ELSEVIER

NeuroImage

www.elsevier.com/locate/ynimg
NeuroImage 22 (2004) 728–739

# Structure-seeking multilinear methods for the analysis of fMRI data

Anders H. Andersen[a,*] and William S. Rayens[b]

[a] *Department of Anatomy and Neurobiology, Magnetic Resonance Imaging and Spectroscopy Center, University of Kentucky, Lexington, KY 40536-0098, USA*
[b] *Department of Statistics, University of Kentucky, Lexington, KY 40536-0098, USA*

Received 22 September 2003; revised 13 February 2004; accepted 19 February 2004

**In comprehensive fMRI studies of brain function, the data structures often contain higher-order ways such as trial, task condition, subject, and group in addition to the intrinsic dimensions of time and space. While multivariate bilinear methods such as principal component analysis (PCA) have been used successfully for extracting information about spatial and temporal features in data from a single fMRI run, the need to unfold higher-order data sets into bilinear arrays has led to decompositions that are nonunique and to the loss of multiway linkages and interactions present in the data. These additional dimensions or ways can be retained in multilinear models to produce structures that are unique and which admit interpretations that are neurophysiologically meaningful. Multiway analysis of fMRI data from multiple runs of a bilateral finger-tapping paradigm was performed using the parallel factor (PARAFAC) model. A trilinear model was fitted to a data cube of dimensions voxels by time by run. Similarly, a quadrilinear model was fitted to a higher-way structure of dimensions voxels by time by trial by run. The spatial and temporal response components were extracted and validated by comparison to results from traditional SVD/PCA analyses based on scenarios of unfolding into lower-order bilinear structures.**


*Keywords:* fMRI; Multilinear method; Analysis

## Introduction

Multilinear modeling is a widely applicable paradigm that could well be a key informatics tool in the analysis of data from functional neuroimaging experiments using MRI as well as other imaging modalities such as PET. This paradigm simply recognizes that in functional imaging studies, the data are acquired from what chemists would call a "higher-order" instrument. For example, in the setting of functional MRI (fMRI) scanning for the mapping of brain activation, the data acquired represent repeated measurements over time of the spatial MR image intensity distribution across a brain volume. The image intensity is sensitive to the local changes in blood flow that accompany primary sensory and motor as well as higher cognitive brain processes. By systematic raster scanning of the spatial-domain data, mimicking the form in which the data are naturally stored in a linear fashion on computer disk media, the measurements can be interpreted as entries in a two-dimensional ("bilinear") unfolded data matrix structure. The true matrix dimensions are "voxels" (volume elements) and "time" samples. A typical data set from an fMRI scan involving human subjects may contain 64 × 64 × 40 voxels sampled at 128 consecutive time instants, producing a single unfolded *matrix* observation with shape 163,840 voxels by 128 time points. Multiple scans on a given subject in the presence of either the same stimulus or across different experimental conditions and co-registered spatially, so as to align voxels across scans, produce an intrinsic higher-dimensional "cube" of data. Increasingly often, complex investigations of human brain function generate data arrays of order three and higher.

Furthermore, a comprehensive neuroimaging study generally involves scans from a large number of subjects so that inferences pertaining to the underlying population can be made (Friston et al., 1999a). Scans from multiple subjects are routinely normalized geometrically by affine transformations to conform to standardized stereotaxic voxel coordinates (Talairach and Tournoux, 1988), thereby adding yet another dimension to the data array. Data sizes are typically quite large as well, about 5–10 GB. Traditionally, the analysis and data reduction have been carried out separately for each brain voxel by using univariate linear regression to "collapse" across time and across subjects with the objective of generating statistical parametric maps of the spatial distribution of brain activation for a particular task or stimulation paradigm (Frackowiak et al., 1997). Such methods are very efficient for analyzing regionally specific effects and for the detection of focal brain activation. However, they fall short in the analysis of functionally connected brain systems ("functional organization"), since information on the multivariate and spatially correlated structure of the data is ignored.

Spatial smoothing has been used commonly to increase the signal-to-noise ratio and as an ad hoc approach for enhancing the depiction of spatially connected voxels (see for instance Kesler-West et al., 2001; Kiebel et al., 1999; Smith et al., 1999). Additionally, cluster analysis methods or Gaussian field theory have been invoked to more formally characterize the spatial extent of brain activation patterns (Friston et al., 1994; Worsley, 1994). However, there is still much to learn on how to jointly characterize

* Corresponding author. Department of Anatomy and Neurobiology, Magnetic Resonance Imaging and Spectroscopy Center, University of Kentucky Medical Center, Room 37 MRISC Building (Davis Mills), 800 Rose Street, Lexington, KY 40536-0098. Fax +1-859-323-1068.
*E-mail address:* anders@mri.uky.edu (A.H. Andersen).
**Available online on ScienceDirect (www.sciencedirect.com.)**

1053-8119/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2004.02.026

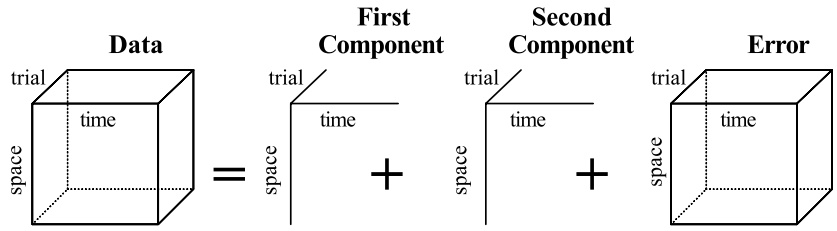


Fig. 1. Decomposing a data cube into a signal matrix composed of two components, each of which consists of three profiles and an error term.

the strength, spatial extent, and functional connectivity of brain activation patterns along with the dependence of these features on subject and task (Friston et al., 1999c). Also, since the data are most often collapsed across subjects within a group by averaging (Fox et al., 1988) or by conjunction (Friston et al., 1999b), linkage information about subject-specific effects and inter-subject heterogeneity is inherently lost.

Likewise, task-specific effects from scans acquired under multiple experimental conditions are difficult to analyze. Some of these issues can be addressed by careful design of the experimental paradigms using for instance hierarchical designs (Martin et al., 1996) or cognitive conjunctions (Price and Friston, 1996). Univariate approaches to data post-processing by voxel-wise one-way ANOVA with tasks as factor levels tend to have a random salt- and-pepper appearance and exhibit very little if any structure due to their failure to capitalize on the spatially extended and correlated nature of the brain activation patterns for the separate tasks.

Multivariate approaches such as principal components analysis (PCA) and the related singular value decomposition (SVD) have been used by several groups as data-driven methods for eliciting information about intrinsic structure, both spatial and temporal, in data sets from fMRI and other neuroimaging modalities (Andersen et al., 1999, 2002; Benali et al., 1995; Friston et al., 1996a; Strother et al., 1995a; Worsley et al., 1997; Zhang et al., 2000). Many of these works have been based on some form of collapsing of the inherently higher-way data sets into two-way bilinear structures. In the original work by Friston et al. (1993) in the context of PET, the data were initially averaged across subjects to create a two-way structure of dimension "voxels by number of repeated trials within subjects." A different variation of PCA, referred to in the neuroimaging literature as the "scaled subprofile model," unfolds an inherently three-way data structure to yield by concatenation an array of dimension "voxels by (subject times number of repeated trials within subjects)" (Hansen et al., 1999; Lautrup et al., 1995; Strother et al., 1995b). This particular scheme of unfolding allows for a partitioning of the total PC variance into contributions from within- and between-subject variability, thus retaining some, but not all, of the information pertaining to the heterogeneity across subjects. A particular attraction of PCA/SVD-based models has been the orthogonality property to where components in the expansion can be computed independently of one another irrespective of the chosen model order.

In summary, data collected from neuroimaging experimental designs are inherently higher-way data. Univariate analyses or bilinear methods are not generally capable of adequately capturing nor uncovering the interrelated and nested structures in these higher-way data, even when the data are reduced to two-way through collapsing in some manner. It has been well established in the multiway literature that forcing a bilinear analysis on genuinely higher-way data can lead to models that are too complex, lack predictive power, and are impossible or very difficult to interpret owing to confounding across collapsed "ways." Most importantly, the multiway linkages are lost in the modeling process and the resulting structures are not unique due to an inherent invariance under rotation. As will be mentioned below, such linkages are preserved in multilinear analyses and the structures that emerge are unique (see e.g., Andersen et al., 2000; Bro, 1997; Field and Graupe, 1991; Möcks, 1988; Pham and Möcks, 1992; Sidiropoulos and Bro, 2000; Smilde, 1992; Smilde and Doornbos, 1992; Wang et al., 2000).

In the present study, we have applied multilinear models to intrinsically higher-way data structures from a functional neuroimaging experiment. Multiway analyses of fMRI data from multiple runs of a bilateral finger-tapping paradigm were performed using the PARAFAC (parallel factor) model. A trilinear model was fitted to a data cube of dimensions voxels by time by run. Similarly, a quadrilinear model was fitted to a higher-way structure of dimensions voxels by time by trial by run. Subsequently, the spatial and temporal response components were compared with results from traditional SVD/PCA analyses based on scenarios of unfolding into bilinear structures.

## Theoretical background

### *Multilinear models*

Extensions of and analogies to the well-known bilinear paradigms have created a substantial literature on so-called multiway

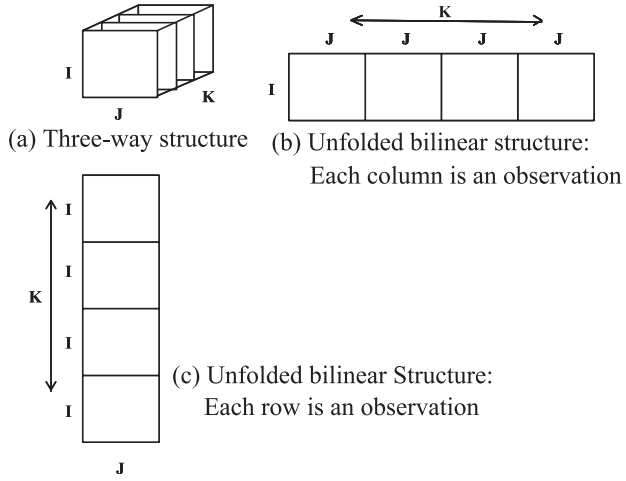


Fig. 2. Illustrating the unfolding of a three-way structure (a) of dimensions $I \times J \times K$ into two-way bilinear structures by either (b) arranging side by side creating a matrix of dimensions $I \times JK$ or by (c) stacking to create a matrix of dimensions $IK \times J$.

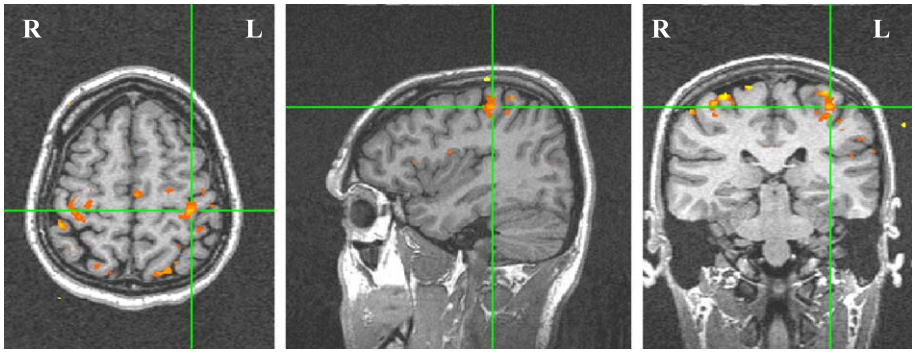


Fig. 3. Illustrating the activation pattern from one run of a bilateral finger tapping experiment in a single subject ( $P < 0.05$, uncorrected). The activation map has been interpolated for purpose of the display, but no spatial smoothing was used. The green lines indicate the position of the corresponding orthogonal sections.

structure-seeking methods. The conceptual similarity to bilinear methods is indicated in the schematic below (Fig. 1), wherein a cube of data is decomposed into two components, each of which consists of three "profiles." Taken as triples in an outer product, the profiles are assumed to characterize the cube. These techniques originated in psychology and continue to be used widely in both that field as well as within chemistry (see e.g., Bro, 1997, 1999; Bro and Heimdal, 1996; Burdick, 1995; Carroll and Chang, 1970; Harshman, 1972; Harshman and Lundy, 1984a,b; Kroonenberg, 1983, 1989; Kruskal, 1984, 1989; Leurgans and Ross, 1992; Mitchell and Burdick, 1994; Rayens and Mitchell, 1997; Sanchez and Kowalski, 1988, 1990; Sidiropoulos and Bro, 2000; Tucker, 1966; Wold et al., 1987).

Multilinear models have been described in the brain mapping literature as well. In particular, trilinear models have been proposed in the analysis of multichannel-evoked potentials in EEG data (Field and Graupe, 1991; Möcks, 1988; Pham and Möcks, 1992; Wang et al., 2000, 2001). In this context, these tools have been called topographic component (TC) models. Although the field is far from unified, with several different multiway models and even different methods of implementing those models, the literature is growing in statistical sophistication and the successes of many applications are undeniable.

*Mathematical notation*

Tensor algebra is needed to fully describe the structure of multiway models. So that the presentation might stay brief and the notation tractable, we will present our technical overview in the context of trilinear methods. Extensions to higher-way arrays are straightforward. Notation from Burdick (1995) is used in the following.

**Definition 1**

- Let $\boldsymbol{x}$ be in $R^I$ and $\boldsymbol{y}$ in $R^J$. A *tensor product* of column vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ is defined as the outer product $\boldsymbol{x} \otimes \boldsymbol{y} = \boldsymbol{x}\boldsymbol{y}^t$, where t denotes the transpose.
- Let $\boldsymbol{z} = (z_k)$ be a vector in $R^K$ and $\boldsymbol{X}$ be an $I$ by $J$ matrix. A *tensor product* of $\boldsymbol{X}$ and $\boldsymbol{z}$ is given by $\boldsymbol{X} \otimes \boldsymbol{z} = (z_1\boldsymbol{X}\ z_2\boldsymbol{X}. . .z_K\boldsymbol{X})_{I \times JK}$.

Typically, a bilinear errors-in-variables model employed to extract structure from $\mathbf{A}_{I \times J}$ has the form of $\mathbf{A}_{I \times J} = \mathbf{S}_{I \times J} + \mathbf{N}_{I \times J}$, which can be interpreted as a signal matrix added to a noise or error matrix. The issue becomes one of how you decide to model the structure in the signal matrix. The typical bilinear approach assumes that there exist vectors $\{\boldsymbol{x}_r\} \subset R^I$, $\{\boldsymbol{y}_r\} \subset R^J$ such that the signal matrix $\mathbf{S}$ can be modeled as a tensor product or outer product expansion $\mathbf{S} = \Sigma_r \boldsymbol{x}_r \otimes \boldsymbol{y}_r$. Of course, this approach suffers from the well-known lack of uniqueness in that there are infinitely many vectors $\boldsymbol{x}_r$ and $\boldsymbol{y}_r$ that can describe the bilinear structure or matrix $\mathbf{S}$ equally well from the point of view of fit. Hence, it becomes very difficult to interpret $\boldsymbol{x}_r$ and $\boldsymbol{y}_r$. This problem is popularly called the rotation problem. When the model is extended to higher-way data structures, the so-called PARAFAC (parallel factor) model emerges (Harshman and Lundy, 1984a,b).


Fig. 4. Illustrating the unique spatial modes from trilinear modeling of a 635 × 110 × 2 data cube. The image gray scale goes from white = max. positive to black = min. negative values.

**Definition 2.** The *PARAFAC* model presumes there exist vectors $\{\boldsymbol{x}_r\} \subset \mathrm{R}^I$, $\{\boldsymbol{y}_r\} \subset \mathrm{R}^J$ and $\{\boldsymbol{z}_r\} \subset \mathrm{R}^K$ such that $\mathbf{S} = \Sigma_r^R \boldsymbol{x}_r \otimes \boldsymbol{y}_r \otimes \boldsymbol{z}_r$, for some appropriate number of tensor products, $R$, which is often called the "rank" of the array.

**Definition 3.** Let $\mathbf{A}$ be an $I$ by $J$ by $K$ array with coordinates $a_{ijk}$. The *Mode X fibers* of $\mathbf{A}$ are the vectors $\boldsymbol{a}_{\cdot jk}$ obtained from $\mathbf{A}$ by fixing the indices $j$ and $k$ and letting the index $i$ vary. Similar for *Mode Y* and *Mode Z fibers*.

**Definition 4.** Let $\mathbf{A}$ be an $I$ by $J$ by $K$ array with coordinates $a_{ijk}$. The *Mode X slabs* of $\mathbf{A}$ are the $J$ by $K$ matrices obtained as subarrays $a_{i\cdot\cdot}$ of $\mathbf{A}$ by fixing the $i$ index and letting $j$ and $k$ vary. Similar for *Mode Y* and *Mode Z slabs*.

Unlike PCA and a host of other bilinear methods, PARAFAC components are not orthogonal and cannot be derived sequentially without causing a potentially serious reduction in model fit. Hence, a four-component model cannot be found by just employing the first four components of a five-component model. This is viewed by some as a desirable aspect of PARAFAC since components now are freer to reflect true underlying structure. Regardless, this makes the determination of the right number of components a more difficult task than with bilinear methods. Practically speaking, however, similar structures are likely to be found in models estimated to have a similar number of components and, hence, appropriately broad-brush conclusions typically will not change. When the purpose of the analysis is focused on the component scores—say, for use in a subsequent regression—then the issue of orthogonality is not important, computational issues aside. If it is necessary to require the component loadings to be orthogonal, and goodness-of-fit is not an issue, then standard algorithms for fitting PARAFAC allow this (e.g., see Andersson and Bro, 2000).

### *Unfolding of three-way arrays to bilinear structures*

When three-way arrays are unfolded, a choice of mode is made and the slabs (subarrays) for that mode are then extracted and arranged side by side or stacked. For instance, $\mathbf{A}$ may be unfolded from an $I \times J \times K$ array to yield a two-way bilinear $I \times JK$ array by extracting the $K$ individual *Mode Z* slabs and juxtaposing them side by side (Fig. 2b) or to a bilinear $IK \times J$ array by stacking (Fig. 2c). It is not hard to see that the trilinear PARAFAC model amounts to a presumption that there is an underlying latent structure common to all of the $Z$-slabs, but invoked in proportions that are $Z$-slab-specific. This is similar to the original Carroll and Chang (1970) INDSCAL model for multidimensional scaling wherein multiple subjects are modeled as having the same so-called group map, but each individual reshapes the dimensions of that map according to these proportions.

In terms of fMRI data sets where $I$, $J$, and $K$ may represent the dimensions of space, time, and trials, respectively, the spatiotemporal patterns in a multilinear model will be common across the $K$ trials but possibly of different amplitude. In the context of bilinear models from unfolded data, however, the spatiotemporal patterns themselves may vary arbitrarily from trial to trial.

Likewise, it is well known that the rank of the bilinear array resulting from unfolding a trilinear array is always less than the rank of the trilinear array, regardless of the mode used to do the unfolding. Hence, one can say that, in this sense, a bilinear model can always be expected to *fit* the unfolded data better than a trilinear model can fit the original array. This is a deceptive virtue, however, since the issue typically is not goodness of fit, but rather the extent of overfit.

Table 1
Amplitude-scaled vector elements $\boldsymbol{c}_r\boldsymbol{z}_{kr}$ for $k$ = 1, 2 (runs) and $r$ = 1, 2, 3 (component number) in a trilinear model

| | Component one | Component two | Component three |
|---|---|---|---|
| 1st run | 0.6373 | 0.8932 | 0.9053 |
| 2nd run | 0.8223 | −0.2683 | 0.2082 |

### *Algorithms*

The PARAFAC multilinear model is usually implemented in one of two ways: using eigenbased methods (Leurgans and Ross,

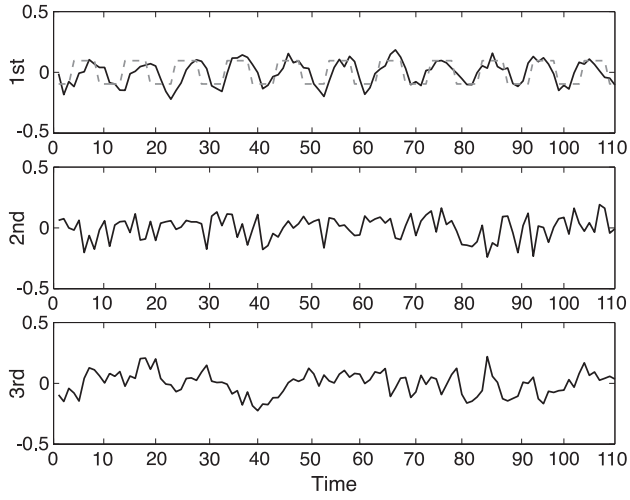


Fig. 5. Illustrating the unit-norm temporal modes from trilinear modeling of a 635 × 110 × 2 data cube.

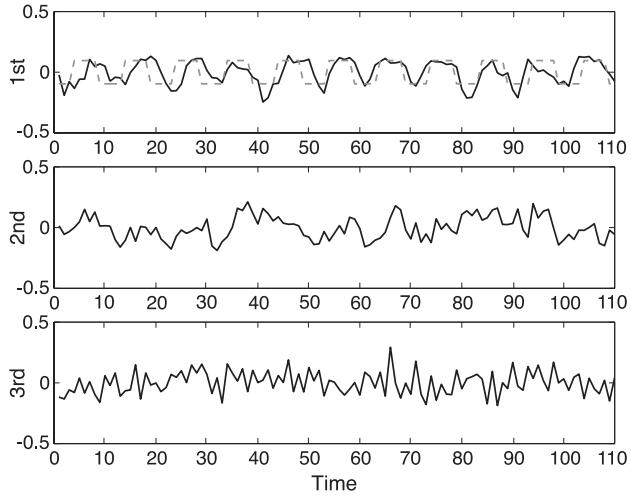


Fig. 6. Illustrating the unit-norm temporal modes from SVD modeling of an unfolded 1270 × 110 bilinear data array. (The amount of variance explained is 3.39%, 2.65%, and 2.22%, respectively.)

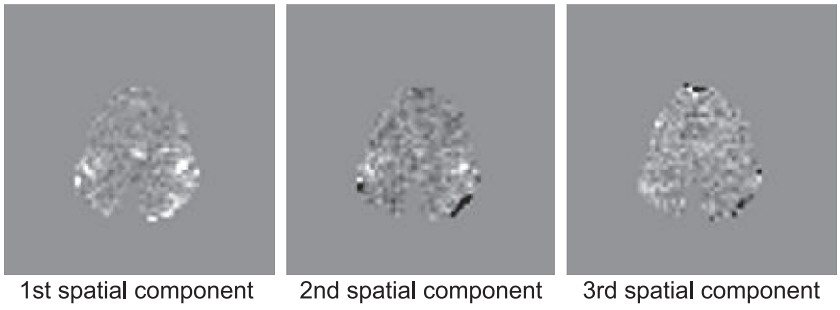


Fig. 7. Illustrating the spatial score maps from SVD modeling of a 1270 × 110 bilinear data array.

1992; Sanchez and Kowalski, 1988, 1990) or by using an alternating least squares routine, unfortunately also called PARAFAC (Appellof and Davidson, 1981; Harshman, 1972; Harshman and Lundy, 1984a,b; Rayens and Mitchell, 1997). The PARAFAC routine exploits the conditional linearity of the model bearing the same name. Two of the so-called factor matrices, say **X** and **Y**, are fixed and linear regression is used to obtain the other factor matrix **Z**. Then **Y** and **Z** are fixed and **X** is estimated, and similar for **Y**. This procedure of alternating fits continues iteratively until some convergence criterion is met. Recent work by Bro and De Jong (1997) and by Bro and Andersson (1998) and others, directed toward speeding up the convergence of PARAFAC, have helped make this recursive solution the most popular.

Yet, there is an exact solution, in a sense. That is, Sanchez and Kowalski (1988) were able to fit the PARAFAC model by solving a generalized eigenanalysis problem. This solution, exact for $K = 2$ slabs and under the assumption of perfect signal, was later adapted to the case of $K > 2$ and an approximate eigensolution derived (Sanchez and Kowalski, 1990). Eigenbased methods are attractive because no iterative scheme is apparent to the user. However, these methods have been found to yield complex eigenstructures if there are significant deviations of the data from the underlying trilinear assumptions. Most often, the eigensolutions are used as intelligent starting points for the iterative PARAFAC. Still, it is useful to recognize that the PARAFAC model can be formulated as an eigenanalysis problem. This underscores the extensions of and analogies to two-way bilinear paradigms, which traditionally have been approached using eigenbased methods such as singular value decomposition (SVD) or PCA.

## Methods

### *Motor activation paradigm*

Functional MRI was used to map brain activation associated with a self-paced bilateral finger-tapping task. A single subject underwent two runs in a single session. Each run contained 11 trials repeated back to back. A trial consisted of rest, finger tapping, and rest periods. fMRI data were collected on a 1.5-T Siemens Vision imager from 24 axial, 3-mm-thick slices with an in-plane resolution of 3.438 × 3.438 mm using a gradient echo EPI sequence (TR/TE = 2500/45 ms, FA 90°, matrix 64 × 64). The duration of each trial corresponded to 10 image time frames. A single trial was defined as consisting of a pattern of task states OFF OFF OFF ON ON ON ON ON OFF OFF to allow for a return to baseline. Changes in task state between rest and activation periods were initiated using an auditory cue. The 3D EPI volumes were motion corrected using SPM99 (Friston et al., 1995). Fig. 3 shows the activation map from the first run and derived from a univariate analysis based on correlation of the observed time series response in each voxel with a fixed-effect square wave reference profile reflecting the blocked experimental design.

Furthermore, we have used these data to demonstrate the potential of multilinear models as appropriate informatics tools for use in fMRI analysis. The resulting spatial patterns or modes are validated by visual inspection and by correlation in terms of the expected result from a traditional univariate analysis model. For comparison, we have also included the results from standard multivariate analyses using SVD/PCA models with unfolded two-way bilinear arrays. A single axial slice at the level in Fig. 3 and containing the motor cortex was chosen, and time series

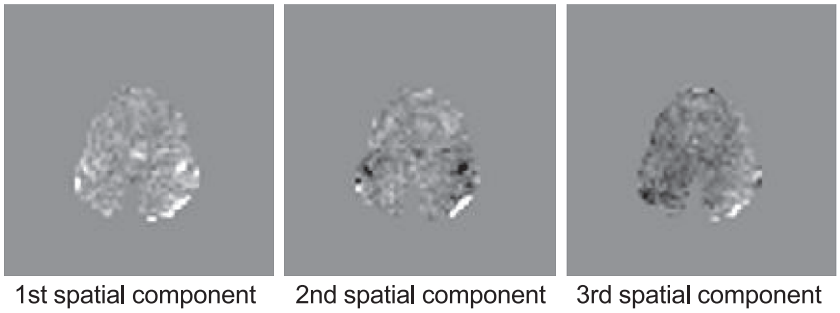


Fig. 8. Illustrating the spatial modes from SVD modeling of a 635 × 220 bilinear data array. (The amount of variance explained is 3.63%, 2.59%, and 2.34%, respectively.)

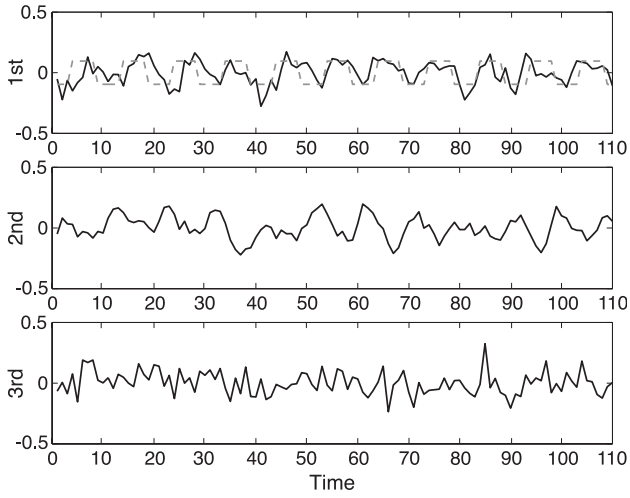


Fig. 9. Illustrating the unit-norm temporal profiles of scores from SVD modeling of an unfolded 635 × 220 bilinear data array.

response data from 635 voxels within the brain were extracted. If we are not interested in analyzing the variability between trials due for instance to motor learning or the interaction of trials with spatial modes, then the data structure can be viewed as a cube of dimensions 635 voxels × 110 time points × 2 runs. This structure can be readily described using a trilinear model as demonstrated below. A scenario where higher-way, multilinear models are used to capture also the variability between trials is presented subsequently.

*Trilinear PARAFAC model of dimensions 635 voxels × 110 time points × 2 runs*

Let **A** be a three-way array of dimensions $I$ by $J$ by $K$ with coordinate elements $a_{ijk}$, where $I$ is the number of brain voxels in each time-frame image slice ($I = 635$), $J$ is the number time points ($J = 110$), and $K$ is the number of repeated runs ($K = 2$). In keeping with the notation of Definition (2), the PARAFAC model presumes that for the appropriate rank $R$, there exist unique unit-norm vectors $\{\boldsymbol{x}_r\} \subset \mathrm{R}^I$, $\{\boldsymbol{y}_r\} \subset \mathrm{R}^J$, and $\{\boldsymbol{z}_r\} \subset \mathrm{R}^K$ such that

$$\mathbf{A} = \sum_{r=1}^{\mathrm{R}} c_r \boldsymbol{x}_r \otimes \boldsymbol{y}_r \otimes \boldsymbol{z}_r \tag{1}$$

where the coefficients $c_r$ are amplitude scaling factors. That is, $\boldsymbol{x}_r$ models the spatial pattern of activation, $\boldsymbol{y}_r$ models the temporal response, and $\boldsymbol{z}_r$ models the effect of run.

*Quadrilinear PARAFAC model of dimensions 635 voxels × 10 time points × 11 trials × 2 runs*

In the context of an event-related experimental design, these data can also be interpreted in a natural way as representing 11 repeated trials of a finger-tapping task per run. This view leads to a four-way data structure of dimensions 635 voxels × 10 time points × 11 trials × 2 runs. It is very compelling here to attempt using a quadrilinear model representation of these data to help explain variations across trials. The vector $\boldsymbol{y}_r$ now models the temporal response within a trial and $\boldsymbol{t}_r$ the effect of trial.

$$\mathbf{A} = \sum_{r=1}^{\mathrm{R}} c_r \boldsymbol{x}_r \otimes \boldsymbol{y}_r \otimes \boldsymbol{t}_r \otimes \boldsymbol{z}_r \tag{2}$$

## Results

*Trilinear PARAFAC model of dimensions 635 voxels × 110 time points × 2 runs*

A three-component ($R = 3$) trilinear PARAFAC model as described in Eq. (1) was fitted to the 635 × 110 × 2 three-way data structure. Fig. 4 shows the resulting unique spatial modes, which have been scaled to unit norm and interpolated for purpose of the display. The spatial pattern of the first expansion component closely resembles the activation map from a univariate analysis model, and the corresponding time-course profile shown in Fig. 5 mimics the ON/OFF pattern of the blocked motor task (dashed line). The second and third components appear to capture BOLD-effect signal from venous drainage vessels on the surface of the brain. The loadings across runs in Table 1 show the first component as capturing features that are common across the two runs, whereas the second and third components largely reflect structure present in the first run but not in the second. The difference between runs may be due to variations in partial voluming for the superficial vessel from slight differences in head position. Data were motion corrected within runs separately but not between runs.

*Unfolded bilinear model of dimensions 1270 voxels × 110 time points*

Traditional multivariate analysis using either an SVD or PCA would be based on an unfolded two-way bilinear structure of dimensions 1270 × 110 corresponding to a 110-point time series response observed across 1270 (635 × 2) voxels. This would

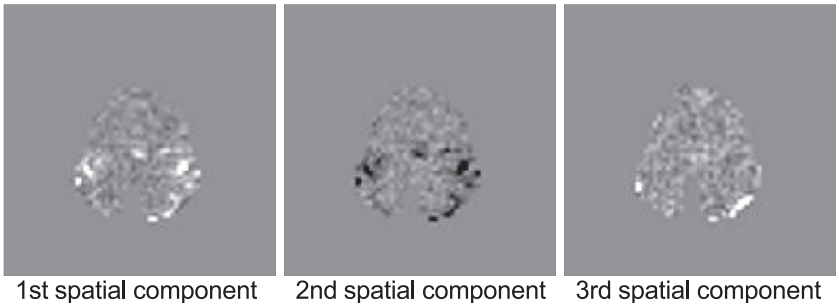


Fig. 10. Illustrating the spatial modes from quadrilinear modeling of a 635 × 10 × 11 × 2 data structure.

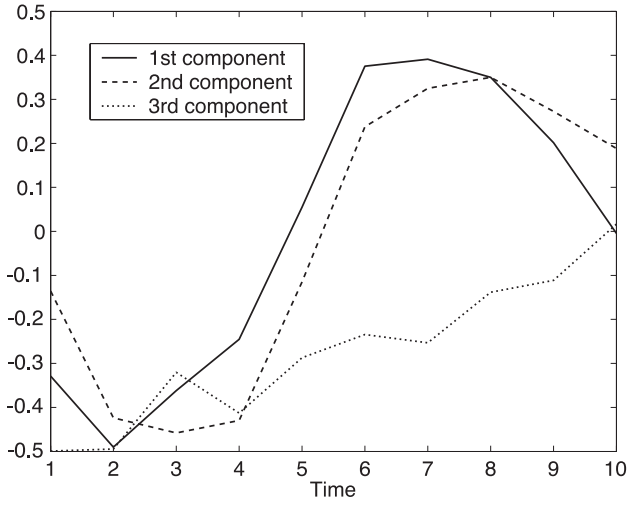


Fig. 11. Illustrating the unit-norm temporal modes $\boldsymbol{y}_r$ from quadrilinear modeling of a 635 × 10 × 11 × 2 higher-way data structure (solid: 1st component; dash: 2nd component; dot: 3rd component).

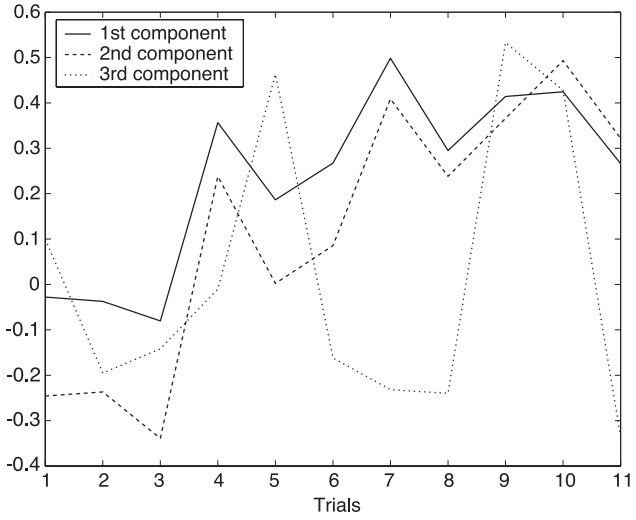


Fig. 12. Illustrating the unit-norm loadings $\boldsymbol{t}_r$ on trials from quadrilinear modeling of a 635 × 10 × 11 × 2 higher-way data structure (solid: 1st component; dash: 2nd component; dot: 3rd component).

represent a scenario of unfolding by stacking as illustrated in Fig. 2c. The loading profiles of the orthonormal eigenvectors are the temporal modes. Fig. 6 shows the first three time-course profiles derived from the SVD model. The corresponding spatial maps depicted in Fig. 7 have been computed as the average values of voxel scores across the two runs. Note that the information on loadings across runs in the trilinear model, and the ability to directly model interactions have been lost due to unfolding of the inherently higher-way data structure. While the ON/OFF pattern of the motor task is reflected in the first time course profile, the spatial distributions of scores suggest that the effect is in fact not captured in a single component but rather is distributed across the first and second components. (An SVD model of the 635 × 110 data subset from only the first run does not reveal the motor activation response manifested in any one component.) A subsequent rotation may help bring the response into the direction of the first component. However, the inherent uniqueness of the trilinear model above has been lost.

*Unfolded bilinear model of dimensions 635 voxels × 220 time points*

If one takes the view that the data structure in the natural order in which it is acquired represents a distribution of MR image voxel intensities observed across time, then a traditional multivariate analysis using either an SVD or PCA would be based on an unfolded two-way bilinear array of dimensions 635 voxels × 220 time points. This would be a scenario of unfolding by concatenation side by side as illustrated in Fig. 2b. The loading profiles of the orthonormal eigenvectors are now the spatial modes. Fig. 8 shows the resulting spatial modes, which have all been scaled to unit norm and interpolated for purpose of the display. The corresponding time-course profiles in Fig. 9 have been computed as the average values of the time-point sample scores across the two runs. The information on loadings across runs in the trilinear model above, along with the ability to directly model interactions, have been lost due to unfolding of the data cube. We notice from the spatial modes as well as from the temporal response profiles that the pertinent cyclic response of the motor task is not manifested in a single component but rather is distributed across several modes due to the rotation indeterminacy of bilinear models.

In a combination of these two scenarios of unfolding, Wang et al. (2000) proposed a trilinear modeling approach based on temporal modes derived from an SVD of the 1270 × 110 voxel-by-time data set unfolded and concatenated as in Fig. 2c paired with spatial modes derived from an SVD of the 635 × 220 voxel-by-time data set unfolded and concatenated as in Fig. 2b. Subsets of the respective modes were subsequently rotated in such a way as to diagonalize the 635 × 110 voxel-by-time matrix obtained by averaging across runs or slabs. The tensor products then became the spatiotemporal components common to all runs.

Table 2
Amplitude-scaled vector elements $\boldsymbol{c}_r\boldsymbol{z}_{lr}$ for $l$ = 1, 2 (runs) and $r$ = 1, 2, 3 (component number) in a quadrilinear model

| | Component one | Component two | Component three |
|---|---|---|---|
| 1st run | 1.0663 | 1.0169 | 0.7424 |
| 2nd run | 1.2583 | 1.1426 | −0.1324 |

*Quadrilinear PARAFAC model of dimensions 635 voxels × 10 time points × 11 trials × 2 runs*

A three-component ($R$ = 3) quadrilinear PARAFAC model as described in Eq. (2) was fitted to the 635 × 10 × 11 × 2 four-way data structure. Fig. 10 shows the resulting unique spatial modes,

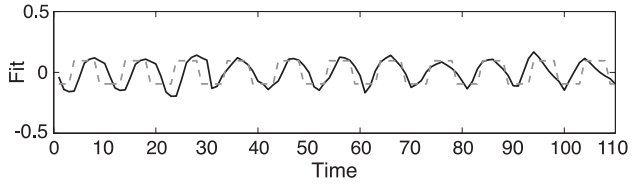


Fig. 13. Illustrating the concatenated trial responses ($\sum_{r=1}^{2} c_r \boldsymbol{y}_r \otimes \boldsymbol{t}_r$) reconstructed from the first two components in the quadrilinear model of a 635 × 10 × 11 × 2 data structure.

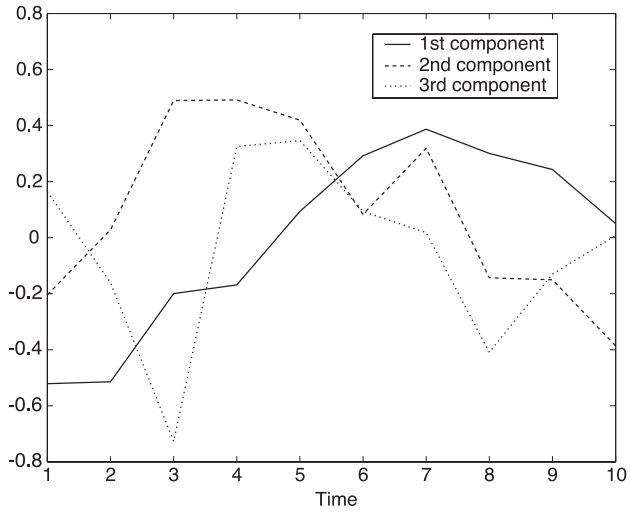


Fig. 14. Illustrating the unit-norm temporal modes from SVD modeling of an unfolded 13970 × 10 bilinear data array (solid: 1st component; dash: 2nd component; dot: 3rd component; the amount of variance explained is 11.81%, 10.60%, and 10.26%, respectively).

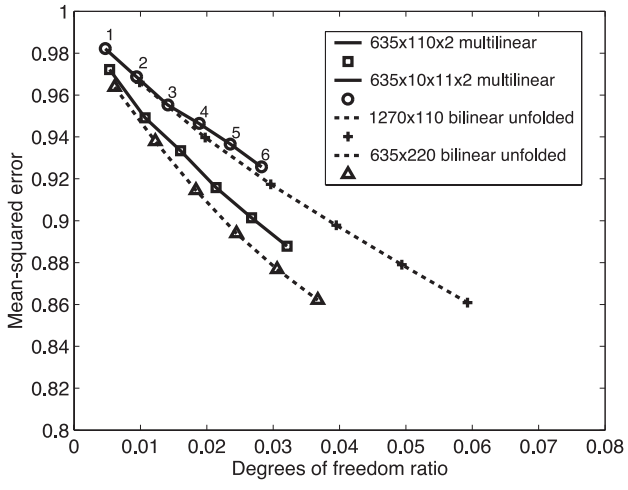


Fig. 16. Illustrating the relative mean-squared goodness-of-fit error vs. the degree of freedom ratio for the bilinear and multilinear models of order 1–6.

which have all been scaled to unit norm and interpolated for purpose of the display. The spatial pattern of the first expansion component closely resembles the activation map from a univariate analysis, and the corresponding time-course profile (solid line) in Fig. 11 mimics the ON/OFF pattern of the motor task (the task states corresponding to the ten acquisition time points across a single trial were OFF OFF OFF ON ON ON ON ON OFF OFF). Note that the response components take on both positive and negative values, rather than riding on a baseline of zero during the OFF state, due to the implicit centering in PARAFAC. The third component appears to capture BOLD-effect signal from a distal venous drainage vessel on the surface of the brain. The loadings across runs listed in Table 2 show that the first and second components capture features that are common across the two runs, whereas the third component largely reflects structure present in the first run but not in the second.

Whereas the component profiles themselves are unique for a chosen number of components $R$ in the expansion, the determination of the appropriate number of components to use depends on the nature of the data and the structure information we seek to extract. We chose $R = 3$ components throughout our analysis scenarios for consistency. In the trilinear model, the structure of the first component remains largely unaffected if the number of components were increased or reduced to $R = 2$. In the quadrilinear model, however, a minimum of $R = 3$ components are required to capture the two unique temporal response components and to model differences in structure between runs.

Fig. 12 depicts the loadings across trials within a run. These may be interpreted naturally as reflecting adaptation or learning of the task. Friston et al. (1996b) used an SVD on an unfolded bilinear data array to illustrate an effect of trial. The somewhat construed unfolding strategy yielded a so-called spatiotemporal by trial array. Illustrated in terms of the present data from a single run, this would correspond to the modeling of a 6350 time-voxels × 11 trials data array. (Frutiger et al. (2000) used a canonical variates analysis to study learning curves across trials.)

From both the spatial and the temporal maps, it appears that the second component is very highly correlated and nearly collinear with the first but of opposite polarity ($r = -0.9322$). This occurrence has been observed as well by investigators who have employed trilinear modeling of event-related EEG data to derive topographic and temporal response components (Field and Graupe, 1991). While this may be an artifact of so-called degenerate solutions caused by the least-squares optimization of the PARAFAC algorithm (Harshman and Lundy, 1984a), it may also reflect actual structure in the data. The temporal response structure explained by the second component is slightly delayed with respect to the first component. We speculate that the two similar components are in fact required to model a variable time-of-onset delay in the response across individual trials, much the same way that sine and cosine components can be combined linearly to model waveforms of arbitrary delay (Field and Graupe, 1991; Hopfinger et al.,

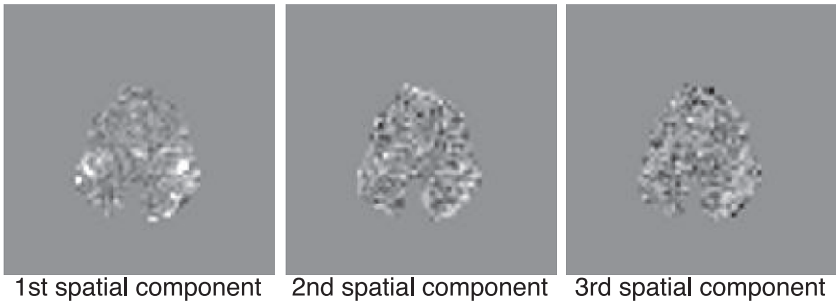


Fig. 15. Illustrating the spatial score maps from SVD modeling of a 13970 × 10 bilinear data array.

Table 3
Correlation coefficients among spatial modes based on the first model component

| | Univariate square wave | Multilinear 635 × 110 × 2 | Bilinear 1270 × 110 | Bilinear 635 × 220 | Multilinear 635 × 10 × 11 × 2 | Bilinear 13,970 × 10 |
|---|---|---|---|---|---|---|
| Univariate square wave | 1.0 | 0.8523 | 0.7738 | 0.5507 | 0.8705 | 0.8577 |
| Multilinear 635 × 110 × 2 | 0.8523 | 1.0 | 0.8355 | 0.5561 | 0.9566 | 0.8998 |
| Bilinear 1270 × 110 | 0.7738 | 0.8355 | 1.0 | 0.8671 | 0.9180 | 0.8502 |
| Bilinear 635 × 220 | 0.5507 | 0.5561 | 0.8671 | 1.0 | 0.6782 | 0.6157 |
| Multilinear 635 × 10 × 11 × 2 | 0.8705 | 0.9566 | 0.9180 | 0.6782 | 1.0 | 0.9423 |
| Bilinear 13,970 × 10 | 0.8577 | 0.8998 | 0.8502 | 0.6157 | 0.9423 | 1.0 |

2000; Möcks, 1986). This is difficult to ascertain, however, due to the coarse temporal resolution of these particular data and the voluntary nature of the task. Also, since we are not imposing any orthogonality constraint on the Y-mode fibers, we would not expect to uncover a temporal mode that is the time-derivative of another component as seen previously in the trilinear modeling of multichannel-evoked potentials from EEG data (Field and Graupe, 1991) and in bilinear modeling of real and simulated hemodynamic response MRI data (Andersen et al., 2001). Fig. 13 illustrates the trial response fits reconstructed from the first two expansion components as $\sum_{r=1}^{2} c_r \boldsymbol{y}_r \otimes \boldsymbol{t}_r$, where $\boldsymbol{y}_r$ models the temporal response and $\boldsymbol{t}_r$ models the effect of trial. The apparent polarity difference of the spatial-mode components seen in Fig. 10 has been incorporated into the amplitude scaling factors $c_r$, and the model fits have been concatenated temporally to allow for comparison with the response profiles of the trilinear model for an entire run in Fig. 5. Note that Fig. 13 exhibits the same temporal delay at the beginning and advance at the end of a run as seen in the first component of the trilinear model.

*Unfolded bilinear model of dimensions 13970 voxel-trials × 10 time points*

The traditional multivariate analysis in the context of an event-related design and using either an SVD or PCA would be based on an unfolded two-way bilinear array of dimensions 13970 × 10 corresponding to a 10-point time series response for each trial observed across 13970 voxel trials. The orthonormal eigenvectors of loading profiles are the temporal modes. Fig. 14 shows the first three time-course profiles derived from the SVD model. The corresponding spatial maps depicted in Fig. 15 have been computed as the average values of voxel scores across the 22 (11 × 2) repeated trials. While the spatial maps depicting the average scores appear to reflect the pattern of the activation map in Fig. 3 obtained from univariate analysis using a fixed-effect model, the information on loadings across trials and runs in the quadrilinear modeling scenario above along with the ability to directly model interactions have been lost due to unfolding.

*Data reduction and goodness of fit*

Fig. 16 illustrates the relative mean-squared error of the respective fits for multilinear and bilinear models of order 1–6. (In comparison, the relative mean-squared error for the fixed-effect univariate analysis with a square-wave reference is 0.9842.) The activation induced signal changes in fMRI experiments are quite weak (about 1% corresponding to typical contrast-to-noise ratio values of 0.5 or less), and the activation is manifested in only a small fraction of the total number of image voxels collected within the brain. Therefore, the amount of variability explained in multivariate models is generally quite low. The goodness-of-fit mean-squared error in turn becomes large and by itself is not an appropriate measure of validity of the models. It is not surprising that the bilinear models from unfolded data achieve lower mean-squared error measures. However, these models tend to "overfit" to where they fit separately to the response in separate trials and/or runs without regard to the presence of common structure across trials and runs. These linkages are preserved in the multilinear models. Also, it is important to consider the amount of data reduction achieved. Data reduction has been quantified by the degrees of freedom ratio (DFR-Carroll and Chang, 1970), defined as the ratio of degrees of freedom for the model (the number of free parameters) to degrees of freedom for the actual data (the total number of data points minus the number determined by preprocessing.) Multilinear models generally achieve a higher degree of data reduction corresponding to smaller DFR values for a particular model order.

In images of the spatial modes and in plots of the temporal response profiles, we have made qualitative comparisons with the activation map derived from the univariate analysis model and with the square-wave pattern of the underlying activation paradigm. Tables 3 and 4 contain quantitative measures of the correlation coefficient between pairs of modes, both spatial and temporal, for the various analysis scenarios. Whether we look at the spatial modes or the temporal response profiles, the first multilinear component in all instances exhibits a higher degree of correlation with the fixed-effect analysis model based on a square-wave

Table 4
Correlation coefficients among temporal modes based on the first model component

| | Univariate square wave | Multilinear 635 × 110 × 2 | Bilinear 1270 × 110 | Bilinear 635 × 220 | Multilinear 635 × 10 × 11 × 2[a] |
|---|---|---|---|---|---|
| Univariate square wave | 1.0 | 0.5683 | 0.4647 | 0.3897 | 0.5438 |
| Multilinear 635 × 110 × 2 | 0.5683 | 1.0 | 0.8220 | 0.7308 | 0.9002 |
| Bilinear 1270 × 110 | 0.4647 | 0.8220 | 1.0 | 0.9380 | 0.8279 |
| Bilinear 635 × 220 | 0.3897 | 0.7308 | 0.9380 | 1.0 | 0.7573 |
| Multilinear 635 × 10 × 11 × 2[a] | 0.5438 | 0.9002 | 0.8279 | 0.7573 | 1.0 |

[a] Using concatenated reference profile from Fig. 13 reconstructed with the first two components in the multilinear model.

reference than do the corresponding unfolded bilinear analysis models. While the bilinear analysis models chase variability in the data, the multilinear models appear to capture salient spatial and temporal features that are common across runs and across trials and in agreement with the underlying activation paradigm.

## Discussion

In this paper, we have presented a compelling application of multilinear models in the analysis of intrinsically higher-way data from a functional neuroimaging experiment. In comprehensive fMRI studies of brain function, the data structures often contain higher-order ways such as trial, task condition, subject, and group in addition to the intrinsic dimensions of time and space. While multivariate bilinear methods such as PCA have been used successfully for extracting information about spatial and temporal features in data from a single fMRI run, the need to unfold higher-order data sets into two-way bilinear arrays leads to decompositions that are unnecessarily complex and badly nonunique. In addition, any hope of understanding the multiway linkages and interactions present in the data is lost. As we have pointed out, these additional dimensions can be retained in multilinear models with interpretations that are neurophysiologically meaningful.

In the analysis scenarios presented above, the extracted multilinear modeling components exhibited these paradigm-related structures first in the hierarchy of component importance, or in the first two components when a subspace was required, while PCA often relegated them to lower echelons or combined them across different components. This was one empirical advantage that the multilinear modeling approach seemed to enjoy. However, it is not necessarily the most important advantage. In fact, many bilinear methods have been adapted to intelligently rearrange component order, or to perhaps direct in a supervised fashion the search for component structures toward meaningful design patterns. Indeed, in most analysis approaches that employ PCA, principal component analysis is only the first step of data reduction. While the salient features may be captured in the subspace spanned by a subset of the PCs, it is highly unlikely that a physiologically meaningful brain response will be captured in the first PC or even that it is manifested in a single PC. When this occurs, ad hoc techniques range from simply extracting the one PC that mostly reflects the experimental paradigm to preprocessing steps for the removal of high-variability structured noise components so as to allow the PC that summarizes task-related variability to "rise to the top."

In working with unfolded data from multiple subjects, Lautrup et al. (1995) identified the PC that reflected a common pattern across subjects as the first component in a set ranked by variance summary that summarized more task-related within-subject variability than between-subject variability. More refined techniques seek to perform an informed rotation within a subspace spanned by a few PCs. Oriented PCA is posed as a generalized eigenvalue problem that seeks to balance variance summary while steering away from a prespecified noise or error subspace (Diamantaras and Kung, 1996; Rayens and Andersen, 2003). Canonical variates analysis (CVA) has been used in this second step to implement a rotation based on canonical vectors that define directions that maximize the between-class variance while minimizing the within-class variance (LaConte et al., 2003). Frackowiak et al. (1997) discussed CVA in the setting of a generalized eigenvalue problem to derive the spatial modes or canonical images that best express the paradigm-related activation effects of interest relative to error effects, akin to denoising techniques in EEG time series data analysis.

Other techniques such as partial least squares (PLS) have been used to focus the structure extraction toward meaningful design profiles, so that seemingly ad hoc rotations after-the-fact could be avoided and the optimality of the extracted components defended. PLS is inherently different from PCA, however, in that two blocks of data are needed to implement PLS, one often thought of as containing the bilinear features (e.g., voxels by time) and the other as having the potential experimental paradigms (e.g., an ON/OFF pattern). The feature block is then used to "predict" the paradigm block, as is the nature of PLS, and the results of the prediction can be viewed as an exploratory means of testing various hypotheses about what the underlying experimental paradigm was perceived to be.

This basic idea was recently extended (Lin et al., 2003) to allow the sample covariance matrix between features and hypotheses, which is the core of a standard PLS analysis under orthogonal constraints, to be further decomposed by either a subsequent PCA or an independent component analysis (ICA). Similar approaches were taken in McIntosh et al. (1996, 1998) and by Lobaugh et al. (2001). Likewise, "target directed PLS" (Rayens and Andersen, 2004) was successful both at focusing the extraction on particular hypothesized "target" profiles and at proving in what sense this would always be a more optimal solution than an after-the-fact rotation or regression involving PCA components. From a slightly different perspective, "oriented PLS" (Rayens and Andersen, 2003) was introduced as a method of further orienting PLS that allows the user to produce PLS-like structures that have been oriented also away from undesirable confounds (e.g., structured noise), confounds that do not have to be orthogonal to the signal of interest. This new method was introduced and demonstrated within the context of a brain mapping study of motor performance that employed functional magnetic resonance imaging (fMRI).

Yet, the purpose of the present study has not been simply to compare the ability of a multiway model to "reorder" components to that of established bilinear methods, which has been done elsewhere (e.g., Linder and Sundberg, 1998, 2002). This was a desirable outcome, granted, that served to illustrate both the usefulness and reasonableness of a multilinear approach. Perhaps, far more important is the observation that bilinear methods have to either operate on data that are inherently bilinear, for example, from a single run in a single subject, or that have been transformed either by concatenation or by collapsing across one or more dimensions, thereby risking the loss of inherent multiway linkages and interactions that are present in the data.

Further, multiway components are typically much easier to interpret than those found from a bilinear analysis of unfolded ways. An excellent example of this phenomenon is provided in the field of chemometrics by Norgaard (1995) who used a PLS-adapted version of multiway to predict the emission–excitation profiles of the attending sugars from those of different juices. In this example, the unfolded bilinear alternative was prohibitively complicated to interpret, although there was less evidence that the unfolded fit would predict badly. However, as Norgaard (1995) pointed out, the ability of multiway fits to predict more robustly than their bilinear counterparts seems to grow with the noise level of the data. This is an important observation with implications for neuroimaging data and especially for design-prediction paradigms, such as mentioned above (Lin et al., 2003), where reliable prediction is key to an empirical assessment of the validity of

the design hypotheses. Nilsson et al. (1997) provide yet another example of the increased prediction stability as well as the increased interpretational simplicity of multilinear models. Future studies based on synthetic neuroimaging data sets that incorporate meaningful response features along the separate ways will be required to further investigate the robustness of multilinear methods to noise, both structured and random.

Last, but certainly not least, multiway structures have a claim to uniqueness that is not available for the bilinear methods discussed above. Rotation flexibility is viewed by some analysts as positive since it obviously affords the user a much greater chance of uncovering features that are interpretable in the context. Such rotation ambiguity, however, has long since been understood to largely undermine a defense of any attending interpretation of the structure. Hence, our examples produced constructs that not only were more cleanly uncovered from intrinsic multiway data than from bilinear alternatives based on unfolding, but constructs that were much less likely to be the result of overfitting and impossible to be merely the result of rotational capriciousness.

## Acknowledgments

The authors wish to thank P. Hardy, R. Greene-Avison, and A. Bognar for contributions to this paper. This work was supported in part by NIH grants R01 NS36660 and P01 AG13494, and by the Vice Chancellor for Research and Graduate Studies, University of Kentucky Medical Center.

## References

Andersen, A.H., Gash, D.M., Avison, M.J., 1999. Principal component analysis of the dynamic response measured by fMRI: a generalized linear systems framework. Magn. Reson. Imaging 17 (6), 795–815.

Andersen, A.H., Rayens, W.S., Li, R.-C., Blonder, L.X., 2000. Mathematical problems in the application of multilinear models to facial emotion processing experiments. In: Wilson, D.C., Tagare, H.D., Bookstein, F.L., Preteux, F.J., Dougherty, E.R. (Eds.), Mathematical Modeling, Estimation, and Imaging. Proc. SPIE, vol. 4121, pp. 77–87.

Andersen, A.H., Rayens, W.S., Greene-Avison, R., Pettigrew, C., Berger, J., Avison, M.J., 2001. PCA for mapping time delays in the transient response from functional MRI studies. Proc. ISMRM 9th Sci., 1720.

Andersen, A., Zhang, Z., Barber, T., Rayens, W., Zhang, J., Grondin, R., Hardy, P., Gerhardt, G., Gash, D., 2002. Functional MRI studies in awake rhesus monkeys: methodological and analytical strategies. J. Neurosci. Methods 118 (2), 141–152.

Andersson, C.A., Bro, R., 2000. The N-way toolbox for MATLAB. Chemom. Intell. Lab. Syst. 52 (1), 1–4.

Appellof, C.J., Davidson, E.R., 1981. Strategies for analyzing data from video fluorometric monitoring of liquid chromatographic effluents. Anal. Chem. 53, 2053–2056.

Benali, H., Anton, J.L., Di Paola, M., Frouin, F., Jolivet, O., Carlier, R.Y., Bittoun, J., Burnod, Y., Di Paola, R., 1995. Conditioned statistical model for functional MRI studies of the human brain. In: Bizais, Y., et al. (Eds.), Inf. Process. Med. Imaging. Kluwer Academic Publishers, Dordtrecht, The Netherlands, pp. 311–322.

Bro, R., 1997. PARAFAC: tutorial and applications. Chemom. Intell. Lab. Syst. 38, 149–171.

Bro, R., 1999. Exploratory study of sugar production using fluorescence spectroscopy and multi-way analysis. Chemom. Intell. Lab. Syst. 46, 133–147.

Bro, R., Andersson, C.A., 1998. Improving the speed of multiway algorithms. Part II: compression. Chemom. Intell. Lab. Syst. 42, 105–113.

Bro, R., De Jong, S., 1997. A fast non-negativity-constrained least squares algorithm. J. Chemom. 11, 393–401.

Bro, R., Heimdal, H., 1996. Enzymatic browning of vegetables. Calibration and analysis of variance by multiway methods. Chemom. Intell. Lab. Syst. 34, 85–102.

Burdick, D., 1995. An introduction to tensor products with applications to multiway data analysis. Chemom. Intell. Lab. Syst. 28, 229–237.

Carroll, J.D., Chang, J.J., 1970. Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition. Psychometrika 35, 283–319.

Diamantaras, K.I., Kung, S.Y., 1996. Principal Component Neural Networks. Wiley, New York.

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

Fox, P.T., Mintum, M.A., Reiman, E.M., Raichle, M.E., 1988. Enhanced detection of focal brain responses using intersubject averaging and change-distribution analysis of subtracted PET images. J. Cereb. Blood Flow Metab. 8, 642–653.

Frackowiak, R.S.J., Friston, K.J., Frith, C.D., Dolan, R.J., Mazziotta, J.C., 1997. Human Brain Function. Academic Press, San Diego.

Friston, K.J., Frith, C.D., Liddle, P.F., Frackowiak, R.S.J., 1993. Functional connectivity: the principal-component analysis of large (PET) data sets. J. Cereb. Blood Flow Metab. 13, 5–14.

Friston, K.J., Worsley, K.J., Frackowiak, R.S.J., Mazziotta, J.C., Evans, A.C., 1994. Assessing the significance of focal activations using their spatial extent. Hum. Brain Mapp. 1, 214–220.

Friston, K.J., Ashburner, J., Poline, J.B., Frith, C.D., Heather, J.D., Frackowiak, R.S.J., 1995. Spatial registration and normalization of images. Hum. Brain Mapp. 2, 165–189.

Friston, K.J., Poline, J.-B., Holmes, A.P., Frith, C.D., Frackowiak, R.S.J., 1996a. A multivariate analysis of PET activation studies. Hum. Brain Mapp. 4, 140–151.

Friston, K.J., Stephan, K.M., Heather, J.D., Frith, C.D., Ioannides, A.A., Liu, L.C., Rugg, M.D., Vieth, J., Keber, H., Hunter, K., Frackowiak, R.S.J., 1996b. A multivariate analysis of evoked responses in EEG and MEG data. NeuroImage 3, 167–174.

Friston, K.J., Holmes, A.P., Worsley, K.J., 1999a. How many subjects constitute a study? NeuroImage 10, 1–5.

Friston, K.J., Holmes, A.P., Price, C.J., Büchel, C., Worsley, K.J., 1999b. Multisubject fMRI studies and conjunction analyses. NeuroImage 10 (4), 385–396.

Friston, K., Phillips, J., Chawla, D., Büchel, C., 1999c. Revealing interactions among brain systems with nonlinear PCA. Hum. Brain Mapp. 8, 92–97.

Frutiger, S.A., Strother, S.C., Anderson, J.R., Sidtis, J.J., Arnold, J.B., Rottenberg, D.A., 2000. Multivariate predictive relationship between kinematic and functional activation patterns in a PET study of visuomotor learning. NeuroImage 12, 515–527.

Hansen, L.K., Larsen, J., Nielsen, F., Strother, S.C., Rostrup, E., Savoy, R., Lange, N., Sidtis, J., Svarer, C., Paulson, O.B., 1999. Generalizable patterns in neuroimaging: how many principal components? NeuroImage 9, 534–544.

Harshman, R.A., 1972. Determination and proof of minimum uniqueness conditions for PARAFAC1. UCLA Work. Pap. Phon. 22, 111–117.

Harshman, R.A., Lundy, M.E., 1984a. The PARAFAC model for three-way factor analysis and multidimensional scaling. In: Law, H.G., Snyder Jr., C.W., Hattie, J.A., Mcdonald, R.P. (Eds.), Research Methods for Multimode Data Analysis. Praeger, New York, pp. 172–215.

Harshman, R.A., Lundy, M.E., 1984b. Data preprocessing and the extended PARAFAC model. In: Law, H.G., Snyder Jr., C.W., Hattie, J.A., Mcdonald, R.P. (Eds.), Research Methods for Multimode Data Analysis. Praeger, New York, pp. 216–284.

Hopfinger, J.B., Büchel, C., Holmes, A.P., Friston, K.J., 2000. A study of

analysis parameters that influence the sensitivity of event-related fMRI analysis. NeuroImage 11, 326–333.
Kesler-West, M.L., Andersen, A.H., Smith, C.D., Avison, M.J., Davis, C.E., Avison, R.G., Kryscio, R.J., Blonder, L.X., 2001. Neural substrates of facial emotion procession using fMRI. Cognit. Brain Res. 11, 213–226.
Kiebel, S.J., Poline, J.-B., Friston, K.J., Holmes, A.P., Worsley, K.J., 1999. Robust smoothness estimation in statistical parametric maps using standardized residuals from the general linear model. NeuroImage 10, 756–766.
Kroonenberg, P.M., 1983. Three-Mode Principal Component Analysis. DSWO Press, Leiden, The Netherlands.
Kroonenberg, P.M., 1989. Singular value decompositions of interactions in three-way contingency tables. In: Coppi, R., Bolasco, S. (Eds.), Multiway Data Analysis. North-Holland, pp. 169–184.
Kruskal, J.B., 1984. Multilinear methods. In: Law, H.G., Snyder Jr., C.W., Hattie, J.A., McDonald, R.P. (Eds.), Research Methods for Multimode Data Analysis. Praeger, New York, pp. 7–18.
Kruskal, J.B., 1989. Rank decomposition and uniqueness for 3-way and N-way arrays. In: Coppi, R., Bolasco, S. (Eds.), Multiway Data Analysis. North-Holland, pp. 36–62.
LaConte, S., Anderson, J., Muley, S., Ashe, J., Frutiger, S., Rehm, K., Hansen, L.K., Yacoub, E., Hu, X., Rottenberg, D., Strother, S., 2003. The evaluation of preprocessing choices in single-subject BOLD fMRI using NPAIRS performance metrics. NeuroImage 18, 10–27.
Lautrup, B., Hansen, K.L., Law, I., Mörch, N., Svarer, C., Strother, S.C., 1995. Massive weight sharing: a cure for extremely ill-posed problems. In: Herman, H.J., et al. (Eds.), Supercomputing in Brain Research: From Tomography to Neural Networks. World Scientific, Singapore, pp. 137–148.
Leurgans, S., Ross, R., 1992. Multilinear models: applications in spectroscopy. Stat. Sci. 7 (3), 289–319.
Lin, F.H., McIntosh, A.R., Agnew, J.A., Eden, G.F., Zeffiro, T.A., Belliveau, J.W., 2003. Multivariate analysis of neuronal interactions in the generalized partial least squares frameworks: simulations and empirical studies. NeuroImage 20, 625–642.
Linder, M., Sundberg, R., 1998. Second order calibration: bilinear least squares regression and a simple alternative. Chemom. Intell. Lab. Syst. 42, 159–178.
Linder, M., Sundberg, R., 2002. Precision of prediction in second-order calibration, with focus on bilinear regression methods. J. Chemom. 16, 12–27.
Lobaugh, N.J., West, R., McIntosh, A.R., 2001. Spatiotemporal analysis of experimental differences in event-related potential data with partial least squares. Psychophysiology 38 (3), 517–530.
Martin, A., Wiggs, C.L., Ungerleider, L.G., Haxby, J.V., 1996. Neural correlates of category-specific knowledge. Nature 379, 649–652.
McIntosh, A.R., Bookstein, F.L., Haxby, J.V., Grady, C.L., 1996. Spatial pattern analysis of functional brain images using Partial Least Squares. NeuroImage 3, 143–157.
McIntosh, A.R., Lobaugh, N.J., Cabeza, R., Bookstein, F.L., Houle, S., 1998. Convergence of neural systems processing stimulus associations and coordinating motor responses. Cereb. Cortex 8, 648–659.
Mitchell, B., Burdick, D., 1994. Slowly converging PARAFAC sequences: swamps and two factor degeneracies. J. Chemom. 8, 155–168.
Möcks, J., 1986. The influence of latency jitter in principal component analysis of event-related potentials. Psychophysiology 23, 480–484.
Möcks, J., 1988. Topographic components model for event-related potentials and some biophysical considerations. IEEE Trans. Biomed. Eng. 35 (6), 482–484.
Nilsson, J., De Jong, S., Smilde, A.K., 1997. Multiway Calibration in 3D QSAR. J. Chemom. 11, 511–524.
Norgaard, L., 1995. Classification and prediction of quality and process parameters of thick juice and beet sugar by fluorescence spectroscopy and chemometrics. Zuckerindustrie 120 (11), 970–981.
Pham, T.D., Möcks, J., 1992. Beyond principal component analysis: a trilinear decomposition model and least squares estimation. Psychometrika 57 (2), 203–215.
Price, C.J., Friston, K.J., 1996. Cognitive conjunction: a new approach to brain activation experiments. NeuroImage 5, 261–270.
Rayens, W.S., Andersen, A.H., 2003. Oriented partial least squares. Invited. Riv. Stat. Appl.-Ital. J. Appl. Stat., RCE Edizioni, Napoli 15 (3), 367–388.
Rayens, W.S., Andersen, A.H., 2004. Partial least squares as a target-directed structure-seeking technique. Chemom. Intell. Lab. Syst. 71, 121–127.
Rayens, W.S., Mitchell, B., 1997. Two-factor degeneracies and a stabilization of PARAFAC. Chemom. Intell. Lab. Syst. 38, 173–181.
Sanchez, E., Kowalski, B.R., 1988. Tensorial calibration: II. Second-order calibration. J. Chemom. 2, 265–280.
Sanchez, E., Kowalski, B.R., 1990. Tensorial resolution: a direct trilinear decomposition. J. Chemom. 4, 29–45.
Sidiropoulos, N.D., Bro, R., 2000. On the uniqueness of multilinear decomposition of N-way arrays. J. Chemom. 14, 229–239.
Smilde, A.K., 1992. Three-way analysis. Problems and prospects. Chemom. Intell. Lab. Syst. 15, 143–157.
Smilde, A.K., Doornbos, D.A., 1992. Simple validatory tools for judging the predictive performance of PARAFAC and three-way PLS. J. Chemom. 6, 11–28.
Smith, C.D., Andersen, A.H., Kryscio, R.J., Schmitt, F.A., Kindy, M.S., Blonder, L.X., Avison, M.J., 1999. Altered brain activation in cognitively intact individuals at high risk for Alzheimer's disease. Neurology 53 (7), 1391–1396.
Strother, S.C., Kanno, I., Rottenberg, D.A., 1995a. Principal component analysis, variance partitioning, and "functional connectivity". J. Cereb. Blood Flow Metab. 15, 353–360.
Strother, S.C., Anderson, J.R., Schaper, K.A., Sidtis, J.J., Liow, J.-S., Woods, R.P., Rottenberg, D.A., 1995b. Principal component analysis and the scaled subprofile model compared to intersubject averaging and statistical parametric mapping: I. "Functional connectivity" of the human motor system studied with [$^{15}$O]Water PET. J. Cereb. Blood Flow Metab. 15, 738–753.
Talairach, J., Tournoux, P., 1988. CoPlanar Stereotaxic Atlas of the Human Brain. Thieme, New York.
Tucker, L.R., 1966. Some mathematical notes on three-mode factor analysis. Psychometrika 31, 279–311.
Wang, K., Begleiter, H., Porjesz, B., 2000. Trilinear modeling of event-related potentials. Brain Topogr. 12 (4), 263–271.
Wang, K., Begleiter, H., Porjesz, B., 2001. Warp-averaged event-related potentials. Clin. Neurophysiol. 112, 1917–1924.
Wold, S., Geladi, P., Esbensen, K., Ohman, J., 1987. Multi-way principal components- and PLS-analysis. J. Chemom. 1, 41–56.
Worsley, K.J., 1994. Local maxima and the expected Euler characteristics of excursion sets of chi2, f, and t fields. Adv. Appl. Probab. 26, 13–42.
Worsley, K.J., Poline, J.-B., Friston, K.J., Evans, A.C., 1997. Characterizing the response of PET and fMRI data using multivariate linear models. NeuroImage 6 (4), 305–319.
Zhang, Z., Andersen, A.H., Avison, M.J., Gerhardt, G.A., Gash, D.M., 2000. Functional MRI of apomorphine activation of the basal ganglia in awake rhesus monkeys. Brain Res. 852, 290–296.