

Decomposing the time-frequency representation of EEG using non-negative matrix and multi-way factorization

Morten Mørup

Department of Signal Processing
Informatics and Mathematical Modeling
Technical University of Denmark
mm@imm.dtu.dk

Lars Kai Hansen

Department of Signal Processing
Informatics and Mathematical Modeling
Technical University of Denmark
lkh@imm.dtu.dk

Josef Parnas

Department of Psychiatry
Hvidovre Hospital,
University Hospital of Copenhagen
Denmark.
jpa@cfs.ku.dk

Sidse M. Arnfred

Department of Psychiatry
Hvidovre Hospital,
University Hospital of Copenhagen
Denmark.
Sidse.Arnfred@hh.hosp.dk

Corresponding author:

Morten Mørup
Department of Signal Processing
Informatics and Mathematical Modeling
Technical University of Denmark
Richard Petersens Plads, Building 321,
DK-2800 Kongens Lyngby, Denmark
Phone: +45 4525 3900
Fax: +45 4587 2599
Email: mm@imm.dtu.dk

Abstract

We demonstrate how non-negative matrix factorization (NMF) can be used to decompose the inter trial phase coherence (ITPC) of multi-channel EEG to yield a unique decomposition of time-frequency signatures present in various degrees in the recording channels. The NMF optimization is easily generalized to a parallel factor (PARAFAC) model to form a non-negative multi-way factorization (NMWF). While the NMF can examine subject specific activities the NMWF can effectively extract the most similar activities across subjects and or conditions. The methods are tested on a proprioceptive stimulus consisting of a weight change in a handheld load. While somatosensory gamma oscillations have previously only been evoked by electrical stimuli we hypothesized that a natural proprioceptive stimulus also would be able to evoke gamma oscillations. ITPC maxima were determined by visual inspection and these results were compared to the NMF and NMWF decompositions. Agreement between the results of the visual pattern inspection and the mathematical decompositions was satisfactory showing two significant coherent activities; the predicted 40Hz activity 60 ms after stimulus onset in the frontal-parietal region contralateral to stimulus side and additionally an unexpected 20Hz activity slightly lateralized in the frontal central region. Consequently, also proprioceptive stimuli are able to elicit evoked gamma activity.

1 Introduction

The analysis of EEG has developed in two major directions; one focusing on dipole or source localization through elaborate statistical models trying to solve the “inverse” electrostatics problem (Koles, Z. J., 1998); another focusing on mathematical decomposition on the data (Dormann, W. U., et al., 1987; Makeig, S., et al., 1997; Rogers, L. J., 1991). Lately there has been a growing interest in assessment of event related electroencephalographic (EEG) activity in the *time-frequency* domain (Duzel, E., et al., 2003; Gruber, T., et al., 2004; Herrmann, C. S., et al., 1999; Jansen, B. H., et al., 2004; Jones, K., et al., 2002; Lachaux, J. P., et al., 2005; Tallon-Baudry, C. and Bertrand, O., 1999). Our aim is here to extend the mathematical decompositions of the EEG to the wavelet transformed multi-channel event related EEG to yield easy interpretable time-frequency plots.

We propose to apply non-negative matrix factorization (NMF) (Lee, D. D. and Seung, H. S., 1999; 2001) to analyze the inter trial phase coherence of multi-channel wavelet transformed EEG given by *channel x time-frequency*. This NMF approach is easily adapted to a parallel factor (PARAFAC) analysis forming a non-negative multi-way factorization (NMWF). The NMWF model enables analysis of EEG data encompassing more modalities such as condition and subject without collapsing these modalities (as is the case for the present multi subject NMF analysis) giving a weighted average of the activity the most similar across subjects and conditions. The PARAFAC model has previously been used to explore the wavelet transformed event related EEG (Mørup, M., et al., 2006). It is however the first

time NMF is used to analyze the wavelet transformed EEG and the novel application of the NMWF includes the creation of time-frequency plots. The decomposition techniques presented herein are proposed to be valuable tools in multi-subject data exploration and analysis since 1. they yield easy interpretable components and 2. they can be formed to give subject specific information within the same scalp region or to capture the activity the most similar across subjects.

To demonstrate the viability of the algorithms the NMF and NMWF models are applied to a data set resulting from a stimulus, consisting of a weight change on a handheld load. As such it is a natural compound somatosensory stimulus but as the major ingredient of the stimulus is the change of applied force on a static muscle contraction, it is primarily conceived as a proprioceptive stimulus (Arnfred, S., et al., 2000; Arnfred, S. M., 2005). Nerve stimulation has been the only type of stimulus previously reported. in scalp recordings of somatosensory gamma band activity magnetoencephalographic (MEG) recordings of gamma synchronization following electric stimulation of the thumb and little finger has shown that thumb stimulation increases synchronization of higher frequencies than little finger stimulation (Tecchio, F., et al., 2003). This has been suggested to be due to more selective neural networks being activated by thumb stimulation (Tecchio, F., et al., 2003). Scalp electroencephalographic (EEG) studies of somatosensory gamma band activity (GBA(30-80Hz)) following electric stimulation have been investigated in the context of pain modulation. Early (<100ms) parietal as well as later more central and frontal (100-300ms) GBA, measured as power or phase coherence, is augmented by pain (Babiloni, C., et al., 2002; Chen, A. C. and Herrmann, C. S., 2001; De Pascalis, V. and Cacace, I., 2005; De Pascalis, V., et al., 2004). In studies of visual processing, GBA increases with perceptual binding load and GBA is suggested to be the electrophysiological manifestation of feature binding (Herrmann, C. S., et al., 2004).

Considering the perceptual binding involved in a complex somatosensory stimulus like a sudden load change, we hypothesised that the proprioceptive stimulus would elicit GBA to be recorded at the scalp above the contralateral primary somatosensory cortex. The results of the NMF and NMWF decompositions of the data from the proprioceptive stimulus are compared to results obtained by visual inspection of the data.

2 Methods

2.1 Algorithms

Traditionally the decomposition of the EEG into components has been based on decomposition techniques such as principal component analysis (PCA) (Collet, W., 1989; Dormann, W. U., et al., 1987; Kayser, J., et al., 2003; Picton, T. W., et al., 2000; Rogers, L. J., 1991) and independent component analysis (ICA) (Makeig, S., et al., 1997; Makeig, S., et al., 1999). When the EEG-data is subjected to the continuous wavelets transformation for

the analysis of frequency changes through time (Herrmann, C. S., et al., 2005), the representation is expanded from *channel x time* to a 3-way array of *channel x frequency x time*. Unfolding this three way array by collapsing the time and frequency dimensions into one dimension of *time-frequency* yields a matrix of *channel x time-frequency* that is analyzable by factor analysis models. These two-way factor analyses yield components consisting of time-frequency signatures with their mixing in the various recording channels. However, additional modalities arise when investigations are performed across subjects, trials or conditions as is commonly the case. Factor analysis models such as ICA, PCA and NMF permit the analysis of such data by further unfolding of these extra modalities. However, unfolding can to some extent hamper interpretation, but more importantly, potentially dismiss modality specific information by mixing information in a given modality with the more or less arbitrary chosen modalities that it has been folded with in a two-way matrix analysis. Furthermore, in this form of analysis the activity that is the most similar across trials or subjects is often the most interesting to access. Consequently, rather than just unfolding these multi-way arrays into matrices we also analyzed this form of data using the multi-way model PARAFAC given in figure 1.

The PARAFAC model, here detailed as the NMWF, creates a weighted average of the most common activity revealing the degree to which this activity is present in the various subjects, trials or conditions.

Two-way analyses, e.g. the NMF, create regional specific components revealing how these differ in time-frequency pattern in the various subjects, trials or conditions.

Consequently, we here use NMWF but also NMF for group analysis. Although the subject specific activity is believed to deviate from the overall mean (as assumed by NMWF) and the placement of the electrodes are not identical across subjects (as assumed by NMF), the decompositions enable an easy method to compare and view the activity across subjects. Furthermore, the activity present in a different subject sample is likely to be captured by the models since the model captures the activity common across the analyzed subjects. However, the caveat must be that the components might be biased by a few subjects having relatively strong signal strength. It follows that any conclusions reached based on the group decompositions has to be validated by single subject analyses. This problem of group analyses is no different from the problems of group averages when analyzing grand averages of the evoked potentials.

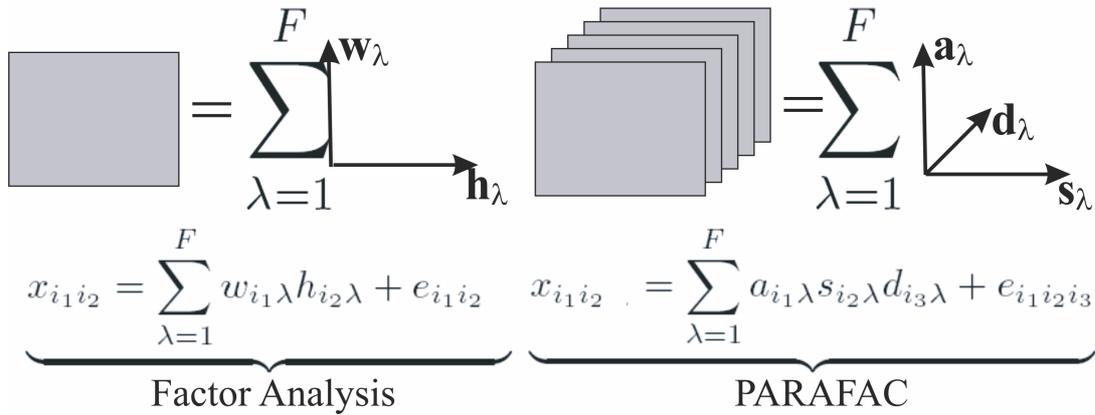


Figure 1: Graphical representation of the factor analysis to the left and the PARAFAC decomposition of a 3-way array to the right. Like the factor analysis, PARAFAC decomposes the data into factor effects pertaining to each modality. F denotes the number of factors.

NMF nor ICA have to our knowledge been used previously to analyze the frequency transformed event related EEG. The use of PARAFAC for analysis of wavelet transformed EEG has recently been advanced. In 2004 Miwakeichi and colleagues (Miwakeichi, F., et al., 2004) suggested the use of PARAFAC on the wavelet transformed ongoing EEG of *channel \times frequency \times time* and the PARAFAC model can be an efficient data explorative tool for the event related wavelet transformed EEG (Mørup, M., et al., 2006). Recently, the PARAFAC model has also been advanced to the analysis of fMRI (Andersen, A. H. and Rayens, W. S., 2004, Beckmann, 2005 #57)

The non-negative matrix factorization (NMF) was introduced by Lee and Seung in 1999 (Lee, D. D. and Seung, H. S., 1999; 2001). They showed how the NMF decomposition gives a more sparse decomposition than PCA and conventional ICA yielding more interpretable components. The NMF has the advantage over PCA and ICA that given the data spans the complete positive octant, no rotation of the factor solutions is possible ensuring uniqueness apart from trivial scaling and permutation, see also (Donoho, D. and Stodden, V., 2004). Although algorithms for positive ICA exist (Højen-Sørensen, P. A. d. F. R., et al., 2002; Oja, E. and Plumbley, M., 2004) these were not considered in the present work as they would give results similar to NMF while being very time consuming.

The parallel factor (PARAFAC) model was independently proposed by Harshman (Harshman, R. A., 1970) and by Carroll and Chang (Carroll, J. D. and Chang, J., 1970), the latter naming it Canonical Decomposition (CANDECOMP). The model is a parsimonious extension of factor analysis to higher dimensional arrays as revealed on Figure 1.

PARAFAC, in contrast to conventional factor models, does not suffer from rotational indeterminacy. As a result, the PARAFAC model is in general unique, apart from scaling and permutation indeterminacies (Kruskal, J. B., 1977; Sidiropoulos, N. D. and Bro, R., 2000). Consequently, the main advantage of PARAFAC over factor analysis models such as PCA, ICA and NMF is that uniqueness is ensured under very mild conditions making it unnecessary to impose constraints in the form of orthogonality, statistical independence / sparsity or requiring the data to span the complete positive octant. The “price paid” for such strong uniqueness is a more restrictive model, that can only capture the activity that is the most similar across trials, subjects and/or conditions. Consider a three way array of size $I \cdot J \cdot K$. A F component PARAFAC model would have $(I+J+K) \cdot F$ free parameters whereas the corresponding unfolded factor analysis model would include $(I+J \cdot K) \cdot F \gg (I+J+K) \cdot F$ free parameters.

Lee and Seung gave two algorithms for NMF both based on gradient descent; one minimizing the squared error the other minimizing the Kullback-Leibler divergence (Lee, D. D. and Seung, H. S., 2001). Both algorithms are easily adapted to the PARAFAC model with non-negativity on all modalities, giving a non negative multi-way factorization NMWF, see also (Hazan, T., et al., 2005; Shashua, A. and Hazan, T., 2005). In the following, the matrices $\mathbf{W}, \mathbf{H}, \mathbf{A}, \mathbf{S}$ and \mathbf{D} will be defined as given in Figure 1.

In the factor analysis we have:

$$\mathbf{X} = \mathbf{W}\mathbf{H}^T + \mathbf{E} \quad \text{or equivalently} \quad \mathbf{X}^T = \mathbf{H}\mathbf{W}^T + \mathbf{E}^T \quad (1)$$

Where \mathbf{E} is the approximation error.

The PARAFAC model can be written in matrix notation by use of the Khatri-Rao product, i.e. $\mathbf{A}|\otimes|\mathbf{S} = [\mathbf{a}_1 \otimes \mathbf{s}_1 \quad \mathbf{a}_2 \otimes \mathbf{s}_2 \quad \dots \quad \mathbf{a}_F \otimes \mathbf{s}_F]$ where F is the number of factors and the n -mode matricizing of the multiway array $\mathcal{X}^{I_1 \times I_2 \times \dots \times I_N}$, i.e. $\mathbf{X}_{(n)} = \mathbf{X}^{I_n \times I_1 I_2 \dots I_{n-1} I_{n+1} \dots I_N}$. This gives the equivalent expressions:

$$\begin{aligned} \mathbf{X}_{(1)} &= \mathbf{A}\mathbf{Z}^{(1)T} + \mathbf{E}_{(1)} \text{ where } \mathbf{Z}^{(1)} = \mathbf{S}|\otimes|\mathbf{D} \\ \mathbf{X}_{(2)} &= \mathbf{S}\mathbf{Z}^{(2)T} + \mathbf{E}_{(2)} \text{ where } \mathbf{Z}^{(2)} = \mathbf{D}|\otimes|\mathbf{A} \\ \mathbf{X}_{(3)} &= \mathbf{D}\mathbf{Z}^{(3)T} + \mathbf{E}_{(3)} \text{ where } \mathbf{Z}^{(3)} = \mathbf{S}|\otimes|\mathbf{A} \end{aligned} \quad (2)$$

In general the PARAFAC model for higher orders than three can be expressed as

$$\begin{aligned} x_{i_1 i_2 \dots i_N} &= \sum_{\lambda=1}^F a_{i_1 \lambda}^{(1)} a_{i_2 \lambda}^{(2)} \dots a_{i_N \lambda}^{(N)} \\ \mathbf{X}_{(n)} &= \mathbf{A}^{(n)} \mathbf{Z}^{(n)T} + \mathbf{E}_{(n)} \text{ where } \mathbf{Z}^{(n)} = \mathbf{A}^{(N)}|\otimes|\mathbf{A}^{(N-1)}|\otimes|\dots|\otimes|\mathbf{A}^{(n+1)}|\otimes|\mathbf{A}^{(n-1)}|\otimes|\dots|\otimes|\mathbf{A}^{(1)} \end{aligned} \quad (3)$$

Notice, in the two-way case the matricizing corresponds to taking the transpose and the formulation in 3 becomes equivalent to the regular factor analysis with $\mathbf{A}^{(1)} = \mathbf{W}, \mathbf{Z}^{(1)} = \mathbf{H}$ and $\mathbf{A}^{(2)} = \mathbf{H}^T, \mathbf{Z}^{(2)} = \mathbf{W}$. In the following we will require $\mathbf{X}, \mathbf{W}, \mathbf{H}, \mathbf{A}, \mathbf{S}$ and \mathbf{Z} to be non-negative.

2.1.1 NMWF based on Least-Squares

Consider the least square cost function C given by:

$$C = \|\mathbf{X} - \mathbf{W}\mathbf{H}^T\|^2 = \|\mathbf{X}^T - \mathbf{W}\mathbf{H}^T\|^2 = \sum_i \sum_j (x_{ij} - (\mathbf{W}\mathbf{H}^T)_{ij})^2 \quad (4)$$

Minimizing C corresponds to maximizing the likelihood of a homoscedatic Gaussian noise model.

Lee and Seung found the following convergent updates for \mathbf{W} and \mathbf{H} , by differentiating C with respect to each element in \mathbf{W} and \mathbf{H} and updating with a gradient based search using a stepsize resulting in multiplicative updates, see (Lee, D. D. and Seung, H. S., 2001) for details:

$$\mathbf{W}_{i\lambda} \leftarrow \mathbf{W}_{i\lambda} \frac{(\mathbf{X}\mathbf{H})_{i\lambda}}{(\mathbf{W}\mathbf{H}^T\mathbf{H})_{i\lambda}}, \quad \mathbf{H}_{j\lambda} \leftarrow \mathbf{H}_{j\lambda} \frac{(\mathbf{X}^T\mathbf{W})_{j\lambda}}{(\mathbf{H}\mathbf{W}^T\mathbf{W})_{j\lambda}} \quad (5)$$

The positivity constraint on \mathbf{W} and \mathbf{H} is insured since these multiplicative updates are bound to remain positive granted \mathbf{X}, \mathbf{W} and \mathbf{H} are positive.

Due to equation 2 the PARAFAC model can be stated as three equivalent least square minimizations giving the following three equivalent cost function expressions for C :

$$C = \|\mathbf{X}_{(1)} - \mathbf{A}\mathbf{Z}^{(1)T}\|^2 = \|\mathbf{X}_{(2)} - \mathbf{S}\mathbf{Z}^{(2)T}\|^2 = \|\mathbf{X}_{(3)} - \mathbf{D}\mathbf{Z}^{(3)T}\|^2 \quad (6)$$

In general following the formulation of equation 3 the cost function for the least square minimization for higher orders can be stated as the equivalent problems:

$$C = \|\mathbf{X}_{(1)} - \mathbf{A}^{(1)}\mathbf{Z}^{(1)T}\|^2 = \|\mathbf{X}_{(2)} - \mathbf{A}^{(2)}\mathbf{Z}^{(2)T}\|^2 = \dots = \|\mathbf{X}_{(n)} - \mathbf{A}^{(n)}\mathbf{Z}^{(n)T}\|^2 \quad (7)$$

As the minimization of the expressions in equation 6 and 7 corresponds to the regular factor analysis the update of each factor is given directly by the NMF updates by interchanging the roles of \mathbf{X}, \mathbf{W} and \mathbf{H} in 5 with that of $\mathbf{X}_{(n)}, \mathbf{A}^{(n)}$ and $\mathbf{Z}^{(n)}$. Consequently each $\mathbf{A}^{(n)}$ is updated according to:

$$\mathbf{A}_{i_n\lambda}^{(n)} \leftarrow \mathbf{A}_{i_n\lambda}^{(n)} \frac{(\mathbf{X}_{(n)}\mathbf{Z}^{(n)})_{i_n\lambda}}{(\mathbf{A}^{(n)}\mathbf{Z}^{(n)T}\mathbf{Z}^{(n)})_{i_n\lambda}} \quad (9)$$

The convergence of the updates follows straight forward from the convergence of regular NMF simply by interchanging the roles of \mathbf{W} and \mathbf{H} with that of $\mathbf{A}^{(n)}$ and $\mathbf{Z}^{(n)}$ in the proof given by Lee and Seung (Lee, D. D. and Seung, H. S., 2001).

Specifically for the three-way PARAFAC model given in figure 1, we get the convergent updates:

$$\mathbf{A}_{1,\lambda} \leftarrow \mathbf{A}_{1,\lambda} \frac{(\mathbf{X}_{(1)}\mathbf{Z}^{(1)})_{i_1\lambda}}{(\mathbf{AZ}^{(1)T}\mathbf{Z}^{(1)})_{i_1\lambda}}, \mathbf{S}_{i_2\lambda} \leftarrow \mathbf{S}_{i_2\lambda} \frac{(\mathbf{X}_{(2)}\mathbf{Z}^{(2)})_{i_2\lambda}}{(\mathbf{SZ}^{(2)T}\mathbf{Z}^{(2)})_{i_2\lambda}}, \mathbf{D}_{i_3\lambda} \leftarrow \mathbf{D}_{i_3\lambda} \frac{(\mathbf{X}_{(3)}\mathbf{Z}^{(3)})_{i_3\lambda}}{(\mathbf{DZ}^{(3)T}\mathbf{Z}^{(3)})_{i_3\lambda}} \quad (8)$$

2.1.2 NMWF based on the divergence approach

Consider the Kullback-Leibler divergence cost function in the Factor Analysis as defined by (Lee, D. D. and Seung, H. S., 2001), i.e.:

$$D(\mathbf{X} \parallel \mathbf{AS}^T) = D(\mathbf{X}^T \parallel \mathbf{SA}^T) = \sum_i^I \sum_j^J \mathbf{X}_{ij} \log \frac{\mathbf{X}_{ij}}{(\mathbf{AS}^T)_{ij}} - \mathbf{X}_{ij} + (\mathbf{AS}^T)_{ij} \quad (10)$$

Using a gradient based search to minimize this divergence with a stepsize yielding multiplicative updates, the following updates of \mathbf{W} and \mathbf{H} forming the NMF-KL algorithm is achieved (Lee, D. D. and Seung, H. S., 2001):

$$\mathbf{W}_{i\lambda} \leftarrow \mathbf{W}_{i\lambda} \frac{\sum_{j=1}^J \mathbf{H}_{j\lambda} \frac{\mathbf{X}_{ij}}{(\mathbf{WH})_{ij}}}{\sum_{j=1}^J \mathbf{H}_{j\lambda}}, \quad \mathbf{H}_{j\lambda} \leftarrow \mathbf{H}_{j\lambda} \frac{\sum_{i=1}^I \mathbf{W}_{i\lambda} \frac{\mathbf{X}_{ij}}{(\mathbf{WH})_{ij}}}{\sum_{i=1}^I \mathbf{W}_{i\lambda}} \quad (11)$$

For the PARAFAC model given in figure 1, the divergence cost function can be stated as the following three equivalent expressions:

$$D(\mathbf{X}_{(1)} \parallel \mathbf{AZ}^{(1)T}) = D(\mathbf{X}_{(2)} \parallel \mathbf{SZ}^{(2)T}) = D(\mathbf{X}_{(3)} \parallel \mathbf{DZ}^{(3)T}) \quad (12)$$

And for higher orders than three as the equivalent problems:

$$D(\mathbf{X}_{(1)} \parallel \mathbf{A}^{(1)}\mathbf{Z}^{(1)T}) = D(\mathbf{X}_{(2)} \parallel \mathbf{A}^{(2)}\mathbf{Z}^{(2)T}) = \dots = D(\mathbf{X}_{(n)} \parallel \mathbf{A}^{(n)}\mathbf{Z}^{(n)T}) \quad (13)$$

As the minimization of the expression in equation 13 again corresponds to the regular factor analysis, the update of each factor is given directly by the NMF-KL updates as:

$$\mathbf{A}_{i_n\lambda}^{(n)} \leftarrow \mathbf{A}_{i_n\lambda}^{(n)} \frac{\sum_{j=1}^{I_1 I_2 \dots I_{n-1} I_{n+1} \dots I_N} \mathbf{Z}_{j\lambda}^{(n)} \frac{\mathbf{X}_{(n) i_n j}}{\left(\mathbf{A}^{(n)} \mathbf{Z}^{(n)T}\right)_{i_n j}}}{\sum_{j=1}^{I_1 I_2 \dots I_{n-1} I_{n+1} \dots I_N} \mathbf{Z}_{j\lambda}^{(n)}} \quad (14)$$

Again, the convergence of these updates follow straight forward from the convergence of the regular NMF-KL updates by interchanging the roles of \mathbf{W} and \mathbf{H} with that of $\mathbf{A}^{(n)}$ and $\mathbf{Z}^{(n)}$ in the proof given by Lee and Seung (Lee, D. D. and Seung, H. S., 2001).

Specifically for the three-way PARAFAC model given in figure 1 the following updates are achieved:

$$\mathbf{A}_{i_1\lambda} \leftarrow \mathbf{A}_{i_1\lambda} \frac{\sum_{j=1}^{I_2 I_3} \mathbf{Z}_{j\lambda}^{(1)} \frac{\mathbf{X}_{(1) i_1 j}}{\left(\mathbf{A}\mathbf{Z}^{(1)T}\right)_{i_1 j}}}{\sum_{j=1}^{I_2 I_3} \mathbf{Z}_{j\lambda}^{(1)}}, \mathbf{S}_{i_2\lambda} \leftarrow \mathbf{S}_{i_2\lambda} \frac{\sum_{j=1}^{I_1 I_3} \mathbf{Z}_{j\lambda}^{(2)} \frac{\mathbf{X}_{(2) i_2 j}}{\left(\mathbf{S}\mathbf{Z}^{(2)T}\right)_{i_2 j}}}{\sum_{j=1}^{I_1 I_3} \mathbf{Z}_{j\lambda}^{(2)}}, \mathbf{D}_{i_3\lambda} \leftarrow \mathbf{D}_{i_3\lambda} \frac{\sum_{j=1}^{I_1 I_2} \mathbf{Z}_{j\lambda}^{(3)} \frac{\mathbf{X}_{(3) i_3 j}}{\left(\mathbf{D}\mathbf{Z}^{(3)T}\right)_{i_3 j}}}{\sum_{j=1}^{I_1 I_2} \mathbf{Z}_{j\lambda}^{(3)}} \quad (15)$$

Model	
$x_{i_1 i_2 \dots i_N} = \sum_{\lambda=1}^F a_{i_1 \lambda}^{(1)} a_{i_2 \lambda}^{(2)} \dots a_{i_N \lambda}^{(N)} + e_{i_1 i_2 \dots i_N}$	
<p style="text-align: center;">Algorithm for NMWF-LS</p> <p>Initialize all $\mathbf{A}^{(n)}$ randomly</p> $C_{new} = \left\ \mathbf{X}_{(N)} - \mathbf{A}^{(N)} \mathbf{Z}^{(N)T} \right\ ^2$ <p>While $\frac{C_{old} - C_{new}}{C_{new}} > \delta$ do</p> <p style="padding-left: 20px;">$C_{old} = C_{new}$</p> <div style="border: 1px solid black; padding: 5px; margin: 5px 0;"> <p style="text-align: center;">For all n do</p> $\mathbf{A}_{i_n \lambda}^{(n)} \leftarrow \mathbf{A}_{i_n \lambda}^{(n)} \frac{(\mathbf{X}_{(n)} \mathbf{Z}^{(n)})_{i_n \lambda}}{(\mathbf{A}^{(n)} \mathbf{Z}^{(n)T} \mathbf{Z}^{(n)})_{i_n \lambda} + \epsilon}$ </div> $C_{new} = \left\ \mathbf{X}_{(N)} - \mathbf{A}^{(N)} \mathbf{Z}^{(N)T} \right\ ^2$	<p style="text-align: center;">Algorithm for NMWF-KL</p> <p>Initialize all $\mathbf{A}^{(n)}$ randomly</p> $C_{new} = D(\mathbf{X}_{(N)} \parallel \mathbf{A}^{(N)} \mathbf{Z}^{(N)T})$ <p>While $\frac{C_{old} - C_{new}}{C_{new}} > \delta$ do</p> <p style="padding-left: 20px;">$C_{old} = C_{new}$</p> <div style="border: 1px solid black; padding: 5px; margin: 5px 0;"> <p style="text-align: center;">For all n do</p> $\mathbf{A}_{i_n \lambda}^{(n)} \leftarrow \mathbf{A}_{i_n \lambda}^{(n)} \frac{\sum_{j=1}^{I_1 I_2 \dots I_{n-1} I_{n+1} \dots I_N} \mathbf{Z}_{j \lambda}^{(n)} \frac{\mathbf{X}_{(n) i_n j}}{(\mathbf{A}_{i_n \lambda}^{(n)} \mathbf{Z}^{(n)T})_{i_n j} + \epsilon}}{\sum_{j=1}^{I_1 I_2 \dots I_{n-1} I_{n+1} \dots I_N} \mathbf{Z}_{j \lambda}^{(n)} + \epsilon}$ </div> $C_{new} = D(\mathbf{X}_{(N)} \parallel \mathbf{A}^{(N)} \mathbf{Z}^{(N)T})$
<p style="text-align: center;">$\mathbf{Z}^{(n)} = \mathbf{A}^{(N)} \otimes \mathbf{A}^{(N-1)} \otimes \dots \otimes \mathbf{A}^{(n+1)} \otimes \mathbf{A}^{(n-1)} \otimes \dots \otimes \mathbf{A}^{(1)}$</p> <p style="text-align: center;">Notice: $\mathbf{X}_{(n)} = \mathbf{X}^{I_n \times I_1 I_2 \dots I_{n-1} I_{n+1} \dots I_N}$</p> <p style="text-align: center;">$\mathbf{X}_{(n)} = \mathbf{A}^{(n)} \mathbf{Z}^{(n)T} + \mathbf{E}_{(n)}$</p>	

The algorithms for NMWF based on least square (LS) and Kulbach-Leibler (KL) divergence minimization. δ was set to 10^{-6} while $\epsilon=10^{-9}$ ensured no division by zero for numerical stability. Notice how the regular NMF algorithm is the special case of the NMWF algorithm given by $x_{i_1 i_2} = \sum_{\lambda=1}^F a_{i_1 \lambda}^{(1)} a_{i_2 \lambda}^{(2)} + e_{i_1 i_2}$.

2.2 The inter trial phase coherence

The inter trial phase coherence (ITPC) is a measure of phase consistency through trials of the continuous wavelet transformed EEG-data. The complex wavelet transform projects the EEG-data onto the complex plane. Define the vector strength as the length of the vector given by the sum of n unit vectors in the complex plane. Then the vector strength measures coherence, i.e. the degree in which the vectors point in the same direction, see also figure 2. The ITPC is a statistical measure of the evoked activity given as the vector strength, i.e. coherence over epochs.

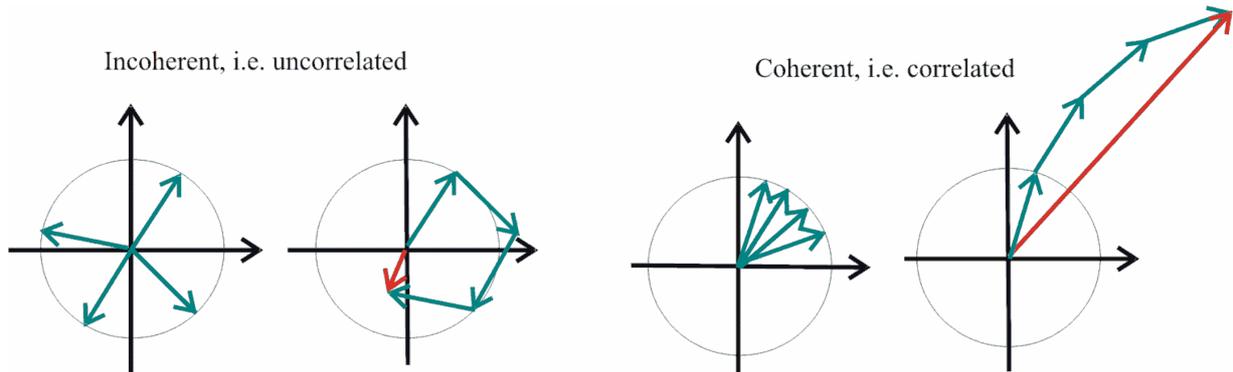


Figure 2: The vector strength is the sum over unit vectors in the complex plane. In regions where the vectors are uncorrelated, i.e. incoherent the vector strength is small compared to region where the vectors are correlated, i.e. coherent yielding a much larger vector strength (compare red vector to the left with red vector to the right). Consequently, the vector strength is a measure of coherence.

Let $X_e(c, f, t)$ be the coefficient of the wavelet transform at channel c at frequency f and time t for epoch e , and let there be a total of n epochs. The ITPC is given by (Delorme, A. and Makeig, S., 2004):

$$ITPC(c, f, t) = \frac{1}{n} \left| \sum_{e=1}^n \frac{X_e(c, f, t)}{|X_e(c, f, t)|} \right| \quad (16)$$

While an area of coherence is approximately normally distributed, random activity/noise is Raleigh distributed (Palva, J. M., et al., 2005) with an average value of approximately $n^{-1/2}$, see Appendix for details. Compared to other measures of coherence such as the avWT (Herrmann, C. S., et al., 2005) the ITPC has two major advantages. 1) Since the statistical properties of random ITPC activity is known the significance of the ITPC activity is easily accessed, see also appendix. 2) Since all epochs are weighted the same the effect of even very noisy trials is limited making the cumbersome work of artifact rejection unnecessary. However, since the ITPC is a signal average over trials it is biased towards the averaged phase. Consequently, some event related brain activities might not be fully explained by phase changes and thereby not be captured by the ITPC. Furthermore, as the ITPC is a statistical measure of phase consistency this makes interpretation in terms of source localization difficult: the propagation factors are largely unknown as the amplitude information is removed in the ITPC measure.

As random ITPC is Raleigh distributed, the significance of the ITPC can be compared to a null hypothesis of random ITPC. The following formula (see appendix for details) can be derived to access the significance of a given ITPC value, x . In the null hypothesis of the ITPC being randomly generated the maximum of N independent ITPC points has probability α of taking a value exceeding x , given by:

$$x = \sqrt{-2\sigma^2 \ln(1 - (1 - \alpha)^{\frac{1}{N}})} \quad (17)$$

2.3 Experimental details

Fourteen healthy subjects (four females) were included after informed consent as approved by the Ethics Committee. They were paid to participate in the experiment. The mean age of the sample was 24.4 years (standard deviation (STD) 3.0) and the mean length of education was 15 years (STD 2.3). The proprioceptive stimulus was delivered by a custom build apparatus (Sv. Christoffersen, Department of Medical Physiology, University of Copenhagen): The subject has a plastic handle in his pronated hand and a minimum static load of 400g was applied through a nylon wire connected to the handle and an electromagnetic servomotor driving a spool. An additional load of 100g was applied with a linear increment of 20g/10ms. The maximum load was sustained for 500ms. The hand was supported by a horizontal cushioned armrest to the level of the metacarpophalangeal articulation of the thumb. A schematic of the set-up is shown in (Arnfred, S., et al., 2000). Stimulus delivery was controlled by the Presentation© software. Alternating between hands, three runs were recorded in both sides. Each run lasted four minutes and consisted of 120 stimuli applied to the same hand with inter stimulus intervals of 1.5s resulting in a total of n=360 epochs. While recording, a monitor showing a fixation cross was placed 50cm in front of the comfortably seated subject, and 76dB masking white noise was delivered through loudspeakers just behind the monitor. The subject was asked to relax and fixate on the monitor and no attempt was made to direct his attention towards the proprioceptive stimuli.

2.3.1 Preprocessing

EEG data was recorded with 64 scalp electrodes (BioSemi Active electrodes system) arranged according to the International 10-10 system. Additional recordings were obtained from earlobes and at the maxillae beneath each eye. The grounding electrodes for the active electrodes (CMS and DRL) were placed centrally, close to POz. Data was recorded continuously at 2048 Hz/channel, band pass 0.1-160 Hz, by a LabView© application (ActivView©) on a Windows© based PC. Off-line processing was performed in EEGLAB for MatLab© (Delorme, A. and Makeig, S., 2004). The data was referenced to digitally linked earlobes and cut into epochs (-250 to +500ms). The data was wavelet transformed using a complex Morlet wavelet (Herrmann, C. S., et al., 1999; Miwakeichi, F., et al., 2004)

with center frequency 1 and bandwidth parameter 2, i.e. $\tilde{\varphi}(t) = \frac{1}{\sqrt{2\pi}} \exp(i2\pi t) \exp\left(-\frac{t^2}{2}\right)$ with

frequencies represented from 15 to 75 Hz with 1 Hz interval. Baseline subtraction was not performed prior to wavelet transformation since the wavelet transform is shift invariant. Since even very noisy epochs might include relevant phase information while having relative little impact on the overall ITPC, no epochs were rejected. This enabled the ITPC to be calculated as an average across all trials, improving signal to noise ratio (SNR). Furthermore, to avoid

reduction of SNR, the data were not normalized across subjects. Normalizing would increase the influence of subjects having less coherence compared to random activity in the analysis.

2.3.2 NMF and NMWF decompositions

Let $\mathcal{X}^{c \times f \times t \times p \times k}$ denote the multi-way array of ITPC activity given by the modalities channels (c), frequency (f), time (t), subjects (p) and conditions (k). Three types of arrays are then analyzed:

- A single subject analysis of *channel x time-frequency* ITPC matrix, i.e. $\mathbf{X}^{c \times f \times t}$.
- A multi-subject analysis of *channel x time – frequency – subject – condition* ITPC matrix, i.e. $\mathbf{X}^{c \times f \times t \times p \times k}$.
- A multi-subject analysis of the 3-way array of *channel x time-frequency x subject - condition* ITPC, i.e. $\mathcal{X}^{c \times f \times t \times p \times k}$.

Decomposing the ITPC given by the matrix $\mathbf{X}^{c \times f \times t}$ of *channel x time-frequency*, i.e.

$$x_{i_1 i_2} = \sum_{\lambda=1}^F a_{i_1 \lambda} s_{i_2 \lambda} + e_{i_1 i_2} \quad (17)$$

corresponds to the assumption that the underlying factors consist of a given time frequency signature \mathbf{s}_λ that has been mixed in the channels by \mathbf{a}_λ . Decomposing the ITPC given by the matrix $\mathbf{X}^{c \times f \times t \times p \times k}$ of *channel x time – frequency – subject – condition* assumes the activity are centered around the same channels but might deviate in onset and frequency through the subjects and conditions.

Since the ITPC by nature is non-negative the decompositions of the ITPC can be based on non-negativity constraints, i.e. $a_{i_1 \lambda}, s_{i_2 \lambda} \geq 0$. Consequently, the non-negative ITPC signatures \mathbf{s}_λ can only be additively mixed in the channels. This is based on the assumption that the coherent activity measured at the scalp stem from the same underlying coherent activities in the brain recorded with varying strength depending on the electrode position. Furthermore, none of these coherent activities measured by the ITPC is allowed to cancel each other. This requires the coherent activities to be separated in either the channel or time-frequency domain. Since the Morlet wavelet transformation is overcomplete and granted the bandwidth of the wavelet is relatively small (here set to 2) the various coherent activities are likely to be separated when lifted to the time-frequency domain.

Restricting the multi-subject analysis to a 3-way array $\mathcal{X}^{c \times f \times t \times p \times k}$ of *channel x time-frequency x subject-condition*, i.e. into the PARAFAC model:

$$x_{i_1 i_2 i_3} = \sum_{\lambda=1}^F a_{i_1 \lambda} s_{i_2 \lambda} d_{i_3 \lambda} + e_{i_1 i_2 i_3} \quad (18)$$

corresponds to the additional assumption that the underlying factors are identical between each subject/condition but present to variable degree given by the score \mathbf{d}_λ , where $d_{i,\lambda} \geq 0$.

Granted NMF captures all “systematic” variation, and since the ITPC is approximately normally distributed in regions of coherence (see appendix 1), the error can be considered normally distributed. Therefore, least square estimation corresponding to maximizing the likelihood of a homoscedastic Gaussian noise model, i.e. the NMF-LS and NMWF-LS algorithms is justified. Appealing to KL minimization is equivalent to assuming a multinomial noise model. Here the residuals are weighted by the relative size of the component. This form of analysis was mainly performed for comparison. Consequently, if the decompositions were too algorithm-dependent this was an indication of unviable results.

The number of factors accepted as the best solution was purely based on visual inspection of the results. It is customary to assess the number of factors in matrix analyses through methods such as Bayesian Information Criterion (Hansen, L. K., et al., 2001), cross-validation and analysis of residuals. Since no factor had to be orthogonal or independent to the remaining factors as is the case for PCA and ICA, respectively, the choice of number of factors used in NMF had little impact on the components found. We judged the amount of components to include by their relative norms and how localized they were. A small norm greatly spread in the channels and time-frequency domain was taken as an indication that too many components were included hence the component was modeling background activity. Each of the decompositions were performed three times and compared to ensure no local optimum was found. The NMF solution is presently compared to an ICA solution based on maximum likelihood as described in (Bell, A. and Sejnowski, T. J., 1995; Hansen, L. K., et al., 2001).

The PARAFAC model is known to suffer from degeneracy and slow convergence (Beckmann, C. F. and Smith, S. M., 2005; Paatero, P., 2000). These problems are, however, circumvented when imposing non-negativity constraints on all modalities (Mørup, M., et al., 2006). While an algorithm for the estimation of PARAFAC under non-negativity constraint has been proposed by (Bro, R., 1998; Bro, R. and Jong, S. D., 1997) the NMWF-LS algorithm yielded equivalent results, but both the LS and KL algorithm for NMWF is easier to implement and to our knowledge also faster in most situations.

The NMF algorithms of Lee and Seung are known to suffer from slow convergence. Consequently, we accelerated the algorithm as devised by (Salakhutdinov, R. and Sam, R., 2003). The NMF decomposition is not unique in general. As mentioned earlier rotational ambiguity is only removed when the data spans the complete positive octant. In order to achieve this, background ITPC activity was removed by subtracting the random coherence

of 0.0465 estimated by bootstrapping ($n=360$). Any values below zero after this subtraction were set to zero. Consequently, all the decompositions shown in the results section have to have this subtracted value added in order to reflect the actual ITPC values. To access the uniqueness properties of the decomposition we analyzed the correlation between signatures of several NMF analysis of the same data, see also table 1.

3 Results and discussion

3.1 Single subject analysis

As seen on figure 3 and 4 the solutions of the ICA, NMF-LS and NMF-KL all include a coherent contralateral parietal-frontal activity and a coherent frontal central activity of lower frequency. Whereas the NMF methods give easily interpretable representations of the activities, the ICA method yields similar results. However, in order to achieve independence regions of the time-frequency signatures have become negative.

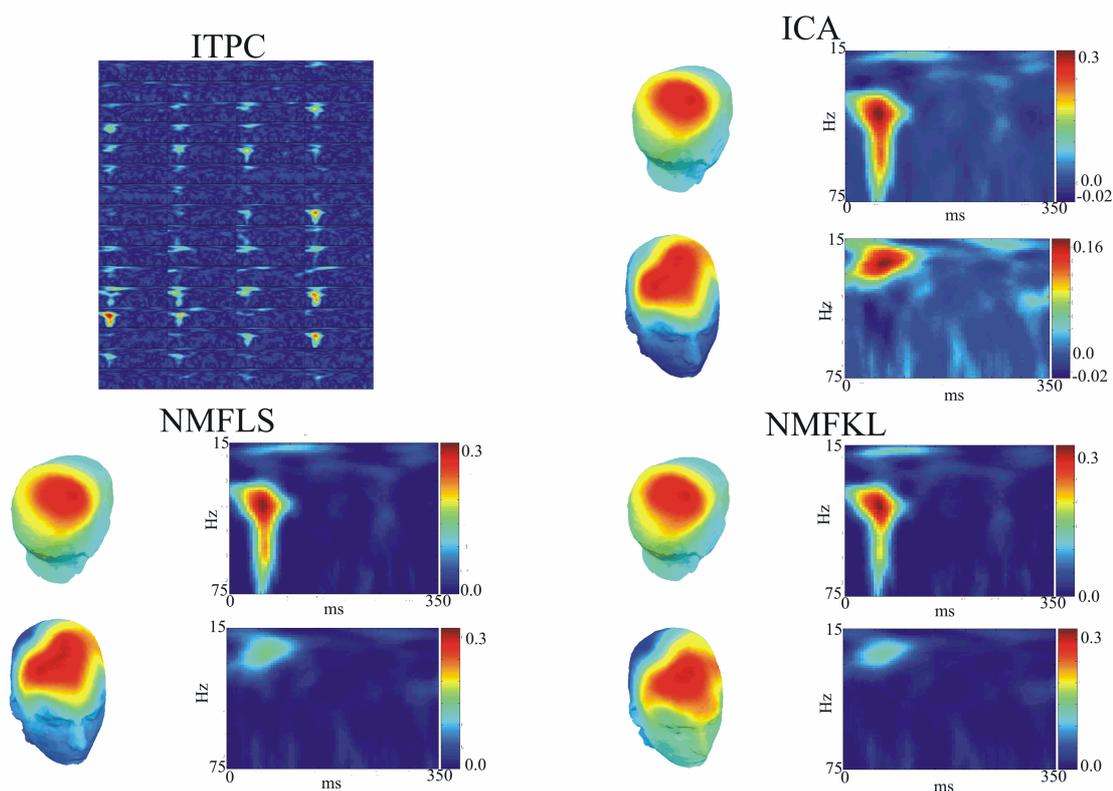


Figure 3: ITPC analysis of one subject (No 2) during left hand stimuli. To the top left is the analyzed ITPC array of channel x time-frequency given by a 16x4 array of a given

channels time-frequency signature of the ITPC. Top right; analysis of the ITPC using independent component analysis. Bottom left; the result of a NMF-LS analysis of the ITPC. Bottom right; the result of a NMF-KL analysis of the ITPC. All three decompositions yield similar results. However, the ICA analysis yields negative results in order to achieve independence which is not physiologically justified. All methods find a strong activity around 40 Hz 50 ms in the right parietal region and a more frontal activity around 20 Hz 70 ms. No apparent difference between the two NMF solutions is observed. Whereas the ICA model explains 72.75 % of the variance the NMFLS and NMFKL analysis explain respectively 72.65 and 69.82 % of the variance. The color axis of the head plots goes from 0 to 1, see also figure 5.

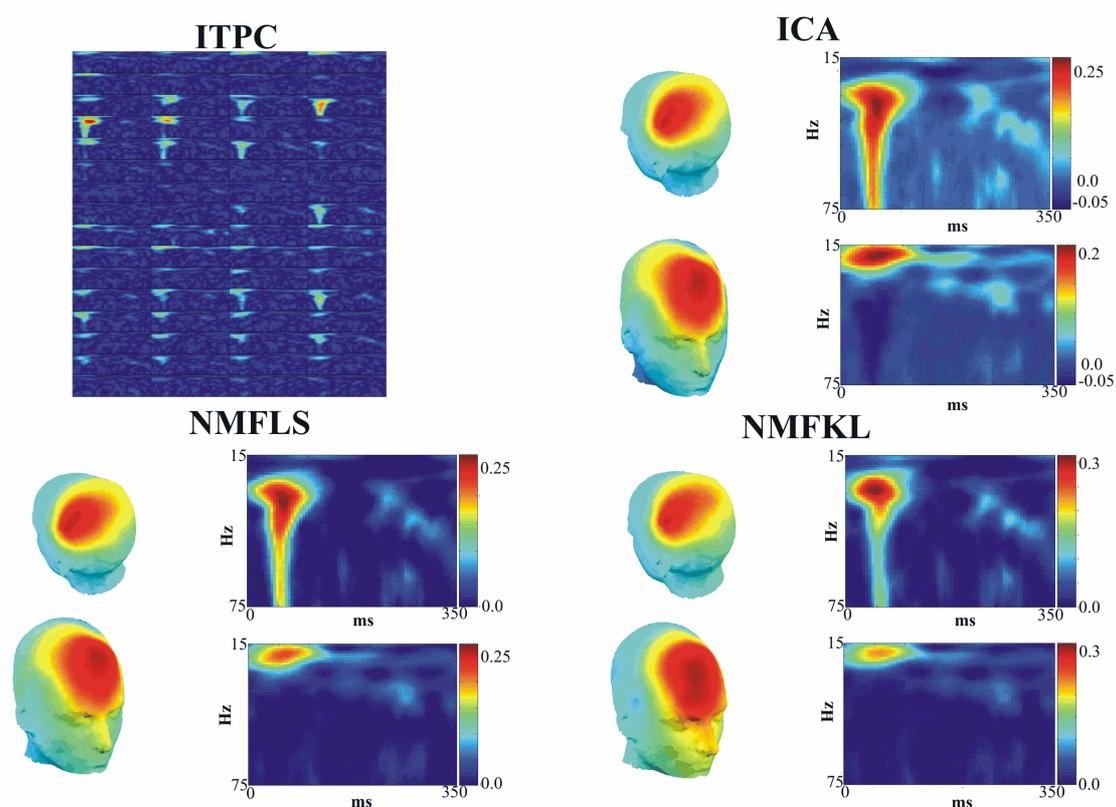


Figure 4: The ITPC of the same subject as figure 4 (No 2) during right hand stimulation. See legends of figure 4 for explanatory details. Again all three methods yield similar results, however in order to achieve independence the ICA method has forced large regions of both components to be negative in order to achieve independence. While the first components in all the analysis reveal a parietal left activity around 40 Hz 60 ms the second component pertains to a lower frequency frontal-central activity around 20 Hz 70 ms.

Whereas the ICA model explains 71.00 % the NMF-LS explains 70.58 % and NMF-KL 67.14 % of the variance. The color axis of the head plots goes from 0 to 1, see also figure 5.

To investigate whether the NMF solutions found are unique we evaluated the NMF analysis of both NMF-LS and NMF-KL on the data. The degree of consistency between the scalp and time-frequency signatures of both NMF decompositions are generally good but becomes perfect when removing the background activity, see table 1. This stems from the fact that removing background activity makes it much more likely that the analyzed data spans the complete octant. Since it is possible to achieve uniqueness without constraints of independence, NMF was superior in the analysis of the ITPC data. Had uniqueness not been achieved, sparseness constraints could have been imposed on the NMF decomposition as proposed by (Hoyer, P. O., 2002) (Eggert, J. and Kömer, E., 2004).

		No background subtraction		Background Subtracted	
		LS2	RS2	LS2	RS2
NMF-LS	a ₁	0.9919	0.9646	1.0000	1.0000
	s ₁	0.9867	0.9934	1.0000	1.0000
	a ₂	0.9979	0.9809	1.0000	1.0000
	s ₂	0.9958	0.9684	1.0000	1.0000
NMF-KL	a ₁	0.9987	0.9974	1.0000	1.0000
	s ₁	0.9981	0.9958	1.0000	1.0000
	a ₂	0.9960	0.9954	1.0000	1.0000
	s ₂	0.9839	0.9917	1.0000	1.0000

Table 1: Mean correlations between signatures of 10 two-component models of NMF-LS and NMF-KL to the mean signature of the 10 analyses. In general, the solutions are close to unique but they become completely unique when removing the background coherence. (LS2=Left hand stimuli subject nr. 2, RS2=Right hand stimuli subject nr. 2).

3.2 Multi-subject analysis

The NMF-LS analysis of the ITPC matrix of *channel x time-frequency – subject – condition* is given in figure 5. A three component analysis explained 73.61 % of the variance in the

data. The third component pertains to some consistent frontal central activity significant ($p < 0.05$) in 14 of the 28 trials (14 subjects \times 2 conditions) and in 9 of these it was even highly significant ($p < 0.001$), see appendix on how this significance level is calculated. The first and second components pertain to the parietal-frontal gamma activity contralateral to stimulus side as the activity is mainly present during right hand stimuli in component one and left hand stimuli in component two. As seen from the subject specific time-frequency maps some variation between the activities of each subject is present. The significance of the two contralateral parietal-frontal activities will be examined in the section on visual inspection of the data.

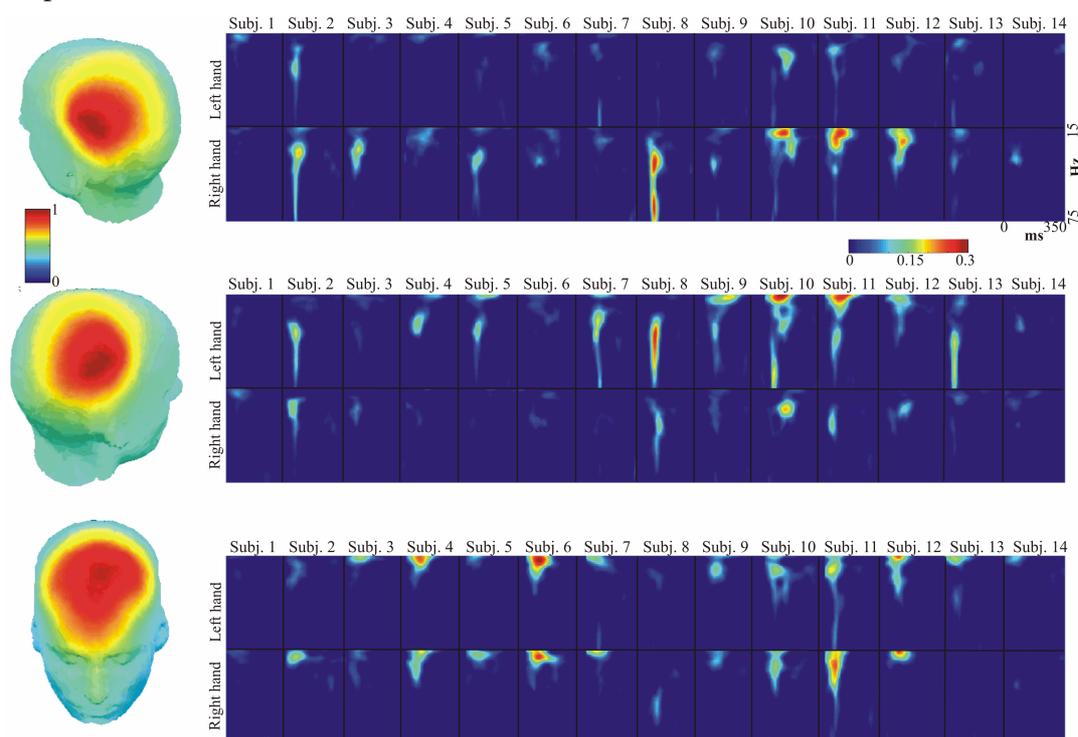


Figure 5: NMF-LS analysis of the ITPC data of channel \times time-frequency-subject-condition (an NMF-KL analysis gave similar results). To the left is the scalp map revealed. To the right is the frequency-time map of each subject during the two conditions. The time-frequency of the top row pertain to left and bottom row to right hand stimuli. From the first factor it is seen that the left parietal-frontal activity is mainly due to gamma activity during right hand stimulation whereas the right parietal-frontal activity revealed in component 2 is mainly due to left hand stimulation. Clearly, the frontal activity given in component 3 is present in almost all subjects in both conditions. The three components explain a total of 73.61 % of the variance.

In the NMWF analysis we assume identical activity through the subjects/conditions and variability only in strength. Two forms of NMWF analysis are performed; one decomposing each stimuli side separately, the other analyzing the frontal activity of the two conditions simultaneously by restricting the analysis to the activity between 15-25 Hz. The result for the first type of decomposition is given in figure 6. Here a two component NMWF analysis (based on NMWF-LS) captures in the first component the parietal contralateral activity and in the second the lower frequent frontal activity.

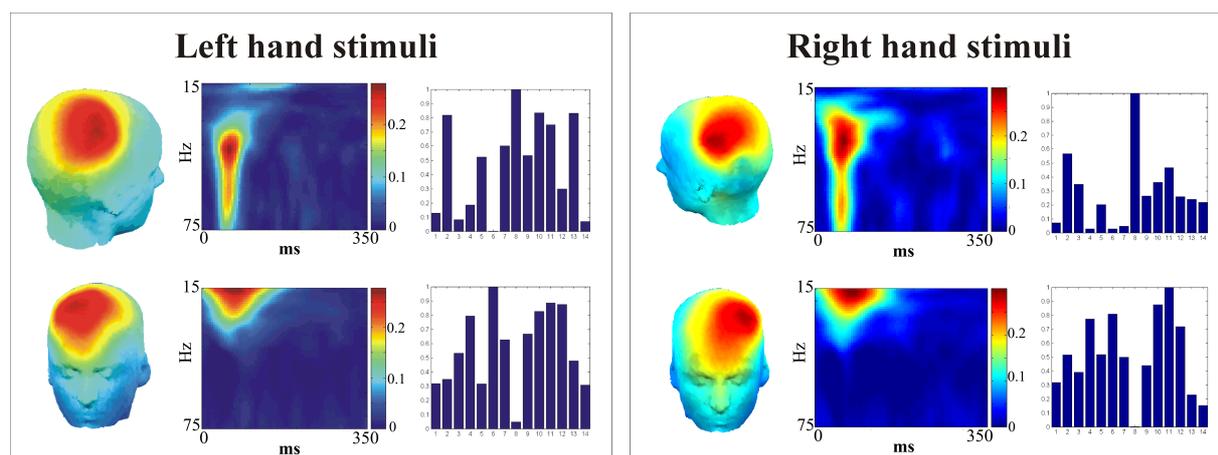


Figure 6: The result of a two component NMWF-LS analysis of the ITPC (NMWF-KL gave similar results) generated from the 14 subjects during left hand stimuli (left panel) and right hand stimuli (right panel). The first component in the left panel pertains to the 40 Hz gamma activity in the right parietal region whereas the first component to the right pertains to the corresponding activity in the left parietal region. Finally the second components in both panels pertains to the more frontal lower frequent activity. While the coherent contralateral parietal-frontal activity is weak in subject 1,3,6 and 14 during left hand stimulation, this activity is weak in subject 1,4,6 and 7 during right hand stimulation. The frontal activity is well present in all but subject 8 both during right and left hand stimulation. While the NMWF model of the left hand stimulation accounts for 49.95 % of the variance the model accounts for 51.82 % of the variance during right hand stimulation. The color axis of the head plots goes from 0 to 1, see also figure 5.

From the decomposition of each side it seems as if the frontal lower frequent activity is slightly lateralized contralateral to stimulus side. Consequently, we analyzed this activity including both conditions simultaneously by restricting the NMWF analysis to the interval 15-25 Hz. As the NMWF captures the activity the most common across subjects and conditions a two component model would separate this activity in an activity present at left hand stimulation and an activity present mainly at right hand stimulation if the activity is dependent on stimulus side. In figure 7 is given this NMWF decomposition using LS as well

as KL. Indeed, the first components in both methods indicate that the left hand stimulation (corresponding to odd numbered bars) was the main contributors to the right central activity, while the right hand stimulation (even numbered bars) was the main contributor to the left central activity given in component two. However, the difference between the condition strengths (even versus odd numbered bars) was not significant.

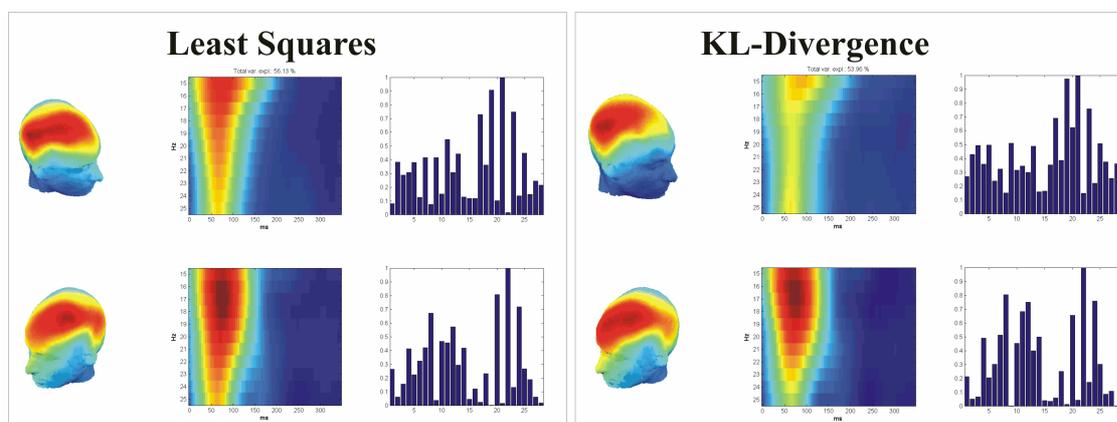


Figure 7: A two component NMWF-LS and NMWF-KL analysis of the lower frequency frontal activity from 15-25 Hz of the 14 subjects during left hand stimuli (odd trials) and right hand stimuli (even trials) giving a total of $14 \cdot 2 = 28$ trials. Clearly both the LS and KL methods indicate that also this activity is lateralized contralateral to stimulus side. While the first component is mainly present in odd trials (left hand stimulation) the second component is mainly present during even trials (right hand stimulation). The two components model explained 56.13 % of the data using the LS method and 53.96 % of the data using the KL method. While second component of both methods are more or less identical, the first component is slightly different since the LS and KL method weights the deviation of the model to the data differently. Color axis are identical to the color axis of figure 6.

The NMWF analysis captures the most common ITPC activity and indicates the strength to which this activity is present in each subject and/or condition. This resulted in an estimate of the subject specific strength relative to the most common activity during each stimulus in the single condition analysis. In the analysis of the lower frequency activity including both conditions the NMWF analysis indicated that the frontal central activity to some extent was condition dependent lateralized contralateral to the stimulus side. Contrary to NMWF, the multi subject NMF analysis captured the subject and condition variability within the same localized scalp region giving three components; a common frontal component and two components each representing the parietal-frontal gamma activity contralateral to the stimulus side.

3.3 Validation by visual inspection

The single condition NMWF weight given in figure 6 (NMWFWeight) as well as the results obtained from a two component multi-subject NMF of each condition (NMFITPC) was compared to the results found by visual inspection of the 16x4 array of channel x time-frequency ITPC activity of each subject. Furthermore, the lateralized ITPC maxima were compared to the ITPC of the same time-frequency point in the corresponding channel across the sagittal midline (controlITPC), i.e. ITPC at FC1; controlITPC at FC2. Since the lateralized components of the multi-subject NMF analysis of each condition separately gave components identical to the lateralized components in figure 5 the figures of this analysis has not been included.

Right hand stimuli

ITPC maxima were observed in the left frontal-parietal central region, most commonly at electrode FC1, but all maxima were within one electrode distance of electrode C1 and none were in mid-line electrodes. Mean frequency was 36Hz (STD 4), latency 60ms (STD 11) and ITPC 0.255 (STD 0.066). The probability of the ITPC in a single channel at one given time frequency point for each subject to exceed 0.255 by random is $2 \cdot 10^{-5}$, see also appendix. The mean NMFITPC was 0.232 (STD 0.071); this was not different from the result of the visual inspection. Again, for each subject to exceed this value by random has a probability of $1 \cdot 10^{-3}$. The mean NMFITPC frequency was lower (33Hz (STD 6) $t = 2.132$, $df = 12$, $p = 0.05$), while the latency did not differ from the result of the visual inspection (73ms (STD 30), $t = -1.776$, $df = 12$, $p = 0.1$). Mean ControlITPC was 0.143 (STD 0.055), a value that could occur at random. The difference between the ITPC and ControlITPC was significant ($t = 8.338$, $df = 13$, $p = 1.4 \times 10^{-6}$). The ITPC was not correlated to the NMWFWeight, but it was correlated to NMFITPC, ($F_{12,12,0}: 7.419$ $\alpha = 0.865$, $p = 0.0008$). ControlITPC was not correlated to NMWFWeight ($F_{13,13,0}: 1.056$ $\alpha = -0.053$, $p = 0.5$), nor to NMFITPC ($F_{12,12,0}: 1.508$ $\alpha = 0.337$, $p = 0.2$).

Left hand stimuli

The ITPC maxima were observed in the right central region, most commonly at electrode C2 and CP4, but all but one of the maxima was within one electrode distance of electrode C2. One maximum was at midline electrode Fz. Mean frequency was 41Hz (STD 9), latency 60ms (STD 10) and ITPC 0.285 (STD 0.072). The chance of exceeding this value by random is less than $5 \cdot 10^{-8}$. Again the NMFITPC was insignificantly smaller (mean 0.245 (STD 0.079) $t = 7.775$, $df = 13$, $p = 3.1 \times 10^{-6}$). The chance of this value occurring at random was $1 \cdot 10^{-4}$. The mean NMF latency was marginally increased (76ms (STD 30), $t = -2.076$, $df = 13$, $p = 0.06$), while the frequency did not differ from the result of the visual inspection (39Hz (STD 13) $t = 1.908$, $df = 13$, $p = 0.08$). Mean ControlITPC was 0.128 (STD 0.056). As with the right hand stimulus the difference between the ITPC and ControlITPC was significant ($t = 6.405$, $df = 13$, $p = 2.3 \times 10^{-5}$). While the ITPC values found by visual

inspection was correlated to NMWFWeight ($F_{13,13,0}$: 2.623 α = 0.619, p = 0.04) and NMFITPC ($F_{13,13,0}$: 13.450 α = 0.925, p < 0.00009), the ControlITPC was not correlated to NMWFWeight ($F_{13,13,0}$: 1.140 α = 0.123, p = 0.4) nor to NMFITPC ($F_{13,13,0}$: 0.832 α = 0.201, p = 0.6).

Apart from the different locations of the maximum activity, neither frequency, latency nor the ITPC value differed between hands of stimulation.

4 Conclusion

Using non-negativity constraints, it is possible to decompose the ITPC to yield easily interpretable time-frequency and scalp maps. The NMF successfully identified the evoked activity of a unilateral stimulus in a single subject analysis but was also capable of capturing the subject and condition variability within the same scalp localized region in a multi-subject, multiple condition analysis. The NMWF gave the subject specific strength to the activity the most similar across the subjects. In the analysis of the lower frequent activity including both conditions the activities would be separated into condition specific components if the similarity across conditions was weaker than the similarity across subjects. This seemed indeed to be the case.

The decomposition findings corresponded well to the results of the visual peak detection of the ITPC in time-frequency plots. Consequently, the present NMF and NMWF analyses efficiently extract the features of interest in the data. Here they reveal how a proprioceptive unilateral stimulus, as predicted, elicits significant evoked gamma activity contralateral to stimulus side from the scalp map likely to stem from somatosensory and motor cortices. Consequently, proprioceptive stimuli are able to elicit evoked gamma activity. Additionally the decompositions revealed a later lower frequent (around 20Hz) activity at a more central frontal location. This activity needs further investigation.

The NMF and NMWF analyses of time-frequency data developed here is obviously also applicable to magnetoencephalographic (MEG) data. Furthermore, the methods used should also be relevant to explore other non-negative measures such as the event related spectral perturbation as well as the other derivatives of EEG wavelet transformations.

Appendix

Background coherence

Let v be the vector strength, i.e. the ITPC, at a given time frequency point defined by averaging N unit vectors in the complex plane, i.e.:

$$v = \frac{1}{N} \left| \sum_n e^{-i\theta_n} \right|$$

Notice further that if θ_n is uniformly distributed over the range $[0;2\pi]$ the following result holds:

$$E(e^{-i\theta_j} e^{i\theta_i}) = E\left(\int_0^{2\pi} \int_0^{2\pi} e^{-i\theta_j} e^{i\theta_i} p(\theta_i) p(\theta_j) d\theta_i d\theta_j\right) = \frac{1}{2\pi} \frac{1}{2\pi} E\left(\int_0^{2\pi} \int_0^{2\pi} e^{i(\theta_i - \theta_j)} d\theta_i d\theta_j\right) = \begin{cases} 0 & \text{if } j \neq i \\ 1 & \text{if } j = i \end{cases}$$

The expected ITPC value at a region without coherence is now given by:

$$E(v) = E\left(\sqrt{\frac{1}{n} \sum_{i=1}^n e^{-i\theta_i} \cdot \frac{1}{n} \sum_{j=1}^n e^{i\theta_j}}\right) = \frac{1}{n} E\left(\sqrt{\sum_{i=1}^n e^{-i\theta_i} \cdot \sum_{j=1}^n e^{i\theta_j}}\right) = \frac{1}{n} E\left(\sqrt{n + \sum_{\substack{i,j \\ i \neq j}}^n e^{i(\theta_j - \theta_i)}}\right) = \frac{1}{n} E\left(\sqrt{n} \sqrt{1 + \frac{1}{n} \sum_{\substack{i,j \\ i \neq j}}^n e^{i(\theta_j - \theta_i)}}\right)$$

Let $\delta = \frac{1}{n} \sum_{\substack{i,j \\ i \neq j}}^n e^{i(\theta_j - \theta_i)}$. By Taylor expansion we get $\sqrt{1 + \delta} \approx 1 + \frac{1}{2} \delta - \frac{1}{4} \delta^2$. We now have

$$E(v) = \frac{1}{n} E\left(\sqrt{n} \sqrt{1 + \delta}\right) \approx \frac{1}{n} E\left(\sqrt{n} \left(1 + \frac{1}{2} \delta - \frac{1}{4} \delta^2\right)\right) = \frac{1}{n} \sqrt{n} \left(1 - \frac{1}{4} E(\delta^2)\right) \approx \frac{1}{\sqrt{n}}$$

since $E(\delta) = 0$ and $E(\delta^2)$ is very small. However, as revealed on figure 9 this approximation is too imprecise to use for statistical evaluation. Consequently, the background average coherence will be estimated by bootstrapping.

Distribution of the ITPC

The ITPC is computed as the mean of n unit vectors in the complex plane. Furthermore, the variance of this distribution is finite since the ITPC takes values between 0 and 1. Consequently, we can appeal to the central limit theorem and expect that the distribution of the ITPC is asymptotically normal for $n \rightarrow \infty$ in regions of coherence, see also (Mardia, K. V. and Jupp, P. E., 1999). However, at regions of random activity the ITPC will tend to zero

approximately by $\frac{1}{\sqrt{n}}$. In this region the ITPC is Raleigh distributed as revealed by bootstrapping on figure 8, i.e. $f(x) = \frac{1}{\sigma^2} x e^{-\frac{x^2}{2\sigma^2}}$. Notice that the Raleigh distribution is completely described by the mean value given by $\sigma\sqrt{\frac{\pi}{2}}$.

Test of significance

A conservative estimate of how significant the found ITPC values are is found by assuming the value of the ITPC at each channel, time and frequency are independent. Assume a total of N ITPC observations. The probability by random that a given ITPC values are less than the observed is given by:

$$P(x_i \leq x) = \int_0^x f(x') dx' = 1 - e^{-\frac{x^2}{2\sigma^2}}$$

Consequently, the probability that the N independent random ITPC values are less than the observed value x is given by:

$$P(x_{\max} \leq x) = P(x_1 \leq x, x_2 \leq x, \dots, x_N \leq x) = (1 - e^{-\frac{x^2}{2\sigma^2}})^N$$

Let this be given at the confidence level α :

$$1 - \alpha = (1 - e^{-\frac{x^2}{2\sigma^2}})^N \Rightarrow x = \sqrt{-2\sigma^2 \ln(1 - (1 - \alpha)^{\frac{1}{N}})}$$

In the null hypothesis of a random ITPC the probability of the maximum exceeding x is α . This is a strong type I control since in reality the ITPC values are not independent, i.e. a given ITPC-value will depend on its neighboring values.

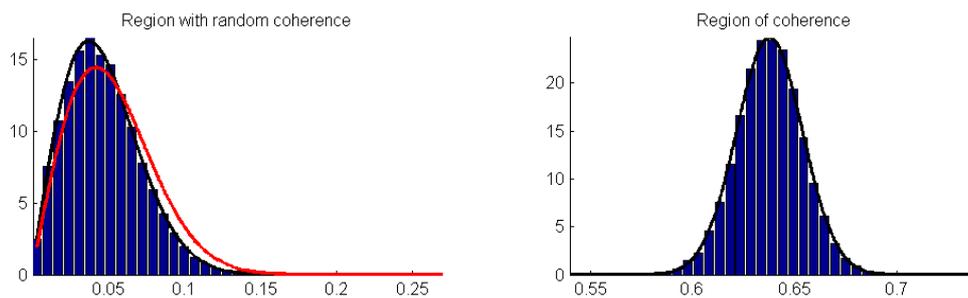


Figure 8: Bootstrap analysis of the ITPC for n=360 epochs. Left panel; the ITPC in regions with no coherence (each epoch were uniformly distributed with angle $[0;2\pi]$). A Raleigh distribution fitted by a bootstrapped mean is given by the black curve,

whereas the distribution corresponding to a mean of $\frac{1}{\sqrt{360}}$ is given by the red curve. Clearly in regions of random coherence the ITPC is Raleigh distributed. However, approximating the mean by $\frac{1}{\sqrt{360}}$ is slightly imprecise. Right panel; The ITPC given in regions of coherence (each epoch were uniformly distributed with angle $[0:\pi]$). Clearly these ITPC values are normally distributed as revealed by the fit of a normal distribution to the bootstrapped ITPC values (black curve).

References

- Andersen, A. H., Rayens, W. S., 2004. Structure-seeking multilinear methods for the analysis of fMRI data. *Neuroimage* 22, 728-739
- Arnfred, S., Chen, A. C., Eder, D., Glenthøj, B., Hemmingsen, R., 2000. Proprioceptive evoked potentials in man: cerebral responses to changing weight loads on the hand. *Neurosci Lett* 288, 111-4
- Arnfred, S. M., 2005. Proprioceptive event related potentials: gating and task effects. *Clin Neurophysiol* 116, 849-60
- Babiloni, C., Babiloni, F., Carducci, F., Cincotti, F., Rosciarelli, F., Arendt-Nielsen, L., Chen, A. C., Rossini, P. M., 2002. Human brain oscillatory activity phase-locked to painful electrical stimulations: a multi-channel EEG study. *Hum Brain Mapp* 15, 112-23
- Beckmann, C. F., Smith, S. M., 2005. Tensorial extensions of independent component analysis for multisubject fMRI analysis. *Neuroimage* 25, 294-311
- Bell, A., Sejnowski, T. J., 1995. An Information-Maximization Approach to Blind Separation and Blind Deconvolution. *Neural Computation* 7, 1129-1159
- Bro, R., 1998. *Multi-way Analysis in the Food Industry: Models, algorithms and Applications*. Amsterdam, Copenhagen.
- Bro, R., Jong, S. D., 1997. A fast non-negativity-constrained least squares algorithm. *J. Chemom.* 11, 393-401
- Carroll, J. D., Chang, J., 1970. Analysis of individual differences in multidimensional scaling via an N-way generalization of 'Eckart-Young' decomposition. *Psychometrika* 35, 283-319
- Chen, A. C., Herrmann, C. S., 2001. Perception of pain coincides with the spatial expansion of electroencephalographic dynamics in human subjects. *Neurosci Lett* 297, 183-6
- Collet, W., 1989. Doubts on the adequacy of the principal component varimax analysis for the identification of event-related brain potential components: a commentary on Glaser and Ruchkin, and Donchin and Heffley. *Biol Psychol* 28, 163-72
- De Pascalis, V., Cacace, I., 2005. Pain perception, obstructive imagery and phase-ordered gamma oscillations. *Int J Psychophysiol* 56, 157-69
- De Pascalis, V., Cacace, I., Massicotte, F., 2004. Perception and modulation of pain in waking and hypnosis: functional significance of phase-ordered gamma oscillations. *Pain* 112, 27-36
- Delorme, A., Makeig, S., 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J Neurosci Methods* 134, 9-21
- Donoho, D., Stodden, V., 2004. When Does Non-Negative Matrix Factorization Give a Correct Decomposition into Parts? *Advances in Neural Information Processing Systems* 17,
- Dormann, W. U., Mundlos, S., Haschke, R., 1987. Comparison of principal components computed with principal factor analysis on the basis of averaged and single-trial ERPs using the Fischer-Roppert procedure. *Int J Psychophysiol* 4, 319-23
- Duzel, E., Habib, R., Schott, B., Schoenfeld, A., Lobaugh, N., McIntosh, A. R., Scholz, M., Heinze, H. J., 2003. A multivariate, spatiotemporal analysis of electromagnetic time-frequency data of recognition memory. *Neuroimage* 18, 185-97
- Eggert, J., Kömer, E., 2004. Sparse coding and NMF. *Proceedings, Neural Networks* 2529-2533

- Gruber, T., Malinowski, P., Muller, M. M., 2004. Modulation of oscillatory brain activity and evoked potentials in a repetition priming task in the human EEG. *Eur J Neurosci* 19, 1073-82
- Hansen, L. K., Larsen, J., Kolenda, T., 2001. Blind Detection of Independent Dynamic Components. *proc. IEEE ICASSP'2001* 5, 3197-3200
- Harshman, R. A., 1970. Foundation of the PARAFAC procedure: models and conditions for an 'explanatory' multi-modal factor analysis. *UCLA Work. Pap. Phon.* 16, 1-84
- Hazan, T., Polak, S., Shashua, A., 2005. Sparse Image Coding using a 3D Non-negative Tensor Factorization. *International Conference of Computer Vision (ICCV)*,
- Herrmann, C. S., Grigutsch, M., Busch, N. A., 2005. EEG oscillations and wavelet analysis.
- Herrmann, C. S., Mecklinger, A., Pfeifer, E., 1999. Gamma responses and ERPs in a visual classification task. *Clin Neurophysiol* 110, 636-42
- Herrmann, C. S., Munk, M. H., Engel, A. K., 2004. Cognitive functions of gamma-band activity: memory match and utilization. *Trends Cogn Sci* 8, 347-55
- Hoyer, P. O., 2002. Non-Negative Sparse Coding. *Neural Networks for Signal Processing XII*
- Højen-Sørensen, P. A. d. F. R., Winther, O., Hansen, L. K., 2002. Mean Field Approaches to Independent Component Analysis. *Neural Computation* 14, 889-918
- Jansen, B. H., Hegde, A., Boutros, N. N., 2004. Contribution of different EEG frequencies to auditory evoked potential abnormalities in schizophrenia. *Clin Neurophysiol* 115, 523-33
- Jones, K., Begleiter, H., Porjesz, B., Wang, K., Chorlian, D., 2002. Complexity measures of event related potential surface Laplacian data calculated using the wavelet packet transform. *Brain Topogr* 14, 333-44
- Kayser, J., Fong, R., Tenke, C. E., Bruder, G. E., 2003. Event-related brain potentials during auditory and visual word recognition memory tasks. *Brain Res Cogn Brain Res* 16, 11-25
- Koles, Z. J., 1998. Trends in EEG source localization. *Electroencephalogr Clin Neurophysiol* 106, 127-37
- Kruskal, J. B., 1977. Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics. *Linear Algebra Appl.* 18, 95-138
- Lachaux, J. P., George, N., Tallon-Baudry, C., Martinerie, J., Hugueville, L., Minotti, L., Kahane, P., Renault, B., 2005. The many faces of the gamma band response to complex visual stimuli. *Neuroimage* 25, 491-501
- Lee, D. D., Seung, H. S., 1999. Learning the parts of objects by non-negative matrix factorization. *Nature* 401(6755), 788-791
- Lee, D. D., Seung, H. S., 2001. Algorithms for non-negative matrix factorization. *Advances in Neural information processing* 13,
- Makeig, S., Jung, T. P., Bell, A. J., Ghahremani, D., Sejnowski, T. J., 1997. Blind separation of auditory event-related brain responses into independent components. *Proc Natl Acad Sci U S A* 94, 10979-84
- Makeig, S., Westerfield, M., Townsend, J., Jung, T. P., Courchesne, E., Sejnowski, T. J., 1999. Functionally independent components of early event-related potentials in a visual spatial attention task. *Philos Trans R Soc Lond B Biol Sci* 354, 1135-44
- Mardia, K. V., Jupp, P. E., 1999. *Directional Statistics*. WILEY & SONS 76-77
- Miwakeichi, F., Martinez-Montes, E., Valdes-Sosa, P. A., Nishiyama, N., Mizuhara, H., Yamaguchi, Y., 2004. Decomposing EEG data into space-time-frequency components using Parallel Factor Analysis. *Neuroimage* 22, 1035-45

- Mørup, M., Hansen, L. K., Herrmann, C. S., Parnas, J., Arnfred, S. M., 2006. Parallel Factor Analysis as an exploratory tool for wavelet transformed event-related EEG. *Neuroimage* vol. 29, 938-947
- Oja, E., Plumbley, M., 2004. Blind Separation of Positive Sources by Globally Convergent Gradient Search. *Neural Computation* 16, 1811-1825
- Palva, J. M., Palva, S., Kaila, K., 2005. Phase Synchrony among Neuronal Oscillations in the Human Cortex. *The Journal of Neuroscience* 25, 3962-3972
- Picton, T. W., Bentin, S., Berg, P., Donchin, E., Hillyard, S. A., Johnson, R., Jr., Miller, G. A., Ritter, W., Ruchkin, D. S., Rugg, M. D., Taylor, M. J., 2000. Guidelines for using human event-related potentials to study cognition: recording standards and publication criteria. *Psychophysiology* 37, 127-52
- Paatero, P., 2000. Construction and analysis of degenerate PARAFAC models. *Journal of Chemometrics* 14, 285-299
- Rogers, L. J., 1991. Determination of the number and waveshapes of event related potential components using comparative factor analysis. *Int J Neurosci* 56, 219-46
- Salakhutdinov, R., Sam, R., 2003. Adaptive Overrelaxed Bound Optimization Methods. Proceedings of the twentieth International Conference on Machine Learning (ICML-2003)
- Shashua, A., Hazan, T., 2005. Non-Negative Tensor Factorization with Applications to Statistics and Computer Vision. International Conference on Machine Learning (ICML)
- Sidiropoulos, N. D., Bro, R., 2000. On the uniqueness of multilinear decomposition of N-way arrays. *J. Chemometrics* 14, 229-239
- Tallon-Baudry, C., Bertrand, O., 1999. Oscillatory gamma activity in humans and its role in object representation. *Trends Cogn Sci* 3, 151-162
- Tecchio, F., Babiloni, C., Zappasodi, F., Vecchio, F., Pizzella, V., Romani, G. L., Rossini, P. M., 2003. Gamma synchronization in human primary somatosensory cortex as revealed by somatosensory evoked neuromagnetic fields. *Brain Res* 986, 63-70