Magnetic Resonance in Medicine 45:323–330 (2001)

Detection of Neural Activity in Functional MRI Using Canonical Correlation Analysis

Ola Friman,[1*] Jonny Cedefamn,[2] Peter Lundberg,[3] Magnus Borga,[1] and Hans Knutsson[1]

A novel method for detecting neural activity in functional magnetic resonance imaging (fMRI) data is introduced. It is based on canonical correlation analysis (CCA), which is a multivariate extension of the univariate correlation analysis widely used in fMRI. To detect homogeneous regions of activity, the method combines a subspace modeling of the hemodynamic response and the use of spatial relationships. The spatial correlation that undoubtedly exists in fMR images is completely ignored when univariate methods such as as *t*-tests, *F*-tests, and ordinary correlation analysis are used. Such methods are for this reason very sensitive to noise, leading to difficulties in detecting activation and significant contributions of false activations. In addition, the proposed CCA method also makes it possible to detect activated brain regions based not only on thresholding a correlation coefficient, but also on physiological parameters such as temporal shape and delay of the hemodynamic response. Excellent performance on real fMRI data is demonstrated. Magn Reson Med 45:323–330, 2001. © 2001 Wiley-Liss, Inc.

Key words: fMRI; canonical correlation analysis (CCA); multivariate analysis; spatial filtering; *t*-test; *F*-test

The functionality of the human brain is still relatively unknown, although much effort has been put into revealing its secrets. A relatively new and promising tool for this purpose is fMRI. The goal in fMRI is to map sensory, motor, and cognitive functions to specific areas in the brain.

Various statistical methods have been used to detect signal variation according to a known paradigm in the fMRI data. This signal variation is often referred to as the blood oxygen level dependent (BOLD) signal because it originates from different levels of blood oxygenation between activated and non-activated regions in the brain. Among the most commonly used methods to detect activity are Student's *t*-test and ordinary correlation analysis, which can be considered equivalent (1). A *t*-test is optimal when detecting a certain, known signal in Gaussian white noise. In fMRI datasets, however, the noise is in general neither Gaussian-distributed nor white (2), and there is no known signal to search for. The hemodynamic response to a task or stimulus is unknown, and it must for this reason be explicitly modeled prior to the analysis. However, experiments have shown that the hemodynamic response differs from person to person, from day to day, and probably also between different regions of the brain (3). Thus, the above methods are suboptimal as tools for robust and sensitive analysis of neurological responses.

If an *F*-test, as outlined in Ref. 1, is applied, the requirement of an explicit model of the hemodynamic response is considerably relaxed. The search will then be focussed on an *unknown* signal in Gaussian white noise. Instead, the signal is assumed to be defined within a known signal subspace. Thus, it is sufficient to model a signal subspace in which we have reason to believe the hemodynamic response should reside.

A characteristic of the univariate methods mentioned above is that single pixels are considered separately. The result is a map of a single statistical parameter. This map is subjected to a threshold chosen to ensure a low probability for false alarms. The pixels above this threshold are then classified as activated, and subject to subsequent quantifications such as volume or pixel count. Due to the high intrinsic noise in the fMR images and the small effect of BOLD on signal intensity, these single-pixel methods often fail to detect homogeneous regions of activity. When considering only one pixel timecourse at a time, all spatial dependencies that exist in an MR image are ignored (4). Sometimes spatial smoothing of the fMR images prior to statistical analysis is applied as a remedy to reduce noise. Postprocessing, such as smoothing of the statistical parameter maps, prior to the thresholding procedure is used as well. This is the approach in the widely used SPM method (5).

We propose a method based on canonical correlation analysis to be used as a neurological activation evaluation tool in fMRI. Canonical correlation analysis (CCA) is an extension of the ordinary correlation analysis, the difference being that CCA operates on multidimensional variables, i.e., it is a multivariate technique. The method combines subspace modeling of the hemodynamic response and the use of spatial dependencies to detect homogeneous maps of brain activity.

It should be pointed out that variations of CCA, as well as other multivariate methods, have earlier been suggested for the analysis of fMRI data (6,7). In contrast to these methods, which use a global view of the detection problem, our approach is more focussed on localized regions.

THEORY

In statistics, canonical correlation analysis is a well known method, developed by Hotelling in 1936 (8).

Consider two multidimensional random vectors $\mathbf{x}$ and $\mathbf{y}$ which are m- respectively n-dimensional. We seek linear combinations

[1]Department of Biomedical Engineering, Linköping University, Linköping, Sweden.

[2]Department of Neuroscience and Locomotion, University Hospital of Linköping, Linköping, Sweden.

[3]Departments of Radiation Physics and Diagnostic Radiology, University Hospital of Linköping, Linköping, Sweden.

Grant sponsors: Swedish Natural Science Research Council; Wallenberg Foundation; Swedish Research Council for Engineering Sciences; University Hospital Research Foundation.

*Correspondence to: Ola Friman, Linköping University, Dept. of Biomedical Engineering, University Hospital, S-581 85 Linköping, Sweden.
E-mail: olafr@imt.liu.se

Received 8 May 2000; revised 18 July 2000; accepted 19 September 2000.

$$X = w_{x_1}x_1 + \cdots + w_{x_m}x_m = \mathbf{w}_x{}^T\mathbf{x} \quad [1]$$

$$Y = w_{y_1}y_1 + \cdots + w_{y_n}y_n = \mathbf{w}_y{}^T\mathbf{y} \quad [2]$$

which are chosen so that X and Y correlate the most. The linear combination coefficients are gathered in the two vectors $\mathbf{w}_x$ and $\mathbf{w}_y$. ^{T} denotes a transpose. X and Y are called *canonical variates.* Thus, we attempt to solve

$$\max_{\mathbf{w}_x,\mathbf{w}_y} \rho = \frac{C[X,\ Y]}{\sqrt{V[X]V[Y]}}. \quad [3]$$

Since both variance (V) and covariance (C) are invariant with respect to an addition of a constant, we can without loss of generality assume a zero mean. The expression for the correlation ρ in Eq. [3] can then be written

$$\begin{aligned}\rho &= \frac{E[XY]}{\sqrt{E[X^2]E[Y^2]}} \\ &= \frac{E[\mathbf{w}_x{}^T\mathbf{x}\mathbf{y}^T\mathbf{w}_y]}{\sqrt{E[\mathbf{w}_x{}^T\mathbf{x}\mathbf{x}^T\mathbf{w}_x,]E[\mathbf{w}_y{}^T\mathbf{y}\mathbf{y}^T\mathbf{w}_y,]}} \\ &= \frac{\mathbf{w}_x{}^T\mathbf{C}_{xy}\mathbf{w}_y}{\sqrt{\mathbf{w}_x{}^T\mathbf{C}_{xx}\mathbf{w}_x\mathbf{w}_y{}^T\mathbf{C}_{yy}\mathbf{w}_y}} \end{aligned} \quad [4]$$

where $\mathbf{C}_{xx}$ and $\mathbf{C}_{yy}$ are the within-set covariance matrices and $\mathbf{C}_{xy}$ is the between-sets covariance matrix. By taking the partial derivatives of Eq. [4] with respect to $\mathbf{w}_x$ and $\mathbf{w}_y$ we arrive in the following system of equations

$$\begin{cases} \mathbf{C}_{xy}\hat{\mathbf{w}}_y = \rho\lambda_x\mathbf{C}_{xx}\hat{\mathbf{w}}_x \\ \mathbf{C}_{yx}\hat{\mathbf{w}}_x = \rho\lambda_y\mathbf{C}_{yy}\hat{\mathbf{w}}_y \end{cases} \quad [5]$$

where

$$\lambda_x = \lambda_y{}^{-1} = \sqrt{\frac{\hat{\mathbf{w}}_y{}^T\mathbf{C}_{yy}\hat{\mathbf{w}}_y}{\hat{\mathbf{w}}_x{}^T\mathbf{C}_{xx}\hat{\mathbf{w}}_x}}. \quad [6]$$

A simple scaling of the linear combination coefficients in Eq. [1] and Eq. [2] would not affect the correlation coefficient. The norms of $\mathbf{w}_x$ and $\mathbf{w}_y$ are for this reason free of choice. A natural choice is $\|\mathbf{w}_x\|_2 = \|\mathbf{w}_y\|_2 = 1$, which explains the normalized vectors $\hat{\mathbf{w}}_x$ and $\hat{\mathbf{w}}_y$ in Eq. [5].

A solution to the equation system in Eq. [5] can be obtained from the equations below:

$$\mathbf{C}_{xx}{}^{-1}\mathbf{C}_{xy}\mathbf{C}_{yy}{}^{-1}\mathbf{C}_{yx}\hat{\mathbf{w}}_x = \rho^2\hat{\mathbf{w}}_x \quad [7]$$

$$\mathbf{C}_{yy}{}^{-1}\mathbf{C}_{yx}\mathbf{C}_{xx}{}^{-1}\mathbf{C}_{xy}\hat{\mathbf{w}}_y = \rho^2\hat{\mathbf{w}}_y. \quad [8]$$

Thus $\mathbf{w}_x$ and $\mathbf{w}_y$ are eigenvectors to the matrices $\mathbf{C}_{xx}{}^{-1}\mathbf{C}_{xy}\mathbf{C}_{yy}{}^{-1}\mathbf{C}_{yx}$ and $\mathbf{C}_{yy}{}^{-1}\mathbf{C}_{yx}\mathbf{C}_{xx}{}^{-1}\mathbf{C}_{xy}$, and the eigenvalues ρ^2 are the squared *canonical correlations.*

In practice, Eq. [7] is solved for $\hat{\mathbf{w}}_x$ and ρ, which are then inserted in Eq. [5] to obtain $\hat{\mathbf{w}}_y$. Note that we obtain several solutions, as there are several eigenvectors to the above matrices. In fact, we obtain $\min\{m, n\}$ solutions, but the eigenvectors $\mathbf{w}_{x,1}$ and $\mathbf{w}_{y,1}$ corresponding to the largest eigenvalue, i.e., the largest canonical correlation, are those that solve Eq. [3]. Generally, there may be several solutions with a high canonical correlation, but in this application we will only consider the solution with the largest canonical correlation.

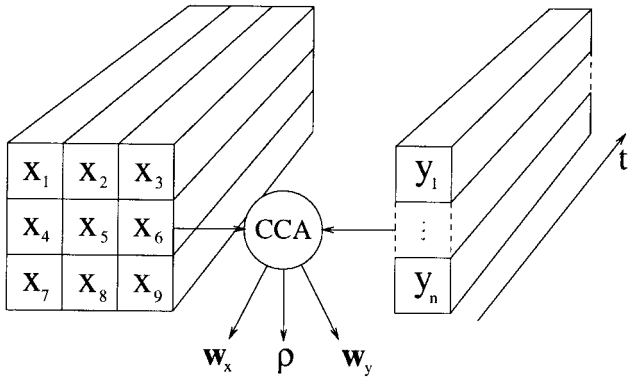

FIG. 1. Canonical correlation analysis between a 3 × 3 region in the fMR image and a set of *n* basis-functions.

Other important properties are that the canonical correlations can always be defined as positive and that the canonical variates are uncorrelated for different solutions, i.e.,

$$\begin{aligned} E[X_iY_j] &= \mathbf{w}_{x,i}{}^T\mathbf{C}_{xy}\mathbf{w}_{y,j} = 0 \\ E[X_iX_j] &= \mathbf{w}_{x,i}{}^T\mathbf{C}_{xx}\mathbf{w}_{x,j} = 0 \quad \text{for } i \neq j \\ E[Y_iY_j] &= \mathbf{w}_{y,i}{}^T\mathbf{C}_{yy}\mathbf{w}_{y,j} = 0. \end{aligned}$$

For more details on canonical correlation analysis, see Refs. 9 and 10.

METHODS

Signal Processing

Assume that an fMRI experiment has been performed in which a number of image slices are acquired at N subsequent timepoints. In each pixel in each image slice we obtain a timeseries of length N. In an activated region of the brain there should be a small signal increase during task performance due to the BOLD effect. We are searching for pixels whose timeseries has a component that follows the paradigm. However, due to the low signal-to-noise ratio we will instead consider a region of pixels to make use of the spatial relationships between pixels. This region becomes our multidimensional **x**-variable in Eq. [1]. In this work, a 3 × 3 region is chosen leading to a nine-dimensional **x**-variable, i.e., $m = 9$ (Fig. 1). The measured timeseries will be referred to as $\mathbf{x}(t) = [x_1(t), \ldots, x_9(t)]^T$.

Next a signal subspace is modeled in such a way that it is able to represent the range of hemodynamic response signals well. This is done by choosing a set of basis-functions that span the signal subspace, here denoted $\mathcal{Y}$. These basis-functions act as **y**-variable on the right side in Fig. 1. Doing so is a rather ad hoc simplification, since it is not well known what the hemodynamic response in an activated region is. Note that if only a single pixel is chosen as **x**-variable, and a single basis-function is used as

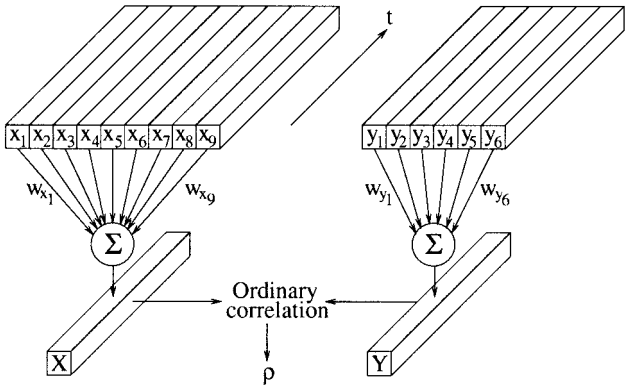

FIG. 2. The CCA finds the linear combination coefficients $w_{x_1} \ldots w_{x_9}$ and $w_{y_1} \ldots w_{y_6}$, which gives the largest correlation between $X(t)$ and $Y(t)$.

y-variable, the method degenerates to become an ordinary correlation analysis.

We use a symmetric square-wave paradigm, i.e., the time periods of rest and task performance are of equal length. It is then likely to assume that the response will have the same fundamental frequency as the paradigm. It is probably also a good choice to include components with frequencies equal to a following few harmonics in the Fourier series expansion of the symmetric square-wave. The use of a truncated Fourier series as basis is by no means new. Such a basis has previously been used (11). To make the following discussions in this work more explicit, we have for the sake of clarity chosen the following simple basis:

$$\mathbf{y}(t) = \begin{pmatrix} \sin(\omega t) \\ \cos(\omega t) \\ \sin(3\omega t) \\ \cos(3\omega t) \\ \sin(5\omega t) \\ \cos(5\omega t) \end{pmatrix}, \quad \omega = \frac{2\pi}{T}, \quad t = 1 \ldots N$$

(i.e., the Fourier-series expansion of a square-wave, which contains odd harmonics only) where T is the period in the paradigm and t is a discrete time index. This choice gives a six-dimensional **y**-variable in Fig. 1, i.e., $n = 6$. The underlying assumption is that the hemodynamic response can be modeled as the output from a linear lowpass system when the square-wave paradigm is used as input. The frequencies appearing on the input and output side are then the same, and there should be little power in the frequencies beyond the fifth harmonic. The signal subspace $\mathcal{Y}$ consists of all signals that can be written as a linear combination of the basis-functions.

Figure 2 elucidates what the CCA will do. Linear combinations of the pixel-timeseries, $X(t) = \mathbf{w}_x{}^T\mathbf{x}(t)$, and basis-functions, $Y(t) = \mathbf{w}_y{}^T\mathbf{y}(t)$, are found so that the correlation between $X(t)$ and $Y(t)$ is the largest achievable.

The sample canonical correlations in each 3×3 neighborhood are found by replacing $\mathbf{C}_{xx}$, $\mathbf{C}_{yy}$, and $\mathbf{C}_{xy}$, with the maximum likelihood estimates $\mathbf{S}_{xx}$, $\mathbf{S}_{yy}$, and $\mathbf{S}_{xy}$ in Eq. [7]. $\mathbf{S}_{yy}$ is constant and can be precalculated. Solving the eigenvalue problem in Eq. [7] will give $\mathbf{w}_x$ and ρ, and inserting these into Eq. [5] gives $\mathbf{w}_y$. This procedure results in six sample canonical correlations, $\hat{\rho}_i$ $i = 1 \ldots 6$, each with a corresponding $\mathbf{w}_x$ and $\mathbf{w}_y$. The largest canonical correlation is assigned to the center pixel in each 3×3 neighborhood, resulting in a map of correlation values similar to the statistical maps obtained when applying any univariate method. The corresponding $\mathbf{w}_x$ and $\mathbf{w}_y$ are also stored for later use. One can view this as finding a filter in each 3×3 neighborhood, and a signal in $\mathcal{Y}$, matched in the sense that the output of this filter on the 3×3 neighborhood correlates the most with the signal found in our signal subspace $\mathcal{Y}$.

Detecting Activated Pixels

Substantially more information is obtained using CCA compared to ordinary correlation analysis. The largest canonical correlation coefficient is a qualitative measure of how well the timeseries in the 3×3 neighborhood corresponded to the optimal signal $Y(t)$, that of the CCA found in $\mathcal{Y}$. A large correlation implies a high degree of similarity, and therefore an activated central pixel. A low correlation value means that it was not possible to find a signal in $\mathcal{Y}$ that had any similarity to the timecourses in the neighborhood. The largest canonical correlation coefficient ρ_1, is thus the most obvious measure for distinguishing non-activated pixels from possibly activated. At each pixel, we would like to test the hypothesis $H_0 : \rho_1 = 0$ against $H_1 : \rho_1 > 0$. A natural decision is to reject H_0 if the largest sample canonical correlation exceeds some threshold ρ_t, i.e., reject H_0 if $\hat{\rho}_1 > \rho_t$.

Postprocessing

The sine/cosine basis gives an opportunity to further reject spurious activated pixels. Noise can just by pure chance create a pixel timeseries in a non-activated area with high correlation to some signal in $\mathcal{Y}$. Apart from the canonical correlations, we obtained the linear combination coefficients in $\mathbf{w}_y$. With the chosen basis, the signal $Y(t)$ in Fig. 2 becomes

$$Y(t) = w_{y_1}\sin(\omega t) + w_{y_2}\cos(\omega t) + w_{y_3}\sin(3\omega t) + w_{y_4}\cos(3\omega t) + w_{y_5}\sin(5\omega t) + w_{y_6}\cos(5\omega t). \quad [9]$$

Applying the basic relationship

$$a\sin(v) + b\cos(v) = r\sin(v + \phi) \quad [10]$$

to Eq. [9] converts each cosine-sine basis pair to an amplitude r, and a phase ϕ, for each frequency component.

$$Y(t) = r_1\sin(\omega t + \phi_1) + r_2\sin(3\omega t + \phi_2) + r_3\sin(5\omega t + \phi_3). \quad [11]$$

The relative amplitudes of the frequency components determine a general shape of $Y(t)$, while the phases can be interpreted as the time delay in each frequency component.

To illustrate the significance of the shape, consider Fig. 3, in which three different signals in $\mathcal{Y}$ are shown. The phase of the three different frequencies are the same, but

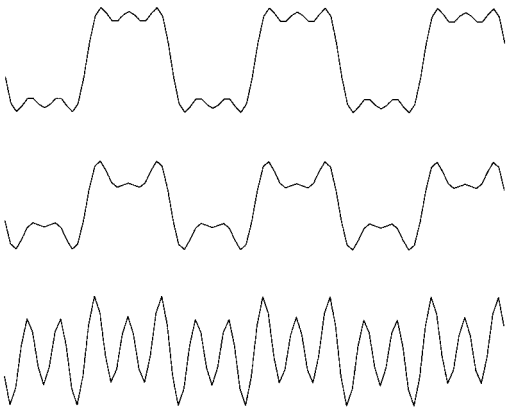

FIG. 3. Three signals that belong to our chosen subspace $\mathcal{Y}$. Top: The square paradigm projected on $\mathcal{Y}$. Middle: A $Y(t)$ which is reasonable as a activation-response to the paradigm. Bottom: A $Y(t)$ from a nonactivated region.

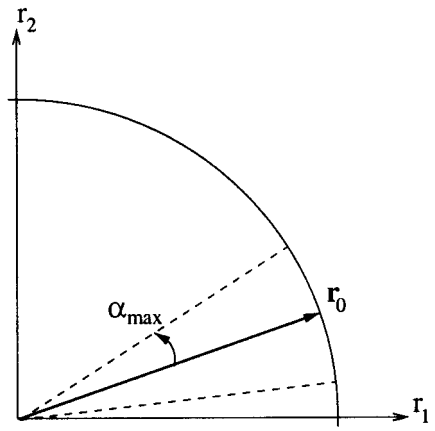

FIG. 4. A measure of similarity between vectors is the angle between them. A normalized vector lies on the unit sphere. We can specify an angle α_{max} as the maximum allowed deviation between a vector $\mathbf{r} = (r_1, r_2)^T$ and a reference vector $\mathbf{r}_0$.

the relative amplitudes of the frequency components differ. Clearly, the bottom plot in Fig. 3 is not a response generally thought to represent true neurological activation. However, it is a perfectly valid signal belonging to $\mathcal{Y}$. To reject unwanted responses a reference response must be used, such as the unknown "ideal" hemodynamic response, from which a measure of distance can be calculated. The true hemodynamic response remains unknown, but in this context it is not critical as we only attempt to *reject* false detections. The square paradigm projected on $\mathcal{Y}$ will be sufficient as a reference response (Fig. 3, top).

First, three dimensional vectors $\mathbf{r}$ of the different frequency amplitudes of $Y(t)$ in Eq. [11] are formed. The vector of amplitudes belonging to the reference response is denoted $\mathbf{r}_0$. Then the angle $\alpha = \arccos \mathbf{r}^T\mathbf{r}_0/(\|\mathbf{r}\|_2\|\mathbf{r}_0\|_2)$ between $\mathbf{r}$ and the reference $\mathbf{r}_0$ becomes a measure of similarity between $\mathbf{r}$ and $\mathbf{r}_0$. For simplicity this is illustrated in two dimensions in Fig. 4. The result is a map of angle values similar to the map of correlation values. We may then specify an angle α_{max}, which is the limiting angle we allow a vector of amplitudes $\mathbf{r}$ to deviate from the reference vector $\mathbf{r}_0$. An amplitude is always positive, and therefore the largest obtainable angle is $\pi/2$ radians. For example, the angles in Fig. 3 between the signals in the middle and bottom panels, and the top panel used as a reference are 0.27 and 1.13 radians, respectively. Obviously, the signal in the middle panel is much more similar to the reference in the top panel than the lower signal.

For obvious reasons there should be no task-related activity in the brain *before* the task has begun. Hemodynamic response delays from 2–6 sec have previously been reported (3). It is natural to constrain the delay we allow. As mentioned previously, the delay of the hemodynamic response can be determined accurately from the phases of the different frequency components used as basis-functions in $\mathcal{Y}$. The phase of fundamental frequency, i.e., ϕ_1 in Eq. [11] has the largest impact on the delay, and therefore we confine ourselves to using the phase difference in the fundamental frequency between the paradigm and $Y(t)$ (Fig. 5). Again, the phases are obtained from the relation in Eq. [10].

If it is assumed that the task-related hemodynamic response would not have a delay larger than δ sec, this restriction is transformed into a maximum allowed phase difference $\Delta\phi$,

$$0 \leq \Delta\phi \leq \delta \frac{2\pi}{T \cdot t_{TR}} \qquad [12]$$

where T is the number of samples in one period of the paradigm, and t_{TR} is the repetition time, i.e., the time delay between two consecutive acquisitions of the same image slice.

In conclusion, we have obtained two parameters that can be adjusted to reject possible false activations. α_{max} rejects pixels based on shape and δ rejects pixels based on response delay. These parameters are calculated from the components in $\mathbf{w}_y$, and the procedure can be seen as if we define a subspace of $\mathcal{Y}$, outside of which we do not consider $Y(t)$ to be a valid response to the paradigm.

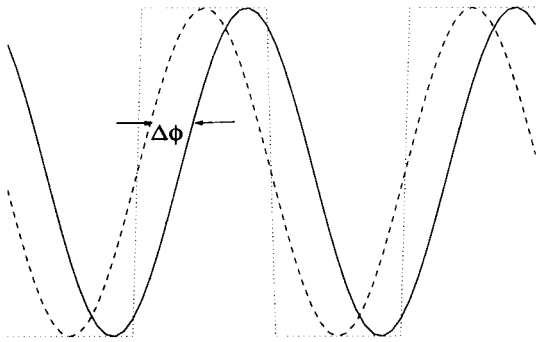

FIG. 5. The difference in phase between the fundamental frequency of $Y(t)$ (solid) and the fundamental frequency of the paradigm (dashed) can be transformed into a time delay of the hemodynamic response. The dotted signal is the square-wave paradigm.

RESULTS

Materials

In order to illustrate the CCA method, it was applied on fMRI data from an experiment performed on a male volunteer. The paradigm used was a single-finger movement using the right-hand index finger, alternated by rest when no task was performed. The test subject was fixated using a custom-made bite device and asked to relax and close his eyes during the experiment.

Functional MRI data were obtained using a GE Signa Horizon Echospeed 1.5 T human scanner. Following the acquisition of sagittal localizer images, 7 oblique (slope: anterior-superior and posterior-inferior) slices were defined. Acquisition was performed using the EPIBOLD sequence (GE Medical). Acquisition parameters were: TR 2 sec; TE 60 msec; acquisition matrix 128 $\times$ 128; FOV 20 cm; slice thickness 3 mm with no interslice spacing; number of time-points 200. Five scans preceded the actual data acquisition in order to allow tissue to reach steady-state magnetization. Ten scans were acquired during each period of rest and task-performance, i.e., the paradigm period was 20 scans. Image registration was performed using the AFNI processing package (12). Data was further processed as described in the previous sections using custom software developed in MATLAB 5.3.

Results

To compare the results of our implementation of the CCA method to existing methods, a *t*-test and an *F*-test were also applied (as outlined in Ref. 1). The reference signal used in the *t*-test was the square paradigm shifted three samples to compensate for the time delay in the hemodynamic response. In the *F*-test, the same subspace $\mathcal{Y}$ was used as in the CCA method. Thresholds for all tests were theoretically calculated to obtain a probability for assigning a non-activated pixel as being activated approximately equal to 10^{-4}. Selecting a threshold for the largest canonical correlation is somewhat more complicated, but, as shown in the Discussion section, $\rho_t = 0.65$ yields an approximate lower bound at 10^{-4}. α_{max} and δ were set so that the shape and delay of the hemodynamic response were not too constraining: $\alpha_{max} = 0.35$ radians and $\delta = 10$ sec.

The result of the three methods applied on three adjacent image slices is shown in Fig 6. The *t*-test is shown in the top panel, the *F*-test in the middle, and the result of the CCA method is shown in the bottom panel. It is obvious that, in contrast to the *t*- and *F*-tests, the CCA method detects homogeneous regions of activity. The emphasis in this work is on the CCA method as such, and therefore the interpretation of the detected activated regions is left for future, more detailed neurological analysis.

Figure 7 shows the detected hemodynamic response obtained as an average of the activated pixel timecourses. The dotted line is the experimental paradigm. The delay in the hemodynamic response is estimated to be approximately 6 sec by calculating the difference in phase of the fundamental frequency of the signals in Fig. 7. This explains the three-sample shift of the reference signal used in the optimized *t*-test (as this corresponded to 6 sec in the experiment setup using TR = 2 sec).

DISCUSSION

The main aim of this work is to introduce a novel method based on CCA as a sensitive, robust, and reliable tool in fMRI data analysis. Robustness and reliability are of particular importance in clinical applications, and sensitivity is required because of the relatively limited signal variation caused by the BOLD effect. We will briefly discuss some important topics and extensions of the method.

Analyzing Regions vs. Single Pixels

When using univariate methods, such as *t*-tests or ordinary correlation analysis, one is often forced to smooth the statistical maps heavily before thresholding to obtain homogeneous regions of neural activity. Another method often used is a so-called "single pixel filter" to remove isolated and apparently false activations.

It can be argued that when analyzing regions instead of single pixels, one may fail to detect highly localized activations. However, suppose that activation is observed in single pixels only, that x_1 in Fig. 1 is such an isolated activated pixel, and the remaining x_2 to x_9 are nonactivated. If $\mathcal{Y}$ is chosen properly, the coefficient $w_{x,1}$ will be large and the center pixel x_5 will probably be erroneously classified as being activated, as the canonical correlation coefficients for this region will be large. Thus, the method will result in the detection of activated regions that appear to be somewhat larger than what is true in a neurological sense. Clearly, the approach will not fail to find such small activated regions. In contrast, this would occur if the data, using a common approach, were spatially smoothed prior to analysis. The underlying assumption when smoothing fMRI data or statistical maps is that the regions we are attempting to detect have some spatial extent. This is plausible since the experimentally observed fMRI-signals originate through influence of blood oxygenation level on intravascular water relaxation time. The blood oxygenation level generally can not change abruptly between pixels. This is exploited by the CCA, which adaptively filters a region in order to reduce noise and to extract a desired signal. An exception is when a vessel, which sometimes can give strong and spatially compact BOLD signals, is encountered. The CCA method will then detect a larger activated region due to the reasons discussed above. Below we give a possible extension of the method which is able to reduce such enlargements.

The coefficients in $\mathbf{w}_x$ determine how much each of the original timeseries contributes to $X(t)$ in Fig. 2. These can be used to increase the spatial resolution of the detection. The reason for not using the coefficients in $\mathbf{w}_x$ in the present method is that the spatial coherence in the fMR images makes the timecourses in the 3 $\times$ 3-pixel neighborhood correlated. This results in ambiguous components in $\mathbf{w}_x$. As a (degenerate) example, suppose that $x_1(t)$ and $x_2(t)$ in Fig. 2 are *exactly* equal. Then the optimal $\mathbf{w}_x$ is not unique because both $\mathbf{w}_x = [1, 0, \ldots]^T$ and $\mathbf{w}_x = [0, 1, \ldots]^T$ will result in the same canonical correlations. Thus, one can not tell with certainty (as one would wish) which pixels in the neighborhood were the most influential just by examining the resulting linear combination coefficients in $\mathbf{w}_x$. For a related discussion see Ref. 13. A

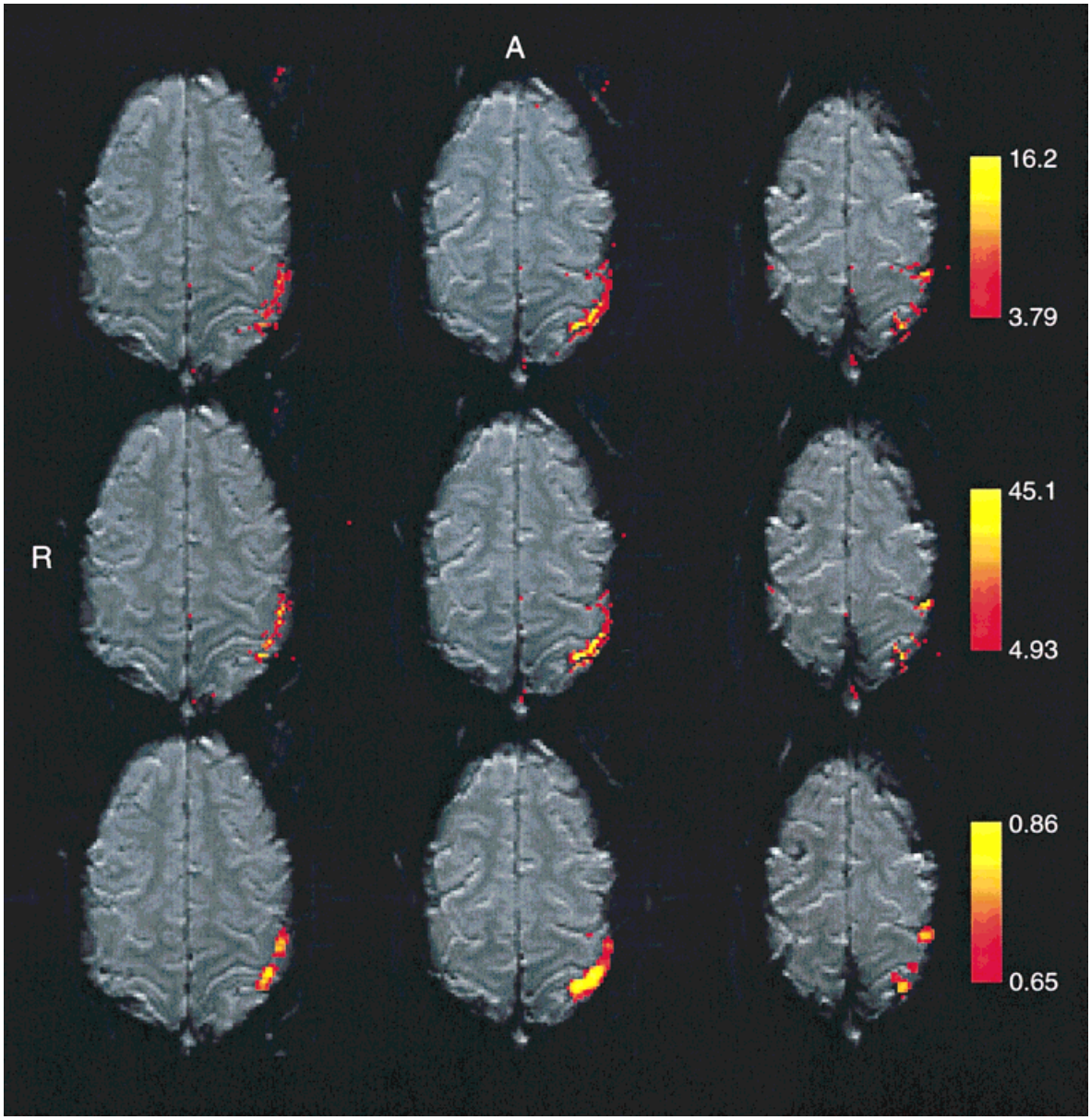

FIG. 6. Three image slices analyzed with three different methods. From the top: a t-test, an *F*-test, and the CCA method described in this work at the bottom row. The legends indicate the color mapping of the *t*-, F-, and correlation values. In contrast to the *t*- and *F*-tests, the CCA method detects homogeneous regions of activity with no obvious spurious activations. Note especially the left-most slice, where the CCA method clearly detects two different regions of activity. Such a pattern can not be detected by a *t*-test or *F*-test.

remedy is to calculate the so-called *canonical loadings,* which often are used to interpret the result from a CCA. The canonical loadings are simply the ordinary correlation coefficients between the resulting $X(t)$ and each of the original timeseries. A large canonical loading implies that the specific timeseries has a lot in common with $X(t)$. It is obviously desirable that the center pixel has a large canonical loading since the final result is assigned to this pixel. Otherwise, we are probably at the edge of an activated region or next to a vessel, and activity should not be assigned to the center pixel. This extension of the method shows promising results. Note that such problems are avoided in $\mathbf{w}_y$ since the sines and cosines are uncorrelated when the paradigm consists of a whole number of periods. The coefficients in $\mathbf{w}_y$ can for this reason not interfere as in $\mathbf{w}_x$, which makes the interpretation more straightforward.

It is also important to consider the size of the region used as $\mathbf{x}$-variable. The size is a tradeoff between sensitivity and specificity. If a smaller image size is used, e.g., 64×64 pixels, one might consider removing the corner pixels of the 3×3 region used here. Other regions can also be chosen. Under suitable experimental conditions (thin slices and no interslice spacing), pixels from the slices above and below can be included. However, the data compilation is then complicated by acquisition delays between neighboring slices.

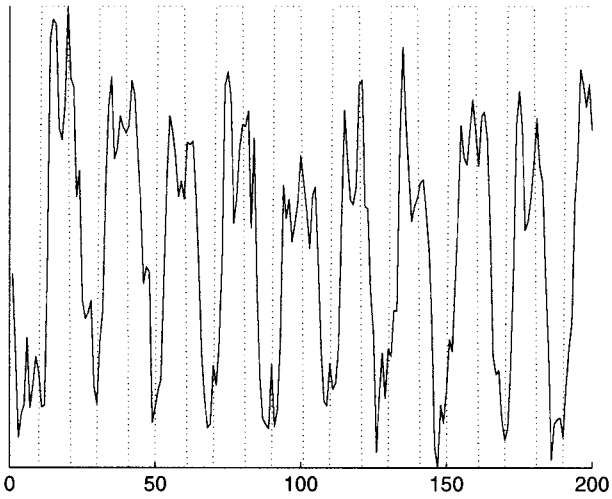

FIG. 7. The detected hemodynamic response in the activated pixels. The dotted line shows the experimental paradigm.

Basis Selection

The sine functions used as basis in $\mathcal{Y}$ were chosen under the assumption that the hemodynamic response is the output from a linear lowpass system using a square-wave input. However, previously published results (5,14) show that the response may be more adequately modeled by a nonlinear system, and for this reason also would contain even harmonics. Even harmonics, as well as any other components that are believed to be a part of the hemodynamic response, can easily be added to the basis. A postprocessing step can, however, become complicated if a non-orthogonal basis is used. Furthermore, if we add more dimensions to $\mathcal{Y}$ by adding more basis-functions, the constraints on the shape of the hemodynamic response would be more relaxed. At the same time, the discriminative power of the correlation coefficient would be weaker. This is a consequence of the additional degrees of freedom to construct arbitrary signals. Those signals may simply be undesired noise or other nuisance timecourses. Consequently, the postprocessing step will become more important when adding more basis functions.

Nuisance Subspace

Subspace methods sometimes introduce a so-called nuisance subspace (11,15), which is intended to represent signals of no neurological interest, such as the cardiac and respiratory cycles or low frequency drifts. The purpose of such an approach is to completely remove the temporal correlation so that the signal (which can not be explained by the basis-functions or nuisance components) consists only of Gaussian-distributed white noise. It is straightforward to incorporate a nuisance subspace into the CCA method. The nuisance components can be placed either with the fMRI-timeseries, (the x-variables) or as an extension to the basis-functions (the y-variables). Adding the nuisance components to the basis-functions will also result in high correlations when there is just nuisance noise in the 3×3 region and no component is evoked by the paradigm. For this reason, it is probably a better choice to add the nuisance components to the fMRI timeseries, in which nuisance noise is selectively removed when necessary. Ambiguity in the canonical correlation coefficients is thereby avoided.

Statistical Significance

An important issue is the statistical significance of the result. A theoretical distribution of the largest canonical correlation coefficient under the assumptions of normality and independent observations exists (16). However, this distribution is intractable. As an alternative to largest canonical correlation statistic, a number of approximate multivariate statistical tests are applicable. One of the most commonly used is Wilks' Lambda (9), which is defined as

$$\Lambda = \Pi_i(1 - \hat{\rho}_i^2)^{-1}. \qquad [13]$$

Here, $i = 1 \ldots 6$. Thus, it forms a statistic based on all the sample canonical correlations. Under H_0 the statistic

$$V = (N - 0.5(m + n + 1))\ln \Lambda \qquad [14]$$

has an approximate $\chi^2(mn)$ distribution for large N. Again, N is the length of the timeseries and m and n are the dimensions on each side of the CCA. Hence, in the present context

$$V = (N - 8)\ln \Lambda \in \chi^2(54). \qquad [15]$$

In order to limit the Type I errors, i.e., the probability for falsely declaring a pixel as being activated, we determine a threshold from the $\chi^2(54)$ distribution which fulfills the desired probability level for false activations. As V in Eq. [15] is *asymptotically* $\chi^2(54)$ distributed it should also be investigated if $N = 200$ is large enough to make such an approximation.

Figure 8 shows the $\chi^2(54)$ distribution together with the estimated distribution of V obtained from a nulldata-set. The nulldata consist of an fMRI experiment acquired under the same experimental conditions as in the Results section, but without any task performed. Clearly, no activation should result. A closer examination reveals that the estimated distribution is somewhat wider, resulting in an overestimation of the significance when choosing a threshold based on the theoretical $\chi^2(54)$ distribution. This is mainly due to the spatial and temporal correlation in the fMRI data, and is not the result of a limited number of samples. The same behavior is observed in both the t-test and the F-test. For this reason, one should be careful when comparing different methods based on thresholds obtained from theoretical distributions. Different tests may behave differently when assumptions, such as independent observations, are violated.

In the Results section we used a threshold for the largest canonical correlation, mainly because it is a more *intuitive* statistic. However, we can obtain an approximate lower bound on the significance when $\rho_t = 0.65$ by setting the remaining canonical correlations to zero and calculating V

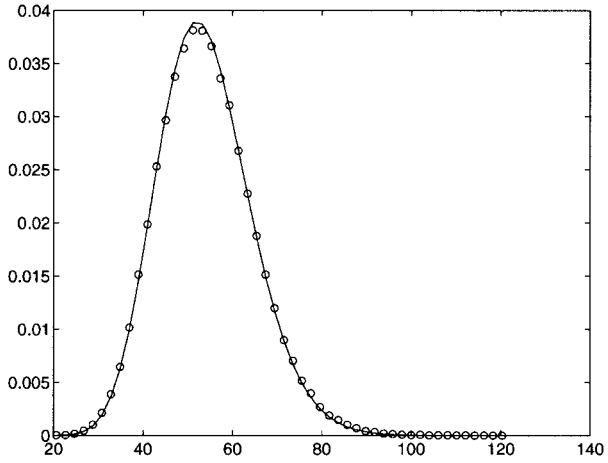

FIG. 8. The solid line shows the theoretical $\chi^2(54)$ distribution and the circles indicate the estimated distribution of V in Eq. [15], obtained from a nulldata-set.

in Eq. [15]. A stated earlier, this gives a lower bound on the significance at approximately 10^{-4}.

Postprocessing

The proposed postprocessing step can be viewed as a separate method to reject spurious activated pixels. It is not specifically required for the CCA method. Any method which uses a subspace spanned by sines and cosines can apply this postprocessing step. It is based on the fact that, regardless of the basis-functions chosen, $\mathcal{Y}$ will in general be too large and contain signals that can not be considered as responses to the paradigm. In our method, we reject a pixel when the corresponding $Y(t)$ falls outside a valid region of $\mathcal{Y}$ in a postprocessing step. It is possible to constrain the search in $\mathcal{Y}$ and thereby incorporate a postprocessing step into the analysis. A CCA method with linear constraints on the coefficients in $\mathbf{w}_x$ and $\mathbf{w}_y$ is reported in Ref. 17. However, in our method the valid region of $\mathcal{Y}$ is instead determined via a nonlinear transformation of the coefficients in $\mathbf{w}_y$. Thus, the method in Ref. 17 could not be applied here to incorporate our postprocessing step into the CCA method.

It is somewhat difficult to evaluate theoretically how the postprocessing affects the statistical significance. The distributions of the delay and angle measures depend on the nontrivial distributions of the components in $\mathbf{w}_y$. Also note that the delay and angle are not independent since they are both functions of $\mathbf{w}_y$. Nevertheless, we find the proposed postprocessing tool valuable for refining the search for activated regions.

CONCLUSIONS

A novel method for detecting neural activity in fMR images was introduced. It combines a subspace modeling of the hemodynamic response with the use of spatial relationships that are present in fMRI data. Unlike univariate methods, it is for this reason capable of detecting homogeneous spatial regions of activity in the human brain, in addition to providing increased sensitivity. Furthermore, a postprocessing method which allows rejection of spurious activated pixels based on shape and delay of the hemodynamic response is proposed. The latter in particular is a very valuable feature, which will be explored in more detail. A method for obtaining estimates of statistical significance has also been discussed.

ACKNOWLEDGMENTS

The authors wish to thank Prof. Birgitta Söderfeldt for valuable discussions.

REFERENCES

1. Ardekani BA, Kanno I. Statistical methods for detecting activated regions in functional MRI of the brain. Magn Reson Imaging 1998;16: 1217–1225.
2. Macovski A. Noise in MRI. Magn Reson Med 1996;36:494–497.
3. Aguirre GK, Zarahn E, D'Esposito M. The variability of human, BOLD hemodynamic responses. Neuroimage 1998;8:360–369.
4. Zarahn E, Aguirre GK, D'Esposito M. Empirical analyses of BOLD fMRI statistics, spatially unsmoothed data collected under null-hypothesis conditions. Neuroimage 1997;5:179–197.
5. Friston KJ, Jezzard P, Turner R. Analysis of functional MRI time-series. Hum Brain Map 1994;1:153–171.
6. Worsley KJ. An overview and some new developments in the statistical analysis of PET and fMRI data. Hum Brain Map 1997;5:254–258.
7. Worsley KJ, Poline J-B, Friston KJ, Evans AC. Characterizing the response of PET and fMRI data using multivariate linear models. Neuroimage 1997;6:305–319.
8. Hotelling H. Relations between two sets of variates. Biometrika 1936; 28:321–377.
9. Anderson TW. An introduction to multivariate statistical analysis. New York: John Wiley & Sons; 1984.
10. Borga M. Learning multidimensional signal processing. PhD thesis, Linköping University, 1998.
11. Ardekani BA, Kershaw J, Kashikura K, Kanno I. Activation detection in functional MRI using subspace modeling and maximum likelihood estimation. IEEE Trans Med Imaging 1999;18:101–114.
12. Cox RW, Hyde JS. Software tools for analysis and visualization of FMRI data. NMR Biomed 1997;10:171–178.
13. Andrade A, Paradis A-L, Rouquette S, Poline J-B. Ambiguous results in functional neuroimaging data analysis due to covariate correlation. Neuroimage 1999;10:483–486.
14. Vazquez AL, Noll DC. Nonlinear aspects of the BOLD response in functional MRI. Neuroimage 1998;7:108–118.
15. Friston KJ, Frith CD, Turner R, Frackowiak RSJ. Characterizing evoked hemodynamics with fMRI. Neuroimage 1995;2:157–165.
16. Constantine AG. Some non-central distribution problems in multivariate analysis. Ann Math Stat 1963;34:1270–1285.
17. Yanai H, Takane Y. Canonical correlation analysis with linear constraints. Linear Algebra and Its Applications 1992;176:75–89.