Analysis of Interactions Among Hidden Components for Tucker Model

Anh Huy Phan and Andrzej Cichocki
Lab for Advanced Brain Signal Processing, BSI, RIKEN, Japan
E-mail: phan, cia@brain.riken.jp

***Abstract*—Tensor representation and tensor decompositions are natural approaches to deal with large amounts of data with multiple aspects and high dimensionality in modern applications, such as environmental analysis, chemometrices, pharmaceutical analysis, spectral analysis, neuroscience. The two most popular decomposition/factorization models for N-th order tensors are the Tucker model and the more restricted PARAFAC model. The Tucker decomposition allows for the extraction of different numbers of factors in each of the modes, and permits interactions within each modality while PARAFAC does not. This advantage, however, is also one of the weakness of this decomposition. The difficult problem is to identify the dominant relationships between components, and to establish unique representation. In this paper, we will introduce a new measure index which is called the Joint Rate (JR) index, in order to evaluate interactions among various components in the general Tucker decomposition. The Hinton diagram is also extended to 3-D visualization. The use of the JR index will be illustrated with the analysis of EEG data for classification and BCI applications.**

I. Tensor Decompositions and Interactive Relation Among their Hidden Components

Standard matrix factorizations, such as PCA/SVD, ICA, NMF, and their variants, are invaluable tools for feature selection, dimensionality reduction, noise reduction, and data mining [1]. However, they have only two modes or 2-way representations, and their use is therefore limited. In many applications such as studies in neuroscience, the data structures often contain higher-order ways (modes) such as trials, task conditions, subjects, and groups together with the intrinsic dimensions of space, time, and frequency. If the data for every subject were analyzed separately by extracting a matrix or slice from a data block, we would lose the covariance information among subjects. To discover hidden components within the data and retain the integrative information, the analysis tools should reflect the multi-dimensional structure of the data [2], [3]. This way all dimensions or modes are retained by virtue of multi-linear models which often produce unique and physically meaningful components.

The two most popular decompositions/factorizations for N-th order tensors are the Tucker model and the more restricted PARAFAC model. The Tucker decomposition is described as a "decomposition of a given N-th order tensor $\underline{\mathbf{Y}} \in \mathbb{R}^{I_1 \times I_2 \cdots \times I_N}$ into an unknown core tensor $\underline{\mathbf{G}} \in \mathbb{R}^{J_1 \times J_2 \cdots \times J_N}$ multiplied by a set of N unknown component matrices, $\mathbf{A}^{(n)} = [\boldsymbol{a}_1^{(n)}, \boldsymbol{a}_2^{(n)}, \ldots, \boldsymbol{a}_{J_n}^{(n)}] \in \mathbb{R}^{I_n \times J_n}$ $(n = 1, 2, \ldots, N)$,

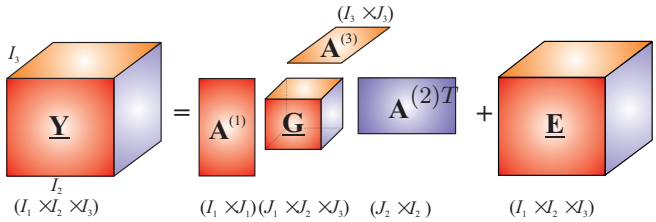


Fig. 1. Illustration and notations used for a higher-order Tucker decomposition; the objective here is to find optimal component (common factor) matrices $\mathbf{A}^{(n)} \in \mathbb{R}^{I_n \times J_n}$ and a core tensor $\underline{\mathbf{G}} \in \mathbb{R}^{J_1 \times J_2 \times \cdots \times J_n}$. We usually impose additional constraints on the component matrices and/or the core tensor such as nonnegativity and sparsity.

representing common factors or loadings" [4], [5], [6]

$$\begin{aligned}\underline{\mathbf{Y}} &= \sum_{j_1=1}^{J_1} \sum_{j_2=1}^{J_2} \cdots \sum_{j_N=1}^{J_N} g_{j_1 j_2 \cdots j_N}\ \boldsymbol{a}_{j_1}^{(1)} \circ \boldsymbol{a}_{j_2}^{(2)} \circ \cdots \circ \boldsymbol{a}_{j_N}^{(N)} + \underline{\mathbf{E}} \\ &= \underline{\mathbf{G}} \times_1 \mathbf{A}^{(1)} \times_2 \mathbf{A}^{(2)} \cdots \times_N \mathbf{A}^{(N)} + \underline{\mathbf{E}} \\ &= \underline{\mathbf{G}} \times \{\mathbf{A}\} + \underline{\mathbf{E}} = \widehat{\underline{\mathbf{Y}}} + \underline{\mathbf{E}}, \end{aligned} \tag{1}$$

where tensor $\widehat{\underline{\mathbf{Y}}}$ is an approximation of tensor $\underline{\mathbf{Y}}$, and tensor $\underline{\mathbf{E}} = \underline{\mathbf{Y}} - \widehat{\underline{\mathbf{Y}}}$ denotes the residual or error tensor (see Fig. 1 as an example of a 3-way Tucker decomposition). There are several existing Tucker decomposition algorithm, such as the Higher-Order Singular Value Decomposition (HOSVD), Higher Order Orthogonal Iteration (HOOI) algorithms which consider orthogonal factors [5], [7], [3], [8]. For analyzing data with nonnegative constraints, the nonnegative Tucker decomposition (NTD) is an extension with multiplicative algorithms [9], [10], [11], [12], The recently proposed hierarchical ALS algorithm for NTD [13] that sequentially estimates components in each factor gives a high performance and is suitable for large scale data.

PARAFAC is a special case of the Tucker decomposition in which the core tensor is a cubical superdiagonal or superidentity tensor with nonzero elements only on the superdiagonal ($J_1 = J_2 = \ldots = J_N$). In PARAFAC, a component in a factor has only one direct relation to components at the same order in the other factors.

For real-world data, there often exist multiple interactive relations among components in different factors. Analysis of a 4-way spectral (Morlet wavelets) EEG tensor (frequency $\times$ time $\times$ channel $\times$ trial) recorded during right and left hand motor imagery is an example. Both left-hand and right-hand imageries are characterized by similar spectral features in the

mean of frequency and time, but spread on different spatial regions over the visual cortex: on the right hemisphere for the left-hand imagery, and on the left hemisphere for the right-hand imagery. This means that this spectral tensor can be explained by only one spectral component (frequency-time) for the first two modes, and two spatial components for the third one. However, analyzing this tensor based on the PARAFAC model needs at least two separate spectral components.

This is the major difference between Tucker and PARAFAC models. The Tucker model allows for the extraction of different numbers of factors in each of the modes, and permits interactions within each modality while the PARAFAC model does not. In other words, the Tucker model encompasses all possible linear interactions between the components (vectors) pertaining to the various modalities of the data. These interactions can be explicitly expressed by a core tensor $\underline{\mathbf{G}}$. Due to these reasons, the Tucker decomposition has high ability to compress data with the minimum number of components.

This advantage, however, is also the greatest weakness of this decomposition. In general, each component in the Tucker decomposition may have J_n different interactions with all the components for all the other factors $\mathbf{A}^{(n)}$. The difficult problem is to identify the dominant relationships between components. For Tucker-3 decomposition, an intuitive method is to use the Hinton diagram to visualize slices of the core tensor $\underline{\mathbf{G}}$. The major interactions for the mode fixed in each illustrated slice can be recognized based on the largest coefficients in that slice [14], [15]. However, this technique is limited for the 3-way tensor, and also misses a quantitive measure to evaluate how strong the relation is. In this paper, we will introduce a new measure index which is called the Joint Rate (JR) index, and allows us to evaluate how strong the interactions among various components in the general Tucker decomposition is. The Hinton diagram is also extended to 3-D visualization. The use of the JR index will be illustrated with the analysis of EEG data for classification and BCI applications.

II. The Joint Rate Index

Let $\underline{\mathbf{Y}}_{\boldsymbol{j}}$, $\boldsymbol{j} = [j_1, j_2, \ldots, j_N]$ denote the rank-one tensor built up from N components $\boldsymbol{a}_{j_1}^{(1)}$, $\boldsymbol{a}_{j_2}^{(2)}$, ..., $\boldsymbol{a}_{j_N}^{(N)}$, that is

$$\underline{\mathbf{Y}}_{\boldsymbol{j}} = g_{\boldsymbol{j}} \boldsymbol{a}_{j_1}^{(1)} \circ \boldsymbol{a}_{j_2}^{(1)} \circ \cdots \circ \boldsymbol{a}_{j_N}^{(N)}. \tag{2}$$

For each pair (k_p, k_q), $1 \leqslant p \neq q \leqslant N$, $k_p = 1, \ldots, J_p$, $k_q = 1, \ldots, J_q$, we define the Joint Rate (JR) index which measure how strongly the component $\boldsymbol{a}_{k_p}^{(p)}$ affects the component $\boldsymbol{a}_{k_q}^{(q)}$.

$$JR_{k_q}^{k_p} = \frac{\sum\limits_{jp=k_p, j_q=k_q} \left\|\underline{\mathbf{Y}}_{\boldsymbol{j}}\right\|_F^2}{\sum\limits_{j_q=k_q} \left\|\underline{\mathbf{Y}}_{\boldsymbol{j}}\right\|_F^2}, \tag{3}$$

The $JR_{j_q}^{j_p}$ index between component $\boldsymbol{a}_{k_p}^{(p)}$, and the component $\boldsymbol{a}_{k_q}^{(q)}$ can be interpreted as the ratio of the total energy of all the rank-one tensors, to which both components $\boldsymbol{a}_{k_p}^{(p)}$ and $\boldsymbol{a}_{k_q}^{(q)}$ contribute, and the total energy of all the rank-one tensors in which the component $\boldsymbol{a}_{k_q}^{(q)}$ interacts.

Assume that all the components $\boldsymbol{a}_{j_n}^{(n)}$ are ℓ_2-norm unit length vectors: $\left\|\boldsymbol{a}_{j_n}^{(n)}\right\|_2^2 = \boldsymbol{a}_{j_n}^{(n)T} \boldsymbol{a}_{j_n}^{(n)} = 1$ for $n = 1, 2, \ldots, N$, the ℓ_2-norm of a rank-one tensor $\underline{\mathbf{Y}}_{\boldsymbol{j}}$ in its vectorized form is given by

$$\begin{aligned} \left\|\underline{\mathbf{Y}}_{\boldsymbol{j}}\right\|_F^2 &= \left\|\mathrm{vec}\left(\underline{\mathbf{Y}}_{\boldsymbol{j}}\right)\right\|_2^2 = \left\|g_{\boldsymbol{j}}\, \boldsymbol{a}_{j_1}^{(1)} \otimes \boldsymbol{a}_{j_2}^{(2)} \otimes \cdots \otimes \boldsymbol{a}_{j_N}^{(N)}\right\|_2^2 \\ &= g_{\boldsymbol{j}}^2 \left(\boldsymbol{a}_{j_1}^{(1)} \otimes \cdots \otimes \boldsymbol{a}_{j_N}^{(N)}\right)^T \left(\boldsymbol{a}_{j_1}^{(1)} \otimes \cdots \otimes \boldsymbol{a}_{j_N}^{(N)}\right) \\ &= g_{\boldsymbol{j}}^2 \left(\boldsymbol{a}_{j_1}^{(1)T} \boldsymbol{a}_{j_1}^{(1)}\right) \cdots \left(\boldsymbol{a}_{j_N}^{(N)T} \boldsymbol{a}_{j_N}^{(N)}\right) = g_{\boldsymbol{j}}^2. \end{aligned} \tag{4}$$

Therefore, for Tucker model, if all components are ℓ_2-norm unit length vectors, the energy of the original data will be expressed only by the coefficients of the core tensor. This leads to a simplified expression of the JR index as follows

$$JR_{k_q}^{k_p} = \frac{\sum\limits_{jp=k_p, j_q=k_q} g_{\boldsymbol{j}}^2}{\sum\limits_{j_q=k_q} g_{\boldsymbol{j}}^2} = \frac{\left\|\underline{\mathbf{G}}_{(j_p,j_q)=(k_p,k_q)}\right\|_F^2}{\left\|\underline{\mathbf{G}}_{j_q=k_q}\right\|_F^2}, \tag{5}$$

where $\underline{\mathbf{G}}_{j_q=k_q}$ is an $(N-1)$-th order subtensor of size $J_1 \times \cdots \times J_{q-1} \times J_{q+1} \times \cdots \times J_N$ obtained by fixing the q-th index of the core tensor $\underline{\mathbf{G}}$ to value k_q, and $\underline{\mathbf{G}}_{(j_p,j_q)=(k_p,k_q)}$ is an $(N-2)$-th order subtensor of size $J_1 \times \cdots \times J_{p-1} \times J_{p+1} \cdots \times J_{q-1} \times J_{q+1} \times \cdots \times J_N$ by fixing the p-th index to value k_p, and the q-th index to value k_q.

For example, for a third-order core tensor, the JR index between the first component $\boldsymbol{a}_1^{(1)}$ of the factor $\mathbf{A}^{(1)}$ and the second component $\boldsymbol{a}_2^{(3)}$ of the factor $\mathbf{A}^{(3)}$ is given by

$$JR_{2_3}^{1_1} = \frac{\sum\limits_{j_2=1}^{J_2} g_{1,j_2,2}^2}{\sum\limits_{j_1=1}^{J_1} \sum\limits_{j_2=1}^{J_2} g_{j_1,j_2,2}^2} = \frac{\left\|\underline{\boldsymbol{g}}_{2_1}\right\|_2^2}{\left\|\mathbf{G}_2\right\|_F^2}, \tag{6}$$

where $\mathbf{G}_2$ is the second frontal slice of core tensor $\underline{\mathbf{G}}$, and $\underline{\boldsymbol{g}}_{2_1}$ is the first row vector of this slice. The index $JR_{k_q}^{k_p}$, $(k_p = 1, 2, \ldots, J_p)$ of a J_p between component $\boldsymbol{a}_{k_p}^{(p)}$ and component $\boldsymbol{a}_{k_q}^{(q)}$ gives the percent rate of the interaction for the each pair. The maximum value of this index is 1, meaning that component $\boldsymbol{a}_{k_q}^{(q)}$ is completely expressed by component $\boldsymbol{a}_{k_p}^{(p)}$, and does not depend on other components of the factor $\mathbf{A}^{(p)}$. Otherwise, the minimum value of this index is 0, meaning that $\boldsymbol{a}_{k_q}^{(q)}$ it is not related to $\boldsymbol{a}_{k_p}^{(p)}$.

We note that the index $JR_{k_q}^{k_p}$ has different meaning to the index $JR_{k_p}^{k_q}$. For the two modes (p, q), we can build a $J_p \times J_q$ matrix of all JR indices among their components.

III. Computation of a Reference Tensor

Most Tucker algorithms do not return ℓ_2-norm unit length components. Hence the core tensor $\underline{\mathbf{G}}$ is often not ready to compute JR indices. Due to this reason, we need to evaluate a reference core tensor. This section describes an efficient method for computing the reference core tensor $\underline{\mathbf{G}}$ by normalizing factors $\mathbf{A}^{(n)}$ to unit length vectors. For a general case, a reference core tensor $\underline{\mathbf{G}}$ can be defined as follows

$$\underline{\mathbf{G}}_f = \underline{\mathbf{G}} \times_1 \mathbf{D}_1 \times_2 \mathbf{D}_2 \cdots \times_N \mathbf{D}_N = \underline{\mathbf{G}} \times \{\mathbf{D}\}, \tag{7}$$

where $\underline{\mathbf{G}}_f$ is the core tensor after fixing, $\mathbf{D}_n$ is the diagonal matrix of ℓ_2-norms of the column vectors in factor $\mathbf{A}^{(n)}$, that is,

$$\begin{aligned} \boldsymbol{d}_n &= \left[\|\boldsymbol{a}_1^{(n)}\|_2, \|\boldsymbol{a}_2^{(n)}\|_2, \ldots, \|\boldsymbol{a}_N^{(n)}\|_2\right], & (8)\\ \mathbf{D}_n &= \operatorname{diag}\{\boldsymbol{d}_n\}, \quad \forall n. & (9) \end{aligned}$$

A considerably simpler and faster method for the reference core tensor is given below. Let $\mathbf{C}$ denote a Kronecker product of two diagonal matrices $\operatorname{diag}\{\boldsymbol{d}_p\}$ and $\operatorname{diag}\{\boldsymbol{d}_q\}$

$$\mathbf{C} = \operatorname{diag}\{\boldsymbol{d}_p\} \otimes \operatorname{diag}\{\boldsymbol{d}_q\} \tag{10}$$

From the definition of the Kronecker product, an element $c_{t_{pq}t_{pq}}$ at the position $t_{pq} = (j_p - 1)\, J_q + j_q$ ($j_p = 1, \ldots, J_p$, $j_q = 1, \ldots, J_q$) on the diagonal of matrix $\mathbf{C}$ is calculated as

$$c_{t_{pq}t_{pq}} = d_{p_{j_p}}\, d_{q_{j_q}}. \tag{11}$$

Thus, the diagonal of matrix $\mathbf{C}$ is exactly the Kronecker product of two vectors $\boldsymbol{d}_p$ and $\boldsymbol{d}_q$

$$\operatorname{diag}\{\mathbf{C}\} = \boldsymbol{d}_p \otimes \boldsymbol{d}_q \tag{12}$$

or

$$\operatorname{diag}\{\boldsymbol{d}_p\} \otimes \operatorname{diag}\{\boldsymbol{d}_q\} = \operatorname{diag}\{\boldsymbol{d}_p \otimes \boldsymbol{d}_q\}. \tag{13}$$

This result can be readily extended to the Kronecker product of N vectors $\boldsymbol{d}_1$, $\boldsymbol{d}_2$, …, $\boldsymbol{d}_N$

$$\operatorname{diag}\{\boldsymbol{d}_1\} \otimes \cdots \otimes \operatorname{diag}\{\boldsymbol{d}_N\} = \operatorname{diag}\{\boldsymbol{d}_1 \otimes \cdots \otimes \boldsymbol{d}_N\}. \tag{14}$$

Equation (7) can be represented in vectorized form as

$$\begin{aligned} \operatorname{vec}\left(\underline{\mathbf{G}}_f\right) &= \left(\mathbf{D}_N \otimes \cdots \otimes \mathbf{D}_1\right) \operatorname{vec}\left(\underline{\mathbf{G}}\right)\\ &= \left(\operatorname{diag}\{\boldsymbol{d}_N\} \otimes \cdots \otimes \operatorname{diag}\{\boldsymbol{d}_1\}\right) \operatorname{vec}\left(\underline{\mathbf{G}}\right)\\ &= \operatorname{diag}\{\boldsymbol{d}_N \otimes \cdots \otimes \boldsymbol{d}_1\} \operatorname{vec}\left(\underline{\mathbf{G}}\right)\\ &= \left(\boldsymbol{d}_N \otimes \cdots \otimes \boldsymbol{d}_1\right) \circledast \operatorname{vec}\left(\underline{\mathbf{G}}\right)\\ &= \operatorname{vec}\left(\boldsymbol{d}_1 \circ \boldsymbol{d}_2 \circ \cdots \circ \boldsymbol{d}_N\right) \circledast \operatorname{vec}\left(\underline{\mathbf{G}}\right) \quad (15) \end{aligned}$$

and its matricization yields the final expression

$$\begin{aligned} \underline{\mathbf{G}}_f &= \left(\boldsymbol{d}_1 \circ \boldsymbol{d}_2 \circ \cdots \circ \boldsymbol{d}_N\right) \circledast \underline{\mathbf{G}}\\ &= \underline{\mathbf{D}} \circledast \underline{\mathbf{G}}. \quad (16) \end{aligned}$$

Finally, computation of the reference core tensor $\underline{\mathbf{G}}_f$ can be achieved conveniently via the Hadamard product of this core tensor and the rank-one tensor built up from N ℓ_2-norm vectors.

TABLE I
Components of Factor $\mathbf{A}^{(3)}$ for the Three Classes of Stimuli.

Component	Coefficients			Class
$\boldsymbol{a}_1^{(3)}$	0. 0	0. 1426	**0. 9898**	3 - Auditory+Visual
$\boldsymbol{a}_2^{(3)}$	**0. 9988**	0. 0489	0. 0	1 - Auditory
$\boldsymbol{a}_3^{(3)}$	0. 0	**0. 9999**	0. 0169	2 - Visual

TABLE II
Frontal Slices of the Estimated Core Tensor $\underline{\mathbf{G}}$.

$\mathbf{G}_1$			$\mathbf{G}_2$			$\mathbf{G}_3$		
0.00	316.66	175.46	50.80	91.92	0.00	79.53	0.00	54.39
74.51	275.45	104.64	157.36	0.00	222.92	85.07	**839.18**	0.00
404.20	0.00	0.00	0.00	188.03	**511.55**	205.80	26.08	32.18

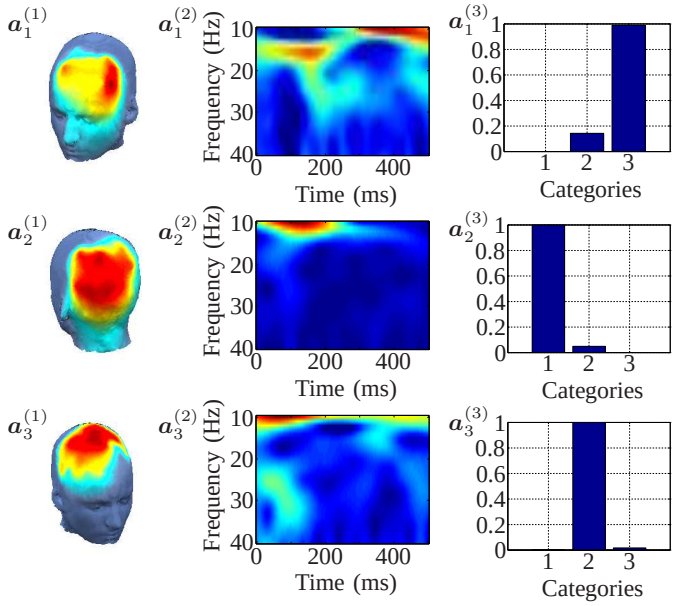
IV. Experiments

A. Example 1

We illustrate the use of the JR index with example of analysis and classification of EEG signals according to the nature of the stimulus for the benchmark `EEG_AV_stimuli` [16], [17], [18], [19]: auditory stimulus with a single tone of 2000 Hz of 30 ms duration; visual stimulus of 30 ms duration in the form of a 5 × 5 checkerboard (600 × 600 pixels) displayed on a LCD screen (32 × 25cm); both the auditory and the visual stimuli simultaneously. In each trial, EEG signals were recorded from 61 channels during 1.5 seconds after stimulus presentation at a sampling rate of 1 kHz. The observed tensor consists the WTav measurements which are the average amplitudes of the oscillations along the 25 trials in the time-frequency domain using the complex Morlet wavelet. The measure tensor corresponds to the time-frequency transformed Event Related Potentials (ERP) [20]. Finally, we vectorized the spectral slices to form a third-order nonnegative tensor $\underline{\mathbf{Y}}$ of the size 61 channels × 3906 frequency-time (31 frequency bins (10-40 Hz) × 126 time frames (0-500ms)) × 3 classes.

The nonnegative Tucker model was chosen to decompose this preprocessed tensor data, in order to find the complex interactions and relationships between components expressing three modes: channels (space), spectra (time frequency representation), and classes (corresponding to three stimuli). The ℓ_2 HALS NTD algorithm [13], [1] was selected to extract underlying components. The number of components was set to three, that is, the size of core tensor was 3 × 3 × 3. The estimated factors are illustrated in Fig. 2. Using such multi-way analysis, the three classes of stimuli were clearly classified, as illustrated by Table I.

The Hinton diagram in Fig. 3 visualizes the estimated core tensor $\mathbf{G}$ whose values are given in Table II. The volume of box is proportional to the intensity of a corresponding coefficient. The dominant coefficients in each slice $\mathbf{G}_j$ indicate the most significant interactions of spatial and spectral components for each j-th category. Using this visualization, $g_{332} = 511.55$ is the major coefficient on the 2-nd frontal slice, the auditory class (corresponding to the component $\boldsymbol{a}_2^{(3)}$) is



(a) Spherically-splined field maps - $\mathbf{A}^{(1)}$ (b) Spectrograms - $\mathbf{A}^{(2)}$ (c) Classes - $\mathbf{A}^{(3)}$

Fig. 2. Visualization of components for example 1. (a) factor $\mathbf{A}^{(1)}$ characterizes spatial components displayed in spherically-spline EEG field maps; (b) spectral components expressed by factor $\mathbf{A}^{(2)}$; (c) expression of factor $\mathbf{A}^{(3)}$ for 3 classes: component $\boldsymbol{a}_1^{(3)}$ - auditory-visual class, component $\boldsymbol{a}_2^{(3)}$ - auditory class, and component $\boldsymbol{a}_3^{(3)}$ - visual class.

therefore mainly characterized by the third spectral component $\boldsymbol{a}_3^{(2)}$ (see in Fig. 2(b)) and spreads on the spatial component $\boldsymbol{a}_3^{(1)}$. Illustration of this relation is given in Fig. 4.

In the next step we specify which spectral components and spatial components mainly affect each class and how they interact with each other by using the JR index. Note that all the component vectors obtained by the ℓ_2 HALS NTD algorithm have been already normalized to unit length vectors, hence the coefficients of the core tensor $\underline{\mathbf{G}}$ express the energy of rank-one tensors which are built up from the basis components $\boldsymbol{a}_j^{(n)}$, $(n = 1, \ldots, N, j = 1, \ldots, J_n)$. In general, we can apply the fast fixing method described in section III.

The Joint Rate (JR) indices computed for all the pairs of components are displayed in Fig. 5. The interactions between spatial and category components are given in Fig. 5(a), whereas we can relations between spectral and category components in Fig. 5(b). From these diagrams, it can be seen that the auditory class (component 2) interacts predominantly with the third spatial component, and the third spectral component, whereas the visual class (component 3) links with the second spatial component, and the second spectral component, and the auditory+visual class (component 1) links with all the spatial and spectral components.

B. Example 2

This example illustrates the analysis of real-world EEG data containing the evoked spectral perturbation (ERSP) measurements of EEG signals recorded from 62 electrodes during right and left hand motor imagery [21]. The observed tensor

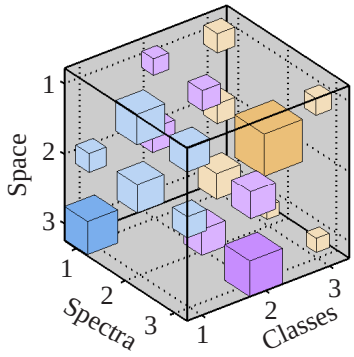
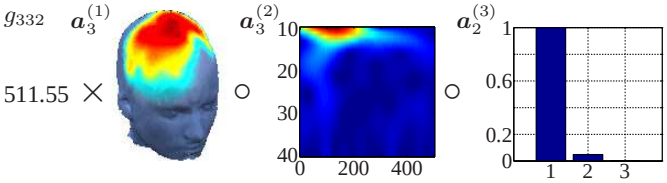
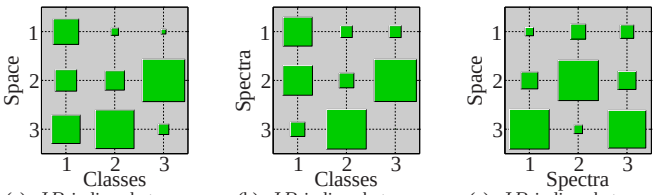


Fig. 3. Illustration of the core tensor via Hinton diagrams. Frontal slices $\mathbf{G}_j = \underline{\mathbf{G}}_{:,:,j}$, $j = 1, 2, 3$ express the interaction of the component $\boldsymbol{a}_j^{(3)}$ with components expressing channel and spectrogram in factors $\mathbf{A}^{(1)}$ and $\mathbf{A}^{(2)}$: auditory class - $\boldsymbol{a}_2^{(3)}$ mainly concentrates on coefficient g_{332}, or spreads on $\boldsymbol{a}_3^{(1)}$ by spectrogram $\boldsymbol{a}_3^{(2)}$; visual class - $\boldsymbol{a}_3^{(3)}$ spreads on $\boldsymbol{a}_2^{(1)}$ by spectrogram $\boldsymbol{a}_2^{(2)}$, auditory + visual class - $\boldsymbol{a}_1^{(3)}$ spreads on $\boldsymbol{a}_3^{(1)}$, $\boldsymbol{a}_2^{(1)}$, and $\boldsymbol{a}_1^{(1)}$ by spectrograms $\boldsymbol{a}_1^{(2)}$, $\boldsymbol{a}_2^{(2)}$ and $\boldsymbol{a}_3^{(2)}$. Darker colors indicate dominant components for each of the three factors.



(a) Channel map (b) Spectrogram (c) Class expression

Fig. 4. Visualization of the three significant components $\boldsymbol{a}_3^{(1)}$, $\boldsymbol{a}_3^{(2)}$, $\boldsymbol{a}_2^{(3)}$ for a dominant rank-one tensor for the 2^{nd} category component - the auditory class.



(a) JR indices between channels and classes (b) JR indices between spectrograms and classes (c) JR indices between spectrograms and classes

Fig. 5. Hinton diagrams of the Joint Rate indices between (a) spatial and category components, (b) spectral and category components and (c) spatial and spectral components.

has size of 62 channels $\times$ 25 frequency bins $\times$ 1000 time frames $\times$ 2 classes (Left/Right). Decomposition of this tensor with the core tensor size of 4 $\times$ 3 $\times$ 3 $\times$ 2 gave us result displayed in Fig. 6. To evaluate how strongly the spectral components $\mathbf{A}^{(2)}$ affect on the spatial components $\mathbf{A}^{(1)}$, we calculate and form their JR matrix which is displayed in Fig. 7. Component $\boldsymbol{a}_1^{(1)}$ illustrates the ERD/ERS phenomena that indicates in the spatial distribution a larger amplitude on the left hemisphere and lower amplitude for the right hemisphere. Whereas component $\boldsymbol{a}_4^{(1)}$ shows ERD on the left hemisphere and ERS on the right hemisphere. Both these components are mainly affected by the spectral component $\boldsymbol{a}_2^{(2)}$ (see Fig. 7). Hence a larger amplitude is shown for class-1 (right-hand

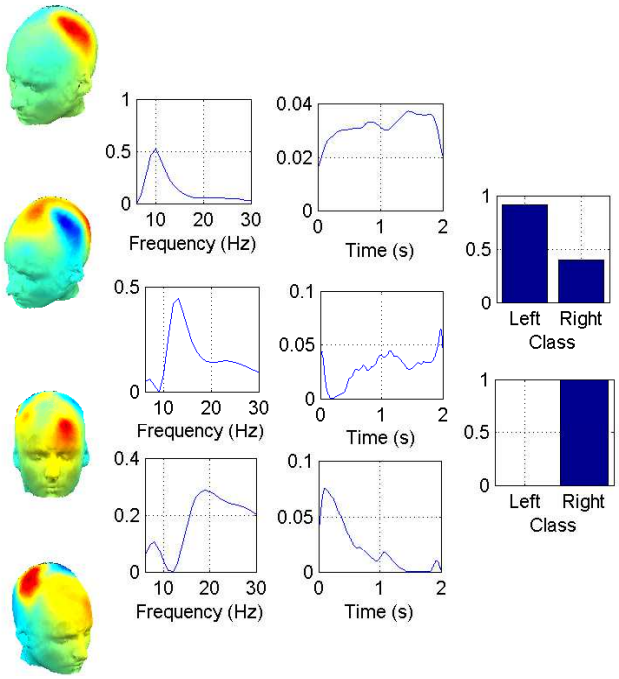


Fig. 6. Factor visualizations for example 2. From left to right, four columns represent the four factors in the analysis.

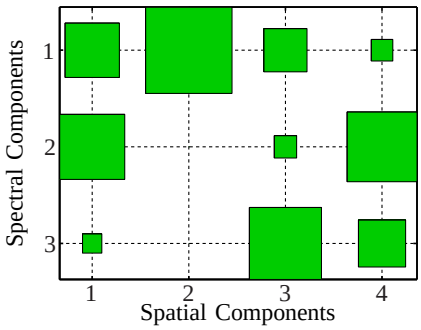


Fig. 7. Hinton diagram of the JR indices between spectral components and spatial components for example 2. Two left and right hemispheres which characterize the left- and right-hand imagery were mainly affected by the second spectral component.

imagery) and lower amplitude on class-2 (left-hand imagery) conditions (column 4 in Fig. 6).

V. Conclusions

We have presented the new index called the Joint Rate index, to measure the interactive relation among components in the Tucker decomposition. With available factors and core tensor obtained by suitable Tucker decomposition algorithms, the JR index can be easily and directly calculated from the core tensor. The fast method for fixing core tensor is also proposed. Moreover, the JR index can also be intuitively illustrated via 2-D and 3-D Hinton diagrams. The experiments with real-world EEG data confirm usefulness and insightful properties of the proposed index.

References

[1] A. Cichocki, R. Zdunek, A.-H. Phan, and S. Amari, *Nonnegative Matrix and Tensor Facorizarions*. Chichester: Wiely, 2009.

[2] F. Miwakeichi, E. Martnez-Montes, P. Valds-Sosa, N. Nishiyama, H. Mizuhara, and Y. Yamaguchi, "Decomposing EEG data into space−time−frequency components using parallel factor analysis," *NeuroImage*, vol. 22, no. 3, pp. 1035–1045, 2004.

[3] T. Kolda and B. Bader, "Tensor decompositions and applications," *SIAM Review*, vol. 51, no. 3, p. (in print), September 2009.

[4] L. Tucker, "Some mathematical notes on three-mode factor analysis," *Psychometrika*, vol. 31, pp. 279–311, 1966.

[5] L. De Lathauwer, B. de Moor, and J. Vandewalle, "A multilinear singular value decomposition," *SIAM Journal of Matrix Analysis and Applications*, vol. 21, pp. 1253–1278, 2001.

[6] T. Kolda, "Multilinear operators for higher-order decompositions," Sandia National Laboratories, Tech. Rep., 2006. [Online]. Available: http://csmr.ca.sandia.gov/ tgkolda/pubs/SAND2006-2081.pdf

[7] L. De Lathauwer, B. D. Moor, and J. Vandewalle, "On the best rank-1 and rank-(R1,R2,. . . ,RN) approximation of higher-order tensors," *SIAM Journal of Matrix Analysis and Applications*, vol. 21, no. 4, pp. 1324–1342, 2000.

[8] M. Friedlander and K. Hatz, "Computing nonnegative tensor factorizations," *Computational Optimization and Applications*, vol. 23, no. 4, pp. 631–647, March 2008.

[9] M. Mørup, L. Hansen, and S. Arnfred, "Algorithms for Sparse Nonnegative Tucker Decompositions," *Neural Computation*, vol. 20, pp. 2112–2131, 2008. [Online]. Available: http://www2.imm.dtu.dk/pubdb/

[10] Y.-D. Kim and S. Choi, "Nonnegative Tucker Decomposition," in *Proc. of Conf. Computer Vision and Pattern Recognition (CVPR-2007)*, Minneapolis, Minnesota, June 2007.

[11] Y.-D. Kim, A. Cichocki, and S. Choi, "Nonnegative Tucker Decomposition with Alpha Divergence," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP2008*, Nevada, USA, 2008.

[12] A. Phan and A. Cichocki, "Fast and efficient algorithms for nonnegative Tucker decomposition," in *Proc. of The Fifth International Symposium on Neural Networks, Springer LNCS-5264*, Beijing, China, 24–28, September 2008, pp. 772–782.

[13] ——, "Local learning rules for nonnegative Tucker decomposition," *(submitted)*, 2009.

[14] M. Mørup, L. Hansen, and S. Arnfred, "ERPWAVELAB a toolbox for multi-channel analysis of time-frequency transformed event related potentials," *Journal of Neuroscience Methods*, vol. 161, no. 2, pp. 361–368, 2006.

[15] M. Mørup, L. K. Hansen, C. S. Herrmann, J. Parnas, and S. M. Arnfred, "Parallel factor analysis as an exploratory tool for wavelet transformed event-related EEG," *NeuroImage*, vol. 29, no. 3, pp. 938–947, 2006.

[16] H. Bakardjian and A. Cichocki, "Extraction and classification of common independent components in single-trial crossmodal cortical responses," in *Proceedings of the 5th Annual Meeting of the International Multisensory Research Forum*, Barcelona, Spain, June 2004, pp. 26–27.

[17] A. Andersen and W. Rayens, "Structure-seeking multilinear methods for the analysis of fMRI data," *NeuroImage*, vol. 22, pp. 728–739, 2004.

[18] E. Martnez-Montes, J. Snchez-Bornot, and P. Valds-Sosa, "Penalized PARAFAC analysis of spontaneous EEG recordings," *Statistica Sinica*, vol. 18, pp. 1449–1464, 2008.

[19] Z. Wang, A. Maier, N. Logothetis, and H. Liang, "Single-trial decoding of bistable perception based on sparse nonnegative tensor decomposition," *Journal of Computational Intelligence and Neuroscience*, vol. 30, pp. 1–10, 2008.

[20] A. Delorme and S. Makeig, "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics," *Journal of Neuroscience Methods*, vol. 134, pp. 9–21, 2004.

[21] A. Cichocki, Y. Washizawa, T. Rutkowski, H. Bakardjian, A.-H. Phan, S. Choi, H. Lee, Q. Zhao, L. Zhang, and Y. Li, "Noninvasive bcis: Multiway signal-processing array decompositions," *Computer*, vol. 41, no. 10, pp. 34–42, 2008.