanov, I., De Lathauwer, L., 2015. Coupled canonical polyadic decompositions and (coupled) decompositions in multilinear rank- $(L_{r,n}, L_{r,n}, 1)$  terms – part II: algorithms. *SIAM J. Matrix Anal. Appl.* 36, 1015–1045. <https://doi.org/10.1137/140956865>.
- Wang, D., Cong, F., Zhao, Q., Toivainen, P., Nandi, A.K., Huotilainen, M., Ristaniemi, T., Cichocki, A., 2016. Exploiting ongoing EEG with multilinear partial least squares during free-listening to music. In: 2016 IEEE 26th International Workshop on Machine Learning for Signal Processing (MLSP). IEEE. pp. 1–6. <https://doi.org/10.1109/MLSP.2016.7738849>.
- Wang, D., Wang, X., Zhu, Y., Toivainen, P., Huotilainen, M., Ristaniemi, T., Cong, F., 2018a. Increasing stability of EEG components extraction using sparsity regularized tensor decomposition. In: International Symposium on Neural Networks. Springer. pp. 789–799. [https://doi.org/10.1007/978-3-319-92537-0\\_89](https://doi.org/10.1007/978-3-319-92537-0_89).
- Wang, D., Zhu, Y., Ristaniemi, T., Cong, F., 2018b. Extracting multi-mode ERP features using fifth-order nonnegative tensor decomposition. *J. Neurosci. Methods* 308, 240–247. <https://doi.org/10.1016/j.jneumeth.2018.07.020>.
- Wang, X., Ristaniemi, T., Cong, F., 2019. Fast implementation of double-coupled non-negative canonical polyadic decomposition. In: International Conference on Acoustics, Speech, and Signal Processing. IEEE. pp. 8588–8592. <https://doi.org/10.1109/ICASSP40910.2019.8851910>.

- 1109/ICASSP.2019.8682737.
- Yokota, T., Cichocki, A., Yamashita, Y., 2012. Linked PARAFAC/CP tensor decomposition and its fast implementation for multi-block tensor analysis. In: International Conference on Neural Information Processing. Springer. pp. 84–91. [https://doi.org/10.1007/978-3-642-34487-9\\_11](https://doi.org/10.1007/978-3-642-34487-9_11).
- Zhou, G., Cichocki, A., Xie, S., 2012. Fast nonnegative matrix/tensor factorization based on low-rank approximation. IEEE Trans. Signal Process. 60, 2928–2940. <https://doi.org/10.1109/TSP.2012.2190410>.
- Zhou, G., Zhao, Q., Zhang, Y., Adali, T., Xie, S., Cichocki, A., 2016. Linked component analysis from matrices to high-order tensors: applications to biomedical data. Proc. IEEE 104, 310–331. <https://doi.org/10.1109/JPROC.2015.2474704>.
- Zhu, Y., Zhang, C., Toivainen, P., Huotilainen, M., Mathiak, K., Ristaniemi, T., Cong, F., 2019. Exploring frequency-dependent brain networks from ongoing EEG using spatial ICA during music listening. bioRxiv 509802.