Chapter 1

EEG CLASSIFICATION OF MILD AND SEVERE ALZHEIMER'S DISEASE USING PARALLEL FACTOR ANALYSIS METHOD

PARAFAC Decomposition of Spectral-Spatial Characteristics of EEG Time Series

Charles-Francois Vincent Latchoumane[1], Francois-Benois Vialatte[2], Jaeseung Jeong[*] and Andrzej Cichocki[†]

[1]*Brain Dynamics Lab., Bio and Brain Engineering Dept., KAIST, Yuseong-gu, Guseong-dong, Daejeon, South Korea,305-701; Tel: +82-42-350-4319; Fax: +82-42-350-4310; Email: kskusanagi@kaist.edu*

[2]*Lab. of Advanced Brain Signal Processing, Brain Science Institute, RIKEN, 2-1 Hirosawa, Wako Saitama, Japan, 351-0198; Tel: +81-48-467-9668; Fax: +81-48-467-9686; Email: fvialatte@brain.riken.jp*

[*]*Brain Dynamics Lab., Bio and Brain Engineering Dept., KAIST, Yuseong-gu, Guseong-dong, Daejeon, South Korea,305-701; Tel: +82-42-350-4319; Fax: +82-42-350-4310; Email: jsjeong@kaist.ac.kr*

[†]*Lab. of Advanced Brain Signal Processing, Brain Science Institute, RIKEN, 2-1 Hirosawa, Wako Saitama, Japan, 351-0198; Tel: +81-48-467-9668; Fax: +81-48-467-9686; Email: cia@brain.riken.jp*

1. INTRODUCTION

1.1 Diagnosis of Alzheimer's disease & EEGs

Alzheimer's disease is the most prevalent neuropathology form leading to dementia; it affects approximately 25 million people worldwide and is expected to have fast recrudescence in near future[1]. The numerous clinical methods that are now available to detect this disease including imaging[2, 3], genetic methods[4], and other physiological markers[5], however, do not allow a mass screening of the population. Whereas psychological tests such as Mini Mental State Evaluation (MMSE) in combination with an electrophysiological analysis (e.g. electroencephalograms or EEG) would be very efficient and inexpensive screening approach to detect the patients damaged by the disease.

EEG recordings are now widely used more and more as a method to assess the susceptibility to Alzheimer's disease, but are often obtained during steady states where temporal information does not easily reveal the relevant features for subject differentiation; interestingly, in those conditions, previous studies could obtain excellent classification results ranging from 84 to 100%, depending on the conditions (i.e. training or validation conditions for the classifier tests) and on the confronted groups (i.e. control subject, mild cognitive impairment (MCI), and different stages of Alzheimer's disease) demonstrating the promising use of resting EEGs for diagnosis of Alzheimer's disease[6-11].

The spectral spatial information estimated during resting states might contain valuable clues for the detection of demented subjects; however, the inter-subject variability, also influenced by the differences in the progression of the disease, might render the study difficult when undertaken subject-by-subject. In that case, a multi-way analysis would allow the extraction of information that is contained across subjects simultaneously considering the spectral spatial information. This methodology has been applied to epilepsy detection and has successfully characterized the epilepsy foci in a temporal-frequency-regional manner[12, 13]. Classification based on multi-way modeling has been performed on continuous EEGs[14] showing the power and versatility of multi-way analyses.

1.2 Multi-way array decomposition

Previous two-way analyses combining PCA-like techniques[6] (i.e. principal component analysis (PCA) and independent component analysis (ICA)) have shown very high performance in the classification of subjects and have assisted in early detection. These methods require data in the form of matrices, and then they limit the order (i.e. number of dimensions or modes) or mix several types of variables (e.g. using the unfolding method, also known as matricization). Either way, the naturally high-order of EEGs and interactions between the ways (or modes) is lost or destroyed, limiting further the understanding of underlying processes in the brain.

The multi-way array decomposition is a modeling tool that conserves the original high dimensional nature of the data. This method uses the common features and interactions between modes present in the data to create a model fit of the same data that can be decomposed into components. The decomposition into components provides the information of interactions between modes in the form of weight, which is relatively easier to interpret. The application of such methods to diagnosis of Alzheimer's disease would allow characterization of the disease based on simple markers or weights.

Thus far, no application of multi-way array decomposition has been made in this type of database, dealing with the classification of Alzheimer's disease subjects based on EEG characteristics.

2. SUBJECTS & EEG RECORDINGS

The subjects analyzed in this study were obtained from a previously studied database[6, 9, 11] and consisted of eyes-open, steady state EEG recordings (20 s in duration), with over 21 leads disposed according to the 10-20 international system and digitalized at 200 Hz. The database contains 38 control (mean age: 71.7±8.3) subjects, 22 mild cognitive impairment (MCI) subjects (mean age: 71.9±10.2) who later contracted Alzheimer's disease, and 23 Alzheimer's disease patients (AD; mean age: 72.9±7.5). The control subjects had no complaints or history of memory problems, and scored over 28 (mean score: 28.5±1.6) on the mini mental state exam (MMSE). The MCI subjects had complaints about memory problems and scored over 24 at the MMSE (mean score: 26±1.8). The inclusion criterion was set at 24 as suggested in[11], therefore encompassing MCI subjects with various cognitive deficits, but in the early stage of Alzheimer's disease. The AD patients scored below 20 on the MMSE and had had full clinical assessment. Thirty-three moderately or severely demented probable AD patients (mean MMSE score: 15.3±6.4; range: 0 to 23) were recruited from the same clinic. After obtaining written informed consent from the patients and controls, all subjects underwent EEG and SPECT examination within one month of entering the study. All subjects were free of acute exacerbations of AD related co-morbidities and were not taking medication. The medical ethical committee of the Japanese National Center of Neurology and Psychiatry approved this study.

3. METHOD

3.1 Three-way tensor for analysis

In this study, we aimed at classifying control subjects from subjects with MCI and from AD. For each subject, we computed the relative Fourier power of five frequency bands δ (1-4 Hz), θ (4-8 Hz), α_1 (8-10 Hz), α_2 (10-12 Hz) and β (12-25 Hz); then for each frequency band, we estimated the mean power of five brain regions: frontal, left temporal, central, right temporal and posterior. There were an equivalent number of electrodes in each of the five regions. This grouping is very useful in normalizing the regional repartition of the information (Figure 1-1).

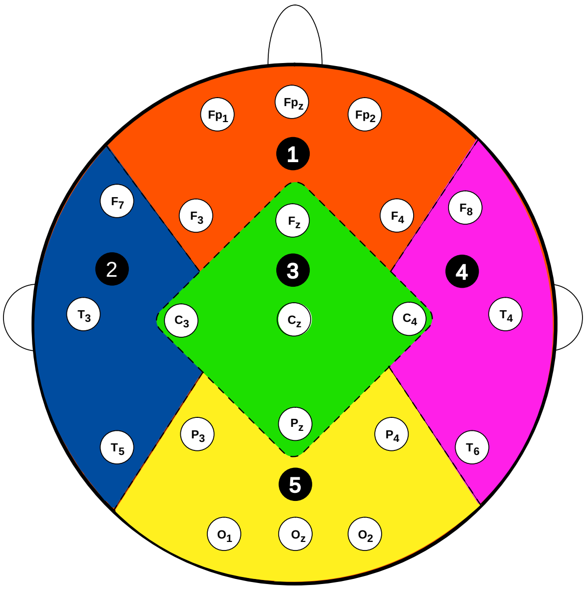

Figure 1-1. Topological grouping of electrodes in five regions; the numbers 1, 2, 3, 4 and 5 denote the frontal, left temporal, central, right temporal and posterior regions, respectively.

We can form a three-way tensor, *Subjects* x *Frequency Band Power* x *Brain Region*, which is the Fourier power by frequency band and by brain region for each subject. From this tensor, we extracted characteristic filters for the classification of subjects using linear and nonlinear classifiers. The detail of the method is described in the next sections.

3.2 Classification in Divide and Conquer Scheme

For simplicity and interpretation, we opted for two-step classification of the subjects (divide and conquer scheme): (1) we compared healthy subjects with the patient group (regrouping MCI and AD); (2) we compared MCI group with AD group. The schematic diagram of our approach is presented in Figure 1-2.

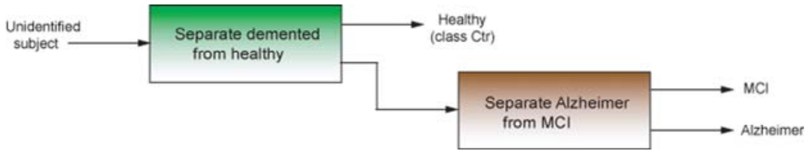

Figure 1-2. Divide and Conquer method for the classification of control subjects vs. MCI vs. AD subjects – cascade of classifiers.

For each step of classification, we extracted the features based on the Parallel Factor Analysis (PARAFAC) decomposition (unsupervised) and the reference filters (supervised). We estimated the accuracy in classification for subjects using a k-fold cross-validation with $k = 1$ (i.e. leave-one-out cross validation) on both a quadratic naïve Bayes classifier (one of the simplest statistical classifier with quadratic boundary) and an artificial neural network (ANN; with $n = 1$ hidden neurons in one hidden layer, i.e. nonlinear classifier).

3.3 The PARAllel FACtor Analysis (PARAFAC)

The important idea underlying multi-way analysis is the extraction of multiple interactions within the structure of the data. Interactions and common patterns between modes of the data are often neglected cause studied separately (e.g. subject by subject, region by region) or folded on the same dimension (i.e. mixed, hence destroying interactions) as for method using two dimensional PCA or ICA. In this study, we constructed a three-way tensor in the form *Subjects* x *Frequency Band Power* x *Brain Region* to preserve relations between spectral and regional characteristics of each subject; we used a common modeling method for N-way analyses to extract common interaction in the tensor with relation to the subjects: the parallel factor analysis (PARAFAC)[15].

The PARAFAC is often referred to as a multilinear version of the bilinear factor models. From a given tensor, $X \in \mathbb{R}^{I \times J \times K}$, this model is able to extract linear decompositions of Rank-1 tensors.

$$\underline{X} = \sum_{r=1}^{R} a_r \circ b_r \circ c_r + \underline{E} \quad , \qquad (1)$$

where $\boldsymbol{a}_r$, $\boldsymbol{b}_r$, and $\boldsymbol{c}_r$ are the r-th column of the component matrices $A \in \mathbb{R}^{I \times R}$, $B \in \mathbb{R}^{J \times R}$, and $C \in \mathbb{R}^{K \times R}$, respectively, and $E \in \mathbb{R}^{I \times J \times K}$ is the residual error matrix (Figure 1-3).

The operator $\circ$ designates the outer product of two vectors. The PARAFAC model under optimal fitting conditions (e.g. core consistency analysis) is capable of providing a model with the assumption of trilinearity relations between the dimensions (also called modes), thus providing a unique fit for the given data.

The fitting of the model depends on the number of components, R, chosen by the users, and for this approach we opted for validation of the model (i.e. number of components) based on the core consistency[16]. The optimal number R is chosen for a core consistency value that is above 90% and with the core consistency value of R+1 under 90%.

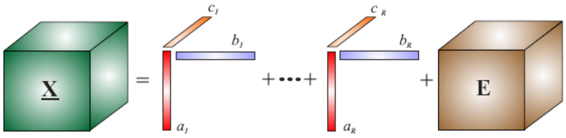

Figure 1-3. PARAFAC modeling of a 3-way tensor; each component (R=2) is the outer product of Rank-1 tensor a, b, and c, and E is a residual tensor.

3.4 Filter estimation and Feature Extraction

3.4.1 PARAFAC Filters

The PARAFAC model with a sufficient number of components and an appropriate fitting (i.e. core consistency[16] and component variability) is capable of fitting a model that decompose data in component of common characteristics with trilinear interaction between modes. We are interested in features that differentiate the subjects according to their clinical group. We assume that the common characteristic contain in the data could significantly discriminate the group based on the difference between the *Frequency Band Power x Brain Region* profile and normal control and demented subjects, i.e. shift in the frequency profile of subjects with AD or “slowing of frequencies”[17, 18] .

The PARAFAC decomposition returns for each component and each mode a Rank-1 vector which can be considered as a weight at each component and each mode, respectively. The mode subject contains the discriminative features or weight of each subject to be compared with the combination of the mode Frequency Band Power and the mode Brain Region. More precisely, the component-wise combination (i.e. for each component) of "*Frequency Band Power*" and "*Brain Region*" could be interpreted as a characteristic filter. Similar to the Fourier Power Distribution which associates a power to each frequency, the decomposition obtained from PARAFAC can represent a distribution of filters (number of filters $R = 3$) associated with a weight (i.e. subject mode). We calculated the characteristic filters for each component, F_r, as described in Eq. 2:

$$F_r = b_r \times c_r^T , \tag{2}$$

where b_r and c_r are the Rank-1 vector of the frequency and region mode in the PARAFAC model, respectively. Following Eq. 2, we obtain a number of filters $R = 3$.

3.4.2 Reference Filters

The reference filter are supervised filters in contrast to the filters obtained using PARAFAC (unsupervised) because they integrate prior knowledge of the membership of the subjects. The reference filters of the control group and demented group were calculated as the average *Frequency Band Power x Brain Region* over the normal subjects and the demented subjects (regrouping MCI and AD), respectively. The reference filters of the MCI group and AD group were calculated as the average *Frequency Band Power x Brain Region* over MCI subjects and AD subjects, respectively.

3.5 From Filter to Feature

The reference filter and the PARAFAC filters contain common characteristics of the group under this study (i.e. control vs. demented or MCI vs. AD). We use a distance measure based on the matrix scalar product also known as normalized Frobenius inner product to compare the estimated filters (i.e. reference filters or PARAFAC filters) with the *Frequency Band Power x Brain Region* profile of each subject. The details on the distance measure are as follows:

$$Dis(F_1, F_2) = \frac{Trace(F_1^T F_2)}{\|F_1\| \|F_2\|} , \tag{3}$$

where F_1 and F_2 are two *Frequency Band Power* x *Brain Region* matrices, ^{T} denotes the conjugate transposition of a matrix and the Trace function returns the trace of a matrix. The function ||.|| indicates the Frobenius norm defined as $\|F\| = \sqrt{Trace(F^T F)}$. For each subject, we obtained *R* features comparing the *R* PARAFAC filter to the subject's *Frequency Band Power* x *Brain Region* profile. Similarly, we obtained two features comparing the references filters to the subject's *Frequency Band Power* x *Brain Region* profile.

4. RESULTS

The best classification results using the artificial neural network (ANN) and quadratic naïve Bayes classifier are displayed in Table 1-1. There was no noticeable difference between the performance of the two classifiers (both nonlinear) using either the reference filters or the PARAFAC filters; however, while comparing MCI with AD subjects, the ANN showed better performance using the PARAFAC filters (75.6%) than using the reference filters (68.9%).

Table 1-1. Classification results obtained using an ANN and quadratic naïve Bayes classifier with the leave-one-out cross-validation method; NC denotes normal controls and Patients denotes patient group (regrouping MCI group and AD group).

	NC vs. Patients (PARAFAC)	NC vs. Patients (reference)	AD vs. MCI (PARAFAC)	AD vs. MCI (reference)
ANN	74.7%	61.4%	75.6%	64.4%
Bayes	74.7%	61.5%	68.9%	64.4%

Generally, the PARAFAC filters showed an improvement in the accuracy of classification as compared with the reference filters and the two classification methods used.

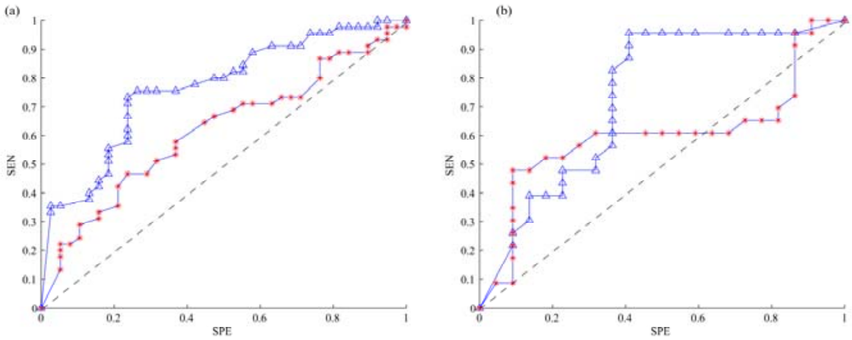

Figure 1-4. ROC curve of classification accuracy of (a) control vs. demented subjects and (b) AD vs. MCI subjects using an ANN and leave-one-out cross-validation method; classification results obtained using the reference filters (stars) and using the filtered data extracted with a three-component PARAFAC ($R = 3$; triangle).

As shown in Figure 1-4, the performance in the classification based on the ANN exhibited higher accuracy using the information from the PARAFAC filters than when using the reference filters. The best performance was found to be 74.7% (75.6% sensitivity, 73.7% specificity) for the Ctr vs. demented classification (Figure 1-4(a)) and 75.6% (95.6% sensitivity, 59.1% specificity) for MCI vs. AD classification (Figure 1-4(b)).

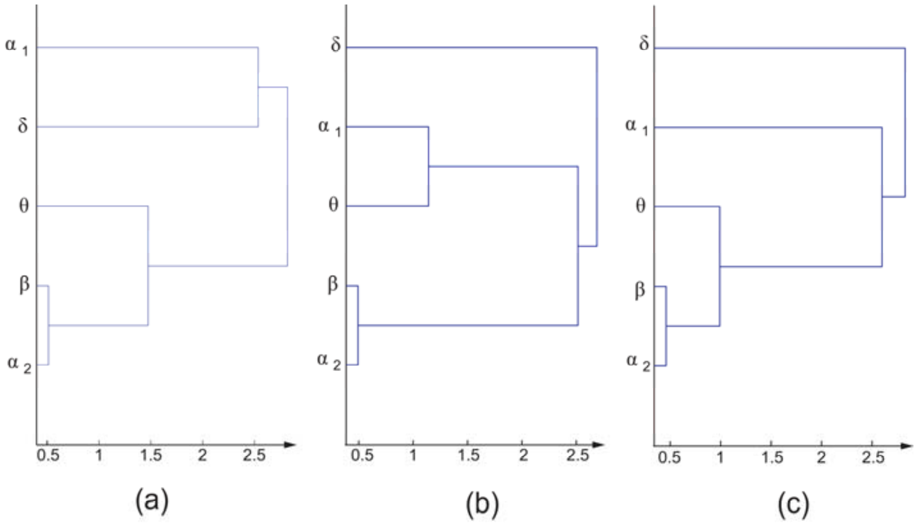

Figure 1-5. Clustering of frequency bands; Dendrograms extracted from the clustering of two-component PARAFAC models for (a) control subject, (b) MCI patients, and (c) Alzheimer patients.

Using PARAFAC on each separate group, a clustering of frequency bands (i.e. frequency band mode) can also be obtained[19] in order to study the interactions between the frequency ranges (Figure 1-6). We observed the evolutions in the inter-relations between the θ range and high frequency (α_2 and β) ranges: seemingly, for Control subjects and AD patients, the θ range activity is clustered with high frequencies (distances of 1.5 and 1, respectively); whereas for the MCI subjects, the θ range activity is much less (a distance of 2.5).

5. DISCUSSION

In this study, we presented a method applied to the classification of subjects based on a PARAFAC decomposition of their EEG features. This type of application of multi-way analysis to EEG features has not yet been implemented for AD diagnosis. We also showed possible interpretation of the fitted model not only based on the spatial-frequency filters, but also based on the unimodal clustering for each group's model. The weight obtained using the PARAFAC decomposition could serve as a marker for differentiation between control and patient group but also between MCI and AD subjects.

Previous studies on Alzheimer's classification and compromising with different degrees of severity of the disease obtained classification accuracies ranging from 84% to 100%[6-11]. The classification results presented in this study (overall good classification of about 74%) are considerably good compared with the classification results using the same database in another study[6] (80%); however, this study proposes a three-class classification using a two step divide and conquer approach.

The capability of the method to make a stepwise differentiation of subjects illustrates the applicability of this study to the clinical domain. The method also demonstrates possibility of mass-screening of subjects for the diagnosis of AD as well as the differentiation of stages in the progression of the disease (MCI developed to turn out to be mild and severe).

In this study, we determined to use a three-way tensor confronting subjects, frequency band power and brain regions allowing simplicity of computation and interpretation. The method using multi-way array decomposition allows the integration of a large number of variables, organized on higher orders (dimensions of the tensor equal to D, typically with D higher than 3). The addition of more variables necessitate that those variables would commonly interact and characterize the data or, in our case, the subjects in the data, otherwise the extracted information might lose its property of separability. More importantly, increasing the number of variables or order would not seriously increase the number of features for classification (i.e. number of components of the PARAFAC model, *R*) which would avoid artificially increasing the separability of the groups and reducing the bias due to overfitting for the classifier.

Moreover, apart from its crucial uniqueness and the resulting easy interpretability[20], the PARAFEC model used might not be the best model to use in this analysis, as it imposes trilinearity conditions. Future investigation include the comparison of the accuracy of other models such as PARAFAC2[21], Tucker3[22], or the nonnegative tensor factorization method[20, 23] and its constrained model on sparsity[24].

ACKNOWLEDGEMENTS

The first author would like to thank the Minister of Information and Technology of South Korea, Institute for Information and Technology Advancement (IITA) for his financial support. The authors would like to thank Dr. T. Musha for his contribution in subjects' analysis and the EEG recordings and Dr. M. Maurice and S. Choe for their help in the editing of the content.

6. REFERENCES

1. Ferri, C.P. et al. Global prevalence of dementia: a Delphi consensus study. *The Lancet* **366**, 2112-2117 (2006).
2. Alexander, G.E. Longitudinal PET Evaluation of Cerebral Metabolic Decline in Dementia: A Potential Outcome Measure in Alzheimer's Disease Treatment Studies. *American Journal of Psychiatry* **159**, 738-745 (2002).
3. Deweer, B. et al. Memory disorders in probable Alzheimer's disease: the role of hippocampal atrophy as shown with MRI. *British Medical Journal* **58**, 590 (1995).
4. Tanzi, R.E. & Bertram, L. New Frontiers in Alzheimer's Disease Genetics. *Neuron* **32**, 181-184 (2001).
5. Andreasen, N. et al. 373-379 (Am Med Assoc, 2001).
6. Cichocki, A. et al. EEG filtering based on blind source separation (BSS) for early detection of Alzheimer's disease. *Clinical Neurophysiology* **116**, 729-737 (2005).
7. Buscema, M., Rossini, P., Babiloni, C. & Grossi, E. The IFAST model, a novel parallel nonlinear EEG analysis technique, distinguishes mild cognitive impairment and Alzheimer's disease patients with high degree of accuracy. *Artificial Intelligence In Medicine* **40**, 127-141 (2007).
8. Huang, C. et al. Discrimination of Alzheimer's disease and mild cognitive impairment by equivalent EEG sources: a cross-sectional and longitudinal study. *Clinical Neurophysiology* **111**, 1961-1967 (2000).
9. Musha, T. et al. A new EEG method for estimating cortical neuronal impairment that is sensitive to early stage Alzheimer's disease. *Clinical Neurophysiology* **113**, 1052-1058 (2002).

10. Pritchard, W.S. et al. EEG-based, neural-net predictive classification of Alzheimer's disease versus control subjects is augmented by non-linear EEG measures. *Electroencephalogr Clin Neurophysiol* **91**, 118-30 (1994).
11. Woon, W.L., Cichocki, A., Vialatte, F. & Musha, T. Techniques for early detection of Alzheimer's disease using spontaneous EEG recordings. *Physiological Measurement* **28**, 335-347 (2007).
12. Acar, E., Aykut-Bingol, C., Bingol, H., Bro, R. & Yener, B. Multiway analysis of epilepsy tensors. *Bioinformatics* **23**, i10 (2007).
13. Acar, E., Bing, C.A., Bing, H. & Yener, B. in Proceedings of the 24th IASTED international conference on Biomedical engineering 317-322 (2006).
14. Lee, H., Kim, Y.D., Cichocki, A. & Choi, S. Nonnegative tensor factorization for continuous EEG classification. *International Journal of Neural Systems* **17**, 305 (2007).
15. Andersson, C.A. & Bro, R. The N-way Toolbox for MATLAB. *Chemometrics and Intelligent Laboratory Systems* **52**, 1-4 (2000).
16. Bro, R. & Kiers, H.A.L. A new efficient method for determining the number of components in PARAFAC models. *Contract* **1999**, 10377.
17. Coben, L.A., Danziger, W.L. & Berg, L. Frequency analysis of the resting awake EEG in mild senile dementia of Alzheimer type. *Electroencephalogr Clin Neurophysiol* **55**, 372-80 (1983).
18. Sloan, E.P., Fenton, G.W., Kennedy, N.S.J. & MacLennan, J.M. Electroencephalography and single photon emission computed tomography in dementia: a comparative study. *Psychological Medicine* **25**, 631 (1995).
19. Atkinson, A.C. & Riani, M. Exploratory tools for clustering multivariate data. *Computational Statistics and Data Analysis* **52**, 272-285 (2007).
20. Bro, R. PARAFAC. Tutorial and applications. *Chemometrics and Intelligent Laboratory Systems* **38**, 149-171 (1997).
21. Al Kiers, H., Ten Berge, J. & Bro, R. PARAFAC2: PART I. a direct fitting algorithm for the PARAFAC2 model. *J. Chemometrics* **13**, 275-294 (1999).
22. Kim, Y.D., Cichocki, A. & Choi, S. in IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP-2008) (IEEE, Las Vegas, Nevada, 2008).
23. Cichocki, A., Zdunek, R. & Amari, S. Nonnegative Matrix and Tensor Factorization. *Signal Processing Magazine, IEEE* **25**, 142-145 (2008).
24. Cichocki, A., Zdunek, R., Plemmons, R. & Amari, S. in ICANNGA-2007 (ed. Science, L.N.i.C.) 271-280 (Springer, Warsaw, Poland, 2007).