NeuroImage **16,** 454–464 (2002)
doi:10.1006/nimg.2002.1067, available online at http://www.idealibrary.com on IDEAL®

Exploratory fMRI Analysis by Autocorrelation Maximization

Ola Friman, Magnus Borga, Peter Lundberg,* and Hans Knutsson

*Department of Biomedical Engineering and *Departments of Radiation Physics and Diagnostic Radiology, Linköping University, University Hospital, 581 85 Linköping, Sweden*

Received June 6, 2001

A novel and computationally efficient method for exploratory analysis of functional MRI data is presented. The basic idea is to reveal underlying components in the fMRI data that have maximum autocorrelation. The tool for accomplishing this task is Canonical Correlation Analysis. The relation to Principal Component Analysis and Independent Component Analysis is discussed and the performance of the methods is compared using both simulated and real data. © 2002 Elsevier Science (USA)

INTRODUCTION

Techniques for analyzing fMRI data can coarsely be divided into hypothesis-driven and data-driven methods. Hypothesis-driven methods require prior knowledge of event timing from which an anticipated hemodynamic response can be modeled. In a vast majority of fMRI experiments, the event timing is predefined and the acquired fMRI data can be subjected to hypothesis-driven analysis such as *t* tests or cross-correlation analysis. Data-driven methods employ a different approach by exploring the fMRI data for, in some sense, interesting components. These components may display structures or patterns, which are difficult to specify *a priori*, such as unexpected activations, drifts and motion related artifacts.

Data-driven methods can explore fMRI data in different ways. The viewpoint taken here is that fMRI data consist of linear mixtures of underlying independent components. The general problem of unmixing and recovering the underlying components is referred to as the Blind Source Separation problem. A popular method for solving this problem is Independent Component Analysis (Bell and Sejnowski, 1995; Hyvärinen and Oja, 2000), which also has been successfully applied on fMRI data (McKeown *et al.*, 1998).

We introduce a new method for data-driven exploration of fMRI data. It relies on the fact that all interesting real signals are autocorrelated, as opposed to white noise which in general is not particularly interesting. When independent autocorrelated signals are linearly mixed the resulting signal generally becomes less autocorrelated than the original signals. Assuming that the observed fMRI data consist of such mixtures of underlying sources, it is potentially possible to recover these underlying sources by finding a new set of signals which are maximally autocorrelated. The linear transformation achieving this can be obtained by Canonical Correlation Analysis (CCA). In the following sections we will describe the novel CCA method and its relation to Principal Component Analysis (PCA) and Independent Component Analysis (ICA).

THEORY

Canonical correlation analysis is a multivariate generalization of the ordinary Pearson correlation (Hotelling, 1936; Anderson, 1984). Consider two zero mean random vectors $\mathbf{x} = [x_1, x_2, \ldots, x_p]^T$ and $\mathbf{y} = [y_1, y_2, \ldots, y_q]^T$ of dimensions p and q. Construct two new scalar random variables x and y as linear combinations of the components in $\mathbf{x}$ and $\mathbf{y}$;

$$x = w_{x_1}x_1 + \ \ldots \ + w_{x_p}x_p = \mathbf{w}_\mathbf{x}^T\mathbf{x}, \tag{1}$$

$$y = w_{y_1}y_1 + \ \ldots \ + w_{y_q}y_q = \mathbf{w}_\mathbf{y}^T\mathbf{y}. \tag{2}$$

The problem is to find the vectors $\mathbf{w}_\mathbf{x} = [w_{x_1}, \ldots, w_{x_p}]^T$ and $\mathbf{w}_\mathbf{y} = [w_{y_1}, \ldots, w_{y_q}]^T$ that maximize the *correlation* ρ between x and y. Such vectors are found by solving the following maximization problem;

$$\max_{\mathbf{w}_\mathbf{x}, \mathbf{w}_\mathbf{y}} \rho\,(\mathbf{w}_\mathbf{x}, \mathbf{w}_\mathbf{y}) = \frac{\mathbf{w}_\mathbf{x}^T\mathbf{C}_\mathbf{xy}\mathbf{w}_\mathbf{y}}{\sqrt{(\mathbf{w}_\mathbf{x}^T\mathbf{C}_\mathbf{xx}\mathbf{w}_\mathbf{x})\,(\mathbf{w}_\mathbf{y}^T\mathbf{C}_\mathbf{yy}\mathbf{w}_\mathbf{y})}}, \tag{3}$$

where $\mathbf{C}_\mathbf{xx}$ and $\mathbf{C}_\mathbf{yy}$ are the within-set covariance matrices of $\mathbf{x}$ and $\mathbf{y}$, respectively, and $\mathbf{C}_\mathbf{xy}$ is the between-sets covariance matrix. In practice, estimated covariance matrices are used. Once an optimal pair of vectors $\mathbf{w}_{\mathbf{x}_1}$ and $\mathbf{w}_{\mathbf{y}_1}$ has been found, which maximize the correlation between the so-called canonical variates $\mathbf{w}_{\mathbf{x}_1}^T\mathbf{x}$ and $\mathbf{w}_{\mathbf{y}_1}^T\mathbf{y}$, it is possible to go on to find a second pair $\mathbf{w}_{\mathbf{x}_2}$

1053-8119/02 $35.00
© 2002 Elsevier Science (USA)
All rights reserved.

and $\mathbf{w}_{\mathbf{y}_2}$ that maximize Eq. (3) with the constraint that the produced variates $\mathbf{w}_{\mathbf{x}_2}^T\mathbf{x}$ and $\mathbf{w}_{\mathbf{y}_2}^T\mathbf{y}$ are uncorrelated with the first pair of variates. Proceeding in this manner, it is possible to find several pairs of vectors that maximize Eq. (3) with the constraint that the corresponding variates are mutually uncorrelated with all the preceedingly found variates. All in all $min(p, q)$ such pairs of vectors $\mathbf{w}_{\mathbf{x}_i}$ and $\mathbf{w}_{\mathbf{y}_i}$ $i = 1 \ldots min(p, q)$ exist along with the corresponding correlation coefficients. We find the $\mathbf{w}_{\mathbf{x}_i}$ vectors by solving the eigenvalue problem;

$$\mathbf{C}_{\mathbf{xx}}^{-1}\,\mathbf{C}_{\mathbf{xy}}\,\mathbf{C}_{\mathbf{yy}}^{-1}\,\mathbf{C}_{\mathbf{yx}}\,\mathbf{w}_{\mathbf{x}_i} = \rho_i^2\,\mathbf{w}_{\mathbf{x}_i}. \tag{4}$$

Thus we find the squared correlations ρ_i^2 and the vectors $\mathbf{w}_{\mathbf{x}_i}$ as the eigenvalues and eigenvectors to the square matrix $\mathbf{C}_{\mathbf{xx}}^{-1}\mathbf{C}_{\mathbf{xy}}\mathbf{C}_{\mathbf{yy}}^{-1}\mathbf{C}_{\mathbf{yx}}$. The $\mathbf{w}_{\mathbf{y}_i}$ vectors are found by exchanging the x and y subscripts in Eq. (4) and solving the corresponding eigenvalue problem.

METHODS

FMRI data can be explored from both a temporal and a spatial point of view. Computationally the basic difference is the organization of the data matrix. The two cases are illustrated in Fig. 1.

Temporal Analysis

In temporal analysis we are searching for interesting timecourses in the fMRI data. Examples are stimulus induced timecourses or timecourses containing pronounced drift or motion artifacts. We assume that the observed voxel timecourses $\mathbf{x}(t) = [x_1(t), x_2(t), \ldots, x_n(t)]^T$, $t = 1 \ldots N$, consist of different mixtures of an equal number of underlying independently generated source signals, $\mathbf{s}(t) = [s_1(t), s_2(t), \ldots, s_n(t)]^T$, $t = 1 \ldots N$,

$$\begin{cases} x_1(t) = a_{11}s_1(t) + \ \ldots\ + a_{1n}s_n(t) = \mathbf{a}_1^T\mathbf{s}(t) \\ x_2(t) = a_{21}s_1(t) + \ \ldots\ + a_{2n}s_n(t) = \mathbf{a}_2^T\mathbf{s}(t) \\ \vdots \\ x_n(t) = a_{n1}s_1(t) + \ \ldots\ + a_{nn}s_n(t) = \mathbf{a}_n^T\mathbf{s}(t), \end{cases}$$

which conveniently can be written in matrix form

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t). \tag{5}$$

Both the mixing matrix $\mathbf{A}$ and the underlying sources $\mathbf{s}(t)$ are unknown. The aim is to recover the underlying $\mathbf{s}(t)$ signals by linearly unmixing the observed fMRI timecourses $\mathbf{x}(t)$,

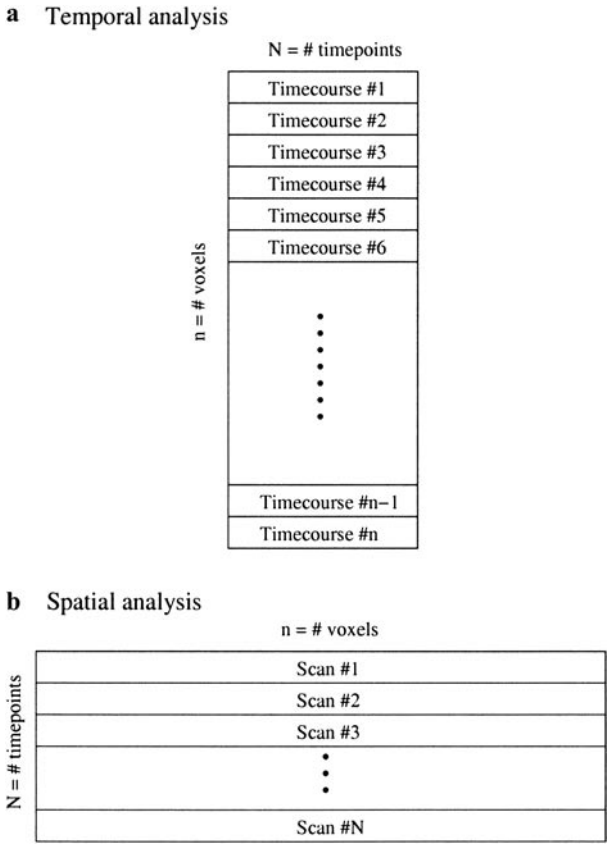

FIG. 1. The organization of the data matrix for (a) temporal and (b) spatial analysis. Each row is considered to be an observed linear mixture of a set of underlying components.

$$\begin{cases} \tilde{s}_1(t) = w_{11}x_1(t) + \ \ldots\ + w_{1n}x_n(t) = \mathbf{w}_1^T\mathbf{x}(t) \\ \tilde{s}_2(t) = w_{21}x_1(t) + \ \ldots\ + w_{2n}x_n(t) = \mathbf{w}_2^T\mathbf{x}(t) \\ \vdots \\ \tilde{s}_n(t) = w_{n1}x_1(t) + \ \ldots\ + w_{nn}x_n(t) = \mathbf{w}_n^T\mathbf{x}(t), \end{cases}$$

or in matrix form

$$\tilde{\mathbf{s}}(t) = \mathbf{W}\mathbf{x}(t). \tag{6}$$

Ideally, $\mathbf{W}$ is the inverse (up to a scaling and permutation) of the unknown mixing matrix $\mathbf{A}$. The common situation in temporal data-driven fMRI analysis is that there are more voxels than timepoints. Often all within-brain voxels are used so that n can be several thousands while the number of acquired timepoints N usually is in the range of 100–200. Hence, there are considerably more mixtures than there are samples of each mixture, making the problem of recovering the underlying sources ill-posed. Therefore, preprocessing and relevant dimensionality reduction are required in order to reduce the number of mixtures.

Spatial Analysis

In spatial analysis we transpose the data matrix, as illustrated in Fig. 1b, and assume that each acquired fMRI image $x_i(k)$, $i = 1 \ldots N$, where k is a voxel index, is a mixture of a set of underlying basis images or 'spatial modes' $s_i(k)$, $i = 1 \ldots N$,

$$\mathbf{x}(k) = \mathbf{A}\mathbf{s}(k). \tag{7}$$

Note that $x_i(k)$ and $s_i(k)$ now denote images. The basis images $s_i(k)$ are assumed to reflect locations of independently occurring processes. In the same manner as in the temporal analysis we try to recover the underlying basis images by linearly unmixing the acquired fMRI images,

$$\tilde{s}_i(k) = \mathbf{w}_i^T\mathbf{x}(k),\ i = 1 \ldots N. \tag{8}$$

By transposing the data matrix we obtain significantly more samples of each mixture and therefore dimension reduction may not be necessary.

Overview of Methods

The three data-driven analysis methods (PCA, ICA, and CCA) considered here have a common principle; they are constrained to find linear transformations $\mathbf{w}_i$ of the original fMRI data that produce mutually *uncorrelated* components s_i, where s_i represents an underlying timecourse in the temporal case and a basis image in the spatial case. Uncorrelatedness is a reasonable constraint since the original sources are assumed to be independent. However, there are an infinite number of linear transformations producing uncorrelated components and therefore an additional objective must be used. PCA, ICA, and the novel CCA method employ different additional objectives. Subject to the constraint of uncorrelated components, PCA maximizes variance or signal energy, ICA minimizes Gaussianity or some related measure and the CCA method maximizes autocorrelation of the resulting components. PCA and ICA are extensively described in the literature (Hyvärinen and Oja, 2000; Bell and Sejnowski, 1995; McKeown *et al.*, 1998; Andersen *et al.*, 1999) and are therefore only briefly reviewed here. The CCA method is described in more detail.

Principal Component Analysis

The tacit assumption made when applying PCA is that the underlying sources s_i have high energy or variability. In practice, the unmixing vectors $\mathbf{w}_i$ are found as the eigenvectors of the estimated covariance matrix $\tilde{\mathbf{C}}$ of the data. In the temporal case $\tilde{\mathbf{C}} = 1/N\Sigma\mathbf{x}(t)\mathbf{x}(t)^T$ and in the spatial case $\tilde{\mathbf{C}} = 1/n\Sigma\mathbf{x}(k)\mathbf{x}(k)^T$, where the mean is assumed removed from each row of the data matrices. Thus $\mathbf{w}_i$ is found by solving the eigenvalue problem

$$\tilde{\mathbf{C}}\mathbf{w}_i = \lambda_i\mathbf{w}_i. \tag{9}$$

Transforming the original fMRI data with these vectors as rows in a transformation matrix $\mathbf{W}$, according to Eq. (6), yields a new set of "eigen-timecourses" or "eigen-images" s_i, which are uncorrelated and sorted by increasing variance.

Independent Component Analysis

The aim of ICA is, as the name implies, to find mutually *statistically* independent components s_i. This goal can, however, not be achieved in practice, a fundamental problem being the lack of general measures of statistical independence. Instead, most ICA methods find transformations that minimize some measure of Gaussianity of the resulting components s_i, which can be interpreted as producing as independent components as possible. Examples of Gaussianity measures or related measures are kurtosis, negentropy and mutual information. Even though ICA in general does not reach its goal, it is still a useful method that has proven to perform well in many applications, fMRI included.

A commonly applied preprocessing step is to prewhite the data, which essentially implies performing a PCA. Most often the dimensionality of the data is also reduced by discarding the principal components with lowest variance, i.e., the number of mixtures is reduced. An unmixing matrix $\mathbf{W}$ is subsequently found by iteration from a random starting point. Traditionally the components found by ICA are presented in random order. However, a natural order is induced by the measure of Gaussianity used for finding the components.

The Canonical Correlation Analysis Method

As mentioned in the introduction, all interesting real signals are autocorrelated. When such signals are mixed the autocorrelation of the resulting mixtures will in general be lower compared to the original source signals. The exploratory method introduced here uses CCA in a special fashion to find unmixing vectors that produce components s_i with maximum autocorrelation (Borga and Knutsson, 2001).

We begin by describing the temporal analysis version of the CCA method and thus assume that the original data matrix is organized as in Fig. 1a. CCA requires two sets of data, here denoted by $\mathbf{x}(t)$ and $\mathbf{y}(t)$. Let $\mathbf{x}(t)$ be the original data matrix, or a dimensionality reduced data matrix obtained by for example PCA, with n mixtures and $t = 1 \ldots N$ samples of each mixture. Construct $\mathbf{y}(t)$ as a temporally delayed version of $\mathbf{x}(t)$,

$$\mathbf{y}(t) = \mathbf{x}(t-1). \tag{10}$$

Subtract the mean of each row from the data matrices $\mathbf{x}(t)$ and $\mathbf{y}(t)$ and estimate the within- and between-set covariance matrices by

$$\tilde{\mathbf{C}}_{\mathbf{xx}} = \frac{1}{N}\Sigma\mathbf{x}(t)\mathbf{x}(t)^T \tag{11}$$

$$\tilde{\mathbf{C}}_{\mathbf{yy}} = \frac{1}{N}\Sigma\mathbf{y}(t)\mathbf{y}(t)^T \tag{12}$$

$$\tilde{\mathbf{C}}_{\mathbf{xy}} = \tilde{\mathbf{C}}_{\mathbf{yx}}^T = \frac{1}{N}\Sigma\mathbf{x}(t)\mathbf{y}(t)^T. \tag{13}$$

Insert the estimated covariance matrices into Eq. (4) and solve the eigenvalue problem to find a set of eigenvectors $\mathbf{w}_{\mathbf{x}_i}$. As pointed out in the Theory section we must solve a similar eigenvalue problem to find the $\mathbf{w}_{\mathbf{y}_i}$ vectors, but since the datasets $\mathbf{x}(t)$ and $\mathbf{y}(t)$ essentially contain the same data if border effects are neglected, $\mathbf{w}_{\mathbf{x}_i}$ and $\mathbf{w}_{\mathbf{y}_i}$ will essentially also contain the same components. Hence it suffices to calculate $\mathbf{w}_{\mathbf{x}_i}$. Therefore the set-index is dropped and we denote the eigenvectors just by $\mathbf{w}_i$, $i = 1 \ldots n$. By construction, the eigenvector $\mathbf{w}_1$ belonging to the largest eigenvalue transforms the original data so that $\tilde{s}_1(t) = \mathbf{w}_1^T\mathbf{x}(t)$ and $\mathbf{w}_1^T\mathbf{y}(t)$ are maximally correlated. But $\mathbf{y}(t)$ is just $\mathbf{x}(t)$ shifted one step so that

$$\mathbf{w}_1^T\mathbf{y}(t) = \mathbf{w}_1^T\mathbf{x}(t-1) = \tilde{s}_1(t-1). \tag{14}$$

Hence, of all linear transformations of the original timecourses in $\mathbf{x}(t)$, the transform given by $\mathbf{w}_1$ yields the timecourse $\tilde{s}_1(t)$ that has maximal temporal autocorrelation (at lag one). The autocorrelation is given as the square root of the corresponding eigenvalue, cf. the Theory section. Proceeding, $\tilde{s}_2(t) = \mathbf{w}_2^T\mathbf{x}(t)$ has maximum autocorrelation of all possible timecourses that can be obtained by a linear transformation and that is uncorrelated with $\tilde{s}_1(t)$. In this manner we find n mutually uncorrelated timecourses $\tilde{s}_i(t) = \mathbf{w}_i^T\mathbf{x}(t)$, $i = 1 \ldots n$, which are maximally autocorrelated. Since the underlying source timecourses were assumed to be independent and to have larger autocorrelation than the original fMRI data in $\mathbf{x}(t)$, it is likely that the CCA method reveals those signals. As in PCA and ICA, the underlying source signals can however only be found up to a multiplicative factor.

The spatial version of the CCA method differs only in the manner in which the $\mathbf{y}(k)$ data are constructed. In images, the autocorrelation function is two-dimensional and there is no unique neighbor that can be used for constructing $\mathbf{y}(k)$. Instead, consider a voxel $x(i, j)$ in an fMRI image, where i and j denote image coordinates. $y(i, j)$ is constructed as the sum of the neighbors to $x(i, j)$,

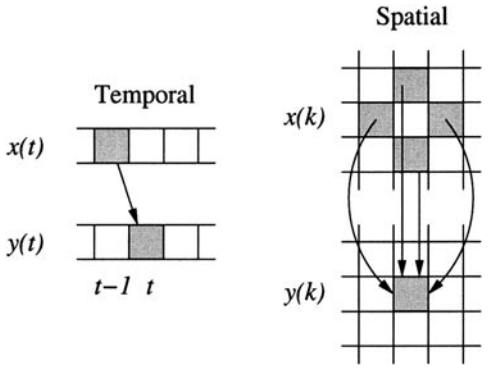

FIG. 2. In temporal CCA analysis, $\mathbf{y}(t)$ is just $\mathbf{x}(t)$ shifted one step. In spatial CCA analysis, $\mathbf{y}(k)$ is constructed as the sum of the voxels surrounding voxel k.

$$y(i, j) = x(i-1, j) + x(i+1, j) + x(i, j-1) + x(i, j+1), \tag{15}$$

which is illustrated in Fig. 2. $x(i, j)$ and $y(i, j)$ are then reorganized as in Fig. 1b for spatial analysis and used as input to the CCA. By the same reasoning as in the temporal analysis case, the CCA method finds mutually orthogonal basis images with maximal spatial autocorrelation. Hence, the spatial CCA method favors components with spatially connected regions. The generalization to 3-D volumes is straightforward by constructing $\mathbf{y}(k)$ as the sum of all neighbors in the 3-dimensional space.

Note that while both PCA and ICA are concerned with the shapes of the distributions of the samples (variance and non-Gaussianity), the CCA method focuses on the temporal or spatial characteristic of the signals in terms of autocorrelation. The usefulness of this approach is shown in the next section. Another point that should be stressed is that PCA and CCA always give the same output when applied on the same input data, while standard implementations of ICA generally give different results when repeatedly applied on the same data.

RESULTS

We use both simulated and experimental fMRI data to demonstrate and compare the PCA, ICA, and CCA methods. The PCA and CCA methods were implemented in MATLAB (The MathWorks, Inc.). For the ICA analysis the FastICA package[1] was used (Hyvärinen, 1999; Hyvärinen and Oja, 2000). FastICA was chosen because it is computationally orders of magni-

[1] Available at http://www.cis.hut.fi/projects/ica/fastica/.

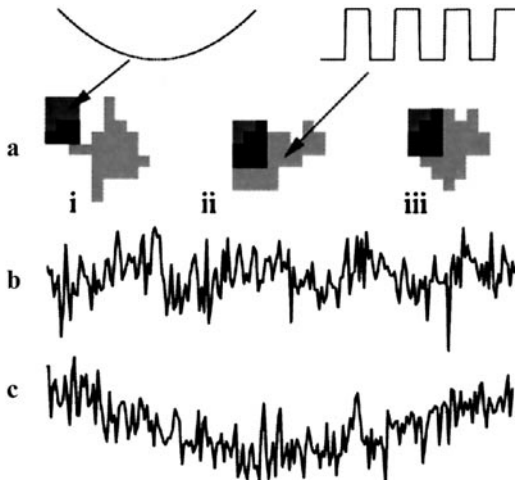

FIG. 3. (a) Three examples of generated regions where the boxcar and quadratic signals were embedded in the noise. The black areas indicate overlaps between the two regions. Examples of boxcar and quadratic timecourses embedded in noise is shown in (b) and (c), respectively.

tude faster than for example the algorithm by Bell and Sejnowski (1995) and with roughly the same performance (Hyvärinen *et al.*, 2001). Still we emphasize that reported ICA results are obtained using FastICA and may not extrapolate to all ICA methods. The best performance with FastICA was obtained using a 'tanh'-nonlinearity and symmetric computation of the independent components (not shown), and this configuration was used in the experiments below.

Simulated Data

The three methods were first applied on simulated data. Regions with boxcar-"activity" were generated and embedded in Gaussian white noise. The signal to noise ratio, defined as the ratio between boxcar and noise standard deviations, was 0.3 (Fig. 3b). Additionally, a patch containing a strong quadratic trend (signal to noise ratio 0.6, Fig. 3c) was superimposed creating different overlaps between the two regions, Fig. 3a shows three representative simulated shapes. The length of each timecourse was 200 timepoints, the size of the boxcar regions was 30 voxels and the size of the quadratic trend patch was 8 voxels. It should be stressed that the simulated data only was used for evaluating the general source separation capabilities of the methods, the intention was not to imitate an actual fMRI situation. Five thousand randomly shaped regions were simulated and both the temporal and spatial versions of the PCA, ICA, and CCA methods were applied on each of the simulated data sets. As the number of voxels in this case was approximately equal to the number of timepoints in each timecourse, dimension reduction was necessary in both the temporal and spatial analysis for ICA and CCA. Dimension reduction was implemented using PCA, i.e., the principal components with highest variance were used as input to the ICA and CCA methods. The number of mixtures was reduced so that there were 20 times more samples of each mixture than the total number of mixtures. As there were 200 samples in each simulated timecourse, the 10 first principal components were used as input to the ICA and CCA methods in the temporal analysis and roughly the same number in the spatial analysis (slightly varying with the spatial size of the simulated dataset). For each simulated case the components for each method that correlated the most with the known temporal or spatial patterns, of which examples are shown in Fig. 3a, were picked out. In Fig. 4, the performance of the temporal analysis over the 5000 simulations are shown as histograms of how well the extracted components correlated with the known ideal timecourses. PCA finds the boxcar and quadratic signals but it mixes them up, which can be seen in the example to the right in Fig. 4. Moderate correlation levels are therefore obtained as the histograms indicate. ICA easily finds the boxcar due to its non-Gaussian nature, but the difficulties ICA has in capturing the quadratic component is apparent. In contrast, the CCA method reliably finds the highly autocorrelated boxcar and quadratic signal in its first and second components, see Fig. 4.

The result from the spatial analysis is shown in Fig. 5. The patch containing the quadratic signal is found most accurately, ICA being the most accurate method as is seen from the histograms. The following components are constrained to be uncorrelated with the patch containing the quadratic timecourse, and the pattern containing the boxcar can therefore not be reliably found where the two regions overlap.

Note that in the examples to the right in Fig. 5, PCA has difficulties separating the two regions just as in the temporal analysis. Moreover, ICA occasionally fails completely in finding the boxcar pattern which is seen in the histogram for the boxcar pattern.

To summarize, PCA performs moderately in both temporal and spatial analysis. ICA can find the underlying components very well but there is also a significant risk of failure. If also taking into account that different results are produced each time ICA is applied on identical data, the performance of ICA is in this case not very satisfying. In contrast, the CCA method finds the underlying components reliably, especially in the temporal analysis.

Experimental Data

The PCA, ICA, and CCA methods were also applied on experimental fMRI data. The experiment was a 180 timepoints long mental calculation task where a volunteer added two and three digit numbers which were projected onto a screen. The paradigm was of a blocked

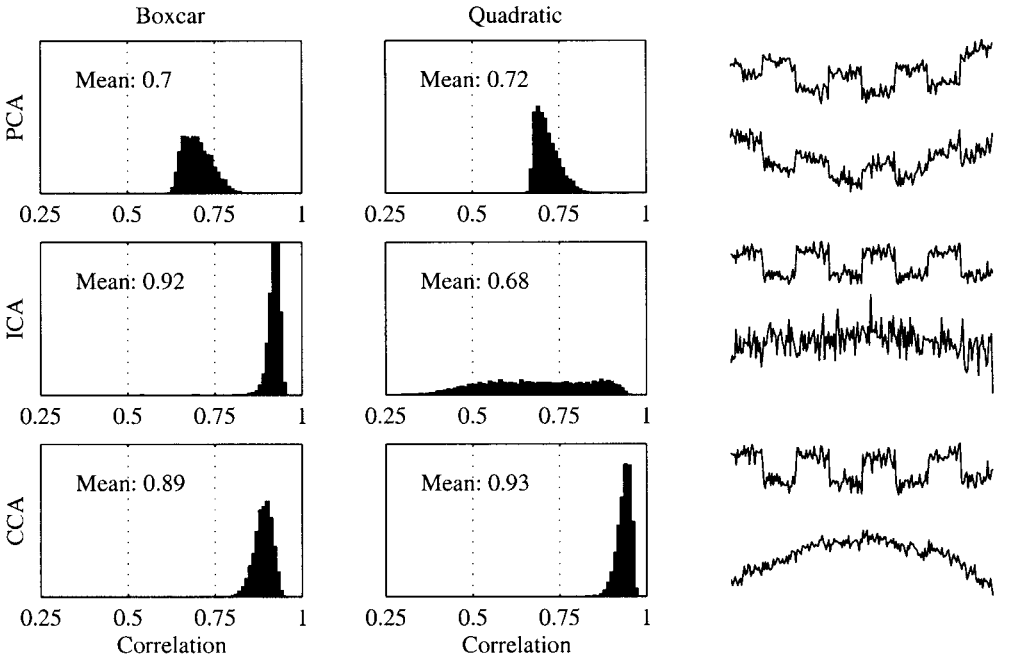

FIG. 4. The histograms show how well the timecourses extracted by the temporal PCA, ICA, and CCA methods correlated with the known temporal boxcar and quadratic signals over the 5000 simulated examples. To the right typical examples of the extracted timecourses for each method are shown.

design with alternating 30-s periods of rest and mental calculation. The image size was 128 × 128 voxels, TR 2 s, TE 60 ms, and slice thickness 6 mm. The images were motion corrected prior to the analysis. A single slice was subjected to both temporal and spatial analysis.

In the temporal analysis all within-brain voxel timecourses were used as input to PCA. The nine first

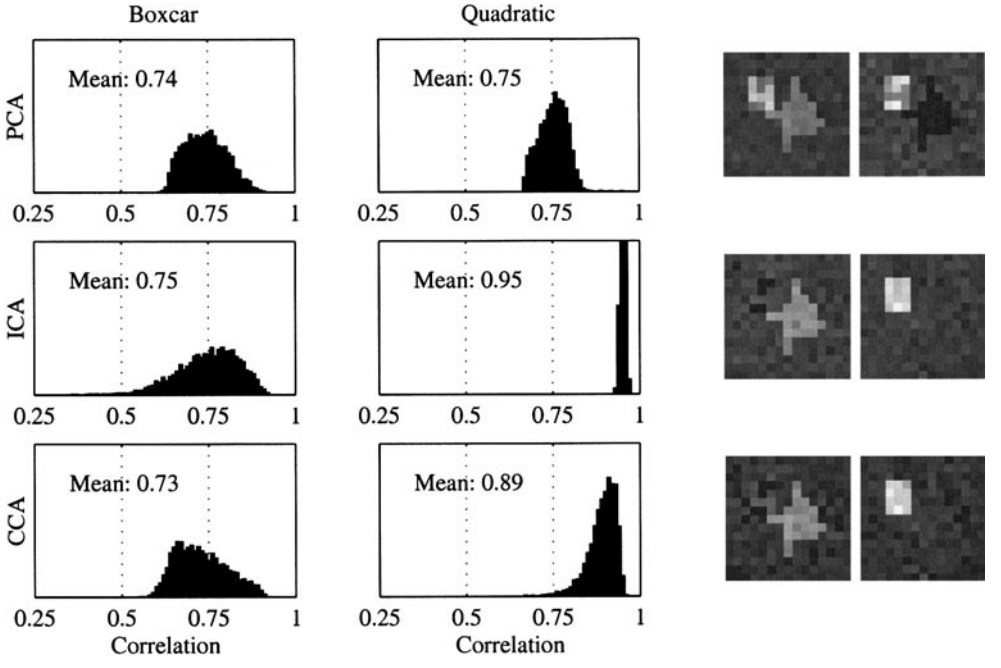

FIG. 5. Histograms of how well the spatial components extracted by the spatial PCA, ICA, and CCA methods correlated with the known shapes containing the boxcar and quadratic timecourses. To the right typical examples of the extracted patterns for each method are shown. The correct spatial pattern is pattern (i) in Fig. 3a.

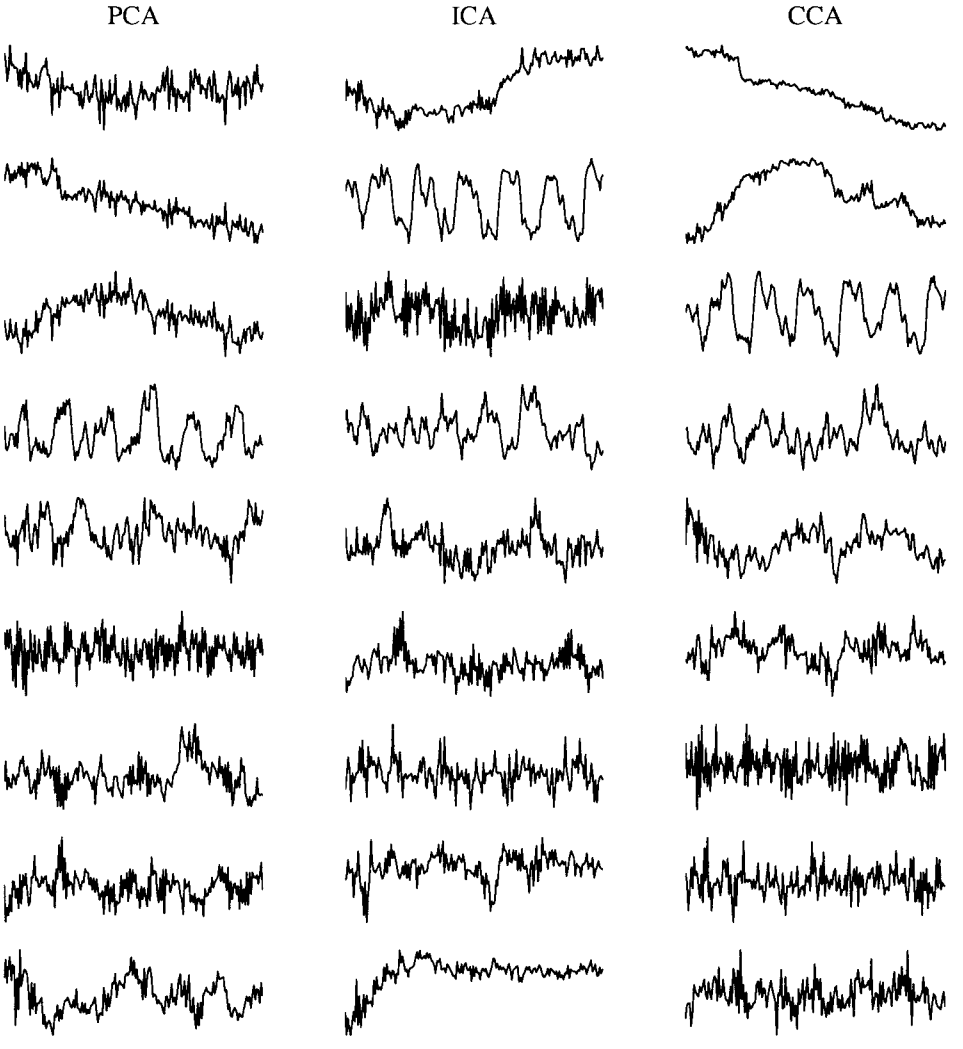

FIG. 6. The timecourses extracted when each method was applied on the experimental fMRI data. The PCA timecourses were used as input to the ICA and CCA methods.

"eigen-timecourses" were subsequently used as input to ICA and the CCA method as there were 180 samples of each mixture. The resulting temporal components are shown in Fig. 6. As discussed earlier, PCA orders the components by energy, CCA by autocorrelation, and the ICA components were post-hoc ordered by the measure of non-Gaussianity used by FastICA. PCA picks up some components which clearly are task related but the noise contamination due to the high variance noise is evident. ICA seems to have isolated a task related timecourse in one or two components and some of the other components show interesting structure. There is, however, still a substantial level of noise left and running ICA again gives a different set of components. The CCA method extracted several interesting timecourses, including a hemodynamic response evoked by the mental calculation task. The first extracted component is a linear trend which is highly autocorrelated and in the subsequent components the autocorrelation decreases. Note that this ordering ensures that the interesting timecourses always end up in the first components while noise timecourses end up in the last components. An interesting question is which timecourses in the original fMRI data contributed to the extracted temporal components. Figure 7 shows the correlation maps when the first four CCA components were correlated with each voxel timecourse in the original data. With this information at hand, the first temporal component indicates a linear drift located near high image gradients (cf. top images in Fig. 8). The second component seem to be a residual motion artifact. The third component is clearly associated with the mental calculation task while the fourth component is not so straightforward to interpret.

In the spatial analysis, the number of samples of each mixture (i.e., the number of within-brain voxels in the image) is significantly larger than the total number of mixtures (i.e., the total number of images which

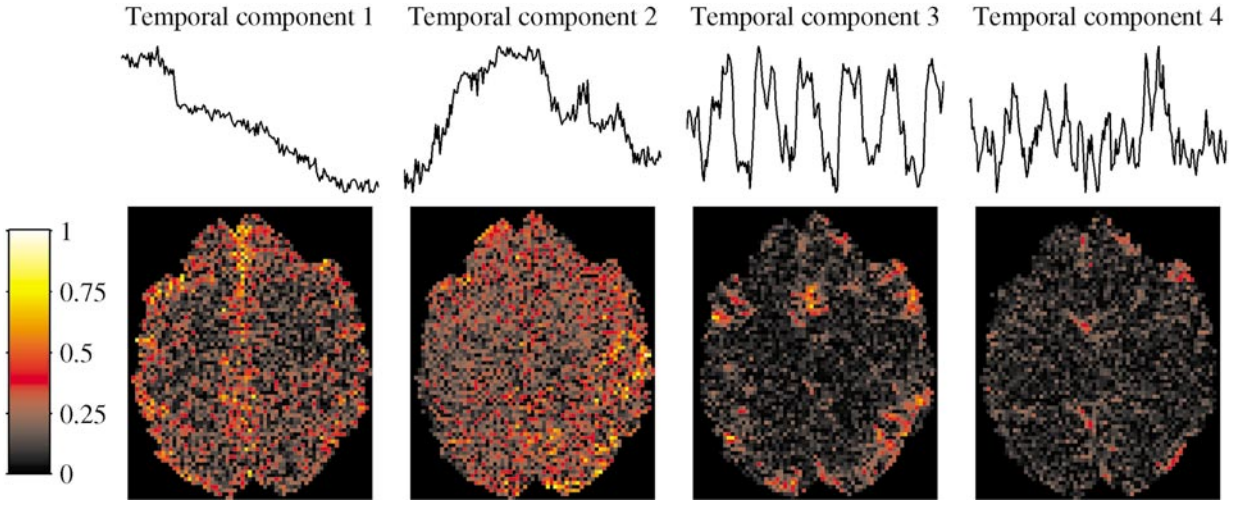

FIG. 7. Correlation maps obtained when the four first temporal CCA components were correlated with the original voxel timecourses. The maps indicate where in the analyzed fMRI slice the extracted CCA timecourses can be found (air voxels were excluded from analysis).

were 180 in this case) but we still reduce the dimensionality by keeping the first 30 "eigen-images" obtained by the spatial PCA analysis. The reason for reducing the dimensionality is partly justified by reliability considerations discussed later. Another reason was that the proposed ordering of the ICA components (by non-Gaussianity) did not produce a satisfactory result and there was a need to screen the output from ICA and PCA manually. An ordering of the components is more important in the spatial case than in the temporal case since in general more components are generated. In Fig. 8, 4 of the 30 spatial components produced by each method are shown. In the CCA case the first four components are shown. The order provided by PCA and ICA did as mentioned not coincide with what visually appeared interesting. Therefore PCA components 1, 4, 9, and 7, and ICA components 5, 1, 2, and 8, were manually selected by a visual interestingness criterion and aligned with the closest CCA components. All three methods give the average image as a component, in the PCA and CCA cases as the first component. In the PCA case this is explained by the large spread of intensity values in the image. The CCA method extracts the average image since, as all images of natural objects, it has a large autocorrelation. The remaining components most likely indicate areas in the brain involved in the mental calculation task or perhaps draining veins. Despite different ordering, there are no dramatic differences between the components. As in the temporal case it is interesting to move over to the dual space, i.e., examining what the timecourses look like if the original fMRI timecourses are weighted according to the voxel values in the spatial components. Figure 9 shows the dual timecourses corresponding to the first four spatial CCA components. Since the first spatial component is the mean image the dual timecourse should be some global signal. The second and third components are probably task-related. The patterns in the fourth spatial component closely

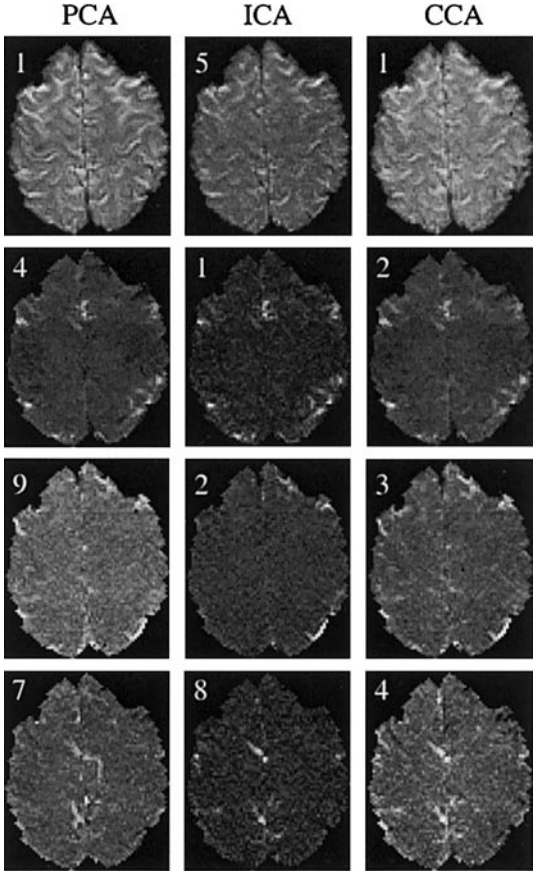

FIG. 8. Spatial components extracted by each method. The numbers indicate the original order. The displayed PCA and ICA components were selected manually.

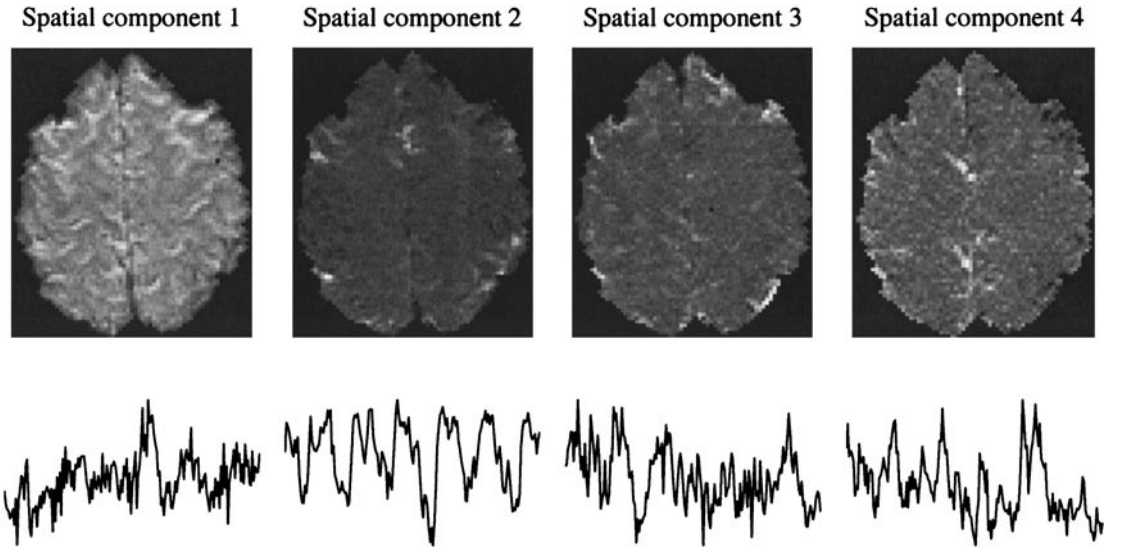

FIG. 9. The corresponding timecourses generated when the original fMRI timecourses are weighted according to the extracted spatial CCA components.

resembles the patterns in the fourth temporal component and is difficult to interpret.

The computational complexity is crucial for practical use of the method. In Table 1 the computational time in seconds for analyzing a single 128 × 128 fMRI slice with 180 timepoints is reported. The computations were performed on a 400 MHz SUN Ultra 10 Workstation. The absolute values are, however, not interesting since they depend on computer power, data size, and the dimension reduction performed. Rather it is the relative values that should be compared. Even though FastICA is a very fast implementation of ICA, the pure CCA method is obviously better in terms of computational complexity by at least an order of magnitude. This computational difference is however negligible in the temporal analysis since PCA is required as preprocessing. In the spatial analysis the CCA method is still much faster than ICA also when the PCA preprocessing is included.

DISCUSSION

Data-driven analysis methods are interesting complements to the more widely used hypothesis-driven methods. Here we have introduced a novel data-driven method for analyzing fMRI data. The appropriate comparison is between ICA and the presented CCA based method since it is quite evident that both methods outperform PCA. PCA should rather be seen as a required preprocessing step. This is in fact a weakness since the success of both the ICA and the CCA method crucially depend on the dimension reduction performed by PCA. In fMRI the interesting signals such as the hemodynamic response may be relatively weak and a PCA based dimension reduction may discard such a signal in favor for high variance noise signals. Thus we wish to keep as many principal components as possible to retain interesting low energy signals in the original data. However, if the number of mixtures is too large compared to the number of samples of each mixture, there is a risk of overfitting. In ICA, overfitting leads to components consisting of just a single spike at different locations (Hyvärinen *et al.*, 1999). Components with just a spike at different positions are uncorrelated, highly non-Gaussian and can be combined to any mixture, which thus is an ideal solution from an ICA view. If the CCA method is given too many mixtures it can combine nonsense noise signals to components with high autocorrelation due to the finite number of samples. Hence, there is a tradeoff between keeping as many principal components from PCA as possible in order not to lose interesting signals while not too many can be kept because of the risk of overfitting. Even if this issue is not specifically addressed here and the number of principal components retained is chosen by practical considerations, the CCA method tended to be more robust with respect to the low sample size in the temporal analysis. The reason may be that measuring autocorrelation does not require as many samples as measuring non-Gaussianity. A more practical consid-

TABLE 1

Computational Time in Seconds for Performing Temporal and Spatial Analysis of a 128 × 128 Voxels Slice

Method	Temporal	Spatial
PCA	40	30
ICA	0.5	110
CCA	<0.01	5

eration related to dimensionality reduction and the number of underlying components revealed is the ordering of the components. Since most ICA methods do not provide an ordering of the components they must be inspected manually. The CCA method orders the components by autocorrelation which is a natural measure of "interestingness" and thus only the first components need to be inspected. It should be noted that there are other similar methods that use temporal or spatial structure to decompose multidimensional data (Switzer and Green, 1984; Molgedey and Schuster, 1994; Ziehe and Müller, 1998; Stone, 2001) and that the CCA method is not novel in that sense. The CCA approach has however not been reported earlier and we believe it is well suited for exploratory analysis of fMRI data where temporal and spatial patterns are important features. Such patterns are ignored by PCA and ICA, which for example produce the same unmixing vectors after a random permutation of the samples in the mixtures.

An important issue that has to be considered before drawing any conclusions about the observed patterns in the obtained components is how well the mixture assumption apply. In the temporal case, the assumption that the observed voxel timeseries are combinations of different independent basis timeseries (hemodynamic response, low frequency drifts, etc.) is generally made and accepted in fMRI data analysis. However, in temporal exploratory analysis we are forced to perform heavy dimension reduction which may discard interesting information. In addition, the constraint to find uncorrelated component timecourses may introduce problems due to the finite number of samples in each timecourse. Even though the underlying components are assumed independent, the *sample* correlation between two finite length observations of the sources need not be zero. This is the case in the simulated experiment where the sample correlation between the boxcar and the quadratic signals is not zero, implying that it is impossible to perfectly separate them when there is a constraint to produce an uncorrelated result. The consequence may be an unsatisfactory separation with underlying signals fractionated into several components. In the spatial analysis the sample size is significantly larger. Hence there is a smaller probability that the sample correlation between two observations of hypothized independent basis images differ significantly from zero. Also, there may not be a need for reducing the dimensionality of the problem. The critical question is instead whether fMRI data can be assumed to consist of a linear mixture of underlying independent basis images, which is not an obvious assumption (McKeown and Sejnowski, 1998; Friston, 1998). The implication from the above discussion is that caution should be taken when interpreting and drawing inferences from the patterns seen in the components. Reported observations such as unexpected activations, transient task-related activations and infered spatial connectivity between different areas in the brain may very well be results of a violation of the mixture assumption, limited sample sizes or other artifacts such as patient motion.

The above issues do, however, by no means imply that exploratory analysis are not useful. The result may for example be used to build more accurate models of the drifts in the fMRI data set at hand, to detect residual motion artifacts or other unexpected components which for example can be included as additional regressors in a hypothesis-driven method (McKeown, 2000).

As a final point, the noise in fMRI data has repeatedly been shown to be temporally autocorrelated (Friston *et al.*, 2000; Zarahn *et al.*, 1997). Since the CCA method explore fMRI data for autocorrelated components the noise characteristics may have implications regarding the performance. The main effect of correlated noise is a reduced robustness against low sample sizes as it becomes easier to combine noise timecourses to something that can appear interesting. However, this is a general issue also hampering for example ICA (Hyvärinen *et al.*, 1999).

CONCLUSIONS

A method for data-driven analysis of fMRI data has been presented and compared primarily with Independent Component Analysis. The novel CCA method has several advantages over ICA. It is computationally efficient, the components are ordered by a natural measure of relevance so that not all components need to be inspected manually, and the same components are always obtained when the method is applied on the same data. Additionally, the CCA method performed robustly in the simulated study and appeared less sensitive to the high noise level when applied on the experimental fMRI data. It is therefore an interesting alternative to ICA as a tool for routinely screening the fMRI data prior to analyzing the data using a hypothesis-driven method.

REFERENCES

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

Anderson, T. W. 1984. *An Introduction to Multivariate Statistical Analysis*, 2nd ed. Wiley.

Bell, A. J., and Sejnowski, T. J. 1995. An information-maximization approach to blind separation and blind deconvolution. *Neural Computat.* **7:** 1129–1159.

Borga, M., and Knutsson, H. 2001. A canonical correlation approach to blind source separation. Technical Report LiU-IMT-EX-0062, Department of Biomedical Engineering, Linköping University.

Friston, K. 1998. Modes or models: A critique on independent component analysis for fMRI. *Trends Cogn. Sci.* **2**(10): 373–375.

Friston, K., Josephs, O., Zarahn, E., Holmes, A., Rouquette, S., and Poline, J. 2000. To smooth or not to smooth? Bias and efficiency in fMRI time-series analysis. *NeuroImage* **12**(2): 196–208.

Hotelling, H. 1936. Relations between two sets of variates. *Biometrika* **28:** 321–377.

Hyvärinen, A. 1999. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Trans. Neural Networks* **10**(3): 626–634.

Hyvärinen, A., Karhunen, J., and Oja, E. 2001. *Independent Component Analysis.* Wiley.

Hyvärinen, A., and Oja, E. 2000. Independent component analysis: Algorithms and applications. *Neural Networks* **13:** 411–430.

Hyvärinen, A., Särelä, J., and Vigaro, R. 1999. Spikes and bumps: Artefacts generated by independent component analysis with insufficient sample size. In *Proc. Int. Workshop on Independent Component Analysis and Signal Separation (ICA '99)*

McKeown, M. 2000. Detection of consistently task-related activations in fMRI data with hybrid independent component analysis. *NeuroImage* **11**(1): 24–35.

McKeown, M., Makeig, S., Brown, G., Jung, T., Kindermann, S., Bell, A., and Sejnowski, T. 1998. Analysis of fMRI data by blind separation into independent spatial components. *Hum. Brain Mapp.* **6**(3): 160–188.

McKeown, M., and Sejnowski, T. 1998. Independent component analysis of fMRI data: Examining the assumptions. *Hum. Brain Mapp.* **6**(5–6): 368–372.

Molgedey, L., and Schuster, H. 1994. Separation of a mixture of independent signals using time delayed correlations. *Phys. Rev. Lett.* **72**(23): 3634–3636.

Stone, J. 2001. Blind source separation using temporal predictability. *Neural Computat.* **13**(7):1559–1574.

Switzer, P., and Green, A. A. 1984. Min/max autocorrelation factors for multivariate spatial imagery. Technical Report 6, Department of Statistics, Stanford University.

Zarahn, E., Aguirre, G., and D'Esposito, M. 1997. Empirical analyses of BOLD fMRI statistics, spatially unsmoothed data collected under null-hypothesis conditions. *NeuroImage* **5**(3): 179–197.

Ziehe, A., and Müller, K.-R. 1998. TDSEP—An efficient algorithm for blind separation using time structure. In *Proc. Int. Conf. on Artificial Neural Networks (ICANN '98),* pp. 675–680.