Increasing stability of EEG Components Extraction Using Sparsity Regularized Tensor Decomposition

Deqing Wang[1,2], Xiaoyu Wang[1], Yongjie Zhu[1,2], Petri Toiviainen[3], Minna Huotilainen[4], Tapani Ristaniemi[2], and Fengyu Cong[1,2](✉)

[1] Department of Biomedical Engineering, Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology, Dalian 116024, China
[2] Faculty of Information Technology, University of Jyväskylä, Jyväskylä 40100, Finland
{deqing.wang, xiaoyu.wang0207, yongjie.zhu}@foxmail.com
cong@dlut.edu.cn, tapani.e.ristaniemi@jyu.fi
[3] Finnish Centre for Interdisciplinary Music Research, Department of Music, University of Jyväskylä, Jyväskylä 40100, Finland
petri.toiviainen@jyu.fi
[4] CICERO Learning network and Cognitive Brain Research Unit, Faculty of Educational Sciences, University of Helsinki., Helsinki 00014, Finland
minna.huotilainen@helsinki.fi

Abstract. Tensor decomposition has been widely employed for EEG signal processing in recent years. Constrained and regularized tensor decomposition often attains more meaningful and interpretable results. In this study, we applied sparse nonnegative CANDECOMP/PARAFAC tensor decomposition to ongoing EEG data under naturalistic music stimulus. Interesting temporal, spectral and spatial components highly related with music features were extracted. We explored the ongoing EEG decomposition results and properties in a wide range of sparsity levels, and proposed a paradigm to select reasonable sparsity regularization parameters. The stability of interesting components extraction from fourteen subjects' data was deeply analyzed. Our results demonstrate that appropriate sparsity regularization can increase the stability of interesting components significantly and remove weak components at the same time.

Keywords: Tensor Decomposition, Sparsity Regularization, Nonnegative Constraints, Ongoing EEG, Stability Analysis.

1 Introduction

In recent years, tensor decomposition [1] has gained more and more popularity for EEG data processing [2]. By time-frequency representation, multi-channel EEG data can be converted into a third-order (time × frequency × channel) tensor. In some EEG experiments, a seventh-order tensor may be generated potentially [2]. There are two basic tensor decomposition methods: CANDECOMP/PARAFAC decomposition and Tucker decomposition. But sometimes these basic decompositions on EEG data can't guarantee meaningful factors. For example, the third-order EEG tensor mentioned

above is a nonnegative tensor essentially. The extracted temporal components should be energy series which are nonnegative. The spectral components should be spectra which are nonnegative and usually very sparse. The spatial components are topographies which are sometimes also sparse. Constrained and regularized tensor decomposition can make the results meaningful and interpretable [1].

Nonnegativity is often achieved by constrained optimization methods. Nonnegative constrains can naturally give sparse results [3], but this sparsity is only a side effect and not controllable. l_1-norm regularization is an effective and widely applied method to impose sparsity explicitly [4], and has been employed for sparse and nonnegative tensor decomposition [5-7]. But these works regarding sparse regularization in tensor decomposition [5-7] only tested a few groups of sparsity regularization parameters, and demonstrated that their model can successfully impose sparsity on factors. No work, as far as we know, had studied how the change of sparsity level affects the results of tensor decomposition and the physical meanings in real application.

In this study, ongoing EEG tensor data under naturalistic modern tango music stimulus was analyzed. Sparse nonnegative CANDECOMP/PARAFAC decomposition was employed to extract groups of interesting components whose temporal components are highly correlated with music features and whose spatial components have dipolar topographies. In order to reveal a clear picture of mathematical properties and components' physical meanings for EEG at different sparsity level, we tested a large range of sparsity regularization parameters. We proposed a method to select reasonable regularization parameters that can best balance the data fitting and sparsity. After careful analyses and comparison, we found that when sparsity regularization is imposed on tensor decomposition, the stability of interesting components was increased significantly. In our results, appropriate sparsity regularization can also remove weak components and weak elements on nonzero sparse components such as spectra.

2 Sparse Nonnegative CANDECOMP/PARAFAC Decomposition

2.1 Notation

In this paper, a boldface lowercase letter, such as $\boldsymbol{x}$, denotes a vector; a boldface uppercase letter, such as $\boldsymbol{X}$, denotes a matrix; and a boldface Euler script letter, such as $\boldsymbol{\mathcal{X}}$, denotes a high order tensor. Operator $\circ$ denotes inner product of vectors, and $[\![\ ]\!]$ denotes Kruskal operator. $\|\cdot\|_F$ represents Frobenius norm, and $\|\boldsymbol{x}\|_1$ represents l_1-norm of vector $\boldsymbol{x}$. We call sparse nonnegative CANDECOMP/PARAFAC decomposition "sparse NCP" and the version without sparsity regularization "NCP" for short.

2.2 Mathematical Model

Given a tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$, sparse NCP is to solve the following minimization problem:

$$\min_{\boldsymbol{A}^{(1)},\cdots,\boldsymbol{A}^{(N)}} \frac{1}{2}\left\|\boldsymbol{\mathcal{X}} - [\![\boldsymbol{A}^{(1)},\cdots,\boldsymbol{A}^{(N)}]\!]\right\|_F^2 + \sum_{n=1}^{N}\beta_n \sum_{j=1}^{K}\left\|\boldsymbol{a}_j^{(n)}\right\|_1 \quad (1)$$

$$\text{s.t. } \boldsymbol{A}^{(n)} \geq 0 \text{ for } n = 1,\cdots,N,$$

where $\boldsymbol{A}^{(n)} \in \mathbb{R}^{I_n \times K}$ for $n = 1,\cdots,N$, and $\boldsymbol{a}_j^{(n)}$ represents the jth column of $\boldsymbol{A}^{(n)}$. β_n for $n = 1,\cdots,N$ are parameters of sparsity regularization items. K is the selected rank-1 tensor number, and the estimated factors in Kruskal operator can be represented by sum of K rank-1 tensors in outer product form:

$$[\![\boldsymbol{A}^{(1)},\cdots,\boldsymbol{A}^{(N)}]\!] = \sum_{j=1}^{K}\boldsymbol{a}_j^{(1)} \circ \cdots \circ \boldsymbol{a}_j^{(N)} \quad (2)$$

2.3 Model solution

Sparse NCP in (1) is a highly nonlinear and nonconvex model, which is non-trivial to solve and converge. Recently, Xu applied an efficient alternating proximal gradient (APG) method to solve nonnegative matrix and tensor decomposition [8]. Later, Xu extended APG method to solve nonnegative Tucker decomposition with l_1-norm sparsity regularizations [7]. Inspired by Xu's works [7], we utilized the same updating method in block coordinate descend (BCD) framework to solve problem (1).

Supposing $\widehat{\boldsymbol{A}}^{(n)}$ is an extrapolated point, $\widehat{\boldsymbol{G}}^{(n)}$ is the block-partial gradient at $\widehat{\boldsymbol{A}}^{(n)}$ and $L^{(n)}$ is a Lipschitz constant, factor matrix $\boldsymbol{A}^{(n)}$ is updated by

$$\boldsymbol{A}^{(n)} \leftarrow \underset{\boldsymbol{A}^{(n)}\geq 0}{\operatorname{argmin}}\left[\langle \widehat{\boldsymbol{G}}^{(n)}, \boldsymbol{A}^{(n)} - \widehat{\boldsymbol{A}}^{(n)}\rangle + \frac{L^{(n)}}{2}\left\|\boldsymbol{A}^{(n)} - \widehat{\boldsymbol{A}}^{(n)}\right\|_F^2 + \beta_n \sum_{j=1}^{K}\left\|\boldsymbol{a}_j^{(n)}\right\|_1\right] \quad (3)$$

which can be written in the closed form

$$\boldsymbol{A}^{(n)} \leftarrow \max\left(0, \widehat{\boldsymbol{A}}^{(n)} - \frac{\widehat{\boldsymbol{G}}^{(n)}}{L^{(n)}} - \frac{\beta_n \mathbf{1}_{I_n \times K}}{L^{(n)}}\right) \quad (4)$$

The detailed solution and convergent properties of APG can be found in [7, 8].

3 Materials and methods

3.1 EEG data and music signal

Data description. The data in this study is ongoing EEG of fourteen right-handed and healthy adults under continuous and naturalistic modern tango music stimulus. Short-time Fourier transform (STFT) was applied to the EEG data, and a third-order tensor of ongoing EEG was created for each subject with size of 510×146×64 on temporal, spectral and spatial mode respectively. In this paper, we represent the estimated factors of the third-order EEG tensor as $\boldsymbol{A}^{(\text{Temporal})}$, $\boldsymbol{A}^{(\text{Spectral})}$ and $\boldsymbol{A}^{(\text{Spatial})}$.

Acoustic features. Five long-term acoustic features, including two tonal features (Mode, Key Clarity) and three rhythmic features (Pulse Clarity, Fluctuation Centroid, Fluctuation Entropy), were extracted from the tango music [9]. STFT is also used for feature extraction and one acoustic feature temporal series contains 510 samples.

The detailed data collection experiment paradigm and data preprocessing procedures can be found in [10, 11]. Detailed acoustic features can be found in [9].

3.2 Correlation Analysis

According to previous studies [9, 10], we hypothesized that acoustic features (tonal and rhythmic components) could activate certain brain areas. We performed correlation analyses (by Pearson's correlation coefficient) between the time series of long-term acoustic features and the time series of temporal components from EEG tensor decomposition to find stimulus-related activations. Monte Carlo method and permutation tests were employed to compute the threshold of correlation coefficient [9, 10]. In the results of EEG tensor decomposition, the temporal components significantly correlated (at level $p < 0.05$) with any of the five acoustic features, and their corresponding spectral and spatial components are recorded for further investigations.

3.3 Sparsity Parameter Selection

When applying sparse NCP model to decompose EEG tensor, a key point is the selection of sparsity regularization parameters β_n in model (1), which balance the data fitting and sparsity level. Data fitting at iteration k is defined by

$$\text{Fit}_k = 1 - \frac{\left\| \mathcal{X} - [\![\boldsymbol{A}_k^{(1)}, \cdots, \boldsymbol{A}_k^{(N)}]\!] \right\|_F}{\|\mathcal{X}\|_F}, \tag{5}$$

which is a measure of similarity of estimated factors to original data tensor. The sparsity level of estimated factor $\boldsymbol{A}_k^{(n)}$ at iteration k is defined by

$$\text{Sparsity}_{\boldsymbol{A}_k^{(n)}} = \frac{\#\left\{ \boldsymbol{A}_k^{(n)}(i,j) < T_S \right\}}{I_n \times K}, \tag{6}$$

where $\#\{\cdot\}$ means the number of elements in factor $\boldsymbol{A}_k^{(n)}$ that satisfy the assumption. Strictly speaking, the factor sparsity should be measured by the number of elements that equal to zero. But, in practice, it is better to select a small positive sparsity threshold T_s. In this study, we select $T_s = 1\text{e} - 6$.

In order to reveal a broad picture of data properties and results at different sparsity levels, we test a large range of sparsity regularization parameters of β_ns for sparse NCP. The procedures are shown in the following steps.

Step 1. For the temporal and spatial factor, we select $\beta_{\text{temporal}} = \beta_{\text{spatial}} = e^{\boldsymbol{\lambda}_1}$ and $\beta_{\text{spectral}} = e^{\boldsymbol{\lambda}_2}$. $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_2$ contain N_1 and N_2 linearized number, then there will be $N_1 \times N_2$ different groups of parameters combination $[\beta_{\text{temporal}}, \beta_{\text{spectral}}, \beta_{\text{spacial}}]$.

Step 2. Using each group of parameters, we run sparse NCP 10 times, and record the average nonzero components number of spectral factor, the average fitting value using (5), and the average sparsity level of spectral factor using (6). Because the spectral components are usually very sparse, we just consider the spectral factor.

Step 3. Reorder all of groups of results based on spectral factor sparsity level in ascend order with in [0,1], and generate the 'Fit-Sparsity' curve which reveal the fitting change at different spectral sparsity value.

Step 4. According to the 'Fit-Sparsity' curve, identify the maximum effective sparsity level based on the relative fitting change (slope) defined as

$$\text{slope} = \frac{\Delta\text{Fit}}{\Delta\text{Sparsity}} = \frac{\text{Fit}(\text{Sparsity}_1) - \text{Fit}(\text{Spasity}_2)}{\text{Sparsity}_1 - \text{Sparsity}_2}. \tag{7}$$

When the slope at some sparsity point is very close to 0.5, the sparsity value and its corresponding sparsity parameters group β_ns are selected. We assume that the slope should not be less than -0.5, because after that the fitting value become poor dramatically and the tensor decomposition results may be not accurate.

3.4 Stability Analysis

By correlation analysis introduce in section 3.3, we can find the interesting components which are assumed to be stimulus-related activations. Tensor decomposition models of NCP with and without sparsity regularization are evaluated respectively. We evaluate the stability of these components using the following steps.

Step 1. Run model 5 times for one subject's EEG tensor, and record those groups of temporal, spectral and spatial components whose temporal courses are highly correlated with any of five music features. Keep those groups of components whose topographies (spatial components) are dipolar as templates [10]. One group components of template can be represented by a rank-1 tensor of inner production of components:

$$\boldsymbol{\mathcal{T}} = \boldsymbol{t}^{(\text{Temporal})} \circ \boldsymbol{t}^{(\text{Spectral})} \circ \boldsymbol{t}^{(\text{Spatial})}$$

Step 2. According to (2), after the tensor decomposition, K rank-1 tensors in outer product form will be obtained. Supposing it is the rth time decomposition, the correlation coefficient of the jth rank-1 tensor and the template is

$$\rho(j,r) = \text{corr}\left(\boldsymbol{a}_j^{(\text{Temporal})}, \boldsymbol{t}^{(\text{Tempral})}\right) \times \text{corr}\left(\boldsymbol{a}_j^{(\text{Spectral})}, \boldsymbol{t}^{(\text{Spectral})}\right) \times \text{corr}\left(\boldsymbol{a}_j^{(\text{Spatial})}, \boldsymbol{t}^{(\text{Spatial})}\right) \tag{8}$$

where $\text{corr}(\cdot,\cdot)$ is the calculation of Pearson's correlation coefficient, $j = 1, \cdots, K$ and $\rho(j,r) \in [0,1]$. Then, calculate the maximum correlation coefficient of the K rank-1 tensors with the template:

$$P(r) = \max_{j=1,\cdots,K} \rho(j,r) \tag{9}$$

Step 3. Run sparse NCP 100 times, and record all $P(r), r = 1, \cdots, 100$. Make a histogram of the 100 $P(r)$ values. Because $P(r) \in [0,1]$, the number of times within $[0.95,1]$ are recorded as stability measurement criterion for further analysis. The higher this criterion number is, the more stable sparse NCP performs.

Step 4. Within the 5 times runs in Step 1 many groups of templates that have very similar temporal, spectral and spatial components will be found. The highest criterion number of these templates after 100 times test will be kept.

4 Experiments and Results

4.1 Tensor Decomposition Implementation

Factor Initialization. All the factors are initialized using normally distributed rand numbers as in [8].

Stop Criteria. We stop the iteration during tensor decomposition by criterion of relative residual change (fit change) according to (5), when the following condition between two iterations are satisfied:

$$T_{\text{stop}} = |\text{Fit}_k - \text{Fit}_{k+1}| < \epsilon, \tag{10}$$

where, $\epsilon = 1\text{e} - 6$ in this study.

Model Order Selection (MOS). Before decompose each participant's EEG tensor, we should determine the initial components number for model (1). A simple and convenient way is employed in this study. We make spatial mode unfolding of the third-order EEG tensor yielding a 64×74460 matrix where temporal and spectral modes are merged. Then we perform PCA along spatial mode on this matrix and records the principal components number that give 99% explained variance for each subject's EEG tensor. The MOS for 14 subjects' EEG tensor are listed in **Table 1**.

Table 1. Model order selection of 14 participants' EEG tensor

Subject	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12	#13	#14
Model Order	37	44	52	34	38	34	51	38	45	31	57	40	53	43

4.2 Sparsity Regularization Parameters

For the temporal and spatial factor, we select $\beta_{\text{temporal}} = \beta_{\text{spatial}} = e^{\boldsymbol{\lambda}_1}$, where vector $\boldsymbol{\lambda}_1 = [-Inf, -5\text{:}\,0.2\text{:}\,0]$ in MATLAB format; for the spectral factor, we select $\beta_{\text{spectral}} = e^{\boldsymbol{\lambda}_2}$, where $\boldsymbol{\lambda}_2 = [-Inf, -5\text{:}\,0.2\text{:}\,1]$. The linear ranges of $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_2$ are selected by try and error method, and be changed slightly for different data. In our EEG data test, we found that using exponential form of a linearized vector parameters, the spectral factor sparsity level are approximately uniformly distributed within [0,1], which helps to analyze the fitting change at different sparsity levels. Vector $\boldsymbol{\lambda}_1$ contains 27 numbers, and $\boldsymbol{\lambda}_2$ contains 32 numbers, so there will be 27×32=864 groups of parameters. We expect to add strong sparsity parameters on spectral factor and weak sparsity parameters on temporal and spatial factors. All groups of parameters with $\boldsymbol{\lambda}_1 > \boldsymbol{\lambda}_2$ are removed, and finally, 509 groups are kept for test.

Each of the 509 parameter groups are tested 10 times on sparse NCP, and the average nonzero components number of spectral factor, tensor fitting value, and sparsity level of spectral factor are recorded. Fig. 1. shows the results of subject #1's tensor data after all the steps in Section 3.4. From Fig. 1. (a) we can find that, when we add more sparsity to spectral factor, some components become zeros which are weak signal components. Fig. 1. (b) shows the tensor decomposition fitting changes at

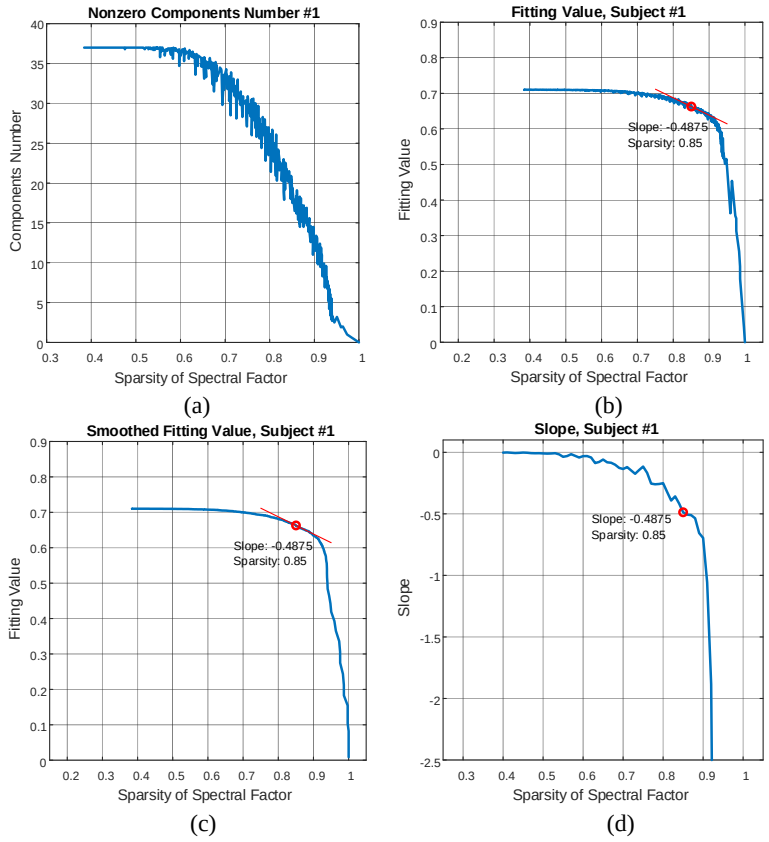


Fig. 1. Sparsity regularization parameters selection for subject #1

Table 2. Results of sparsity regularization parameters selection

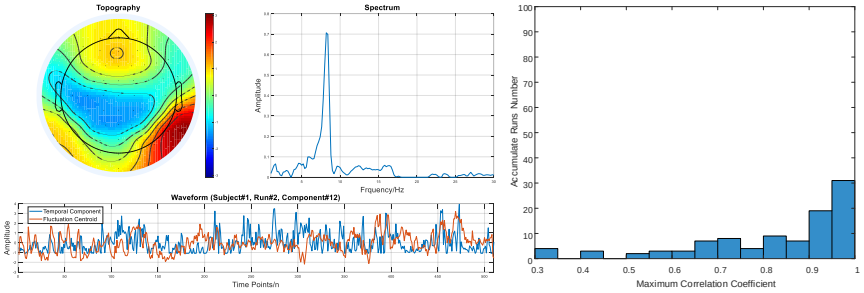
Subject	Identified Sparsity	Nonzero Comps	Fitting Value	$[e^{\lambda_1}, e^{\lambda_2}, e^{\lambda_1}]$	$[\lambda_1, \lambda_2, \lambda_1]$
#1	0.85	20.4	0.6667	[0.0608,1.2214,0.0608]	[-2.8,0.2,-2.8]
#2	0.87	28.5	0.6991	[0.0123,0.3012,0.0123]	[-4.4,-1.2,-4.4]
#3	0.86	42.5	0.6726	[0.0000,2.7183,0.0000]	[-Inf,1.0,-Inf]
#4	0.84	28.7	0.7683	[0.0183,0.2466,0.0183]	[-4.0,-1.4,-4.0]
#5	0.90	18.4	0.7241	[0.0150,0.1653,0.0150]	[-4.2,-1.8,-4.2]
#6	0.86	17.4	0.8084	[0.0123,0.0907,0.0123]	[-4.4,-2.4,-4.4]
#7	0.90	31.8	0.8095	[0.0067,0.1353,0.0067]	[-5.0,-2.0,-5.0]
#8	0.89	24.5	0.6870	[0.0183,1.8221,0.0183]	[-4.0,0.6,-4.0]
#9	0.88	28.8	0.6811	[0.0334,0.3679,0.0334]	[-3.4,-1.0,-3.4]
#10	0.88	19.5	0.7435	[0.0498,0.6703,0.0498]	[-3.0,-0.4,-3.0]
#11	0.93	37.2	0.7188	[0.0498,0.4493,0.0498]	[-3.0,-0.8,-3.0]
#12	0.90	19.8	0.7110	[0.0150,1.4918,0.0150]	[-4.2,0.4,-4.2]
#13	0.92	33.4	0.6992	[0.0224,0.2466,0.0224]	[-3.8,-1.4,-3.8]
#14	0.90	20.4	0.6899	[0.0123,0.6703,0.0123]	[-4.4,-0.4,-4.4]

different spectral factor sparsity levels. When sparsity increases, fitting value decreases gradually, because some weak components are removed. But after some sparsity point, the fitting value drops dramatically. In order to find this point, we smooth the fitting curve in (b) by low-pass filter, as is shown in (c). Then, we compute the slope according to equation (7), as is shown in (d). Based on curve (d), the largest sparsity level value 0.85 before slope -0.5 is selected. Then we search in all groups of sparsity regularization parameters and find the group $[\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \boldsymbol{\lambda}_1] = [-2.8, 0.2, -2.8]$ can attain the sparsity level 0.85 best. All subjects' results are shown in Table 2.

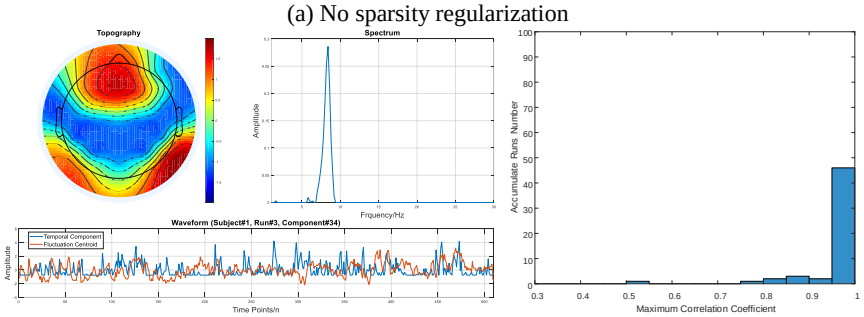
4.3 Sparsity and Stability Comparison

From Table 2 we identified the sparsity regularization parameters group $[\beta_{\text{temporal}}, \beta_{\text{spectral}}, \beta_{\text{spacial}}] = [e^{\boldsymbol{\lambda}_1}, e^{\boldsymbol{\lambda}_2}, e^{\boldsymbol{\lambda}_1}]$ for each subject's tensor data, using which we run tensor decomposition model (1). In order to make a comparison, we also test the original NCP without any sparsity regularizations by setting $[\beta_{\text{temporal}}, \beta_{\text{spectral}}, \beta_{\text{spacial}}] = [0,0,0]$.

Fig 2 shows two templates of subject #1 whose temporal components are both correlated with fluctuation centroid music feature series. One template has sparsity regularization imposed, while another doesn't. The histograms of stability analyses are also included. The template in Fig 2 (a) was selected from the 5th run of NCP and the 14th rank-1 components, and has the highest stability measurement number (the number between [0.95,1] on histogram) than other similar templates in the 5 runs.



(a) No sparsity regularization



(b) Sparsity regularization imposed

Fig. 2. Fluctuation centroid templates and their stability analysis from subject #1

Table 3. Stability analyses of sparsity regularization for all subjects

Subject	Music Feature	No Sparsity Regularization		With Sparsity Regularization	
		Template Index	Stability	Template Index	Stability
#1	PulseClarity	[Run#5, Comp#14]	42%	[Run#5, Comp#23]	**74%**
	FlucCentroid	[Run#2, Comp#12]	31%	[Run#3, Comp#34]	**46%**
#4	FlucCentroid	[Run#3, Comp#9]	86%	[Run#1, Comp#24]	**97%**
#7	Key	[Run#1, Comp#21]	57%	[Run#2, Comp#30]	**63%**
#8	Mode	[Run#4, Comp#35]	29%	[Run#1, Comp#33]	**60%**
	Mode	[Run#5, Comp#3]	16%	[Run#3, Comp#15]	**68%**
	Key	[Run#2, Comp#37]	17%	[Run#1, Comp#22]	**57%**
#9	Mode	[Run#3, Comp#37]	10%	[Run#4, Comp#23]	**54%**
#10	PulseClarity	[Run#4, Comp#27]	48%	[Run#2, Comp#25]	**49%**
	Key	[Run#4, Comp#1]	**57%**	[Run#2, Comp#20]	51%
#11	FlucCentroid	[Run#5, Comp#55]	14%	[Run#3, Comp#40]	**55%**
	Mode	[Run#3, Comp#1]	**73%**	[Run#1, Comp#56]	48%
#13	FlucEntropy	[Run#2, Comp#17]	29%	[Run#1, Comp#34]	**58%**
	FlucCentroid	[Run#1, Comp#5]	16%	[Run#5, Comp#52]	**37%**
	Key	[Run#3, Comp#51]	30%	[Run#2, Comp#20]	**72%**

Note: The stability is measured by number of maximum correlation coefficient within $[0.95, 1]$. But there are two exceptions of "#11-FlucCentroid" and "#13-FlucCentroid", which are measured within $[0.9, 0.95]$, because no significant value appear within $[0.95, 1]$.

The template in Fig 2 (b) was selected from the 5th run of sparse NCP and the 23th rank-1 components. All of the results for 14 subjects are summarized in Table 3 with selected templates and stability analyses. The selected pairs of templates for comparison had very similar topographies and spectra, and appeared in both situations with and without sparsity imposed on spectral components.

Comparing the histograms in Fig 2 (a) and (b), we found that, with sparsity regularization the stabilities of selected templates components have obvious increases. Table 3 further demonstrated that, for most of the subjects, when sparsity regularization is imposed on NCP for EEG data, the stability of extracting interesting components highly correlated with some certain music features are improved significantly.

From Fig 1 (a) we observed that when high sparsity regularization is imposed on spectral factor, more components of full zeros appear. This is a general phenomenon for all subjects according to the nonzero components number in Table 2 and initial model orders in Table 1. By carefully observing the spectrum in Fig 2 (b) compared to that in (a), we also find that small elements are suppressed when sparsity regularization is imposed. We believe adding sparsity regularization not only can remove weak components but also help to suppress weak information on nonzero components for nonnegative tensor decomposition.

5 Conclusion

In this study, we applied sparse nonnegative CANDECOMP/PARAFAC decomposition to ongoing EEG tensor data collected under naturalistic music stimulus. Interest-

ing temporal components correlated music features and corresponding spectral and spatial components are extracted. Mathematical properties and physical meanings of the decomposition with sparsity regularization are deeply analyzed in a large range of sparsity levels. We proposed a method to select reasonable sparsity regularization parameters based on the derivative of fitting-sparsity curve. It can be concluded from our results that appropriate sparsity regularization on tensor decomposition can improve the stability of interesting components and suppress weak signals.

Acknowledgements. This work was supported by the National Natural Science Foundation of China (Grant No. 81471742), the Fundamental Research Funds for the Central Universities [DUT16JJ(G)03] in Dalian University of Technology in China, and the scholarships from China Scholarship Council (Nos. 201600090043 and 201600090042).

References

1. Sidiropoulos, N.D., De Lathauwer, L., Fu, X., et al.: Tensor decomposition for signal processing and machine learning. IEEE Transactions on Signal Processing. 65(13), 3551-3582 (2017).
2. Cong, F., Lin, Q.-H., Kuang, L.-D., et al.: Tensor decomposition of EEG signals: a brief review. Journal of neuroscience methods. 248, 59-69 (2015).
3. Lee, D.D. and Seung, H.S.: Learning the parts of objects by non-negative matrix factorization. Nature. 401(6755), 788-791 (1999).
4. Donoho, D.L.: For most large underdetermined systems of linear equations the minimal L1 ‐ norm solution is also the sparsest solution. Communications on pure and applied mathematics. 59(6), 797-829 (2006).
5. Mørup, M., Hansen, L.K., and Arnfred, S.M.: Algorithms for sparse nonnegative Tucker decompositions. Neural computation. 20(8), 2112-2131 (2008).
6. Liu, J., Liu, J., Wonka, P., et al.: Sparse non-negative tensor factorization using columnwise coordinate descent. Pattern Recognition. 45(1), 649-656 (2012).
7. Xu, Y.: Alternating proximal gradient method for sparse nonnegative Tucker decomposition. Mathematical Programming Computation. 7(1), 39-70 (2015).
8. Xu, Y. and Yin, W.: A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion. SIAM Journal on imaging sciences. 6(3), 1758-1789 (2013).
9. Alluri, V., Toiviainen, P., Jääskeläinen, I.P., et al.: Large-scale brain networks emerge from dynamic processing of musical timbre, key and rhythm. Neuroimage. 59(4), 3677-3689 (2012).
10. Cong, F., Alluri, V., Nandi, A.K., et al.: Linking brain responses to naturalistic music through analysis of ongoing EEG and stimulus features. IEEE Transactions on Multimedia. 15(5), 1060-1069 (2013).
11. Wang, D., Cong, F., Zhao, Q., et al.: 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), pp. 1-6. IEEE, Salerno, Italy (2016).