



Published in final edited form as:

IEEE Signal Process Mag. 2010 ; 27(4): 39–50. doi:10.1109/MSP.2010.936725.

Canonical Correlation Analysis for Data Fusion and Group

Inferences:

Examining applications of medical imaging data

Nicolle M. Correa, Tülay Adalı, Yi-Ou Li, and Vince D. Calhoun

Data-driven analysis methods, such as blind source separation (BSS) based on independent component analysis (ICA), have proven very useful in the study of brain function, in particular when the dynamics are hard to model and underlying assumptions about the data have to be minimized. Many problems in medical data analysis involve the analysis of multiple data sets, either of the same type as in a group study where inferences are based on the same modality, e.g., group inferences from functional magnetic resonance imaging (fMRI) data collected from multiple subjects, or from different modalities as in the case of data fusion where inferences have to be drawn from data collected from multiple modalities such as fMRI, electroencephalography (EEG), and structural MRI (sMRI), for the same group of subjects. Canonical correlation analysis (CCA) [1], another data-driven approach, and its extension to multiple data sets—multiset CCA (M-CCA) [2]—provide a natural framework for both types of study. In this article, we show how CCA and M-CCA can be used for the analysis of data from a single modality for group inferences as well as fusion of data from multiple modalities using a feature-based approach, discuss the advantages of the CCA-based approach, and compare its performance to ICA that has been successfully applied to both types of study.

BACKGROUND

Analysis of multiple sets of data, either of the same type as in multitask or multisubject data, or of different type or nature as in multimodality data, is inherent to many fields and is a particularly challenging problem in biomedical image analysis because of the rich nature of the data made available by different imaging modalities. Analysis of multiple sets of same type of data is an integral part of biomedical imaging studies, e.g., when estimating brain activations in fMRI data from a group of subjects or when analyzing data from two different experimental conditions such as fMRI data from subjects scanned at different alcohol levels while performing a given task. Fusion of data from different modalities promises to provide a better understanding of the problem at hand since each modality has its own advantages as well as limitations. In the case of biomedical imaging, an increasing number of studies are collecting multiple measurements, e.g., fMRI data, sMRI data, EEG data, genetic data, and

©2010IEEE

CALLOUTS

Many problems in medical data analysis involve the analysis of multiple data sets, either of the same type as in a group study where inferences are based on the same modality.

The main purpose of analyzing multiple modalities is to utilize the common as well as unique information from complementary modalities to better understand neuronal activity.

M-CCA provides an effective tool to perform group BSS while maintaining the correspondence of the source estimates across different data sets and retaining the intersubject source variability.

The analysis of more than two brain imaging modalities collectively, e.g., fMRI, sMRI, and EEG, can help identify interesting associations across brain structure and function.

CCA- and M-CCA-based methods provide attractive solutions to data fusion and group analysis.

others from the same participants. Efficient use of all this information for inference, while minimizing assumptions made about the underlying nature of the data and relationships, is an arduous task but is one that promises significant gains in understanding of the human brain function. The main purpose of analyzing multiple modalities is to utilize the common as well as unique information from complementary modalities to better understand neuronal activity. For example, the fusion of fMRI data and EEG data—fMRI having good spatial resolution and EEG having high temporal resolution but poor spatial localization of brain activity—provides a better spatio-temporal mapping of the brain function. In this article, we address both types of problems: fusion of data sets collected from multiple modalities and the analysis of multisubject data from the same modality.

Approaches to solve these multidata set problems can be broadly classified as being either model based or data driven. Model-based approaches investigate the goodness-of-fit of the data to the prior knowledge about the experimental paradigm and the properties of the data, for example the general linear model approach [3] for the analysis of fMRI data utilizes the prior knowledge of the hemodynamic properties of the data and the task. While model-based approaches have been extensively used in biomedical data analysis, their use is limited when the dynamics of the experiment become hard to model, e.g., when studying rest state or naturalistic paradigms such as driving or watching a movie. Data-driven methods are suitable for the analysis of such complex paradigms as they minimize the assumptions on the underlying properties of the data by decomposing the observed data based on a generative model. The most common decomposition is given by $\mathbf{X} = \mathbf{AS}$ (with the possibility of including an additive noise term), where \mathbf{X} is the mixture that is factorized into latent variables through two matrices—a mixing matrix \mathbf{A} and a component (source) matrix \mathbf{S} . For uniqueness of the decomposition (subject to scaling and permutation ambiguity), constraints are applied to the two matrices such as sparsity or independence of the components. Model-based approaches provide a similar decomposition, however they differ from data-driven methods in their modeling of the matrix \mathbf{A} , which is based on prior knowledge of the experiment and data in the form of regressors. ICA is a popular data-driven BSS technique that imposes the constraint of statistical independence on the components, i.e., source distributions. It has been successfully applied to a number of biomedical data such as fMRI [4], [5] and EEG [6]. A number of approaches have been proposed to solve the ICA problem, a popular one being the maximum likelihood estimation technique that finds an approximation of the underlying sources by using the maximum likelihood estimator of the demixing matrix \mathbf{W} such that $\hat{\mathbf{S}} = \mathbf{WX}$. Second-order data-driven methods have also been used for biomedical data analysis such as linear discriminant analysis, partial least squares, CCA, and source-separation algorithms such as the Molgedey Schuster algorithm [7]. CCA has been used to find latent sources in single subject fMRI data by taking advantage of the spatial or temporal autocorrelation in the data [8]. An extension of the Molgedey Schuster algorithm has also been used to extract sources from two or more data sets based on the temporal autocorrelation in fMRI data [9]. In this article, we focus on reviewing CCA and M-CCA methods for data fusion and multisubject analysis of biomedical data and putting these into perspective through comparisons with closely related ICA-based methods.

DATA FUSION

Data-driven fusion of multimodality data is an especially challenging problem since brain imaging data types are intrinsically dissimilar in nature, making it difficult to analyze them together without making a number of assumptions, most often unrealistic about the nature of the data. Unlike data integration methods, which tend to use information from one modality to improve the other, data fusion techniques incorporate both modalities in a combined analysis, thus allowing for true interaction between the different data types [10]. Instead of

entering the entire data sets into a combined analysis, an alternate approach is to reduce each modality to a feature, which is a lower-dimensional representation of selected brain activity or structure, and then to explore associations across these feature data sets through variations across individuals. Investigating variations across subjects or between patients and controls at the feature-level provides a natural way to find multimodality associations [11] and also alleviates the difficulty of fusing data types of different dimensionality and nature as well as those that have not been recorded simultaneously. Feature-level analysis has been successfully used in data-driven fusion techniques such as joint-ICA (jICA) [11] and CCA-based fusion [12]. Given two feature data sets \mathbf{X}_1 and \mathbf{X}_2 , the jICA approach involves concatenating the data sets alongside each other and then performing ICA on the concatenated data set as in $[\mathbf{X}_1 \mathbf{X}_2] = \mathbf{A}[\mathbf{S}_1 \mathbf{S}_2]$. Joint-ICA assumes that the sources have a common modulation profile \mathbf{A} across subjects, which is a strong constraint considering that the data come from two different modalities. Parallel-ICA (paraICA) [13] is another ICA-based feature-level fusion approach that has been successful in identifying relationships between neuroimaging data types as well as between genetic and phenotypic data. The method performs separate ICA on the different modalities in parallel while enhancing intrinsic correlations across the modalities. While jICA requires the two modalities to have relatively similar dimensions, the two separate ICA in paraICA is more flexible in that it allows the modalities to have different dimensions. For a review of feature-based fusion methods using ICA and their application to biomedical imaging, refer to [14] and [15].

Recently, it has been shown that CCA and M-CCA can allow a more flexible approach to the fusion problem. The CCA-based fusion method also adopts a feature-based approach and similarly models the feature data set from each modality as a linear mixture of components with varying levels of activations for different subjects. Thus, the relationship between modalities is based on intersubject covariations as shown in Figure 1(a). The scheme is flexible as the connections are based only on the linear mixing model and intersubject covariances across modalities, and the modulation profiles of components are not constrained to be exactly the same as in jICA. Successful application of CCA to feature-based fusion of two modalities has been shown in [12] and of M-CCA to the fusion of three brain imaging modalities has been shown in [16]. CCA and M-CCA have also been successfully used for fusion in other fields such as remote sensing [17] and pattern recognition [18].

MULTISUBJECT ANALYSIS

M-CCA can be used to perform group BSS, i.e., source separation of single modality data from multiple subjects [19]. While it is straightforward to apply data-driven techniques such as BSS to each subject's data separately, the challenge lies in matching the separated sources across different data sets, which is straightforward in model-based methods. M-CCA provides an effective tool to perform group BSS while maintaining the correspondence of the source estimates across different data sets and retaining the intersubject source variability. The generative model for M-CCA for BSS is shown in Figure 1(b). A number of data-driven methods have been proposed for achieving group BSS and can be broadly categorized into two different approaches. One approach is to concatenate multiple data sets to aggregate the common features, perform analysis in the common feature space to estimate group components, and back-project the estimated group components into each data set to obtain individual components with cross-data set correspondence. Group ICA [20] and tensorial ICA [21] fall into this category. The other approach assumes a generative model on latent components with cross-data set correspondence and perform group component extraction using statistical measures of correspondence. M-CCA and independent vector analysis (IVA) [22] fall into this category; however, M-CCA and IVA are complementary in modeling component correspondence [19]. Compared with methods based on data set

concatenation, M-CCA is more flexible in identifying cross-data set variation of the components, which can be used for making group level inferences in different ways. M-CCA has been shown to be successful for the group analysis of fMRI data in [19].

OUTLINE

In this article, we begin with a description of the main statistical tools for fusion and multisubject analysis—CCA and M-CCA. We then provide a brief introduction to the medical imaging data that we will be using to demonstrate CCA- and ICA-based analysis techniques. There are two formats in which the data will be utilized by the approaches we present: 1) the whole multidimensional data, e.g., for fMRI data, both the time and the volume information contained in all image slices of the brain, and 2) lower-dimensional features extracted from the whole data to represent certain aspects of the data, e.g., for fMRI data, areas of the brain where neuronal activity due to task is exhibited excluding the time information for these areas. The presented fusion approaches are carried out at the feature-level using CCA, M-CCA, or jICA while the group analysis is carried out on whole multidimensional data sets using M-CCA and group-ICA. To further illustrate the use of the methods, we provide a number of examples of medical imaging applications for the presented techniques.

CCA

CCA has been traditionally used to analyze relationships between two sets of variables [1]. CCA seeks two sets of transformed variates such that the transformed variates assume maximum correlation across the two data sets, while the transformation within each data set are uncorrelated. CCA is an attractive analysis tool, based on second-order statistics and is less stringent than those based on stronger statistical measures such as ICA. Also, being multivariate, it can provide increased statistical power over univariate methods.

Given two data sets $\mathbf{X}_1 \in \mathbb{R}^{n \times p}$ and $\mathbf{X}_2 \in \mathbb{R}^{n \times q}$, CCA finds the linear combinations $\mathbf{X}_1 \mathbf{W}_1$ and $\mathbf{X}_2 \mathbf{W}_2$ that maximize the pair-wise correlations across the two data sets. \mathbf{A}_1 and $\mathbf{A}_2 \in \mathbb{R}^{n \times d}$, $d \leq \min(\text{rank}(\mathbf{X}_1, \mathbf{X}_2))$, are known as canonical variates and $\mathbf{W}_1 \in \mathbb{R}^{p \times d}$ and $\mathbf{W}_2 \in \mathbb{R}^{q \times d}$ are the canonical coefficients vectors.

In the deflationary approach, the method finds the first pair of canonical coefficient vectors $\mathbf{w}_1^{(1)}$ and $\mathbf{w}_2^{(1)}$, ($\mathbf{w}_1^{(1)} \in \mathbb{R}^{p \times 1}$, $\mathbf{w}_2^{(1)} \in \mathbb{R}^{q \times 1}$) that maximize linear combinations of the two data sets given by

$$\max_{\mathbf{w}_1^{(1)}, \mathbf{w}_2^{(1)}} \text{corr}(\mathbf{X}_1 \mathbf{w}_1^{(1)}, \mathbf{X}_2 \mathbf{w}_2^{(1)})$$

to obtain the first pair of canonical variates given by

$$\mathbf{a}_1^{(1)} = \mathbf{X}_1 \mathbf{w}_1^{(1)} \quad \text{and} \quad \mathbf{a}_2^{(1)} = \mathbf{X}_2 \mathbf{w}_2^{(1)}.$$

The remaining $d - 1$ canonical variates can be calculated similarly, with the following additional constraints on the columns of the \mathbf{A} matrices, i.e., $\mathbf{a}_k^{(i)}$ ($i=1, \dots, d, k=1, 2$):

- The canonical variates are uncorrelated within each data set and have zero mean and unit variance, i.e.,

$$\mathbf{A}_k^T \mathbf{A}_k = \mathbf{I}, \quad k=1, 2, \quad (1)$$

- The canonical variates have nonzero correlation only on their corresponding indices, and have correlation coefficients, $r_{k,l}^{(1)} \geq r_{k,l}^{(2)} \geq \dots \geq r_{k,l}^{(d)}$, where $r_{k,l}^{(i)} = \mathbf{a}_k^{(i)T} \mathbf{a}_l^{(i)}$, i.e.,

$$\mathbf{A}_k^T \mathbf{A}_l = \mathbf{R}_{k,l}, \quad k \neq l, k, l=1, 2, \quad (2)$$

where $\mathbf{R}_{k,l} = \text{diag}(r_{k,l}^{(1)}, \dots, r_{k,l}^{(d)})$.

The CCA problem can be posed as a constrained optimization problem using Lagrange multipliers and the canonical covariates can be calculated by solving a generalized eigenvalue solution, where the columns of \mathbf{W}_1 and \mathbf{W}_2 are the eigenvectors of the two matrices

$$\begin{aligned} (\mathbf{C}_{X_2}^{-1} \mathbf{C}_{X_2 X_1} \mathbf{C}_{X_1}^{-1} \mathbf{C}_{X_1 X_2} - \lambda \mathbf{I}) \mathbf{W}_1 &= 0 \quad \text{and} \\ (\mathbf{C}_{X_1}^{-1} \mathbf{C}_{X_1 X_2} \mathbf{C}_{X_2}^{-1} \mathbf{C}_{X_2 X_1} - \lambda \mathbf{I}) \mathbf{W}_2 &= 0, \end{aligned}$$

where λ is vector of eigenvalues or squared canonical correlations, $\mathbf{C}_{X_1 X_2}$ is the cross-correlation matrix of \mathbf{X}_1 and \mathbf{X}_2 ($\mathbf{C}_{X_2 X_1} = \mathbf{C}_{X_1 X_2}^T$), and \mathbf{C}_{X_1} and \mathbf{C}_{X_2} are the autocorrelation matrices of \mathbf{X}_1 and \mathbf{X}_2 , respectively.

M-CCA

The CCA problem can be extended to multiple data sets using the framework developed in [2]. In contrast to CCA where correlation between two canonical variates is maximized, M-CCA optimizes an objective function of the correlation matrix of the canonical variates from multiple random vectors such that the canonical variates achieve maximum overall correlation. Furthermore, due to the consideration of multiple random vectors, M-CCA can not be solved by a simple eigenvalue decomposition problem as in the case of CCA. Instead, M-CCA takes multiple stages such that in each stage, one group of canonical variates is obtained by optimizing the objective function with respect to a set of transformation vectors. For the second stage and higher stages in M-CCA, the estimated canonical variates are constrained to be uncorrelated to the ones estimated in the previous stages. M-CCA reduces to CCA when the number of random vectors is two. Given K data sets, the canonical variates

$$\mathbf{A}_k = \mathbf{X}_k \mathbf{W}_k, \quad k=1, 2, \dots, K \quad (3)$$

can be estimated through a deflationary approach such that we first determine the initial K vectors corresponding to the first source from each of the K data sets using

$$\{\mathbf{w}_1^{(1)}, \mathbf{w}_2^{(1)}, \dots, \mathbf{w}_K^{(1)}\} = \underset{\mathbf{w}}{\text{argmax}} J(\mathbf{r}_{k,l}^{(1)})$$

and then the next vectors using the same procedure such that $\mathbf{w}_k^{(i)}$ is orthogonal to the previous estimates, i.e., to $\{\mathbf{w}_k^{(1)}, \mathbf{w}_k^{(2)}, \dots, \mathbf{w}_k^{(i-1)}\}$, $k = 1, 2, \dots, K$. Here, $r_{k,l}^{(i)}$ is the correlation between the i th canonical variates, from the k th and l th data sets, estimated in the final

decomposition and $J(\cdot)$ is an appropriately chosen cost. The canonical correlations can be obtained by optimizing a number of cost functions proposed in [2], e.g., maximizing the sum of squared correlations among the canonical variates.

We can summarize the M-CCA procedure based on the sum of squares correlation (SSQCOR) cost as

■ Stage 1

$$\{\mathbf{w}_1^{(1)}, \mathbf{w}_2^{(1)}, \dots, \mathbf{w}_K^{(1)}\} = \underset{\mathbf{w}}{\operatorname{argmax}} \left\{ \sum_{k,l=1}^K |r_{k,l}^{(1)}|^2 \right\} \quad (4)$$

■ Stage 2 to d for $i = 2:d$

$$\begin{aligned} \{\mathbf{w}_1^{(i)}, \mathbf{w}_2^{(i)}, \dots, \mathbf{w}_K^{(i)}\} &= \underset{\mathbf{w}}{\operatorname{argmax}} \left\{ \sum_{k,l=1}^K |r_{k,l}^{(i)}|^2 \right\} \\ \text{s.t. } \mathbf{w}_k^{(i)} &\perp \{\mathbf{w}_k^{(1)}, \mathbf{w}_k^{(2)}, \dots, \mathbf{w}_k^{(i-1)}\}, k=1, 2, \dots, K \end{aligned}$$

end

For M-CCA, up to d canonical variates can be calculated iteratively, where $d \leq \min(\operatorname{rank}(\mathbf{X}_k))$. In [2], Stage 1 is solved by first calculating the partial derivative function of the SSQCOR cost with respect to each $\mathbf{w}_k^{(1)}$ and equating it to zero to find the stationary point. Since the SSQCOR cost is a quadratic function of each $\mathbf{w}_k^{(1)}$, the partial derivative is a linear function of $\mathbf{w}_k^{(1)}$ and hence, the closed-form solution can be derived. Starting from an initial point, each $\mathbf{w}_k^{(1)}$ vector is updated in sequel to guarantee an increase in the cost function and a sweep through all the $\mathbf{w}_k^{(1)}$ constitutes one step of the iterative maximization procedure. The iterations are stopped when the cost convergence criterion is met and the resulting $\mathbf{w}_k^{(1)}$ vectors are taken as the optimal solution. Stage 2 and higher stages are solved in a similar manner with the cost function replaced by a Lagrangian incorporating the orthogonality constraints on the canonical coefficient vectors.

Next, we explain the biomedical imaging modalities, the generative models, and some examples of the application of CCA and M-CCA for data fusion and source-separation applications.

MEDICAL IMAGING MODALITIES AND FEATURE GENERATION

In this article, we demonstrate the effectiveness of the CCA-based approach using three modalities: fMRI, sMRI, and EEG. Each of these modalities provides limited information about the human brain. FMRI is a noninvasive brain imaging technique that provides information about brain function by measuring the changes in blood-oxygenation in the brain. SMRI provides information about the tissue type of the brain—gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF). EEG records brain function by measuring the brain electrical field through the scalp. For fusion, we adopt a feature-level analysis to derive a lower dimensional feature from the imaging data as in [14] and [20]. A feature is a subdata set extracted from one type of data, related to a selected brain activity or structure. These features can then be analyzed to integrate or fuse the information across multiple modalities. Next, we briefly introduce the data types, the preprocessing used for

each of these data types, and the types of features we generate for the fusion analysis from each of the data sets.

fMRI

fMRI data provide a measure of brain function on a millimeter spatial scale and a subsecond (and delayed) temporal scale. The data consists of repeatedly imaging the 3-D volume of the brain slice-by-slice, usually while the subject performs a particular task. A number of preprocessing steps are important for fMRI—slice-timing correction to correct for the sequential acquisition of the slices, registration to correct for subject motion in the scanner, spatial filtering to reduce noise, and spatial normalization to compare brains across different individuals and to use standardized atlases to identify particular brain regions. For fMRI data, we use the task-related spatial activity map as calculated by the GLM approach as the spatial feature for the fusion analysis.

sMRI

We define sMRI analysis as the acquisition and processing of T1-, T2-, and/or proton density-weighted images. Multiple structural images are often collected to enable multispectral segmentation approaches. The primary outcome measure in a structural image may include a measure of a particular structure (e.g., volume or surface area) or a description of the tissue type, (e.g., GM or WM). There are many methods for preprocessing sMRI data that may include bias field correction [intensity changes caused by radio frequency (RF) or main magnetic field (B_0) inhomogeneities] [23], spatial linear or nonlinear [24] filtering, and normalization. MR images are typically segmented using a tissue classifier producing images showing the spatial distribution of GM, WM, and CSF. Both supervised and automated segmentation approaches have been developed for sMRI analysis [25]–[27], and each technique is optimized to detect specific features. We use probabilistically segmented GM images as features of sMRI data for the fusion analysis.

EEG

EEG is a technique that measures brain function by recording and analyzing the scalp electrical activity generated by brain structures. Like MRI, it is a noninvasive procedure that can be applied repeatedly in patients, normal adults and children, with virtually no risks or limitations. Local current flows are produced when brain cells are activated. It is believed that contributions are primarily driven by large synchronous populations of firing neurons. The recorded electrical signals are then amplified, digitized, and stored.

Event-related potentials (ERPs) are small voltage fluctuations resulting from evoked neural activity and are one of many ways to process EEG data. These electrical changes are extracted from scalp recordings by computer averaging epochs (recording periods) of EEG time locked to repeated occurrences of sensory, cognitive, or motor events. The spontaneous background EEG fluctuations, which are typically random relative to when the stimuli occurred, are averaged out, leaving the event-related brain potentials. These electrical signals reflect only that activity that is consistently associated with the stimulus processing in a time-locked way. The ERP thus reflects, with high temporal resolution, the patterns of neuronal activity evoked by a stimulus. Due to their high temporal resolution, ERPs provide unique and important timing information about brain processing and are an ideal methodology for studying the timing aspects of both normal and abnormal cognitive processes. More recently, ICA has been used to take advantage of EEG activity that may be averaged out by computing an ERP [6]. For the feature-based fusion analysis we use ERPs as the EEG feature.

DATA FUSION

In this section, we present the data fusion scheme at the feature level using CCA and M-CCA. We explain the data generation model, the modeling assumptions for feature-based fusion using CCA methods, and the dimension reduction step that is used to avoid overfitting. Additionally, we demonstrate the use of the method by presenting two examples and compare the CCA-based methods with ICA-based fusion method, jICA.

GENERATIVE MODEL FOR DATA FUSION

We develop the following generative model for data fusion. Given two feature data sets \mathbf{X}_1 and \mathbf{X}_2 , we seek to decompose them into two sets of components, \mathbf{C}_1 and \mathbf{C}_2 , and corresponding modulation profiles (intersubject variations), \mathbf{A}_1 and \mathbf{A}_2 as shown in Figure 1(a). The connection across the two modalities can be evaluated based on correlations of modulation profiles of one modality with those of the other. If the modulation profiles are uncorrelated within each modality, each component can be associated with only one component across modalities. This one-to-one correspondence aids in the examination of associations across modalities. The generative model is thus given by

$$\mathbf{X}_k = \mathbf{A}_k \mathbf{C}_k, \quad \text{for } k=1, 2,$$

where $\mathbf{X}_k \in \mathbb{R}^{n \times v_k}$, $\mathbf{A}_k \in \mathbb{R}^{n \times d}$, $\mathbf{C}_k \in \mathbb{R}^{d \times v_k}$, v_k is the number of variables in \mathbf{X}_k , n is the number of observations in \mathbf{X}_k , and $d \leq \min[\text{rank}(\mathbf{X}_1, \mathbf{X}_2)]$. The modeling assumptions imply that the modulation profiles, given by columns of \mathbf{A}_1 and \mathbf{A}_2 satisfy the constraints given by (1) and (2). In the feature-based fusion approach [12], the intersubject covariations across the two modalities, i.e., the correlations across the modulation profiles are identified using CCA as described in the section ‘‘Canonical Correlation Analysis.’’ The feature-based fusion scheme models the modulation profiles \mathbf{A}_1 and \mathbf{A}_2 as the canonical variates obtained by CCA, and based on the modulation profiles identified, the associated components can be calculated using least squares approximations given by

$$\widehat{\mathbf{C}}_k = (\mathbf{A}_k^T \mathbf{A}_k)^{-1} \mathbf{A}_k^T \mathbf{X}_k, \quad \text{for } k=1, 2.$$

Thus, this fusion approach identifies the cross-modality covariations, and based on these, it decomposes each feature data set into a set of components—such as spatial areas for fMRI/sMRI or temporal segments for EEG.

Typically, the number of variables (voxels/time points) in the feature data sets is much larger than the number of observations (subjects). Due to the high dimensionality and high noise levels in the brain imaging data, order selection is critical to avoid overfitting the data. Transforming each set of features to a subspace with smaller number of variables helps reduce any redundancy in the analysis. The dimension is chosen to fall in a range where the results are stable and most of the variance in the data can be retained. Dimension reduction is performed on the feature data set using singular value decomposition (SVD), and we perform CCA on the dimension-reduced data sets. We assume a noiseless generative model since we perform dimension reduction. i.e., the assumption in the SVD-based dimension reduction scheme is that small singular values of the matrix that are discarded correspond to additive noise.

The generative model we have described with respect to two data sets can be extended to multiple data sets. For example, for three data sets the CCA fusion method again models the

modulation profiles \mathbf{A}_1 , \mathbf{A}_2 , and \mathbf{A}_3 as the canonical variates—however, it is worth noting that in this case the canonical variates are obtained using M-CCA. The procedure for M-CCA is as described in the section “Multiset CCA.” The calculation of the components as well as the dimension reduction steps for the features are the same as described above.

COMPARISON WITH jICA

Joint-ICA has been successfully used for the fusion of data from two modalities such as fMRI, EEG, and sMRI data [11], [28]. The jICA approach is similar to the CCA-based fusion approach in that it is a second-level analysis based on lower-dimensional features of the data and the associations across the two modalities are based on intersubject covariations. However, there are a number of differences in the modeling assumptions of the two methods. Most importantly, jICA assumes that the sources share a common modulation profile while CCA-based fusion models the modulation profiles of each modality to be separate. Given the diverse nature of the two modalities, assuming that the modulations are exactly the same across different modalities can be a very strong constraint. Another important difference between the methods is that the associations across modalities in CCA-based fusion are solely based on intersubject covariations whereas the associations in jICA are based on the assumptions of common profiles as well as statistical independence among the joint sources. While CCA provides a relatively less constrained solution to the fusion problem, jICA utilizes higher-order statistical information by employing ICA. When correlations are strong between the two modalities, the assumption of a common mixing matrix may be justified and the jICA technique could potentially improve approximation of the joint sources by employing higher-order statistics in the estimation. However, by allowing for separate mixing matrices, CCA-based fusion promises to identify common as well as distinct components and reliably estimates the amount of association between the two modalities. For a detailed comparison of the two models as well as experimental results based on simulated fMRI-like and ERP-like data, refer to [12].

APPLICATION OF CCA TO FUSION OF TWO MODALITIES

The CCA-based fusion approach has been successfully used to analyze the spatio-temporal associations between fMRI data and EEG data, and also, to detect functional and structural relationships between fMRI data and sMRI data in [12]. Here we discuss few of the key findings on the fusion of fMRI and sMRI data and compare the results to those obtained using the jICA method presented in [12].

FUSION OF fMRI AND sMRI—fMRI data provides information about brain function while sMRI data contains information about brain structure. Fusing information from the two modalities could help to understand the link between brain structure and function. We demonstrate the fusion approach on sMRI data and fMRI data from 37 patients with schizophrenia and 36 healthy controls carrying out an auditory sensorimotor task consisting of patterns of eight tones, alternately increasing and decreasing in pitch. The subjects are instructed to press a button with their right thumb for each presented tone. Details of the experimental setup are given in [29]. The fMRI data and sMRI data are converted into lower-dimensional features using the preprocessing techniques described in the section “Medical Imaging Modalities and Feature Generation.” The reduced dimension for both features was empirically chosen as 18.

The pair of components corresponding to profiles showing the strongest correlation ($r_{1,2}^{(1)}=0.87$) across the two data sets demonstrate significant group differences ($\alpha \leq 0.05$: $t_{\text{fMRI}} = -4.92$ and $t_{\text{sMRI}} = -4.80$) between patients with schizophrenia and healthy controls (fMRI map, sMRI map, and scatter plots of the profiles are shown in Figure 3). The fMRI

component map shows that healthy controls have more functional activity in the temporal areas (activations enclosed in blue box) and less motor activity (activations enclosed in red box) compared to patients with schizophrenia. The GM map shows that healthy controls have more GM compared to the patients in frontal (activations enclosed in purple box) and temporal areas (activations enclosed in blue box). These results reveal associations between the modalities in adjacent or close sets of voxels as well as remotely located voxels. This is consistent with previous studies showing changes in both brain structure and brain function in frontal and temporal lobe regions in schizophrenia and is also in agreement with previous studies on fusion of fMRI and GM [28], [30], [31].

We also perform jICA on this data set using the fusion ICA toolbox (http://icatb.sourceforge.net/fusion/fusion_startup.php). The result obtained using jICA, shown in Figure 4, is similar to the one obtained using CCA (Figure 3). However, CCA shows additional motor and temporal areas. Also the structural regions are more localized and well defined in the CCA-based result. In general, however, the jICA components are mostly sparse, at least sparser than components obtained using CCA-based fusion, due to the non-Gaussian emphasis by ICA algorithms such as [32]. In the fusion example presented here, this was seen for the fMRI result but not for the sMRI result. The sparseness of the sMRI results for CCA is hence interesting and may be due to the fact that CCA relaxes the strong constraint of common profiles for the pair of components.

APPLICATION OF M-CCA FOR FUSION OF MULTIPLE MODALITIES

A number of approaches have been proposed to integrate or fuse multitask or multimodality data. However, these have mostly been limited to two modalities or multiple data sets from the same modality. In [16], M-CCA was demonstrated to be successful in fusing data from three modalities. Next, we highlight the key findings from [16] and present new results that demonstrate the increase in sensitivity of the analysis with addition on more modalities.

FUSION OF fMRI, sMRI, AND EEG—The analysis of more than two brain imaging modalities collectively, e.g., fMRI, sMRI, and EEG, can help identify interesting associations across brain structure and function. Performing M-CCA on multiple data sets can be more restrictive since we are requiring covariation of all three modalities, however, this is also informative since we find changes that are related across the three modalities. An interesting point to note is that M-CCA-based fusion allows for associations in local voxels as well as remotely located voxels, thus enabling discoveries of structural changes causing compensatory functional activation in distant, but connected, regions.

Again, we decompose the data into sets of components and their corresponding modulation profiles across the subjects as shown in Figure 5. The data fusion scheme determines the linear transformation that maximizes the intersubject covariations across the three modalities using M-CCA, and based on these covariations, the associations among the components across modalities are determined. As an example for multimodality fusion using M-CCA, consider the fusion of three brain imaging modalities: fMRI, sMRI, and EEG. The MRI and EEG data are acquired from 36 subjects (22 healthy controls and 14 schizophrenia patients). The fMRI and EEG data were collected while the subjects performed an auditory oddball (AOD) task that required them to press a button when they detect a particular infrequent sound among three kinds of auditory stimuli. Details of the task design and the participants are given in [33]. The data was preprocessed and features were obtained as described in the section “Medical Imaging Modalities and Feature Generation.” EEG features, or ERPs, are calculated from the midline central position (Cz) because it appeared to be the best single channel to detect both anterior and posterior sources for the given task.

We perform CCA on the dimension-reduced fMRI, sMRI, and ERP data to estimate 15 sets of components that contain interesting associations across the modalities. The results identify changes in the motor and temporal areas associated with the N2/P3 complex in the ERP (the EEG feature) as shown in Figure 6, areas that have been also previously noted as affected in schizophrenia. On examining the intersubject modulation in conjunction with the spatial and temporal components, the results imply that subjects with schizophrenia have less functional activity and less GM in the areas detected in this component and also a part of the ERP response appeared to be affected. Note that in the section “Application of CCA to Fusion of Two Modalities,” the sensorimotor task showed an increase in motor activity for patients with schizophrenia, while the current result from an auditory oddball task shows a decrease in motor activity. The change in direction is likely due to the significant attentional component in the auditory oddball task. In contrast, the sensorimotor task is predictable, and increase in motor activity in patients performing similar tasks, have been noticed in neuroimaging literature.

We also perform CCA-based fusion on the fMRI and sMRI data sets while excluding the EEG data. Comparing the results of the three-way analysis with those from the two-way analysis (results not shown), we find that for both experiments the areas detected in the fMRI and sMRI component are very similar for the component that showed significant differences between the two groups, with the two-way analysis showing some areas of deactivation. Additionally, we note that the statistical significance of the difference between healthy controls and patients increased significantly with the use of three modalities in the analysis as compared to two modalities (Table 1) confirming the expectation that increased number of modalities do help identify more discriminative features increasing the overall sensitivity of the analysis. Also, if the tests are corrected for Type-I errors using the Bonferroni correction, the significance threshold would be 0.003 and we can see in Table 1 that the results of the three modality fusion satisfy this threshold for at least two modalities while the two modality fusion results do not pass the threshold. Also, note that the Bonferroni correction may be too conservative and instead a less conservative false discovery rate threshold can also be used to check the significance of the results.

In the previous sections, we have presented the use of CCA and M-CCA for the fusion of data from different modalities. Next, we present a related but different framework for the use of M-CCA to perform group study of data from the same modality.

MULTISUBJECT DATA ANALYSIS

In biomedical applications, it is common to study data from a number of subjects under identical experimental conditions and to make inferences based on group analysis—simply looking for occurrences that can be said to be true for the group. In this section, we present the M-CCA based group analysis method for multisubject fMRI data analysis introduced in [19] and highlight one of the key results from the paper along with a comparison with the group ICA analysis technique.

GENERATIVE MODEL AND M-CCA FOR GROUP ANALYSIS

For a group of K data sets, each data set $\mathbf{X}_k = [\mathbf{x}_k^{(1)}, \mathbf{x}_k^{(2)}, \dots, \mathbf{x}_k^{(N)}]^T$, $k = 1, 2, \dots, K$ contains linear mixtures of N sources given in given the source vector $\mathbf{S}_k = [\mathbf{s}_k^{(1)}, \mathbf{s}_k^{(2)}, \dots, \mathbf{s}_k^{(N)}]^T$, mixed by a nonsingular matrix, \mathbf{A}_k i.e.,

$$\mathbf{X}_k = \mathbf{A}_k \mathbf{S}_k, \quad (5)$$

where $\mathbf{X}_k, \mathbf{S}_k \in \mathbb{R}^{N \times Q}$ form the mixture data set and source data set respectively, $\mathbf{A}_k \in \mathbb{R}^{N \times N}$ is a nonsingular square matrix. Note that the mixture data set for multisubject BSS is the whole multidimensional data set and is different from the feature data sets used in the previous sections for data fusion. Sources are uncorrelated within each data set and have zero mean and unit variance, i.e., $E\{\mathbf{S}_k\} = \mathbf{0}$, $k = 1, 2, \dots, K$ and $E\{\mathbf{S}_k \mathbf{S}_k^T\} = \mathbf{I}$, $k = 1, 2, \dots, K$ where \mathbf{I} is the identity matrix. Sources from any pair of data sets $k \neq l$; $k, l \in \{1, 2, \dots, K\}$ have nonzero correlation only on their corresponding indices. Without loss of generality, we assume that the magnitude of correlation between corresponding sources are in nondecreasing order, i.e., $r_{k,l}^{(1)} \geq r_{k,l}^{(2)} \geq \dots \geq r_{k,l}^{(N)}$, where $r_{k,l}^{(i)} = E\{\mathbf{s}_k^{(i)} (\mathbf{s}_l^{(i)})^T\}$.

This assumed correlation pattern for latent sources in the generative model can be effectively used to construct a multisubject separation scheme using M-CCA, which is shown in Figure 1(b). In this scheme, the group of sources that have the maximal between-set correlation values are first extracted from the data sets. By removing the estimated sources from the data sets and repeating the correlation maximization procedure, subsequent procedures can extract groups of corresponding sources from each data set in decreasing order of between-set correlation values. This procedure is described in the section ‘‘Multiset CCA,’’ however, in this case the canonical variates will not be defined as in (3) and instead

they will be defined as $\mathbf{S}_k = \mathbf{W}_k \mathbf{X}_k$, for $k = 1, 2, \dots, K$ and $r_{k,l}^{(i)} = E\{\mathbf{s}_k^{(i)} (\mathbf{s}_l^{(i)})^T\}$.

In [19], the study of source separability conditions based on a flexible generative model shows that the method can be used to achieve successful source separation under mild conditions. The conditions depend on the chosen cost function, e.g., we show that

$J(\mathbf{r}_{k,l}) = \sum_{k,l=1}^K |\mathbf{r}_{k,l}|^2$ is a practical choice for the cost and that it leads to robust separation performance and a separability condition that is easily satisfied especially when the number of observations increases. The superior performance of M-CCA is shown for group source separation for large number of data sets, robustness to outliers, and robustness to complex-valued data distributions, when compared with data-driven methods that assume a non-Gaussian model [19].

COMPARISON WITH OTHER APPROACHES

Group ICA achieves source separation of multiple data sets by first reducing the dimensionality of data from each subject, followed by reducing these reduced data sets to a common subspace, then performing ICA on this common subspace, and finally back-reconstructing the subject-specific source estimates. The data reduction steps used to obtain the common subspace reduce the amount of subject variability in the estimated subject-specific source estimates. M-CCA, on the other hand, performs BSS after a subject-level data reduction stage and does not work on a common subspace. This allows M-CCA to retain much of the intersubject variability. Tensorial ICA also specifies a common signal subspace model that is similar to that of group ICA. Additionally, in tensorial ICA, each group of the corresponding mixing vectors is represented by a common mixing vector associated with a cross-subject variation vector using a rank-one approximation. In this way, the group data sets are decomposed into a three-way tensor product of the common sources, common mixing vectors, and the associated cross-subject variation vectors. IVA uses a mutual information-based formulation to perform source separation across multiple data sets; however, the algorithmic development of the method involves the simplifying assumption that the estimated sources are uncorrelated across data sets, which is an unrealistic assumption for many applications including analysis of biomedical data sets. For

a more detailed comparison of these group analysis techniques, refer to [19], and for review of ICA-based multisubject analysis methods in fMRI, refer to [15].

APPLICATION OF M-CCA TO MULTISUBJECT DATA ANALYSIS

The multiple data set extension of CCA, M-CCA reveals relationship among the hidden factors in multiple data sets. In this section, we show how M-CCA can be used for source separation across multiple data sets.

MULTISUBJECT ANALYSIS OF fMRI—Twelve right-handed participants with normal vision—six females, six males, average age 30 years—participated in the study. Subjects performed a visuomotor task involving two identical but spatially offset, periodic, visual stimulus, shifted by 20 s from one another. A total of 12 data sets are jointly analyzed. Each data set is preprocessed according to typical fMRI analysis procedures consisting of slice-timing correction, image registration, motion correction, smoothing, whitening, and dimension reduction. Thirty-two normalized principal components (PCs) are retained for each data set and M-CCA is applied to the 12 sets of retained PCs. The optimal number of PCs are selected using an information theoretic criterion with correction for sample dependence; for implementation details refer to [19].

We present a source of interest from the M-CCA and group ICA estimation results. The M-CCA result shows activation at inferior parietal lobule, posterior cingulate, and medial frontal gyrus—this set of regions is called the “default mode” network that tends to be less active during the performance of a task [34]—as well as deactivation in motor, temporal, and visual regions. The group ICA result, on the other hand, focuses on the default mode activity. The estimated mean activation maps over all data sets, image of the cross-subject source correlation matrices, and the mean time course are displayed in Figure 7. The right- and left-side visuomotor task paradigm is overlaid onto the estimated time courses for reference.

The estimated sources by M-CCA and group ICA are shown in Figure 7. It is observed that the spatial map estimated by M-CCA shows higher cross-subject correlation level than group ICA. The time courses of default mode estimated by M-CCA and group ICA both show expected negative correlation against the onset of the visuomotor task. Furthermore, a multiple linear regression is performed on the estimated time course with the right (R) and left (L) visuomotor paradigm regressors. It is observed that time course estimated by M-CCA has more significant regression coefficients with the task paradigms, i.e., the values are for M-CCA (R): -0.52 with estimated confidence interval (CI): $[-0.38, -0.65]$ and (L): -0.87 CI: $[-0.74, -1.01]$; group ICA (R): -0.45 CI: $[-0.28, -0.62]$ and (L) -0.60 CI: $[-0.43, -0.77]$. Hence, M-CCA achieves higher consistency on spatial activation region and also the time courses show a higher correlation with the task paradigm. The agreement of the spatial and temporal features suggests that default mode network is a common feature across all subjects that is driven by both the left and right visuomotor task.

DISCUSSION

We have presented two CCA-based approaches for data fusion and group analysis of biomedical imaging data and demonstrated their utility on fMRI, sMRI, and EEG data. The results show that CCA and M-CCA are powerful tools that naturally allow the analysis of multiple data sets. The data fusion and group analysis methods presented are completely data driven, and use simple linear mixing models to decompose the data into their latent components. Since CCA and M-CCA are based on second-order statistics they provide a relatively less constrained solution as compared to methods based on higher-order statistics such as ICA. While this can be advantageous, the flexibility also tends to lead to solutions

that are less sparse than those obtained using assumptions of non-Gaussianity—in particular super-Gaussianity—at times making the results more difficult to interpret. Thus, it is important to note that both approaches provide complementary perspectives, and hence it is beneficial to study the data using different analysis techniques.

Though, in general, the tendency in data analysis is to try to minimize the assumptions on the nature of data, certain assumptions may be suited to the data being studied, and strong assumptions such as independence or sparsity might help improve robustness of the solutions. Thus, the performance of a method should be judged on the overall properties rather than a simple optimality criterion, while taking the underlying assumptions into account. Especially in the case of the study of brain structure and function, since the ground truth is seldom available, the assumptions in most cases cannot be verified. Thus, fully exploiting the complementary nature of different methods becomes especially important, as now noted in most neuroimaging literature. As we demonstrate in this article, CCA- and M-CCA-based methods provide attractive solutions to data fusion and group analysis, and their true power might be realized when they are used in conjunction with other methods that are complementary in nature. Also, their extensions to incorporate sparsity and higher-order statistical information can help improve their utility.

AUTHORS

Nicolle M. Correa (nicolle1@umbc.edu) received a bachelor's degree in electronics and telecommunications engineering from St. Francis Institute of Technology, Mumbai University, India, in 2003. She received a master's degree in electrical engineering from the University of Maryland Baltimore County (UMBC), in 2005. She is currently a Ph.D. candidate in the Machine Learning for Signal Processing Laboratory at UMBC. Her research interests include statistical signal processing, machine learning, and their applications to medical image analysis, genetics, and proteomics.

Tülay Adalı (adali@umbc.edu) <AU: please check that e-mail address is correct> received the Ph.D. degree in electrical engineering from North Carolina State University, Raleigh, in 1992 and joined the faculty at UMBC as a professor that same year. She was the general cochair of the IEEE International Workshop on Neural Networks for Signal Processing (2001–2003); technical chair of the IEEE International Workshop on Machine Learning for Signal Processing (MLSP) (2004–2008); publicity chair of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) (2000 and 2005); and publications cochair of ICASSP 2008. She chaired the IEEE Signal Processing Society (SPS) MLSP Technical Committee (2003–2005) and serves on the IEEE SPS MLSP and the Signal Processing Theory and Methods Technical Committees. She was an associate editor for both *IEEE Transactions on Signal Processing* and *Signal Processing* and is currently an associate editor for *IEEE Transactions on Biomedical Engineering*, *IEEE Journal of Selected Topics in Signal Processing*, and *Journal of Signal Processing Systems*. She is a Fellow of the IEEE and the AIMBE and a recipient of an NSF CAREER Award. Her research interests are in the areas of statistical signal processing, MLSP, and biomedical data analysis.

Yi-Ou Li (liyoiu1@umbc.edu) <AU: please check that e-mail address is correct> received his B.S. degree in electrical engineering from Beijing University of Posts and Telecommunications. He graduated from UMBC with a Ph.D. degree in electrical engineering. He joined the Neural Connectivity Lab in the Department of Radiology and Biomedical Imaging at the University of California, San Francisco, in 2009. His research interests include statistical signal processing, multivariate analysis, and quantitative evaluation of data-driven methods and their applications to functional MRI data analysis.

Vince D. Calhoun (vcalhoun@unm.edu) <AU: please check that e-mail address is correct> received his Ph.D. degree in electrical engineering from UMBC in 2002. He is the chief technology officer at the Mind Research Network and is an associate professor at the University of New Mexico. He is the author of over 100 journal articles and 200 conference proceedings. He is on the editorial board of *Human Brain Mapping* and *Neuroimage*. His research interests include the development of data-driven methods for brain imaging and genetics in the areas of image processing and pattern recognition. He is a Senior Member of the IEEE.

REFERENCES

1. Hotelling H. Relations between two sets of variates. *Biometrika* 1936;28:321–377. <AU: Kindly provide the issue number.>.
2. Kettenring J. Canonical analysis of several sets of variables. *Biometrika* 1971;58:433–451. <AU: Kindly provide the issue number.>.
3. Friston K, Jezzard P, Turner R. Analysis of functional MRI time-series. *Hum. Brain Mapp* 1994;1:153–171. <AU: Kindly provide the issue number.>.
4. McKeown MJ, Makeig S, Brown GG, Jung TP, Kindermann SS, Sejnowski TJ. Analysis of fMRI by blind separation into independent spatial components. *Hum. Brain Mapp* 1998;6:160–188. <AU: Kindly provide the issue number.>. [PubMed: 9673671]
5. Calhoun VD, Adali T. Unmixing fMRI with independent component analysis. *IEEE Eng. Med. Biol. Mag* Mar./Apr.;2006 25(2):79–90. [PubMed: 16568940]
6. Jung TP, Makeig S, Westerfield M, Townsend J, Courchesne E, Sejnowski TJ. Analysis and visualization of single-trial event-related potentials. *Hum. Brain Mapp* 2001;14:166–185. <AU: Kindly provide the issue number.>. [PubMed: 11559961]
7. Molgedey L, Schuster H. Separation of independent signals using time-delayed correlations. *Phys. Rev. Lett* 1994;723:3634–3637. <AU: Kindly provide the issue number.>. [PubMed: 10056251]
8. Friman O, Borga M, Lundberg P, Knutsson H. Exploratory fMRI analysis by autocorrelation maximization. *Neuroimage* 2002;16:454–464. <AU: Kindly provide the issue number.>. [PubMed: 12030831]
9. Lukic, AS.; Wernick, MN.; Hansen, LK.; Anderson, J.; Strother, SC. A spatially robust ICA algorithm for multiple fMRI data sets. *Proc. IEEE Int. Symp. Biomedical Imaging (ISBI)*; 2002; p. 839–842.
10. Savopoul, F.; Armenakis, C. Mergine of heterogeneous data for emergency mapping: Data integration or data fusion?. *Proc. ISPRS; Ottawa, ON, Canada. 2002.* <AU: Kindly provide the complete page range.>
11. Calhoun VD, Adali T, Pearlson GD, Kiehl KA. Neuronal chronometry of target detection: Fusion of hemodynamic and event-related potential data. *Neuroimage* 2006;30(2):544–553. [PubMed: 16246587]
12. Correa N, Li Y-O, Adali T, Calhoun VD. Canonical correlation analysis for feature-based fusion of biomedical imaging modalities and its application to detection of associative networks in schizophrenia. *IEEE J. Select. Topics Signal Process* Dec.;2008 2(6):998–1007.
13. Liu J, Pearlson GD, Windemuth A, Ruano G, Perrone-Bizzozero NI, Calhoun VD. Combining fMRI and SNP data to investigate connections between brain function and genetics using parallel ICA. *Hum. Brain Mapp* 2009;30(1):241–255. [PubMed: 18072279]
14. Calhoun VD, Adali T. Feature-based fusion of medical imaging data. *IEEE Trans. Inform. Technol. Biomed* 2009;13:1–10. <AU: Kindly provide the issue number.>.
15. Calhoun VD, Sui J, Adali T. A review of group ICA for fMRI data and ICA for joint inference of imaging, genetic, and ERP data. *Neuroimage* Mar.;2009 45(1)(Suppl. 1):S163–S172. [PubMed: 19059344]
16. Correa, N.; Li, Y-O.; Adali, T.; Calhoun, VD. Fusion of fMRI, sMRI, and EEG data using canonical correlation analysis. *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*; 2009. <AU: Kindly provide the complete page range.>

17. Nielsen AA. Multiset canonical correlation analysis and multispectral, truly multitemporal remote sensing data. *IEEE Trans. Image Process* 2002;11(3):293–305. [PubMed: 18244632]
18. Sun, Q-S.; Zeng, S-G.; Heng, P-A.; Xia, D-S. Feature fusion method based on canonical correlation analysis and handwritten character recognition. *Proc. Int. Conf. Control, Automation, Robotics, and Vision*; 2004. <AU: Kindly provide the complete page range.>
19. Li Y-O, Wang W, Adali T, Calhoun VD. Joint blind source separation by multi-set canonical correlation analysis. *IEEE Trans. Signal Processing* 2009;57(10):3918–3929.
20. Calhoun VD, Adali T, Pekar JJ, Pearlson GD. A method for making group inferences from functional MRI data using independent component analysis. *Hum. Brain Mapp* 2001;14:140–151. <AU: Kindly provide the issue number.>. [PubMed: 11559959]
21. Beckmann CF, Smith SM. Tensorial extensions of independent component analysis for group fMRI data analysis. *Neuroimage* 2005;25(1):294–311. [PubMed: 15734364]
22. Kim T, Attias HT, Lee SY, Lee TW. Blind source separation exploiting higher-order frequency dependencies. *IEEE Trans. Acoust., Speech, Signal Processing* 2007;15:70–79. <AU: Kindly provide the issue number.>.
23. Cohen MS, DuBois RM, Zeineh MM. Rapid and effective correction of RF inhomogeneity for high field magnetic resonance imaging. *Hum. Brain Mapp* 2000;10:204–211. <AU: Kindly provide the issue number.>. [PubMed: 10949058]
24. Gerig G, Martin J, Kikinis R, Kubler O, Shenton ME, Jolesz F. Unsupervised tissue type segmentation of 3D dual-echo MR head data. *Image Vis. Comput* 1992;10(6) <AU: Kindly provide the complete page range.>.
25. Zhang Y, Brady M, Smith S. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Trans. Med. Imaging* 2001;20:45–57. <AU: Kindly provide the issue number.>. [PubMed: 11293691]
26. Wells WM III, Grimson WEL, Kikinis R, Jolesz FA. Adaptive segmentation of MRI data. *IEEE Trans. Med. Imaging* 1996;15(4):429–442. [PubMed: 18215925]
27. Bezdek JC, Hall LO, Clarke LP. Review of MR image segmentation techniques using pattern recognition. [Review]. *Med. Phys* 1993;20:1033–1048. <AU: Kindly provide the issue number.>. [PubMed: 8413011]
28. Calhoun VD, Adali T, Giuliani NR, Pekar JJ, Kiehl KA, Pearlson GD. Method for multimodal analysis of independent source differences in schizophrenia: combining gray matter structural and auditory oddball functional data. *Neuroimage* 2006;30:544–553. <AU: Kindly provide the issue number.>. [PubMed: 16246587]
29. Machado, G.; Juarez, M.; Clark, VP.; Gollub, R.; Magnotta, V.; White, T.; Calhoun, VD. Probing schizophrenia using a sensorimotor task: Large-scale (N=273) independent component analysis of first episode and chronic schizophrenia patients. *Proc. Society for Neuroscience*; San Diego, CA. 2007. <AU: Kindly provide the complete page range.>
30. Pearlson GD, Marsh L. Structural brain imaging in schizophrenia: A selective review. *Biol. Psychiatry* 1999;46:627–649. <AU: Kindly provide the issue number.>. [PubMed: 10472416]
31. Pearlson GD, Calhoun VD. Structural and functional magnetic resonance imaging in psychiatric disorders. *Can. J. Psychiatry* 2007;52 <AU: Kindly provide the issue number and complete page range.>.
32. Bell AJ, Sejnowski TJ. An information-maximization approach to blind separation and blind deconvolution. *Neural Comput* 1995;7(6):1004–1034.
33. Kiehl KA, Stevens M, Laurens KR, Pearlson GD, Calhoun VD, Liddle PF, et al. An adaptive reflexive processing model of neurocognitive function: Supporting evidence from a large scale (n=100) fMRI study of an auditory oddball task. *Neuroimage* 2005;25:899–915. <AU: Kindly provide all the author names and issue number.>. [PubMed: 15808990]
34. Raichle ME, MacLeod AM, Snyder AZ, Powers WJ, Gusnard DA, Shulman GL. A default mode of brain function. *Proc. Nat. Acad. Sci* 2001;98:676–682. <AU: Kindly provide the issue number.>. [PubMed: 11209064]

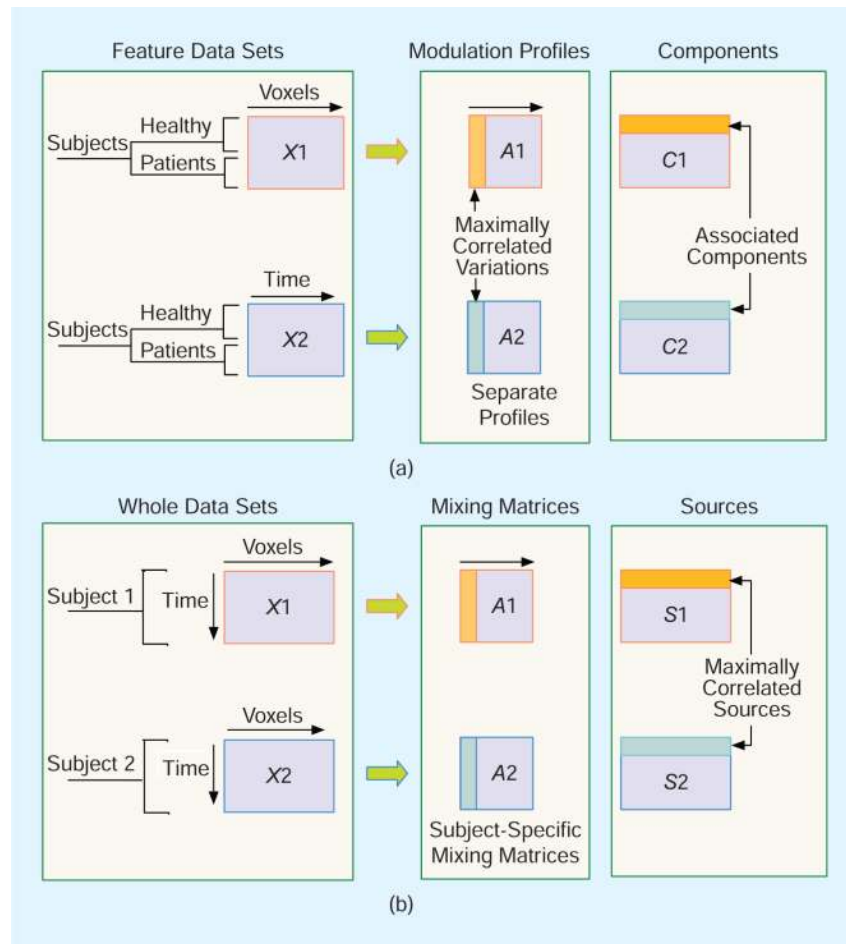


FIG1. Generative models for fusion and source separation are shown in (a) and (b). To avoid overfitting, typically dimension reduced data matrices are used instead of the original high-dimensional data X , X_1 , and X_2 . For data fusion, the spatial or temporal dimension is reduced. For group analysis, the temporal dimension is reduced.

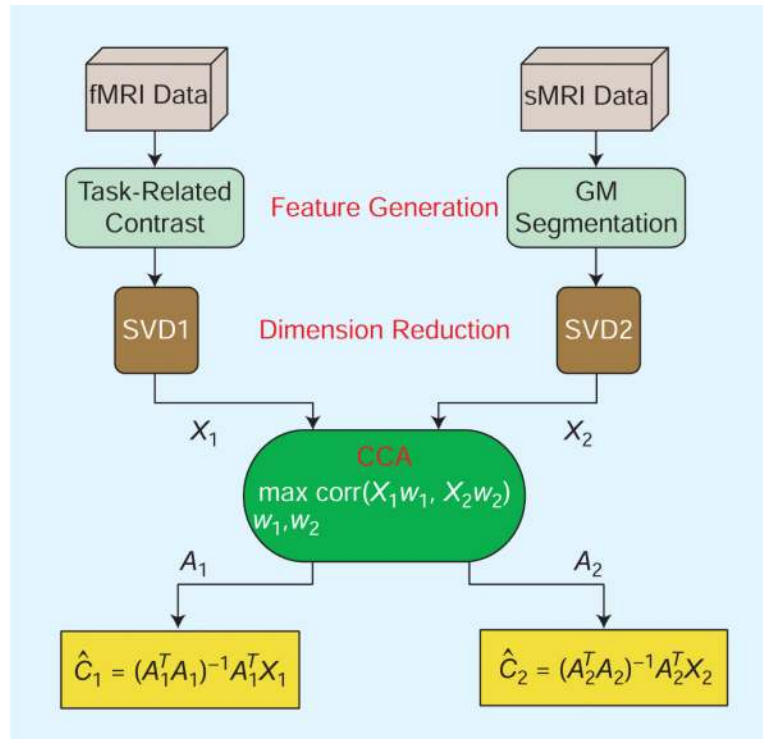
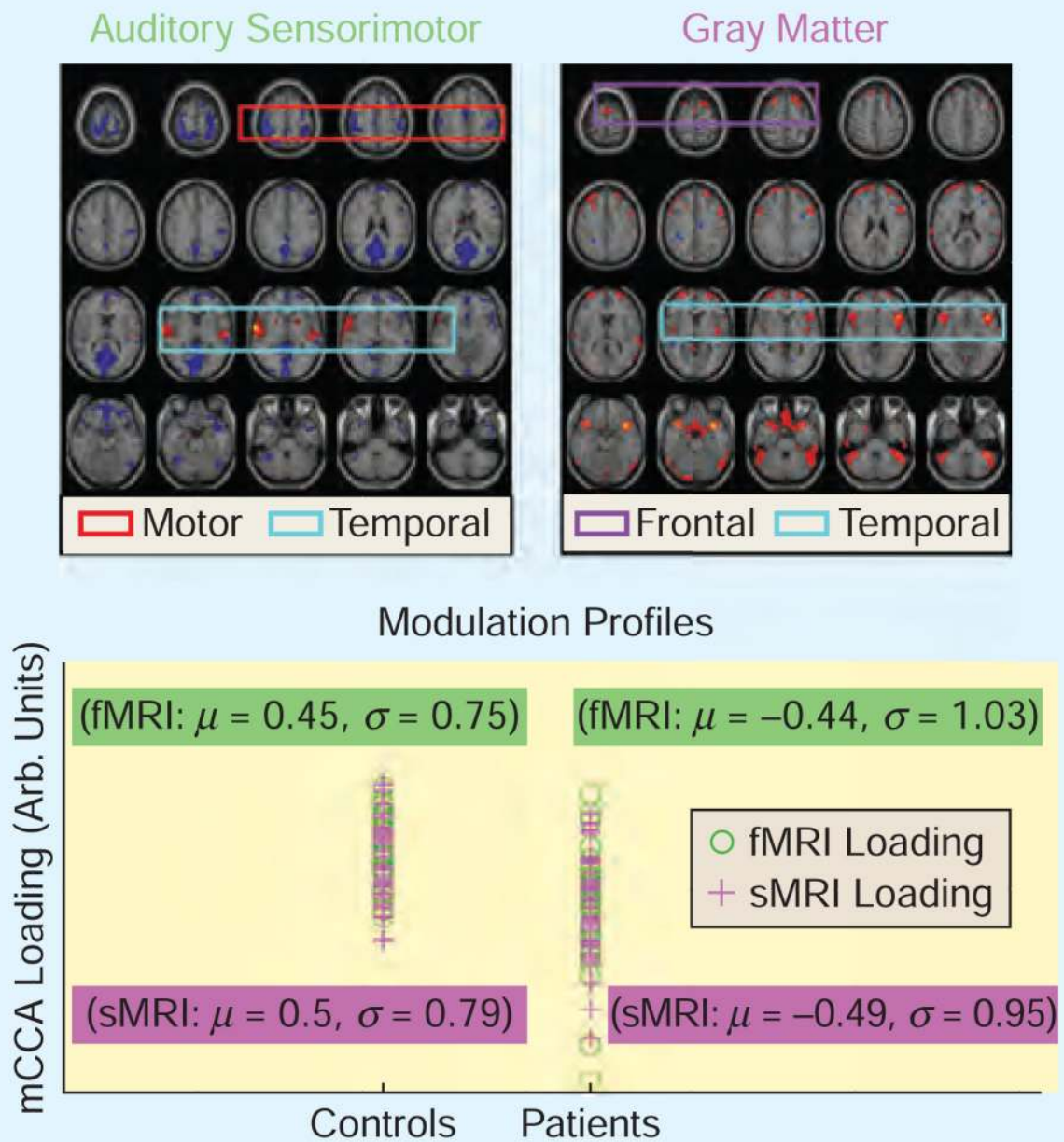


FIG2.
Implementation steps for CCA-based fusion.

**FIG3.**

The fMRI component, sMRI component, and scatter plots of profiles for pair of components identified by CCA as maximally correlated. The profiles for both fMRI and sMRI are significantly different ($\alpha \leq 0.05$) between patients and controls. Patients with schizophrenia show more functional activity in motor areas and less activity in temporal areas associated with less gray matter as compared to healthy controls. The activation maps are scaled to Z values and thresholded at $Z = 3.5$.

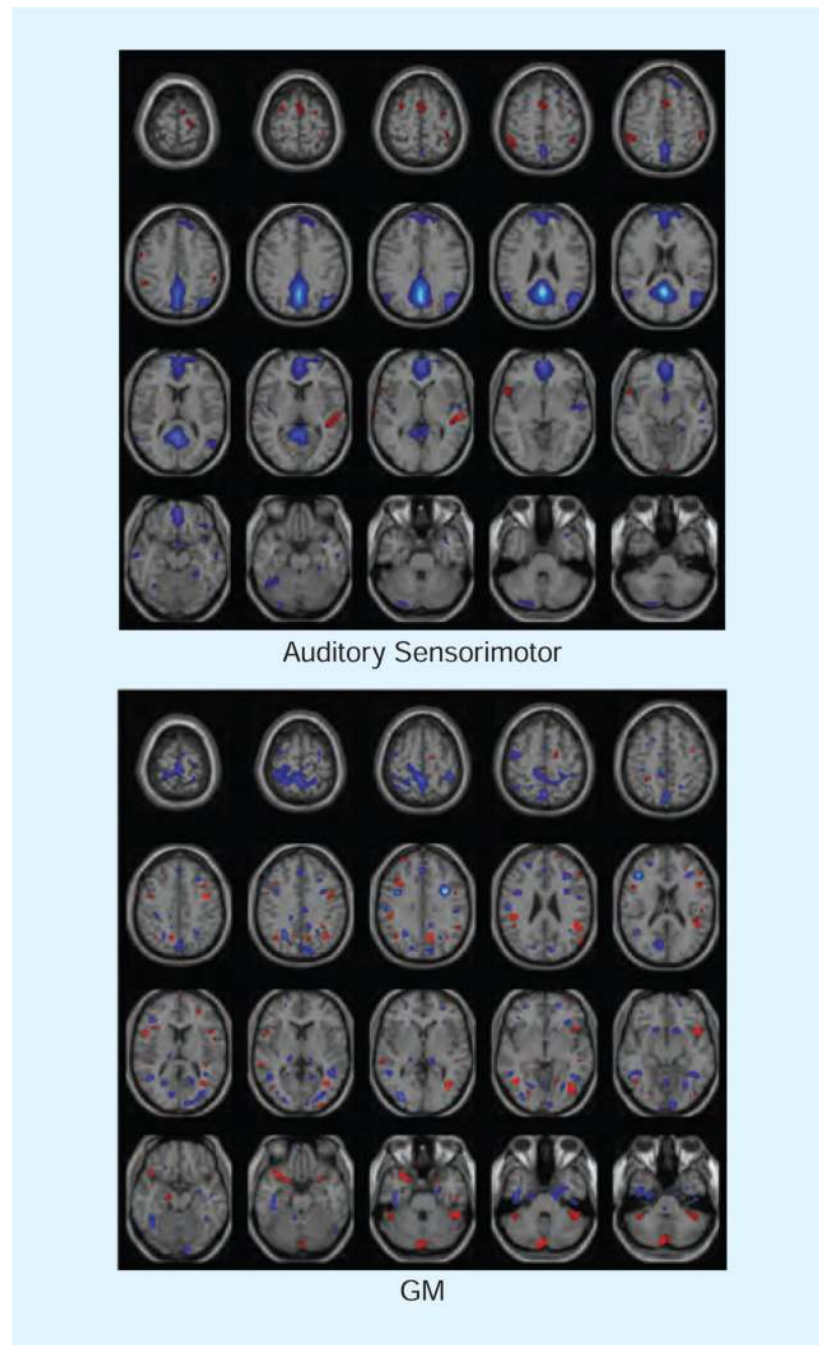


FIG4. Joint components estimated by jICA corresponding to common profile demonstrating significant difference between patients and controls. The activation maps are scaled to Z values and thresholded at $Z = 3.5$.

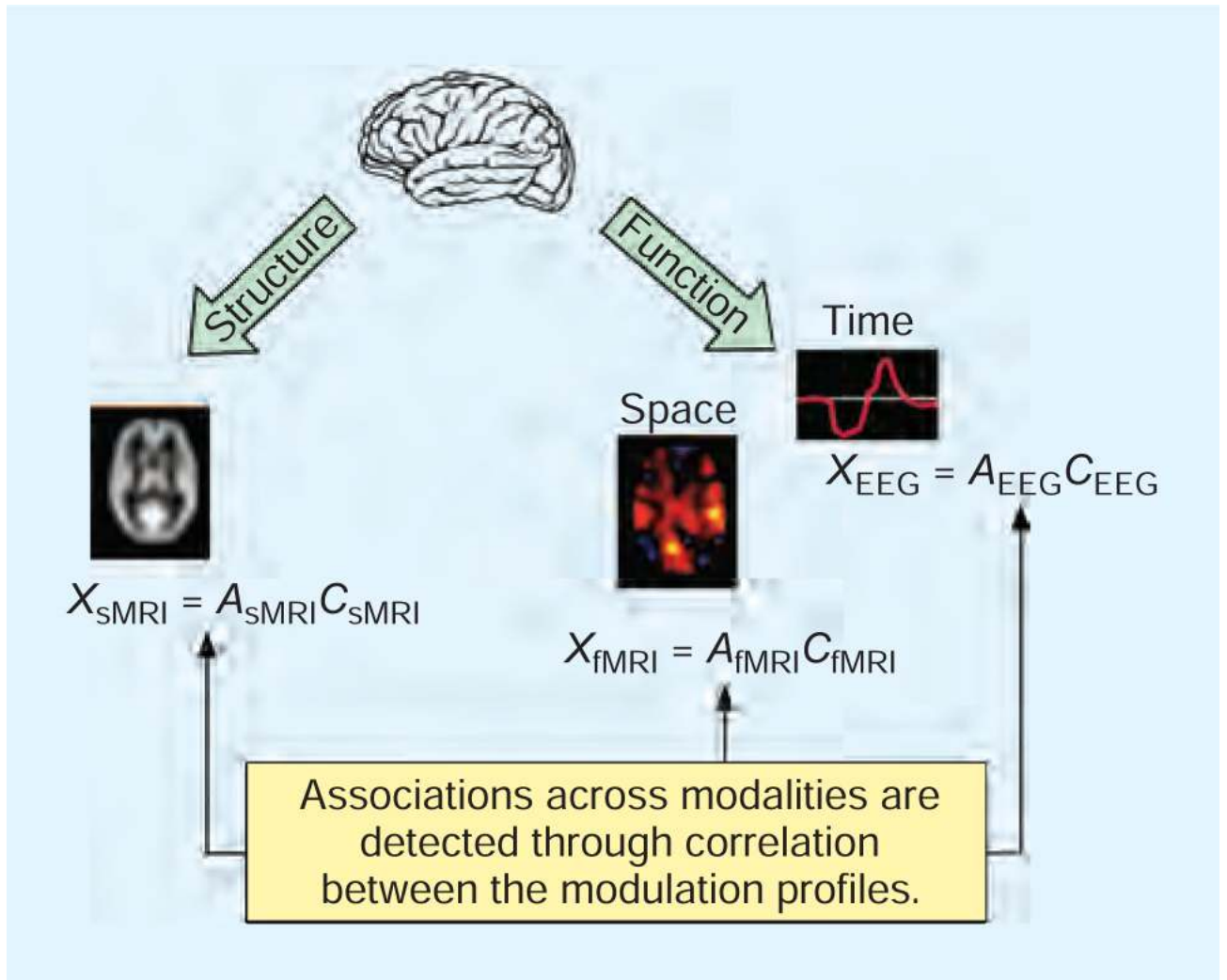


FIG5.
Data model for fusion of brain structure and function.

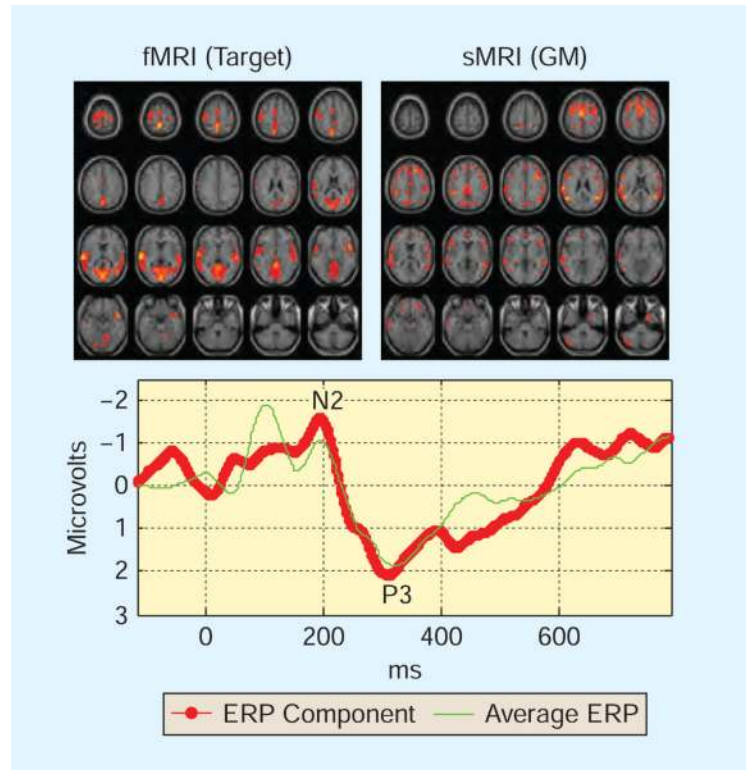
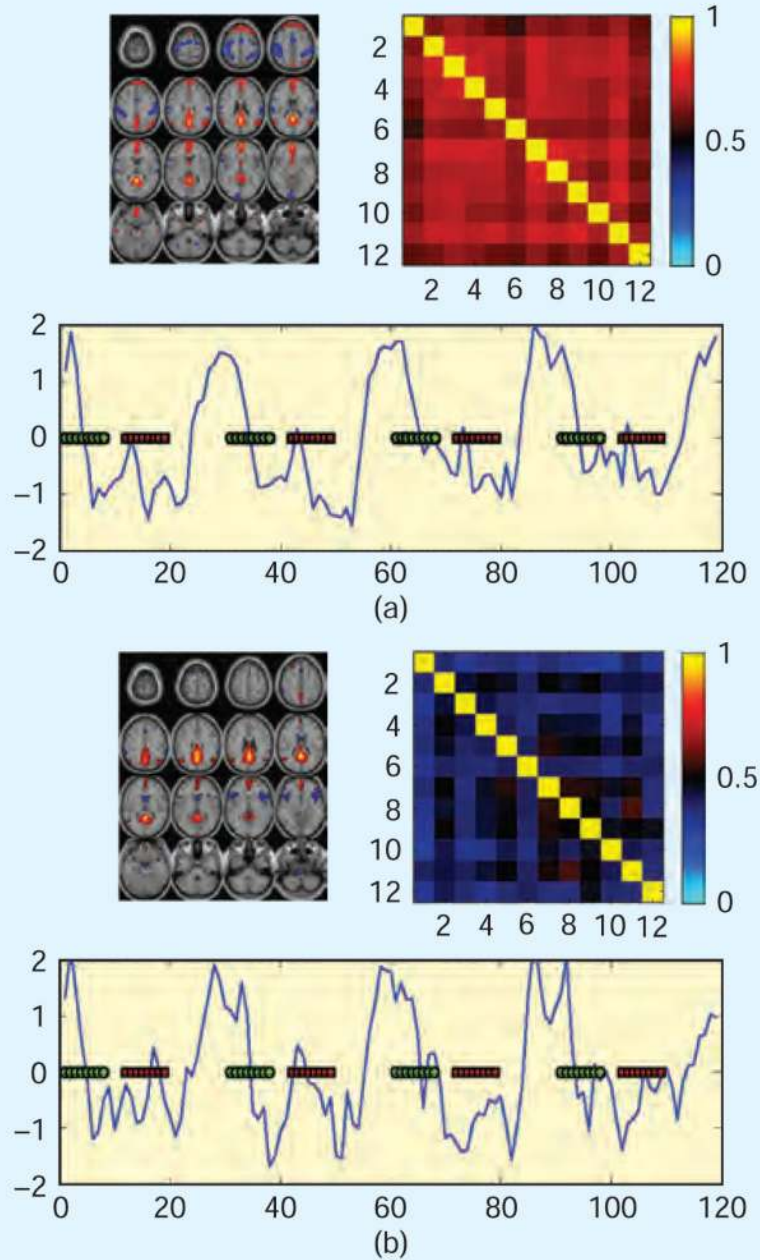


FIG6. Set of associated components estimated by M-CCA that showed significantly different loading for patients versus controls.

**FIG7.**

Estimated mean activation maps (top left), source correlation between subjects (top), and time course (bottom) of the default mode by (a) M-CCA and (b) group ICA. The right (green circle) and left (red block) visuomotor task paradigm is overlaid onto the estimated time courses for reference.

TABLE 1

COMPARISON OF t -TESTS FOR THREE MODALITY (fMRI, sMRI, AND EEG) VERSUS TWO MODALITY (fMRI AND sMRI) ANALYSES FOR COMPONENT. THE t -TESTS ARE PERFORMED ON THE LOADINGS FROM THE MODULATION PROFILES OF HEALTHY AND SCHIZOPHRENIC SUBJECTS.

MODALITIES	THREE MODALITIES		TWO MODALITIES	
	t	α	t	α
fMRI	3.45	0.002	2.17	0.038
sMRI	2.86	0.001	2.20	0.034
EEG	3.61	0.007	–	–