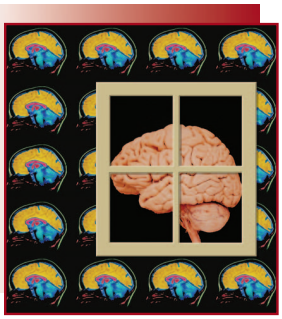

Increasing the Effect Size in Event-Related fMRI Studies

Getting More in Less Time with ICA Denoising

BY MARTIN J. McKEOWN, Z. JANE WANG, RAFEEF ABUGHARBIEH, AND TODD C. HANDY

Since it was first introduced to functional magnetic resonance imaging (fMRI) in 1998 [1]–[4], independent component analysis (ICA) has proved to be a powerful method for exploratory analysis of fMRI data. It has been used to uncover unexpected activations in fMRI data derived from brain activation in response to very complex stimuli, including driving [5] or watching a movie [6], and to infer connectivity between brain regions [6]. ICA has been used to characterize other sources of variability in the fMRI signal besides task-related activity [2], [3] as well as challenging some of the assumptions inherent in other fMRI analysis methods [7]–[9].

As a data-driven fMRI analysis technique, the philosophy of ICA is often in disagreement with hypothesis-driven methods, such as multivariate linear regression [10]. In data-driven methods, as few assumptions as possible are made to reveal what is in the data. In hypothesis-driven methods, fairly rigid assumptions—e.g., all the assumptions associated with determining the least squares (LS) estimates of the coefficients in multiple linear regression [11]—are specified, and specific hypotheses are tested. As most scientific inquiry is hypothesis-driven, this latter approach is clearly desirable. However, as we still do not currently understand the underlying nature of all the sources of fluctuation within fMRI data [12], there is the risk of substantially violating one or more of the underlying assumptions, possibly resulting in misinterpretations [13]. Until fMRI data are sufficiently well understood so that the major sources of variability are well characterized and models are developed that incorporate these sources of variability, data-driven methods including ICA will continue to play an important role in fMRI analysis.

ICA in fMRI

Let the fMRI signal be represented by the space-time data matrix of measurements $X_{j,t}$, where $j = 1, \ldots, V$; $t = 1, \ldots, T$; V is the number of voxels; and T is the number of time samples, respectively. In the linear mixing case, we assume that the matrix can be modeled as follows:

$$X_{jt} = \sum_{k=1}^{N} A_{jk} S_{kt} + \eta_{jt}, \qquad (1)$$

where A and S are formed by the N independent components (ICs) of the process, and η_{jt} is spatially and temporally white noise.

In spatial ICA, we assume that the columns of the matrix $\boldsymbol{A} = [A_{jk}]$ are statistically independent processes while in temporal ICA, the rows of $\boldsymbol{S} = [S_{kt}]$ are assumed independent. While both types of ICA have been applied to fMRI, spatial ICA tends to be more popular [1], [14]–[16]. The most popular algorithms for calculating ICs include the Bell-Sejnowski Infomax algorithm [17] and Hyvarinen's FastICA algorithm [18]. Attias's algorithm combining higher-order statistics and decorrelation can extract temporally and spatially independent patterns [19] but at considerably higher computational cost.

Task-Related Components

The results of applying ICA are often a multitude of ICA components that may be difficult to interpret. Various methods have been proposed to select which components are interesting, such as the task-related ones (i.e., ones whose time courses are consistent with the underlying behavioral experiment). The selection of task-related components includes simple visual inspection of the static spatial map and the associated time course to determine if they are consistent with prior knowledge about brain activation [2], [3] as well as selection of components based on cross-correlations with the underlying behavioral experiment [14] or spectral content [20].

A potential problem of using ICA to infer the spatial extent of task-related activity is the risk of overfitting. As was noticed in the early papers on the subject, there is a risk that ICA may fragment a large area of activation into several sparse spatial maps, each with similar corresponding time courses [2], [3]. This may in part be related to the fact that the underlying sources of ICA, in this case spatial map, assumes a non-Gaussian distribution [17]. Consider the following Gedanken experiment: an fMRI study is conducted resulting in robust activation in one part of the brain, during which ten slices spanning the entire region of task-related brain activity are collected. The experiment is then repeated, resulting in identical brain activation; only this time, 30 slices are collected, again encompassing the entire activated region. In the former case, because of the ICA algorithm's tendency to search for sparse spatial maps, there may be

0739-5175/06/$20.00©2006IEEE

As a data-driven fMRI analysis technique, the philosophy of ICA is often in disagreement with hypothesis-driven methods.

more than one task-related component separated, each with highly similar time courses. This feature of ICA, namely that the robustness of separated activity is a function of the fraction of voxels that are task related, has probably been inadequately explored to date.

The problem of overfitting in ICA has been approached in a number of ways. The HYBICA scheme proposed in [21] suggests a general linear model approach for postprocessing of ICA results. This article also addresses the critical issue of how many components to keep in the analysis, employing predictability of the statistical model as a suitable criterion. A similar approach can be used in principal component analysis (PCA) to determine the underlying dimensionality of the data [22]. Beckmann et al. suggested a Bayesian approach to estimate the true underlying dimensionality of the data, assuming a Gaussian noise model [23]. More generally, numerous methods have been proposed for statistical model selection, such as Akaike's information criterion (AIC) and Bayesian information criterion (BIC) [24]. However, empirically, we have found that these tend to discard components that appear to have clear neuroscience meaning. Perhaps this is based on inherent underlying assumptions of Gaussian noise models, which may be less appropriate for fMRI data.

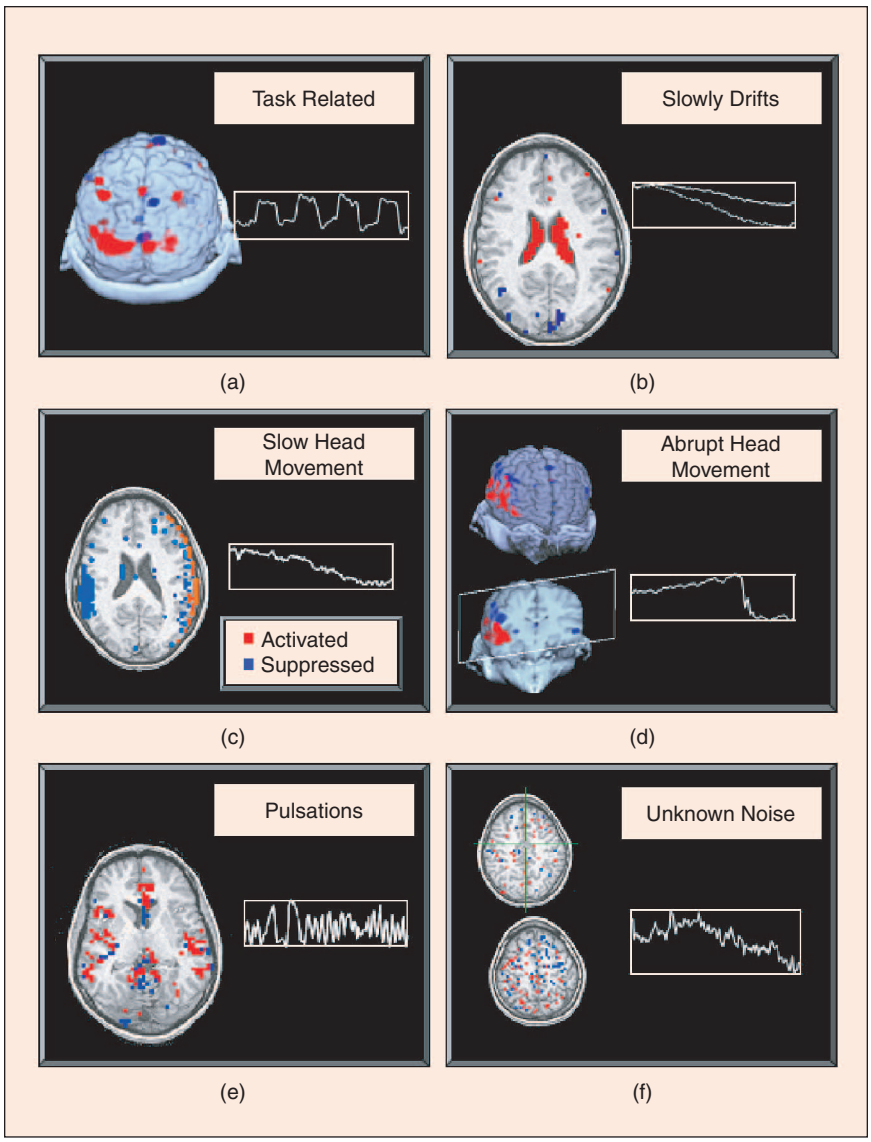

Fig. 1. ICA components from a typical fMRI experiment may reflect (a) task-related activity, (b) slow drifts in the fMRI signal, (c) and (d) slow and abrupt head movement, (e) pulsations of cardiac and respiratory origin as well as (f) other sources of noise of indeterminate origin. Figure adapted from (3).

Noise Components

Perhaps one of the most interesting byproducts of the burgeoning use of ICA in the analysis of fMRI data has been the interesting insights into the underlying nature of the fMRI signal [12]. Various ICA components appear related to head movement, local brain distortions related to cardiac and respiratory pulsations, slow drifts in the signal (which are also present during cadaver recordings [25]), and brain-related activity [2], [3] (Figure 1). Since the vast majority of statistical signal-processing models assume that there is some underlying signal that is corrupted by (possibly white) random Gaussian noise [12], determining which features of fMRI data are stochastic and which are deterministic is important for the development of suitable statistical models [9]. The fact that ICA extracts numerous, robust, and reproducible components suggests that many of the nonbrain-related signals in fMRI data are not random; rather, these nonbrain signals have significant spatial and temporal stucture [9].

The ability of ICA to remove structured signals suggests that this method would be useful for noise reduction in fMRI [26]. Thomas et al. have compared ICA and PCA in separating the blood oxygen level dependent (BOLD) signal change from structured and random noise and, as expected, found out that PCA was superior to ICA when random noise corrupted the data, but ICA was superior when structured noise was present [26].

One important source of variability in fMRI data with nonrandom spatial structure is related to motion artifact. The extreme sensitivity of ICA in detecting a broad range of spatially nonrandom signals makes ICA exquisitely sensitive to motion. Even ICA analysis on standard motion-corrected fMRI data results in ICA components that are still consistent with motion [12], [27]. Recently, we have exploited this sensitivity to motion to propose a motion correction scheme based on ICA, called motion-corrected ICA (MCICA) [27]. It is based on the observation that movement during acquisition of an fMRI time series corrupts the statistics of the acquired data, resulting in a simultaneous increase in the joint entropy of the data and a decrease in the entropy of a nonlinear function of the ICs. Motion effects can therefore be mitigated by spatially adjusting the data to maximize the entropy difference between the data and the nonlinear function of the calculated ICA components without explicitly estimating motion parameters (Figure 2).

ICA Analysis of Different fMRI Designs

fMRI experiments are typically performed in block designs [28] or event-related designs [29]–[33]. In block fMRI designs, the subject is instructed to perform experimental (E) and control (C) tasks in an alternating sequence of 15–40 s blocks (e.g., CECECEC…). In event-related designs, the subject is instructed to respond singly to intermittently presented stimuli. When event-related studies are analyzed, the fMRI data are divided into epochs time-locked to stimulation presentation, which are then typically averaged to obtain a mean response to the stimuli. The benefit of this approach is that it allows the robust isolation of possibly highly spatially overlapping brain areas demonstrating significantly different magnitudes of fMRI BOLD response [34]. This method of analysis implicitly assumes that the data can be accurately modeled as a deterministic signal that is precisely time-locked to stimulus presentation and is corrupted with random noise that will tend to zero when averaged over many trials; i.e., all signals not precisely time-locked to stimulus presentation, including other brain signals and decaying responses from previous stimuli, are assumed to be noise. However, if underlying components of the data are not completely random with respect to stimulus presentation, they may tend to average to values other than zero, thus introducing biases in the estimates of stimulus-locked signals. For example, since reaction times to stimuli are distributed over a temporal range, averaging the data time-locked to response time will tend to accentuate components of the data related to response, but this will result in temporal misalignment of features time-locked to the immediate stimulus recognition and vice versa.

ICA has proved useful in exploring the validity of some of the assumptions underlying event-related fMRI analysis. Analysis of fMRI data with ICA has suggested that there can be considerable trial-to-trial variability in the fMRI BOLD response [7]. Moreover, other ICs besides the dominant task-related component may still be significantly task related but with different latencies and variability in latencies [8], [9] (Figure 3).

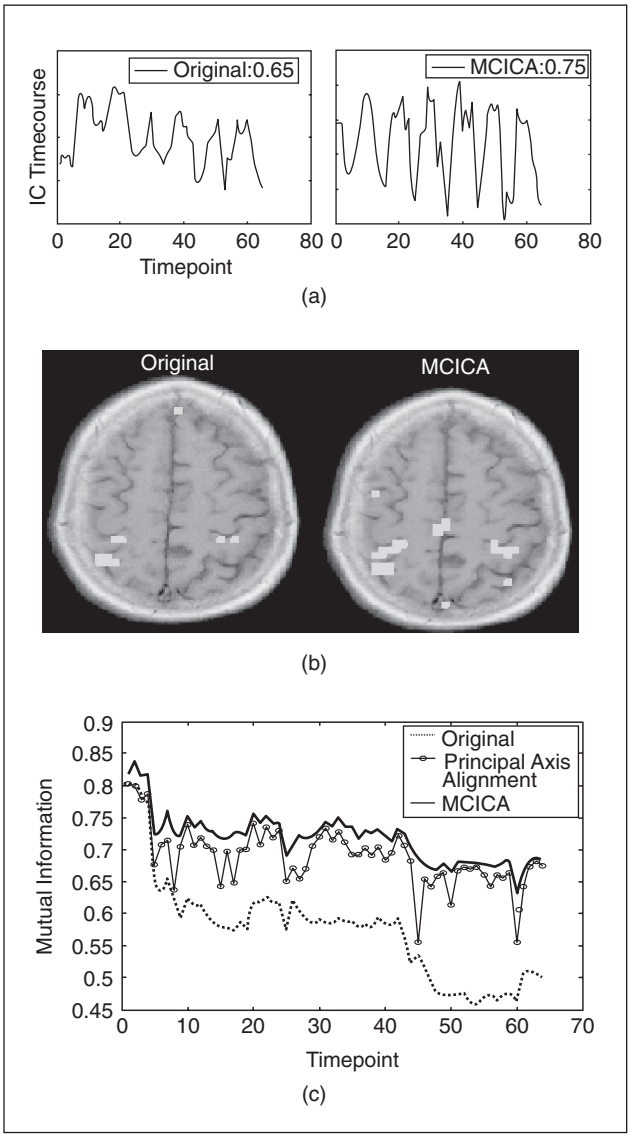

Fig. 2. The MCICA approach to motion correction. The sensitivity of ICA to motion artifact can be utilized for motion correction. (a) A motor experiment was performed with a subject moving both hands simultaneously and resting in alternate blocks. An IC was separated that significantly correlated ($r = 0.65$) with the behavioral task. After using MCICA to correct for motion and reanalyzing with ICA, the task-related component became more correlated the task ($r = 75$). (b) The associated spatial map (thresholded for $z > 2$) after MCICA revealed robust activation in motor and supplementary motor cortices. (c) Although mutual information is not explicitly optimized, the mutual information between the first volumes and subsequent volumes is increased after motion correction with MCICA. Using the centroid and principal axes of activation for alignment results in increases in mutual information between that of the original data and MCICA. Figure adapted from (27).

As the evoked hemodynamic response (HDR) is typically small in fMRI data, numerous stimuli must be presented to achieve a given level of statistical significance. Since imaging is typically expensive and subjects can only tolerate being in the scanner for reasonable lengths of time, stimuli are presented as fast as practically possible. However, the sluggishness of the HDR requires that either the stimuli are presented infrequently enough to prevent excessive overlapping of responses, thus limiting the total number of trials, or attempts must be made to analytically deconvolve the possibly nonlinear overlapping responses in subsequent analysis.

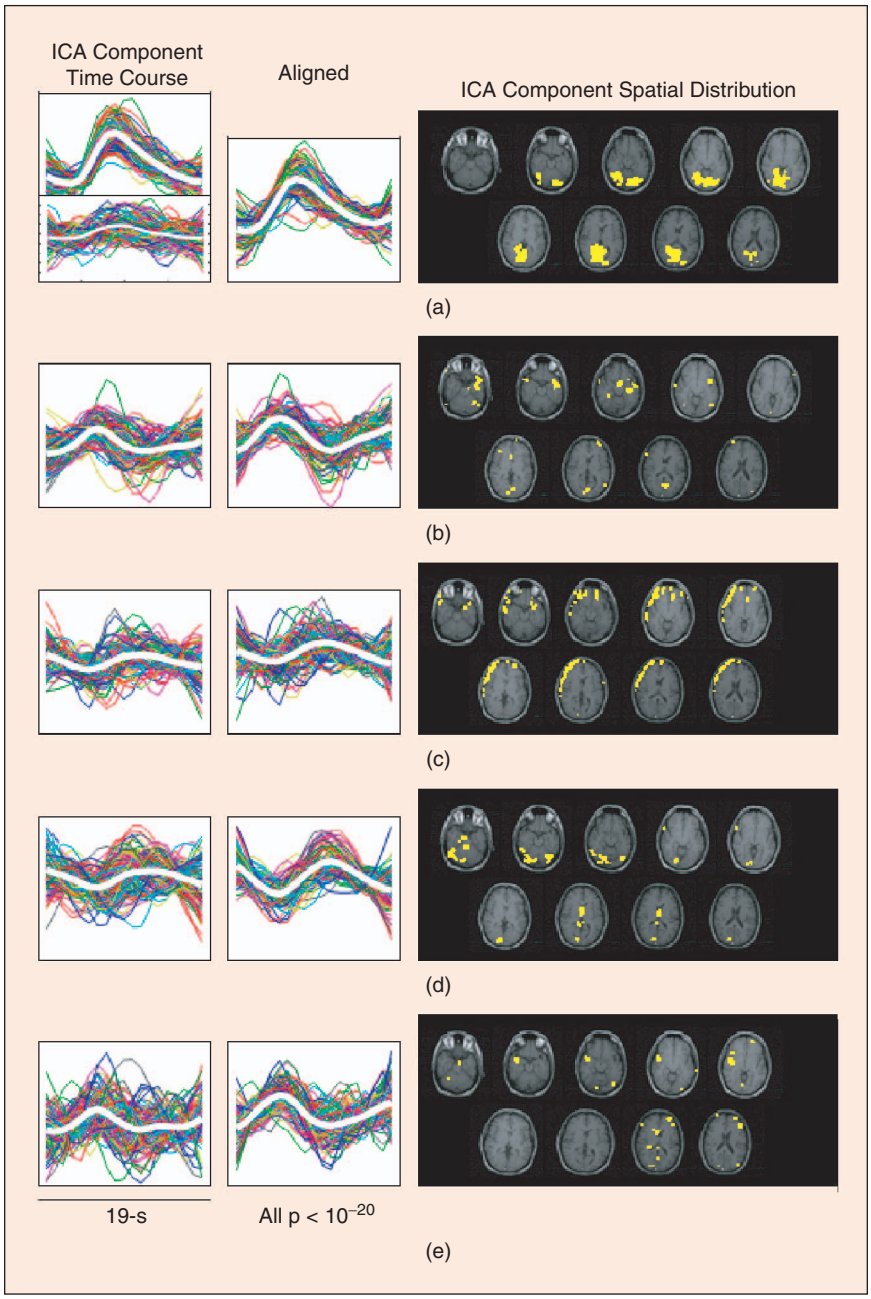

Fig. 3. A subject was asked to view 95 visual stimuli consisting of black and white radial checkerboards that subtended about 20° by 15° of visual angle, presented singly for 500 ms. The interval between successive stimuli varied randomly between 14–18 s. Five ICs were most significantly (all $p < 10^{-20}$) affected by task. The time courses of these components (the first column) were divided into 19-time-point epochs that were time-locked to stimulus presentation, linearly detrended, and overlaid. The heavy white lines correspond to the mean response.The raw time course of a typical activated voxel ($p < 10^{-13}$, Bonferroni corrected) ((a), inset) demonstrated much more trial-to-trial variability than the respective ICA component time courses. Component time courses (the middle column) were aligned by cross-correlating the IC time course from an individual stimulus presentation with the mean from all stimulus presentations. Note the minimal effect the alignment procedure had on component (a) compared to the other components (b)–(e). (Figure adapted from (47)).

Accurate estimation of the form of possibly overlapping, often miniscule stimulus-evoked HDR in the face of underlying noise is challenging. A linear model has been proposed to isolate the stimulus-induced fMRI BOLD response [31] despite some evidence for nonlinear interactions in fMRI responses (e.g., [35]). In the linear approach, a stimulus convolution matrix based on the timing of stimuli presentation is derived, and the HDR is estimated by multivariate regression using ordinary LS [36]. By estimating the noise covariance, the maximum likelihood (ML) estimate of the HDR was studied in [36]. Furthermore, by exploiting the prior knowledge of the anticipated HDR shape, hemodynamic basis functions can be incorporated into the estimation process to improve the estimator efficiency [37]. Here, efficiency is defined as the reciprocal of the HDR estimator variance.

The choice of a suitable noise model may enhance the sensitivity and the accuracy of estimating the evoked HDR. Much prior work has assumed a white Gaussian noise model (e.g., [22] and [23]). However, as the MRI image is transformed to a magnitude-phase presentation, it has been suggested that noisy MRI data may follow non-Gaussian distributions, such as a Rician distribution [38]. Assuming a Rician noise model, a denoising scheme was developed in [39]. Wavelets [40], spectrum subtraction [39], and component analysis [26] have all been suggested as suitable methods for noise reduction in fMRI data. However, the extent to which evoked brain estimates are biased by the choice of inappropriate noise models has yet to be fully explored.

In this article, we present work that elucidates the aforementioned concepts. We perform ICA on fMRI data from a simple event-related motor task. We show that ICA may overfit

In data-driven methods, as few assumptions as possible are made to reveal what is in the data.

the data as several components appear task related. However, we demonstrate that projecting the original data onto the linear subspace defined by these task-related components essentially denoises the signal and results in significant improvement in the effect size. We then perform a number of simulations using a number of different noise models. We demonstrate that the ICA denoising procedure compares favorably to previously described regression methods for determining overlapping responses and is robust to a variety of realistic noise models.

Methods

fMRI Data

The data were collected from a single 24-year-old normal male subject on a Philips 3T magnet, with a time to repetition interval of 1,000 ms, 4 mm slice thickness with a 1-mm slice gap, matrix size 128 × 128, and a field of view (FOV) of 240 mm (see Figure 4). Nineteen axial slices were acquired. Each slice was acquired at 80 × 78 resolution that was then reconstructed to 128 × 128 using standard Phillips software.

Slices were collected in interleaved fashion (Figure 4). A total of 360 volumes were acquired for each run, and a total of six runs were performed and the data concatenated together. The interstimulus interval varied from 13–18 s, with a total of 126 stimuli presented. The subject was asked to press a button with his right hand in response to visual stimulus. In between stimuli, he was asked to fixate on the screen.

The fMRI data were first preprocessed using the BrainVoyager software. Each run was corrected for slice timing delays, linearly detrended, and then a three-dimensional rigid-body registration was used for motion correction.

To reduce computation, voxels assumed to be involved in the task were first estimated by averaging the data in all voxels time-locked to the presentation of the stimuli using a 16-s window. Those voxels whose mean activation correlated above an arbitrary threshold ($r = 0.7$) with a previously published generic HDR [41] were retained for further analysis. This resulted in a matrix size of 2,160 (360/run × 6 runs) × 7,846.

Linear Model for Overlapping HDRs

We start with a linear model for the observed fMRI event-related response [36]. For simplicity, we consider the case of single type of event rather than multiple interleaved stimuli, each evoking a different response. For each voxel i, $i = 1, \ldots, N$, the fMRI observation $y_i(t)$ is modeled as

$$y_i(t) = x(t) * h_i(t) + n_i(t), \tag{2}$$

where the event sequence $x(t)$ is a sum of time-shifted delta functions determined by the stimulus intervals, $h_i(t)$ is the evoked HDR of each voxel i, which is to be estimated, $n_i(t)$ represents additive noise, and * represents the linear convolution operator. In a discrete version, we define the corresponding vectors $\mathbf{y}_i$, $\mathbf{h}_i$, and $\mathbf{n}_i$, thus the model becomes

$$\mathbf{y}_i = \mathbf{X}\mathbf{h}_i + \mathbf{n}_i, \tag{3}$$

where $\mathbf{X}$ is the so-called stimulus convolution matrix determined by the event sequence $x(t)$. Our purpose is to estimate the HDRs $\{h_i\}$ based on the fMRI observations $\{y_i\}$.

Proposed Scheme to Estimate the Evoked HDRs

The basic idea of the proposed scheme is to first denoise the observations by utilizing the results from an initial ICA analysis. The data are first analyzed by ICA, and the event-related components are identified. The original data is then projected into the linear subspace spanned by the components of interest. The proposed scheme is summarized as follows:

Step 1: Apply PCA to achieve dimension reduction [42], [43]. PCA seeks to achieve dimension reduction by projecting the original data into a few orthogonal linear combinations [the principal components (PCs)], thus it provides the best linear dimension reduction performance in the sense of mean square error. Define $\boldsymbol{Y}$ to be an $n \times N$ data matrix of observed voxel time courses, where $n = 2{,}160$ is the number

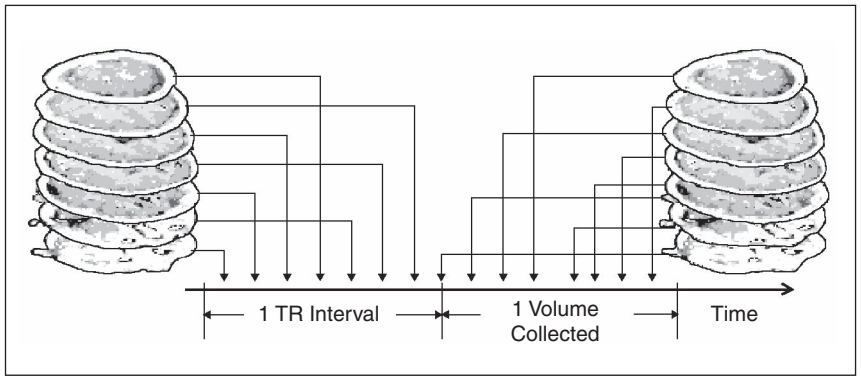

Fig. 4. FMRI data are acquired in slices, typically in interleaved fashion. The TR interval refers to the period of time through which an entire volume is collected.

A potential problem of using ICA to infer the spatial extent of task-related activity is the risk of overfitting.

of temporal observations in each voxel, and N is the number of total voxels considered in the experiment. The PCA procedure reduces the number of voxels from N to p (e.g., 50) [9]. This procedure can be written as

$$\mathbf{Y_1} = \mathbf{Y}\Lambda,$$

where $\mathbf{Y}_1$ is $n \times p$, and Λ is the $N \times p$ matrix of eigenvectors used for the PCA reduction.

Step 2: Apply ICA and choose the most related components. The ICA can be written as

$$\mathbf{S} = \mathbf{Y_1}\mathbf{B} = \mathbf{Y}\Lambda \cdot \mathbf{B},$$

where $\mathbf{B}$ is the $p \times p$ unmixing matrix. In the current situation, the first 50 PCs were subsequently used as input to ICA, as suggested elsewhere [9]. Infomax ICA [17] was then used to separate the eigenimages into spatial ICs [2]. The associated time courses were then estimated.

Step 3: Select components deemed task related by model-fitting. Denote the jth component by the vector $\boldsymbol{f}_j$. If $\boldsymbol{f}_j$ is event related, according to the signal model (3), we expect that $\boldsymbol{f}_j$ follows a similar formulation in the form $\mathbf{f}_j = \mathbf{X}e_j + \mathbf{n}_j$, where e_j is the original event-related signal. Therefore, we propose to select components deemed task-related by fitting each component vector to the above model to estimate $\boldsymbol{e}_j$ in the LS sense ($\hat{\mathbf{e}}_j$), and using the corresponding relative fitting error, defined as $d_j = (||\mathbf{f}_j - \mathbf{X}\hat{\mathbf{e}}_j||^2)/(||\mathbf{f}_j||^2)$, as a criterion to determine whether a component $\boldsymbol{f}_j$ is task related or not. After this model fitting procedure, the reduction in the number of components from p to M can be written in matrix form. Without loss of generality, let the column components in S be sequentially ordered from most to least task related by the d_j criterion, then the selected components can be written as

$$\mathbf{S_1} = \mathbf{S}\Omega, \text{ with } \Omega = (\mathbf{I}_M, \mathbf{0}) \text{ being size } p \times M,$$

where the M most task-related components are retained via the identity matrix I_M.

Step 4: Denoise observations. The idea is to denoise the observation by projecting the original data onto the selected ICs. Without loss of generality, suppose the components $\boldsymbol{f}_j$, for $j = 1, \ldots M$, are determined as task related in Step 3. One simple way to denoise the original voxel vector $\boldsymbol{y}_i$ is to project it onto the above signal subspace S_1 spanned by the components $\boldsymbol{f}_j$, $j = 1, \ldots M$. Therefore, the denoised observations can be expressed as

$$\hat{\mathbf{Y}} = [\hat{\mathbf{y}}_\mathbf{1}, \hat{\mathbf{y}}_\mathbf{2}, \ldots, \hat{\mathbf{y}}_\mathbf{N}] = \mathbf{S_1}(\mathbf{S_1^T}\mathbf{S_1})^{-1}\mathbf{S_1^T}\mathbf{Y}. \quad (4)$$

One can see that if $\Lambda = \mathbf{B} = \Omega = \mathbf{I}_N$, then $\hat{\mathbf{Y}} = \mathbf{Y}$.

Step 5: Estimate the impulse HDR. For each voxel i, estimate the impulse (HDR) $\boldsymbol{h}_i$, based on the denoised vector $\hat{\mathbf{y}}_i$. The estimated HDRs can be written in matrix form $\hat{\mathbf{H}} = [\hat{\mathbf{h}}_1, \hat{\mathbf{h}}_\mathbf{2}, \ldots, \hat{\mathbf{h}}_N]$ with size $q \times N$.

Then, we apply the LS approach to estimate the HDRs, expressed as

$$\hat{\mathbf{H}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\hat{\mathbf{Y}}. \quad (5)$$

Suppose that the noise after the denoising process is temporally uncorrelated, then the ML estimator for our problem reduces to the above ordinary LS estimate.

It is worth mentioning that in Step 4, we denoised the signal by projecting the raw data into a signal subspace derived from ICA (4). This is different from previous methods of removing artifacts within the ICA components by simply setting those components to zero [26]. If we consider the data Y, we are modeling the ICA data as:

$$Y = [S \quad N]\begin{bmatrix} C_s \\ C_n \end{bmatrix},$$

where $\mathbf{Y}$ is the observation matrix $\mathbf{S}$ is the signal subspace, and $\mathbf{N}$ is the noise subspace. We propose $\hat{\mathbf{Y}} = \mathbf{S}(\mathbf{S}^T\mathbf{S})^{-1}\mathbf{S}^T\mathbf{Y}$, as opposed to $\hat{\mathbf{Y}} = [\mathbf{S}|\mathbf{N}]\,(\mathbf{C_s}/\mathbf{0})$ as suggested elsewhere [26]. For comparison, we compared the HDR estimate calculated from the original data:

$$\tilde{\mathbf{h}}_i = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T y_i. \quad (6)$$

To determine if any of the time points contained in $\boldsymbol{h}_i$ has statistical significance, a null distribution for each voxel (both before and after denoising) was determined by computing averages from randomly placed 16-s windows containing the same number of stimuli as in the original case. One thousand draws were used to calculate the null distribution.

Simulations Incorporating Several Noise Models

To test the robustness of the proposed denoising procedure to noise, we consider the addition of several types of noise to idealized underlying signal. The noise parameters are chosen to yield a signal-to-noise ratio (SNR) of around –15 dB in order to emulate the low SNR observed in real fMRI data, where SNR is defined as

$$10\log_{10}\left(\frac{\text{signal power}}{\text{noise power}}\right).$$

The simulation of ideal error-free, task-related fMRI observations is performed as follows. First, the unbiased ML HDR estimator described in [36] was applied to a real event-related fMRI data. Then, using the convolution matrix $\mathbf{X}$ and

The ability of ICA to remove structured signals suggests that this method would be useful for noise reduction in fMRI.

such estimated HDRs, the voxel event-related time courses were reconstructed to create a true underlying signal that was then corrupted with different types of noise. As we based our simulations on our collected data, the data size settings are as follows: the number of total voxels of interest is 7,846, the number of stimulus is 126 with durations randomly distributed between 13–17, the temporal length of the observation is 2,160, and the duration of HDR is assumed to be captured in a 16-s window. Based on the additive noise model in (3), the choices of noise models are as follows:

• *Case 1: Spatially and temporally white Gaussian noise*. The noise covariance matrix is assumed as $\mathbf{R}_n = \sigma^2\mathbf{I}$. The SNR was chosen to be around −15 dB.

• *Case 2: Temporally correlated Gaussian noise with 0-mean and covariance matrix* $\mathbf{R}_n$. The noise covariance $\mathbf{R}_n$ was estimated from the real fMRI data, and the SNR was around −13 dB.

• *Case 3: Independent identically distributed Rayleigh noise*. The noise is assumed to follow a Rayleigh distribution (i.e., special case of Rician, with the parameter A = 0). The SNR was around −12 dB.

• *Case 4: Rician noise with time-varying parameter* A. The noise is assumed to follow a distribution Rician (A, σ^2), defined as

$$f(t) = \frac{t}{\sigma^2}\exp(-(t^2 + A^2)/\sigma^2) I_0\left(\frac{At}{\sigma^2}\right),$$

where A and σ^2 are the distribution parameters and $I_0(.)$ is the modified zero-order Bessel function of the first kind. The noise at time j follows a Rician distribution Rician (A_j, σ^2), with the time-varying parameter A_j/σ randomly chosen from 0.3–0.9. For the present study, the overall SNR was around −12 dB.

It has been known that fMRI measurements appear in both the real and imaginary channels due to phase imperfection, and nearly all fMRI studies have been based on the magnitude-only image data. Assuming zero-mean Gaussian noise in both the real and imaginary channels, the fMRI magnitude-only data was shown to obey a Rician distribution [38], [44]. Unlike additive Gaussian noise, this so-called Rician noise is signal-dependent, since the magnitude-only measurement for one voxel can be expressed as

$$y_m = \left\{y^2 + \left(\varepsilon_R^2 + \varepsilon_I^2\right) + 2y(\varepsilon_R\cos\theta + \varepsilon_I\sin\theta)\right\}^{1/2}, \quad (7)$$

where ε_R and ε_I are the Gaussian noise in the real and imaginary channels respectively, y is the signal of interest, and θ is the phase imperfection introduced to the voxel. The above magnitude is Rician distributed. Based on the signal-dependent noise model in (7), we also study the following case:

Case 5: Embedded Rician noise. The observation is simulated according to (7), where the variance white Gaussian noises ε_R and ε_I are chosen to yield a SNR around −12 dB. In this case, SNR is defined as

$$10\log_{10}\left(\frac{\text{signal power}}{\text{overall power} - \text{signal power}}\right).$$

One may argue that the additive Rician noise models of Case 3 and Case 4 are against the embedded Rician noise model in Case 5. It is worth mentioning that Case 3 and Case 4 are studied to simulate the scenarios that the error-free fMRI data are a combination of the signal due to resting brain and the purely task-related signal, meaning

$$y = y_{\text{rest}} + y_{\text{task}},$$

where y, y_{rest}, y_{task} represent the error-free fMRI data, the resting brain signal, and the purely task-related signal, respectively. Since the signal of interest is the purely task-related signal, the resting brain signal is also counted as an additional noise source. The joint contribution of the resting brain signal and the Gaussian noise is modeled as an additive Rician noise, as studied in Case 4.

In all cases, the estimation performance of the proposed scheme was compared with that of the direct LS approach (6) when no denoising process was performed.

Estimation Efficiency

Due to the complex nature of ICA and the unknown noise model of the event-related fMRI data, it is impossible to characterize the estimation performance of the proposed scheme analytically. Hence, performance demonstrations are based on simulations. In this study, we aim to estimate HDRs as efficiently as possible. To evaluate the estimation performance with regard to this objective, we may examine various statistical criteria, such as the estimator efficiency $\boldsymbol{E}$ defined in [36] and [45]. Let $\boldsymbol{h}$ and $\hat{\mathbf{h}}$ be the true and estimated HDR with length n, respectively. To evaluate the estimation performance, we calculate the correlation coefficient (CC) between the estimated HDRs and the true ones. We also study the relative estimation residual power (i.e., the estimator variance) defined as $r = E((||\mathbf{h} - \hat{\mathbf{h}}||^2)/(||\mathbf{h}||^2))$, since it is desirable for an estimator to fit the real HDR curve in an LS sense, where $E(.)$ denotes the expectation operation. Note that this estimation criteria r is closely related to the estimation efficiency E in [36], which is defined as the reciprocal of estimator variance. Assuming the HDRs are normalized, r defined here is the reciprocal of $\boldsymbol{E}$ defined in [36]. The reason why we chose to study the criteria r instead of the criteria $\boldsymbol{E}$ is because r is generally restricted to the range

0–1. The larger the CC and the smaller the relative residual power *r,* the better the estimation performance.

Results

Simulation Results

Figure 5 shows the true HDRs and the HDRs estimated by the proposed scheme under the presumed realistic Rician noise model [38] (Case 4) for six typical voxels. The proposed scheme using denoising more closely followed the underlying signal compared to the original estimate, with original (denoted by LS-R) and denoised (denoted by LS-D) representing the LS approaches described in (6) and (5), respectively. Unlike the estimate from the proposed denoised scheme, the original estimate occasionally demonstrated more than one peak in the estimate.

To further evaluate the estimation performance statistically, we studied the performance measures in terms of correlation between the true HDR and the estimated HDR (CC) and the statistical efficiency (r) discussed above. Table 1 shows the statistical results of estimating the HDRs, where the empirical means and standard deviations of these two performance measures are reported. The proposed ICA-based scheme provided consistently superior estimate performance over the original LS approach without ICA denoising. Both methods demonstrated significantly degraded performance with the temporally correlated Gaussian noise model (Case 2) and the embedded Rician noise model (Case 5), although the original LS method was more severely affected.

fMRI Data Results

Figure 6 is a scatterplot comparing the z-score of the fourth time point (i.e., 4 s after stimulus presentation) for each voxel for the LS-D and LS-R cases. The fourth time point was chosen because each time point within the estimated HDR had an associated z-score, and the peak was typically at timepoints 4–6. The other scatterplots for timepoints 5 and 6 (not shown) were qualitatively the same. Almost all (93%) voxels had greater z-scores using the LS-D estimate compared to the LS-R case, as evidenced by the position of the point above the 45° line. The distribution of the z-scores across all active voxels

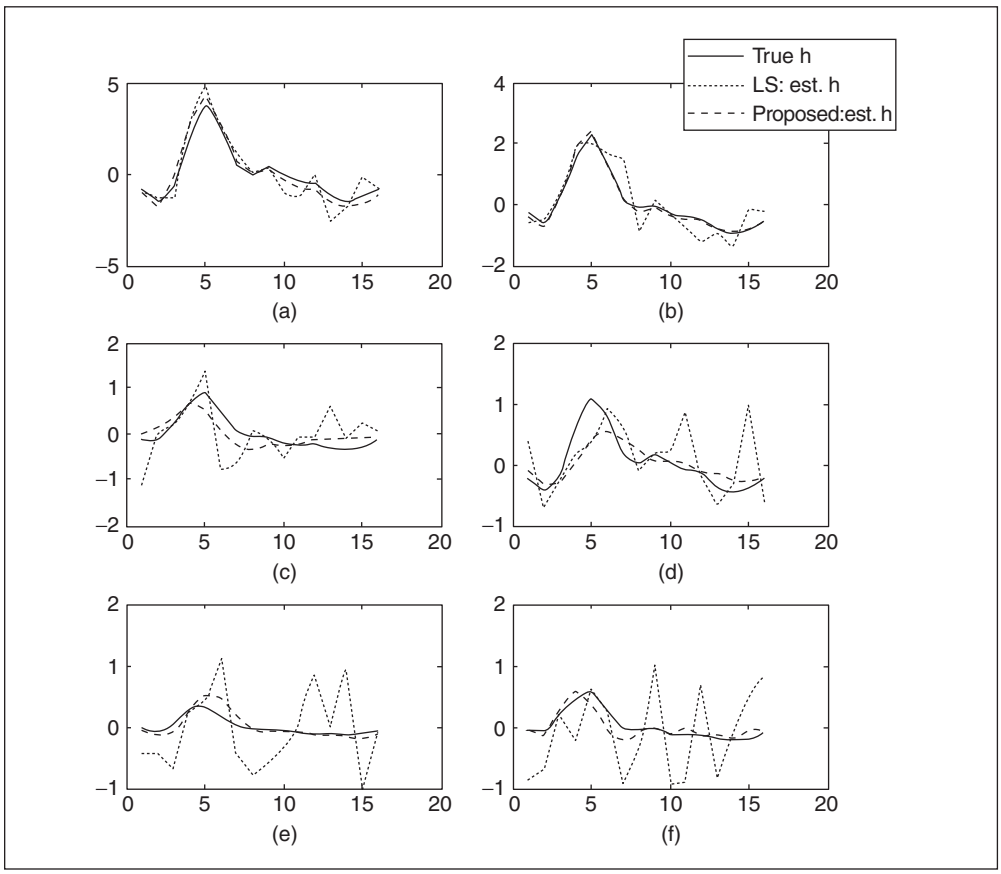

Fig. 5. HDR simulation examples. The true HDR from a random collection of six voxels is shown. The LS estimate is shown in blue, and the proposed method employing ICA denoising is shown in black. Note the spurious multiple peaks present in the LS estimate, presumably because of the mismatch between the implicit Gaussian model with the LS and the other noise model used in the simulation.

Table 1. Estimation efficiency of the proposed scheme in estimating HDRs, in terms of the empirical mean and standard deviation of the correction CC and the efficiency *r*.

	Case 1		Case 2		Case 3		Case 4		Case 5	
	White Gaussian		*Temporally Correlated Gaussian*		*Rayleigh Noise*		*Nonstationary Rician Noise*		*Embedded Rician Noise*	
Noise Model	**Original**	**Denoised**	**Original**	**Denoised**	**Original**	**Denoised**	**Original**	**Denoised**	**Original**	**Denoised**
Correlation	(0.900, 0.059)	(0.960, 0.036)	(0.856, 0.176)	(0.934, 0.164)	(0.952, 0.029)	(0.982, 0.014)	(0.936, 0.037)	(0.979, 0.017)	(0.70, 0.182)	(0.92, 0.060)
Efficiency	(0.269, 0.160)	(0.161, 0.085)	(0.362, 0.356)	(0.154, 0.150)	(0.114, 0.067)	(0.067, 0.036)	(0.153, 0.090)	(0.078, 0.045)	(0.71, 0.088)	(0.35, 0.070)

was $\bar{\mathbf{z}} = 4.42$, $\sigma_{\mathbf{z}} = 2.3$ for LS-R and $\bar{\mathbf{z}} = 8.90$, $\sigma_{\mathbf{z}} = 3.8$ for LS-D. The spatial distribution of activated voxels using the original and denoised data was qualitatively similar [Figure 7(a)]. Both methods reliably demonstrated activation in the left primary motor (right side of image, the solid arrow) and supplementary motor cortices (open arrow). The LS-D estimator detected additional significant activation in the contralateral primary motor cortex. To ensure that any spatial differences between the estimates based on the raw and denoised data were not the result of arbitrarily chosen threshold levels, we plotted activated voxels under a range of different thresholds [Figure 7(b)–(d)]. Again the spatial distribution of activation was qualitatively similar across the two datasets.

Discussion

We proposed a method that uses ICA to define a linear signal subspace and then project the data into that subspace. This resulted in areas of significant activation that were qualitatively similar to the denoised case (Figure 7). However, the effect size, or significance of activation, was much higher in the denoised case (Figure 6). As scanner costs are considerable, this suggests that the same level of statistical significance could be obtained with substantially fewer stimuli presentations, resulting in less scanner time or fewer subjects being scanned, in order to obtain a desired level of statistical significance.

To take into consideration the overfitting issue, we denoised the signal by projecting the raw data into a signal subspace derived from ICA (4). This is in contrast to previous methods of removing artifacts within the ICA components by simply setting those components to zero, as suggested in [26] and [46].

The vast majority of signal processing models assumes a Gaussian noise model. However, this appears to be less suitable for fMRI data, where nontask-related activity is not random unstructured noise but appears to have specific spatial and temporal structure (Figures 1 and 3). We note that our simulation results imply that the choice of a suitable noise model is imperative for accurate assessments of evoked hemodynamic activity if common LS estimates are used (Table 1 and Figure 5). In contrast, the proposed scheme was robust to the exact type of underlying noise model used in the simulations.

Another frequent assumption of many signal processing and statistical approaches in analyzing fMRI data (for example,

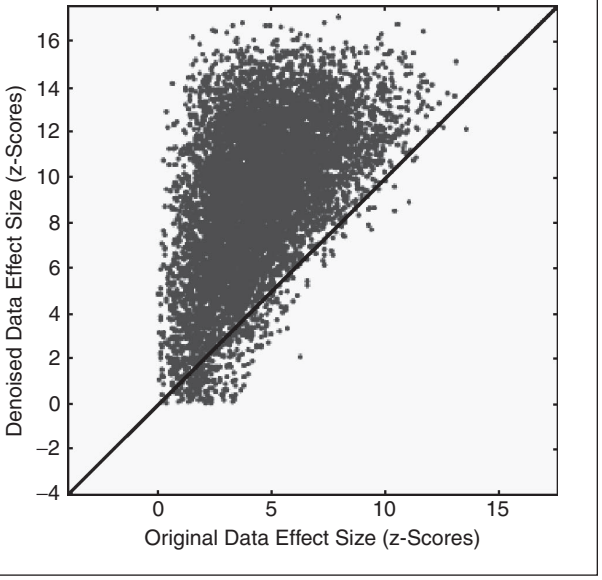

Fig. 6. The comparison of the level of significance (effect size) before and after denoising. The level of significance of evoked response 4 s after the onset of the stimuli is shown. Monte Carlo estimates were used to estimate the null distribution of mean activity. Note that the vast majority of voxels have increased levels of significance after the proposed denoising procedure, evidenced by the fact that they lie above the diagonal line (same horizontal and vertical scales).

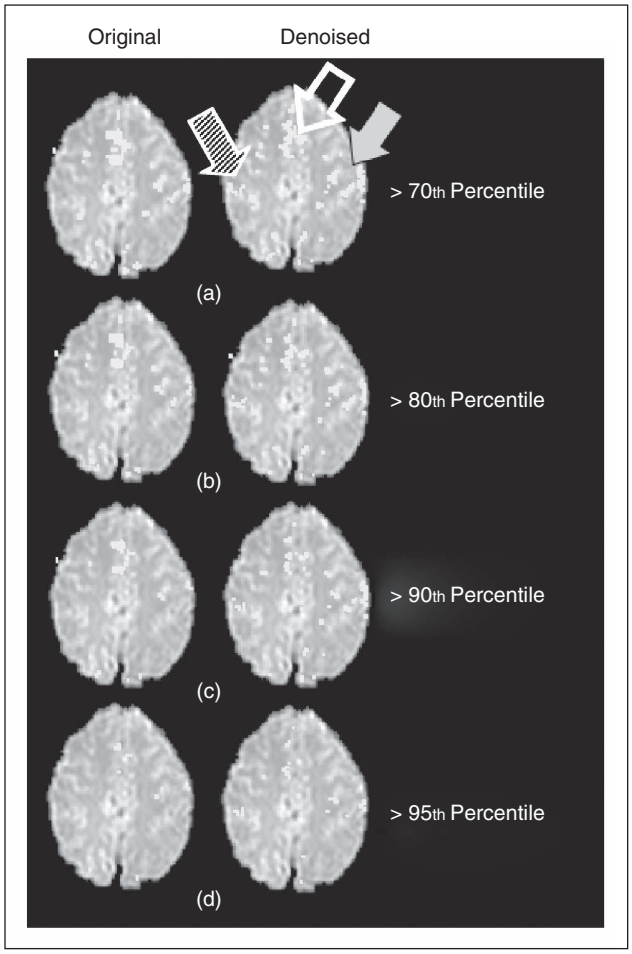

Fig. 7. Spatial distribution of activated voxels before and after denoising. Although the denoised voxels have greater effect size (Figure 5), the spatial distribution of activated voxels is qualitatively similar between the denoised data and the original data. Note the evidence of activation in the primary motor cortex (solid arrow), the supplementary motor cortex (open arrow), although the denoised data also detected activity in the ipsilateral motor cortex (patterned arrow). Voxels with *z*-scores above the 70th, 80th, 90th, and 95th percentiles are shown. The activation maps were not spatially smoothed. The images follow a radiographic presentation such that the left side of the brain is on the right side of the image and vice versa.

multivariate linear regression) is to examine each voxel in isolation instead of an independent realization of some underlying statistical process. The success of ICA in enhancing estimates of evoked activity (Figure 3) affirms the benefits of examining the time course of specific groups of voxels rather than examining them singly.

In the current approach, we used PCA as a dimension reduction technique, as this is probably the simplest and most widely used technique. We note that PCA, which assumes a multivariate Gaussian model, may be less appropriate for dimension reduction in fMRI data. Further work is required to determine if other methods of dimension reduction, including nonlinear approaches, are more suitable.

With the present simulations and fMRI data, we have concentrated on the effects of different types of noise in a setting where there was minimal temporal overlap between stimuli. Although the proposed approach is quite general, a suitable extension would be to determine how much the combined effects of temporal overlap and different noise models degrade the LS estimates of the evoked HDR.

As in previous research, we have assumed that the evoked HDR of each trial was constant and needed to be estimated in low SNR conditions. We are currently investigating a more general model in which HDR of each trial could have different magnitude over time.

Conclusions

By exploiting the fact that much of fMRI data has deterministic spatial-temporal structure, we have proposed a scheme employing ICA denoising and LS estimation of the evoked HDR. Simulations suggest that the method is more robust to different noise models compared to naïve application of LS. The result is a considerably increased level of significance of activation for a given voxel but still qualitatively similar spatial distribution of activations over all voxels. We suggest that the proposed method has the potential to substantially reduce total scanning time requirements to achieve the same level of statistically significant activation.

Martin J. McKeown graduated summa cum laude in engineering physics from McMaster University, Canada, in 1986, and subsequently earned his M.D. degree from the University of Toronto, Canada, in 1990. From 1990–1994 he specialized in neurology at the University of Western Ontario and later did a fellowship in clinical electrophysiology. He is board certified in neurology in Canada and in the United States and has been an examiner for the U.S. Neurology Board certification. From 1995–1998, he was a research associate at the Computational Neurobiology Lab, Salk Institute for Biology Studies, San Diego, California. From 1998–2003 he was an assistant professor of Medicine at Duke University, core faculty at Duke's Brain Research and Analysis Center, as well as an adjunct professor in biomedical engineering and Duke's Center for Neural Analysis. He is currently an associate professor of medicine (neurology) and faculty member of the Pacific Parkinson's Research Center, and the Brain Research Centre at the University of British Columbia.

Z. Jane Wang received the B.Sc. degree from Tsinghua University, China, in 1996, with the highest honor, and the M.Sc. and Ph.D. degrees from the University of Connecticut in 2000 and 2002, respectively, all in electrical engineering. While at the University of Connecticut, she received the Outstanding Engineering Doctoral Student Award. She has been a research associate of the Electrical and Computer Engineering Department and Institute for Systems Research at the University of Maryland, College Park. Since August 2004, she has been with the Department Electrical and Computer Engineering at the University of British Columbia (UBC), Canada, as an assistant professor. Her research interests are in the broad areas of statistical signal processing, with applications to information security, biomedical imaging, genomic, and wireless communications. She coreceived the EURASIP *Journal on Applied Signal Processing* (JASP) Best Paper Award 2004, and she is an associate editor for the EURASIP *Journal on Bioinformatics and Systems Biology*. She coedited the book *Genomic Signal Processing and Statistics* and coauthored *Multimedia Fingerprinting Forensics for Traitor Tracing*.

Rafeef Abugharbieh is currently an assistant professor at the Department of Electrical and Computer Engineering, University of British Columbia (UBC) in Vancouver, Canada. She is the cofounder and codirector of the Biomedical Signal and Image Computing Laboratory (BiSICL), a multidisciplinary research laboratory at UBC dedicated to computational research in biomedical applications. She received her Ph.D., Technical Licentiate and M.Sc. (with distinction) degrees from the School of Electrical and Computer Engineering, Chalmers University of Technology in Goteborg, Sweden, in 2001, 1999, and 1996, respectively. She was a member of the Imaging and Image Analysis Group at the Department of Signals and Systems at Chalmers University from 1996–2001. In 1995, she obtained her B.Sc. degree in electrical engineering from the Faculty of Engineering and Technology at the University of Jordan in Amman, Jordan.

Todd C. Handy is currently an assistant professor in the Department of Psychology, University of British Columbia, Vancouver, Canada.

Address for Correspondence: Martin J. McKeown, FRCP(C), Pacific Parkinson's Research Centre, University of British Columbia, M31, Purdy Pavilion, University Hospital, UBC Site, 2221 Wesbrook Mall, Vancouver, British Columbia, V6T 2B5 Canada. Phone: +1 604 822 7516. Fax: +1 604 822 7866. E-mail: mmckeown@interchange.ubc.ca.

References

[1] M.J. McKeown, S. Makeig, G. Brown, S. Kindermann, and T.J. Sejnowski, "Modes or models: A critique on independent component analysis for fMRI—reply to commentary by Prof Friston," *Trends Cognitive Scis.*, vol. 2, p. 375. 1998.

[2] M.J. McKeown, T.P. Jung, S. Makeig, G. Brown, S.S. Kindermann, T.W. Lee, and T.J. Sejnowski, "Spatially independent activity patterns in functional MRI data during the stroop color-naming task," *Proc. Nat. Acad. Sci. U.S.A.,* vol. 95, no. 3, pp. 803–810, 1998.

[3] M.J. McKeown, S. Makeig, G.G. Brown, T.P. Jung, S.S. Kindermann, A.J. Bell, and T.J. Sejnowski, "Analysis of fMRI data by blind separation into independent spatial components," *Human Brain Mapping*, vol. 6, no. 3, pp. 160–88, 1998.

[4] M.J. McKeown, and T.J. Sejnowski, "Independent component analysis of fMRI data: Examining the assumptions," *Human Brain Mapping*, vol. 6, no. 5–6, pp. 368–372, 1998.

[5] V.D. Calhoun, J.J. Pekar, V.B. McGinty, T. Adali, T.D. Watson, and G.D. Pearlson, "Different activation dynamics in multiple neural systems during simulated driving," *Human Brain Mapping*., vol. 16, no. 3, pp. 158–167. 2002.

[6] S. Zeki, R.J. Perry, and A. Bartels, "The processing of kinetic contours in the brain," *Cerebral Cortex*., vol. 13, no. 2, pp. 189–202, 2003.

[7] J.R. Duann, T.P. Jung, W.J. Kuo, T.C. Yeh, S. Makeig, J.C. Hsieh, and T.J. Sejnowski, "Single-trial variability in event-related BOLD signals," *Neuroimage*, vol. 15, no. 4, pp. 823–835, 2002.

[8] M.J. McKeown, V. Varadarajan, S. Huettel, and G. McCarthy, "Deterministic and stochastic features of fMRI data: Implications for analysis of event-related experiments," *J. Neurosci. Methods*., vol. 118, no. 2, pp. 103–113, 2002.

[9] M. McKeown, "Deterministic and stochastic features of fMRI data: Implications for data averaging," in *Exploratory Analysis and Data Modeling in Functional Neuroimaging*, F.T. Sommer and A. Wichert, Eds. Cambridge, MA: MIT Press, 2003, pp. 63–76.

[10] K.J. Friston, "Statistical parametric mapping and other analyses of functional imaging data," in *Brain Mapping, the Methods,* A.W. Toga and J.C. Mazziotta, Eds. San Diego, CA: Academic, 1996, pp. 363–396.

[11] R.H. Myers, *Classical and Modern Regression with Applications*, 1st ed. Boston, MA: PWS-Kent*,* 1986.

[12] M.J. McKeown, L.K. Hansen, and T.J. Sejnowsk, "Independent component analysis of functional MRI: What is signal and what is noise?" *Current Opinion Neurobiology.*, vol. 13, no. 5, pp. 620–629, 2003.

[13] E. Zarahn, G.K. Aguirre, and M. D'Esposito, "Empirical analyses of BOLD fMRI statistics, I. Spatially unsmoothed data collected under null-hypothesis conditions," *Neuroimage*, vol. 5, no. 3, pp. 179–197, 1997.

[14] K. Petersen, L. Hansen, T. Koleda, E. Rostrup, and S. Strother, "On the independent components of functional neuroimages," in *Proc. 3rd Int. Conf. Independent Component Analysis Blind Source Separation (ICA2000)*. 2000, pp. 615–620.

[15] K.J. Friston, "Modes or models: A critique on independent component analysis for fMRI," *Trends Cognitive Sci.*, vol. 2, no. 10, pp. 373–375, 1998.

[16] V.D. Calhoun, T. Adali, G.D. Pearlson, and J.J. Pekar, "Spatial and temporal independent component analysis of functional MRI data containing a pair of task-related waveforms," *Human Brain Mapping*, vol. 13, no. 1, pp. 43–53, 2001.

[17] A.J. Bell and T.J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural Comput.,* vol. 7, no. 6, pp. 1129–1159, 1995.

[18] A. Hyvarinen and E. Oja, "A fast fixed-point algorithm for independent component analysis," *Neural Comput.*, vol. 9, no. 7, pp. 1483–1492, 1997.

[19] H. Attias and C.E. Schreiner, "Blind source separation and deconvolution by dynamic component analysis," in *Proc. 1997 IEEE Workshop Neural Networks Signal Processing VII*, 1997, pp. 456–465.

[20] C.H. Moritz, B.P. Rogers, and M.E. Meyerand, "Power spectrum ranked independent component analysis of a periodic fMRI complex motor paradigm," *Human Brain Mapping*., vol. 18, no. 2, pp. 111–122, 2003.

[21] M.J. McKeown, "Detection of consistently task-related activations in fMRI data with HYBrid Independent Component Analysis (HYBICA)," *NeuroImage*, vol. 11, no. 1, pp. 24–35, 2000.

[22] L.K. Hansen, J. Larsen, F.A. Nielsen, S.C. Strother, E. Rostrup, R. Savoy, N. Lange, J. Sidtis, C. Svarer, and O.B. Paulson, "Generalizable patterns in neuroimaging:" how many principal components?" *Neuroimage.,* vol. 9, no. 5, pp. 534–544, 1999.

[23] C.F. Beckmann and S.M. Smith, "Probabilistic independent component analysis for functional magnetic resonance imaging," *IEEE Trans. Med. Imag.*, vol. 23, no. 2, pp. 137–152, 2004.

[24] E.B. Hook and R.R. Regal, "Validity of methods for model selection, weighting for model uncertainty, and small sample adjustment in capture-recapture estimation," *Amer. J. Epidemiology*, vol. 145, no. 12, pp. 1138–1144, 1997.

[25] A.M. Smith, B.K. Lewis, U.E. Ruttimann, F.Q. Ye, T.M. Sinnwell, Y. Yang, J.H. Duyn, and J.A. Frank, "Investigation of low frequency drift in fMRI signal," *Neuroimage*, vol. 9, no. 5, pp. 526–533, 1999.

[26] C.G. Thomas, R.A. Harshman, and R.S. Menon, "Noise reduction in BOLD-based fMRI using component analysis," *Neuroimage.*, vol. 17, no. 3, pp. 1521–1537, 2002.

[27] R. Liao, J. Krolik, and M.J. McKeown, "An Information-Theoretic Criterion for Intra-subject Alignment of FMRI Time Series: Motion-Corrected Independent Component Analysis (MCICA)," *IEEE Trans. Med. Imag.*, vol. 24, pp. 29–43, 2005.

[28] P.A. Bandettini, E.C. Wong, R.S. Hinks, R.S. Tikofsky, and J.S. Hyde, "Time course EPI of human brain function during task activation," *Magn. Reson. Med.,* vol. 25, no. 2, pp. 390–397, 1992.

[29] P.A. Bandettini, and R.W. Cox, "Event-related fMRI contrast when using constant interstimulus interval: Theory and experiment," *Magn. Reson. Med.*, vol. 43, no. 4, pp. 540–548, 2000.

[30] R.L. Buckner, "Event-related fMRI and the hemodynamic response," *Hum. Brain Mapp.*, vol. 6, no. 5–6, pp. 373–377, 1998.

[31] A.M. Dale and R.L. Buckner, "Selective averaging of rapidly presented individual trials using fMRI," *Hum. Brain Mapp.*, vol. 5, pp. 329–340, 1997.

[32] K.J. Friston, O. Josephs, G. Rees, and R. Turner, "Nonlinear event-related responses in fMRI," *Magn. Reson. Med.*, vol. 39, no. 1, pp. 41–52, 1998.

[33] A.L. Paradis, P.F. Van de Moortele, D. Le Bihan, and J.B. Poline, "Slice acquisition order and blood oxygenation level dependent frequency content: An event-related functional magnetic resonance imaging study," *Magma*, vol. 13, no. 2, pp. 91–100, 2001.

[34] S.A. Huettel, J.D. Singerman, and G. McCarthy, "The effects of aging upon the hemodynamic response measured by functional MRI," *Neuroimage*, vol. 13, no. 1, pp. 161–175, 2001.

[35] S.A. Huettel, M.J. McKeown, A.W. Song, S. Hart, D.D. Spencer, T. Allison, and G. McCarthy, "Linking hemodynamic and electrophysiological measures of brain activity: Evidence from functional MRI and intracranial field potentials," *Cerebral Cortex*, vol. 14, 2004, pp. 165–173.

[36] A.M. Dale, "Optimal experimental design for event-related fMRI," *Human Brain Mapping*, vol. 8, no. 2–3, pp. 109–114. 1999.

[37] R.N.A. Henson, M.D. Rugg, and K.J. Friston, "The choice of basis functions in event-related fMRI," *Neuroimage,* vol. 13, suppl., no. 1, p. 127, 2001.

[38] H. Gudbjartsson and S. Patz, "The Rician distribution of noisy MRI data," *Magn. Reson. Med.*, vol. 34, no. 6, pp. 910–914, 1995.

[39] Y.M. Kadah, S. LaGonte, and X. Hu. "Denoising of Event-Related fMRI Data Based on a Rician Noise Model For Robust Analysis Using ICA," in *Proc. Int. Soc. Magnetic Resonance Medicine*, Honolulu, 2002*,* vol. 10.

[40] S.M. LaConte, S.C. Ngan, and X. Hu, "Wavelet transform-based Wiener filtering of event-related fMRI data," *Magn. Reson. Med.*, vol. 44, no. 5, pp. 746–757, 2000.

[41] G. Glover and S. Lai, "Reduction of susceptibility effects in BOLD fMRI using tailored RF pulses," in *Proc. ISMRM*, 1998, p. 298.

[42] J.E. Jackson, *A User's Guide to Principal Components.* New York: Wiley, 1991.

[43] I.K. Fodor, "A survey of dimension reduction techniques," Lawrence Livermore Nat. Lab., UCRL-ID-148494, 2002.

[44] D.B. Rowe and B.R. Logan, "A complex way to compute fMRI activation," *Neuroimage*, vol. 23, no. 3, pp. 1078–1092, 2004.

[45] A. Mechelli, C.J. Price, R.N. Henson, and K.J. Friston, "Estimating efficiency a priori: A comparison of blocked and randomized designs," *Neuroimage*, vol. 18, no. 3, pp. 798–805, 2003.

[46] S. Makeig, T.P. Jung, A.J. Bell, D. Ghahremani, and T.J. Sejnowski, "Blind separation of auditory event-related brain responses into independent components," *Proc. Nat. Acad. Sci. U.S.A.*, vol. 94, no. 20, pp. 10979–10984, 1997.

[47] M. McKeown, V. Varadarajan, S. Huettel, and G. McCarthy, "Deterministic and stochastic features of fMRI data: Implications for analysis of event-related experiments," *J Neurosci. Methods*, vol. 118, no. 2, pp. 103–113, 2002.