



ACADEMIC  
PRESS

Available online at [www.sciencedirect.com](http://www.sciencedirect.com)

SCIENCE @ DIRECT®

NeuroImage

NeuroImage 20 (2003) 625–642

[www.elsevier.com/locate/ynimg](http://www.elsevier.com/locate/ynimg)

# Multivariate analysis of neuronal interactions in the generalized partial least squares framework: simulations and empirical studies

Fa-Hsuan Lin,<sup>a,b,\*</sup> Anthony R. McIntosh,<sup>c</sup> John A. Agnew,<sup>d</sup> Guinevere F. Eden,<sup>d</sup>  
Thomas A. Zeffiro,<sup>d</sup> and John W. Belliveau<sup>b</sup>

<sup>a</sup> Harvard-MIT Division of Health Sciences and Technology, Charlestown, MA 02446, USA

<sup>b</sup> MGH-MIT-HMS Athinoula A. Martinos Center for Biomedical Imaging, Charlestown, MA 02446, USA

<sup>c</sup> Rotman Research Institute of Baycrest Center, University of Toronto, Ontario, M6A 2E1, Canada

<sup>d</sup> Center for the Study of Learning, Georgetown University, Washington, DC 20057, USA

Received 26 August 2002; revised 28 April 2003; accepted 21 May 2003

## Abstract

Identification of spatiotemporal interactions within/between neuron populations is critical for detection and characterization of large-scale neuronal interactions underlying perception, cognition, and behavior. Univariate analysis has been employed successfully in many neuroimaging studies. However, univariate analysis does not explicitly test for interactions between distributed areas of activity and is not sensitive to distributed responses across the brain. Multivariate analysis can explicitly test for multiple statistical models, including the designed paradigm, and allows for spatial and temporal model detection. Here, we investigate multivariate analysis approaches that take into consideration the 4D (time and space) covariance structure of the data. Principal component analysis (PCA) and independent component analysis (ICA) are two popular multivariate approaches with distinct mathematical constraints. Common difficulties in using these two different decompositions include the following: classification of the revealed components (task-related signal versus noise), overall signal-to-noise sensitivity, and the relatively low computational efficiency (multivariate analysis requires the entire raw data set and more time for model identification analysis). Using both Monte Carlo simulations and empirical data, we derived and tested the generalized partial least squares (gPLS) framework, which can incorporate both PCA and ICA decompositions with computational efficiency. The gPLS method explicitly incorporates the experimental design to simplify the identification of characteristic spatiotemporal patterns. We performed parametric modeling studies of a blocked-design experiment under various conditions, including background noise distribution, sampling rate, and hemodynamic response delay. We used a randomized grouping approach to manipulate the degrees of freedom of PCA and ICA in gPLS to characterize both paradigm coherent and transient brain responses. Simulation data suggest that in the gPLS framework, PCA mostly outperforms ICA as measured by the receiver operating curves (ROCs) in SNR from 0.01 to 100, the hemodynamic response delays from 0 to 3 TR in fMRI, background noise models of Gaussian, sub-Gaussian, and super-Gaussian distributions and the number of observations from 5, 10, to 20 in each block of a six-block experiment. Further, due to selective averaging, the gPLS method performs robustly in low signal-to-noise ratio ( $<1$ ) experiments. We also tested PCA and ICA using PLS in a simulated event-related fMRI data to show their similar detection. Finally, we tested our gPLS approach on empirical fMRI motor data. Using the randomized grouping method, we are able to identify both transient responses and consistent paradigm/model coherent components in the 10-epoch block design motor fMRI experiment. Overall, studies of synthetic and empirical data suggest that PLS analysis, using PCA decomposition, provides a stable and powerful tool for exploration of fMRI/behavior data.

© 2003 Elsevier Inc. All rights reserved.

## Introduction

The analysis of functional magnetic resonance imaging (fMRI) data (Belliveau et al., 1990, 1991; Kwong et al., 1992; Ogawa et al., 1990) is intended to reveal the under-

lying spatiotemporal interactions of neuronal populations. This includes the identification and characterization of orchestrated brain areas in both spatial and temporal domains. Traditionally, univariate statistical methods (Bandettini et al., 1993; Friston et al., 1995b, 1995c; Worsley and Friston, 1995), such as the  $t$  test or nonparametric Kolmogorov–Smirnov test and their associated statistical parametric maps, have been used quite often to detect cognitive task-

\* Corresponding author. MGH-NMR Center, Rm. 2301, Bldg. 149  
13th St. Charlestown MA 02129.

E-mail address: [fhlin@mit.edu](mailto:fhlin@mit.edu) (F.-H. Lin).

related components. A reference temporal profile, which usually is associated with the experiment stimulus paradigm, is required for hypothesis testing. Such univariate approaches test brain voxels distributed over the brain independently, and they ignore the anatomical connections among brain loci. This lack of consideration of spatial interactions inspired the utilization of multivariate methods for detection and estimation of spatial activation and temporal dynamics of the brain (Bullmore et al., 1996; Friston et al., 1993, 1995a; McIntosh et al., 1996; McKeown et al., 1998). Multivariate statistics regard the whole spatiotemporal data as an entity during estimation of active brain areas. We investigated two different approaches (PCA and ICA) for data decomposition within our generalized Partial Least Squares framework.

Principal component analysis (PCA) and associated eigenimage (Bullmore et al., 1996; Friston, et al., 1993, 1995a) are examples of multivariate methods to decompose the data into subspaces. These subspaces account for the total spatiotemporal variance in the order of explained variance. The first eigenimage in PCA represents the spatial brain patterns and their associated timing, which explains the most of the variance. The subsequent eigenimages explain the residual variance using the constraint of orthonormality to the preceding eigenimages. PCA is thus a successive variance decomposition process with an orthonormality constraint among eigenimages.

Independent component analysis (ICA) (Bell and Sejnowski, 1995; McKeown et al., 1998) is an alternative approach. Unlike PCA, which assumes an orthonormality constraint, ICA assumes statistical independence among a collection of spatial patterns. ICA uses a linear mixture assumption to decompose the original data into spatially statistically independent components. Both PCA and ICA are model-free methods (independent of task paradigm design) to explore the structure of the data without a priori information. This data-driven property excludes biases and enables the detection of transient responses and artifacts inside the data.

The difficulties of applying PCA and ICA to fMRI analysis include the need to identify separated components and computational requirements. Owing to the size of collected fMRI data, PCA and ICA both identify a large collection of intrinsic structures, which makes it difficult for researchers to establish direct correspondence between the identified components and the experiment hypotheses. Both PCA and ICA require high computational power to decompose the data matrix. ICA is especially computationally expensive compared to PCA, even when utilizing an advanced algorithm (Hyvarinen, 1999).

Regardless of the analysis approach used to explore the data matrix, prior information about the design of the experiment is valuable for detection and estimation of brain activity. The univariate general linear model (GLM) framework (Friston, et al., 1995c) utilizes experimental paradigms and presumed basis functions to encode multiple null

hypotheses and confounds in the design matrix as regressors. On the other hand, PCA and ICA make use of the information about the experimental design *after* the decomposition (Friston et al., 1995a; McKeown et al., 1998) instead of incorporating such information *before* the decomposition, like GLM. Thus, PCA and ICA are more data-driven methods due to the post hoc identification of decomposed components.

Partial least squares (PLS) (McIntosh et al., 1996) is a compromise between hypothesis-driven and data-driven approaches. PLS uses an intermediate step of selective averaging of the experimental design to incorporate hypotheses explicitly. PCA is then used to decompose the collapsed data to reveal intrinsic structures. PLS provides advantages of dimensional reduction of the data and signal-to-noise enhancement due to the selective averaging. The dimension reduction process relieves the difficulties of posterior interpretation for numerous components resulting from direct multivariate decomposition by PCA.

Traditionally PLS has utilized PCA for decomposition. Here, we extended the original partial least square framework (McIntosh et al., 1996) to incorporate both PCA and ICA as data decomposition alternatives. Furthermore, a randomized grouping/selective averaging approach was used to generalize the a priori contrast matrix to investigate possible transient and time-locked activities in both simulations and empirical studies. To quantitatively evaluate the use of either PCA or ICA in this generalized PLS (gPLS) framework, we performed parametric simulations for a blocked-design fMRI experiment at conditions of various SNR, time of repetition (TR), background noise models, experimental paradigms, and hemodynamic response delays. Also, we presented the simulation results for an event-related fMRI experiments to test the differences when PCA or ICA is utilized as the decomposition in PLS. In addition to modeling studies, we analyzed an empirical blocked-design fMRI experiment of voluntary finger tapping. We examined the capability of gPLS to detect consistently task-related components and transient responses in both task-control single-contrast and parametrically designed multiple-contrast experiments. Our generalized partial least squares approach proposed here takes advantage of the a priori task paradigm design yet allows for detection of potential brain activity unassociated with the a priori contrast matrix. The pros and cons of using either PCA or ICA decomposition schemes to optimize fMRI signal detection are discussed.

## Theory

### *Multivariate approach to reveal the functional connectivity: PCA*

The collected data in functional brain imaging studies can be collapsed into a two-dimensional matrix, **D**, which includes both spatial and temporal information and is re-

ferred to the data matrix. Without loss of generality, we assume here that each row encodes one time-point/condition acquisition for the whole brain and each column represents one voxel's time series or voxel's recording across different subjects and different conditions. For example,  $\mathbf{D}_{ij}$  represents the value of voxel  $j$  in time point  $i$  acquisition.

Different decomposition procedures can be applied to the data matrix to reveal the internal structures. Here we refer to total decomposition (TD) as the approach that directly decomposes the spatiotemporal data matrix,  $\mathbf{D}$ , by multivariate tools. PCA decomposes the two-dimensional matrix into orthogonal subspaces, which are termed latent variables (LV) in the gPLS framework. The outer product of the left singular vector and the right singular vector, weighted by the corresponding singular value, results in each latent variable expressed as:

$$\mathbf{D} = {}_{\text{PCA}}\mathbf{U} \times {}_{\text{PCA}}\mathbf{S} \times {}_{\text{PCA}}\mathbf{V}^T \quad (1)$$

$${}_{\text{PCA}}\mathbf{U}^T {}_{\text{PCA}}\mathbf{U} = {}_{\text{PCA}}\mathbf{U} {}_{\text{PCA}}\mathbf{U}^T = \mathbf{I} \quad (2a)$$

$${}_{\text{PCA}}\mathbf{V}^T {}_{\text{PCA}}\mathbf{V} = {}_{\text{PCA}}\mathbf{V} {}_{\text{PCA}}\mathbf{V}^T = \mathbf{I}. \quad (2b)$$

Here  $\mathbf{I}$  denotes the identity matrix. Each LV consists of a singular value (a diagonal entry of  ${}_{\text{PCA}}\mathbf{S}$ ), one left singular vector (a column of  ${}_{\text{PCA}}\mathbf{U}$ ), and one right singular vector (a row of  ${}_{\text{PCA}}\mathbf{V}^T$ ). The left singular vector in each LV ( ${}_{\text{PCA}}\mathbf{U}$ ), which is termed the “design LV” in the following application of PCA in gPLS, gives the loading of different time points/conditions to maximize the explained variance in the associated LV under the orthonormality constraint to the remaining design LVs. In addition, the right singular vector ( ${}_{\text{PCA}}\mathbf{V}$ ), which is named “brain LV” in the following application of PCA in gPLS, gives the loading of each voxel to maximize the explained variance in the associated LV subjected to the orthonormality constraint to other right singular vectors. Singular values,  ${}_{\text{PCA}}\mathbf{S}$ , are metrics to quantify the significance of each LV. Larger singular values represent a more dominant contribution from the corresponding LV to the total variance in the data matrix. The proportion of the square of one singular value over the sum of squares of singular values provides the quantitative significance of the latent variable. Each design LV provides the physiological inference of the latent variable, and the brain LV represents the spatial loading of the effect on different voxels.

The data matrix can be reconstructed from the latent variables created by the outer product of the corresponding left singular vector and the right singular vector, weighted by the singular value. The sequential sum of all latent variables constitutes the least squares fit of the data matrix in terms of minimizing the mean square error. The spectrum theorem (ref. linear algebra) describes the reconstruction procedure from the decomposed components as follows:

$$\mathbf{D} = \sum_i {}_{\text{PCA}}\mathbf{S}_{ii} ({}_{\text{PCA}}\tilde{\mathbf{U}}_{\cdot i} \times {}_{\text{PCA}}\tilde{\mathbf{V}}_{i \cdot}) \quad (3)$$

where  ${}_{\text{PCA}}\tilde{\mathbf{U}}_{\cdot i}$  denotes the  $i$ th column of matrix  ${}_{\text{PCA}}\mathbf{U}$  and  ${}_{\text{PCA}}\tilde{\mathbf{V}}_{i \cdot}$  denotes the  $i$ th row of matrix  ${}_{\text{PCA}}\mathbf{V}$ .

PCA is the mathematical process to reorganize the total variance in the new coordinate system by orthogonal rotation. PCA provides such a rotation subject to the constraints on orthonormality among coordinate axes and maximal variance of the projected raw data on the new coordinate system after rotation.

#### *Multivariate approach to reveal the functional connectivity: ICA*

Independent component analysis (ICA) is an alternative multivariate brain-imaging data-analysis tool. Instead of decomposing the data into orthogonal subspaces, ICA minimizes the mutual information among “channels,” which refers to rows in the data matrix. The critical assumption of ICA is that the recorded signal is the linear time-invariant mixture of several statistically independent components. The ICA algorithm estimates the mixture matrix and it searches an “unmixing” linear operator to restore these spatially independent components. This is formulated as follows:

$$\mathbf{D} = \mathbf{W} \times \mathbf{X}. \quad (4)$$

Each row of  $\mathbf{X}$  represents one spatially independent component (IC). Observed data matrix,  $\mathbf{D}$ , is generated by the linear mixture of these independent components via the linear mixing operator,  $\mathbf{W}$ .

Independent components are found by reversing the mixture process as follows:

$$\hat{\mathbf{X}} = (\hat{\mathbf{W}})^{-1} \times \mathbf{D}. \quad (5)$$

The similarity between the independent components in ICA and the latent variables in PCA can be compared by Eq. (1) and Eq. (5). This generates the following analogy between PCA and ICA brain latent variables and design latent variable:

$${}_{\text{ICA}}\tilde{\mathbf{U}}_{\cdot i} = \frac{\mathbf{W}_{\cdot i}}{\|\mathbf{W}_{\cdot i}\|} \quad (6a)$$

$${}_{\text{ICA}}\tilde{\mathbf{V}}_{i \cdot}^T = \frac{\hat{\mathbf{X}}_{i \cdot}}{\|\hat{\mathbf{X}}_{i \cdot}\|} \quad (6b)$$

$${}_{\text{ICA}}\mathbf{S}_{ii} = \sqrt{\frac{\|\hat{\mathbf{W}}_{\cdot i} \times \hat{\mathbf{X}}_{i \cdot}\|}{\|{}_{\text{ICA}}\tilde{\mathbf{U}}_{\cdot i} \times {}_{\text{ICA}}\tilde{\mathbf{V}}_{i \cdot}^T\|}}, \quad (6c)$$

where the  $\|\cdot\|$  operator denotes the root-mean-square of the vector. The defined design LV,  ${}_{\text{ICA}}\tilde{\mathbf{U}}$ , and brain LV,  ${}_{\text{ICA}}\tilde{\mathbf{V}}$ , are normalized to unit power as they are in PCA. A diagonal singular matrix can also be constructed by placing defined ICA singular values in Eq. (6c) at diagonal entries. Similar to the spectrum theorem in PCA, the derived design and brain latent variables in ICA reconstruct the original data matrix by summing up the subspaces by the cross-product of the ICA design and the brain LV weighted by the ICA singular value as follows:

$$D = \sum_i \text{ICA} S_{ii} (\text{ICA} \tilde{U}_i \times \text{ICA} \tilde{V}_i^T). \quad (7)$$

### Functional connectivity analysis by generalized partial least squares

PLS (McIntosh et al., 1996) is an alternative multivariate approach for analyzing the functional neuroimaging matrices within the new space of proposed contrasts (or hypotheses) of interest. The goal here is to generalize the PLS algorithm (gPLS) to use both PCA and ICA as decomposition tools. Like the original PLS approach, generalized PLS constructs an effect space of reduced dimension via the interaction of the data matrix,  $\mathbf{D}$ , and the contrast matrix,  $\mathbf{C}$ , which encodes multiple comparisons mathematically.

For a data matrix  $\mathbf{D}$  of  $n$  time points (or conditions) and  $m$  voxels, a contrast matrix  $\mathbf{C}$  of size  $c$  by  $n$  can be constructed to incorporate  $c$  contrasts, each of which represents one null hypothesis. The cross-product of these two matrices creates the effect space  $\mathbf{E}$  as follows:

$$\mathbf{E} = \mathbf{C}^T \times \mathbf{D}. \quad (8)$$

By use of the contrast matrix, the dimensionality of the original spatiotemporal neuroimaging recordings is decreased dramatically compared to the total decomposition of the data matrix in most fMRI experiment scenarios. This leads to the advantages of gPLS for increased power for signal detection and estimation, resulting from the weighted averaging provided by the contrast matrix, and for decreased complexity in interpreting the revealed structures within the data due to the reduction of dimension.

One would like to contrast effects in the temporal domain without temporal overlapping. In a task-control experiment to detect baseline-activation differences or in parametrically designed experiments to reveal condition-related effects, orthonormal Helmert contrasts (McIntosh et al., 1996) can be used as the contrast matrix for multiple conditions comparison. A Helmert contrast matrix of temporal dimension  $n$  is written explicitly as follows:

$$H_{ij} = \begin{cases} 0, & \text{when } i < j \\ \frac{1}{\sqrt{1 + 1/(n-j)}}, & \text{when } i = j \\ -\frac{1}{(n-j)\sqrt{1 + 1/(n-j)}}, & \text{when } i > j. \end{cases} \quad (9)$$

Each column of the Helmert contrast matrix can be used to represent one null hypothesis. Two properties of the Helmert basis function are useful for detecting activation: first, the fact that each basis with sum of entries equal to zero explicitly implements a matched filter by providing weightings to various temporal observations. Each basis tests one single null hypothesis in the selective averaging process by correlating the contrast matrix and the data matrix. Second, all Helmert bases are of unit variance and orthogonal among each other such that no bias toward any

comparison encoded by an individual basis is generated in the subsequent decomposition.

To perform multiple comparisons, observations are temporally segregated into groups, and the associated Helmert bases are created to encode the differences among them. The following decomposition algorithm identifies the coefficients for these bases by maximizing the total effect created by individual comparison using either PCA or ICA. For a blocked-design experiment, groups can be either different epochs for a single subject experiment or different condition/subject indices for a cross-subject, parametrically designed study.

In the case of testing a single hypothesis (e.g., detection of baseline-activation contrast in blocked-design experiment) only one associated contrast vector is created and traditional PLS fails in this degenerated case, because of the rank deficiency of the effect space. To resolve this problem, we propose “randomized grouping,” which is essentially making use of the repeated observations by posing multiple null hypotheses to test differences between partitioned groups. In practice, we may assume there are  $g$  groups in a  $t$ -epoch blocked-design experiment, where  $g \leq t$ . A supplementary contrast matrix encoding the difference of  $g$  groups can be created by randomly partitioning epochs using Helmert bases. The new contrast matrix, consisting of the baseline-activation contrast vector as well as the epoch difference encoded in the supplementary contrast matrix, enriches the content of the effect space. The dimension of the effect space after randomized grouping is still dramatically smaller than that of the data matrix in conventional experiment setups. This property can be exploited for iterations of randomized grouping to estimate the errors in the grouping process.

Using the total decomposition, the dimension of the matrix fed into the multivariate decomposition is often the number of temporal observations, because an fMRI experiment generates many more spatial voxels than temporal scans. In contrast, the PLS and gPLS frameworks create the effect space with dimension equal to the number of contrasts to be tested. This reduction in dimension provides the advantages of not only increased power for signal detection/estimation (due to the explicit selective averaging by columns of the contrast matrix) but also decreased computational complexity due to the reduced row dimension of the effect space. Randomized grouping further provides the flexibility to manipulate the size and the rank of the effect space. Various dimension-reduced effect spaces constitute the sampling pool to estimate the robustness of the subsequent decomposition.

Either PCA or ICA can be applied to the effect space to explore the intrinsic structures associated with the proposed contrasts. In PCA, revealed latent variables sequentially account for the total variances in the effect space  $\mathbf{E}$ , subject to the orthonormality constraint. In ICA, however, the mutual information among spatial patterns of different effects and comparisons is minimized among rows of  $\mathbf{E}$ . Latent

variables are generated by either PCA or ICA decomposition. Each design LV represents the loadings of the proposed contrasts, and each brain LV summarizes the spatial loading of different voxels in the associated LV, which imply the characteristic brain activity patterns. For both PCA and ICA, the loading at different time points/conditions can be obtained by calculating the design score ( $S_{\text{design}}$ ), which is defined as the product of the contrast matrix and the design LV as follows:

$$S_{\text{design}} = C \times U \quad (10)$$

Design scores indicate the physiological implications of the latent variables because the temporal loadings can be covaried with either the experiment paradigm or the temporal confounds. The loading of all voxels for each revealed latent variable can be represented by the brain score ( $S_{\text{brain}}$ ), which is defined as the product of the data matrix and the brain LV as follows:

$$S_{\text{brain}} = D \times V. \quad (11)$$

Each column of the brain score constitutes the loading of single brain LV's at different time points/conditions.

The name generalized partial least squares (gPLS) is derived from the process to decompose the effect space, which is the temporal covariance between the contrast designs and the observed data. In total decomposition, people deal with the complete covariance matrix of data, and thus it is computationally expensive. The effect space is a subset of the complete covariance matrix, correlating all spatio-temporal observations and null hypotheses encoded in the contrast matrix. gPLS deals with only a fraction of the complete covariance, enabling the fast computation of latent variables by reducing the dimension of the matrix to be decomposed. Note that the generalized partial least squares and total decomposition are equivalent when the contrast matrix is identity.

## Methods

### *Quantifying results from PCA and ICA by receiver operating curve analysis*

Using simulation data, quantitative comparison of PCA and ICA decomposition was performed by using receiver operating curves (ROC) to access the detection-cost characteristics of these two decomposition approaches in the gPLS and TD framework. The ROC area index is defined as the area under each ROC. An ROC area around 0.5 represents an inferior separation of signal from noise by the testing procedure, whereas an area close to 1 implies a good differentiation. Figure 1 illustrates the ROCs and their underlying areas when various thresholds are set to distinguish two Gaussian distributions of identical unit variance and separated at different means (D).

Given a single revealed brain latent variable and the

activated voxel indices, an ROC area metric (the ROC index) can be obtained. To characterize all latent variables for detection, a “weighted ROC index” for brain latent variables is defined as the weighted sum of all ROC areas from all LVs. One possible choice of the weighting factor for the latent variable  $i$ ,  $W_i$ , is the latent variable specific “power fraction” equal to the fraction of the sum-of-squares of singular values.

$$W_i = \frac{S_{ii}^2}{\sum_k S_{kk}^2}. \quad (12)$$

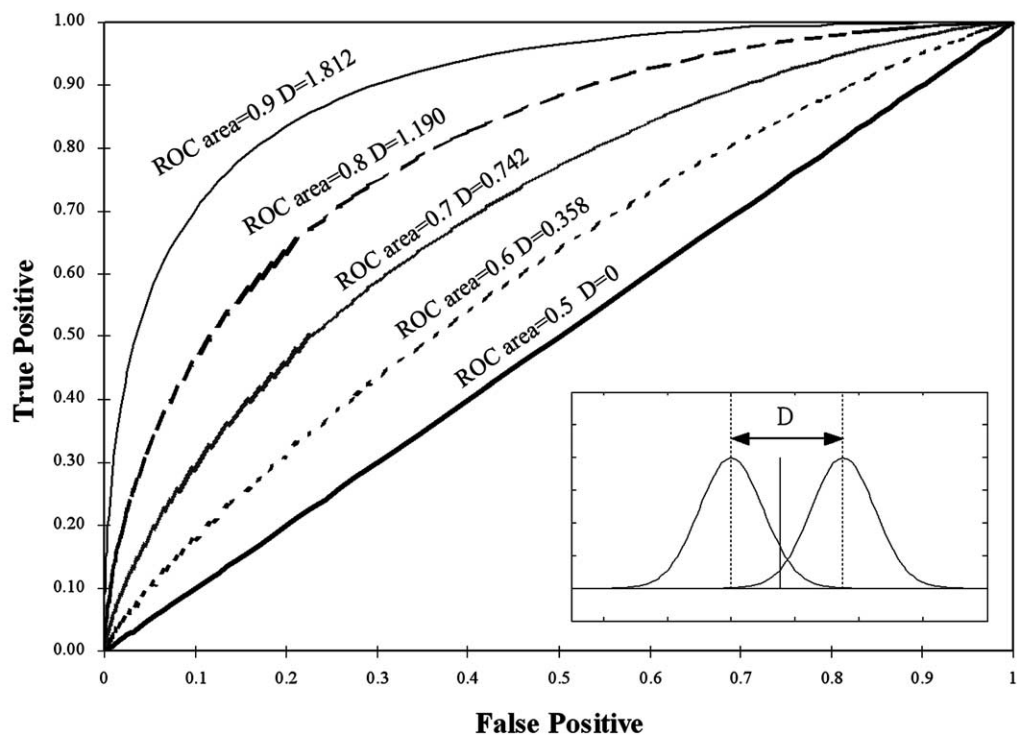
These two ROC area metrics enable the assessment of detection power by either an individual latent variable or the ensemble of LVs. Ideally, a perfect detection without either Type I or Type II error makes both the ROC index and the weighted ROC index equal to 1.

The brain LV from either PCA or ICA can be regarded as a multidimensional vector. Due to the normalization in Eq. (2) and Eq. (6), each brain LV has unit power, and we propose an angle metric  $\theta$  to quantify the (dis-)similarity between two brain LVs by the multidimensional inner-product as follows:

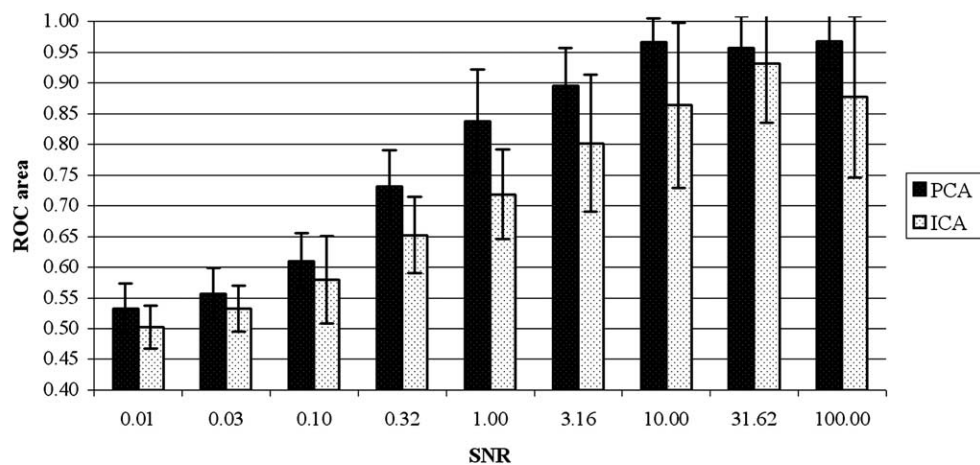
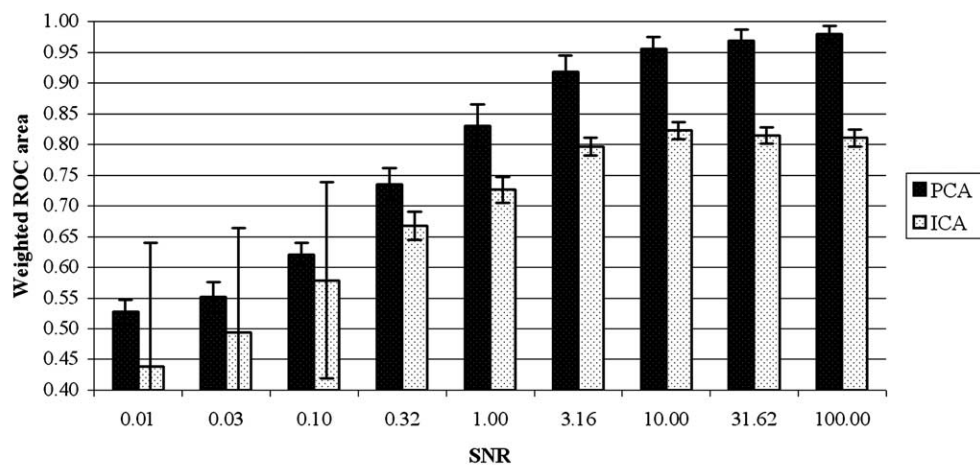
$$\theta(\tilde{V}_1, \tilde{V}_2) = \tilde{V}_1^T \cdot \tilde{V}_2. \quad (13)$$

### *Simulation*

Using a blocked-design experiment, a data matrix of dimension 120 rows and 10,000 columns was created to simulate the spatiotemporal observations of 6 epochs, each of which had a baseline of 10 time points and a stimuli of 10 time points. Values of the time series of each voxel consist of foreground signal and background noise at different signal-to-noise ratio (SNR), which is defined as the power ratio of the artificially created signal patterns and the background noise realizations. Two categories of signal patterns of activations were created: orthogonal activation patterns and nonorthogonal activation patterns. In a task-control blocked-design experiment, orthogonal patterns arise from the perfect alignment of a voxel's activating response to the timing of the experiment stimuli. Additionally, multiple subjects and/or multiple conditions in the experiments also create temporally, and individually, orthogonal time series when there is no interaction among conditions and subjects. Note that nonorthogonal activation delay (NOAD) can be present in the realistic fMRI data due to the hemodynamic delay of the onset of brain response to the stimuli. We varied NOAD between 0 to 3 TR in fMRI acquisition to simulate the delays of the activity relative to the orthogonal activation due to experimental paradigms. In the conventional TR = 2 s setting, this is equivalent to accommodating 0–6 s hemodynamic response delay. The background noise for each voxel was assumed to follow three probability distribution functions (PDF): Gaussian distribution, super-Gaussian distribution, and sub-Gaussian distribution. The Gaussian background noise was created by the normal dis-



1



2

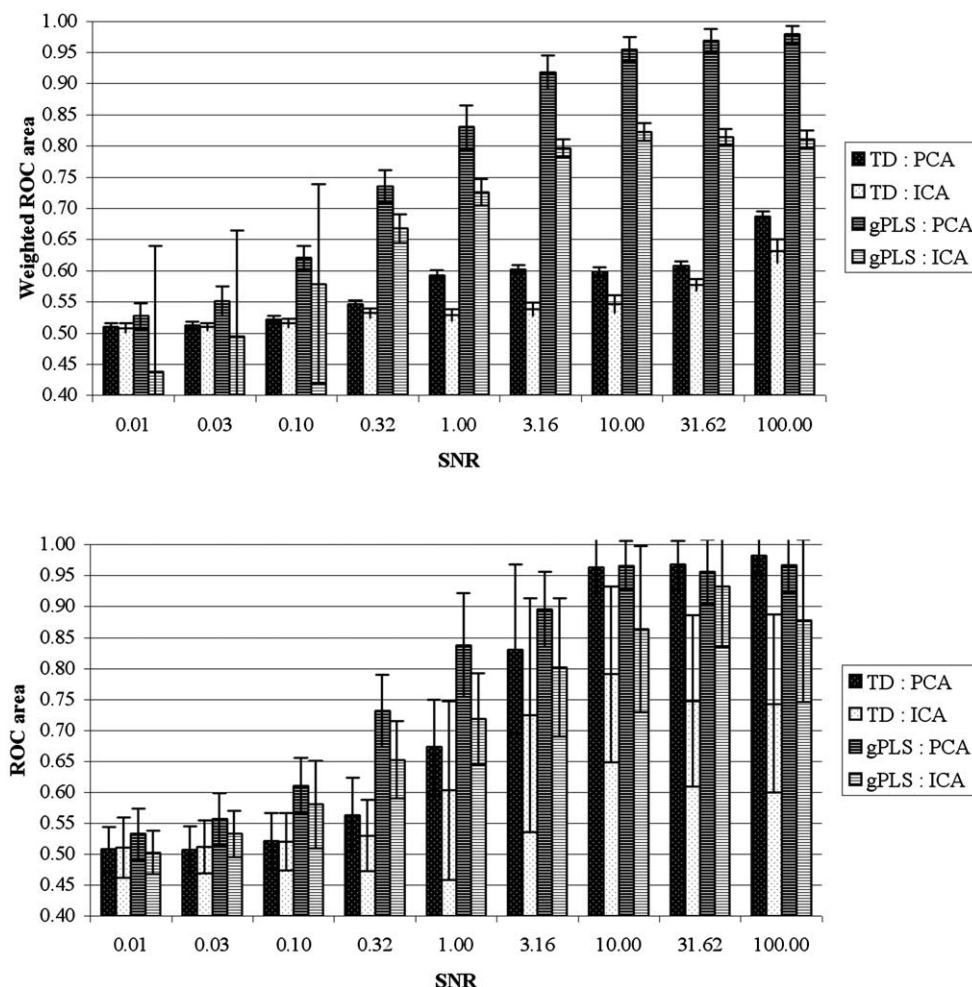


Fig. 3. The weighted ROC area (upper panel) and the most correlated LV (lower panel) by PCA and ICA using gPLS and TD with Gaussian background noise at various SNR. Five orthogonal activation patterns with six epochs and 10 samples in each epoch were simulated here.

tribution with zero mean and  $\sigma_i^2$  variance, denoted by  $N(0, \sigma_i^2)$ , where the variance follows another normal distribution  $\sigma_i^2 \sim N(5, 1)$ . A super-Gaussian noise model was implemented using an exponential distribution with zero mean, which was equivalent to a kurtosis of 6. We also adjusted the variance of each voxel's time series to follow another exponential distribution, with the mean equal to 5 and the variance equal to 1. A sub-Gaussian noise model was created by uniform distribution between  $-0.5$  and  $0.5$ , with the variance of each voxel's time series following another uniform distribution between 3 and 7.

The duration of activation for each activated pattern may affect the signal detection. Thus, we varied the temporal sampling rate (TSR), which is defined as the number of

scans in a fMRI experiment within each epoch/condition, to test the difference in the detection power. TSR changed from 5, 10, and 20, in the six-block simulations with corresponding changes of the number of rows in the data matrix. We performed iterative analyses by parametric variations over TSR, SNR, orthogonal and nonorthogonal patterns, noise models, as well as decomposition schemes by PCA or ICA. Each combination of simulation parameters was repeated for 30 iterations to evaluate the averaged performance. This also generated the error estimates for each set of the simulation parameters. For each gPLS simulation, a contrast matrix used the Helmert basis Eq. (9) to encode the orthogonal contrasts. Total decomposition of the data matrix by either PCA or ICA at the same parameter

Fig. 1. The ROCs and their area metrics from varying thresholds to distinguish two Gaussian distributions with unit variance and various mean differences. When the distance of the mean of two Gaussians is more than 1.8, the ROC area is higher than 0.9.

Fig. 2. (Top panel) The weighted ROC area by PCA and ICA using gPLS from Gaussian background noise at various SNRs. (Bottom panel) The ROC area from the most correlated latent variable by PCA and ICA using gPLS from Gaussian background noise at various SNRs. Five orthogonal activation patterns with six epochs and 10 samples in each epoch were simulated here.

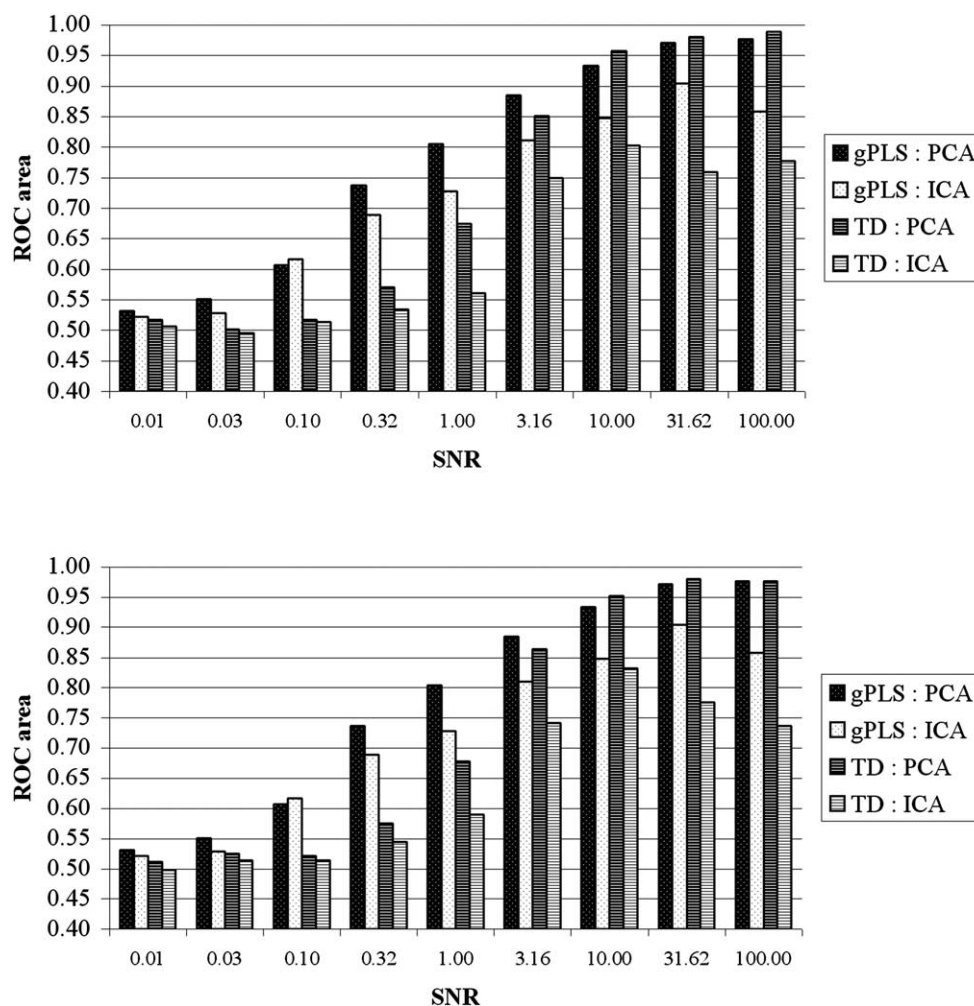


Fig. 4. The most correlated LV by PCA and ICA using gPLS and TD on super-Gaussian (top panel) and sub-Gaussian (bottom panel) background noise at various SNR. Five orthogonal activation patterns with six epochs and 10 samples in each epoch were simulated here.

settings was also simulated to contrast the detection power of gPLS.

We compared the outcome of applying PCA versus ICA decompositions to a simulated event-related fMRI data set. The data were created by the same methods as outlined in Della-Maggiore et al. (2002). Briefly, baseline fMRI time-series data were extracted from actual EPI MRI. First, we generated the baseline activity of the simulated data sets by using a first-order autoregressive plus white-noise model derived empirically by (Purdon and Weisskoff, 1998). Then, for defined epoch in the time series for five “subjects,” a 2% signal change was added to three voxel clusters ( $3 \times 3 \times 2$  voxels/cluster). This signal change was represented by a modeled hemodynamic response function (Cohen, 1997). Three tasks were modeled for each subject, one without any change in activation (baseline) and two tasks showing unique activation patterns.

These data were analyzed using the spatiotemporal variation of PLS (Lobaugh et al., 2001). Here the voxel time series are expressed in the same dimension so that the resulting data matrix has space and time extending along the

rows of the matrix and observations/subjects along the column dimension. This enables the same decomposition method to be used on the data matrix as for PLS on block-design fMRI, but identifies both the timing and location of task-related differences. The first latent variables from both ICA and PCA decompositions are rendered to illustrate the most significant activation.

Independent component analysis was implemented by the FastICA algorithm (Hyvärinen, 1999). We used the Matlab R11 (Waltham, MA) built-in function to calculate PCA. The Intel Pentium-III 450 MHz PC (Santa Clara, CA) was used as the hardware platform for the simulation and the data analysis.

#### *fMRI experiments of voluntary finger movement*

A right-handed subject executed a button press using the left hand in response to a visual stimulus appearing at three different frequencies: 0.3, 1, and 3 Hz during the task conditions using a block design, which consisted of six task time points (TR) and six baseline time points per epoch in



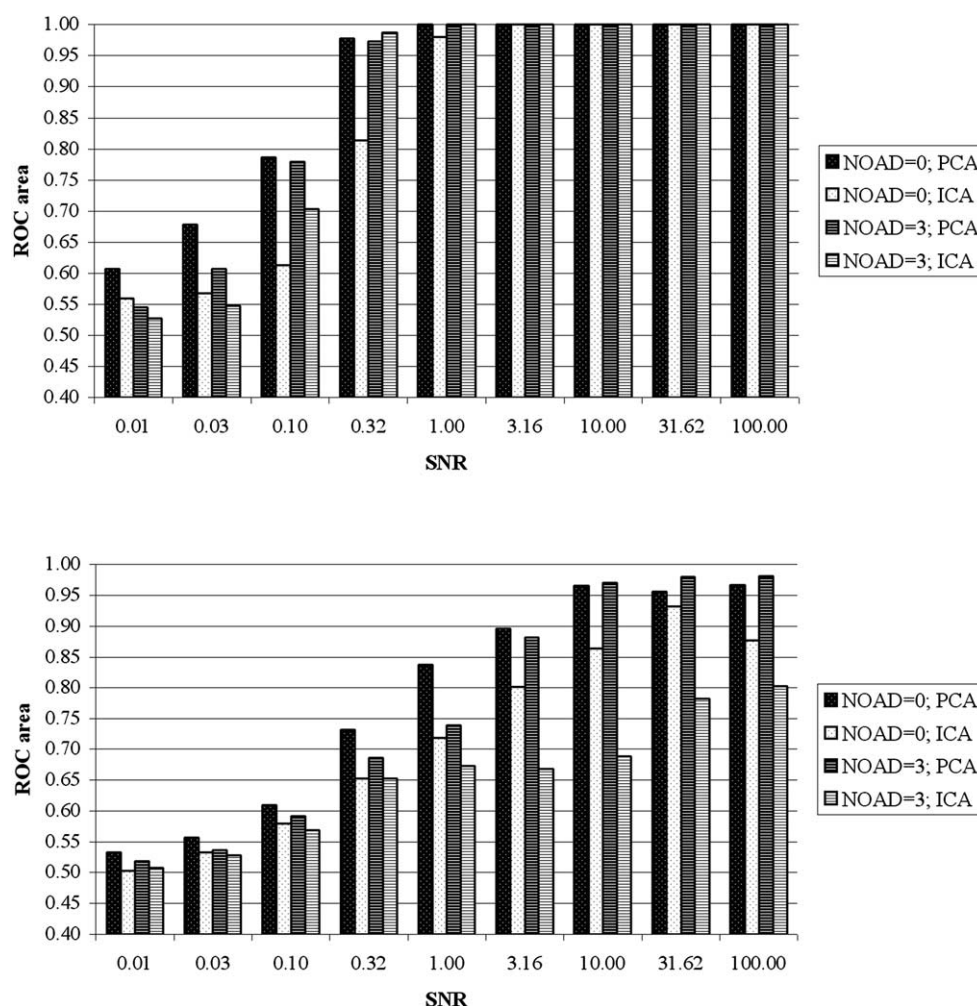
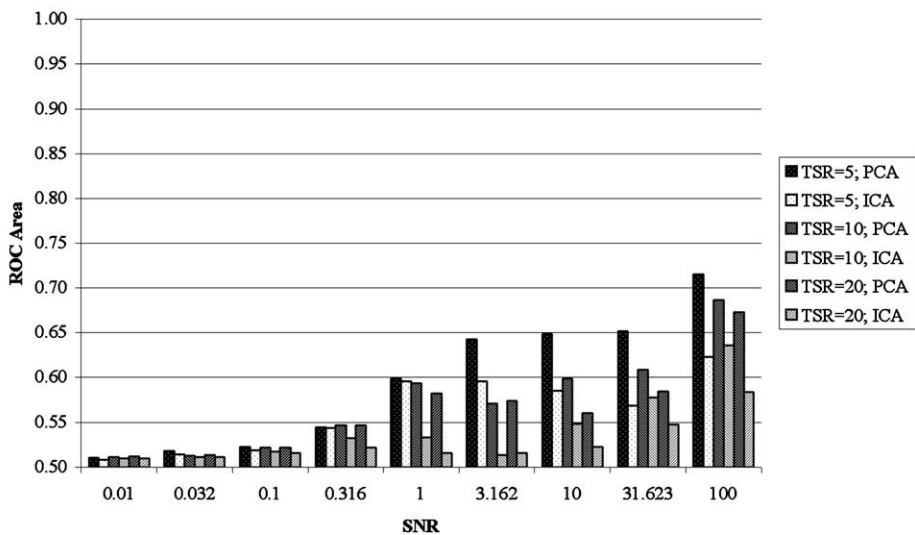
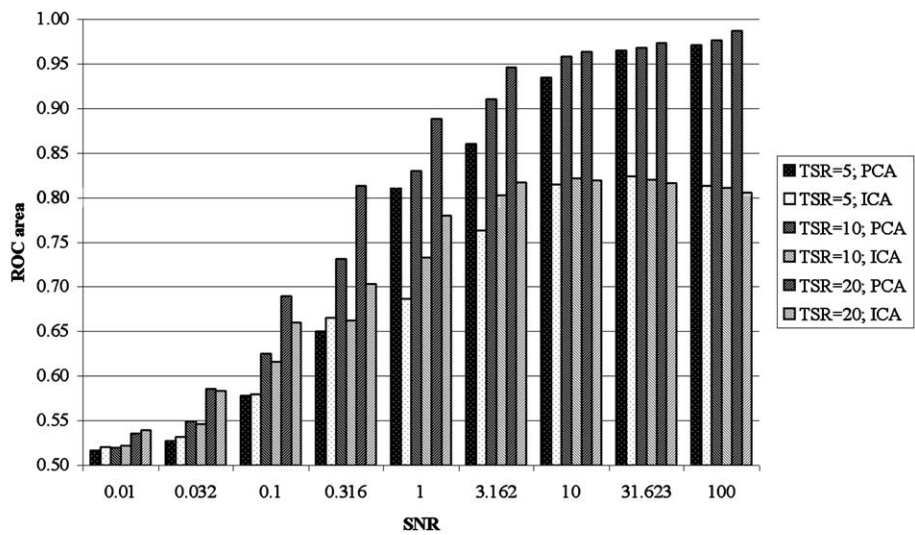


Fig. 5. The most correlated LV by PCA and ICA using gPLS on Gaussian background noise at various SNRs. Single (top panel) or five (bottom panel) orthogonal activation patterns with six epochs and 10 samples in each epoch were simulated. Delays between the experimental paradigm and the onset of the voxel activity varied between 0 to 3 time points (NOAD = 0 and NOAD = 3)

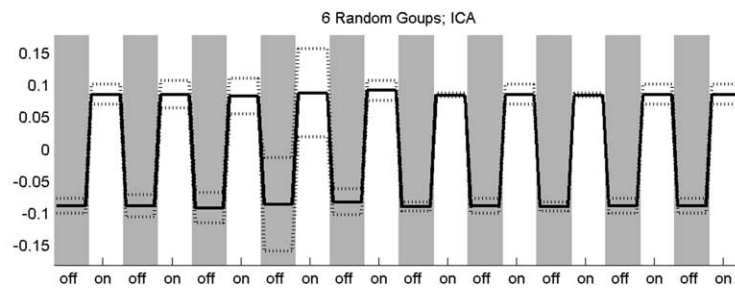
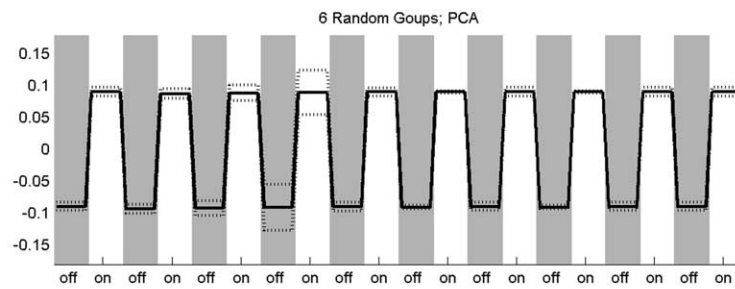
the 10 epochs. In baseline, a crosshair was shown at the center for visual fixation. Multislice echoplanar image (EPI) acquisition was used (1.5-tesla scanner, SIEMENS, Erlangen Germany) (43-ms TE, 4.2-s TR,  $64 \times 64$  matrix, 230-mm FOV, 46 axial slices, 3.6-mm cubic voxels, 128 time points per run). The time series were processed using MEDx (Sensor Systems, Sterling, VA). To correct for within-run interscan head motion, each EPI volume was registered to the mean of its time series using a linear six-parameter rigid-body transformation model employing a least squares cost function. Image volumes were resampled using scanline chirp- $z$  interpolation. Global intensity variations were corrected by global image intensity rescaling, performed by computing the ratio that relates the mean image intensity in a particular volume to an arbitrary value of 1000. Low-frequency temporal signal fluctuations were removed by the application of a high-pass filter with a cutoff of twice the period length. Next, a 3D Gaussian filter (FWHM 6 mm in all dimensions) was applied to each volume in the time series for spatial smoothing. Images

were then spatially normalized with a 3D warp to an EPI template in Talairach space. The full-time series of the subject's tapping at the rate of 1 Hz was used to demonstrate the capability of gPLS to identify task-related spatiotemporal structures. To illustrate the power of multiple comparisons using gPLS, the data of the same subject tapping the left hand at three different frequencies were analyzed to highlight the frequency-dependent activities. All baseline time points and all task time points in each tapping frequency were averaged to generate a single fixation mean and a task mean, respectively, in order to reduce the dimension of the data and to minimize the variability of different epochs.

Two gPLS analyses were separately computed to differentiate the task-control contrast and finger flexion rate-dependent effects. The first gPLS made use of a contrast matrix, including a paradigm coherent vector representing the "on" time points, when the subject tapped his fingers at the rate cued by the visual stimuli, and the "off" time points, when the subject maintained visual fixation only. Addition-



6



8

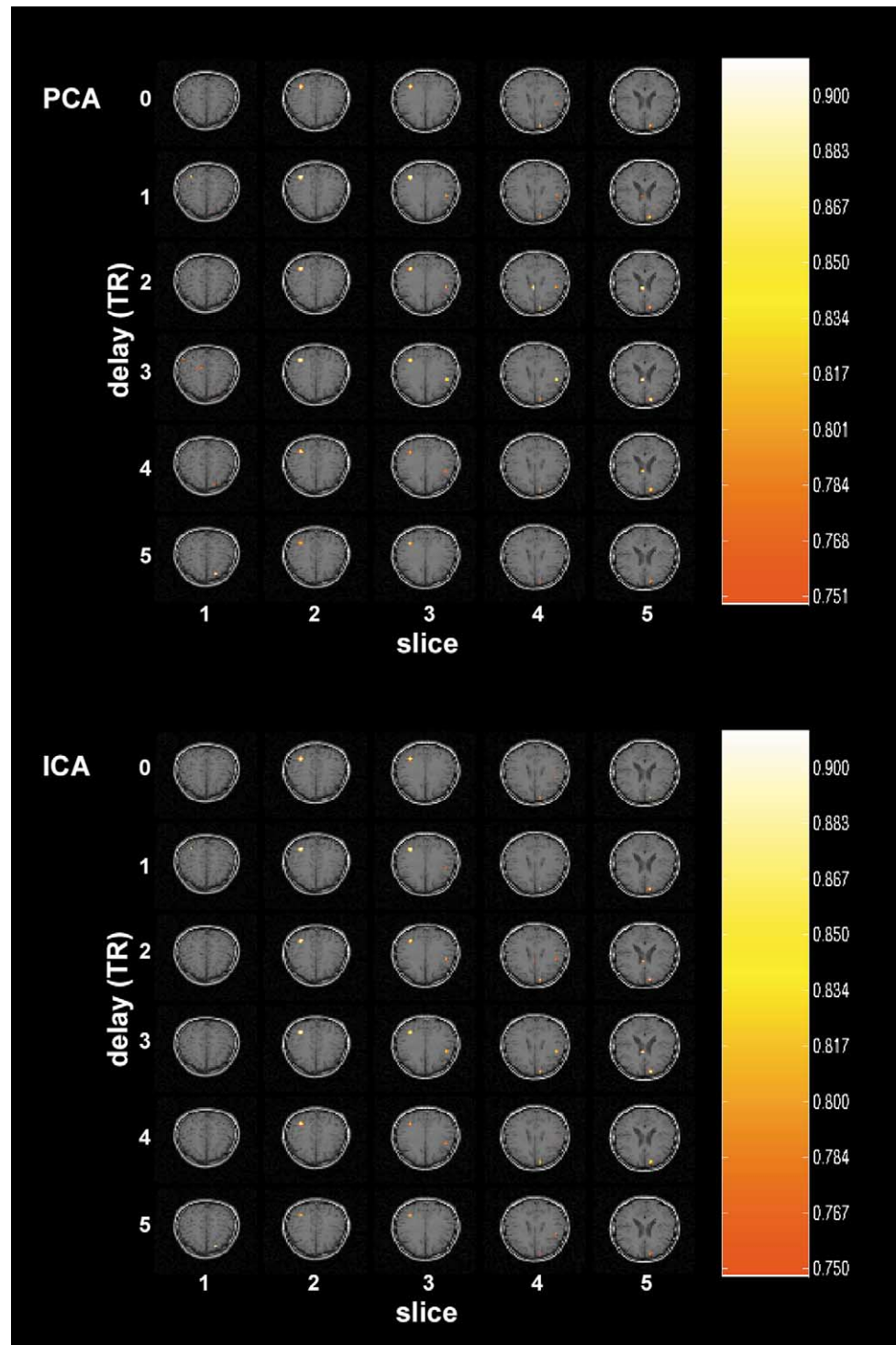


Fig. 7. The first brain LV of PCA (top panel) and ICA (bottom panel) decomposition from an event-related fMRI simulation data.

Fig. 6. The total weighted ROC area in gPLS (top panel) and TD (bottom panel) using PCA and ICA for various temporal sampling rates (TSRs) at different SNR in the Gaussian background with five orthogonal activations.

Fig. 8. The design scores of six random groupings on different epochs of the left-hand tapping data. They reveal the task-related components as well as transient variation in these task-related components. Epoch 4 shows instability in identifying consistently task-related components.

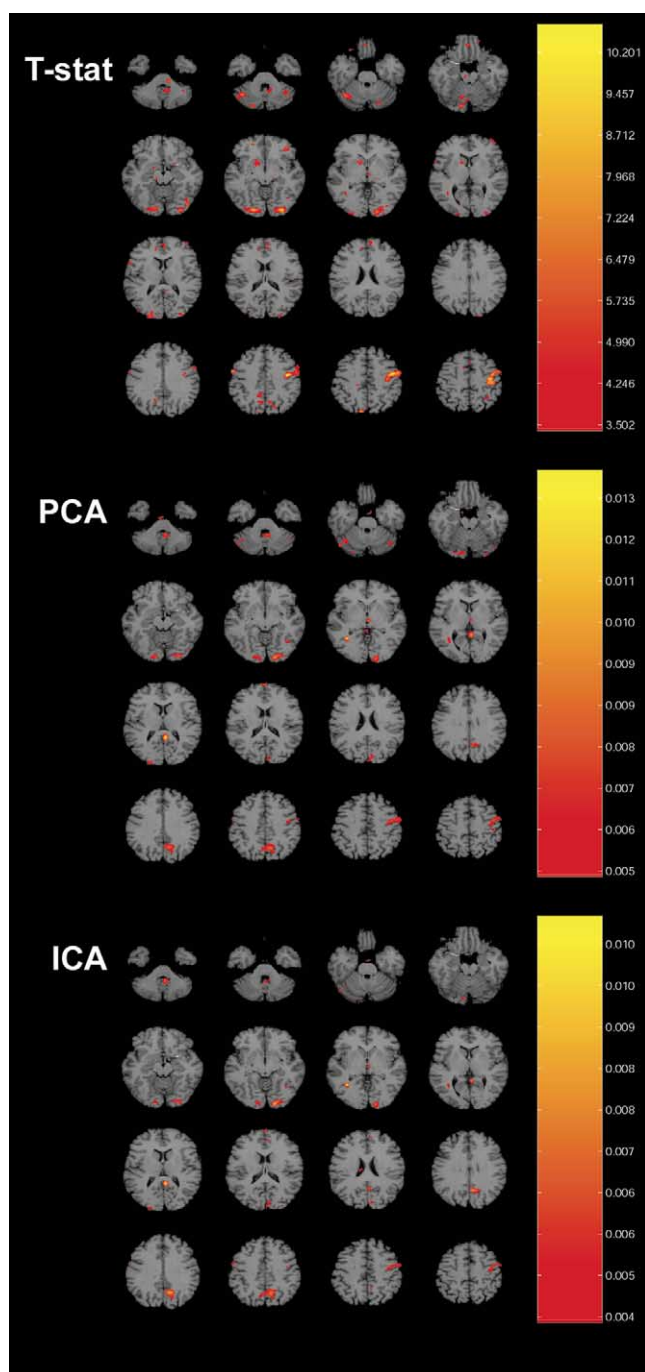


Fig. 9. *t* Test brain activity statistical maps and brain latent variables from gPLS using PCA and ICA decomposition of the effect space with six random groups in the contrast matrix. In these maps, the right hemisphere is on the right.

ally, we manipulated the degree of freedom of the contrast matrix by segregating different epochs of the experiment paradigm into two to six groups to differentiate the epoch-related effects. The second gPLS detected additional finger flexion rate-dependent effects by using Helmert basis to construct a contrast matrix for comparing three finger tapping frequencies. Also, we choose eight randomized groups inside the contrast matrix to test the robustness of the

revealed eigenstructures. Both gPLS analyses used PCA and ICA separately, using 100 iterations to generate error estimates.

## Results

### Simulation studies

#### Signal-to-noise ratio

Figure 2 shows the weighted ROC indices for Gaussian background noise with five orthogonal activations in gPLS. Generally, as the SNR increases, the weighted ROC area from all latent variables using either PCA or ICA increases monotonically. Given the same SNR, PCA decomposition always gives better detection than ICA in the simulation. Note the high variability of the ICA results at low SNR ( $<0.1$ ).

When the SNR is equal to 1, which corresponds to the same level of paradigm-coherent activation signal and background noise, PCA has a weighted ROC area more than 0.82, whereas the weighted ROC area of ICA is only 0.73. The discrepancy between PCA and ICA prevails across various SNRs. When the SNR is higher than 10, PCA always has weighted ROC areas of more than 0.95, and ICA has values of around 0.81. The plateau of ICA detection is about 0.8 when the SNR is greater than 3.16, whereas PCA stabilizes the detection power around 0.95 as the SNR is higher than 10.

#### Most correlated latent variable

The bottom panel of Fig. 2 also reports the ROC area detection metric of the “most correlated” latent variable, whose design score has the highest absolute value of correlation coefficient between the experimental paradigm and the revealed design score. The observation that higher SNRs correspond to higher ROC areas is still valid in the most correlated LV metrics for either PCA or ICA in the gPLS framework. When the SNR is higher than 10, the averaged ROC area by PCA is over 0.95, whereas ICA has variable ROC areas between 0.93 and 0.86. Low SNR decreases the distinction between PCA and ICA decomposition. As the error bars show, there is no significant difference in the ROC area metric from PCA and ICA when the SNR is lower than 0.1. The similar detection power of the single most correlated LV to that of weighted ROC areas from all LVs at different SNR demonstrates that a single LV with the most correlated design score and experimental paradigm is capable of detecting underlying spatiotemporal patterns of activation.

#### Total decomposition versus generalized partial least squares

Figure 3 shows the comparison of the detection using gPLS and TD using Gaussian background noise and five synthetic orthogonal signals. The detection by the weighted

Table 1

The inner product of the first latent variable revealed by PCA and ICA at various numbers of randomized grouping

		PCA (no. of groups)					ICA (no. of groups)				
		2	3	4	5	6	2	3	4	5	6
PCA (groups)	2	0.9992	0.9987	0.9982	0.9976	0.9973	0.8673	0.7589	0.7160	0.6749	0.6547
	3		0.9984	0.9981	0.9978	0.9976	0.8678	0.7600	0.7168	0.6761	0.6557
	4			0.9980	0.9979	0.9977	0.8675	0.7600	0.7163	0.6761	0.6547
	5				0.9980	0.9979	0.8675	0.7605	0.7166	0.6768	0.6553
	6					0.9979	0.8674	0.7602	0.7159	0.6756	0.6535
ICA (groups)	2						0.7736	0.6812	0.6482	0.6103	0.5984
	3							0.6085	0.5826	0.5568	0.5493
	4								0.5663	0.5413	0.5390
	5									0.5404	0.5467
	6										0.5774

Note. PCA shows consistent similarity at different randomized group numbers. ICA decomposition varies more greatly than PCA with a significant difference between two-group and six-group random grouping. Between PCA and ICA, the similarity is low and becomes lower when the number of random grouping increases.

ROC area in TD is lower than that in gPLS at SNRs ranging from 0.1 to 100. The gPLS approach performs better than TD when using either PCA and ICA decompositions. Presumably this advantage comes from the selective averaging step. When the SNR equals 1, TD using PCA has a maximally weighted ROC area of 0.58, whereas TD using ICA is 0.54. gPLS at the same SNR ( $\text{SNR} = 1$ ) has the weighted ROC area of 0.82 and 0.73 by PCA and ICA, respectively. This comparison quantitatively shows the advantage of gPLS over TD for higher signal detection. For the ROC area of the most correlated LV (bottom panel of Fig. 3), PCA is of similar detection (ROC area over 0.95) using either gPLS

or TD at high SNR ( $>10$ ). ICA in gPLS demonstrates higher detection than in TD by a larger ROC area at high SNR. When SNR is lower than 0.03, gPLS or TD provides almost the same insufficient detection, no matter whether PCA or ICA is utilized. In general, the variability of the reported ROC metrics from the most correlated LV is smaller in PCA than ICA when SNR is higher than 1.

#### Background noise

In addition to the Gaussian distribution, super-Gaussian and sub-Gaussian probability distribution functions were used to simulate background noises, as shown in Fig. 4. With super-

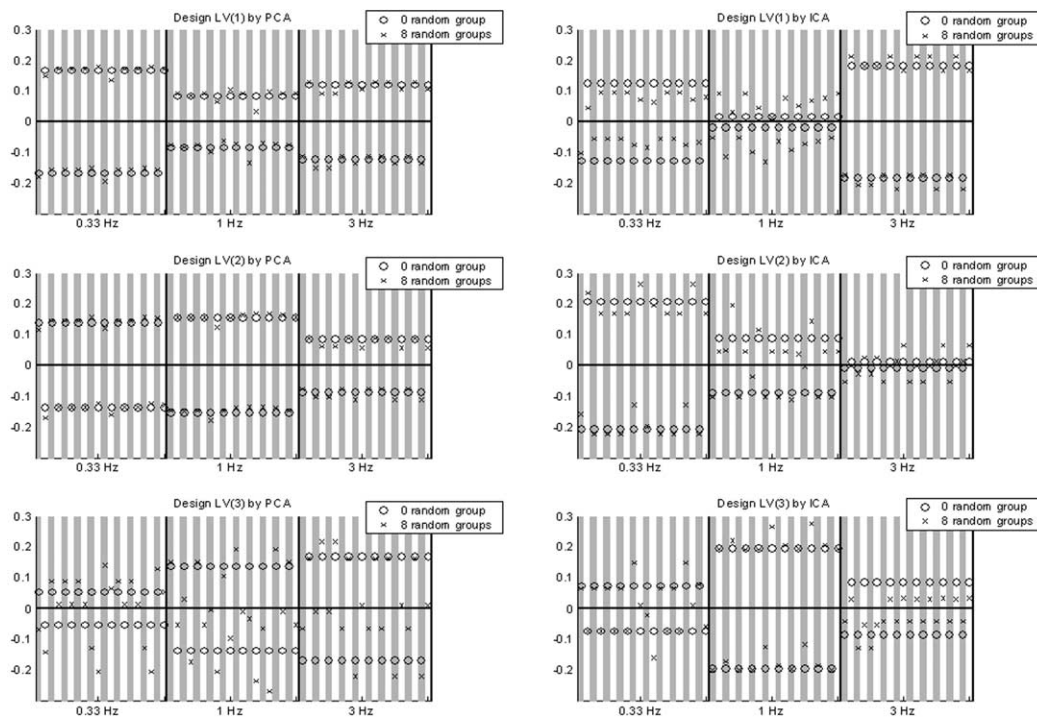


Fig. 10. Design scores of the first three latent variables in gPLS using PCA and ICA for multiple comparisons, including three different tapping frequencies and task-control contrast. The first three LVs account for the majority of the total variance in the effect space. And they represent the task-control contrast and tapping frequency-dependent brain maps.

Gaussian background noise and using the most correlated latent variable metric, TD is slightly better than gPLS at high SNR regions ( $\text{SNR} > 10$ ) when PCA is employed. In the lower SNR ( $\text{SNR} < 3$ ) contexts, gPLS is superior to TD using PCA. When ICA is adopted as the decomposition tool, gPLS always outperforms TD at all SNRs. The sub-Gaussian background noise study (bottom panel of Fig. 4) shows the higher detection using gPLS over TD and PCA over ICA when the SNR is lower than 1. With the sub-Gaussian background noise, the weighted ROC area metrics at SNR larger than 1 are slightly higher than the most correlated LV ROC index (not shown). This suggests that when the background noise follows a sub-Gaussian distribution, such as uniform distribution, the ensemble of multiple latent variables has higher detection power than the single most correlated LV.

#### *Effect of multiple contrasts and hemodynamic delays*

The gPLS structure can be used for either multiple hypotheses/activations comparison/detection as well as a single contrast identification, as illustrated in Fig. 5. In this Gaussian background test bed, when there is only one paradigm-related activation signal, detection using only the most paradigm correlated latent variable is quite efficient ( $>0.99$ ) when the SNR is greater than 1. PCA outperforms ICA when the SNR is less than 1, except that ICA detects slightly better when SNR equals 0.3 and the delayed activations exist ( $\text{NOAD} = 3 \text{ TR}$ ). Simulation results show that when the number of contrasts is low, a single latent variable is adequate to examine the spatiotemporal structures associated with the posed hull hypothesis encoded in the contrast matrix. Multiple contrasts detection by gPLS (bottom panel of Fig. 5) is illustrated using five orthogonal activations and 0 to 3 TR delayed signals in the data (to model hemodynamic delays) at various SNRs. PCA has better detection than ICA for all SNR and time delays.

The hemodynamic responses in the realistic fMRI might contain voxels with different delays. Thus, we parametrically simulated the delayed activations between 0 and 3 time points, which are 0 to 3 TR delays in fMRI acquisition for the epoch length of 10 TR. If the data contain only a single paradigm-coherent component, such delays decrease the detection by PCA in low SNR contexts, but it enhances ICA detection at two simulation SNRs ( $\text{SNR} = 0.1$  and  $\text{SNR} = 0.32$ ). However, PCA provides higher detection power than ICA with or without hemodynamic delays in all conditions. In data containing multiple hypothesis-related signals, the delays hinder the PCA detection in low SNR regions ( $\text{SNR} < 1$ ), but they do not change significantly when SNR is greater than 10. ICA degrades the detection significantly when there are multiple delayed activations in higher SNR contexts ( $\text{SNR} > 3$ ). It is worth noticing that ICA has lower detection when delays exist in the multiple contrasts data. With sufficient SNR, gPLS using PCA is insensitive to the hemodynamic delays in the data, providing a higher detection power compared to ICA decomposition.

#### *Temporal sampling rate*

Various temporal sampling rates (TSR) were simulated for 5, 10, and 20 scans for each epoch in a blocked-design experiment. Results suggest that a higher TSR always has a higher ROC area metric in PCA decomposition at different SNRs, as shown in Fig. 6. ICA has a similar tendency to have higher detection power as the TSR increases. When the SNR is higher than 3, more temporal samplings produce relatively fewer ROC increments. Increasing the TSR is only advantageous for gPLS, not for TD. The increased sampling does not enrich the content of data, and both PCA and ICA suffer from the increased degrees of freedom in the decomposition, as shown in the bottom panel of Fig. 6.

#### *Event-related fMRI simulation*

Figure 7 shows the most significant 25% of the first brain latent variables in an event-related fMRI experiment when PCA and ICA were used to decompose the PLS effect space. Qualitatively, there was no appreciable difference in the results using PCA vs ICA decomposition. Both show similar within-voxel activation duration and the location of the activation loci. This latent variable differentiated the two conditions with the added activations from each other and the “baseline” condition.

#### *fMRI motor system study*

##### *Movement contrast detection*

We found from the singular values that the first LV using PCA and ICA accounts for 90 and 81%, respectively, of the total effect space variance. The variation of the singular value from ICA (8.5%) is higher compared to the variation from PCA (0.5%) in the 100 iterative analyses.

Figure 8 shows the design scores of the most dominant (the first) latent variable in gPLS using PCA and ICA. The corresponding LVs are well correlated to the on/off experiment paradigm. Either PCA or ICA decomposition can robustly detect the contrast between the baseline and task conditions in the gPLS framework. Multiple (100) iterations were used to investigate the variability of the design scores in different epochs. Comparing PCA and ICA, the design scores show that gPLS is able to detect the consistently task-related structures inside the data matrix, as well as the transient responses. The fourth epoch of gPLS by both PCA and ICA using six randomized groups demonstrated higher variability, representing lower confidence during task-related component identification.

The spatial patterns of the first latent variable are shown in Fig. 9. For comparison,  $t$  test uncorrected  $P$  values are also shown to contrast the detection by univariate and multivariate approaches. Thresholds are set to 3.5 for base 10 logarithm  $P$  values for  $t$  test and 40% of latent variable maximum for gPLS. These values are chosen for optimal visual comparison between approaches. The cerebellum, visual cortex, primary motor area, cingulate gyrus, and the medial superior parietal lobule are shown activated in PCA



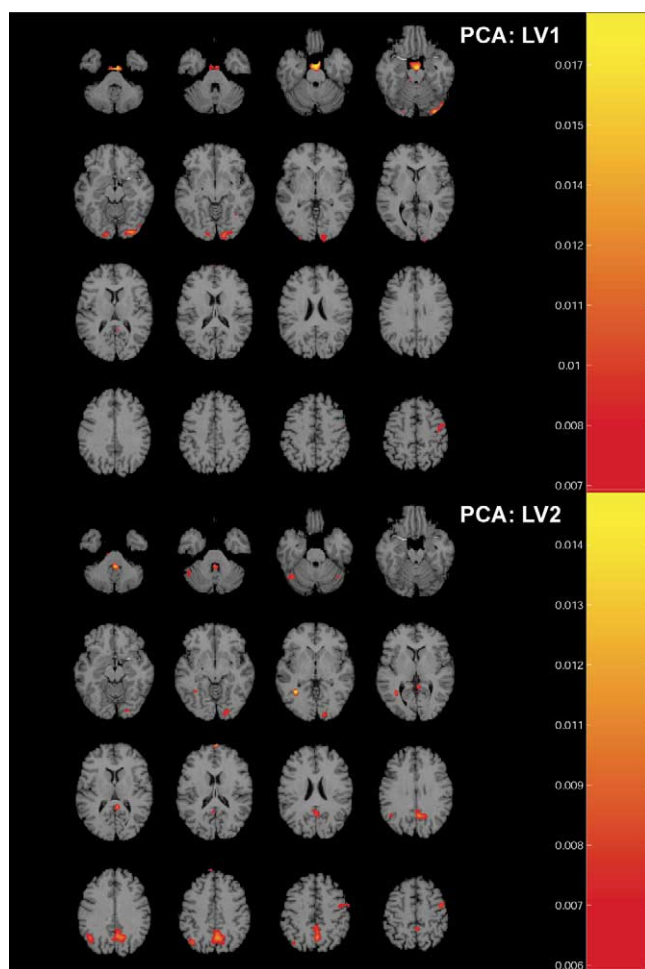


Fig. 11. The first brain latent variable from PCA decomposition of the effect space, representing the areas that are more active during task than control conditions across all three tapping frequencies. The second brain latent variable from PCA (similar to the third LV from ICA) highlights the areas that are more active during high-frequency tapping (weighted average of 3 and 1 Hz) than low-frequency tapping (0.1 Hz).

and ICA decomposition using six random groups and the experiment paradigm to constitute the contrast matrix. A large amount of similarity and overlapping between PCA and ICA decompositions is observed. The differences between the  $t$  test statistical map and gPLS brain LVs are most significant in spatial extension and distribution of estimated active areas in cerebellum, internal capsule, and superior parietal lobule.

Different random groupings reveal similar significant latent variables in gPLS. Table 1 lists the angles between the most dominant brain LVs as multidimensional vectors. The inner product of two high-dimensional brain LVs quantifies the alignment and the similarity between them. PCA random grouping is more robust than ICA for the first LV because of the high inner product ( $>0.999$ ). Brain LVs from ICA vary at different numbers of randomized grouping with a maximal similarity of inner product of 0.77, which is analogous to 39 degrees in two-dimensional space. In contrast, 0.992 inner product is equivalent to 2.9 degrees in two-dimensional space.

Using total decomposition, PCA and ICA generate complex time courses for different latent variables. The maximum of the correlation coefficient between the design score and the experiment paradigm is 0.5204 for PCA and 0.3441 for ICA. The associated spatial patterns are complicated by ventricle false alarms, distributed local activities, and motion artifacts (not shown) and therefore further processing, for example, canonical variance analysis (Fletcher et al., 1996), is necessary to make use of these eigenstructures.

#### *Multiple comparisons between finger tapping rates*

The design score plots (Fig. 10) for the multiple comparisons of different tapping frequencies illustrate identification of both task-control contrast and tapping frequency contrast. The distance between the individual mark and zero is proportional to the contribution of that observation to the contrast revealed by the latent variable. Note that a design score may represent an interaction effects. For example, a design score with “on/off” contrast (“on” positive and “off” negative) and “off/on” contrast (“off” positive and “on” negative) at different frequencies, characterizes the differential “on/off” contrasts due to distinct motor frequencies.

To contrast the “on/off,” the first LV revealed by both PCA and ICA using either no random groups or eight random groups identifies the contrast between task and control states across three tapping frequencies. The second LV of PCA decomposition represents the interaction of the generalized high tapping rate versus low tapping rate contrast (weighted average of 3 and 1 Hz versus 0.3 Hz) and the task-control contrast. Similar results from ICA decompositions are also present but appeared in the third LV. Interestingly, the second LV from ICA also represents a contrast between task and control, which is consistent with the inference from the first LV. Note that the tapping frequency contrast is the interaction/modulation of task-control differences over blocks of different tapping frequencies.

The first three LVs explain all variances in the effect space when no random group is applied. And they account for 76.5 and 58.3% of total variance in eight random grouping gPLS using PCA and ICA, respectively. Random grouping detects the transient responses and variations of different epochs in three tapping frequencies, as shown in Fig. 11.

Examining spatial patterns of the brain LVs excludes ventricle activations in the third LV in PCA and second LV in ICA (not shown). The first LVs using PCA and ICA demonstrate the areas that are more active during task than control conditions regardless of finger tapping rates. These include the visual cortex, cerebellum hemispheres, and contralateral primary motor cortex. The contrast between the higher tapping rate and the lower rates revealed by the second LV in PCA and the third LV in ICA in gPLS suggests that the cerebellum hemisphere, visual cortex, and medial superior parietal lobule are more activated at high-frequency finger tapping. Because areas indicated by LVs with PCA and ICA are visually similar, only PCA results are illustrated.

## Discussion

Univariate approaches, such as correlation coefficient analysis and statistical parametric mapping (Bandettini et al., 1993; Friston et al., 1995b, 1995c; Worsley and Friston, 1995), are model-driven. Reference functions are provided to build models for the observed neuroimaging data. Because multiple null hypotheses have been explicitly constructed inside the model, identification of components related to the questions of interest is straightforward. However, the segregation of image voxels during estimation is problematic because these methods ignore possible anatomically interconnected neurons and associated activation. It is also possible that voxels might correlate quite well to the reference function, but they are not significantly correlated with each other.

We have presented the theory and algorithms to use multivariate data analysis technique to explore brain mapping based on functional MRI data. Compared to univariate approaches, multivariate techniques facilitate the identification of interaction among neuronal populations in the data analysis. No additional clustering or corrected statistics for the estimated active regions are necessary. The multivariate approach has been verified and adopted for functional connectivity analysis to investigate temporal interactions among distributed regions in the brain during cognitive tasks, such as the orchestration of human memory and learning systems. A comparison study of the memory system even supports the assertion that more information is revealed by multivariate analysis than by univariate processing streams (Fletcher et al., 1996).

The conventional multivariate technique is a data-driven model subjected to different mathematical constraints. The data-driven characteristics make the identification of components of interest in the revealed structures difficult. This is especially discussed in independent component analysis (ICA) because of the high variability of the correlation coefficient between the provided reference function and the temporal independent component, which is equivalent to the design score in the generalized partial least squares (gPLS) formulation. The proposed gPLS framework moderates the difficulty of identifying task-related components by selective averaging to increase the SNR in the data. gPLS is an extension of partial least squares in the sense of identical utilization of the contrast matrix. The rank-reduced effect space in gPLS has advantages of both increased SNR as well as decreased computation time. The higher power of detection in low SNR data derives from the construction of an effect space by explicit incorporation of columns of the contrast matrix as averaging factors. The subsequent PCA/ICA decomposition allows the data itself to reorganize into separate components under distinct mathematical constraints, which are either orthogonality constraint in PCA or statistical independence in ICA. In gPLS, the size of effect space is dramatically smaller than the original data matrix. Dimension reduction in gPLS facilitates efficient calculation

for latent variables. Without gPLS, the size of the matrix is huge for a multiple-subject multiple-task full-time series analysis. In our simulation, even with the improved fast algorithm (Hyvarinen, 1999), ICA is much slower than PCA in the same computational environment. The difference in computational loading between PCA and ICA is manyfold (PCA is about 10 to 30 s depending on the size of the matrix; ICA takes more than 2 to 15 min for the data of the same size). Thus, generalized partial least squares has the combined advantages of being model-driven from univariate statistical procedures by the essence of contrast matrix correlation and data-driven characteristics via PCA/ICA decomposition to explore the interactions inside these data in an integrated way.

Another feature of gPLS is efficient multiple comparisons, which have been reported in studies of the memory system (McIntosh et al., 1996) and interactions of sensory systems (McIntosh et al., 1999, 1998). In our simulations and realistic fMRI data, we observe that gPLS is capable of detecting effects of different conditions. This is essentially accomplished by providing different “contrast basis functions” in the contrast matrix. Subsequent PCA/ICA application is then used to estimate the relative contributions of these contrast bases. Without gPLS, multiple subjects and multiple conditions comparison is difficult, because of the large data size and unconstrained decomposition. gPLS is thus a convenient tool for experiments with parametric designs and those reporting multiple-subjects commonality.

Yet another advantage of gPLS is the simplified identification of latent variables. Because the data are constructed within hypotheses of interest, the revealed design latent variable explicitly gives the weighting factors of different hypotheses under the test. The associated design scores immediately show the temporal representation of the corresponding structure. A simple correlation coefficient can be applied to categorize latent variables as either task-related or task-independent as shown in our simulation studies. The associated singular values quantify the significance of latent variables in terms of variance partitioning. Conventional total decomposition interpretation of the revealed latent variables suffers because the correlation coefficient between the reference function and LV associated temporal components is highly variable and small. A final complication of using total decomposition comes from the need to do another canonical variate analysis following the first multivariate decomposition.

In a full-rank effect space, the number of latent variables (or principal components in PCA and independent components in ICA) is determined by the minimum of the dimension of the effect space. This fixed-dimension property poses another constraint in addition to the mathematical requirements of the decomposition algorithm. The fixed number of decomposed components may consume the degrees of freedom by partitioning total variance in the data into either hypothesis-related components or confounds. Thus, the separation of confounds from pertinent contrasts



is essential in analysis of functional imaging data, regardless of using either total decomposition or generalized partial least squares. However, this problem can be partially alleviated by varying the number of groups, which collapses the nonparadigm related data components together. For example, we dealt with the single subject blocked-design fMRI experiment by randomly partitioning different epochs in the data into various numbers of groups. The flexible manipulation of randomized grouping in the construction of the contrast matrix enables the identification of not only constantly task-related responses, but also transiently varying responses, as shown in Figs. 7 and 9. This randomization releases the constraint on the fixed number of components by either PCA or ICA. Various numbers of random groups enable another statistical inference about the robustness of the detected neural activity structures. This randomization can be combined with the bootstrap and permutation tests in the partial least squares approach to provide further confidence in the estimation of the revealed neuronal interaction.

Between PCA and ICA, it has been claimed that ICA is superior for task-related components, especially transient responses, and noise detection. And it is claimed that ICA identifies components following a super-Gaussian distribution more efficiently than PCA. Recently, it has been reported that the BOLD fMRI follows a super-Gaussian distribution (Hanson and Bly, 2001), which suggests an advantage of ICA over PCA. But in the gPLS framework, the selective averaging implemented by the contrast matrix over the raw data compromises such claims; because the Central Limit Theorem suggests that the mixture of independently identically distributed components appears to be Gaussian. Although the number of averaged samples might not be large enough to be Gaussian for all cases, and the observations of fMRI time series are not temporally independent, gPLS still generates a sufficient amount of averaging to increase detection power. This contributes to the better performance of PCA over ICA in our generalized partial least squares framework. Such a statement is valid for blocked-design experiments and has been verified in our simulations of both blocked-design and event-related fMRI. From our fMRI data, it is evident that PCA is more robust for the estimation. The revealed most-significant latent variables at different numbers of random grouping are very similar (inner product  $>0.99$ ). Our simulation also shows a smaller standard deviation in PCA compared to ICA. ICA is sensitive to the number of components to be decomposed, which is equal to the minimum of the number of rows and columns of the matrix. The observed high variability between the most significant LVs at different numbers of random grouping decreases the confidence of estimated spatial patterns derived by ICA.

Realistic hemodynamic responses of voxels have various waveforms, including different delays and shapes, as identified by previous studies (Kruggel and von Cramon, 1999;

Rajapakse et al., 1998). Versatile waveforms deviating from the boxcar reference paradigm in blocked-design fMRI experiments have been modeled at different SNR in generalized partial least squares. Different onset delays have also been simulated by assuming different numbers of NOADs. We simulated that delays may be up to 3 TR in our test bed of 10 TR, which is equivalent to 30% temporal incoherence. The results suggest that in most SNR contexts, a single LV from PCA has better detection power than one from ICA in the gPLS framework. In addition, the most correlated LV might be quite sufficient to detect task-related components, which include transient variations if random grouping is employed. Multiple-subjects multiple-conditions comparison favors PCA over ICA.

The general conclusion that in the PLS framework, PCA and ICA are comparable to reveal the significant latent variables is also validated in our event-related fMRI simulations. This is entirely consistent with our findings in the block design simulation for the similarity of PCA and ICA detection power in the PLS framework. We suspect that part of the reason for the agreement comes from the constraint of the solution space to those spanned by task differences rather than all dimensions in the data.

To evaluate the power of hypothesis testing under the Neyman–Pearson framework (Casella and Berger, 2002), we used receive operation curves (ROC) to assess the trade-off between Type I and Type II error. ROC has been adopted for assessment of fMRI signal detection (Constable et al., 1995; Skudlarski et al., 1999). The ROC area and weighted ROC area for all LVs in our simulations are neither unique nor globally optimal methods for assessing the power of detection. However, singular values are good metrics for quantifying the amount of variance explained in the model. And, ROC areas correspond conveniently to the trade-off between Type I and Type II errors. However, if people are more concerned about one hypothesis testing error than the other, other metrics should be employed.

Here, we used a linear decomposition approach to reveal the structures in the data. Nonlinear decomposition can also be used in the gPLS framework as well. PCA has been extended for identifying nonlinear interactions in brain systems. The same approach could be applied to the nonlinear identification of different functional areas when it is applied on the effect space in the future.

In summary, adoption of the multivariate analysis tool depends on the spatiotemporal structure of the data and the experimental questions. We show gPLS provides computational efficiency and flexibility for testing hypotheses at different levels. And gPLS can be used for either single or multiple hypotheses testing by two alternative decompositions: PCA and ICA. Although PCA seems to outperform ICA in several of our scenarios, a conservative conclusion would suggest that neither can be favored when applied in the PLS framework.

## Acknowledgments

We thank Dr. Wilkin Chau for providing event-related fMRI simulation data set and Dr. Dave S. Tuch's comments on the article. We thank the funding supports from National Institutes of Health (RO1 NS37462, RO1 HD/DC40712, and NCRR P41 RR14075), Mental Illness and Neuroscience Discovery (MIND) Institute, and MGH-MIT-HMS Athinoula A. Martinos Center for Biomedical Imaging. We also thank NRSA grant funding for data collection and the National Institutes of Health for their support through Grants MH12584 and HD40095.

## References

- Bandettini, P.A., Jesmanowicz, A., Wong, E.C., Hyde, J.S., 1993. Processing strategies for time-course data sets in functional MRI of the human brain. *Magn. Reson. Med.* 30, 161–173.
- Bell, A.J., Sejnowski, T.J., 1995. An information-maximization approach to blind separation and blind deconvolution. *Neural Comput.* 7, 1129–1159.
- Belliveau, J.W., Rosen, B.R., Kantor, H.L., Rzedzian, R.R., Kennedy, D.N., McKinstry, R.C., Vevea, J.M., Cohen, M.S., Pykett, I.L., Brady, T.J., 1990. Functional cerebral imaging by susceptibility-contrast NMR. *Magn. Reson. Med.* 14, 538–546.
- Belliveau, J.W., Kennedy Jr., D.N., McKinstry, R.C., Buchbinder, B.R., Weisskoff, R.M., Cohen, M.S., Vevea, J.M., Brady, T.J., Rosen, B.R., 1991. Functional mapping of the human visual cortex by magnetic resonance imaging. *Science* 254, 716–719.
- Bullmore, E.T., Rabe-Hesketh, S., Morris, R.G., Williams, S.C., Gregory, L., Gray, J.A., Brammer, M.J., 1996. Functional magnetic resonance image analysis of a large-scale neurocognitive network. *NeuroImage* 4, 16–33.
- Cohen, M.S., 1997. Parametric analysis of fMRI data using linear systems methods. *NeuroImage* 6, 93–103.
- Constable, R.T., Skudlarski, P., Gore, J.C., 1995. An ROC approach for evaluating functional brain MR imaging and postprocessing protocols. *Magn. Reson. Med.* 34, 57–64.
- Della-Maggiore, V., Chau, W., Peres-Neto, P.R., McIntosh, A.R., 2002. An empirical comparison of SPM preprocessing parameters to the analysis of fMRI data. *NeuroImage* 17, 19–28.
- Fletcher, P.C., Dolan, R.J., Shallice, T., Frith, C.D., Frackowiak, R.S., Friston, K.J., 1996. Is multivariate analysis of PET data more revealing than the univariate approach? Evidence from a study of episodic memory retrieval. *NeuroImage* 3, 209–215.
- Friston, K.J., Frith, C.D., Liddle, P.F., Frackowiak, R.S., 1993. Functional connectivity: the principal-component analysis of large (PET) data sets. *J. Cereb. Blood Flow Metab.* 13, 5–14.
- Friston, K.J., Frith, C.D., Frackowiak, R.S., Turner, R., 1995a. Characterizing dynamic brain responses with fMRI: a multivariate approach. *NeuroImage* 2, 166–172.
- Friston, K.J., Frith, C.D., Turner, R., Frackowiak, R.S., 1995b. Characterizing evoked hemodynamics with fMRI. *NeuroImage* 2, 157–165.
- Friston, K.J., Holmes, A.P., Poline, J.B., Grasby, P.J., Williams, S.C., Frackowiak, R.S., Turner, R., 1995c. Analysis of fMRI time-series revisited. *NeuroImage* 2, 45–53.
- Hanson, S.J., Bly, B.M., 2001. The distribution of BOLD susceptibility effects in the brain is non-Gaussian. *NeuroReport* 12, 1971–1977.
- Hyvärinen, A., 1999. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transact. Neural Networks* 10, 626–634.
- Kruggel, F., von Cramon, D.Y., 1999. Modeling the hemodynamic response in single-trial functional MRI experiments. *Magn. Reson. Med.* 42, 787–797.
- Kwong, K.K., Belliveau, J.W., Chesler, D.A., Goldberg, I.E., Weisskoff, R.M., Poncelet, B.P., Kennedy, D.N., Hoppel, B.E., Cohen, M.S., Turner, R., et al., 1992. Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. *Proc. Natl. Acad. Sci. USA* 89, 5675–5679.
- Lobaugh, N.J., West, R., McIntosh, A.R., 2001. Spatiotemporal analysis of experimental differences in event-related potential data with partial least squares. *Psychophysiology* 38, 517–530.
- McIntosh, A.R., Bookstein, F.L., Haxby, J.V., Grady, C.L., 1996. Spatial pattern analysis of functional brain images using partial least squares. *NeuroImage* 3, 143–157.
- McKeown, M.J., Makeig, S., Brown, G.G., Jung, T.P., Kindermann, S.S., Bell, A.J., Sejnowski, T.J., 1998. Analysis of fMRI data by blind separation into independent spatial components. *Hum. Brain Mapp.* 6, 160–188.
- Ogawa, S., Lee, T.M., Kay, A.R., Tank, D.W., 1990. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proc. Natl. Acad. Sci. USA* 87, 9868–9872.
- Purdon, P.L., Weisskoff, R.M., 1998. Effect of temporal autocorrelation due to physiological noise and stimulus paradigm on voxel-level false-positive rates in fMRI. *Hum. Brain Mapp.* 6, 239–249.
- Rajapakse, J.C., Kruggel, F., Maisog, J.M., von Cramon, D.Y., 1998. Modeling hemodynamic response for analysis of functional MRI time-series. *Hum. Brain Mapp.* 6, 283–300.
- Skudlarski, P., Constable, R.T., Gore, J.C., 1999. ROC analysis of statistical methods used in functional MRI: individual subjects. *NeuroImage* 9, 311–329.
- Worsley, K.J., Friston, K.J., 1995. Analysis of fMRI time-series revisited—again. *NeuroImage* 2, 173–181.