Shift-Invariant Multilinear Decomposition of Neuroimaging Data

Morten Mørup [a] Lars Kai Hansen [a] Sidse Marie Arnfred [b]
Lek-Heng Lim [c] Kristoffer Hougaard Madsen [a,d,*]

[a] *Informatics and Mathematical Modelling, Technical University of Denmark, Lyngby, Denmark*

[b] *Department of Psychiatry, Hvidovre hospital, University Hospital of Copenhagen, Denmark*

[c] *Department of Mathematics, University of California, Berkeley, CA, USA*

[d] *Danish Research Centre for Magnetic Resonance, Copenhagen University Hospital, Hvidovre, Denmark*

Abstract

We present an algorithm for multilinear decomposition that allows for arbitrary shifts along one modality. The method is applied to neural activity arranged in the three modalities space, time, and trial. Thus, the algorithm models neural activity as a linear superposition of components with a fixed time course that may vary across either trials or space in its overall intensity or latency. Its utility is demonstrated on simulated data as well as actual EEG, and fMRI data. This work shows how pseudo-multilinear decompositions of multiway data can successfully cope with variable latencies in data derived from neural activity — a problem that has caused degenerate solutions especially in modeling neuroimaging data with instantaneous multilinear decompositions. Our algorithm is available for download at `www.erpwavelab.org`.

Key words: parallel factor analysis (PARAFAC), canonical decomposition (CANDECOMP), shift invariance, shifted CP (SCP), CP-degeneracy, fMRI, EEG, hypermatrix, multilinear, uniqueness, event related averaging, retinotopic mapping.

* To whom correspondence should be addressed:
Kristoffer H. Madsen
Informatics and Mathematical Modelling, Technical University of Denmark
Richard Petersens Plads, 2800 Kgs. Lyngby, Denmark
Phone: +45 4525 3900
Fax: +45 4587 2599

Email addresses: `mm@imm.dtu.dk` (Morten Mørup), `khm@imm.dtu.dk` (Kristoffer

1 Introduction

Neuroimaging data such as electroencephalography (EEG) and functional magnetic resonance imaging (fMRI) data is heavily impacted by noise. In general, it is difficult if not impossible to extract underlying neural activity from noisy raw data. To overcome the poor signal to noise ratio (SNR) it is customary to average over repeated trials. Such practice improves the SNR as noise is often unrelated to the event and hence self-cancels (has zero mean) in theory. As such, the power of white noise decreases as $\sigma^2_{\text{noise}}/N_{\text{epoch}}$, where σ^2_{noise} is the variance of the noise and is assumed constant over the trials while N_{epoch} is the number of trials. (Notice that the relation is only correct for white noise — in general EEG noise is correlated in a way that the true falloff will differ). As SNR varies through the trials it would be desirable to give more importance to trials with high SNR than to trials with low SNR.

To extract the underlying neural activity it is customary, prior to or post averaging, to decompose the data using various types of decompositions based on the bilinear factor analytic model

$$\mathbf{X}_{i,j} = \sum_{d=1}^{D} \mathbf{A}_{i,d}\mathbf{B}_{j,d} + \mathbf{E}_{i,j}$$

where $\mathbf{E}_{i,j}$ denotes the residual (Donchin and Heffley, 1978; Makeig *et al.*, 1996, 1997, 2002; McKeown *et al.*, 1998, 2003). For EEG and fMRI data the recorded data may be represented by the *channel/voxel* $\times$ *time* matrix $\mathbf{X} \in \mathbb{R}^{I\times J}$. The decomposition above then describes the data as a sum of components separated into time profiles $\mathbf{B}_1, \ldots, \mathbf{B}_D$ with corresponding spatial topographies $\mathbf{A}_1, \ldots, \mathbf{A}_D$. However, since modeling data by a factor analytic type decomposition is ambiguous, additional constraints that are often not physiologically justified have to be imposed. For example, in Singular Value Decomposition (SVD) and Principal Component Analysis (PCA), components are assumed to be uncorrelated; whereas in Independent Component Analysis (ICA), components are assumed to be statistically independent.

If the data are recorded over K trials of varying strength, we obtain a *channel/voxel* $\times$ *time* $\times$ *trial* hypermatrix $\mathcal{X} \in \mathbb{R}^{I\times J\times K}$. Forming a trilinear decomposition as in factor analysis yields the CANDECOMP/PARAFAC (CP) model Carroll and Chang (1970); Harshman (1970). The CP model is given by

$$\mathcal{X}_{i,j,k} = \sum_{d=1}^{D} \mathbf{A}_{i,d}\mathbf{B}_{j,d}\mathbf{C}_{k,d} + \mathcal{E}_{i,j,k}.$$

Here $\mathbf{C}_d$ gives the degree in which the profile time series $\mathbf{B}_d$ with spatial topography $\mathbf{A}_d$ is present throughout the various trials. The model is illustrated in Figure 1. As Kruskal proved in (Kruskal, 1977), the CP model is unique under mild conditions. Conditions that, in the presence of noise in the data, are practically always

Hougaard Madsen).

satisfied. Consequently, modeling repeated trials by CP in theory not only improves the component identification but also resolves the ambiguities encountered when modeling the averaged data by (bilinear) factor analysis (Kruskal, 1977). Notice that the application of CP to EEG was already suggested in the original paper on CP (Harshman, 1970) and was later reinvented in Möcks (1988) under the name *topographic component analysis*. In Andersen and Rayens (2004) it was further demonstrated how the CP model is useful in the analysis of neuroimaging data such as fMRI (Andersen and Rayens, 2004). Additional applications of multilinear (also called multiway) modeling in EEG and fMRI include (Möcks, 1988; Field and Graupe, 1991; Wang *et al.*, 2000; Beckmann and Smith, 2005; Miwakeichi *et al.*, 2004; Mørup *et al.*, 2006; De Vos *et al.*, 2007; Acar *et al.*, 2007; Dyrholm *et al.*, 2007).

Time shifts occur naturally in fMRI data. For instance, these could be due to hemodynamic delay (Buxton *et al.*, 1998) or they could arise in stimuli studies (Sereno *et al.*, 1995), where delays play a particularly important role. For EEG data, onset changes of prominent physiological activity not phase locked to the event (such as eye blinks and alpha activity) are known to cause delays across trials. Extending the CP model to incorporate delays form the shifted CP model (shifted over third mode), henceforth denoted as SCP model,

$$\mathcal{X}_{i,j,k} = \sum_{d=1}^{D} \mathbf{A}_{i,d}\mathbf{B}_{j-\boldsymbol{\tau}_{k,d},d}\mathbf{C}_{k,d} + \mathcal{E}_{i,j,k}. \tag{1}$$

Here, each time profile $\mathbf{B}_d$ is shifted according to the vector $\boldsymbol{\tau}_{k,d}$ that represents time-samples dependent on the k index of the third mode. Hence, the shifts will be along the j index. See also Figure 1. Data generated from the SCP model is no longer multilinear and therefore the CP model is no longer a valid model for the data. When data violates multilinearity, 'CP-degenerate' solutions are known to occur. Roughly speaking, this refers to solutions in which some component loadings are highly correlated in all modes and the elements of these components become arbitrarily large (Stegeman, 2007). CP-degeneracy makes the estimation unstable, the algorithm slow to converge (or even diverge), and the result difficult to interpret — largely because the model is plaqued by strong between-component cancellation (Harshman and Lundy, 1984). For a mathematically precise discussion, the reader is referred to de Silva and Lim (2008). To avoid CP-degeneracy in the CP model, artificial restrictions in the form of orthogonality (Field and Graupe, 1991; Lundy *et al.*, 1989) or independence (Beckmann and Smith, 2005) have been imposed; alternatively, the signal is analyzed via purely additive models based on analysis of amplitudes in a spectral representation (Miwakeichi *et al.*, 2004; Mørup *et al.*, 2006). We found these ad hoc measures unsatisfactory. Rather than restricting the CP model, we propose a *pseudo-mutilinear* model using the unambiguous CP model combined with a time-shift accounting for explicit delays. We claim that this will alleviate the issues inherent to previous matrix and multilinear decompositions of neuroimaging data. We will support our claim in this paper with several experiments with actual neuroimaging data.

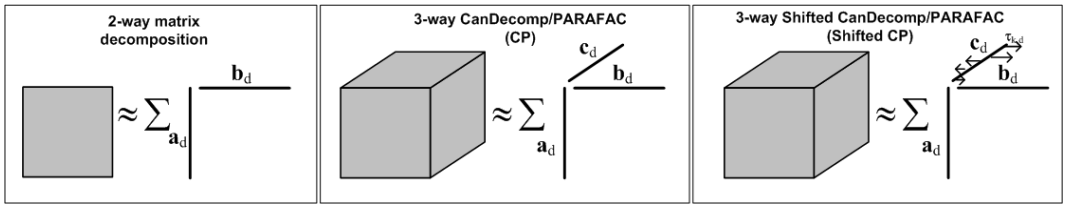


Fig. 1. The CP model can be considered a straightforward generalization of 2-way (matrix) decomposition (left panel) to arrays of more than two modalities (middle panel). Thus, the data is described by an outer product of factors pertaining to each of the modalities. The shifted CP model allow shifts to occur over the second mode such that for each index of the third mode the component of the second mode is shifted a given amount.

Our modeling of delays is further motivated by a number of papers that explain how degenerate solutions might be caused by component delays (Field and Graupe, 1991; Andersen and Rayens, 2004; Harshman *et al.*, 2003a; Hong and Harshman, 2003). Indeed, if shifts are causing CP-degeneracy, then it is more natural to extend the CP model to accomodate shifts rather than resorting to orthogonality or independence constraints that may not be physiologically justified. Furthermore, a decomposition into profiles resembling pairs of functions and their derivatives, e.g. pairs of cosine and sine functions in (Field and Graupe, 1991), provides strong evidence that neuroimaging data should be decomposed by a model accounting for shifts rather than models based on instantaneous mixing.

We stress here that CP-degeneracy should not be assumed to be a consequence of poor modeling. It has been shown in de Silva and Lim (2008) that CP-degeneracy, a manifestation of the ill-posedness of the best rank-r approximation problem for hypermatrices, is not an isolated phenomenon. Two results are worth highlighting here: (1) CP-degeneracy occurs on a set of positive volume, i.e. if we pick a random $\mathcal{X} \in \mathbb{R}^{I \times J \times K}$, there is a non-zero probability that it will be CP-degenerate; (2) a non-degenerate case can become CP-degenerate under small noisy perturbation. As such, it is important to have safeguards against CP-degeneracy built into CP-type models. The SCP model in this paper is proposed with this in mind. We demonstrate that the CP model, when suitably modified to allow for time shifts, will accurately capture the features of time-delayed data sets and at the same time avoid the CP-degeneracy pitfall.

The paper is structured as follows. We first derive the algorithm for SCP and demonstrate the algorithm on synthetic and real EEG data. Next, we validate the usefulness of our SCP decomposition for fMRI data based on a retinotopic mapping paradigm where delay modeling is used when forming the retinotopic maps (Sereno *et al.*, 1995). In particular, we will investigate the following questions:

(1) Can shifts improve component identification?
(2) Can shifts yield a more compact representation of neuroimaging data?
(3) Does the SCP model alleviate degeneracy?

2 Methods

Factor analysis with shifts have been treated in numerous papers (Bell and Sejnowski, 1995; Torkkola, 1996; Emile and Comon, 1998; Yeredor, 2003; Harshman *et al.*, 2003a,b; Truccolo *et al.*, 2003; Mørup *et al.*, 2007a,b). Shifts based on the CP model has previously been treated in (Hong and Harshman, 2003; Knuth *et al.*, 2006). Unfortunately, the algorithms devised are prohibitively slow for large scale problems such as EEG and fMRI and does not allow for non-integer shifts. Presently, we derive an efficient algorithm for SCP with the following benefits

- Closed form solutions are obtained for all modes while keeping the remaining modes fixed.
- Integer shifts are estimated by cross-correlation rather than the exhaustive searches used in (Hong and Harshman, 2003).
- Non-integer shifts can be found by iterative methods in the frequency domain.

Non-integer shifts are in particular important for fMRI data due to low temporal resolution.

2.1 Notations

In the following $\mathcal{U}_{i,j,k}$, $\mathbf{U}_{i,j}$ and $\tilde{\mathcal{U}}_{i,f,k}$, $\tilde{\mathbf{U}}_{i,f}$ will denote the same hypermatrix and matrix of size $I \times J \times K$ and $I \times J$ in the time and frequency domain respectively using the discrete fourier transform (DFT) and inverse fourier transform along the second modality indexed by j. Furthermore, $\mathbf{U} \bullet \mathbf{V}$ denotes the direct product, i.e. element-wise multiplication. Let $\boldsymbol{\tau}$ be a matrix containing the delays pertaining to each index of $\mathbf{C}$. Notice, shifting the dth component time series τ samples, i.e. $\mathbf{X}_{i-\tau,d}$ corresponds in the frequency domain to the complex multiplication $\tilde{\mathbf{X}}_{f,d} e^{-2\frac{f-1}{J}\pi\imath\tau}$. Note that we write $\imath := \sqrt{-1}$ to distinguish it from the letter i used in the indices. Thus, a combined shift and mixing matrix at frequency f can be stated as $\tilde{\mathbf{C}}^{(f)} = \mathbf{C} \bullet e^{-2\pi(f-1)\imath\boldsymbol{\tau}/J}$ where $e^{\boldsymbol{\tau}}$ denotes element-wise exponentiating the elements of $\boldsymbol{\tau}$. The ith row of a matrix will be denoted $\mathbf{U}_{i,:}$. The n-mode matricizing operation unfolds the hypermatrix $\mathcal{U}^{I_1 \times I_2 \times \ldots \times I_N}$ into a matrix $\mathbf{U}_{(n)}^{I_n \times I_1 \cdot I_2 \cdots I_{n-1} \cdot I_{n+1} \cdots I_N}$ while the folding (i.e. hypermatrizing) of $\mathbf{U}_{(n)}$ to the hypermatrix $\mathcal{U}$ denotes the opposite operation. Finally, the Khatri-Rao product is given by $\mathbf{A} \odot \mathbf{B} = [\mathbf{A}_1 \otimes \mathbf{B}_1 \ \mathbf{A}_2 \otimes \mathbf{B}_2 \ \cdots \ \mathbf{A}_D \otimes \mathbf{B}_D]$ where $\otimes$ is the regular Kronecker product. D in the CP and SCP model will denote the number of components. Only if the CP model is exact will D also denote the rank of the modeled hypermatrix.

2.2 CP model estimation

The CP model is normally estimated by alternating least squares (Bro and Andersson, 2000) which solves for $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ according to

$$\begin{aligned}
\mathbf{X}_{(1)} &= \mathbf{A}(\mathbf{C} \odot \mathbf{B})^{\mathsf{T}} + \mathbf{E}_{(1)} \quad \text{via} \quad \mathbf{A} \leftarrow \mathbf{X}_{(1)}(\mathbf{C} \odot \mathbf{B})^{\mathsf{T}\dagger}, \\
\mathbf{X}_{(2)} &= \mathbf{B}(\mathbf{C} \odot \mathbf{A})^{\mathsf{T}} + \mathbf{E}_{(2)} \quad \text{via} \quad \mathbf{B} \leftarrow \mathbf{X}_{(2)}(\mathbf{C} \odot \mathbf{A})^{\mathsf{T}\dagger}, \\
\mathbf{X}_{(3)} &= \mathbf{C}(\mathbf{B} \odot \mathbf{A})^{\mathsf{T}} + \mathbf{E}_{(3)} \quad \text{via} \quad \mathbf{C} \leftarrow \mathbf{X}_{(3)}(\mathbf{B} \odot \mathbf{A})^{\mathsf{T}\dagger},
\end{aligned}$$

i.e. by formulating the estimation problem as regular matrix analysis problems using the matricizing operation and Khatri-Rao product and alternatingly solving for the loadings of each mode via Moore-Penrose inverses.

The CP model is unique if

$$k_{\mathbf{A}} + k_{\mathbf{B}} + k_{\mathbf{C}} \geq 2D' + 2 \tag{2}$$

where D' is the rank of the hypermatrix and $k_{\mathbf{A}}$ is the Kruskal rank denoting the smallest subset of columns of $\mathbf{A}$ that is guaranteed to be linearly independent (Kruskal, 1977). Thus, $k_{\mathbf{A}} \leq \operatorname{rank}(\mathbf{A})$. When modeling the data using a low rank approximation ($D < D'$) the above criterion guarantees that the residuals are uniquely defined. In the presence of noise both $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ will have full rank and uniqueness is also guaranteed by proofs given in (Harshman, 1972; Möcks, 1988).

2.3 SCP model

In (Hong and Harshman, 2003) the SCP model was proposed and an algorithm devised based on exhaustive integer searches over all possible shifts. This is however very expensive making the estimation infeasible when including many shifts. Thus, we here propose to solve the model in the frequency domain rather than the time-domain. The attractive property being that each integer delay $\boldsymbol{\tau}_{k,d}$ has a closed form solution while keeping the remaining delays fixed given by calculating cross-correlations which is inexpensive in the frequency domain. Furthermore, in a frequency representation non-integer delays can be estimated using gradient based searches. Finally, in a frequency representation $\mathbf{B}$ has a closed form solution.

In the frequency domain the SCP model is given by

$$\tilde{\mathcal{X}}_{i,f,k} = \sum\nolimits_{d=1}^{D} \mathbf{A}_{i,d}\tilde{\mathbf{B}}_{f,d}\mathbf{C}_{k,d}e^{-2\pi(f-1)\imath\boldsymbol{\tau}_{k,d}/J} + \tilde{\mathcal{E}}_{i,f,k}.$$

Thus, the sources $\mathbf{B}_d$ are assumed to be periodic such that shifts $\boldsymbol{\tau}_{k,d}$ correspond to the complex multiplication of $\tilde{\mathbf{B}}_d$ with the factor $\exp[-2(f-1)\pi\imath\boldsymbol{\tau}_{k,d}/J]$. Thus, we assume that the data can be arranged such that each source time course in each

epoch is periodic, if this is not the case the periodicity can be enforced by introducing a temporal windowing function. Notice, due to Parseval's identity there is a one-to-one correspondence between the least squares error in the time and frequency domain such that the least squares minimization can be performed arbitrarily between the two domains

$$\sum_{i,j,k} \|\mathcal{E}_{i,j,k}\|^2 = \frac{1}{J} \sum_{i,f,k} \|\tilde{\mathcal{E}}_{i,f,k}\|^2.$$

2.3.1 $\mathbf{A}$, $\mathbf{B}$ *and* $\mathbf{C}$ *update*

Let $\tilde{\mathbf{B}}^{(k)}_{f,d} = \tilde{\mathbf{B}}_{f,d} \bullet \exp[-2(f-1)\pi\imath\boldsymbol{\tau}_{k,d}/J]$, i.e. $\tilde{\mathbf{B}}$ componentwise shifted according to the delays to the kth channel. Let further $\mathbf{Z}_{j+k(J-1),d} = \mathbf{C}_{k,d}\mathbf{B}^{(k)}_{j,d}$, i.e. the Khatri-Rao product between $\mathbf{C}$ and the shifted version of $\mathbf{B}$.

Using the n-mode matricizing and the Khatri-Rao product we can again state the estimation of $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ by ordinary factor analysis

$$\begin{aligned}
\mathbf{X}_{(1)} &= \mathbf{A}\mathbf{Z}^{\mathsf{T}} + \mathbf{E}_{(1)} \quad \text{via} \quad \mathbf{A} \leftarrow \mathbf{X}_{(1)}\mathbf{Z}^{\mathsf{T}\dagger} \\
\tilde{\mathbf{X}}_{(2)_{f,:}} &= \tilde{\mathbf{B}}_{f,:}(\tilde{\mathbf{C}}^{(f)} \odot \mathbf{A})^{\mathsf{T}} + \tilde{\mathbf{E}}_{(2)_{f,:}} \quad \text{via} \quad \tilde{\mathbf{B}}_{f,:} \leftarrow \tilde{\mathbf{X}}_{(2)_{f,:}}(\tilde{\mathbf{C}}^{(f)} \odot \mathbf{A})^{\mathsf{T}\dagger} \\
\mathbf{X}_{(3)_{k,:}} &= \mathbf{C}_{k,:}(\mathbf{B}^{(k)} \odot \mathbf{A})^{\mathsf{T}} + \mathbf{E}_{(3)_{k,:}} \quad \text{via} \quad \mathbf{C}_{k,:} \leftarrow \mathbf{X}_{(3)_{k,:}}(\mathbf{B}^{(k)} \odot \mathbf{A})^{\mathsf{T}\dagger}
\end{aligned}$$

Notice, where as $\mathbf{A}$ and $\mathbf{C}$ are updated in the real domain $\mathbf{B}$ is updated in the complex domain. $\mathbf{B}$ is only real valued in the time domain if the following relation holds in the frequency domain

$$\tilde{\mathbf{B}}_{J-f+1,d} = \tilde{\mathbf{B}}^{*}_{f,d}, \tag{3}$$

such that $\tilde{\mathbf{B}}$ is conjugate symmetric. This constraint is enforced by updating the first $\lfloor J/2 \rfloor + 1$ elements, i.e. up to the Nyquist frequency, while setting the remaining elements according to equation 3. Since the estimation is stated as regular factor analysis problems non-negativity constraints for $\mathbf{A}$ and $\mathbf{C}$ can be imposed using the active set procedure given in (Bro and de Jong, 1997).

2.3.2 $\boldsymbol{\tau}$ *update*

Let

$$\mathbf{R}_{(3)^{d'}_{k,:}} = \mathbf{X}_{(3)_{k,:}} - \sum\nolimits_{d \neq d'} \mathbf{C}_{k,d}(\mathbf{B}^{(k)}_{d} \odot \mathbf{A}_d)^{\mathsf{T}},$$

i.e. $\mathbf{R}^{d'}_{(3)_{k,:}}$ is the remaining signal at the kth row when projecting all but the d'th source out of $\mathbf{X}_{(3)}$. Notice, with this notation the least squares error can be rewritten

as

$$\begin{aligned}
&\textstyle\sum_k \|\mathbf{X}_{(3)_{k,:}} - \sum_d^D \mathbf{C}_{k,d}(\mathbf{B}_d^{(k)} \odot \mathbf{A}_d)^\mathsf{T})\|^2 \\
=&\textstyle\sum_k \|\mathbf{R}^{d'}_{(3)_{k,:}} - \mathbf{C}_{k,d'}(\mathbf{B}_{d'}^{(k)} \odot \mathbf{A}_{d'})^\mathsf{T}\|^2 \\
=&\|\mathbf{R}^{d'}_{(3)_{k,:}}\|^2 - \mathbf{C}_{k,d'} \textstyle\sum_j \mathbf{B}_{j-\tau_{k,d'},d'} \sum_i \mathcal{R}^{d'}_{i,j,k}\mathbf{A}_{i,d'} + \|\mathbf{C}_{k,d'}(\mathbf{B}_{d'}^{(k)} \odot \mathbf{A}_{d'})^\mathsf{T}\|^2.
\end{aligned}$$

The first and third term is independent of $\tau_{k,d'}$. Thus, the least square error is minimized when the second term is maximized. Since $\mathbf{C}_{k,d'}$ is a constant this can be omitted such that we get

$$\begin{aligned}
\mathbf{r}_j^{(k,d')} &= \textstyle\sum_i \mathcal{R}^{d'}_{i,j,k}\mathbf{A}_{i,d'} \\
\tilde{\mathbf{c}}_{k,d'}(f) &= \tilde{\mathbf{r}}_f^{(k,d')^*}\tilde{\mathbf{B}}_{f,d'}.
\end{aligned}$$

$\boldsymbol{\tau}_{k,d'}$ can now be estimated as

$$\begin{aligned}
\hat{\boldsymbol{\tau}}_{k,d'} &= \operatorname*{argmax}_t |\mathbf{c}_{k,d'}(t)| \\
\boldsymbol{\tau}_{k,d'} &= \hat{\boldsymbol{\tau}}_{k,d'} - (J+1).
\end{aligned}$$

I.e. as the delay corresponding to maximum absolute cross-correlation between $\mathbf{r}^{(k,d')}$-the time profile of the residual for the d' component and $\mathbf{B}_{d'}$-the component time profile. The value of $\mathbf{C}_{k,d'}$ corresponding to this delay is given by

$$\mathbf{C}_{k,d'} = \frac{\mathbf{c}_{k,d'}(\hat{\boldsymbol{\tau}}_{k,d'})}{\mathbf{B}_{d'}^\mathsf{T}\mathbf{B}_{d'}}.$$

If $\mathbf{C}$ is constrained positive only positive values of $\mathbf{c}_{k,d}(t)$ are considered. The above procedure can only estimate integer delays. However, by minimizing the least squares error in the complex domain with respect to $\boldsymbol{\tau}$ a gradient and Hessian can be calculated such that non-integer delays can be estimated for instance by the Newton-Raphson procedure. For details on this see (Mørup *et al.*, 2007b,a).

The convergence criterion of the algorithm was set to a relative change in fit less than 10^{-6} or when the algorithm had run for 1000 iterations. The parameters in the Shifted CP model were updated in the sequence; $\mathbf{B}$, $\boldsymbol{\tau}$, $\mathbf{A}$, $\mathbf{C}$ and the updates were accelerated using the approach suggested in (Tomasi, 2006). The number of components were estimating using an adapted version of the core consistency diagnostic (Bro and Kiers, 2003).

2.3.3 Uniqueness

Unfortunately, the rigorous proof of uniqueness by Kruskal using Kruskal-rank given in equation 2 is involved. However, the uniqueness assuming $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ all have full rank can be proven by considering the CP model in a slab representation

(Harshman, 1972; Möcks, 1988). For the kth slab the CP model reads

$$\begin{aligned}\mathbf{X}_{:,:,k} &\approx \mathbf{A}\,\mathrm{diag}(\mathbf{C}_{k,:})\mathbf{B}^{\mathsf{T}} \\ &= \mathbf{A}\mathbf{P}[\mathbf{P}^{-1}\,\mathrm{diag}(\mathbf{C}_{k,:})\mathbf{Q}]\mathbf{Q}^{-1}\mathbf{B}^{\mathsf{T}} \\ &= \hat{\mathbf{A}}\,\mathrm{diag}(\hat{\mathbf{C}}_{k,:})\hat{\mathbf{B}}^{\mathsf{T}}.\end{aligned}$$

Thus, if two solutions $\mathbf{A}, \mathbf{B}, \mathbf{C}$ and $\hat{\mathbf{A}}, \hat{\mathbf{B}}, \hat{\mathbf{C}}$ exists there must be a mapping from one solution to the other given by $\mathbf{P}$ and $\mathbf{Q}$. However, for this mapping the term $\mathbf{P}^{-1}\,\mathrm{diag}(\mathbf{C}_{k,:})\mathbf{Q}$ has to be diagonal for all k which when $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ have full rank restricts $\mathbf{P}$ and $\mathbf{Q}$ to be simple scale and permutation matrices (Harshman, 1972; Möcks, 1988). For the SCP model we instead have

$$\begin{aligned}\mathbf{X}_{:,:,k} &\approx \mathbf{A}\,\mathrm{diag}(\mathbf{C}_{k,:})\mathbf{B}^{(k)\mathsf{T}} \\ &= \mathbf{A}\mathbf{P}[\mathbf{P}^{-1}\,\mathrm{diag}(\mathbf{C}_{k,:})\mathbf{Q}]\mathbf{Q}^{-1}\mathbf{B}^{(k)\mathsf{T}} \\ &= \hat{\mathbf{A}}\,\mathrm{diag}(\hat{\mathbf{C}}_{k,:})\hat{\mathbf{B}}^{(k)\mathsf{T}}\end{aligned}$$

Where $\mathbf{B}^{(k)}_d = \mathbf{B}_{j-\boldsymbol{\tau}_{k,d},d}$. Although, the CP model is extended such that $\mathbf{B}$ is shifted $\mathbf{P}^{-1}\,\mathrm{diag}(\mathbf{C}_{k,:})\mathbf{Q}$ still has to remain diagonal for all values of k. This again strongly restricts $\mathbf{P}$ and $\mathbf{Q}$. The obvious ambiguities are scaling, permutation, relative shift and onset as well as period of the time-series. We tested the uniqueness of the decomposition by investigating the similarity of 250 decompositions randomly initialized. We found that the 50 decompositions with lowest least square error were identical.

2.4 Experimental details

2.4.1 EEG data

A healthy subject was enrolled after informed consent as approved by the Ethics Committee. EEG 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 lateral canthus/brow for each eye. The grounding electrodes for the active electrodes (CMS and DRL) were placed centrally, close to POz. Data were recorded continuously at 2048 Hz/channel, band pass 0.1–760 Hz, by a LabView© application (ActivView©). After down sampling to 512 Hz/channel, 50 Hz electronic noise was projected out using a multiple linear regression filter in intervals of 2 seconds. Further data processing was done in EEGLAB for MATLAB© (Delorme and Makeig, 2004). The data were referenced to digitally linked earlobes, high-pass filtered > 3 Hz and cut into epochs (-500 to $+1500$ ms). The stimulus paradigm was taken from (Herrmann *et al.*, 2004b,a). Briefly, it consists of two types of black and white drawings: (1) objects (Ob), which are easily recognizable everyday type of objects like a chair, a number or a pipe, and (2) non-objects (Nob), which are chaotic re-arrangements of the Ob

drawings. Each stimulus category included 313 events and an object was presented up to three times. Stimulus delivery was controlled by the Presentation© software. Each stimulus was presented for 1 s, and randomized inter stimulus interval was jittering between 1.3 to 1.7 s. The stimulus presentation monitor was placed 75 cm in front of the comfortably seated subject. To keep stimuli in focus yet keep the object/non-object dichotomy unaware, the subject was instructed to respond with a mouse button press depending on his judgment of the drawings as primarily having round contours or primarily having edges. The present analysis is based on the Ob condition thus the size of the data was $I = 64$ channel $\times$ $J = 1024$ time points $\times$ $K = 313$ epochs.

2.4.2 fMRI data

A healthy subject was enrolled after informed consent as approved by the Ethics Committee. The fMRI dataset consisted of 381 volumes and was acquired on a 3T (Siemens Magnetom Trio) scanner using the standard birdcage head coil. Two datasets were collected from a normal subject using an EPI GRE sequence with 40 slices acquired in interleaved order with the following acquisition parameters: Repetition time (TR) 2370 ms, echo time (TE) 30 ms, flip angle (FA) 90 degrees, field of view (FOV) 192×192 mm, 64×64 acquisition matrix. For visualization purposes a high resolution anatomical scan were obtained using a magnetization prepared rapid gradient echo (MPRGE) sequence with 192 sagittal slices and 1 mm isotropic resolution. Additional sequence parameters were as follows: TR $= 15.4$ ms, TE $= 3.93$ ms, flip angle FA $= 9$ degrees, FOV $= 256 \times 256$. In order to demonstrate the usefulness of including shifts in the analysis we employed a simple retinotopic mapping paradigm where the subject was stimulated visually during scanning by an 8 Hz reversing checkerboard (expanding ring and rotating wedge) (Sereno *et al.*, 1995), see Figure 2. This should result in a specific relative delay in the visual areas of the brain according to the movement of the ring and wedge in the visual field. The checkers was scaled approximately by the cortical magnification factor (Slotnick *et al.*, 2001) and each rotation/expansion lasted 30 seconds. Notice, the shifts are over the voxel mode such that the SCP model should be able to captures the paradigm in a single component. The fMRI time series was co-registered to the anatomical scan and realigned using the SPM5 software package `http://www.fil.ion.ucl.ac.uk/spm/software/spm5/`, additionally the time series were filtered prior to analysis to account for various noise sources. Several nuisance contributions (Lund *et al.*, 2006) was projected out using a univariate multiple linear regression model with parameters determined by maximum likelihood estimation. The filter consisted of the following factors: High pass filter by discrete cosine transform basis functions up to a cut off frequency of $1/128$ Hz (18 parameters to account for hardware drift), Volterra expansion of motion parameters including spin-history effects (Friston *et al.*, 1996) (24 parameters), Fourier expansion of the aliased cardiac and respiratory cycles using the RETROICOR method by Glover *et al.* (2000) based on recording of the cardiac cycle by pulse oximetry and

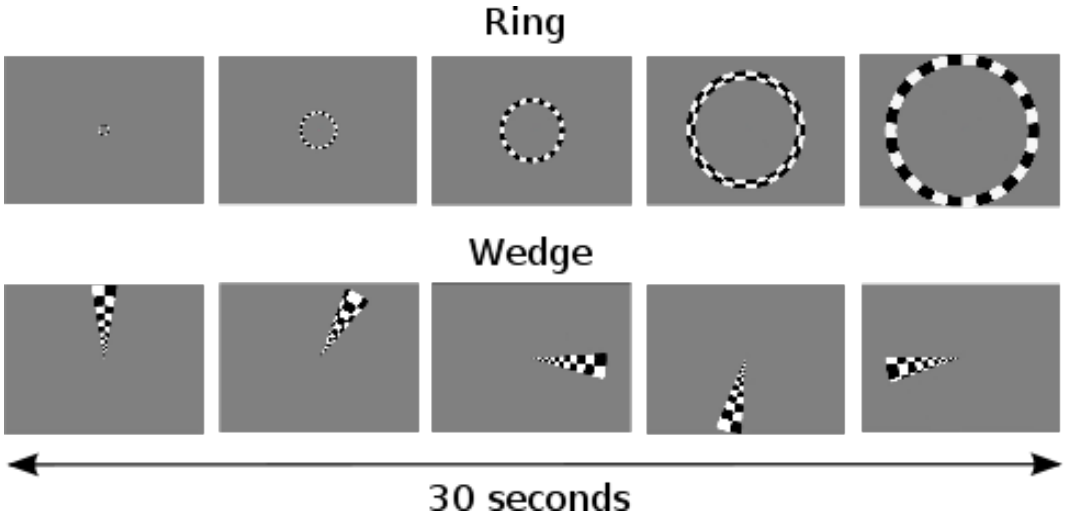


Fig. 2. Illustration of the fMRI paradigm. The top row of images illustrate how a ring of flickering (8 Hz reversal rate) checkers expands in the visual field to generate activation in corresponding locations in the visual cortex. This expansion lasts 30 seconds and is repeated 30 times in order to generate an eccentricity mapping of the visual cortex. Likewise the second row of images illustrate the polar mapping experiment where a wedge of flickering checkers rotates in the visual field.

respiratory cycles using a pneumatic belt (16 parameters), and simple subtraction of the voxel-wise mean (1 parameter). Subsequently the data was divided into epochs according to the stimulus cycle to form a three-way array consisting of $I = 30$ epochs each containing $J = 14$ time points recorded over $K = 48435$ voxels (after masking with a rough brain mask). Due to the fact that the TR does not match the stimulus cycle precisely the input time series within each epoch was shifted (non integer shift) to make the time points of each epoch the same relative to the stimulus onset.

3 Results

The SCP algorithm was first tested on a synthetic EEG data set and then used to analyze the event related EEG and fMRI data described in section 2.4. For the analysis of EEG $\mathbf{A}$ will pertain to spatial activity (i.e., the electrodes), $\mathbf{B}$ to the temporal profile while $\mathbf{C}$ will give the epoch strength. $\mathbf{C}$ is constrained non-negative such that only activities that are similar across epochs are estimated. In the analysis of the fMRI data the roles of $\mathbf{A}$ and $\mathbf{C}$ are switched such that delays occur over the spatial mode. Thus, $\mathbf{A}$ will denote the epoch strength (constrained non-negative to find activities that are similar), $\mathbf{B}$ the temporal profiles and $\mathbf{C}$ will denote the spacial activities (i.e., voxel strengths). $\mathbf{C}$ is also constrained non-negative such that the estimated BOLD response in $\mathbf{B}$ is assumed to have similar temporal profile across

the voxels.

3.1 Synthetic data

A synthetic event related EEG data set of 64 channels, of 1 second of EEG activity sampled at 512 Hz over 105 trials was generated. The generated EEG consisted of a 20 Hz occipital burst, a 12 Hz frontal burst, a 50 Hz constant oscillation most prominent in the central area of the scalp and a 4 Hz slow wave. All signals were uniformly randomly delayed across each trials by ± 100 ms violating a trilinear structure. The components of the synthetic data as well as a decomposition found by regular CP and the proposed SCP algorithm is given in Figure 3. The core consistency diagnostic of the SCP model indicated a 4 component model. For comparison we used the same number of components for the instantaneous CP, despite that the core consistency diagnostic here indicated a 1 component CP model.

3.2 Event related EEG data

When modeling event related data nuisance due to prominent consistent activity over the epochs unrelated to the event such as alpha activity prior and post the event as well as eye blinks can confound the identification of event related components. To capture a strictly event related component in the data we constrained the last of the SCP components in the present analysis to have a fixed delay of zero. The remaining components including shifts should then model consistent confounding effects not phase locked to the event. The core consistency diagnostic indicated that 5 components adequately described the data using this SCP model. Thus, in Figure 4 is given the results obtained fitting a 5 component SCP model (including the instantaneous component) as well as the corresponding instantaneous CP model. The figure also shows the time $\times$ trial ERP-image as well as the evoked potential at the channel of maximal activity. These ERP-images are given for both the raw unshifted data as well as the data shifted according to each of the estimated component delays.

In Figure 5 the estimated visually evoked potentials given by component 5 in the SCP (i.e., the instantaneous component) as well as instantaneous CP is given.

3.3 fMRI data

Traditionally, a retinotopic map of the visual cortex has been constructed by estimating the relative delays in the individual voxels caused by the sweep of the

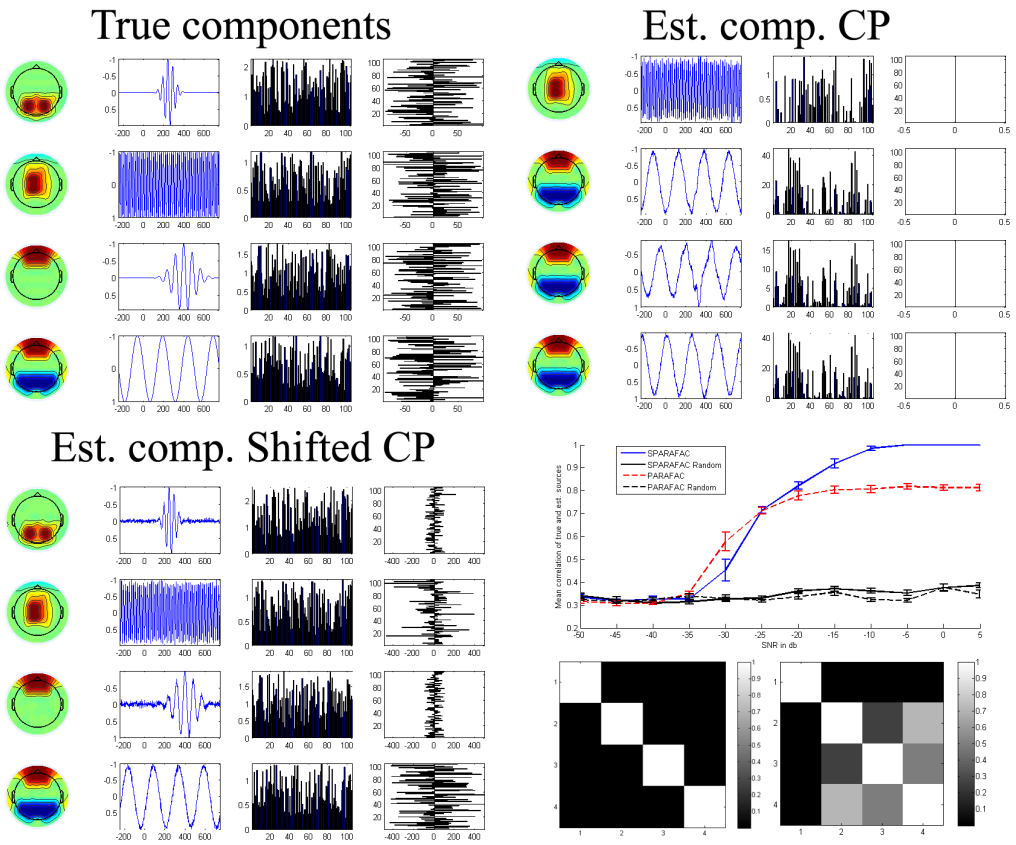


Fig. 3. An analysis of synthetic EEG data. **Top left panel:** The four components forming the synthetic dataset; a 20 Hz occipital burst, a 50 Hz constant oscillation most prominent in the central area of the scalp, a 12 Hz frontal burst and a 4 Hz frontal-parietal slow wave. **Top right panel:** results obtained by a regular CP analysis with a SNR $= -10$ dB. Clearly, the CP model has degenerated due to the violation of trilinearity, thus, forming 3 components of more or less identical scalp maps and component time-series while the strength of each of these components over trials are arbitrarily large counteracting the effects of the other components. **Bottom left panel:** Result obtained using SCP with a SNR $= -10$ dB. Here, all four components are correctly recovered. Notice, the delays for the 50 Hz and 20 Hz components are ambiguous up to a period of the oscillations. **Bottom right panel:** The top graph shows the mean correlation of the estimated sources to the true sources for various SNR, the bottom part the absolute correlation of the components for the SCP model (to the left) and instantaneous CP (to the right). Whereas the CP model fails in identifying the correct components the SCP model is able to identify the sources perfectly down to about a SNR of -10 dB. From the correlation of the components it is seen that whereas the epoch strengths in the SCP model is uncorrelated, strong correlations are found between the second, third and fourth component of the CP model resulting in degenerate solutions.

stimulus in the visual field. Typically these types of paradigms are analyzed using a DFT for each voxel identifying the phase (Sereno *et al.*, 1995). This phase can then be used to identify which part of the visual field the voxels receive their inputs from. As such, delay modeling is crucial in the analysis of these types of

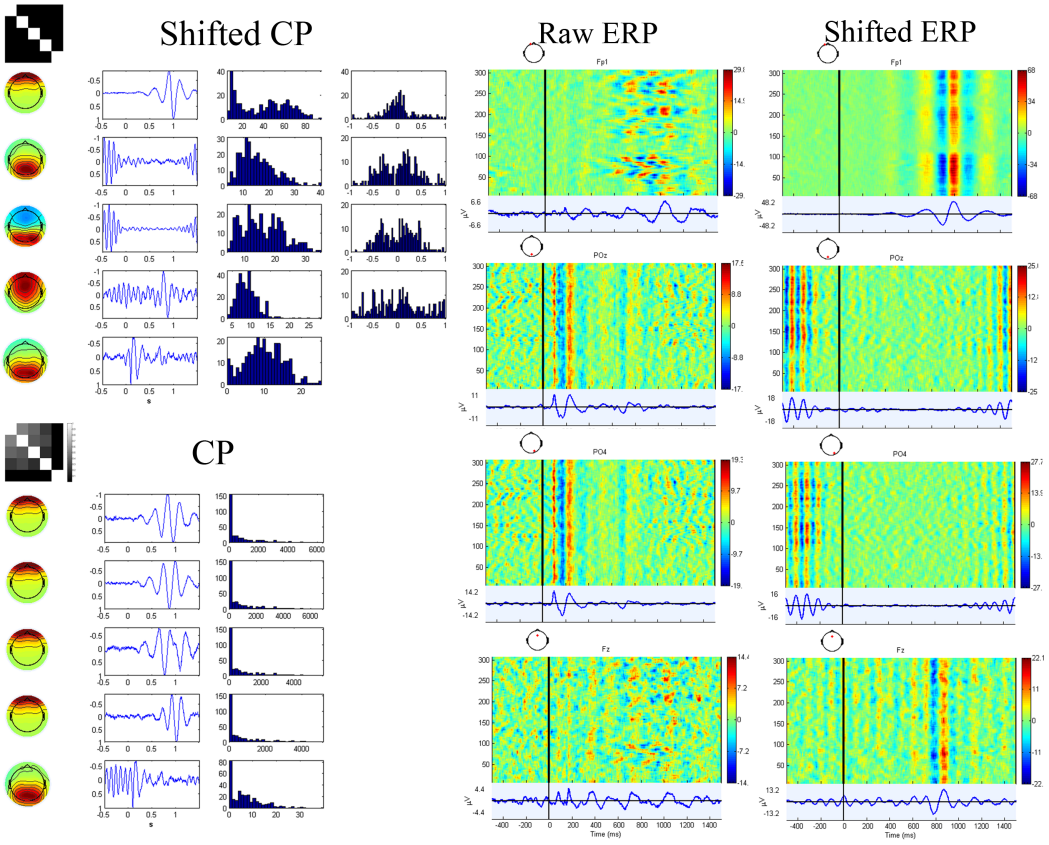


Fig. 4. **Left panel:** (Top) A 5 component shifted CP model of a 64 channel $\times$ 1024 time–points $\times$ 313 trials event related EEG data set as well as the corresponding instantaneous CP analysis (bottom). For each component the spatial map $\mathbf{A}_d$, time series $\mathbf{B}_d$, histogram of trial strengths $\mathbf{C}_d$ and histogram of delays $\boldsymbol{\tau}_d$ are given. The absolute correlation between the various components are shown above the decompositions. Clearly, the instantaneous CP model has found a degenerate solution in which the activity of the eye-blink has been captured in the four first components. Thus whereas the correlation between the factors is very small for the SCP model a degenerate solution is obtained in the instantaneous CP analysis (see correlation matrices in the top left corner of each decomposition). **Right panel:** ERP images and event related potential for the channel having the maximal activity in each of the four shifted components of the SCP model given in the left panel as well as the activity when shifting each trial of the EEG data according to the estimated component delays $\boldsymbol{\tau}_d$ (the ERP images are smoothed with a gaussian window $\sigma = 10$). Whereas the SCP model accounts for 36 % of the variance the instantaneous CP model only accounts for 21 % of the variation in the data. Notice, since the data is high-pass filtered ($>$ 3 Hz) only the high-frequency component of the eye-blink is modeled.

data. The benefits of the present SCP analysis are that the shape of the bold signal is estimated from the data (i.e. not assumed sinusoidal) while amplitude changes over the repeats are modeled. Ideally, the stimuli paradigm for the fMRI data should be modeled by a single component SCP model. Thus, a one component SCP model was estimated. To emphasize the importance of delay modeling we included for

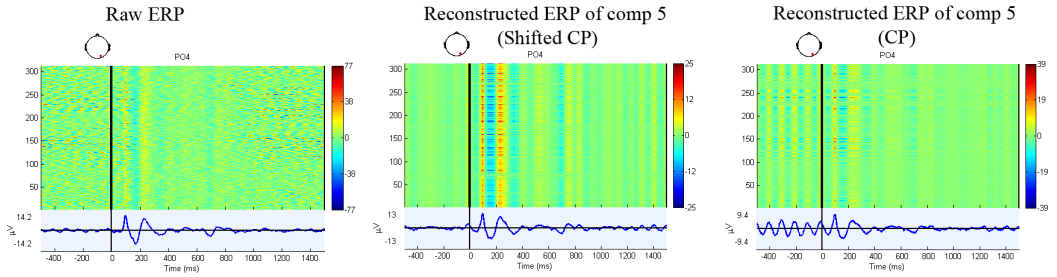


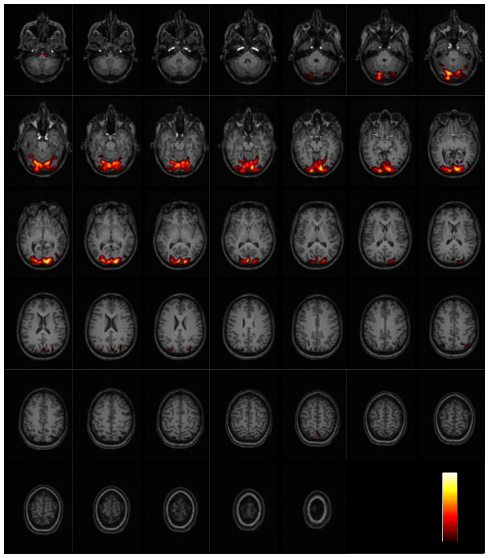
Fig. 5. Investigation of the visual evoked potential (EP) (notice, the ERP images are not smoothed). **Left panel:** EP of the raw EEG data at channel PO4. **Middle panel:** EP found by the SCP model here shown when reconstructing the data from the visual evoked component (component 5) in channel PO4. **Right panel:** EP found by the instantaneous CP model here also shown when reconstructing the data from the visual evoked component obtained (component 5) in channel PO4. Clearly, the estimated visual evoked component of the SCP model has captured more of the EP of the data than the corresponding component of the regular CP analysis. In particular the P100, N200 and P300 complex is easily identified in the shifted CP model whereas it is more impacted by noise in the regular CP model. The latter being more influenced by alpha activity prior to the event. Notice how each trial is weighted by the epoch strengths estimated in the CP models ($\mathbf{C}$) such that trials heavily impacted by noise contributes less in the component identification while the corresponding spatial map of the activity can be inspected in figure 4.

comparison a one component regular instantaneous CP model.

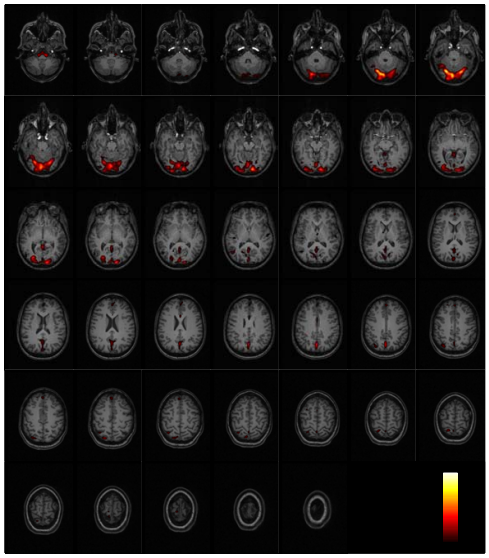
4 Discussion and Conclusion

From the artificial event related EEG data (see Figure 3) it was seen that the SCP model was able to correctly identify the components of the data whereas the corresponding instantaneous CP analysis resulted in degenerate solutions due to the violation of trilinearity in the data. Thus, if delays in data are causing violation of trilinearity a viable approach to avoid degeneracy is extending the CP model to the SCP. Thereby alleviating drawbacks such as slow convergence (the SCP model took between 20–60 iterations in order to converge whereas the regular CP took in the order 200–300 iterations to converge) and poor interpretability of the components (strong cancelation effect present in the instantaneous CP decomposition).

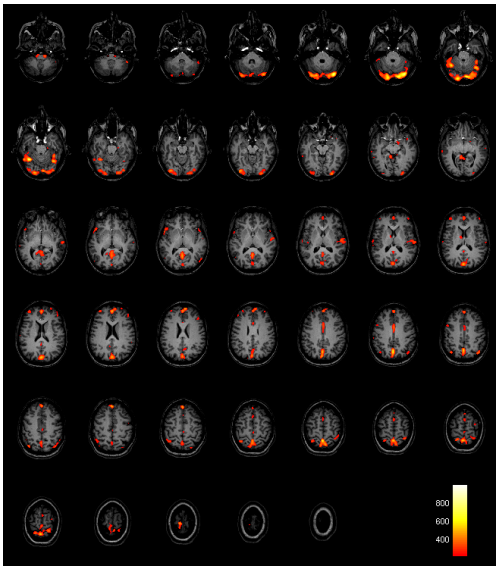
From the real event related EEG data (see Figure 4) 5 components was identified by the SCP model. The first component found corresponds to eye blinks that consistently occur on average about 1 second post the stimulus varying across the epochs ± 200 ms. The time series of component 2 and 3 both resemble alpha activity prior and post the event as seen from the bimodal histogram of the delays. However, the spatial location is slightly more frontal central in the third component. Neither the time-delays nor epoch strengths are correlated thus these two components are not



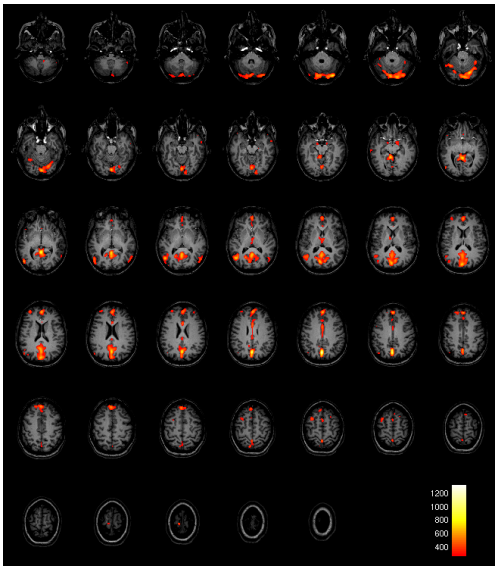
(a) Ring: Est. voxel Act. $\mathbf{C}$ (SCP)



(b) Wedge: Est. voxel Act. $\mathbf{C}$ (SCP)

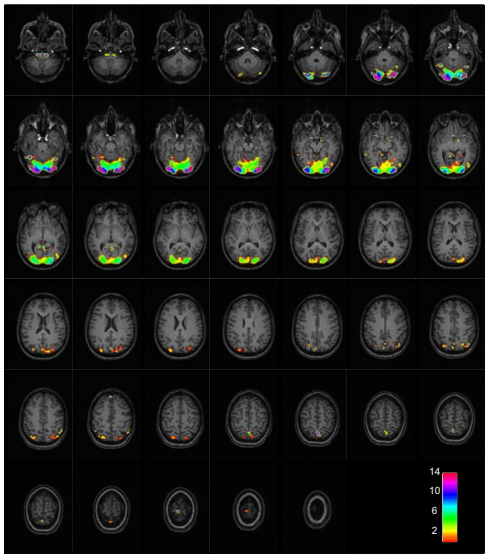


(c) Ring: Est. Voxel Act. $\mathbf{C}$ (CP)

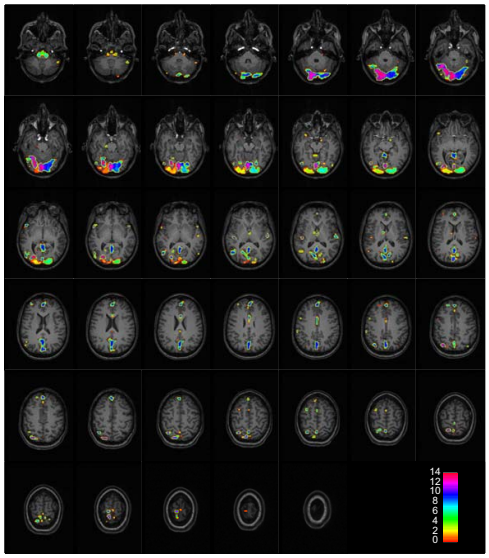


(d) Wedge: Est. Voxel Act. $\mathbf{C}$ (CP)

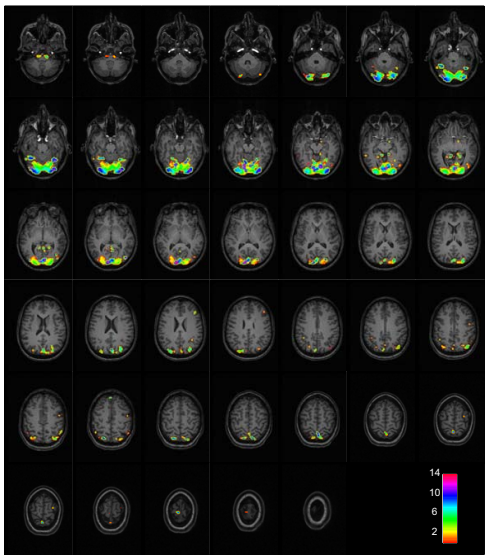
Fig. 6. The component strength ($\mathbf{C}$) over voxels overlayed on the high resolution structural scan for the one component SCP model as well as the corresponding one component instantaneous CP. The map was threshold such that the 5% of the voxels with the largest voxel score $\mathbf{C}$ are shown, a standard Z-transform is not meaningful because $\mathbf{C}$ is constrained non-negative. **Top left panel:** Clearly, the most prominent activity found by SCP on the ring paradigm corresponds well with areas pertaining to visual information processing, i.e. visual cortex. **Top right panel:** Also for the wedge paradigm the most prominent activity for the one component SCP model pertains to visual cortex. **Bottom left and right panel:** Since delay modeling is crucial in the present paradigms the most prominent activity estimated by the one component instantaneous CP model for the ring and wedge paradigms are no longer confined to the visual cortex.



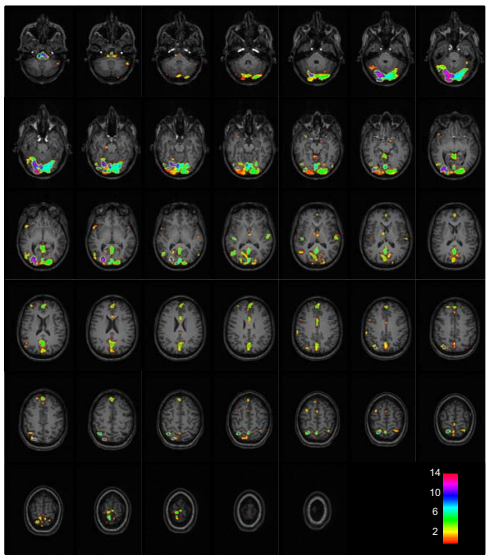
(a) Ring: Est. delays τ (SCP)



(b) Wedge: Est. delays τ (SCP)

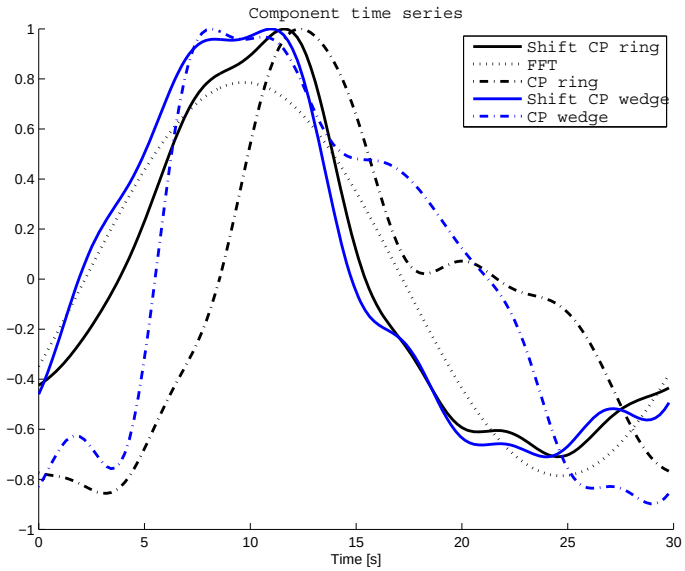


(c) Ring: Est. delays τ (DFT)



(d) Wedge: Est. delays τ (DFT)

Fig. 7. The estimated temporal delays based on the one component SCP model as well as the traditional DFT analysis for the ring and wedge paradigms. **Top left panel:** Delays estimated by SCP for the ring paradigm. Clearly, the delays are symmetric across the two hemispheres.**Top right panel:** Delays estimated by SCP for the wedge paradigm. Clearly, there is a difference in the delays between right and left visual field, i.e. right and left hemisphere. **Bottom left and right panel:** Delays estimated based on the phases obtained by a voxel-wise DFT analysis of the ring and wedge paradigms. Similar symmetries are found as obtained by the SCP, however, the maps are not as smooth as the delay maps found by SCP thus appear somewhat more confounded by noise.



(a) Time profiles SCP

Fig. 8. The extracted component time-series for the ring (black line) and wedge task (blue line) for the SCP model (solid), the corresponding instantaneous CP (dash-dotted) and the sinusoidal time series used by the DFT based approach to estimate the delays (dotted). Clearly, the estimated BOLD signals are more consistent between the two paradigms for the SCP model than the instantaneous CP model while the DFT based approach does not model the actual BOLD pattern but correlates the times series to the sinusoidal time-series corresponding to the stimuli frequency.

degenerated. Finally, the 4th component corresponds to frontal alpha activity while the 5th component models the EP. The importance of event related modulation of alpha and alpha-like oscillations for cognition has been emphasized lately (Basar *et al.*, 2000; Pineda, 2005; Palva and Palva, 2007). Event related de-synchronization of alpha oscillations (as seen in component 2 and 3) has been described in a visual paradigm resembling the presently used (Gruber and Müller, 2006), and in a source location study of visual attention two sources of alpha oscillations were attenuated (Gómez *et al.*, 2006) — an occipital and a right-left occipital-temporal, which is in agreement with the topography of the captured components, component 3 and 2 respectively. The frontal alpha oscillations (as seen in component 4) is suggested to reflect higher cognitive processes (Pineda, 2005). The instantaneous CP did not facilitate a five-component model due to CP-degeneracy. Hence, the first four components were used to span the activity of the eye blinks varying in onset over the trials. The last component pertained to the visually evoked potential, however, contrary to the SCP and regular EP the component was confounded by alpha activity in particular prior to the event, see Figure 5. Thus, not only are the components of the SCP model in accordance with previous findings but the added flexibility of the SCP model account for more of the variation (36% versus 21% using $4 \times 313/[5 \times (64 + 1024 + 313)] = 17.9\%$ more parameters). Furthermore, the SCP model could more easily be interpreted since the decomposition did not degenerate. Finally, the CP model allowed trials dominated by noise to be given

little weight in the average whereas trials where the evoked activity was prominent could be given more weight (see Figure 5). Thus, the model facilitates a weighted average over trials and electrodes when estimating the evoked potential while the components including delays in the SCP can remove prominent systematic nuisance that are not time locked to the event such as eye-blink and alpha activity.

For the fMRI data the SCP model found both the regions relevant for the visual stimuli as well as the signal time course and the delay across voxels pertaining to the stimuli unsupervised. The model further allowed the activity to be more prominent in some epochs than others. The delays obtained had the expected symmetry around the mid-sagittal plane for the ring paradigm while the wedge paradigm resulted in a difference in delays between right and left visual field, i.e. the right and left hemisphere. These delays corresponded to the delays obtained from a traditional voxel-wise DFT analysis (Sereno *et al.*, 1995; Engel *et al.*, 1997; Warnking *et al.*, 2002). DFT assumes the time series for the delay modeling is sinusoidal and constant in strength over the epochs. Thus the benefits of the SCP are that noisefull epochs are given less importance in the estimation of the delays while a more complex pattern of the time series improves the delay estimation. As such, the delays estimated by the SCP model experience a more smooth behavior as would be expected than the DFT based analysis while the expected delay symmetries were more prominent. While the DFT based approach uses 2×48435 the SCP uses $2 \times 48435 + 14 + 30$, i.e. 0.05% more parameters. The estimated BOLD time series found by SCP were consistent across stimuli paradigms and with what is typically reported in the literature (Boynton *et al.*, 1996). Presently we considered stimuli induced delays - however, in general, the SCP model should also be able to capture delays in the hemodynamic response caused by other effects for instance caused by local differences in physiology or sequential information processing in the brain.

Extending the analysis of 2-way data such as *channel* $\times$ *time* averages to *channel* $\times$ *time* $\times$ *trials* the 3-way decomposition is generally unique and does not require additional constraints such as independence which is not necessarily physiologically justified. Furthermore, weighting the components over trials makes it possible to reduce the influence of trials where the presence of the component is weak, for instance due to corruption by noise. Thus the extension of two-way factor analysis to three-way SCP indeed forms a model that can take into account trial specific information as well as decompose the data unambiguously such that interpretability is improved. Finally, the degeneracies occurring when previously modeling the neuroscience data by CP have given strong indication that this type of model is inadequate. By the present extension of the CP model allowing for shifts we have demonstrated that CP-degeneracy is no longer a major concern, thus, shifts seem to be an important cause of violation of trilinearity in the data. Presently, we extended the CP model to include shifts over one mode. The model can be extended to include shifts over additional modes and can also be generalized to data of more modalities than three which naturally arises for instance when including modes such as subjects, conditions or runs (Andersen and Rayens, 2004). Presently, the focus was set

on neuroscience data such as EEG and fMRI, however, the model should be readily applicable to other types of neuroimaging data such as magnetoencephalography (MEG) and positron emission tomography (PET). Presently, the model was used unsupervised, however the model can also be used supervised for instance to find the delays of given, known activities.

Acknowledgements

This work was financially supported by the Lundbeck Foundation, partly through individual founding and partly through the Center for Integrated Molecular Brain Imaging (`www.cimbi.dk`). Further funding was supplied by the Gangsted Foundation, the Novo Nordic Foundation and the Danish Research Council. L.-H. Lim gratefully acknowledges the warm hospitality of the Intelligent Signal Processing group at DTU Denmark during his visit.

References

Acar, E., Aykut-Bingol, C., Bingol, H., Bro, R., and Yener, B. (2007). Multiway analysis of epilepsy tensors. *Bioinformatics*, **23**, 10–18.

Andersen, A.H. and Rayens, W.S. (2004). Structure-seeking multilinear methods for the analysis of fMRI data. *Neuroimage*, **22**, 728–739.

Basar, E., Basar-Eroglu, C., Karakas, S., and Schurmann, M. (2000). Brain oscillations in perception and memory. *Int. J. Psychophysiol.*, **35**, 95–124.

Beckmann, C. and Smith, S. (2005). Tensorial extensions of independent component analysis for multisubject fMRI analysis. *Neuroimage*, **25**, 294–311.

Bell, A.J. and Sejnowski, T. J. (1995). An information maximization approach to blind source separation and blind deconvolution. *Neural Comput.*, **7**, 1129–1159.

Boynton, G., Engel, S., Glover, G., and Heeger, D. (1996). Linear systems analysis of functional magnetic resonance imaging in human V1. *J. Neurosci.*, **16**, 4207–4221.

Bro, R. and Andersson, C.A. (2000). The n-way toolbox for MATLAB. *Chemometrics Intelligent Lab. Syst.*, **52**, 1–4.

Bro, R. and de Jong, S. (1997). A fast non-negativity-constrained least squares algorithm. *J. Chemometrics*, **11** (5), 393–401.

Bro, R. and Kiers, H.A.L. (2003). A new efficient method for determining the number of components in PARAFAC models. *J. Chemometrics*, **17** (5), 274–286.

Buxton, R. B., Wong, E. C., and Frank, L. R. (1998). Dynamics of blood flow and oxygenation changes during brain activation: the balloon model. *Magn. Reson. Med*, **39** (6), 855–864.

Carroll, J.D. and Chang, J.J. (1970). Analysis of individual differences in multidi-

mensional scaling via an N-way generalization of "Eckart-Young" decomposition. *Psychometrika*, **35**, 283–319.

de Silva, V. and Lim, L.-H. (2008). Tensor rank and the ill-posedness of the best low-rank approximation problem. *SIAM J. Matrix Anal. Appl.*, to appear.

De Vos, M., Vergult, A., De Lathauwer, L., De Clercq, W., Van Huffel, S., Dupont, P., Palmini, A., and Van Paesschen, W. (2007). Canonical decomposition of ictal scalp EEG reliably detects the seizure onset zone. *Neuroimage*, **37** (3), 844–854.

Delorme, A. and 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.

Donchin, E. and Heffley, E. (1978). Multivariate analysis of event-related potential data: A tutorial review. *Multidisciplinary perspectives in event-related brain potential research*, 555–572.

Dyrholm, M., Christoforou, C., and Parra, L.C. (2007). Bilinear discriminant component analysis. *J. Mach. Learn. Res.*, **8**, 1097–1111.

Emile, B. and Comon, P. (1998). Estimation of time delays between unknown colored signals. *Signal Process.*, **68** (1), 93–100.

Engel, S., Glover, G., and Wandell, B. (1997). Retinotopic organization in human visual cortex and the spatial precision of functional MRI. *Cereb Cortex.*, **7** (2), 181–192.

Field, A.S. and Graupe, D. (1991). Topographic component (parallel factor) analysis of multichannel evoked potentials: Practical issues in trilinear spatiotemporal decomposition. *Brain Topogr.*, **3** (4), 407–423.

Friston, K.J., Williams, S., Howard, R., Frackowiak, R.S., and Turner, R. (1996). Movement-related effects in fMRI time-series. *Magn. Reson. Med.*, **35** (3), 346–355.

Glover, G.H., Li, T.Q., and Ress, D. (2000). Image-based method for retrospective correction of physiological motion effects in fMRI: RETROICOR. *Magn. Reson. Med.*, **44** (1), 162–167.

Gómez, C., Marco-Pallarés, J., and Grau, C. (2006). Location of brain rhythms and their modulation by preparatory attention estimated by current density. *Brain Res.*, **1107**, 151–160.

Gruber, T. and Müller, M. (2006). Oscillatory brain activity in the human EEG during indirect and direct memory tasks. *Brain Res.*, **1097**, 194–204.

Harshman, R., Hong, S., and Lundy, M. (2003a). Shifted factor analysis part I: Models and properties. *J. Chemometrics*, **17**, 363–378.

Harshman, R., Hong, S., and Lundy, M. (2003b). Shifted factor analysis part II: Algorithms. *J. Chemometrics*, **17**, 379–388.

Harshman, R.A. (1970). Foundations of the PARAFAC procedure: Models and conditions for an "explanatory" multi-modal factor analysis. *UCLA Working Papers in Phonetics*, **16**, 1–84.

Harshman, R.A. (1972). Determination and proof of minimum uniqueness conditions for PARAFAC1. *UCLA Working Papers in Phonetics*, **22**, 111–117.

Harshman, R.A. and Lundy, M.E. (1984). Data preprocessing and the extended PARAFAC model. In: Law, H.G., Snyder, Jr., C.W., Hattie, J.A., and McDonald,

R.P. (Eds.), *Research Methods for Multimode Data Analysis,* 216–281, Praeger.

Herrmann, C., Munk, M., and Engel, A. (2004a). Cognitive functions of gamma-band activity: memory match and utilization. *Trends Cogn. Sci.*, **8** (8), 347–355.

Herrmann, C., Lenz, D., Junge, S., Busch, N., and Maess, B. (2004b). Memory-matches evoke human gamma-responses. *BMC Neurosci.*, **5** (13).

Hong, S. and Harshman, R.A. (2003). Shifted factor analysis part III: N-way generalization and application. *J. Chemometrics*, **17**, 389–399.

Knuth, K., Shah, A., Truccolo, W., Ding, M., Bressler, S., and Schroeder, C. (2006). Differentially variable component analysis: Identifying multiple evoked components using trial-to-trial variability. *J. Neurophysiol.*, **95**, 3257–3276.

Kruskal, J. (1977). Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics. *Linear Algebra Appl.*, **18**, 95–138.

Lund, T., Madsen, K., Sidaros, K., Luo, W., and Nichols, T. (2006). Non-white noise in fMRI: does modelling have an impact? *Neuroimage*, **29**, 54–66.

Lundy, M.E., Harshman, R.A., and Kruskal, J.B. (1989). A two-stage procedure incorporating good features of both trilinear and quadrilinear methods. In R. Coppi and S. Bolasco (Eds.), *Multiway data analysis*, 123–130*,* Elsevier.

Makeig, S., Bell, A.J., Jung, T.-P., and Sejnowski, T.J. (1996). Independent component analysis of electroencephalographic data. In D.S. Touretzky, M.C. Mozer, and M.E. Hasselmo (Eds.), *Adv. Neural Inform. Process. Syst.*, **8**, 145–151. MIT Press.

Makeig, S., Jung, T.-P., and Bell, A., Ghahremani, D., and Sejnowski, T. (1997). Blind separation of auditory event-related brain responses into independent components. *Proc. Natl. Acad. Sci.*, **94**, 10979–10984.

Makeig, S., Westerfield, M., Jung, T.-P., Enghoff, S., Townsend, J., Courchesne, E., and Sejnowski, T.J. (2002). Dynamic brain sources of visual evoked responses. *Science*, **295** (5555), 690–694.

McKeown, M.J., Jung, T.P., Makeig, S., Brown, G., Kindermann, S.S., Lee, T.W., and Sejnowski, T.J. (1998). Spatially independent activity patterns in functional MRI data during the stroop color-naming task. *Proc. Natl. Acad. Sci.*, **95** (3), 803–810.

McKeown, M. J., Hansen, L.K., and Sejnowski, T.J. (2003). Independent component analysis of functional MRI: what is signal and what is noise? *Curr. Opin. Neurobiol.*, **13** (5), 620–629.

Miwakeichi, F., Martínez-Montes, E., Valdés-Sosa, P.A., Nishiyama, N., Mizuharam, H., and Yamaguchi, Y. (2004). Decomposing EEG data into space time frequency components using parallel factor analysis. *Neuroimage*, **22** (3), 1035–1045.

Möcks, J. (1988). Topographic components model for event-related potentials and some biophysical considerations. *IEEE Trans. Biomed. Eng.*, **35**, 482–484.

Mørup, M., Hansen, L.K., Herrmann, C.S., Parnas, J., and Arnfred, S.M. (2006). Parallel factor analysis as an exploratory tool for wavelet transformed event-related EEG. *Neuroimage*, **29** (3), 938–947.

Mørup, M., Madsen, K.H., and Hansen, L.K. (2007a). Shifted independent compo-

nent analysis. *Indep. Component Anal.*, to appear.

Mørup, M., Madsen, K.H., and Hansen, L.K. (2007b). Shifted non-negative matrix factorization. *Mach. Learn. Signal Process.*, to appear.

Palva, S. and Palva, J. (2007). New vistas for alpha-frequency band oscillations. *Trends Neurosci.*, **30**, 150–158.

Pineda, J. (2005). The functional significance of mu rhythms: translating "seeing" and "hearing" into "doing". *Brain Res. Rev.*, **50**, 57–68.

Sereno, M.I., Dale, A.M., Reppas, J.B., Kwong, K.K., Belliveau, J.W., Brady, T.J., Rosen, B.R., and Tootell, R.B. (1995). Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science*, **268** (5212), 889–893.

Slotnick, S. D., Klein, S. A., Carney, T., and Sutter, E. E. (2001). Electrophysiological estimate of human cortical magnification. *Clin. Neurophysiol.*, **112** (7), 1349–1356.

Stegeman, A. (2007). Degeneracy in CANDECOMP/PARAFAC and INDSCAL explained for several three-sliced arrays with a two-valued typical rank. *Psychometrika*, **72** (4), 601–619.

Tomasi, G. (2006). *Practical and computational aspects in chemometric data analysis*. Ph.D. thesis, The Royal Veterinary and Agricultural University, Frederiksberg, Denmark.

Torkkola, K. (1996). Blind separation of delayed sources based on information-maximization. *Int. Conf. Acoustics, Speech, Signal Process.*, **6**, 3509–3512.

Truccolo, W., Knuth, K., Shah, A., Bressler, S., Schroeder, C., and Ding, M. (2003). Estimation of single-trial multicomponent ERPs: differentially variable component analysis (dVCA). *Biol. Cybern.*, **89**, 426–438.

Wang, K., Begleiter, H., and Porjesz, B. (2000). Trilinear modeling of event-related potentials. *Brain Topogr.*, **12**(4), 263–271.

Warnking, J., Dojat, M., Guérin-Dugué, A., Delon-Martin, C., Olympieff, S., Richard, N., Chéhikian, A., and Segebarth, C. (2002). fMRI retinotopic mapping—step by step. *Neuroimage*, **17**, 1665–1683.

Yeredor, A. (2003). Time-delay estimation in mixtures. *Int. Conf. Acoustics, Speech, Signal Process.*, **5**, 237–240.