Published in final edited form as:

Neuroimage. 2010 May 15; 51(1): 123–134. doi:10.1016/j.neuroimage.2010.01.069.

A CCA+ICA based model for multi-task brain imaging data fusion and its application to schizophrenia

Jing Sui^{a,*}, Tülay Adalı^b, Godfrey Pearlson^{c,d}, Honghui Yang^{e,f}, Scott Sponheim^{g,h}, Tonya White^{h,i}, and Vince D. Calhoun^{a,c,d,e}

Jing Sui: jsui@mrn.org; Vince D. Calhoun: vcalhoun@unm.edu

^a The Mind Research Network, 1101 Yale Blvd, NE, Albuquerque, NM 87106, USA ^b Department of CSEE, University of Maryland, Baltimore County, Baltimore, MD 21250, USA ^c Olin Neuropsychiatry Research Center, Hartford, CT 06106, USA ^d Department of Psychiatry, Yale University, New Haven, CT 06519, USA ^e Department of ECE, University of New Mexico, Albuquerque, NM 87131, USA ^f Department of Environment Engineering, Northwestern Polytechnical University, 710072, China ^g VA Medical Center, Minneapolis, MN 55417, USA ^h Department of Psychiatry, University of Minnesota, Minneapolis, MN 55455, USA ⁱ Department of Child and Adolescent Psychiatry, Erasmus Medical Centre, Rotterdam, Netherlands

Abstract

Collection of multiple-task brain imaging data from the same subject has now become common practice in medical imaging studies. In this paper, we propose a simple yet effective model, “CCA+ICA”, as a powerful and new method for multi-task data fusion. This joint blind source separation (BSS) model takes advantage of two multivariate methods: canonical correlation analysis and independent component analysis, to achieve both high estimation accuracy and to provide the correct connection between two datasets in which sources can have either common or distinct between-dataset correlation. In both simulated and real fMRI applications, we compare the proposed scheme with other joint BSS models and examine the different modeling assumptions. The contrast images of two tasks: sensorimotor (SM) and Sternberg working memory (SB), derived from a general linear model (GLM), were chosen to contribute real multi-task fMRI data, both of which were collected from 50 schizophrenia patients and 50 healthy controls. When examining the relationship with duration of illness, CCA+ICA revealed a significant negative correlation with temporal lobe activation. Furthermore, CCA+ICA located sensorimotor cortex as the group-discriminative regions for both tasks and identified the superior temporal gyrus in SM and prefrontal cortex in SB as task-specific group-discriminative brain networks. In summary, we compared the new approach with competitive methods with different assumptions, and found consistent results regarding each of their hypotheses on connecting the two tasks. Such an approach fills a gap in existing multivariate methods for identifying biomarkers from brain imaging data.

Keywords

Independent component analysis (ICA); Canonical correlation analysis (CCA); fMRI; Schizophrenia; Multi-task; Brain imaging data fusion; Joint blind source separation

*Corresponding author. Fax: +1 505 272 8002.

Introduction

Collection of multi-task data from the same subjects has become common practice primarily due to the desire to select tasks that probe a particular cognitive function or maximize clinical sensitivity. Typically these data are analyzed separately; however separate analyses do not enable the examination of cross-information among the tasks. In this paper, we present a new approach that enables joint analysis of multi-task data and investigation of both “common” and “unique” information among different tasks (or different conditions for the same task, e.g., target vs. novel in auditory oddball task), where the descriptor “common” refers to co-variation in both tasks and “distinct” denotes changes in only one of the tasks.

We focus on multi-task brain imaging data fusion in this paper, which is a second-level analysis based on “features”. In the case of fMRI, a “feature” is a summary image such as a contrast image (3D spatial map) computed via the GLM, or a component image such as the “default mode” resulting from a first-level ICA. It contributes as an input vector for each task and each subject. We note that this definition of “feature” is somewhat different than what is used in traditional machine learning algorithms (Blum and Langley, 1997). In our case, a “feature” is a distilled dataset representing the task-related activations and tends to be more tractable than working with the original 4D data due to the reduced dimension (Calhoun and Adali, 2009).

For analyzing a single brain imaging dataset, the most widely used approach is voxel-wise multiple regression using regressors derived from ideal task conditions convolved with a canonical hemodynamic response function (Rajapakse et al., 1998). More recently, a variety of data-driven approaches have been developed, including independent component analysis (ICA) (Beckmann and Smith, 2005; Calhoun et al., 2001; Guo and Pagnoni, 2008; Lee et al., 2008; Smith et al., 2009), partial least squares (PLS) (Lin et al., 2003; McIntosh et al., 1996), clustering (Esposito et al., 2005), local linear discriminant analysis (McKeown et al., 2007), and support vector machines (Wang et al., 2007). All of these approaches can be considered to be first-level processing, i.e., calculated using single subject spatio-temporal fMRI data, and while they are useful for generating summary results reporting on a single modality, they do not report on the combination of data from multiple fMRI tasks.

Methods for joint analysis of two brain imaging datasets include the use of contrast images in a second-level GLM, e.g., controlling for structural information in the form of a covariate (Oakes et al., 2007). Such an approach however does not provide information about inter-voxel relationships. Methods such as path analysis or structural equation modeling have also been used when it is possible to employ a small summary of measures from a limited number of regions of interest (Bullmore et al., 2000; Ecker et al., 2008; Schlosser et al., 2003); However, in this case one may miss some of the relationships, since not all voxels are included in the analysis and averages across a small number of regions of interest are typically utilized.

More recently, multivariate data-mining based approaches for data fusion have been developed. These methods have primarily been focused upon second-level analysis (e.g. calculated from the first level output such as contrast maps from the GLM) including (1) joint ICA (jICA) (Calhoun et al., 2006), (2) multimodal canonical correlation analysis (mCCA) (Correa et al., 2007), and (3) some of their extensions (e.g., parallel ICA (Liu et al., 2009) and CC-ICA (Sui et al., 2009b)). In addition, a first-level PLS approach has been proposed (Grady et al., 2006).

All of the above multivariate methods analyze data via blind/semi-blind source separation techniques, which enable the cross-information between datasets to be utilized during analysis. They provide complimentary perspectives on data fusion via different optimization assumptions. We review these methods as below:

- jICA assumes that two or more features share the same mixing coefficient matrix and maximizes the independence among joint components. It is suitable for examining the common connection between features.
- mCCA maximizes the inter-subject covariation across two sets of features and generates two linked variables, one from each dataset, i.e., canonical variants (CVs); which correlate each other only on the same indices (rows) and their corresponding correlation values are called canonical correlation coefficients (CCC). This scheme allows for common as well as distinct level of connection between two features, but the identified CVs may not be sufficiently sparse and the performance might suffer when the canonical correlation coefficients are not sufficiently distinct.
- Parallel ICA and CC-ICA both need prior information to solve a constrained optimization problem. They are semi-blind source separation approaches.
- PLS is based on the definition of a linear relationship between a dependent variable and predictor variables, and hence the goal is to determine which aspect of a set of observations (e.g., imaging data) are related directly to another set of data (e.g., experimental design, behavior). Even though PLS has some similarity to CCA in that it also finds between-set connections, PLS is based on the definition of a dependency such that one dataset is dependent on another in one aspect, e.g., the temporal signatures of the concurrent fMRI and EEG (Martinez-Montes et al., 2004). Thus, PLS is not suitable for our case, since the two tasks SM and SBP are temporally distinct so that we cannot define or assume such a dependency.

In summary, all of methods mentioned above are based on linear mixture model and maximize (1) inter-subject/direction covariation or (2) statistical independence among components, or both to connect two datasets. However, such a requirement may not be met in practice, thus the two assumptions cannot be satisfied simultaneously or without the use of any prior information; consequently, resulting in a tradeoff solution. Therefore, we propose a new joint blind source separation (BSS) model whose assumptions are less stringent and to take maximally advantage of the data at hand. In the case of fMRI, we expect to identify brain regions that are commonly activated in multiple tasks and also regions that are uniquely engaged in each task.

The scheme we propose is based on a combination of CCA and ICA, which we refer to as, “CCA+ICA”. It works on the second-level features, e.g., the contrast images of fMRI. The basic idea can be summarized as follows: first, adopt CCA to obtain two CVs, which are the estimated sources of two datasets (these may still contain mixtures of real independent components) and correlate to each other with a unique profile. Next we perform a joint ICA on the concatenated CVs and decompose the remained mixtures into joint independent components. In conclusion, CCA automatically links two datasets, compared to separately analyzing each dataset, while jICA further decomposes the remained mixtures in CVs and relaxes the constraint that CCA has on the distinctiveness of the results.

Therefore, CCA+ICA works under a less-constrained condition than mCCA, CCA or jICA and can precisely extract both shared (highly correlated) and distinct (weakly correlated) sources across features and their mixing profiles. Fig. 1 illustrates how the above mentioned four BSS methods work on two features. Note that in our method, CCA is performed differently from its application in mCCA. In particular, the CVs are generated from sources rather than mixing matrices, thus we denote it as sCCA. In this paper, we compare CCA+ICA with the other four BSS models via the simulated images under different signal to noise ratios and provide detailed analysis on their performances with respect to their assumptions.

One area that can benefit greatly from the fusion of multi-task data is the study of schizophrenia, on which the brain imaging studies have shown a wide variety of functional abnormalities. Many neuroimaging projects have now used multiple tasks to probe potential biomarkers in schizophrenia patients. Our analysis took advantage of this by extracting features from two well-known paradigms: an auditory sensorimotor task (Mattay et al., 1997) and a Sternberg working memory task (Johnson et al., 2006; Manoach et al., 1999) using GLM from 50 chronic patients with schizophrenia and 50 age-matched controls as part of the Mind Clinical Imaging Consortium (MCIC). Both tasks have been found to reveal robust activation differences in schizophrenia patients (Kim et al., 2009; Liddle et al., 2006). The performance of CCA+ICA on jointly analyzing these two features was compared with that of joint-ICA and mCCA. As expected, the three methods successfully extract different views of the data with CCA+ICA appearing to highlight both task-common and task-distinct aberrant brain regions in schizophrenia. Furthermore, via CCA+ICA, we found a significant negative correlation between activation in temporal lobe and the duration of illness (DOI) of the schizophrenia patients. In addition, CCA+ICA provides a source correlation profile, which can determine the activation similarities across tasks on brain maps. The linked aspects of activations represent joint information between tasks which is not provided in a separate analysis of each task. The benefits of our approach are demonstrated in a simulation and also via application to human fMRI data.

Materials and methods

Generative model

We assume that the multi-task dataset $\mathbf{x}_k$, $k=1, 2$, is a linear mixture of m_k sources $\mathbf{s}_k$, with a nonsingular mixing matrix $\mathbf{A}_k$ for each,

$$\mathbf{x}_k = \mathbf{A}_k \mathbf{s}_k \quad (1)$$

where $\mathbf{x}_k$ is in form of subjects by voxels, as the “feature” shown in Fig. 1. Sources $\mathbf{s}_k$ are distinct within each dataset, while $\mathbf{s}_1$ and $\mathbf{s}_2$ have high correlation only on their corresponding indices, namely

$$\begin{aligned} E \{ \mathbf{s}_1 \mathbf{s}_2^T \} \\ = \mathbf{R}_{1,2} \approx \text{diag}(r_1, r_2 \dots r_M) \quad M = \min(M_1, M_2) \end{aligned} \quad (2)$$

The correlation values $r_1, r_2 \dots r_M$ can be either common or different from each other, and likewise for the corresponding columns of $\mathbf{A}_1$ and $\mathbf{A}_2$. This assumption is more flexible compared to that of both joint ICA and mCCA. Because joint ICA requires $\mathbf{A}_1$ and $\mathbf{A}_2$ to be exactly the same in order to find the common sources; and mCCA performs well only when strict inequality holds for the correlation coefficients between columns of $\mathbf{A}_1$ and $\mathbf{A}_2$. If it is assumed that $\text{corr}\{\mathbf{A}_{1,i}, \mathbf{A}_{2,i}\} = c_i$, $i=1, 2 \dots M$, then $c_1, c_2 \dots c_M$ must be distinct enough to ensure the recovery of $\mathbf{A}_k$. Furthermore, due to the potential common correlation values among $r_1, r_2 \dots r_M$, applying individual ICA within each dataset may introduce ambiguity in feature source matching via cross-correlation. Next we will describe how CCA+ICA accomplish joint BSS under such assumptions.

Introduction of CCA+ICA

Assume there are N subjects, according to Akaike's information criterion (Akaike, 1974) or minimum description length (MDL) criterion (Rissanen, 1978), M_k number of independent components are estimated out for each dataset, $M_k < N$. Principal component analysis (PCA) is performed independently to each dataset as the first step to reduce the dimension of $\mathbf{x}_k$ from N to M_k . PCA is a typical preprocessing step in multivariate data analysis to decompose the data into a set of uncorrelated principal components (PCs) ordered by the variance of each component. For data with low contrast to noise ratio and large dimensionality, such as the fMRI feature matrix $\mathbf{x}_k$ resulted from the first-level GLM analysis), dimension reduction is usually performed based on the variance order to remove the insignificant PCs and noise. The dimension-reduced data are then whitened by normalizing each PC to unit variance. Given that the true mixing matrices are nonsingular, the whitening step can always be achieved.

Assume $\mathbf{\Lambda}_k$ and $\mathbf{B}_k$ are, respectively, the diagonal eigenvalue and eigenvector matrix of the covariance matrix $E\{\mathbf{X}_k \cdot \mathbf{X}_k^T\}$, $\mathbf{\Lambda}_k = \text{diag}\{\lambda_{k1}, \lambda_{k2}, \dots, \lambda_{kN}\}$. The top M_k eigenvectors with the largest λ_{ki} are selected, thus $\mathbf{\Lambda}$ is changed to $\mathbf{\Lambda}'_k$, an $M_k \times M_k$ diagonal matrix, and $\mathbf{B}_k$ is reduced to a corresponding $M_k \times N$ matrix $\mathbf{B}'_k$. Therefore the whitened principal components from each dataset $\mathbf{y}_k$ are given by

$$\begin{aligned}\mathbf{y}_k &= \mathbf{\Lambda}'_k{}^{-1/2} \cdot \mathbf{B}'_k \cdot \mathbf{x}_k \\ &= \mathbf{D}_k \cdot \mathbf{x}_k, k=1, 2\end{aligned}\quad (3)$$

where the $M_k \times N$ matrix $\mathbf{D}_k$ is called the whitening matrix, which ensures that the covariance matrix of $\mathbf{y}_k$ is given by the identity matrix.

CCA is then adopted to link the principal components from each dataset one by one via the canonical correlation; further (linear) transformations on $\mathbf{y}_k$ are motivated to remove the between-set cross-correlations while preserving the whiteness of the principal components within each dataset. Suppose $\mathbf{z}_k$ is the canonical variant, we aim to realize

$$\mathbf{z}_k = \mathbf{E}_k \cdot \mathbf{y}_k, E\{\mathbf{z}_k \mathbf{z}_k^T\} = \mathbf{I}, k=1, 2 \quad (4)$$

$$\begin{aligned}E\{\mathbf{z}_1 \mathbf{z}_2^T\} &= E\{\mathbf{z}_2 \mathbf{z}_1^T\} \\ &= \mathbf{L} = \text{diag}(l_1 \\ &\quad, l_2, \dots, l_M) \quad M = \min(M_1, M_2)\end{aligned}\quad (5)$$

Where $\mathbf{E}_1$ and $\mathbf{E}_2$ are two canonical transformation matrices that are calculated by solving the following eigenvalue decomposition problem

$$\mathbf{E}_1 \cdot E\{\mathbf{y}_1 \mathbf{y}_2^T\} E\{\mathbf{y}_2 \mathbf{y}_1^T\} \cdot \mathbf{E}_1^T = \mathbf{L}^2 \quad (6)$$

$$\mathbf{E}_2 = \mathbf{L}^{-1} \mathbf{E}_1 \cdot E\{\mathbf{y}_1 \mathbf{y}_2^T\} \quad (7)$$

It is easy to verify that $E\{\mathbf{z}_1 \mathbf{z}_2^T\} = E\{\mathbf{E}_1 \mathbf{y}_1 \cdot \mathbf{y}_2^T \mathbf{E}_2^T\} = \mathbf{L}$ by substituting Eqs. (6) and (7). However, this eigenvalue-based method fails to separate sources whose correlation coefficients are equal or very close, which occurs frequently in real applications of brain imaging data analysis. So $\mathbf{z}_1$ and $\mathbf{z}_2$ can be regarded as the incompletely decomposed sources whose correlation provides an approximate estimate of the linkage between two datasets.

Finally, joint ICA is performed on the concatenated CVs: $[\mathbf{z}_1, \mathbf{z}_2]$, to maximize the independence among joint components by reducing their second and higher order statistical dependencies, namely

$$[\mathbf{s}_1, \mathbf{s}_2] = \mathbf{W} \cdot [\mathbf{z}_1, \mathbf{z}_2] \quad (8)$$

ICA as a central tool for BSS has been studied extensively and many algorithms have been developed based on different patterns of cost function (Bell and Sejnowski, 1995; Cichocki et al., 2007; Comon, 1994; Hyvarinen et al., 2001). We utilized COMBI (Tichavsky et al., 2008) in our work due to its flexibility in adapting to different source distributions and its fast speed.

Thus two joint independent components (ICs) $\mathbf{s}_1$ and $\mathbf{s}_2$ are achieved, with their corresponding rows linked via a unique correlation profile. According to Eqs. (1), (3), (4) and (8), the proposed CCA+ICA scheme can be summarized as

$$\begin{aligned} \mathbf{s}_k &= (\mathbf{W} \mathbf{E}_k \mathbf{D}_k) \cdot \mathbf{x}_k \\ &= \mathbf{W} \cdot \mathbf{x}_k \\ \mathbf{A}_k &= (\mathbf{W} \mathbf{E}_k \mathbf{D}_k)^{-1}, k=1, 2 \end{aligned} \quad (9)$$

We then correlate $\mathbf{s}_1$ with $\mathbf{s}_2$, generating a correlation profile, i.e. $r_1, r_2 \dots r_M$ in Eq. (2), which could contain identical values and are sorted from high to low. Namely, IC 1 has the highest $\mathbf{s}_1$ – $\mathbf{s}_2$ correlation (task shared source), and IC M has the least similarity between $\mathbf{s}_1$ and $\mathbf{s}_2$. Note in this paper, a similar sorting is performed by mCCA, where correlation coefficients $c_1, c_2 \dots c_M$ are computed between the columns of $\mathbf{A}_1$ and $\mathbf{A}_2$.

Synthetic image data

By simulating two groups of image sources as two features, we investigate the joint BSS performance of CCA+ICA on simulated data and compare it to that of joint ICA, mCCA, sCCA and separate ICA on two datasets. As shown in Fig. 1, six image sources in dimension of 256×256 are simulated for each feature as true sources $\mathbf{S}_1$ and $\mathbf{S}_2$ (in dimension of $6 \times 65,536$); the correlations between their corresponding rows are set from high to low with increase of the source index, namely $[0.99 \ 0.88 \ 0.77 \ 0.45 \ 0.38 \ -0.004]$, representing both shared and distinct parts of two features. Further, the mixing matrices of each feature, i.e. $\mathbf{A}_1$ and $\mathbf{A}_2$ (in dimension of 20×6), have diverse correlations between their corresponding columns too, as listed in Table 1, $\mathbf{A}_1$ – $\mathbf{A}_2$ correlation = $[0.99 \ 0.78 \ 0.68 \ 0.59 \ 0.22 \ 0.88]$. Therefore, 20 noisy mixed images $\mathbf{X}_k$ are generated for each feature according to $\mathbf{X}_k = \mathbf{I}_k + \mathbf{N}_k = \mathbf{A}_k \mathbf{S}_k + \mathbf{N}_k$, $k=1,2$. We considered a wide range of noisy conditions by adding different random Gaussian noise $\mathbf{N}_k$ to every pure signal mixtures $\mathbf{I}_k$, resulting in 11 noisy cases with the mean peak signal-to-noise

ratios (PSNR) in range of $[-1\ 33]$ dB. The PSNR is a most commonly used measure of image quality after corruption or recovery, which in our case is defined as Eq. (10) for the j th mixed 8-bit images of feature k , with total l pixels. Typical PSNR value for the acceptable image quality is about 30 dB; the lower the value, the more degraded the image (Thomos et al., 2006).

$$\begin{aligned} \text{PSNR}(k, j) &= 10 \log_{10} \left[\frac{(2^{\text{bit}} - 1)^2}{\frac{1}{l} \sum_{i=1}^l |\mathbf{X}_k(i) - \mathbf{I}_k(i)|^2} \right] \\ &= 20 \log_{10} \left[\frac{255}{\frac{1}{l} \sum_{i=1}^l |\mathbf{X}_k(i) - \mathbf{I}_k(i)|^2} \right] \end{aligned} \quad (10)$$

$$j = 1, 2, \dots, 20, l = 256 \times 256 = 65536, k = 1, 2, \text{bit} = 8$$

Four joint BSS models plus separate ICA are implemented on the two datasets respectively under every PSNR for 10 runs. The decomposed components are paired with the true sources via cross-correlation automatically within each feature. Except two-separate-ICA, all models already auto-link the components between features via their assumptions. We adopted two metrics to estimate the joint BSS performance of every method. One is the average correlations of the $6 \times 2 = 12$ estimated source components with the ground truth; another is the inter-symbol interference (ISI) (Amari et al., 1996; Macchi and Moreau, 1995), which is defined below as d , where $\mathbf{W}$ is the demixing matrix used to recover the estimates of sources and $\mathbf{W}_0 = (\mathbf{A})^{-1}$ is the inverse of the true mixing matrix. ISI is invariant to permutation and scaling of the columns of $\mathbf{W}$ ($\mathbf{W}_0$). Its value is always in range of $[0, 1]$ and is equal to zero if and only if $\mathbf{W}_0$ and $\mathbf{W}$ represent the same components.

$$\begin{aligned} \text{ISI}(\mathbf{W}_0, \mathbf{W}) &= \frac{1}{2M(M-1)} \\ &\times \left[\sum_{i=1}^M \left(\frac{\sum_{j=1}^M |a_{ij}|}{\max_j |a_{ij}|} - 1 \right) + \sum_{j=1}^M \left(\frac{\sum_{i=1}^M |a_{ij}|}{\max_i |a_{ij}|} - 1 \right) \right], \\ a_{ij} &= (\mathbf{W}_0 \mathbf{W}^{-1})_{ij} \end{aligned} \quad (11)$$

Human fMRI data

Participants—Fifty chronic schizophrenia patients (SZ, age 36 ± 11 , 10 females) and with 50 matched healthy controls (HC, age 34 ± 12 , 14 females) provided written informed consent for the MIND Clinical Imaging Consortium. They are selected from three sites: the Universities of New Mexico (UNM), Minnesota (MINN) and Iowa (IOWA) based on their age, sex and accuracy of performing the tasks, and represent a matched subgroup of the larger Mind multi-site study. All patients were stabilized on medication prior to the fMRI scan session in this study. Wide range achievement test (WRAT) (Cochran and Pedrini, 1969) scores showed significant IQ differences between patients and controls (mean HC = 51, mean SZ = 48, $p = 0.015$), which are expected due to debilitating cognitive effects of schizophrenia and curtailed education due to illness. Patients and controls were age matched, thus there were no significant differences ($p = 0.38$) between the two groups on this measure. The duration of illness

in patients was in range of 0.75–42 years, with a mean of 14 years. For the Sternberg task, both groups averaged greater than 95% in the overall accuracy of probe responses. The auditory sensorimotor task was a block design and we verified during and after data collection that subjects were performing the task.

Sternberg working memory task (SB) requires subjects to memorize a list of digits (displayed simultaneously) and following a delay, to identify if a ‘probe’ digit was in the list. Three working memory loads high (five digits), medium (three digits) and low (one digit) were used in this paradigm. Each run contained two blocks of each of the three loads in a pseudorandom order. Three runs of the Sternberg task were performed. Subjects were asked to respond with their right thumb if the probe digit was a target (digits previously displayed) and with their left thumb for a foil. We use the averaged contrast images of the three loads in this paper.

Sensorimotor task (SM) consisted of an on/off block design, each with a duration of 16 s. During the on-block cycles of eight ascending-pitched and eight descending-pitched, 200 ms tones were presented. There were three runs each with duration of 4 min. The participant was instructed to press the right thumb of the input device after each tone was presented.

Imaging parameters—All sites, except for UNM, utilized a Siemens 3 T Trio Scanner, while UNM utilized a Siemens 1.5 T Sonata. The scanners were equipped with a 40 mT/m gradient and a standard quadrature head coil. The fMRI pulse sequence parameters were identical for both tasks as following single-shot echo planar imaging (EPI); scan plane=oblique axial (AC–PC); time to repeat (TR)=2 s; echo time (TE)=30 ms (for UNM, TE=40 ms); field of view (FOV)= 22 cm, matrix = 64×64 ; flip angle = 90° ; voxel size = $3.4 \times 3.4 \times 4 \text{ mm}^3$; slice thickness=4 mm; slice-gap=1 mm; number of slices=27; slice acquisition=ascending.

fMRI preprocessing data were preprocessed using the software package SPM5. Images were realigned using INRIalign, a motion correction algorithm unbiased by local signal changes (Freire et al., 2002). Data were spatially normalized into the standard MNI space (Friston et al., 1995), smoothed with a 9 mm^3 full width at half-maximum Gaussian kernel. The data, originally $3.75 \times 3.75 \times 4 \text{ mm}$, were slightly subsampled to $3 \times 3 \times 3 \text{ mm}$, resulting in $53 \times 63 \times 46$ voxels.

Feature extraction—A GLM approach using SPM5 was used to find task-associated brain regions, labeled as contrast maps which were then used as features within our CCA+ICA analysis. Specifically, a separate GLM analysis was performed for each task (SB, SM) and consisted of a univariate multiple regression of each voxel’s time-course with an experimental design matrix, generated by the convolution of the task onset times with a hemodynamic response function. This resulted in a set of beta-weight maps associated with each parametric regressor for each task. The subtraction of one beta-weight map with another is often referred to as a contrast map, which represents the effect of a task in relation to an experimental baseline. For demonstration of our method we utilize the average probe effect for the SB task and the tapping effect for the SM task.

Results

Simulation

Fig. 2 shows the two performance metrics: mean source correlations and ISI, of the four joint BSS methods under different PSNRs. The proposed scheme CCA+ICA outperforms other methods in all noisy conditions. Note that when PSNR=−1 dB, i.e., noise is bigger than signal, all four methods can still have the mean source correlations higher than 0.5. In order to investigate the BSS performance of all methods on every source which is simulated based on diverse assumptions, we focus on one case (PSNR=10 dB) and list all measurements for this

case in Table 1. Fig. 3 displays part of the mixed noisy images, true sources and the corresponding estimated results from all models. Specifically, the between-feature assumptions on $\mathbf{A}$ and $\mathbf{S}$ are as follows:

- Source 1: $\mathbf{A}_1$ – $\mathbf{A}_2$ correlation=0.988, $\mathbf{S}_1$ – $\mathbf{S}_2$ correlation=0.987; two features share almost the same source and mixing matrix.
- Source 6: $\mathbf{A}_1$ – $\mathbf{A}_2$ correlation=0.881, $\mathbf{S}_1$ – $\mathbf{S}_2$ correlation=0.004; two features share very similar mixing matrix, but the sources are quite different.
- Source 2: $\mathbf{A}_1$ – $\mathbf{A}_2$ correlation=0.781, $\mathbf{S}_1$ – $\mathbf{S}_2$ correlation=0.88; two features share very similar source as well as the mixing matrix.
- Sources 3–5: Two features have the sources and mixing matrices with variant degree of dissimilarities. The correlation values may not be distinct enough between each other.
- Source 5 in feature 1 has higher correlation with source 6 than source 5 in feature 2 (0.42>0.38).

Table 1 compares five models and indicates that CCA+ICA and 2-separate-ICA outperform other methods on both metrics (shown in yellow background), whereas the latter may introduce ambiguity on between-feature source match via cross-correlation, especially when the estimated component number is high. Here from indexes 1 to 6, the estimated between-feature source correlations are [0.97 0.88 0.67 0.47 0.45 –0.02] for CCA+ICA and [0.97 0.87 0.69 0.48 0.41 0.006] for 2-separate-ICA, which is very similar to the truth [0.99 0.88 0.77 0.45 0.38 –0.004]; however, source 5 of feature 1 is misconnected to source 6 of feature 2 in separate ICA (red background), since $\text{corr}(S_{1,5}, S_{2,6}) = 0.41 > \text{corr}(S_{1,5}, S_{2,5}) = 0.35$. Moreover, sources 1, 2 and 6 have high $\mathbf{A}_1$ – $\mathbf{A}_2$ correlation values, thus joint ICA works well, in accordance with its hypothesis; consequently, the performance of mCCA degrades in sources 3–6 due to the requirement of sufficiently distinct mixing profile correlations. When the between-feature source correlations are close (e.g., sources 3, 4, green background), the proposed CCA+ICA mitigates the performance deficits occurred by using only sCCA. In summary, CCA+ICA succeeds in separating sources accurately and linking them correctly in a less-constrained condition, where both sources and mixing profiles can have either common or distinct between-dataset correlation.

Multi-task fusion of fMRI data

Three joint BSS models: CCA+ICA, Joint ICA and multimodal CCA are applied to two fMRI features: SM and SB, collected from 100 subjects, with the goal to find the aberrant functional brain activity in schizophrenia and the relationships between these areas and duration of illness. Sixteen components were estimated for each feature according to an improved MDL criterion (Li et al., 2007). Before starting the analysis, each feature is first reshaped into a matrix with dimensions of $100 \times [\text{number of voxels}]$, divided by its standard derivation. We analyzed the results in the following three aspects.

Group differences

For every component and every task, we tested the variance of loading parameters for the 50 controls and 50 patients (i.e., each column of $\mathbf{A}_k$, $k=1,2$) and found that most ICs had unequal variances between groups. Therefore, two sample t -tests with unequal variances are performed between each column of $\mathbf{A}_1$ and $\mathbf{A}_2$. Note that in joint ICA, $\mathbf{A}_1 = \mathbf{A}_2$; whereas in CCA+ICA and mCCA, $\mathbf{A}_1$ is different from $\mathbf{A}_2$, we identified indices of the significant components for each task among 12 or 13 non-artifactual components ($p < 0.05$, usually three or four ICs are shown as artifacts among 16), resulting in ind_1 and ind_2 . For example, if $\text{ind}_1 = [6, 8]$ and $\text{ind}_2 = [8, 11]$, then the 8th component is significant for both tasks, thus it is referred as “task-

common” group-discriminative component (green frame as shown in Fig. 4); the residual significant components (the 6th IC in task1 and the 11th IC in task2) are denoted as “task-distinct” (pink frame in Fig. 4). For the data we analyzed in this paper, mCCA happened to have only “task-common” significant component and did not reveal “task-distinct” information; while it may generate such information in other cases. As shown in Fig. 4 and Table 2, five p values pass the false discovery rate (FDR) correction and another three passed at an uncorrected $p < 0.05$.

The Talairach labels of the identified group-discriminative components are listed in Table 2. It is indicated that both SM and SB tasks show strong activations on primary somatosensory cortex (Brodmann Area [BA] 3 1 2) and the motor cortex (BA 4 6) in CCA+ICA as well as mCCA. Furthermore, CCA+ICA along with joint ICA detect common brain activations in superior and transverse temporal gyrus for both features. Finally, results from all three models indicate that significant group differences are reflected from either SM or SB task in superior/inferior parietal lobule (BA 7 40) and the frontal gyrus (BA 6 8 10 32).

Study on duration of illness

We are interested in exploring the relationship of the fMRI changes to the duration of illness (DOI) in schizophrenia. To achieve this, for each task and each algorithm, we correlate the patients’ loadings of every meaningful component (exclude the components showing artifacts) with their DOI, and find that only one component derived from CCA+ICA in the SM feature shows a significant negative correlation ($p < 0.05$, FDR corrected for multiple comparisons), as illustrated in Fig. 5. As DOI is also correlated with age, and the patients’ ages are found to be correlated with the same loadings (although less so than DOI), we evaluated the correlation of controls’ ages to their loadings of the same component. We did not find a significant correlation, which suggests the association is primarily driven by DOI and not by age.

Verification of assumptions for each model

In addition to decomposing data within each feature, mCCA and jICA link two tasks via their mixing matrices, while CCA+ICA associates two tasks by providing a correlation vector on the corresponding spatial maps. The top three most similarly activated components between SM and SB tasks are illustrated in Fig. 6a, i.e., artifacts caused by movement, the default mode network and visual cortex.

Moreover, in order to compare the effects caused by different assumptions, from each method, we pick one paired components with strong activations in temporal lobe shown in SM feature. The reason we choose such a joint component is that the sensorimotor paradigm is a simple yet effective auditory task that implies strong activations in superior temporal gyrus (STG), which is a region that has been implicated in a number of studies of schizophrenia. As seen from Fig. 6b, different spatial maps are shown in SB feature based on the model assumption, though their paired SM components share quite similar activations in temporal lobe.

Note that we also performed a separate ICA for each task individually using 16 components, and matched the results with those from CCA+ICA via cross-correlation, the averaged source similarity is 0.81. However, when linking the sources between tasks, four pairs are mismatched in separate ICA compared to CCA+ICA.

Discussion

We propose a simple yet effective new joint blind source separation model, ‘CCA+ICA’, for multi-task brain imaging data fusion. This model can achieve both high estimation accuracy and the correct source linking between tasks, as shown in the simulation. Based on a second-

level fMRI analysis framework, we apply the proposed model to the contrast images derived from two tasks: sensorimotor and Sternberg working memory task, both of which were collected from 50 patients with schizophrenia and 50 age-matched healthy controls. We explore the joint as well as distinct information existing between the two tasks, in order to discover abnormal regions that can best discriminate patients from controls. To examine results deriving from various fusion methods and assumptions, we compare the proposed scheme with several other joint BSS models in both simulated and real multi-task fMRI data.

One of the major strengths of this model is that the combination of sCCA and joint ICA improves the performance of joint BSS significantly under a less-constrained condition. As seen in Table 1, both mCCA and sCCA require a strict distinction between the canonical correlation coefficients (though mCCA on mixing matrix and sCCA on source maps), which, even if too close in values, may degrade the BSS performance seriously. By contrast, the CCA+ICA method is not constrained by this requirement, i.e., different components can have the same CCC as there is a subsequent ICA step to follow. Moreover, joint ICA and mCCA seek similarities in the component mixing profile between features, without considering the source maps. Hence, when their assumptions are not true, the joint component maps can suffer from ambiguity and misinterpretation, as displayed in Fig. 3, e.g., source 4 in joint ICA and the source 3 in mCCA, where mCCA leads to spatial maps that are not very sparse and hence not easy to interpret as in the results of an ICA algorithm. On the other hand, our method focuses on source maps, aiming to explore both similarity and discrepancy between the feature maps, and can also recover the mixing matrix precisely, see ISI in Fig. 2 and Table 1. Finally, compared to other ICA-based methods, the CCA+ICA method does not increase the computational load appreciably; however it achieves the best performance under very flexible conditions. This is likely because sCCA automatically links two datasets with respect to two-separate-ICA discarding the joint information, while joint ICA further decomposes the remained mixtures in canonical variants and relaxes the constraint that CCC should be dissimilar enough. This constraint is not easy to be satisfied when the component number is large, e.g. if $M=20$ the mean distance between CCC is smaller than 0.05.

A chief purpose of multi-task fMRI data fusion is to access the joint information provided by multiple tasks, which in turn can be useful for identifying dysfunctional regions implicated in brain disorders. As we mentioned above, detecting group difference via a two sample *t*-test on the component loadings enables more flexibility in mCCA and CCA+ICA, resulting in either task-common (similar to joint component in jICA) or task-specific group-discriminating components. In Fig. 4, all models depict a set of well known regions that have been previously implicated in schizophrenia, including STG, prefrontal cortex, sensorimotor cortex and inferior parietal lobule (IPL) (Shenton et al., 2001), but also with heterogeneity in the spatial variability of the maps. We highlight a few of the key findings below:

- It is notable that a joint component engaged in both CCA+ICA and mCCA shows activity in the motor cortex [BA 1–4, 6], compatible with both tasks in requiring subjects to push buttons. And the somatosensory cortex has shown differences in schizophrenia patients versus controls in both sMRI and fMRI studies (Schroder et al., 1999; Thoma et al., 2007).
- In addition, the superior temporal gyrus plays a prominent role in schizophrenia. It is the most group-discriminating region for healthy controls versus schizophrenia patients in auditory tasks such as sensorimotor paradigm (Demirci et al., 2008; Sui et al., 2009a) and its dysfunction has been related to the auditory hallucinations that are common in schizophrenia patients (Barta et al., 1990; Calhoun et al., 2004). It is encouraging to see that the results derived from both joint ICA and CCA+ICA engage the significant group difference in this area selectively in SM feature.

- Furthermore, prefrontal dysfunction in schizophrenia is well-delineated in many fMRI and structural MRI studies using a variety of approaches (Goldman-Rakic and Selemon, 1997; Johnson et al., 2006; Kim et al., 2009; Manoach et al., 1999). There is considerable evidence that the dorsolateral prefrontal cortex (DLPFC, BA 6, 9 and 46) is tightly linked to the “on-line” maintenance of working memory items (Goldman-Rakic, 1994). As we expected, this region is represented in significant components from all three models in the SB task, implicating aberrant processing in high-level executive functions and memory retrieval in schizophrenia.
- In addition, the IPL (BA 7 and 40) which is involved in multiple functions including attention, sensory integration (Andersen and Buneo, 2002; Assaf et al., 2006) and decision making (McClure et al., 2004) is identified in both tasks from all methods, especially significantly in Sternberg working memory, since IPL has been known to be intricately connected to the DLPFC (Cole and Schneider, 2007; Pearlson et al., 1996).

Therefore, in differentiating two groups, the components derived from CCA+ICA implicate a varied but well defined set of brain regions that have often been linked to schizophrenia, and the feature-specific component reveals the previously reported, task-associated dysfunctional regions in schizophrenia. Our results thus appear valid in elucidating the pathophysiology of schizophrenia and this complementary method may aid in search for meaningful biological markers and cross-information in future multi-task neuroimaging studies. Future studies utilizing the CCA+ICA approach in conditions such as bipolar affective disorder will be important for assessing the specificity of these dysfunctional regions.

In comparing spatial activation with illness duration in the patients, only one component with meaningful activations from CCA+ICA showed a significant negative correlation ($r=-0.33$, $p=0.0105$, FDR corrected), suggesting that the activations of STG, (a key region that is both task-related as well as compromised in schizophrenia), decrease with longer DOI in schizophrenia.

When determining similarity between tasks, the proposed model generates a correlation profile of the corresponding spatial maps between tasks, and a dividing line is needed to separate “similar” and “distinct”. We utilized an information theoretic criteria, the minimum description length (MDL) (Li et al., 2007), to select the cutoff quantitatively. Three ICs are selected out via the MDL criterion and all of them are noted to have high between-task source correlation ($r>0.7$). Results in Fig. 6a confirmed the hypotheses of CCA+ICA, as we found the top two most correlated joint components are intrinsically similar brain networks regardless of task, i.e. ventricle and movement-related artifacts, and the default mode network, which latter is defined as a baseline condition of brain function and is also implicated in schizophrenia (Garrity et al., 2007; Raichle et al., 2001). Primary visual cortex is engaged in the third most correlated component pair, implicating great similarities in visual activations for both task-related and non task-related features, in agreement with (Calhoun et al., 2008; Smith et al., 2009).

Fig. 6b provides an illustration of data fusion via different optimization assumptions, reflecting complimentary perspectives in interpreting multi-task data. CCA+ICA seek similarities in spatial maps; thus both features share STG regions but also with some variability. Joint ICA suggests that the STG in SM feature and the DLPFC in SB feature have very similar modulation across subjects. In mCCA, despite the different positive activations, both features represent negative activations of the default mode network, resulting in a high correlation on modulation of 0.82, supporting the idea that this network is not singular, but a conglomeration of multiple subnetworks that work in conjunction with one another (Uddin et al., 2009). Hence it is important to choose an appropriate fusion method according to the purpose of the analysis, to help better elucidate the relationship between multi-task datasets.

We would like to address some additional aspects of our study. First, we performed PCA for each feature to reduce the dimensionality and obtain principal components, and then applied CCA to the principal components to obtain M canonical variants, on which the joint ICA was implemented. For the purpose of obtaining unique solution, we chose $M = \min(M_1, M_2)$, where M_1 and M_2 were an estimated number of independent components for each feature. We note that if we fix M and change the number of PCs, the decomposition results can vary; and with the increase of the number of PCs, greater resemblance between tasks is detected. Second, our method currently is only suitable for fusion of two same-type datasets, both of which need to have the same dimension on maps (voxels/signals) after preprocessing. However, the data type can be diverse; besides fMRI, EEG, MEG, NIRS and genetic data can be analyzed as well.

In addition to working with multi-task brain imaging data, CCA+ICA can be used to examine group similarities, e.g. schizophrenia and bipolar patients, or to investigate similar networks between task-related and resting-state brain imaging data. Similar studies have been done by many other groups (Curtis et al., 2001; Hill et al., 2008; Smith et al., 2009) through inspecting associations of two datasets (via a two sample t -test or cross-correlation) after individually processing each separately, thus arguing for the widespread utilization of the CCA+ICA method. Future work might include improvement and modification of the model (e.g., using mCCA+jICA) to carry out the fusion analysis of multimodal brain imaging data.

In conclusion, the proposed “CCA+ICA” model takes advantage of two multivariate analyses to guide the joint source extraction. It enables more flexibility on the statistical assumptions and can serve as a valuable alternative to multi-task fusion framework in the neuroimaging community. We find a striking consistency in the identified group-discriminating components compared to previous reports; a correlation study on the duration of illness of schizophrenia further confirms its effectiveness in data mining. The strong performance of CCA+ICA on joint decomposition suggests wider applications to neuroimaging studies and may aid in identifying potential brain illness biomarkers in the future.

Acknowledgments

This work was supported by the National Institutes of Health grants 1 R01 EB 006841 and 1 R01 EB 005846 (to Vince D. Calhoun) and MH43775, MH074797 and MH077945 (to Godfrey D. Pearlson). We thank the research staff at the University of New Mexico, Minnesota, Iowa and the Mind Research Network who helped collect and process the data. We also appreciate the valuable advice given by Jingyu Liu at the Mind Research Network and Yi-Ou Li at University of California, San Francisco.

References

- Akaike H. A new look at the statistical model identification. *IEEE Trans Automat Contr* 1974;19 (6): 716–723.
- Amari, S.; Cichocki, A.; Yang, HH. *Advances in Neural Information Processing Systems*. MIT Press; Cambridge, MA: 1996. A New Learning Algorithm for Blind Signal Separation.
- Andersen RA, Buneo CA. Intentional maps in posterior parietal cortex. *Annu Rev Neurosci* 2002;25:189–220. [PubMed: 12052908]
- Assaf M, Rivkin PR, Kuzu CH, Calhoun VD, Kraut MA, Groth KM, Yassa MA, Hart J Jr, Pearlson GD. Abnormal object recall and anterior cingulate overactivation correlate with formal thought disorder in schizophrenia. *Biol Psychiatry* 2006;59 (5):452–459. [PubMed: 16199012]
- Barta PE, Pearlson GD, Powers RE, Richards SS, Tune LE. Auditory hallucinations and smaller superior temporal gyrus volume in schizophrenia. *Am J Psychiatry* 1990;147 (11):1457–1462. [PubMed: 2221156]
- Beckmann CF, Smith SM. Tensorial extensions of independent component analysis for multisubject fMRI analysis. *NeuroImage* 2005;25 (1):294–311. [PubMed: 15734364]

- Bell AJ, Sejnowski TJ. An information-maximization approach to blind separation and blind deconvolution. *Neural Comput* 1995;7 (6):1129–1159. [PubMed: 7584893]
- Blum AL, Langley P. Selection of relevant features and examples in machine learning. *Artif Intell* 1997;97 (1–2):245–271.
- Bullmore E, Horwitz B, Honey G, Brammer M, Williams S, Sharma T. How good is good enough in path analysis of fMRI data? *NeuroImage* 2000;11 (4):289–301. [PubMed: 10725185]
- Calhoun VD, Adali T. Feature-based fusion of medical imaging data. *IEEE Trans Inf Technol Biomed* 2009;13 (5):711–720. [PubMed: 19273016]
- Calhoun VD, Adali T, Pearlson GD, Pekar JJ. A method for making group inferences from functional MRI data using independent component analysis. *Hum Brain Mapp* 2001;14 (3):140–151. [PubMed: 11559959]
- Calhoun VD, Kiehl KA, Liddle PF, Pearlson GD. Aberrant localization of synchronous hemodynamic activity in auditory cortex reliably characterizes schizophrenia. *Biol Psychiatry* 2004;55 (8):842–849. [PubMed: 15050866]
- Calhoun VD, Adali T, Kiehl KA, Astur R, Pekar JJ, Pearlson GD. A method for multitask fMRI data fusion applied to schizophrenia. *Hum Brain Mapp* 2006;27 (7):598–610. [PubMed: 16342150]
- Calhoun VD, Kiehl KA, Pearlson GD. Modulation of temporally coherent brain networks estimated using ICA at rest and during cognitive tasks. *Hum Brain Mapp* 2008;29 (7):828–838. [PubMed: 18438867]
- Cichocki, A.; Amari, SSK.; Tanaka, T.; AHP. ICALAB Toolboxes. 2007.
- Cochran ML, Pedrini DT. The concurrent validity of the 1965 WRAT with adult retardates. *Am J Ment Defic* 1969;73 (4):654–656. [PubMed: 5762263]
- Cole MW, Schneider W. The cognitive control network: integrated cortical regions with dissociable functions. *NeuroImage* 2007;37 (1):343–360. [PubMed: 17553704]
- Comon P. Independent component analysis, a new concept ? *Signal Processing, Special issue on Higher-Order Statistics* 1994;36 (3):287–314.
- Correa N, Adali T, Calhoun VD. Performance of blind source separation algorithms for fMRI analysis using a group ICA method. *Magn Reson Imaging* 2007;25 (5):684–694. [PubMed: 17540281]
- Curtis VA, Dixon TA, Morris RG, Bullmore ET, Brammer MJ, Williams SC, Sharma T, Murray RM, McGuire PK. Differential frontal activation in schizophrenia and bipolar illness during verbal fluency. *J Affect Disord* 2001;66 (2–3):111–121. [PubMed: 11578663]
- Demirci O, Clark VP, Magnotta VA, Andreasen NC, Lauriello J, Kiehl KA, Pearlson GD, Calhoun DC. A review of challenges in the use of fMRI for disease classification/characterization and a projection pursuit application from multi-site fMRI schizophrenia study. *Brain Imaging and Behavior* 2008;2 (3):207–226.
- Ecker C, Brammer MJ, Williams SC. Combining path analysis with time-resolved functional magnetic resonance imaging: the neurocognitive network underlying mental rotation. *J Cogn Neurosci* 2008;20 (6):1003–1020. [PubMed: 18211236]
- Esposito F, Scarabino T, Hyvarinen A, Himberg J, Formisano E, Comani S, Tedeschi G, Goebel R, Seifritz E, Di Salle F. Independent component analysis of fMRI group studies by self-organizing clustering. *NeuroImage* 2005;25 (1):193–205. [PubMed: 15734355]
- Freire L, Roche A, Mangin JF. What is the best similarity measure for motion correction in fMRI time series? *IEEE Trans Med Imaging* 2002;21 (5):470–484. [PubMed: 12071618]
- Friston KJ, Ashburner J, Frith CD, Poline JP, Heather JD, Frackowiak RS. Spatial registration and normalization of images. *Hum Brain Mapp* 1995;2:165–189.
- Garrity AG, Pearlson GD, McKiernan K, Lloyd D, Kiehl KA, Calhoun VD. Aberrant “default mode” functional connectivity in schizophrenia. *Am J Psychiatry* 2007;164 (3):450–457. [PubMed: 17329470]
- Goldman-Rakic PS. Working memory dysfunction in schizophrenia. *J Neuropsychiatry Clin Neurosci* 1994;6 (4):348–357. [PubMed: 7841806]
- Goldman-Rakic PS, Selemon LD. Functional and anatomical aspects of prefrontal pathology in schizophrenia. *Schizophr Bull* 1997;23 (3):437–458. [PubMed: 9327508]
- Grady CL, Springer MV, Hongwanishkul D, McIntosh AR, Winocur G. Age-related changes in brain activity across the adult lifespan. *J Cogn Neurosci* 2006;18 (2):227–241. [PubMed: 16494683]

- Guo Y, Pagnoni G. A unified framework for group independent component analysis for multi-subject fMRI data. *NeuroImage* 2008;42 (3):1078–1093. [PubMed: 18650105]
- Hill SK, Harris MS, Herbener ES, Pavuluri M, Sweeney JA. Neurocognitive allied phenotypes for schizophrenia and bipolar disorder. *Schizophr Bull* 2008;34 (4):743–759. [PubMed: 18448479]
- Hyvarinen, A.; Karhunen, J.; Oja, E. Independent Component Analysis. John Wiley & Sons; 2001.
- Johnson MR, Morris NA, Astur RS, Calhoun VD, Mathalon DH, Kiehl KA, Pearlson GD. A functional magnetic resonance imaging study of working memory abnormalities in schizophrenia. *Biol Psychiatry* 2006;60 (1):11–21. [PubMed: 16503328]
- Kim DI, Manoach DS, Mathalon DH, Turner JA, Mannell M, Brown GG, Ford JM, Gollub RL, White T, Wible C, et al. Dysregulation of working memory and default-mode networks in schizophrenia using independent component analysis, an fBIRN and MCIC study. *Hum Brain Mapp*. 2009
- Lee JH, Lee TW, Jolesz FA, Yoo SS. Independent vector analysis (IVA): multivariate approach for fMRI group study. *NeuroImage* 2008;40 (1):86–109. [PubMed: 18165105]
- Li YO, Adali T, Calhoun VD. Estimating the number of independent components for functional magnetic resonance imaging data. *Hum Brain Mapp* 2007;28 (11):1251–1266. [PubMed: 17274023]
- Liddle PF, Laurens KR, Kiehl KA, Ngan ET. Abnormal function of the brain system supporting motivated attention in medicated patients with schizophrenia: an fMRI study. *Psychol Med* 2006;36 (8):1097–1108. [PubMed: 16650349]
- Lin FH, McIntosh AR, Agnew JA, Eden GF, Zeffiro TA, Belliveau JW. Multivariate analysis of neuronal interactions in the generalized partial least squares framework: simulations and empirical studies. *NeuroImage* 2003;20 (2):625–642. [PubMed: 14568440]
- Liu J, Pearlson G, Windemuth A, Ruano G, Perrone-Bizzozero NI, Calhoun V. 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]
- Macchi, O.; Moreau, E. Self-adaptive Source Separation by Direct or Recursive Networks. Cyprus; Limasol: 1995. p. 122–129.
- Manoach DS, Press DZ, Thangaraj V, Searl MM, Goff DC, Halpern E, Saper CB, Warach S. Schizophrenic subjects activate dorsolateral prefrontal cortex during a working memory task, as measured by fMRI. *Biol Psychiatry* 1999;45 (9):1128–1137. [PubMed: 10331104]
- Martinez-Montes E, Valdes-Sosa PA, Miwakeichi F, Goldman RI, Cohen MS. Concurrent EEG/fMRI analysis by multiway Partial Least Squares. *NeuroImage* 2004;22 (3):1023–1034. [PubMed: 15219575]
- Mattay VS, Callicott JH, Bertolino A, Santha AK, Tallent KA, Goldberg TE, Frank JA, Weinberger DR. Abnormal functional lateralization of the sensorimotor cortex in patients with schizophrenia. *NeuroReport* 1997;8 (13):2977–2984. [PubMed: 9376542]
- McClure SM, Laibson DI, Loewenstein G, Cohen JD. Separate neural systems value immediate and delayed monetary rewards. *Science* 2004;306 (5695):503–507. [PubMed: 15486304]
- McIntosh AR, Bookstein FL, Haxby JV, Grady CL. Spatial pattern analysis of functional brain images using partial least squares. *NeuroImage* 1996;3 (3 Pt 1):143–157. [PubMed: 9345485]
- McKeown MJ, Li J, Huang X, Lewis MM, Rhee S, Young Truong KN, Wang ZJ. Local linear discriminant analysis (LLDA) for group and region of interest (ROI)-based fMRI analysis. *NeuroImage* 2007;37 (3):855–865. [PubMed: 17627850]
- Oakes TR, Fox AS, Johnstone T, Chung MK, Kalin N, Davidson RJ. Integrating VBM into the General Linear Model with voxelwise anatomical covariates. *NeuroImage* 2007;34 (2):500–508. [PubMed: 17113790]
- Pearlson GD, Petty RG, Ross CA, Tien AY. Schizophrenia: a disease of heteromodal association cortex? *Neuropsychopharmacology* 1996;14 (1):1–17. [PubMed: 8719025]
- Raichle ME, MacLeod AM, Snyder AZ, Powers WJ, Gusnard DA, Shulman GL. A default mode of brain function. *Proc Natl Acad Sci U S A* 2001;98 (2):676–682. [PubMed: 11209064]
- Rajapakse JC, Kruggel F, Maisog JM, von Cramon DY. Modeling hemodynamic response for analysis of functional MRI time-series. *Hum Brain Mapp* 1998;6 (4):283–300. [PubMed: 9704266]
- Rissanen J. Modeling by shortest data description. *Automatica* 1978;14 (5):465–471.

- Schlosser R, Gesierich T, Kaufmann B, Vucurevic G, Hunsche S, Gawehn J, Stoeter P. Altered effective connectivity during working memory performance in schizophrenia: a study with fMRI and structural equation modeling. *NeuroImage* 2003;19 (3):751–763. [PubMed: 12880804]
- Schroder J, Essig M, Baudendistel K, Jahn T, Gerdson I, Stockert A, Schad LR, Knopp MV. Motor dysfunction and sensorimotor cortex activation changes in schizophrenia: a study with functional magnetic resonance imaging. *NeuroImage* 1999;9 (1):81–87. [PubMed: 9918729]
- Shenton ME, Dickey CC, Frumin M, McCarley RW. A review of MRI findings in schizophrenia. *Schizophr Res* 2001;49 (1–2):1–52. [PubMed: 11343862]
- Smith SM, Fox PT, Miller KL, Glahn DC, Fox PM, Mackay CE, Filippini N, Watkins KE, Toro R, Laird AR, et al. Correspondence of the brain's functional architecture during activation and rest. *Proc Natl Acad Sci U S A* 2009;106 (31):13040–13045. [PubMed: 19620724]
- Sui J, Adali T, Pearlson GD, Calhoun VD. An ICA-based method for the identification of optimal fMRI features and components using combined group-discriminative techniques. *NeuroImage* 2009a;46 (1):73–86. [PubMed: 19457398]
- Sui J, Adali T, Pearlson GD, Clark VP, Calhoun VD. A method for accurate group difference detection by constraining the mixing coefficients in an ICA framework. *Hum Brain Mapp* 2009b;30(9)
- Thoma RJ, Hanlon FM, Huang M, Miller GA, Moses SN, Weisend MP, Jones A, Paulson KM, Irwin J, Canive JM. Impaired secondary somatosensory gating in patients with schizophrenia. *Psychiatry Res* 2007;151 (3):189–199. [PubMed: 17412427]
- Thomos N, Boulgouris NV, Strintzis MG. Optimized transmission of JPEG2000 streams over wireless channels. *IEEE Trans Image Process* 2006;15(1)
- Tichavsky P, Koldovsky Z, Yeredor A, Gomez-Herrero G, Doron E. A hybrid technique for blind separation of non-Gaussian and time-correlated sources using a multicomponent approach. *IEEE Trans Neural Netw* 2008;19 (3):421–430. [PubMed: 18334362]
- Uddin LQ, Kelly AM, Biswal BB, Xavier Castellanos F, Milham MP. Functional connectivity of default mode network components: correlation, anticorrelation, and causality. *Hum Brain Mapp* 2009;30 (2):625–637. [PubMed: 18219617]
- Wang Z, Childress AR, Wang J, Detre JA. Support vector machine learning-based fMRI data group analysis. *NeuroImage* 2007;36 (4):1139–1151. [PubMed: 17524674]

Fig. 1.

This figure illustrates how the four BSS methods work on two datasets of the same type. Note that sCCA is different from mCCA in the part to generate CVs.

Fig. 2. BSS performance comparison of four joint BSS methods using simulated images. (a) shows the average absolute correlations of the $6 \times 2 = 12$ estimated source components with the ground truth under different PSNRs; (b) uses mean ISI of the two features as the measure of estimation accuracy of the mixing matrix, the lower the better.

Fig. 3.

The true sources, part of mixed noisy images under PSNR=10 dB, and the corresponding estimated results from four joint BSS models: CCA+ICA, joint ICA, mCCA and sCCA. Note that CCA+ICA succeeds in recovering sources accurately and linking them correctly, while the joint components derived from jICA, mCCA and sCCA may suffer from ambiguity and misinterpretation when their assumptions are not satisfied, e.g., source 4 in jICA, source 3 in mCCA and source 5 in sCCA.

Fig. 4.

Spatial maps of the identified significant group-discriminative components from three methods. The IC indices and their corresponding p values of the two sample t -test on group loadings are displayed below. Note that in joint ICA, two tasks share the same mixing matrix, while in CCA+ICA and mCCA, each task has its own mixing matrix, thus we can find both task-common (green frame) and task-distinct (pink frame) group discriminative components. The ICs with ** pass the FDR correction and the ICs with * pass an uncorrected level of $p < 0.05$. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Fig. 5.
The identified component whose loadings show a significant negative correlation with duration of illness (DOI) in schizophrenia, where $r=-0.33$, $p=0.0105$, FDR corrected.

Fig. 6.

Verification that the detected components of interest from each model are consistent with its own respective statistical assumptions. (a) For CCA+ICA, the top three paired components with the highest between-task source correlation are displayed. (b) One joint component from each method is selected with strong activations in STG shown in SM task; note that different spatial maps are demonstrated in SB task based on the model assumption.

Table 1

Joint BSS performance comparison via simulations.

Source index	S1-S2 corr	A1-A2 corr	Absolute correlation between sources inference and the ground truth under PSNR=10dB				
			JICA	mCCA	sCCA	CCA+ICA	2ICA
1	0.987	0.988	0.967±0.004	0.892±0.084	0.980±0.000	0.990±0.000	0.990±0.003
2	0.88	0.781	0.976±0.005	0.941 ±0.005	0.871 ±0.000	0.995±0.000	0.980±0.000
3	0.76	0.678	0.435±0.041	0.330±0.077	0.630±0.007	0.820±0.001	0.817±0.001
4	0.451	0.590	0.490±0.020	0.538±0.024	0.728±0.013	0.924±0.001	0.892±0.001
5	0.383	0.217	0.820±0.001	0.578±0.035	0.577±0.003	0.916±0.001	0.461±0.003
6	0.004	0.881	0.700±0.034	0.514±0.055	0.821 ±0.001	0.840±0.001	0.278±0.001
Inter-symbol Inference A1			0.184±0.019	0.251±0.021	0.199±0.002	0.051±0.001	0.034±0.005
Inter-symbol Inference A2			0.166±0.014	0.248±0.007	0.315±0.002	0.081 ±0.001	0.067±0.001
X1=A1-S1, X2=A2-S2 statistical assumption for every method			A1 and A2 should be identical	"A1-A2 corr" should be distinct	"S1-S2 corr" should be distinct enough	"S1-S2 corr" can be either common or distinct	No assumption to link two data sets.

BS S performance degrades due to unsatisfaction of the requirements

Sources are linked correctly and performance is the best among the five methods

Sources are mis-linked via cross-correlation between two data sets.

Table 2

Talairach table of the group-discriminative components identified by three methods.

Talairach label	Brodman areas	R/L (mm ³)	R/L Max Z (MNI Coord.)	R/L (mm ³)	R/L Max Z (MNI Coord.)
		SM		SBP	
CCA+ICA		IC 8**		IC 8*	
Postcentral gyrus	3 2 40 1 5 7 43	4.1/4.3	4.1(-59, -21,40)/7.3(45, -20,56)	4.5/4.7	5.6(-48, -26,57)/4.2(59, -24,40)
Precentral gyrus	4 6 44	0.7/1.2	4.0(-59, -18,40)/6.0(39, -23,56)	2.3/1.6	4.8(-45, -12,56)/3.4(36, -26,57)
Medial frontal gyrus	6 32 10	1.7/0.7	4.6(-3, -6,56)/4.7(3, -3,55)	1.0/0.7	5.1(0,0,50)/4.7(3,0,50)
Superior/middle frontal gyrus	6 10 8	0.4/0.3	3.5(-3,5,49)/4.3(36, -3,55)	0.1/0.4	3.2(-3,5,49)/2.7(3,5,49)
Superior/inferior parietal lobule	7 40	0.3/0.6	2.3(-56, -47,41)/3.7(33, -52,58)	1.5/3.9	4.9(-45, -32,57)/4.2(42, -38,57)
Cingulate gyrus	24 32	1.4/0.2	3.3(-6,5,44)/2.9(3, -4,44)	0.7/0.2	4.0(0,2,44)/3.7(3, -1,44)
Superior temporal gyrus	21 22 41 42 38	1.1/1.6	2.9(-56,3, -5)/3.5(59, -23,7)	0.1/0.3	2.5(-68, -35,2)/2.8(50,14, -11)
Inferior frontal gyrus	45 44	0.1/0.5	2.3(-53,12,10)/2.7(59,13,21)		
Transverse temporal gyrus	41 42			0.3/0.0	2.7(-56, -17,12)/NA
Occipital gyrus/cuneus	18 19			0.7/0.2	2.8(-36, -90,10)/2.3(6, -73, -1)
Precuneus	7			0.3/0.1	2.2(-9, -54,36)/2.4(3, -50,41)
		IC 6**		IC 11*	
Superior temporal gyrus	22 41 42 38 39 13	2.3/0.5	4.5(-50, -17,6)/3.5(48, -34,13)	4.4/0.5	4.8(-59, -6,6)/2.6(62, -26,12)
Transverse temporal gyrus	41	0.4/0.0	2.5(-50, -20,12)/NA	0.8/0.1	3.7(-59, -11,12)/2.3(56, -23,12)
Precuneus	7 19	0.4/0.3	5.5(-6, -79,43)/3.5(6, -77,43)	0.2/0.0	2.7(-33, -71,42)/2.1(30, -50,52)
Occipital gyrus/cuneus	18 19 37	1.2/0.6	4.5(-48, -70, -7)/3.5(3, -77,37)	0.5/0.8	2.6(-30, -93,5)/2.9(21, -99,10)
Postcentral gyrus	3 1 2 43 40	0.2/1.7	2.5(-65, -13,23)/4.1(42, -20,56)	1.4/0.4	3.9(-65, -25,18)/2.5(62, -5,17)
Precentral gyrus	4 6 43 13	0.0/0.7	NA/4.0(39, -20,56)	1.9/0.2	4.3(-59,0,6)/2.2(62, -2,17)
Angular gyrus	39	0.5/0.0	2.4(-50, -68,31)/NA		
Middle temporal gyrus	21 39 37	0.7/0.1	3.4(-62, -12, -4)/2.4(59, -46,8)		
Fusiform gyrus	19 37 20	1.7/0.7	4.8(-45, -73, -11)/2.8(39, -76, -11)		
Superior/inferior parietal lobule	7 40			1.2/0.5	3.7(-36, -56,53)/3.3(39, -56,53)
Medial/inferior frontal gyrus	11 10 9 6 8 45 46 47			3.4/0.6	3.5(-3,52, -13)/3.2(3,49, -13)
Insula	13			0.9/0.0	3.0(-45, -17,12)/NA
Joint ICA		IC 6**		IC 6**	
Superior/inferior parietal lobule	40 7 39	2.9/1.0	6.4(-45, -53,52)/3.9(56, -42,24)	1.4/2.8	3.7(-65, -25,23)/3.7(45, -50,52)

Talairach label	Brodman areas	R/L (mm ³)	R/L Max Z (MNI Coord.)	R/L (mm ³)	R/L Max Z (MNI Coord.)
Superior temporal gyrus	41 42 38 39 22 13	1.3/2.9	3.2(-53, -38.7)/4.2(53, -43.19)	2.7/2.1	4.4(-62, -23.12)/3.6(50, -3.0)
Insula	13	0.2/1.0	2.9(-45, -3.6)/4.0(50, -40.19)	1.1/0.3	3.0(-45.3, 0)/3.1(45.6, -5)
Transverse temporal gyrus	41 42	0.1/0.4	2.2(-50, -26.12)/3.7(39, -34.13)	0.5/0.0	3.7(-56, -20.12)/NA
Middle temporal gyrus	21 39	1.4/0.3	3.0(-56, -52.8)/2.6(48, -38.2)	0.0/1.5	NA/3.1(59, -41, -8)
Precuneus	7 19 31	0.6/0.0	3.3(-6, -74.42)/NA	0.9/0.4	3.6(-3, -56.42)/2.7(3, -51.36)
Precentral gyrus	4 6 44	0.3/1.0	3.1(-45, -6.6)/4.4(39, -23.56)	0.2/1.9	4.3(-59.0, 6)/2.2(62, -2.17)
Postcentral gyrus	3 40 2	0.0/1.1	NA/4.4(42, -23.56)	1.4/0.0	3.9(-65, -25.18)/NA
Middle frontal gyrus	6 8	0.8/0.0	3.0(-42.8, 49)/NA	0.0/1.9	NA/3.3(53, 16.27)
Medial/inferior frontal gyrus	11 10 9 6 8 40			0.8/2.4	3.5(-3, -9.56)/4.0(56, 10.27)
Angular gyrus	39			0.0/0.5	NA/3.1(53, -56.36)
mCCA			IC 12 *		IC 12 **
Occipital gyrus	18 37 19	3.2/2.9	6.7(-30, -91, -6)/8.1(48, -73, -6)	0.2/0.1	3.8(-39, -80.32)/2.8(33, -77.31)
Superior/middle frontal gyrus	6 10 9 46 8	1.8/3.4	5.6(-21.62, 19)/5.2(33, -6.56)	8.2/0.0	6.0(-39.45, 20)/NA
Cuneus/lingual gyrus	19 18 7 30 17	6.5/8.1	6.4(-3, -80.32)/7.2(9, -77.31)	0.4/1.0	3.0(-3, -80.37)/3.6(24, -83.37)
Medial/inferior frontal gyrus	10 6 9 46 44 45 32 47	3.4/3.1	7.3(-3, -6.56)/5.9(3, -6.50)	7.4/0.7	5.6(-3, -6.56)/4.5(3, -6.50)
Precentral gyrus	6 4 3	6.1/4.1	6.5(-45, -9.56)/5.5(36, -6.56)	6.0/0.9	6.3(-48, -12.56)/3.1(56, -19.34)
Postcentral gyrus	3 40 1 2 7 5 43	5.9/4.8	6.1(-45, -15.56)/5.0(50, -16.39)	7.5/4.3	7.7(-39, -32.62)/4.2(53, -27.40)
Superior/inferior parietal lobule	7 40	2.4/2.6	5.4(-21, -55.58)/4.5(39, -35.52)	4.4/2.8	7.2(-48, -32.57)/3.9(56, -27.35)
Cingulate gyrus/anterior cingulate	24 32 31 33	1.1/0.2	5.5(-6.2, 44)/4.5(3, -7.45)	2.3/0/3	4.7(-6.5, 44)/2.9(3, -7.45)
Precuneus	7 19 31	3.5/2.6	5.2(-9, -77.43)/4.9(3, -74.37)	1.1/1.5	5.2(-9, -73.48)/4.0(21, -70.48)
Middle/inferior temporal gyrus	18 37 19 21 39	1.9/1.7	5.3(-59, -58, -2)/6.2(50, -73, -1)		
Fusiform gyrus	19 18 37 20	1.6/0.6	4.8(-21, -67, -7)/4.8(48, -65, -12)		
Posterior cingulate	30 31 23	0.4/0.1	3.5(-15, -61.9)/3.2(3, -61.9)		
Superior/transverse temporal gyrus	42 41 22 38 13			4.6/1.0	4.9(-62, -20.12)/3.8(59, -34.18)
Insula	13 47 45			2.8/0.5	3.9(-45, -17.12)/3.7(56, -34.18)

Note that

** means the IC is FDR corrected for multiple comparisons and

* means uncorrected.