JYX JYVÄSKYLÄN YLIOPISTO UNIVERSITY OF JYVÄSKYLÄ

This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.

Author(s): Cong, Fengyu; Zhou, Guoxu; Astikainen, Piia; Zhao, Qibin; Wu, Qiang; Nandi, Asoke; Hietanen, Jari K.; Ristaniemi, Tapani; Cichocki, Andrzej

Title: Low-rank approximation based non-negative multi-way array decomposition on event-related potentials

Year: 2014

Version: Published version

Copyright: © 2014 the Authors

Rights: CC BY 4.0

Rights url: https://creativecommons.org/licenses/by/4.0/

Please cite the original version:

Cong, F., Zhou, G., Astikainen, P., Zhao, Q., Wu, Q., Nandi, A., Hietanen, J. K., Ristaniemi, T., & Cichocki, A. (2014). Low-rank approximation based non-negative multi-way array decomposition on event-related potentials. International Journal of Neural Systems, 24(8), Article 1440005. https://doi.org/10.1142/S012906571440005X

International Journal of Neural Systems, Vol. 24, No. 8 (2014) 1440005 (19 pages)
© The Authors
DOI: 10.1142/S012906571440005X

Int. J. Neur. Syst. 2014.24. Downloaded from www.worldscientific.com
by 81.197.1.127 on 09/15/21. Re-use and distribution is strictly not permitted, except for Open Access articles.

LOW-RANK APPROXIMATION BASED NON-NEGATIVE MULTI-WAY ARRAY DECOMPOSITION ON EVENT-RELATED POTENTIALS

FENGYU CONG*
Department of Biomedical Engineering, Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology, China, and Department of Mathematical Information Technology University of Jyväskylä, Finland
cong@dlut.edu.cn; FengyuCong@gmail.com

GUOXU ZHOU
Laboratory for Advanced Brain Signal Processing RIKEN Brain Science Institute, Japan
zhouguoxu@brain.riken.jp

PIIA ASTIKAINEN
Department of Psychology, University of Jyväskylä, Finland
piia.astikainen@jyu.fi

QIBIN ZHAO
Laboratory for Advanced Brain Signal Processing RIKEN Brain Science Institute, Japan
qbzhao@brain.riken.jp

QIANG WU
School of Information Science and Engineering, Shandong University Jinan, Shandong, China
wuqiang@sdu.edu.cn

ASOKE K NANDI
Department of Electronic and Computer Engineering Brunel University, Uxbridge, Middlesex, UB8 3PH, UK Department of Mathematical Information Technology University of Jyväskylä, Finland
asoke.nandi@brunel.ac.uk

JARI K. HIETANEN
Human Information Processing Laboratory School of Social Sciences and Humanities University of Tampere, Finland
jari.hietanen@uta.fi

TAPANI RISTANIEMI
Department of Mathematical Information Technology University of Jyväskylä, Finland
tapani.ristaniemi@jyu.fi

*Corresponding author.

ANDRZEJ CICHOCKI
Laboratory for Advanced Brain Signal Processing
RIKEN Brain Science Institute, Japan and
Systems Research Institute in Polish Academy of Science
Warsaw, Poland
cia@brain.riken.jp

Accepted 26 March 2014
Published Online 9 May 2014

Int. J. Neur. Syst. 2014.24. Downloaded from www.worldscientific.com
by 81.197.1.127 on 09/15/21. Re-use and distribution is strictly not permitted, except for Open Access articles.

Non-negative tensor factorization (NTF) has been successfully applied to analyze event-related potentials (ERPs), and shown superiority in terms of capturing multi-domain features. However, the time-frequency representation of ERPs by higher-order tensors are usually large-scale, which prevents the popularity of most tensor factorization algorithms. To overcome this issue, we introduce a non-negative canonical polyadic decomposition (NCPD) based on low-rank approximation (LRA) and hierarchical alternating least square (HALS) techniques. We applied NCPD (LRAHALS and benchmark HALS) and CPD to extract multi-domain features of a visual ERP. The features and components extracted by LRAHALS NCPD and HALS NCPD were very similar, but LRAHALS NCPD was 70 times faster than HALS NCPD. Moreover, the desired multi-domain feature of the ERP by NCPD showed a significant group difference (control versus depressed participants) and a difference in emotion processing (fearful versus happy faces). This was more satisfactory than that by CPD, which revealed only a group difference.

Keywords: Event-related potential; low-rank approximation; multi-domain feature; non-negative canonical polyadic decomposition; non-negative tensor factorization; tensor decomposition.

1. Introduction

Event-related potentials (ERPs) have been extensively used in cognitive neuroscience research.[1] The peak amplitude of an ERP is the common feature used to represent brain activity corresponding to an event, and is measured sequentially across multiple channels and participants. Statistical analyses of these data are often carried out to detect differences at the group or condition level.[1] Group analyses are of particular importance in paradigms where the signal-to-noise-ratio (SNR) is very low. For example, in the passive oddball paradigm, in which participants ignore the predominant stimuli, ERP parameters/features have to be analyzed at the group-level.[2] Tensor decomposition is a signal processing method that has recently been developed and applied to group-level analyses of ERPs.[3,4] This method provides a novel approach for investigating brain activity simultaneously in multiple domains of cognitive neuroscience.[5–9]

In contrast to the conventional measurement of peak amplitudes, tensor decomposition allows concurrent extraction of new features to represent brain activity across multiple subjects. Moreover, these new features can span the same bases in multiple domains across multiple subjects. For example, ERP data can be represented by a third-order ERP tensor (three-way array) to include waveforms from ERP data across multiple channels and multiple trials. New features can be extracted by tensor decomposition simultaneously for multiple trials.[9] The new extracted features show variations across multiple trials, with these multiple trials spanning the same temporal and spatial components.[9] Moreover, following a transformation of the ERP data into the time–frequency domain, a fourth-order ERP tensor (with four modes: time, frequency, space, and subject) can be formulated. This tensor consists of a time–frequency representation (TFR) of ERPs from multiple channels and multiple participants.[7,8] After the tensor is decomposed, the feature components encompass the variation of multiple participants, whilst spanning the same temporal, spectral, and spatial components.[7,8] This means that the new features extracted by tensor decomposition reveal information regarding brain activity in multiple

Int. J. Neur. Syst. 2014.24. Downloaded from www.worldscientific.com
by 81.197.1.127 on 09/15/21. Re-use and distribution is strictly not permitted, except for Open Access articles.

domains simultaneously.[5–9] Such new features have been referred to as multi-domain features.[7,8] Herein after, when a third- or fourth-order tensor is mentioned without any special notation, the two tensors referred to consist of waveforms of ERPs and TFR of waveforms of ERPs, respectively.

It should be noted that: (1) tensor decomposition with the non-negative constraint for a non-negative tensor has been named as NTF (non-negative tensor factorization),[3,4] and that (2) tensor decomposition includes two basic models. These models are the canonical polyadic (CP) model[10] and the Tucker model.[11] The CP model has also been referred to as PARAFAC.[12,13] NTF thus consists of non-negative canonical polyadic decomposition (NCPD) and non-negative Tucker decomposition (NTD).[3,4] It is well known that the CP decomposition (CPD) is more straightforward than the Tucker decomposition, at least from the perspective of mathematical model complexity.[3,4]

The current trend in cognitive neuroscience research using ERPs in EEG is to utilize a high-density sensor array and high sampling frequency for data collection. For group-level analysis of ERPs, the afore-mentioned high-order ERP tensor data can be large-scale (e.g. hundreds of megabytes). This makes benchmark tensor decomposition algorithms (e.g. alternating least squares (ALS) and hierarchical alternating least squares (HALS).[3,4,14]) very slow to decompose the data.

Low-rank approximation (LRA)-based sequential non-negative Tucker decomposition (LRAS NTD)[15] has recently been developed. This algorithm has been shown to be much faster than benchmark algorithms without the loss of accuracy in decomposition.[15] We have previously shown that the extracted components from the fourth-order ERP tensor by LRAS NTD and the benchmark algorithm referred to as HALS NTD[16] were highly similar.[17] However, the former algorithm was much faster than the latter.[17] This finding motivated us to study the fast NTF algorithm for the CP model, since only the Tucker model was investigated previously.[15,17] Here, we developed a fast tensor decomposition algorithm referred to as a low-rank approximation (LRA)-based hierarchical alternating least squares non-negative canonical polyadic decomposition (LRA-HALS NCPD). This new algorithm is designed to extract multi-domain features of ERPs from the fourth-order tensor mentioned previously.

In order to demonstrate the effectiveness of the LRAHALS NCPD, we studied the fourth-order tensor including TFR of visual ERP data in a passive oddball paradigm. Moreover, we compared the new algorithm with the benchmark algorithm (HALS NCPD[18]). In order to show the superiority of NCPD on the fourth-order tensor, CPD on the third-order tensor consisting of the visual ERP waveforms was also analyzed. The third-order tensor involves the three modes of time, space (i.e. channel) and subject. Finally, key issues including determination of the appropriate number of extracted components, selection of the desired multi-domain feature of the ERP, and robustness of the desired multi-domain feature for NCPD to extract multi-domain features of ERPs are discussed.

2. Method

2.1. *ERP data description*

50 adults participated in an experiment conforming to a passive oddball paradigm. 21 healthy adults (17 females and 4 males, age range 30–58 years, mean 46.8 years) were included in a control group (CONT). The remaining 29 adults (24 females and 5 males, age range 29–61 years, mean 49.1 years) were included in a group with depressive symptoms (DEPR). Experimental conditions were similar to Ref. 19. Pictures of faces with different expressions were presented for 200 ms, and subtended a visual angle of $4 \times 5°$, Neutral facial expressions (probability of presentation = 0.8) were considered standard stimuli. Happy and fearful expressions were rarely presented deviant stimuli (probability = 0.1 for each) (henceforth referred to as Fear and Happy for fearful faces and happy faces, respectively). The stimulus-onset asynchrony (SOA) was 500 ms. During the experiment, at least two standards were presented between randomly presented deviants. Altogether, 1600 face stimuli were presented (1280 neutral, 160 fear, 160 happy). For the recordings, participants were seated in a chair and saw the screen presenting faces. They were instructed to pay no attention to the visual stimuli but to focus on listening to a radio play presented via loud

Int. J. Neur. Syst. 2014.24. Downloaded from www.worldscientific.com
by 81.197.1.127 on 09/15/21. Re-use and distribution is strictly not permitted, except for Open Access articles.

speakers. Previous research using this type of oddball paradigm has shown enhanced face sensitive N170 responses elicited for emotional faces.[19,20]

EEG was recorded using 14 electrodes placed at Fz, F3, F4, Cz, C3, C4, Pz, P3, P4, P7, P8, Oz, O1, and O2 according to the international 10–20 system. An average reference was used. The sampling rate was 1000 Hz. EEG data were digitally filtered from 0.1 to 100 Hz in real time. The continuous EEG data were segmented into single trials including 200 ms pre-stimulus period and 500 ms after the stimulus onset. The baseline was corrected based on the average amplitude of the 200 ms pre-stimulus period. Trials with signal amplitudes beyond the range between -100 and $100\,\mu$V in any recording channel were rejected. The number of trials kept for averaging was about 100 for each deviant type. The recordings of the artifact-free single trials were then averaged to obtain the ERP data, in line with conventional ERP data processing methods. In order to reduce noise further, the ERP data were filtered by a fast Fourier Transform (FFT) filter (number of points for FFT was 10,000) with a pass band of 1–30 Hz.

In the present study, the data from responses to the fearful and happy deviants were used in the analysis. The peak measurements including amplitudes and latencies of N170 were then obtained for P7 and P8. These were used because the two scalp locations are most representative for N170.[21] The analysis employed was a three-factor [group (CONT versus DEPR) × emotion (Fear versus Happy) × hemisphere (Left versus Right)] repeated measurements ANOVA. Here, we expected to find a difference in the N170 between the two groups of participants, as well as a difference in N170 between the two emotions.

2.2. *Tensor representation of ERPs*

In order to extract multi-domain features of ERPs by tensor decomposition for a group-level analysis, different high-order ERP tensors can be formulated. Irrespective of the way the tensor is organized, it can be regarded as a mixture that includes different kinds of brain activity, artifacts, interference, and noise. Using tensor decomposition, it was expected that the desired brain activity would be extracted out from the mixture.

2.2.1. *Third-order ERP tensor of waveforms*

After ERP data are preprocessed, a third-order tensor is naturally produced. It includes the ERP waveforms of multiple channels, conditions, and participants. After the condition mode and the participant mode are combined together, the third-order tensor includes the three modes of time, space, and subject. When single-trial ERP data are analyzed, the subject denotes a single-trial.[9]

2.2.2. *Fourth-order ERP tensor of TFR*

A fourth-order tensor can be generated after the ERP waveform is transformed into the time–frequency domain.[7,8] This tensor consists of the four modes of time, frequency, space, and subject. The high-order tensor based on the TFR of ERPs is non-negative. Therefore, tensor decomposition with a non-negative constraint was applied for the group-level analysis of ERPs.[5–8]

This study also used the Morlet wavelet to transform the average over EEG single trials to obtain the TFR of an ERP as previously described.[7,8] For the Morlet, the half wavelet length was set to be 6 for the optimal resolutions of frequency and time.[22] The frequency range was set from 1 to 15 Hz. This is the frequency band of an ERP. 71 frequency bins were uniformly distributed within this frequency range. Subsequently, the fourth-order tensor was formulated. It included 71 bins in the frequency mode, 700 samples in the time mode, 14 channels in the space mode, and 100 subjects (2 conditions × 50 participants) in the subject mode. It should be noted that the size of tensor is very large. Under the 'double' precision in MATLAB (v.R2010b), the size of tensor is approximately 500 megabytes.

2.3. *NCPD of the fourth-order ERP tensor*

2.3.1. *LRA-based non-negative matrix factorization*

Non-negative matrix factorization (NMF) is the basis for NTF.[3] Hence, low-rank approximation-based NMF (LRANMF)[15] is first introduced as follows. For a given large-scale non-negative matrix $\mathbf{Y} \in \mathfrak{R}_+^{M\times N}$, NMF attempts to find the basis matrix $\mathbf{A} \in \mathfrak{R}_+^{M\times J}$ and encoding matrix $\mathbf{B} \in \mathfrak{R}_+^{N\times J}$ by minimizing the distance $D(\mathbf{A},\mathbf{B}) = \|\mathbf{Y} - \mathbf{A}\mathbf{B}^T\|_F^2$.

Int. J. Neur. Syst. 2014.24. Downloaded from www.worldscientific.com
by 81.197.1.127 on 09/15/21. Re-use and distribution is strictly not permitted, except for Open Access articles.

LRANMF conforms to another objective function reading as

$$D(\tilde{\mathbf{A}}, \tilde{\mathbf{B}}; \mathbf{A}, \mathbf{B}) = \|\mathbf{Y} - \tilde{\mathbf{A}}\tilde{\mathbf{B}}^T\|_F^2 + \|\tilde{\mathbf{A}}\tilde{\mathbf{B}}^T - \mathbf{A}\mathbf{B}^T\|_F^2, \quad (1)$$

where $\tilde{\mathbf{A}} \in \mathfrak{R}_+^{M\times P}$, $\tilde{\mathbf{B}} \in \mathfrak{R}_+^{N\times P}$, $M \leq N$, $P = \mu J \ll M$, and $\mu \geq 1$ is a small positive constant and typically set to 1 for standard NMF. Then, two steps can be used to minimize the above model: (1) LRA by minimizing $\|\mathbf{Y} - \tilde{\mathbf{A}}\tilde{\mathbf{B}}^T\|_F^2$, and (2) NMF for $\|\tilde{\mathbf{A}}\tilde{\mathbf{B}}^T - \mathbf{A}\mathbf{B}^T\|_F^2$ with fixed $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{B}}$. This procedure is known as LRANMF.[15] After introducing LRA techniques, LRANMF can enjoy two major advantages. First, the original large matrix $\mathbf{Y}$ is replaced by two much smaller matrices $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{B}}$ iteratively. This therefore provides significantly reduced computational complexity. Second, the LRA procedure is helpful for filtering out all kinds of noise. For details of the LRANMF algorithm, please see Ref. 15.

2.3.2. *LRA-based HALS NCPD*

Given an Nth-order tensor $\underline{\mathbf{Y}} \in \mathfrak{R}_+^{I_1\times I_2\times\cdots\times I_N}$, the NCPD model[3,4] can be illustrated as follows,

$$\underline{\mathbf{Y}} \approx \hat{\underline{\mathbf{Y}}} = \sum_{j=1}^{J} \mathbf{u}_j^{(2)} \circ \mathbf{u}_j^{(3)} \circ \cdots \circ \mathbf{u}_j^{(N)} = \underline{\mathbf{I}} \times_1 \mathbf{U}^{(2)} \times_2 \mathbf{U}^{(3)} \cdots \times_N \mathbf{U}^{(N)}, \quad (2)$$

where $\mathbf{U}^{(n)} = [\mathbf{u}_1^{(n)}, \mathbf{u}_2^{(n)}, \ldots, \mathbf{u}_J^{(n)}] \in \mathfrak{R}_+^{I_n\times J}$ denote component matrices (also called factors or loadings), $n = 1, 2, \ldots, N$, $\hat{\underline{\mathbf{Y}}}$ approximates the tensor $\underline{\mathbf{Y}}$, and $\underline{\mathbf{I}} \in \mathfrak{R}_+^{I_1\times I_2\times\cdots\times I_N}$ is a diagonal tensor whose diagonal entries are 1. For the definition of outer product and mode-n tensor matrix product, please refer to Appendix A.

Practical NCPD algorithms are usually explained in terms of minimization of some distance between the tensor data and the used model. The distance is subject to the non-negativity constraints[3] as the following

$$D(\underline{\mathbf{Y}} \,|\, \{\mathbf{U}\}) = \frac{1}{2}\|\underline{\mathbf{Y}} - \underline{\mathbf{I}} \times_1 \mathbf{U}^{(2)} \times_2 \mathbf{U}^{(3)} \cdots \times_N \mathbf{U}^{(N)}\|_F^2. \quad (3)$$

ALS algorithm[4] is the conventional method for NCPD.[3] In practice, when the data for decomposition are large, ALS becomes impractical due to the immense computational load. The HALS algorithm is another benchmark method for NCPD and has been shown to be faster than ALS.[3,18] Here, based on HALS NCPD[16] and LRAS NTD,[15] we developed the LRAHALS NCPD algorithm to pursue an even faster NCPD. In LRAHALS NCPD, the nonnegative factors are extracted through two steps[15]: (1) running unconstrained CPD on the tensor $\underline{\mathbf{Y}}$ to achieve its LRA such that $\mathbf{Y} \approx [\![\tilde{U}_1, \tilde{U}_2, \ldots, \tilde{U}_n]\!]$ (in this step we obtain compressed and noise reduced data), and (2) updating each column $\mathbf{u}_j^{(n)}$ of each non-negative $\mathbf{U}^{(n)}$ by solving the following optimization model

$$\min \frac{1}{2}\left\| \begin{array}{l} \sum_{j=1}^{J} \tilde{\mathbf{u}}_j^{(1)} \circ \tilde{\mathbf{u}}_j^{(2)} \circ \cdots \circ \tilde{\mathbf{u}}_j^{(N)} \\ -\sum_{j=1}^{J} \mathbf{u}_j^{(2)} \circ \mathbf{u}_j^{(3)} \circ \cdots \circ \mathbf{u}_j^{(N)} \end{array} \right\| \quad (4)$$
$$\text{s.t. } \mathbf{u}_j^{(n)} \geq 0, \quad j = 1, 2, \ldots, J, \;\; n = 1, 2, \ldots, N.$$

This thereby leads to the following update rule for each n and j:

$$\mathbf{u}_j^{(n)} \leftarrow [\gamma \mathbf{u}_j^{(n)} + \tilde{\mathbf{U}}_{(n)}^T \tilde{\mathbf{b}}_j - \mathbf{U}^{(n)} \mathbf{b}_j]_+, \quad (5)$$

where $[x]_+ = \max(0, x)$, $\gamma = \prod_{k\neq n} u_j^{(n)T} u_j^{(n)}$, $\mathbf{b}_j$ and $\tilde{\mathbf{b}}_j$ are the jth column of matrices $\mathbf{B}$ and $\tilde{\mathbf{B}}$, respectively. After the update, the components are normalized by using $\mathbf{u}_j^{(n)} \leftarrow \mathbf{u}_j^{(n)}/\|\mathbf{u}_j^{(n)}\|_2$ for all $n \neq N$ (when $n = N$, $\gamma = \|\mathbf{u}_j^{(N)}\|^2$). Due to the LRA in step 1, these two matrices can be very efficiently computed as $\mathbf{B} = \circledast_{k\neq n}(\mathbf{U}^{(k)T}\mathbf{U}^{(k)})$ and $\tilde{\mathbf{B}} = \circledast_{k\neq n}(\tilde{\mathbf{U}}^{(k)T}\mathbf{U}^{(k)})$ (where $\circledast$ is the element-wise product of matrices). This process only involves multiplications of very small matrices. Consequently, LRAHALS NCPD turns out to be surprisingly efficient, especially for large-scale problems. For details of LRAHALS NCPD, please refer to the laboratory of tensor decomposition and analysis, TDALAB.[23]

In order to minimize the distance in Eq. (3), initializing component matrices is often the first step in the iteration. Usually, there are several initialization approaches, including singular value decomposition (SVD), randomization, and fiber.[3] The latter two are determined methods. As this was the first use of LRAHALS NCPD for analysis of ERPs, we choose random initialization to establish a benchmark for future comparisons with different initialization methods.

Int. J. Neur. Syst. 2014.24. Downloaded from www.worldscientific.com
by 81.197.1.127 on 09/15/21. Re-use and distribution is strictly not permitted, except for Open Access articles.

2.3.3. *Multi-domain feature of an ERP: extraction, selection and analysis*

Through NCPD, a fourth-order tensor of time–frequency transformed ERP data can be decomposed into spectral, temporal, spatial, and subject factors[7,8]:

$$\underline{\mathbf{Y}} \approx \sum_{j=1}^{J} \mathbf{u}_j^{(f)} \circ \mathbf{u}_j^{(t)} \circ \mathbf{u}_j^{(c)} \circ \mathbf{f}_j = \underline{\mathbf{I}} \times_1 \mathbf{U}^{(f)} \times_2 \mathbf{U}^{(t)} \times_3 \mathbf{U}^{(c)} \times_4 \mathbf{F}. \quad (6)$$

The last component matrix $\mathbf{F}$ consists of the extracted J multi-domain features of brain responses. Each column of $\mathbf{F}$ corresponds to one feature. The component matrices $\mathbf{U}^{(f)}$, $\mathbf{U}^{(t)}$, and $\mathbf{U}^{(c)}$ respectively denote the J extracted spectral, temporal, and spatial components. Each column of the three matrices represents one extracted component. In the CP model, $\mathbf{f}_j$, $\mathbf{u}_j^{(f)}$, $\mathbf{u}_j^{(t)}$, and $\mathbf{u}_j^{(c)}$ with the same subscript are associated with each other. Since they share the same subscript j, they can also be referred to as parallel with each other. The $\mathbf{u}_j^{(f)}$, $\mathbf{u}_j^{(t)}$, and $\mathbf{u}_j^{(c)}$ reveal the spectral, temporal and spatial properties of brain activity. Given the fourth-order tensor in Eq. (6), they are common across different subjects in the subject mode. The multi-domain feature $\mathbf{f}_j$ carries the variation of different subjects given $\mathbf{u}_j^{(f)}$, $\mathbf{u}_j^{(t)}$, and $\mathbf{u}_j^{(c)}$.

After J multi-domain features are extracted, it is necessary to choose the desired features for ERPs of interest. In an experiment utilizing a well-known ERP, the latency of the ERP and the structure of its spectrum are usually known. For example, in this study, the ERP component of interest is N170, which peaks around 170 ms after the stimulus onset, at around 7 Hz.[21] Using this information, the multi-domain features with the similar temporal and spectral components can be first selected, while the features that do not concurrently possess the temporal and the spectral properties of the ERP of interest can be rejected. For the selected multi-domain features, statistical tests were performed to determine the feature(s) showing the expected effect(s) of the experiment.

Each multi-domain feature $\mathbf{f}_j$ is represented by a vector. When the ERP tensor for decomposition includes data from multiple groups and multiple experimental conditions, this vector can be reshaped to a three-way array with three modes, including the number of participants, group, and experimental condition. In this study, the three modes are number of participants, group (CONT versus DEPR), and emotion (Fear versus Happy). Therefore, a two-factor (group by emotion) statistical analysis was performed to examine any differences in the selected multi-domain feature(s) between the two groups, the two emotions, and their interaction. Both ANOVA and the Kruskal–Wallis test[24] were used here. ANOVA is the preferred method when the data meet the assumptions for a parametric test. However, when the data do not meet these assumptions, in this case due to outliers, the Kruskal–Wallis test can more accurately reveal an effect.[24] For the same data, the smaller p-value derived from the two statistical methods was reported. This reporting method was used in the previous studies for NTF on ERPs.[7,8]

2.3.4. *Number of extracted components for LRAHALS NCPD*

For tensor decomposition, it is necessary to determine the number of extracted components in each factor. For the CP model, only one parameter, J, in Eq. (6) should be selected. DIFFIT[25,26] was used here. DIFFIT measures the change of fits of different models to determine the model with the smallest increment of fit. Fit can be defined as fit $= 1 - \|\underline{\mathbf{Y}} - \hat{\underline{\mathbf{Y}}}\|^2 / \|\underline{\mathbf{Y}}\|^2$ where $\|\cdot\|^2$ is the norm-2. For different LRAHALS NCPD models, the numbers of temporal components here ranged from 2 to 80. Due to the random initialization used for NCPD here, the LRAHALS NCPD was run 30 times under each NPCD model. The mean fit over 30 rounds for each NCPD model was used for DIFFIT.

2.3.5. *Uniqueness of LRAHALS NCPD*

The uniqueness of LRAHALS NCPD on the fourth-order tensor was examined by the variant of Kruskal's theorem[27,28]:

$$\sum_{n=1}^{N} k_{\mathbf{U}^n} \geq 2J + N - 1, \quad (7)$$

where $k_{\mathbf{U}^n}$ is rank of the component matrix, $\mathbf{U}^{(n)}$ in Eq. (2), J is the number of extracted components in each mode, and N (= 4 here) is the number of modes of a given tensor.

Int. J. Neur. Syst. 2014.24. Downloaded from www.worldscientific.com
by 81.197.1.127 on 09/15/21. Re-use and distribution is strictly not permitted, except for Open Access articles.

2.3.6. *Robustness of the desired multi-domain feature extracted by LRAHALS NCPD*

Since multiple NCPD models were being applied, it was necessary to examine whether the desired multi-domain feature of an ERP could be extracted by many models or not. Furthermore, given a selected NCPD model, the results from multiple rounds of the NCPD model should also be investigated.

First, the appropriate NCPD model is suggested by DIFFIT. Then, the desired multi-domain feature of an ERP is selected according to the methods introduced in Sec. 2.3.3. Next, a rank-one fourth-order tensor can be formulated as a template rank-one tensor:

$$\underline{\mathbf{Y}}_{\text{template}} = \mathbf{u}^{(f)}_{\text{template}} \circ \mathbf{u}^{(t)}_{\text{template}} \circ \mathbf{u}^{(c)}_{\text{template}} \circ \mathbf{f}_{\text{template}}. \quad (8)$$

Here, $\mathbf{f}_{\text{template}}$ is the desired multi-domain feature of an ERP, $\mathbf{u}^{(f)}_{\text{template}}$, $\mathbf{u}^{(t)}_{\text{template}}$, $\mathbf{u}^{(c)}_{\text{template}}$ are spectral, temporal and spatial components of the multi-domain feature. It is then necessary to examine empirically two issues: (1) whether the template rank-one tensor can be extracted by other NCPD models or not, and (2) whether it can be stably extracted when one NCPD model is run multiple times with random initialization.

The robustness analysis used here is for the desired multi-domain feature extracted by LRAHALS NCPD. Each extracted rank-one tensor in each NCPD model of each round decomposition is correlated with the template rank-one tensor as

$$\begin{aligned}\rho(j,J,r) = {} & [(\mathbf{u}^{(f)}_{(j,J,r)})^T \mathbf{u}^{(f)}_{\text{template}}] \\ & \cdot [(\mathbf{u}^{(t)}_{(j,J,r)})^T \mathbf{u}^{(t)}_{\text{template}}] \\ & \cdot [(\mathbf{u}^{(c)}_{(j,J,r)})^T \mathbf{u}^{(c)}_{\text{template}}] \\ & \cdot [(\mathbf{f}_{(j,J,r)})^T \mathbf{f}_{\text{template}}],\end{aligned}$$

where $J = 2, 3, \ldots, 80$, $j = 1, 2, \ldots, J$ and $r = 1, 2, \ldots, 30$ in this study. Each component was normalized by its standard deviation and its nonzero mean was subtracted. Then, given one J and one r, J correlation coefficients were obtained. Next, the maximal coefficient among them was chosen as

$$\begin{aligned} q(J,r) &= \rho(k,J,r) \\ &= \max[\rho(1,J,r), \rho(2,J,r), \ldots, \rho(J,J,r)], \quad (9)\end{aligned}$$

where $k \in [1, J]$.

If the robustness of NCPD exists for the desired multi-domain feature, for different J, $q(J,r)$ can be different. If the LRAHALS NCPD is stable in decomposing the tensor, given one J, $q(J,r)$ should not vary much across different r. If the desired multi-domain feature is extracted by NCPD, $q(J,r)$ approaches 1, otherwise, $q(J,r)$ converges to 0.

2.3.7. *Comparison between LRAHALS NCPD and HALS NCPD*

Due to the large-scale ERP data tensor used here, HALS NCPD was very slow for decomposition. The stability of HALS NCPD has been examined previously and it has been found that HALS NCPD is stable in decomposing time–frequency transformed ERP data.[7] Hence, HALS NCPD with random initialization was only run once in this study with the numbers of components ranging from 2 to 80.

The template in Eq. (8) can be used to examine whether the rank-one tensor similar to the template tensor can be extracted by HALS NCPD. For this purpose, Eq. (9) can be used with $r = 1$.

2.4. *CPD of the third-order ERP tensor*

In Ref. 9, CPD was successfully applied on the third-order tensor to extract the multi-domain features of single-trial ERP signals for the classification of different tasks. We also applied CPD on a third-order tensor here. The three modes of the tensor were time, space (i.e. channel), and subject.

In Ref. 9, the orthogonal constraint was applied for the constrained CPD. Indeed, it may not be reasonable to assume that two topographies (i.e. spatial components) of any two kinds of brain activity in EEG data are not correlated with each other. It is also not appropriate to suppose that the two feature components of multiple single-trials are orthogonal. In the time mode, it seems to be plausible to assume the two temporal components are not correlated. The independent assumption for electrical sources of brain activity is often made when independent component analysis (ICA) is applied to EEG data.[29,30] Therefore, the orthogonal constraint was also applied on the time mode for the constrained CPD here. For comparison, the unconstrained CPD was used. As in Ref. 9, the PARAFAC algorithm in the N-way toolbox[31] was used.

Int. J. Neur. Syst. 2014.24. Downloaded from www.worldscientific.com
by 81.197.1.127 on 09/15/21. Re-use and distribution is strictly not permitted, except for Open Access articles.

It was necessary to make a fair comparison between CPD of the third-order tensor (time by space by subject) and NCPD of the fourth-order tensor (time by frequency by space by subject). For this, the LRAHALS CPD was applied. When the non-negative constraint is not used, LRAHALS NCPD becomes LRAHALS CPD.

The numbers of components ranged from 2 to 80 for the LRAHALS CPD, and for each selected number, the LRAHALS CPD was run 30 times. For demonstration, the number of components for PARAFAC algorithm in the N-way toolbox was set to be 10, as used previously in Ref. 9.

3. Results

3.1. *Conventional ERP analysis*

3.1.1. *Peak amplitude analysis*

Figure 1(a) shows the grand average of ERPs at P7 and P8. They are located at left and right temporo-parietal sites, respectively. Figure 1(b) shows the topography of N170 in the grand average. As expected, the N170 was elicited by the experimental paradigm. Its peak amplitudes at P7 or P8 were the largest (absolute value) among the 14 scalp locations. As mentioned earlier, P7 and P8 are the two most representative scalp locations for N170.[21] Figure 1(c)

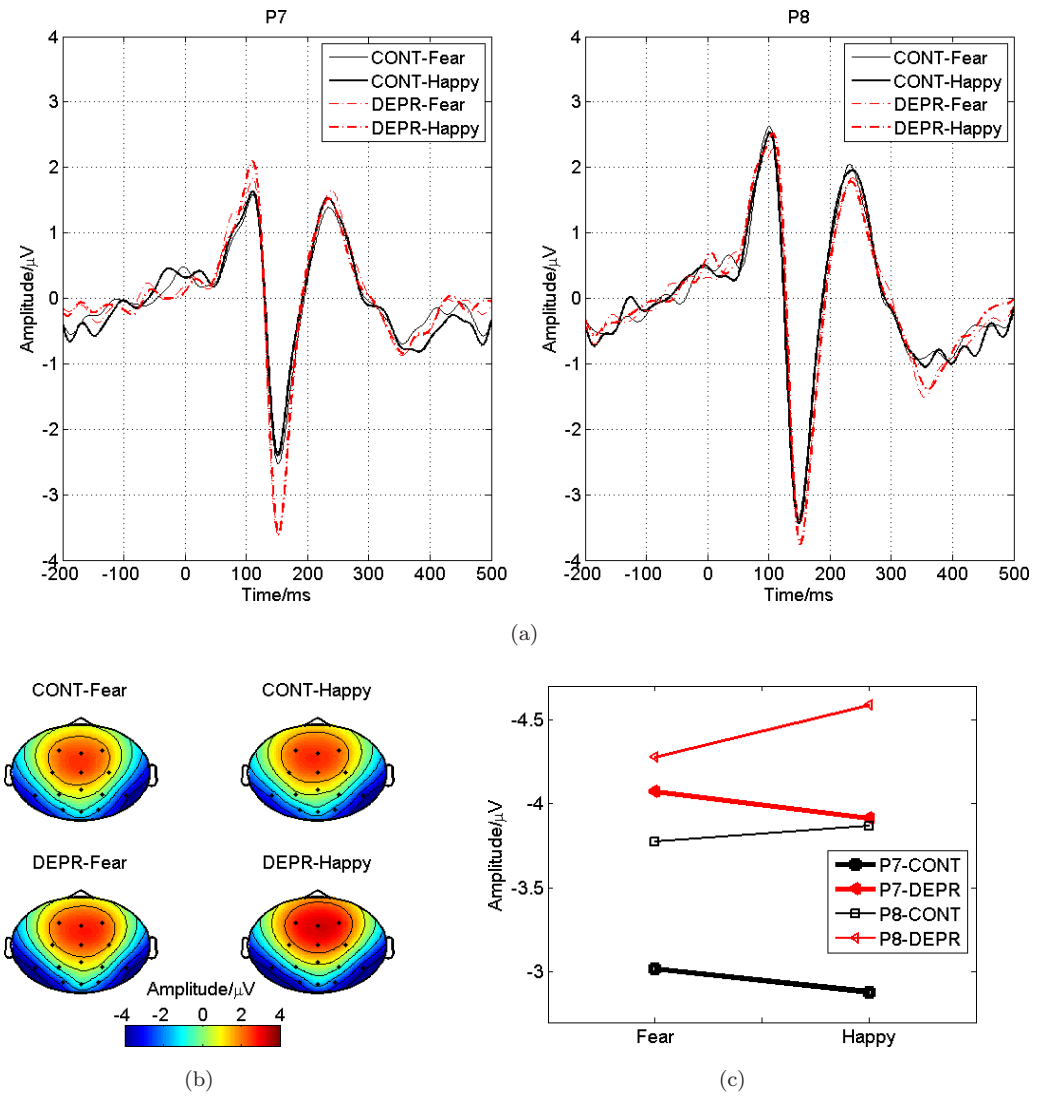


Fig. 1. (a) Grand average of ERPs at P7 and P8 electrodes. (b) Topography of N170 in the grand average. (c) Grand averaged peak amplitudes across all participants of each group under two deviants and two electrodes.

Int. J. Neur. Syst. 2014.24. Downloaded from www.worldscientific.com
by 81.197.1.127 on 09/15/21. Re-use and distribution is strictly not permitted, except for Open Access articles.

shows the grand-averaged peak amplitudes across all participants in both groups for the two deviants at the two locations.

A three-factor ANOVA (mixed design) revealed a significant main effect of hemisphere ($F(1, 48) = 7.7, p = 0.008$). In addition, the interaction between hemisphere and emotion was significant ($F(1, 48) = 8.4, p = 0.006$) as shown in Fig. 1(c). The main effect of 'group' or any interaction involving 'group' was not significant. However, there was a tendency for the N170 amplitudes of the DEPR group to be stronger than those of CONT (shown in Fig. 1(c)).

Since there was a hemisphere by emotion interaction but no effects involving 'group', paired t-tests were applied to the data across all subjects. This was to investigate further differences between the left and right hemispheres in response to each emotion, and also to examine the difference in responses between happy and fearful expressions within each hemisphere. For Happy, N170 amplitude at the right hemisphere was significantly larger than that at the left ($t(49) = 2.9$, $p = 0.006$). For Fear, this difference approached significance, ($t(49) = 1.9$, $p = 0.06$). In the right hemisphere (P8), N170 for Happy was significantly larger than that of 'Fear' ($t(49) = 2.6$, $p = 0.013$). In the left hemisphere (P7), N170 did not significantly differentiate between Fear and Happy.

3.1.2. *Peak latency analysis*

A three-factor ANOVA (mixed design) revealed no significant main effects or interactions even though some differences in latencies were observed as shown in Fig. 1(a).

Indeed, this finding was a necessary condition for the application of NCPD on the fourth-order ERP tensor. The same temporal component extracted by NCPD is spanned by different groups of subjects. When different groups of subjects are significantly different in their peak latencies, it is difficult to find the desired temporal component.

3.2. *Multi-domain feature extracted by NCPD*

3.2.1. *Fits of LRAHALS NCPD and appropriate number of components*

Figure 2 shows the mean, maximum, and minimum fits across 30 rounds for each LRAHALS NCPD model. Obviously, their difference is small, indicating the high stability of LRAHALS NCPD based on the fit of raw data.

After DIFFIT was applied on the mean fits of NCPD models ranging from 2 to 80, $J = 50$ was suggested as the appropriate number of components for LRAHALS NCPD.

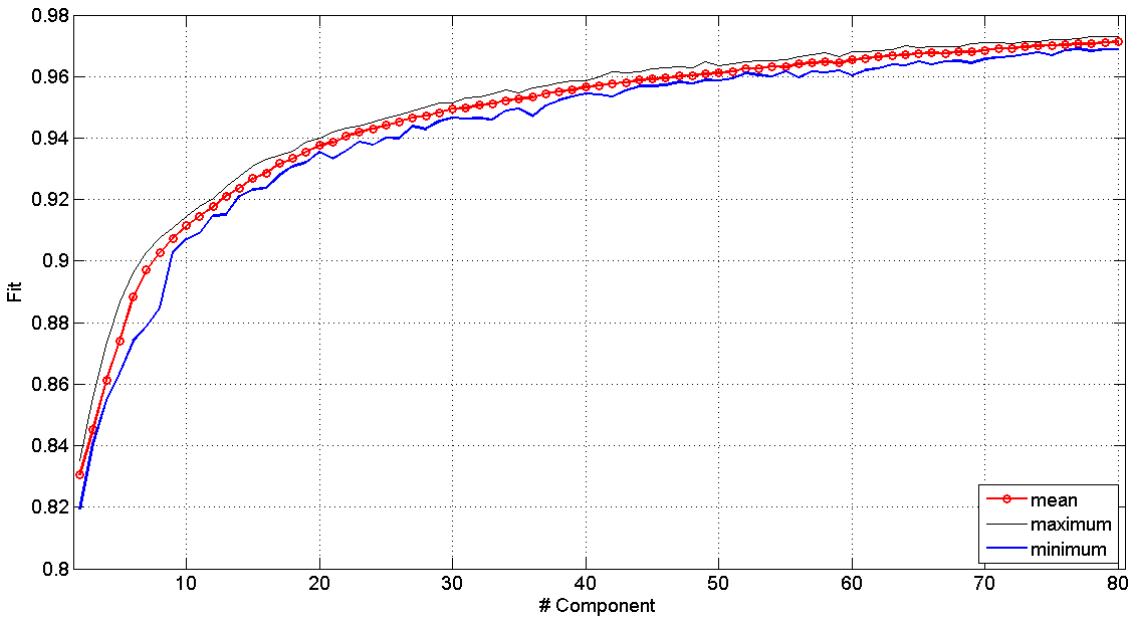


Fig. 2. Mean, maximum and minimum of fits of 30 rounds for each LRAHALS NCPD model.

Int. J. Neur. Syst. 2014.24. Downloaded from www.worldscientific.com
by 81.197.1.127 on 09/15/21. Re-use and distribution is strictly not permitted, except for Open Access articles.

3.2.2. *Multi-domain feature analysis*

For $J = 50$, results of the second round among 30 rounds of LRAHALS NCPD were taken as the example for further analysis. This is because the fit of the second round was closest to the averaged fit over 30 rounds for $J = 50$.

Figure 3 shows the mean of each of two selected multi-domain features, and their temporal, spectral, and spatial components of two rank-one tensors. They were selected since only they simultaneously possessed the properties of N170 in the time and frequency domains. This means that the temporal component peaked at about 150 ms and the spectrum peaked around 7 Hz.

Next, the two multi-domain features were tested using a two-way (group by emotion) ANOVA (mixed design). The multi-domain feature #1 revealed a significant main effect of emotion, $F(1, 48) = 20.6$, $p < 0.001$. The data indicated that the magnitude of the multi-domain feature to Fear was significantly larger than to Happy. No main effect of group or interaction was found.

The multi-domain feature #2 also revealed that a significant main effect of emotion, $F(1, 48) = 15.9$, $p < 0.001$. The same relationship between Fear and Happy was found as that in the feature #1. Moreover, the main effect of group was significant, $F(1, 48) = 11.2$, $p = 0.002$, with the magnitude of multi-domain feature of DEPR being significantly larger than that of CONT. There was no interaction between group and emotion. Henceforth, the multi-domain feature #2 is defined as the desired multi-domain feature due to two reasons: One is that the corresponding temporal and spectral components matched the properties of N170 in the time and frequency domains. The second is that it differentiated the two groups and the two emotions, revealing the experimental design.

For the other 48 features among all 50 features, no main effect of group was observed.

3.2.3. *Interpretation of the desired multi-domain feature*

The magnitude of the multi-domain feature extracted by NCPD from the fourth-order tensor denotes the strength of brain activity. The corresponding temporal, spectral, and spatial components characterize the temporal, spectral, and spatial structures of this brain activity.

Here, the desired multi-domain feature of N170 revealed the group difference and the emotion difference. In other words, the group difference and

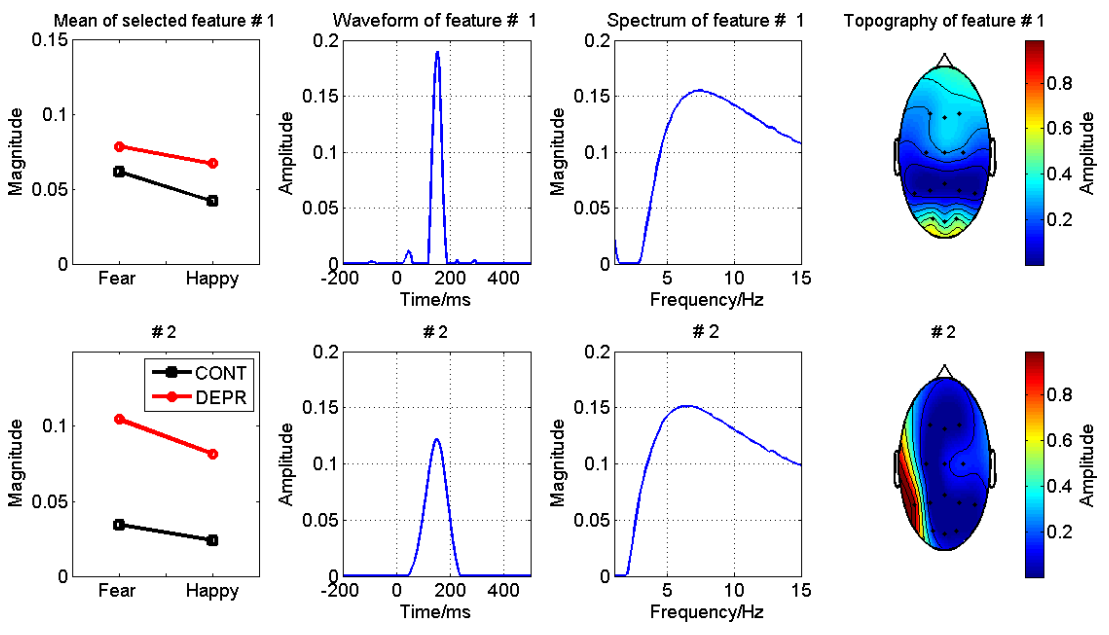


Fig. 3. Mean of selected multi-domain extracted feature and their corresponding temporal, spectral, and spatial components extracted by LRAHASL NCPD ($J = 50$). These were selected based on the properties of N170 in the time and frequency domains.

Int. J. Neur. Syst. 2014.24. Downloaded from www.worldscientific.com
by 81.197.1.127 on 09/15/21. Re-use and distribution is strictly not permitted, except for Open Access articles.

emotion difference appeared when brain activity possessed those temporal, spectral and spatial structures presented on the bottom row of Fig. 3. As mentioned earlier in Sec. 3.2.2, the temporal and spectral components of the desired multi-domain feature did show the properties of N170 in the time and frequency domains. In particular, the corresponding temporal component is very similar to the successfully extracted temporal component of N170 by ICA from the optimally filtered waveforms of ERPs (elicited by the same experimental paradigm).[20,32] In contrast to the topographies of N170 amplitudes shown in Fig. 1(b), the corresponding spatial component of the desired multi-domain feature showed a different pattern of topography. The spatial component had the largest signature at P7, which is the representative location for N170.[21] This means that the group difference and the emotion difference could most probably appear at P7 in this study.

3.2.4. *Uniqueness of LRAHALS NCPD*

In the 50-component model of LRAHALS NCPD for the second round, the ranks of the temporal, spectral, spatial and feature components matrices shown in Eq. (6) were 50, 50, 15, and 50. This means that the left side of Eq. (7) was 165, and the right side was 103. Such results met the variant of Kruskal's theorem according to Eq. (7)[27,28] for the uniqueness of tensor decomposition with the CP model.

3.2.5. *Robustness of the desired multi-domain feature extracted by LRAHALS NCPD*

Figure 4 shows that the desired multi-domain feature was hard to extract when the number of components was smaller than 15. This was because the maximal correlation coefficient [$q(J, r)$ in Eq. (9), $J < 15$] was very small between any of the extracted rank-one tensors in one NCPD model and the template rank-one tensor. This resulted from the fact the mixtures represented by the fourth-order tensor were not well separated. The desired multi-domain feature was extracted when the numbers of components ranged from about 50 to about 60. This was because the maximal correlation coefficient was close to 1 between any of the extracted rank-one tensors in one NCPD model and the template rank-one tensor.

Furthermore, the maximal correlation coefficient between any of the extracted rank-one tensors in one NCPD model and the template rank-one tensor did not vary much across different rounds of LRAHALS NCPD.

These results highlight two issues: (1) LRAHALS NCPD was stable enough to decompose the ERP data, and (2) extraction of the desired multi-domain

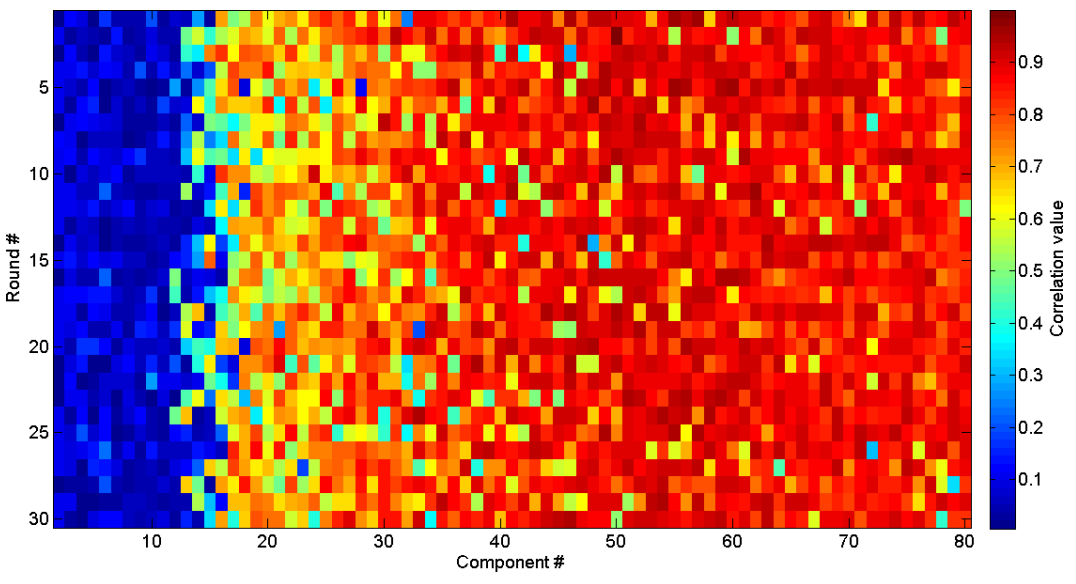


Fig. 4. Robustness analysis [according to Eq. (8)] of the desired multi-domain feature extracted by LRAHALS NCPD.

Int. J. Neur. Syst. 2014.24. Downloaded from www.worldscientific.com
by 81.197.1.127 on 09/15/21. Re-use and distribution is strictly not permitted, except for Open Access articles.

feature from the fourth-order ERP tensor was robust as long as the number of components was approximately chosen.

3.2.6. *Comparison between LRAHALS NCPD and HALS NCPD*

LRAHALS NCPD is based on LRA and the benchmark HALS NCPD. We compared the two NCPD approaches from three aspects as follows:

3.2.6.1. Comparison in extracting the desired multi-domain feature

The template rank-one tensor is defined in Eq. (8). In order to check whether it was extracted by HALS NCPD, each extracted rank-one tensor by HALS NCPD was correlated with the template tensor. Given a selected number of extracted components, the maximum correlation coefficient was chosen. Then, the robustness of the HALS NCPD in extracting the desired multi-domain feature was obtained. For LRAHALS NCPD, the robustness analysis of the first round of 30 (i.e. the first row of Fig. 4) was taken for comparison.

Figure 5(a) shows that there is no significant difference between the two methods (one-way Kruskal–Wallis test: $\mathrm{F}(1, 157) = 1.52$, $p = 0.22$) in extracting the desired multi-domain feature. This can be observed in particular when the number of components is about larger than 40.

3.2.6.2. Fit comparison

Figure 5(b) shows that the loss of fit was very tiny in LRAHALS NCPD compared to HALS NCPD. Specifically, for the difference in fit between the two NCPD approaches, the median difference was 0.0086, the maximal was 0.0223, and the minimal was 0.0046. It should be noted that the decomposed fourth-order tensor was about 500 megabytes in size.

For NCPD with the same number of extracted components, it is understandable that the fit of the LRAHALS is smaller than the benchmark HALS. This is because LRAHALS is based on LRANMF. For LRANMF, the LRA is first applied according to Eq. (1). There is fit loss in this step. Indeed, as long as the fit loss is small enough, the influence on the accuracy of tensor decomposition can be ignored.

3.2.6.3. Computing comparison

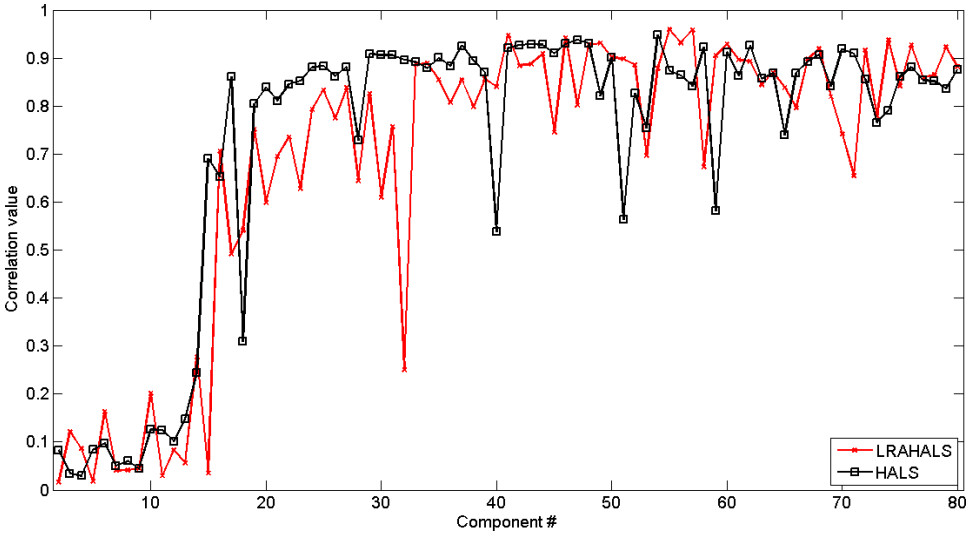
The computing speed for LRAHALS NCPD was more than 70 times faster than HALS NCPD on a local workstation for $J = 50$ (737 s versus 58,034 s). Specific parameters of this workstation are OS — 64-bit Win7 pro, CPU — Intel i7 X980 3.33 GHZ, and Memory — 24 GB. For the LRAHALS NCPD algorithm, the number of iterations was 5000. For the HALS NCPD algorithm[23] the maximal number of iterations was also 5000, and the tolerance to stop iteration was 10-e8 for the increase of fit.[18]

Using LRAHALS NCPD, the sizes of tensor are first reduced by LRA. Then, tensor decomposition is applied for the size-reduced tensor. Faster computing by LRAHALS NCPD would be expected compared to HALS NCPD. As mentioned earlier, LRAHALS NCPD is indeed based on LRANMF. In theory, LRANMF can be about M/J times faster than the standard NMF[15] (M is the size of one mode of matrix, and J is the number of extracted components). As described in Sec. 2.3.1, $P = \mu J \ll M$, μ is no smaller than 1. Here, the size of the time mode is 700, and the number of extracted components is 50. Temporal samples can be highly correlated with each other in ERP data due to the high resolution of time in EEG. Thus, there is much space in the time mode for LRA. Therefore, P can be bigger than J (number of extracted components), but can be much smaller than M (number of temporal samples) for the ERP data.

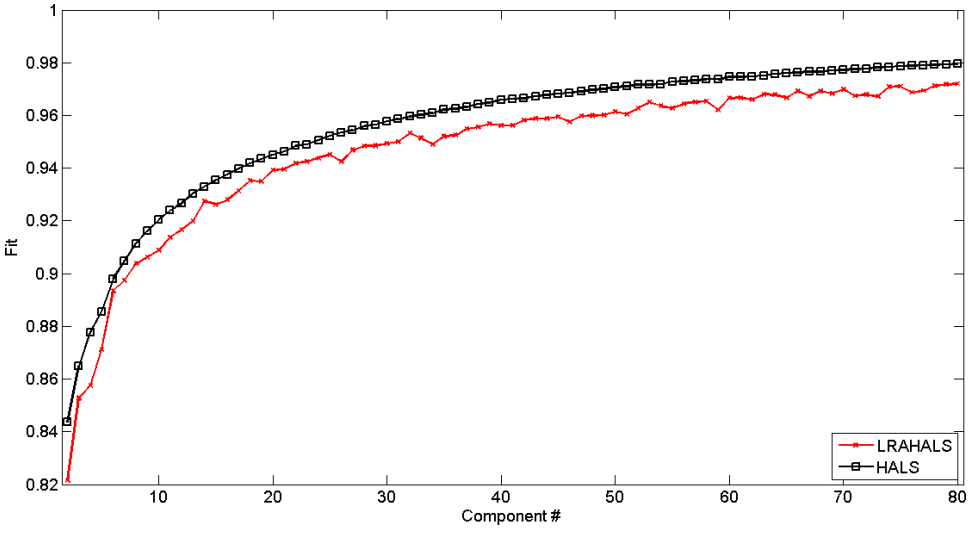
3.3. *Multi-domain feature extracted by CPD*

Given the number of extracted components as 10, the significant difference ($p = 0.001$) between two groups of participants in only one multi-domain feature was found by the unconstrained CPD. Regarding the constrained CPD orthogonal to the time mode, there was also only one multi-domain feature showing a significant group difference ($p = 0.019$). The two features were then selected for CPD and constrained CPD on the third-order tensor. Figure 6 shows the mean of selected features with corresponding temporal and spatial components. No difference between emotions was found. Moreover, the selected temporal components extracted by the unconstrained CPD and the constrained CPD were not well separated out from mixtures of ERP waveforms. This was because

Int. J. Neur. Syst. 2014.24. Downloaded from www.worldscientific.com
by 81.197.1.127 on 09/15/21. Re-use and distribution is strictly not permitted, except for Open Access articles.



(a)



(b)

Fig. 5. Comparison between LRAHALS NCPD and HALS NCPD with single-round decomposition: (a) maximal value of j (j is the number of extracted components, $j \in [2, 80]$) correlation coefficients which were between extracted rank-one tensor and the template tensor; and (b) Fits of NCDP models with the number of components ranging from 2 to 80.

they did not match very well with the desired N170 temporal component.[20,32] The extracted temporal components not only contained the N170-relevant brain activity but also included other components. This means that they were still mixtures of brain activity. Furthermore, the largest signatures (absolute value) of the selected spatial maps appeared at O1 and Fz for the unconstrained CPD and the constrained CPD, respectively. The maximal signature did not appear at P7 or P8 (the two electrodes are the most representative for N170[21]). Thus, the extracted spatial maps were not satisfactory either.

Next, results by LRAHALS CPD with the numbers of components ranging from 2 to 80 and with 30 rounds were examined. It should be noted that the estimation of the number of extracted components by DIFFIT was 74.

When employing the third-order tensor by CPD with any number of components (between 2 and 80), we did not find any satisfactory extracted

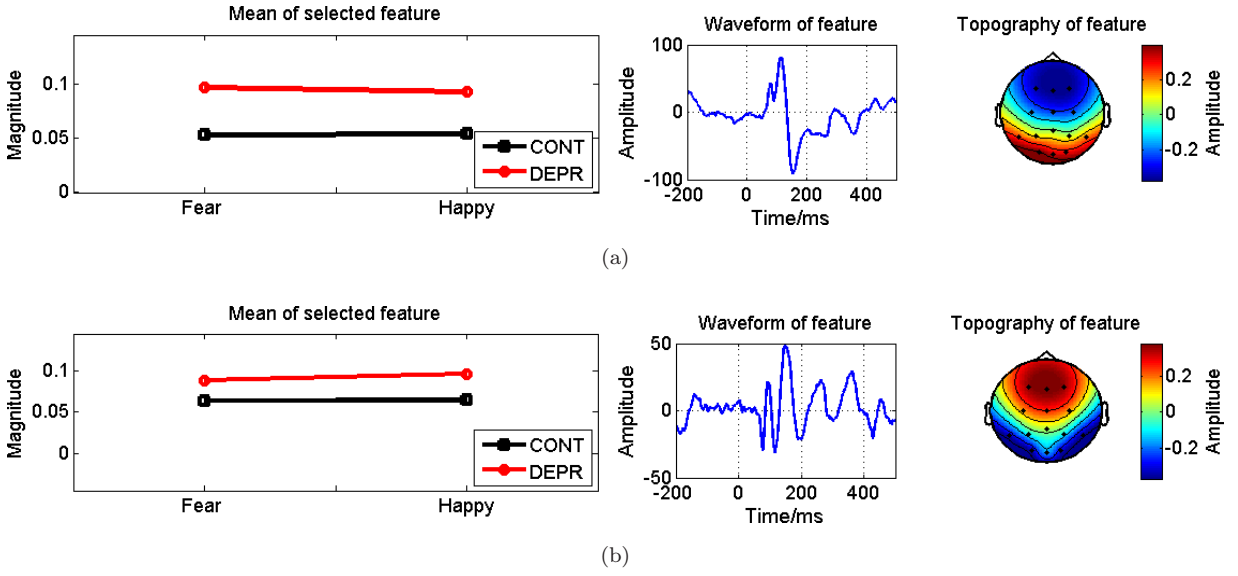


Fig. 6. Mean of selected multi-domain feature and its corresponding temporal and spatial components: (a) unconstrained CPD, (b) constrained CPD with the orthogonal in the time mode.

multi-domain features. This means no features showed any significant difference between the two emotions when considering temporal and spatial properties of the ERP, N170.

When CPD was applied on the ERP data here, the extracted temporal components always included more than one ERP component, as shown in Fig. 6. This means the extracted component did not purely represent one ERP component. This leads to difficulty in choosing the multi-domain feature of N170. However, we do have prior knowledge based on the results of NCPD on the TFR of ERP data. For example, the multi-domain feature of N170 may show a group difference or an emotion difference. Therefore, we used such prior knowledge to choose the most possible components of interest in CPD on ERP data. Without this, it would have been very difficult to use CPD on ERP data for the comparison with NCPD on TFR of ERP data.

4. Discussion

NTF has been shown to be superior in processing and analyzing ERPs.[5–8] Due to the large-scale high-order ERP tensor (for example, about 500 megabytes here) for group-level analysis, it is more realistic and practical to apply fast NTF algorithms. We have shown the LRA-based NTF algorithms are much faster than the bench-mark HALS NTF algorithms both in CP (shown here) and Tucker[17] models. Some conventionally used tensor decomposition algorithms have been applied to study spontaneous EEG for epilepsy diagnosis,[33–38] single-trial ERPs in Brain–computer Interface (BCI),[16,39–42] and multispectral images.[43,44] Therefore, LRA-based NTF is also recommended for research in these fields, as well as LRANMF for NMF on EEG.[45]

In this study, the magnitude of the desired multi-domain feature was significantly larger in depressive participants (DEPR) than in healthy control participants (CONT). Moreover, the feature was significantly larger for fearful faces (Fear) than for happy faces (Happy). Such a relationship between Fear and Happy in N170 was also reported previously using the oddball paradigm.[46–48] The grand averaged ERP amplitudes showed such tendencies at P7 in Fig. 1(c). However, no such statistically significant effects were observed in the conventional ERP data analysis. This indeed shows superiority of NTF in processing and analyzing ERPs. The results also suggest that the multi-domain feature extracted by NTF is a more sensitive indicator of information processing differences between depressives and controls.

As shown in Fig. 3, the multi-domain features #1 and #2 differ in revealing the contrast between the two groups. Such difference stems from the corresponding spatial components. These show foci in occipital sites and the temporo-parietal site,

Int. J. Neur. Syst. 2014.24. Downloaded from www.worldscientific.com
by 81.197.1.127 on 09/15/21. Re-use and distribution is strictly not permitted, except for Open Access articles.

Int. J. Neur. Syst. 2014.24. Downloaded from www.worldscientific.com
by 81.197.1.127 on 09/15/21. Re-use and distribution is strictly not permitted, except for Open Access articles.

respectively. This indeed suggests that the group difference in N170 can occur in the left temporo-parietal electrode site. For peak amplitudes of N170, such a tendency can be observed in Fig. 1(c); the difference was statistically marginal. We speculate that the difference mainly comes from the different lateralization of N170 in the CONT and DEPR groups. This is because, for the conventional ERP data in Fig. 1, there is a tendency of right hemisphere dominance in CONT, which is not seen in DEPR. DERP showed no hemispheric dominance. We will be seeking to validate this speculation in future research.

At P8, the peak amplitude of N170 to Happy was significantly larger than that to Fear. This agrees with our previous study[8] in which the desired multi-domain feature was extracted by NTD from difference wave (DW) of the ERP data. However, such an effect was not observed in the multi-domain feature of N170 in this study. As introduced in Sec. 2.1, the ERP data were filtered by a 1–30 Hz band-pass filter in the conventional ERP data processing. In Ref. 8, the frequency band of the TFR was also 1–30 Hz. Here, the frequency band of the TFR is 1–15 Hz. Hence, we speculate that the effect of emotions in the peak amplitude of N170 at P8 in this study is in the higher-frequency range.

In this study, fearful and happy faces served as deviant stimuli, and neutral faces as standard stimuli, in the passive oddball paradigm.[19,20] Indeed, in such a paradigm, both responses to deviant stimuli as well as DW of responses between the deviant stimuli and the standard stimuli can be of interest. This depends on different purposes for the research. DW is obtained by subtracting responses to standard stimuli from responses to deviant stimuli. Specifically, using DW, the mismatch between the deviant and the standard stimuli is of interest.[2] In the previous study,[8] DW was examined for healthy participants (whose responses to deviant stimuli are used here). DW can be much noisier than responses to deviant stimuli[49] since subtraction tends to produce higher frequency noise. This can apparently be observed from the waveform of responses to deviant stimuli (Fig. 1(a)) and that of DW (Fig. 1 in Ref. 8). The desired multi-domain features of ERPs in DW were not extracted by NCPD due to a higher level of noise, but by NTD (NTF based on the Tucker model) in Ref. 8. This is because the Tucker model can provide many more possibilities for decomposing a tensor than the CP model.[3,4,33]

Nevertheless, the Tucker model is more complicated than the CP model, which introduces practical difficulties. For example, in the Tucker model, the numbers of components can be different for different factors.[4] However, to estimate those parameters precisely can be very challenging for extracting a multi-domain feature of an ERP for cognitive neuroscience research.[8] For example, in another previous study,[7] the multi-domain feature of mismatch negativity (MMN) elicited by an auditory passive oddball paradigm was successfully extracted by NCPD from DW. Based on our experience, DW for auditory MMN often possesses higher SNR than that for visual MMN. Based on this, when NTF is applied to extract the multi-domain feature of an ERP, we recommend using the CP model first. If the desired feature cannot be extracted out, the Tucker model can then be implemented.

The fourth-order tensor of ERP data consists of the TFR of ERP data of multiple channels and multiple participants. Using NTF, the temporal, spectral, and spatial components can be regarded as the bases of the multi-domain feature of an ERP. The bases are common to all participants, and the multi-domain feature of the ERP contains the individual/group differences.[50] In other words, the multi-domain feature determines the difference between/among groups in terms of power of brain activity. Moreover, the bases determine the conditions of the occurrence of this difference (when, at which frequency, and where along the scalp the difference can occur). Therefore, the multi-domain feature is cognitive. Indeed, EEG feature extraction based on various nonlinear methods[51–59] has been extensively studied for clinical disease diagnosis in the field of computer science and artificial intelligence. However, such features are difficult to apply to cognitive neuroscience research, because they are difficult to interpret in the traditional cognitive neuroscience.

Apart from the fourth-order tensor of TFR of ERPs, there are many ways to organize the TFR of EEG data in an ERP experiment.[5,6] For example, the time and frequency modes were vectorized in Ref. 6. Alternatively, the subject and space modes were merged but the participants and experimental conditions were not combined in Ref. 5. The way to organize the ERP tensor is thus based on the research

Int. J. Neur. Syst. 2014.24. Downloaded from www.worldscientific.com
by 81.197.1.127 on 09/15/21. Re-use and distribution is strictly not permitted, except for Open Access articles.

question of interest when applying tensor decomposition.

Here, the EEG data were collected using a low-density array of only 14 electrodes. Nowadays, it is more common to use high-density arrays consisting of over 100 electrodes in neuroscience research (although low-density arrays continue to be widely used in clinical practice). This means that the size of high-order ERP tensor can be much larger, therefore developing a faster NTF algorithm is highly important. In this study, the fast NCPD algorithm is implemented by combining LRA and a basic NCPD algorithm (the benchmark HALS is used here). Therefore, it is expected that a faster LRA-based NCPD algorithm will be developed with a faster base NCPD algorithm. For example, the block principal pivoting method-based NCPD[60] has been shown to be several times faster than HALS NCPD.

NCPD on the fourth-order tensor and CPD on the third-order tensor were compared in this study. Using the latter approach, we did not find satisfactory multi-domain features. We think the key reason for this is that CPD did not separate well the mixtures of brain activity as shown in Fig. 6. The fourth-order tensor data of TFR of ERPs are non-negative. Hence, it is natural and objective to add the non-negative constraint for data decomposition. In blind separation of non-negative mixtures of non-negative sources, NMF has shown great superiority over many blind source separation methods.[3] Therefore, the tensor decomposition with non-negative constraints may extract more reasonable multi-domain features of ERPs from the non-negative TFR of ERP data.

The TFR of an ERP component is very sparse. It is not difficult to examine the sparse property if the temporal and spectral components in Fig. 3 are multiplied according to the outer product. Indeed, forcing non-negative constraints implicitly adds the sparse constraint in this study. Such implicit addition of sparse constraints benefits the separation of mixtures by ICA.[61] We think this is the main reason that the non-negative data (ERP's TFR) are better for tensor decomposition to extract the multi-domain feature of an ERP. Furthermore, real brain activity in EEG may be dipolar.[62] This means that the topography of real brain activity can be sparse. It has been shown that the sparse constraint is explicitly and implicitly useful in blind source separation and ICA to extract sparse sources.[61,63–65] Consequently, it is interesting to examine whether the sparse constraint on the spatial components of ERPs can benefit NCPD on the fourth-order tensor of ERPs. For tensor decomposition, adding constraint may result in loss of fit of raw tensor.[66] Therefore, the trade-off between adding sparse constraint and keeping enough fit is worthy of investigation.

5. Conclusion

The novel LRAHALS NCPD algorithm (NCPD based on LRA and HALS) can extract well the desired multi-domain feature of an ERP from ERP's TFR. In contrast to the benchmark HALS NCPD, LRAHALS NCPD is much faster in computational terms, and there is no significant loss of accuracy in decomposing the high-order ERP tensor.

Acknowledgments

This work was partially supported by TEKES (Finland) grant 40334/10 'Machine Learning for Future Music and Learning Technologies', and supported by the Fundamental Research Funds for the Central Universities (China, DUT14RC(3)037). A.K. Nandi would like to thank TEKES for their award of the Finland Distinguished Professorship. F. Cong thanks the Research and Innovation Office of the University of Jyväskylä for the international mobility grant (2009, 2010) and thank Dr. Anh Huy Phan (BSI RIKEN, Japan) for helping in learning tensor decomposition. Q. Zhao was partly supported by JSPS Grants-in-Aid for Scientific Research (Grant No. 24700154) and the National Natural Science Foundation of China (Grant No. 61202155). F. Cong and G. Zhou contribute equally to this work.

Appendix A

Outer product of vectors

Given three vectors $\mathbf{a} \in \mathbb{R}^{\mathbf{I}}$, $\mathbf{b} \in \mathbb{R}^{\mathbf{J}}$ and $\mathbf{c} \in \mathbb{R}^{\mathbf{Q}}$, their outer product yields a third-order rank-one tensor:

$$\underline{\mathbf{Z}} = \mathbf{a} \circ \mathbf{b} \circ \mathbf{c} \in \mathbb{R}^{\mathbf{I}\times\mathbf{J}\times\mathbf{Q}},$$

where $z_{ijq} = a_i b_j c_q$.

Mode-n tensor matrix product

The mode-n product $\underline{\mathbf{Y}} = \underline{\mathbf{G}} \times_n \mathbf{A}$ of a tensor $\underline{\mathbf{G}} \in \mathbb{R}^{J_1\times J_2\times\cdots\times J_N}$ and a matrix $\mathbf{A} \in \mathbb{R}^{I_n\times J_n}$ is a tensor

Int. J. Neur. Syst. 2014.24. Downloaded from www.worldscientific.com
by 81.197.1.127 on 09/15/21. Re-use and distribution is strictly not permitted, except for Open Access articles.

$\underline{\mathbf{Y}} \in \mathbb{R}^{J_1 \times J_2 \times \cdots \times J_{n-1} \times I_n \times J_{n+1} \times \cdots \times J_N}$, with elements

$$y_{j_1 j_2 \cdots j_{n-1} i_n j_{n+1} \cdots j_N} = \sum_{j_n=1}^{J_n} g_{j_1 j_2 \cdots j_N} a_{i_n, j_n}.$$

References

1. S. J. Luck, *An Introduction to the Event-Related Potential Technique* (The MIT Press, 2005).
2. R. Näätänen, T. Kujala, K. Kreegipuu, S. Carlson, C. Escera, T. Baldeweg and C. Ponton, The mismatch negativity: An index of cognitive decline in neuropsychiatric and neurological diseases and in ageing, *Brain* **134** (2011) 3432–3450.
3. A. Cichocki, R. Zdunek, A. H. Phan and S. Amari, *Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-Way Data Analysis* (John Wiley, 2009).
4. T. Kolda and B. Bader, Tensor decompositions and applications, *SIAM Rev.* **51** (2009) 455–500.
5. M. Morup, 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* **29** (2006) 938–947.
6. M. Morup, L. K. Hansen and S. M. Arnfred, ERPWAVELAB a toolbox for multi-channel analysis of time-frequency transformed event related potentials, *J. Neurosci. Methods* **161** (2007) 361–368.
7. F. Cong, A. H. Phan, Q. Zhao, T. Huttunen-Scott, J. Kaartinen, T. Ristaniemi, H. Lyytinen and A. Cichocki, 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** (2012) 1–19.
8. F. Cong, A. H. Phan, P. Astikainen, Q. Zhao, Q. Wu, J. K. Hietanen, T. Ristaniemi and A. Cichocki, Multi-domain feature extraction for small event-related potentials through nonnegative multi-way array decomposition from low dense array EEG, *Int. J. Neural Syst.* **23** (2013) 1–18.
9. K. Vanderperren, B. Mijovic, N. Novitskiy, B. Vanrumste, P. Stiers, B. R. Van den Bergh, L. Lagae, S. Sunaert, J. Wagemans, S. Van Huffel and M. De Vos, Single trial ERP reading based on parallel factor analysis, *Psychophysiology* **50** (2013) 97–110.
10. F. L. Hitchcock, The expression of a tensor or a polyadic as a sum of products, *J. Math. Phys.* **6** (1927) 164–189.
11. L. R. Tucker, Some mathematical notes on three-mode factor analysis, *Psychometrika* **31** (1966) 279–311.
12. R. A. Harshman, Foundations of the PARAFAC procedure: Models and conditions for an "explanatory" multi-modal factor analysis, *UCLA Working Papers Phonetics* **16** (1970) 1–84.
13. J. D. Carroll and J. Chang, Analysis of individual differences in ultidimensional scaling via an n-way generalization of 'Eckart–Young' decomposition, *Psychometrika* **35** (1970) 283–319.
14. A. Cichocki, R. Zdunek and S. Amari, Hierarchical ALS algorithms for nonnegative matrix and 3D tensor factorization, eds. M. E. Davies *et al.*, Lecture Notes in Computer Science, Vol. 4666, ICA 2007 (2007), pp. 169–176.
15. G. Zhou, A. Cichocki and S. Xie, Fast nonnegative matrix/tensor factorization based on low-rank approximation, *IEEE Trans. Signal Process.* **60** (2012) 2928–2940.
16. A. H. Phan and A. Cichocki, Extended HALS algorithm for nonnegative Tucker decomposition and its applications for multiway analysis and classification, *Neurocomputing* **74** (2011) 1956–1969.
17. F. Cong, G. Zhou, Q. Zhao, Q. Wu, A. K. Nandi, T. Ristaniemi and A. Cichocki, Sequential nonnegative Tucker decomposition on multi-way array of time-frequency transformed event-related potentials, in *Proc. 2012 IEEE Workshop on Machine Learning for Signal Processing* (*MLSP*) (2012), pp. 1–6.
18. A. Cichocki and A. H. Phan, Fast local algorithms for large scale nonnegative matrix and tensor factorizations, *IEICE Trans. Fund. Electron. Commun. Comput. Sci.* **92-A** (2009) 708–721.
19. P. Astikainen and J. K. Hietanen, Event-related potentials to task-irrelevant changes in facial expressions, *Behav. Brain Funct.* **5** (2009) 30.
20. P. Astikainen, F. Cong, T. Ristaniemi and J. K. Hietanen, Event-related potentials to unattended changes in facial expressions: Detection of regularity violations or encoding of emotions? *Front. Hum. Neurosci.* **7** (2013) 557.
21. B. Rossion and C. Jacques, Does physical interstimulus variance account for early electrophysiological face sensitive responses in the human brain? Ten lessons on the N170, *Neuroimage* **39** (2008) 1959–1979.
22. C. Tallon-Baudry, O. Bertrand, C. Delpuech and J. Pernier, Stimulus specificity of phase-locked and non-phase-locked 40 Hz visual responses in human, *J. Neurosci.* **16** (1996) 4240–4249.
23. G. Zhou and A. Cichocki, Laboratory for tensor decomposition and analysis, TDALAB Ver1.0 (2013), Available at http://bsp.brain.riken.jp/TDALAB/.
24. R. V. Hogg and J. Ledolter, *Engineering Statistics* (MacMillan, New York, 1987).
25. M. Mørup and L. K. Hansen, Automatic relevance determination for multiway models, *J. Chemometr.* **23** (2009) 352–363.
26. M. E. Timmerman and H. A. Kiers, Three-mode principal components analysis: Choosing the numbers of components and sensitivity to local optima, *Br. J. Math. Stat. Psychol.* **53** (Pt 1) (2000) 1–16.
27. J. B. Kruskal, Three-way arrays: Rank and uniqueness of trilinear decompositions, with application

Int. J. Neur. Syst. 2014.24. Downloaded from www.worldscientific.com
by 81.197.1.127 on 09/15/21. Re-use and distribution is strictly not permitted, except for Open Access articles.

to arithmetic complexity and statistics, *Linear Alg. Appl.* **18** (1977) 95–138.
28. N. D. Sidiropoulos and R. Bro, On the uniqueness of multilinear decomposition of N-way arrays, *J. Chemometr.* **14** (2000) 229–239.
29. S. Makeig, T. P. Jung, A. J. Bell, D. Ghahremani and T. J. Sejnowski, Blind separation of auditory event-related brain responses into independent components, *Proc. Natl. Acad. Sci. U. S. A.* **94** (1997) 10979–10984.
30. S. Makeig, M. Westerfield, T. P. Jung, J. Covington, J. Townsend, T. J. Sejnowski and E. Courchesne, Functionally independent components of the late positive event-related potential during visual spatial attention, *J. Neurosci.* **19** (1999) 2665–2680.
31. C. A. Andersson and R. Bro, The N-way Toolbox for MATLAB, *Chemometr. Intell. Lab. Syst.* **52** (2000) 1–4.
32. F. Cong, P. H. Leppanen, P. Astikainen, J. Hämäläinen, J. K. Hietanen and T. Ristaniemi, Dimension reduction: Additional benefit of an optimal filter for independent component analysis to extract event-related potentials, *J. Neurosci. Methods* **201** (2011) 269–280.
33. E. Acar and B. Yener, Unsupervised multiway data analysis: A literature survey, *IEEE Trans. Knowl. Data Eng.* **21** (2009) 6–20.
34. E. Acar, C. Aykut-Bingol, H. Bingol, R. Bro and B. Yener, Multiway analysis of epilepsy tensors, *Bioinformatics* **23** (2007) i10–8.
35. E. Acar, C. A. Bingöl and H. Bingöl, Computational analysis of epileptic focus localization, in *Proc. Fourth IASTED Int. Conf. Biomedical Engineering* (2006), pp. 317–322.
36. M. De Vos, L. De Lathauwer, B. Vanrumste, S. Van Huffel and W. Van Paesschen, Canonical decomposition of ictal scalp EEG and accurate source localisation: Principles and simulation study, *Comput. Intell. Neurosci.* (2007) 58253.
37. M. De Vos, A. Vergult, L. De Lathauwer, W. De Clercq, S. Van Huffel, P. Dupont, A. Palmini and W. Van Paesschen, Canonical decomposition of ictal scalp EEG reliably detects the seizure onset zone, *Neuroimage* **37** (2007) 844–854.
38. H. Lee, Y. D. Kim, A. Cichocki and S. Choi, Nonnegative tensor factorization for continuous EEG classification, *Int. J. Neural Syst.* **17** (2007) 305–317.
39. A. Cichocki, Y. Washizawa, T. M. Rutkowski, H. Bakardjian, A. H. Phan, S. Choi and Q. Zhao, Noninvasive BCIs: Multiway signal-processing array decompositions, *Computer* **41** (2008) 34–42.
40. A. H. Phan and A. Cichocki, Tensor decomposition for feature extraction and classification problem, *IEICE Trans. Fund. Electron. Commun. Comput. Sci.* **1** (2010) 37–68.
41. J. Li and L. Q. Zhang, Regularized tensor discriminant analysis for single trial EEG classification in BCI, *Pattern Recogn. Lett.* **31** (2010) 619–628.
42. J. Li, L. Q. Zhang, D. Tao, H. Sun, Q. B. Zhao, A prior neurophysiologic knowledge free tensor-based scheme for single trial EEG classification, *IEEE Trans. Neural Syst. Rehab. Eng.* **17** (2009) 107–115.
43. I. Kopriva, M. Hadzija, M. Popovic Hadzija, M. Korolija and A. Cichocki, Rational variety mapping for contrast-enhanced nonlinear unsupervised segmentation of multispectral images of unstained specimen, *Am. J. Pathol.* **179** (2011) 547–554.
44. I. Kopriva, A. Persin, N. Puizina-Ivic and L. Miric, Robust demarcation of basal cell carcinoma by dependent component analysis-based segmentation of multi-spectral fluorescence images, *J. Photochem. Photobiol. B.* **100** (2010) 10–18.
45. A. Cichocki and R. Zdunek, Multilayer nonnegative matrix factorization using projected gradient approaches, *Int. J. Neural Syst.* **17** (2007) 431–446.
46. M. Batty and M. J. Taylor, Early processing of the six basic facial emotional expressions, *Brain Res. Cogn. Brain Res.* **17** (2003) 613–620.
47. J. M. Leppanen, M. C. Moulson, V. K. Vogel-Farley and C. A. Nelson, An ERP study of emotional face processing in the adult and infant brain, *Child Dev.* **78** (2007) 232–245.
48. M. Rossignol, P. Philippot, C. Douilliez, M. Crommelinck and S. Campanella, The perception of fearful and happy facial expression is modulated by anxiety: An event-related potential study, *Neurosci. Lett.* **377** (2005) 115–120.
49. I. Kalyakin, N. Gonzalez, J. Joutsensalo, T. Huttunen, J. Kaartinen and H. Lyytinen, Optimal digital filtering versus difference waves on the mismatch negativity in an uninterrupted sound paradigm, *Dev. Neuropsychol.* **31** (2007) 429–452.
50. A. Cichocki, Tensors decompositions: New concepts for brain data analysis? *J. Control Meas., Syst. Integr.* **6** (2013) 507–517.
51. U. R. Acharya, S. V. Sree, S. Chattopadhyay, W. Yu and P. C. Ang, Application of recurrence quantification analysis for the automated identification of epileptic EEG signals, *Int. J. Neural Syst.* **21** (2011) 199–211.
52. U. R. Acharya, S. V. Sree and J. S. Suri, Automatic detection of epileptic EEG signals using higher order cumulant features, *Int. J. Neural Syst.* **21** (2011) 403–414.
53. H. Adeli and S. Ghosh-Dastidar, *Automated EEG-based Diagnosis of Neurological Disorders — Inventing the Future of Neurology* (CRC press, Florida, USA, 2010).
54. H. Adeli, S. Ghosh-Dastidar and N. Dadmehr, A spatio-temporal wavelet-chaos methodology for

EEG-based diagnosis of Alzheimer's disease, *Neurosci. Lett.* **444** (2008) 190–194.

55. M. Ahmadlou and H. Adeli, Functional community analysis of brain: A new approach for EEG-based investigation of the brain pathology, *Neuroimage* **58** (2011) 401–408.
56. O. Faust, U. R. Acharya, L. C. Min and B. H. Sputh, Automatic identification of epileptic and background EEG signals using frequency domain parameters, *Int. J. Neural Syst.* **20** (2010) 159–176.
57. L. J. Herrera, C. M. Fernandes, A. M. Mora, D. Migotina, R. Largo, A. Guillen and A. C. Rosa, Combination of heterogeneous EEG feature extraction methods and stacked sequential learning for sleep stage classification, *Int. J. Neural Syst.* **23** (2013) 1350012.
58. R. J. Martis, U. R. Acharya, J. H. Tan, A. Petznick, L. Tong, C. K. Chua and E. Y. Ng, Application of intrinsic time-scale decomposition (ITD) to EEG signals for automated seizure prediction, *Int. J. Neural Syst.* **23** (2013) 1350023.
59. Z. Sankari, H. Adeli and A. Adeli, Intrahemispheric, interhemispheric, and distal EEG coherence in Alzheimer's disease, *Clin. Neurophysiol.* **122** (2011) 897–906.
60. J. Kim, H. Park, Fast nonnegative tensor factorization with an active-set-like method, in *High-Performance Scientific Computing*, eds. M. W. Berry, K. A. Gallivan, E. Gallopoulos *et al.* (Springer, London, 2012), pp. 311–326.
61. I. Daubechies, E. Roussos, S. Takerkart, M. Benharrosh, C. Golden, K. D'Ardenne, W. Richter, J. D. Cohen and J. Haxby, Independent component analysis for brain fMRI does not select for independence, *Proc. Natl. Acad. Sci. U. S. A.* **106** (2009) 10415–10422.
62. A. Delorme, J. Palmer, J. Onton, R. Oostenveld and S. Makeig, Independent EEG sources are dipolar, *PLoS One* **7** (2012) e30135.
63. Z. Yang, G. Zhou, S. Xie, S. Ding, J. M. Yang and J. Zhang, Blind spectral unmixing based on sparse nonnegative matrix factorization, *IEEE Trans. Image Process.* **20** (2011) 1112–1125.
64. G. Zhou, S. Xie, Z. Yang, J. M. Yang and Z. He, Minimum-volume-constrained nonnegative matrix factorization: Enhanced ability of learning parts, *IEEE Trans. Neural Netw.* **22** (2011) 1626–1637.
65. G. Zhou, Z. Yang, S. Xie and J. M. Yang, Online blind source separation using incremental nonnegative matrix factorization with volume constraint, *IEEE Trans. Neural Netw.* **22** (2011) 550–560.
66. R. Bro, Multi-way analysis in the food industry — models, algorithms, and applications, PhD thesis, University of Amsterdam, Holland (1998).

Int. J. Neur. Syst. 2014.24. Downloaded from www.worldscientific.com
by 81.197.1.127 on 09/15/21. Re-use and distribution is strictly not permitted, except for Open Access articles.