



A review of group ICA for fMRI data and ICA for joint inference of imaging, genetic, and ERP data

Vince D. Calhoun^{a,b,*}, Jingyu Liu^{a,b}, Tülay Adalı^c

^a The Mind Research Network, Albuquerque, NM 87131, USA

^b Department of ECE, University of New Mexico, Albuquerque, NM, USA

^c University of Maryland Baltimore County, Department of CSEE, Baltimore, MD 21250, USA

ARTICLE INFO

Article history:

Received 15 September 2008

Accepted 15 October 2008

Available online 13 November 2008

Keywords:

fMRI

SNP

ERP

Genetics

Independent component analysis

ABSTRACT

Independent component analysis (ICA) has become an increasingly utilized approach for analyzing brain imaging data. In contrast to the widely used general linear model (GLM) that requires the user to parameterize the data (e.g. the brain's response to stimuli), ICA, by relying upon a general assumption of independence, allows the user to be agnostic regarding the exact form of the response. In addition, ICA is intrinsically a multivariate approach, and hence each component provides a grouping of brain activity into regions that share the same response pattern thus providing a natural measure of functional connectivity. There are a wide variety of ICA approaches that have been proposed, in this paper we focus upon two distinct methods. The first part of this paper reviews the use of ICA for making group inferences from fMRI data. We provide an overview of current approaches for utilizing ICA to make group inferences with a focus upon the group ICA approach implemented in the GIFT software. In the next part of this paper, we provide an overview of the use of ICA to combine or fuse multimodal data. ICA has proven particularly useful for data fusion of multiple tasks or data modalities such as single nucleotide polymorphism (SNP) data or event-related potentials. As demonstrated by a number of examples in this paper, ICA is a powerful and versatile data-driven approach for studying the brain.

© 2008 Elsevier Inc. All rights reserved.

Introduction and background

Independent component analysis (ICA) is increasingly utilized as a tool for evaluating the hidden spatiotemporal structure contained within brain imaging data. In this paper, we first provide a brief overview of ICA and ICA as applied to functional magnetic resonance imaging (fMRI) data. Next, we discuss group ICA and ICA for data fusion with an emphasis upon the methods developed within our group and also discuss within a larger context of the many alternative approaches that are currently in use.

ICA is a statistical method used to discover hidden factors (sources or features) from a set of measurements or observed data such that the sources are maximally independent. Typically, it assumes a generative model where observations are assumed to be linear mixtures of independent sources, and unlike principal component analysis (PCA) which only uncorrelates the data, ICA works with higher-order statistics to achieve independence. Uncorrelatedness is only partway to independence, if two random variables are independent they are uncorrelated, however not all uncorrelated random variables are independent. An intuitive example of ICA can be given by a scatter-plot of two independent signals s_1 and s_2 . Fig. 1a (left, middle) show

the projections for PCA and ICA, respectively, for a linear mixture of s_1 and s_2 and Fig. 1a (right) shows a plot of the two independent signals (s_1, s_2) in a scatter-plot. PCA finds the orthogonal vectors u_1, u_2 , but cannot identify the independent vectors. In contrast, ICA is able to find the independent vectors a_1, a_2 of the linearly mixed signals (s_1, s_2), and is thus able to restore the original sources.

A typical ICA model assumes that the source signals are not observable, statistically independent and non-Gaussian, with an unknown, but linear, mixing process. Consider an observed M -dimensional random vector denoted by $\mathbf{x} = [x_1, x_2, \dots, x_M]^T$, which is generated by the ICA model:

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

where $\mathbf{s} = [s_1, s_2, \dots, s_N]^T$ is an N -dimensional vector whose elements are the random variables that refer to the independent sources and $\mathbf{A}_{M \times N}$ is an unknown mixing matrix. Typically $M \geq N$, so that \mathbf{A} is usually of full rank. The goal of ICA is to estimate an unmixing matrix $\mathbf{W}_{N \times M}$ such that \mathbf{y} given by

$$\mathbf{y} = \mathbf{W}\mathbf{x} \quad (2)$$

is a good approximation to the 'true' sources: \mathbf{s} .

Since to achieve ICA, statistical information higher than second order is needed, it can either be generated using nonlinear functions or can be explicitly calculated. Algorithms that use nonlinear functions

* Corresponding author. The Mind Research Network, 1101 Yale Blvd NE, Albuquerque, NM 87131, USA.

E-mail address: vcalhoun@unm.edu (V.D. Calhoun).

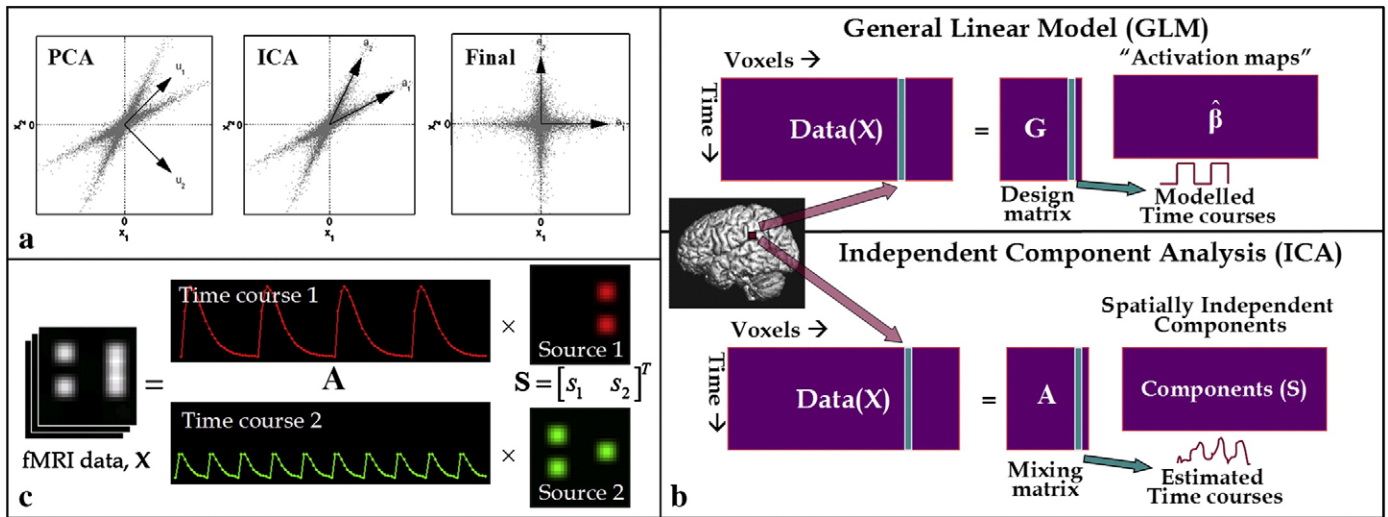


Fig. 1. a) Illustration of the need for higher order statistics: principle component analysis (PCA) identifies orthogonal directions which capture the most variance (a second order statistic) whereas ICA finds maximally independent directions using higher order statistics, b) Comparison of GLM and spatial ICA for fMRI data: the GLM requires the specification of the temporal model in the design matrix, whereas ICA estimates the timecourses from the data by maximizing independences between the component images, and c) Illustration of spatial ICA of fMRI data: the fMRI data is assumed to be comprised of linearly mixed sources, which are extracted via ICA along with their corresponding timecourses.

to generate higher-order statistics have been the most popular ICA approaches and there are a number of algorithms derived based on maximum likelihood estimation, maximization of information transfer, mutual information minimization, and maximization of non-Gaussianity. The first three approaches are equivalent to each other, and they coincide with maximization of non-Gaussianity when the unmixing matrix W is constrained to be orthogonal (Adali et al., 2008). The algorithms derived within these formulations have optimal large sample properties in the maximum likelihood sense when the nonlinearity within each algorithm is chosen to match the source density. Two commonly used ICA algorithms derived within these formulations are Infomax (Bell and Sejnowski, 1995; Lee et al., 1999) and FastICA (Hyvarinen and Oja, 1997). Another popular algorithm is joint approximate diagonalization of eigenmatrices (JADE) (Cardoso and Souloumiac, 1993), which relies on explicit computation of fourth-order statistical information. Both Infomax and FastICA typically work with a fixed nonlinearity or one that is selected from a small set, e.g., two in the case of extended Infomax (Bell and Sejnowski, 1995; Lee et al., 1999). These algorithms typically work well for symmetric distributions and are less accurate for skewed distributions and for sources close to Gaussian. Since optimality condition requires the nonlinearities to match the form of source distributions, there are a number of adaptation strategies that are developed. A flexible ICA using generalized Gaussian density model method is introduced in (Choi et al., 2000). Other flexible extensions of ICA include non-parametric ICA (Boscolo et al., 2001) and kernel ICA (Bach and Jordan, 2002) as well as approaches introduced in (Hong et al., 2005; Vlassis and Motomura, 2001). The variety of recent approaches for performing ICA and its applications in areas as diverse as biomedicine, astrophysics, and communications demonstrates the vitality of research in this area.

ICA of fMRI data

Following its first application to fMRI (McKeown et al., 1998), ICA has been successfully utilized in a number of exciting fMRI applications and especially in those that have proven challenging with the standard regression-type approaches for a recent collection of examples see (Calhoun and Adali, 2006; McKeown et al., 2003). Spatial ICA finds systematically non-overlapping, temporally coherent brain regions without constraining the shape of the temporal

response. The temporal dynamics of many fMRI experiments are difficult to study with fMRI due to the lack of a well-understood brain-activation model whereas ICA can reveal inter-subject and inter-event differences in the temporal dynamics. A strength of ICA is its ability to reveal dynamics for which a temporal model is not available (Calhoun et al., 2002). A comparison of the GLM approach and ICA as applied to fMRI analysis is shown in Fig. 1b.

Independent component analysis is used in fMRI modeling to study the spatio-temporal structure of the signal, and it can be used to discover either spatially or temporally independent components (Jung, 2001). Most applications of ICA to fMRI use the former approach and seek components that are maximally independent in space. In such a setting, we let the observation data matrix be X , an $N \times M$ matrix (where N is the number of time points and M is the number of voxels) as shown in Fig. 1b. The aim of fMRI component analysis is then to factor the data matrix into a product of a set of time courses and a set of spatial patterns. An illustration of how ICA decomposes the data into a parsimonious summary of images and time courses is shown in Fig. 1c. The number of components is a free parameter, which has previously been either empirically determined or estimated. There are a number of approaches for estimating the number of components using information theoretic approaches (Beckmann and Smith, 2004; Li et al., 2007).

Since the introduction of ICA for fMRI analysis, the choice of spatial or temporal independency has been controversial. However, the two options are merely two different modeling assumptions. McKeown et al. argued that the sparse distributed nature of the spatial pattern for typical cognitive activation paradigms would work well with spatial ICA (sICA). Furthermore, since the proto-typical confounds are also sparse and localized, e.g., vascular pulsation (signal localized to larger veins that are moving as a result of cardiac pulsation) or breathing induced motion (signal localized to strong tissue contrast near discontinuities: "tissue edges"), the Infomax algorithm with a sparse prior is very well suited for spatial analysis (Petersen et al., 2000) and has also been used for temporal ICA (Calhoun et al., 2001c) as have decorrelation-based algorithms (Petersen et al., 2000). Stone et al., proposed a method which attempts to maximize both spatial and temporal independence (Stone et al., 1999). An interesting combination of spatial and temporal ICA was pursued by Seifritz et al. (2002); they used an initial sICA to reduce the spatial dimensionality of the data by locating a region of interest in which they then subsequently

performed temporal ICA to study in more detail the structure of the non-trivial temporal response in the human auditory cortex.

Group ICA of fMRI data

Unlike univariate methods (e.g., regression analysis, Kolmogorov–Smirnov statistics), ICA does not naturally generalize to a method suitable for drawing inferences about groups of subjects. For example, when using the general linear model, the investigator specifies the regressors of interest, and so drawing inferences about group data comes naturally, since all individuals in the group share the same regressors. In ICA, by contrast, different individuals in the group will have different time courses, and they will be sorted differently, so it is not immediately clear how to draw inferences about group data using ICA. Despite this, several ICA multi-subject analysis approaches have been proposed (Beckmann and Smith, 2005; Calhoun et al., 2001a; Calhoun et al., 2001b; Calhoun et al., 2004a; Esposito et al., 2005; Guo and Giuseppe, In Press; Lukic et al., 2002; Schmithorst and Holland, 2004; Svensen et al., 2002). The various approaches differ in terms of how the data is organized prior to the ICA analysis, what types of output are available (e.g. single subject contributions, group averages, etc), and how the statistical inference is made.

A summary of several group ICA approaches is provided in Fig. 2. Approaches can be grouped into five categories. Fig. 2a illustrates approaches that perform single-subject ICA and then attempt to combine the output into a group post hoc by using approaches such as self-organized clustering or spatial correlation of the components (Calhoun et al., 2001a; Esposito et al., 2005). This has the advantage of allowing for unique spatial and temporal features, but has the disadvantage that since the data are noisy the components are not necessarily unmixed in the same way for each subject. The other four approaches involve an ICA computed on the group data directly.

Temporal concatenation Fig. 2b and spatial concatenation Fig. 2c have both been examined. The advantage of these approaches is that they perform one ICA, which can then be divided into subject specific parts, hence the comparison of subject differences within a component is straightforward. The temporal concatenation approach allows for unique time courses for each subject, but assumes common group maps whereas the spatial concatenation approach allows for unique maps but assumes common time courses. Although they are really just two different approaches for organizing the data, temporal concatenation appears to work better for fMRI data (Schmithorst and Holland, 2004) most likely because the temporal variations in the fMRI signal are much larger than the spatial variations, and has been widely used for group ICA of fMRI data.

The temporal concatenation approach is implemented in the MELODIC software (<http://www.fmrib.ox.ac.uk/fsl/>) and also the GIFT Matlab software (<http://icatb.sourceforge.net/>). The GIFT software additionally implements a back-reconstruction step which produces subject specific images (Calhoun et al., 2001b). This enables a comparison of both the time courses and the images for one group or multiple groups (Calhoun et al., 2008) (see simulation in (Calhoun et al., 2001b) which shows ICA with temporal concatenation plus back-reconstruction can capture variations in subject specific images). The approach implemented in GIFT thus trades-off the use of a common model for the spatial maps against the difficulties of combining single subject ICA. An in-between approach would be to utilize temporal concatenation separately for each group (Celone et al., 2006), although in this case matching the components post hoc becomes again necessary. The approach in Fig. 2d involves averaging the data prior to performing ICA. This approach is less computationally demanding, but makes a more stringent assumption that requires a common time course and a common spatial map. Finally, the tensorial approach in Fig. 2e (implemented in MELODIC) involves estimating a common time course and a common image for each

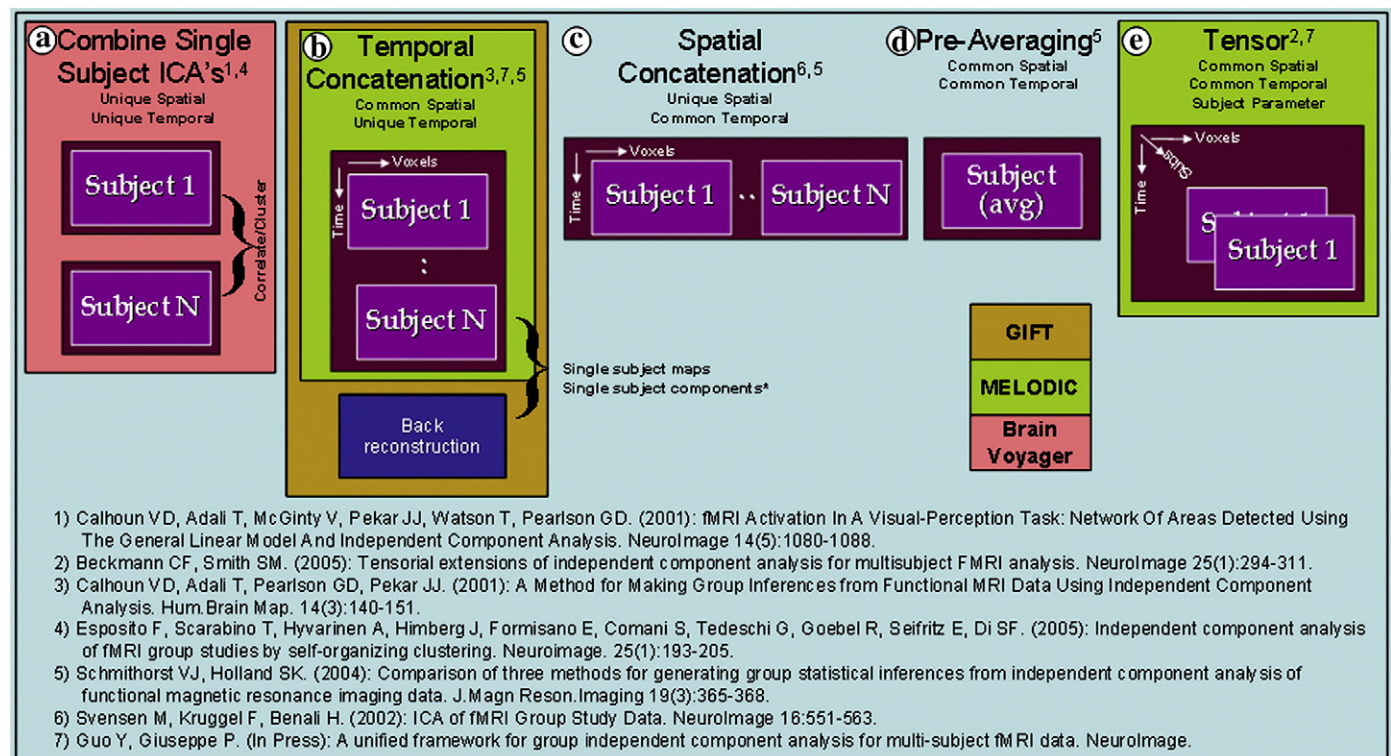


Fig. 2. Several Group ICA Approaches: A comparison of 5 group ICA approaches and some of the software packages which implement these methods as a primary pipeline. a) separate ICA analyses run on each subjects, followed by correlation or clustering to enable group inference, b) temporal concatenation followed by an aggregate ICA analysis is a popular approach which also can include a back-reconstruction step to compute single subject maps and timecourses, c) spatial concatenation or d) pre-averaging prior to ICA have also been proposed. Finally, e) tensor based approaches stack the data into a cube.

component but allows for a subject specific parameter to be estimated.

Higher order tensor decompositions (also known as multidimensional, multi-way, or n-way), have received renewed interest recently, although their adaptation to group and multi-group fMRI data is still being explored. Fig. 2e shows an approach based upon a three-dimensional tensor that has been developed to estimate a single spatial, temporal, and subject-specific ‘mode’ for each component to attempt to capture the multidimensional structure of the data in the estimation stage (Beckmann and Smith, 2005). This approach however may not work as well (without additional preprocessing) if the time courses between subjects are different, such as in a resting state study. A detailed comparison of several group ICA approaches including temporal concatenation and tensor ICA is provided in a recent paper (Guo and Giuseppe, *In Press*).

In the remainder of this paper, we focus on the group ICA approach implemented in the GIFT software (Calhoun et al., 2001b), which uses multiple data reduction steps following data concatenation to reduce the computational load, along with back-reconstruction and statistical comparison of individual maps and time courses following ICA estimation. An example group ICA analysis of nine subjects performing a four cycle alternating left/right visual stimulation task is presented in Fig. 3 (from Calhoun et al., 2001b). Separate components for primary visual areas on the left and the right visual cortex (depicted in red and blue, respectively) were consistently task-related with respect to the appropriate stimulus. A large region (depicted in green) including occipital areas and extending into parietal areas appeared to be sensitive to changes in the visual stimuli. Additionally some visual association areas (depicted in white) had time courses that were not task related. As we discuss later, group inference or comparison of groups can be performed by performing statistics on either the ICA images or the time courses.

ICA for data fusion

Many studies are currently collecting multiple types of imaging data from the same participants. Each imaging method reports on a

limited domain and typically provides both common and unique information about the problem in question. Approaches for combining or fusing data in brain imaging can be conceptualized as having a place on an analytic spectrum with meta-analysis (highly distilled data) to examine convergent evidence at one end and large-scale computational modeling (highly detailed theoretical modeling) at the other end (Husain et al., 2002). In between are methods that attempt to perform a direct data fusion (Horwitz and Poeppel, 2002). One promising data fusion approach is to first process each image type and extract features from different modalities. These features are then examined for relationships among the data types at the group level (i.e., variations among individuals or between patients and controls). This approach allows us to take advantage of the ‘cross’-information among data types and when performing multimodal fusion provides a natural link among different data types (Ardnt, 1996; Calhoun et al., 2006c; Savopol and Armenakis, 2002).

A natural set of tools for performing data fusion include those that transform data matrices into a smaller set of modes or components. Such approaches include those based upon singular value decomposition (SVD) (Friston et al., 1996; McIntosh et al., 1996) as well as ICA (McKeown et al., 1998). An advantage of ICA over variance-based approaches like SVD or PCA is the use of higher-order statistics to reveal hidden structure. In this paper, we describe two approaches for data fusion, joint ICA and parallel ICA. We show two examples, the first one involving event-related potential (ERP) and fMRI data and a second one on fMRI and genetic data.

Theory and implementation

In this section, we review the methods behind group ICA, joint ICA, and parallel ICA.

Group ICA of fMRI

As we mentioned earlier, the group ICA approach implemented in GIFT incorporates temporal concatenation plus back-reconstruction. Fig. 4 (top) provides a graphical representation of the GIFT approach

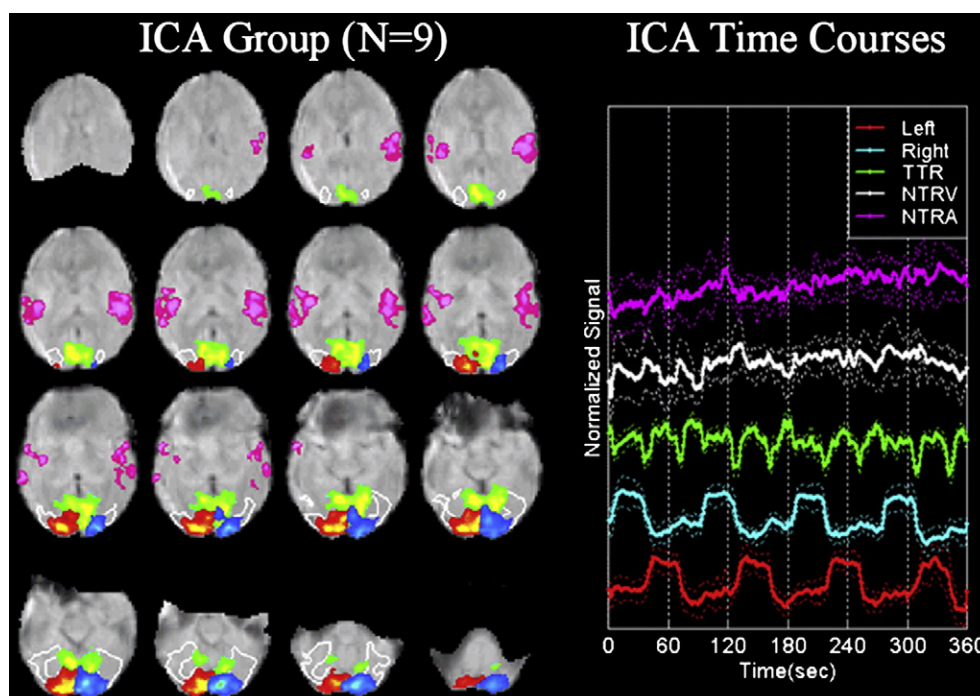


Fig. 3. fMRI Group ICA results (from Calhoun et al., 2001b): Group ICA identifies temporally coherent networks which are spatially distinct. In a relatively simple visual stimulation paradigm ICA identified strongly task-related networks (blue, red) as well as transient and non-task related networks (green, white, pink).

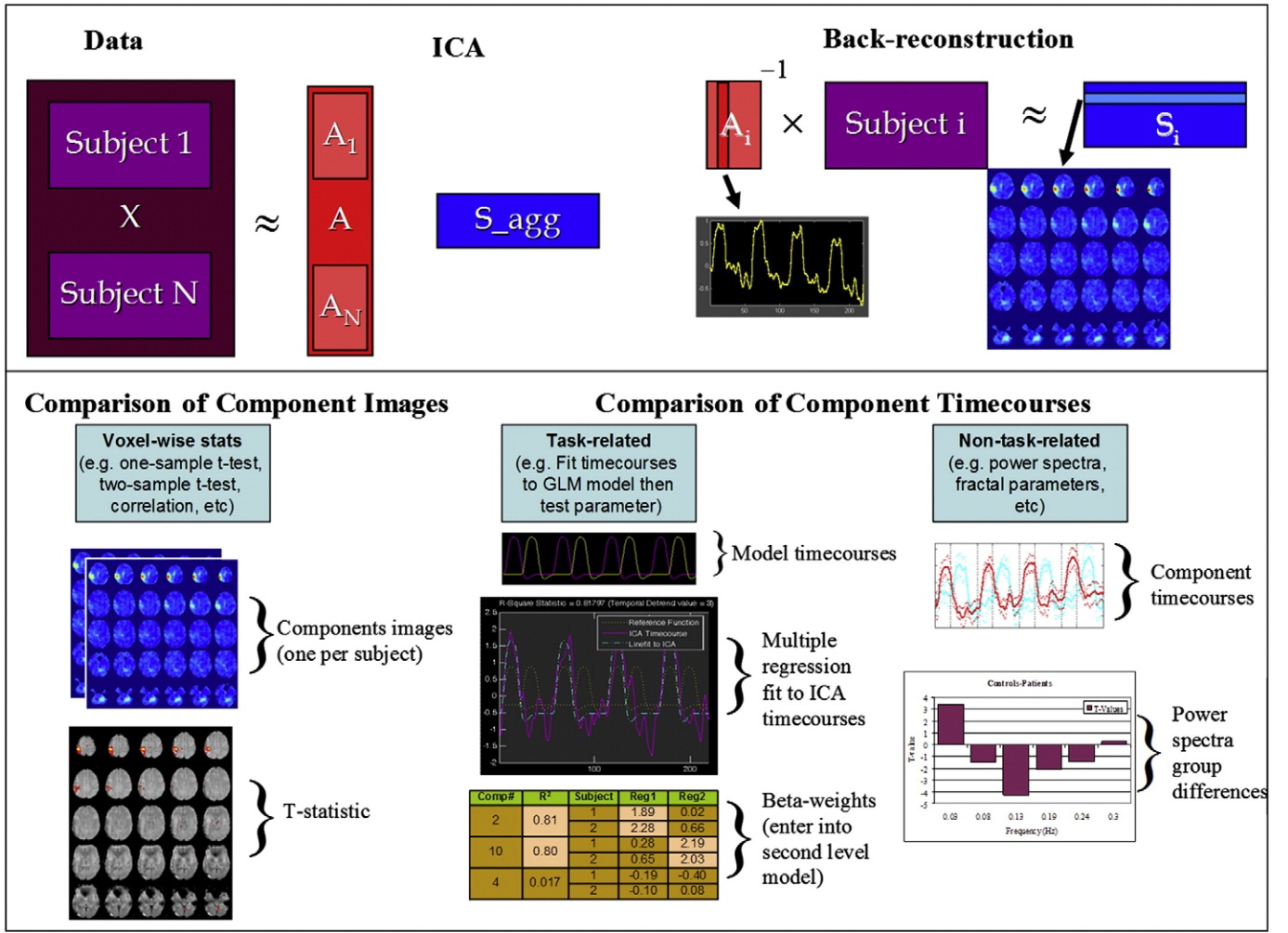


Fig. 4. Graphical Illustration of Group ICA as implemented in GIFT: Group ICA as implemented in GIFT incorporates temporal concatenation plus a back-reconstruction step to produce single subject maps and timecourses. The individual subject data is projected onto the subject-specific partition of the mixing matrix to compute the corresponding single-subject component image (top panel). Which of these components is of interest depends upon the question being asked which can draw upon comparisons of the component images or timecourses. Group ICA enables voxel-wise testing of the components images or fitting of a model to the component timecourses (bottom panel).

that essentially involves estimating a mixing matrix which has partitions that are unique to each subject. Once the mixing matrix is estimated, the component maps for each subject can be computed by projecting the single subject data onto the inverse of the partition of the mixing matrix that corresponds to that subject. In the end this provides subject specific time courses and images which can be used to make group and inter-group inferences.

An additional aspect to consider is that GIFT performs multiple data reduction steps, typically using PCA, primarily for computational reasons to reduce the amount of required memory. Mathematically, if we let $X_i = F_i^{-1} Y_i$ be the $L \times V$ reduced data matrix from subject i , where Y_i is the $K \times V$ data matrix (containing the preprocessed and spatially normalized data), F_i^{-1} is the $L \times K$ reducing matrix (determined by the PCA decomposition), V is the number of voxels, K is the number of fMRI time points and L is the size of the time dimension following reduction. The reduced data from all subjects is concatenated into a matrix and reduced using PCA to N dimensions (the number of components to be estimated). The $LM \times V$ reduced, concatenated matrix for the M subjects is

$$X = G^{-1} \begin{bmatrix} F_1^{-1} Y_1 \\ \vdots \\ F_M^{-1} Y_M \end{bmatrix} \quad (3)$$

where G^{-1} is an $N \times LM$ reducing matrix (also determined by a PCA decomposition) and is multiplied on the right by the $LM \times V$

concatenated data matrix for the M subjects. Following ICA estimation, we can write $X = AS$, where A is the $N \times N$ mixing matrix and S is the $N \times V$ component map. Substituting this expression for X into Eq. (3) and multiplying both sides by G , we have

$$GAS = \begin{bmatrix} F_1^{-1} Y_1 \\ \vdots \\ F_M^{-1} Y_M \end{bmatrix} \quad (4)$$

Partitioning the matrix G by subject provides the following expression.

$$\begin{bmatrix} G_1 \\ \vdots \\ G_M \end{bmatrix} \hat{AS} = \begin{bmatrix} F_1^{-1} Y_1 \\ \vdots \\ F_M^{-1} Y_M \end{bmatrix} \quad (5)$$

We then write the equation for subject i by working only with the elements in partition i of the above matrices such that

$$G_i \hat{AS}_i = F_i^{-1} Y_i \quad (6)$$

The matrix S_i in Eq. (6) contains the single subject maps for subject i and is calculated from the following equation

$$\hat{S}_i = (G_i \hat{A})^{-1} F_i^{-1} Y_i \quad (7)$$

We now multiply both sides of Eq. (6) by F_i and write

$$Y_i \approx F_i G_i \hat{AS}_i \quad (8)$$

which provides the ICA decomposition of the data from subject i , contained in the matrix \mathbf{Y}_i . The $N \times V$ matrix \mathbf{S}_i contains the N source maps and the $K \times N$ matrix $\mathbf{F}_i \mathbf{G}_i \mathbf{A}$ is the single subject mixing matrix and contains the time course for each of the N components.

Group inferences can be made by analyzing the subject specific time courses and spatial maps. Fig. 4 (bottom) categorizes these analyses into three main areas. To evaluate spatial properties of a given component statistically, one can perform voxel-wise tests on the spatial maps (Fig. 4; bottom left). The time courses can be analyzed by fitting to a GLM (the same model one would use for a GLM analysis; e.g. multiple regression), but instead of fitting to the voxel-wise data the ICA time courses are the dependent variable (Fig. 4; bottom middle). The estimated parameters can then be entered into a second level statistical analysis to make inferences about how much each component is modulated by a given stimulus, whether one component is modulated more by one stimulus than another, whether one group shows a stronger task-modulation than another, etc (Stevens et al., 2007). This provides a powerful way to make inferences about the components. Finally, one may be interested in the non-task-related components (or components in a resting-state study). In this case, one can evaluate differences in the spectral power between groups (Garrity et al., 2007) or compute additional parameters such as the fractal dimension of the subject component time courses (Fig. 4; bottom right).

Joint and parallel ICA

Next, we introduce two approaches for performing data fusion with ICA, joint ICA and parallel ICA (both of which are implemented in the Matlab-based Fusion ICA Toolbox (FIT: <http://icatb.sourceforge.net>)).

Joint ICA

Joint ICA is an approach that enables us to jointly analyze multiple modalities which have all been collected in the same set of subjects. In our development, we primarily consider a set of extracted features from each subject's data, these data form the multiple observations—the vertical dimension in our group dataset. Given two sets of group data (can be more than two, for simplicity,

we first consider two), \mathbf{X}_F and \mathbf{X}_G , we concatenate the two datasets side-by-side to form \mathbf{X}_J and write the likelihood as

$$L(\mathbf{W}) = \prod_{n=1}^N \prod_{v=1}^V p_{J,n}(u_{J,v}), \quad (9)$$

where $\mathbf{u}_J = \mathbf{W}\mathbf{x}_J$. Here, we use the notation in terms of random variables such that each entry in the vectors \mathbf{u}_J and \mathbf{x}_J correspond to a random variable, which is replaced by the observation for each sample $n=1, \dots, N$ as rows of matrices \mathbf{U}_J and \mathbf{X}_J . When posed as a maximum likelihood problem, we estimate a *joint* unmixing matrix \mathbf{W} such that the likelihood $L(\mathbf{W})$ is maximized.

Let the two datasets \mathbf{X}_F and \mathbf{X}_G have dimensionality $N \times V_1$ and $N \times V_2$, then we have

$$L(\mathbf{W}) = \prod_{n=1}^N \left(\prod_{v=1}^{V_1} p_{F,n}(u_{F,v}) \prod_{v=1}^{V_2} p_{G,n}(u_{G,v}) \right), \quad (10)$$

Depending on the data types in question, the above formula can be made more or less flexible.

This formulation characterizes the basic jICA approach and assumes that the sources associated with the two data types (F and G) modulate the same way across N subjects (see Fig. 5a). The assumption of the same linear covariation for both modalities is fairly strong, however it has the advantage of providing a parsimonious way to link multiple data types and has been demonstrated in a variety of cases with meaningful results (Calhoun et al., 2006a; Calhoun et al., 2006b; Calhoun et al., 2006c; Eichele et al., 2008; Moosmann et al., 2008).

There are different ways to relax the assumptions made in the formulation above, such as instead of constraining the two types of sources to share the same mixing coefficients, i.e., to have the same modulation across N samples, we can require that the form of modulation across samples for the sources from two data types to be correlated but not necessarily the same (Correa et al., 2008). The approach we discuss next, called parallel ICA provides this additional flexibility in modeling (Liu and Calhoun, 2006; Liu and Calhoun, 2007).

Parallel ICA

As noted earlier, the strong regularization imposed by the jICA framework can be relaxed in a number of ways to allow for more

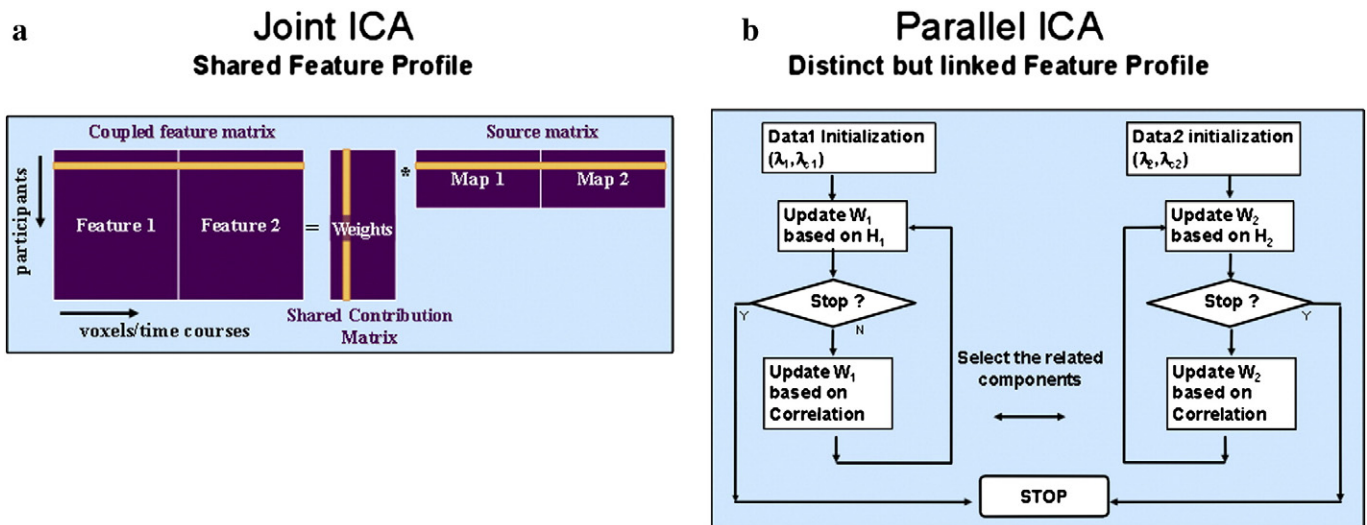


Fig. 5. Illustration of joint ICA and Parallel ICA models: Joint ICA (left) assumes a shared contribution matrix for the two modalities. Parallel ICA (right) updates separate ICA processes using the correlation between the subject profiles for the two modalities.

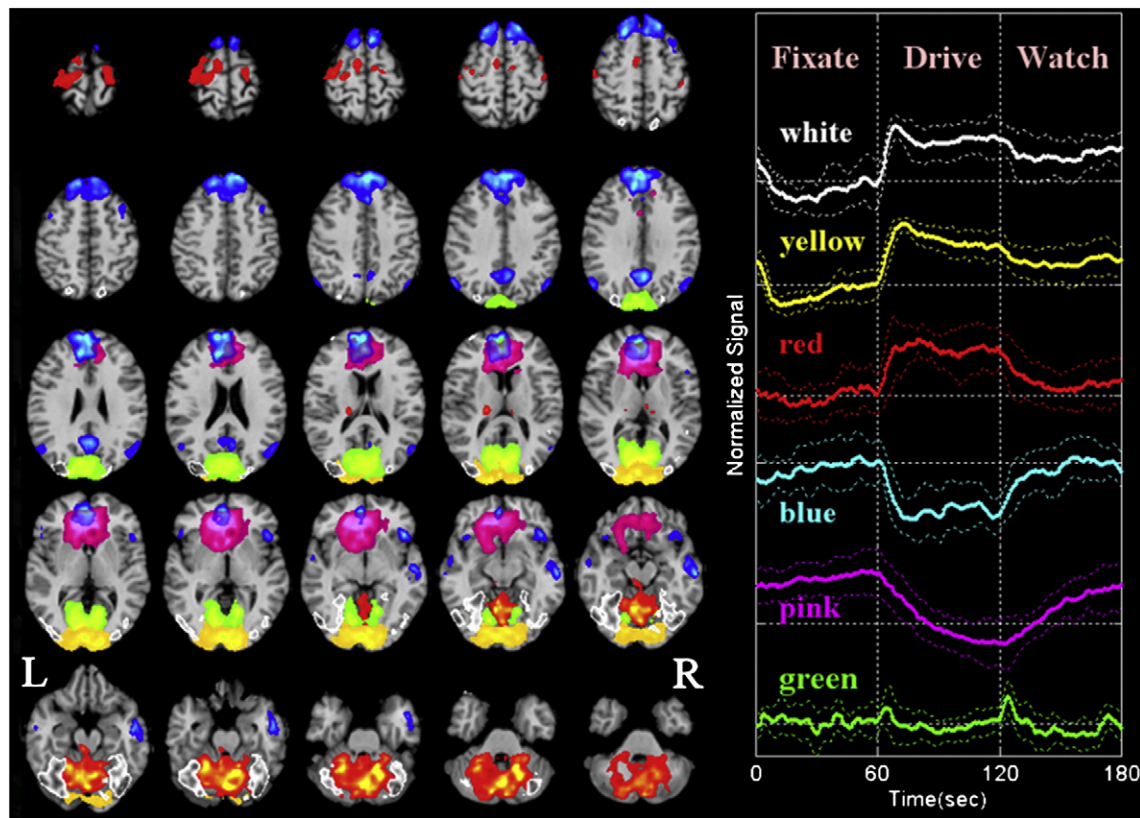


Fig. 6. Naturalistic driving (from Calhoun et al., 2002): Multiple networks identified during simulated driving. ICA enables us to study the complex and overlapping dynamics that occur during a naturalistic task.

flexibility in the estimation. One such approach we developed is called parallel independent component analysis (paralICA). As a framework to investigate the integration of data from two imaging modalities, this method identifies components of both modalities and connections between them through enhancing intrinsic interrelationships (see Fig. 5b). We have applied this approach to link fMRI/ERP data and

also fMRI and genetic data (single nucleotide polymorphism arrays) (Liu and Calhoun, 2006; Liu and Calhoun, 2007; Liu et al., In Press). Results show that paralICA provides stable results and can identify the linked components with a relatively high accuracy.

In our initial application of paralICA, we defined a genetic independent component as a specific SNP association, i.e., a group of

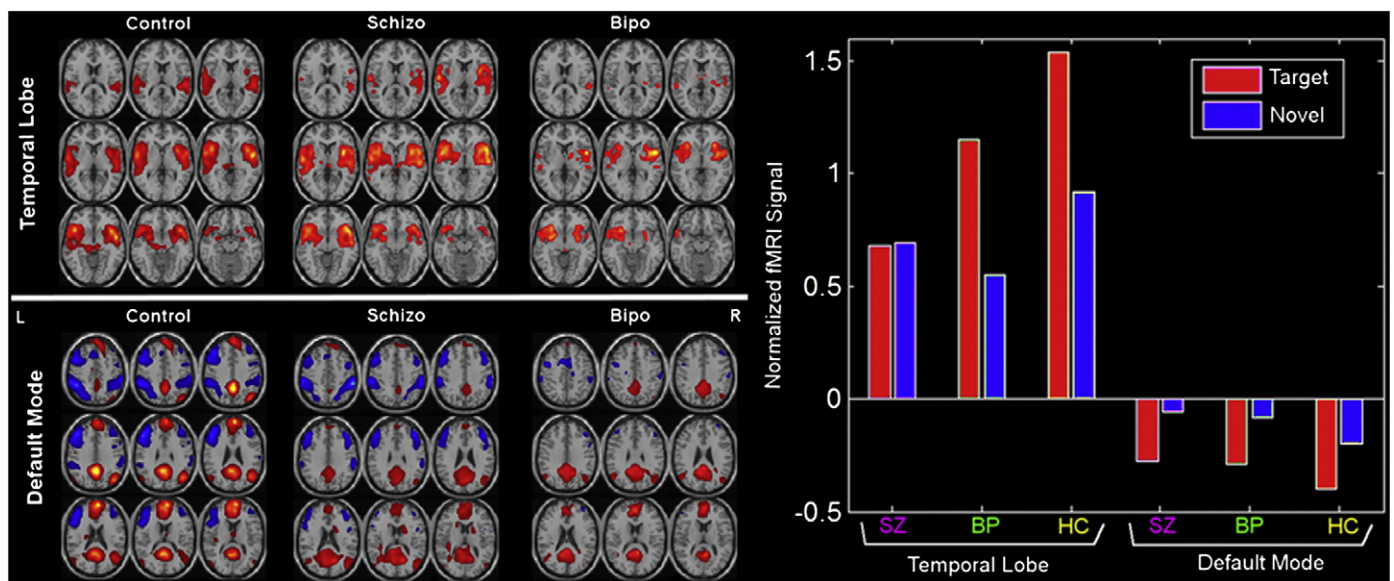


Fig. 7. Pair-wise comparisons of the Control, Schizophrenia, and Bipolar Groups (from Calhoun et al., 2008): Two-sample *t*-tests were performed to illustrate most significant differences for each pair-wise comparison (left). Note that these maps are generated from all subjects and actual classification regions will be slightly different due to the leave-1-out approach. On the right is plotted the average beta weights for the stimuli broken out by group.

SNPs with various degrees of contribution, which partially determines a specific phenotype or endophenotype. This association can be modeled as a linear combination of SNP genotypes (Dawy et al., 2005; Lee and Batzoglou, 2003),

$$s = \beta_1 \cdot \text{snp}_1 + \beta_2 \cdot \text{snp}_2 + \dots + \beta_{n1} \cdot \text{snp}_n; \quad (11)$$

where, snp is a genotype at a given locus and β is a weight contributed from a SNP to the genetic association. Beside the independent component, the weight itself is also of interest, implying the influence factor and type, i.e., inhibitory or excitatory to a phenotype. With the assumption that each component has an independent distribution pattern in 367 SNPs, we constructed the SNP data matrix, X , in a participant-by-SNP direction. The mixing process is presented in Eq. (12),

$$\begin{aligned} X_s &= [x_{s1}, x_{s2}, x_{s3}, \dots, x_{sn}]^T, \quad S_s = [s_{s1}, s_{s2}, s_{s3}, \dots, s_{sm}]^T; \\ S_s &= W_s X_s; \\ A_s &= W_s^{-1}; \quad A_s = [a_{s1}, a_{s2}, a_{s3}, \dots, a_{sn}]^T \end{aligned} \quad (12)$$

where, n is the number of participants and m is the number of components. x_{si} is a vector of 367 SNP genotypes for one participant. s_{si} is a vector of 367 SNP weights for one genetic component. A_s is the matrix

of the loading parameters, presenting the influence of each SNP component on participants.

In our current formulation, the relationship between brain function and the genetic component is calculated as the correlation between the columns of the fMRI A_f matrix and the SNP A_s matrix (note this can also be defined using other criteria such as mutual information, to identify nonlinear coupling between fMRI and SNP data). Thus, we have the correlation term and the maximization function based upon entropy. The procedure of parallel ICA is illustrated in Fig. 5b, where data 1 is the fMRI data and data 2 is the SNP data. The algorithm proceeds such that two demixing matrices W are updated separately during which the component with highest correlation from each modality is selected and used to modify the update of the demixing matrix based on the correlation value using appropriate stopping criteria.

Examples

In this section, we present examples of results from previous work using group ICA, joint ICA, and parallel ICA. The first example shows an analysis of a simulated driving paradigm, a case in which ICA is particularly useful as it is a naturalistic task that is difficult to

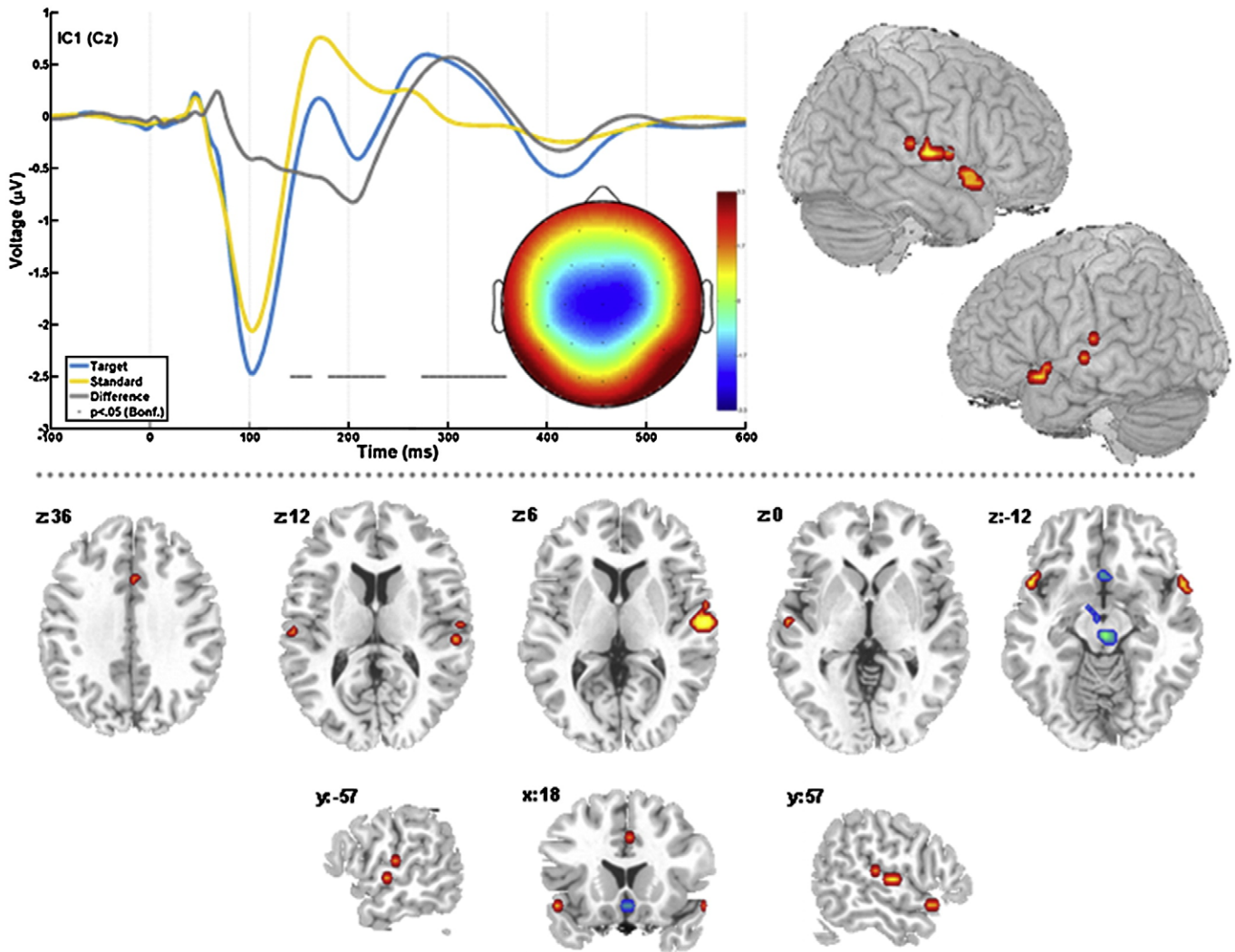


Fig. 8. Fusion of ERP and fMRI data (from Eichele et al., 2008): Time course and topography for EEG-tIC1 for standard and target epochs as well as the difference wave between them. The difference wave was subjected to a pointwise one-sample t -test, black dots indicate timeframes with significant difference from zero at $p < .05$, Bonferroni corrected for 512 tests ($t > 6.93$). The bilateral temporal activation in the linked fMRI component is shown as a surface rendering (top right). Additional slices in the lower half illustrate the overall spatial pattern. The maps are thresholded at 1% false discovery rate, cluster extent 5 voxels. Positive correlation is plotted in red, inverse correlation in blue.

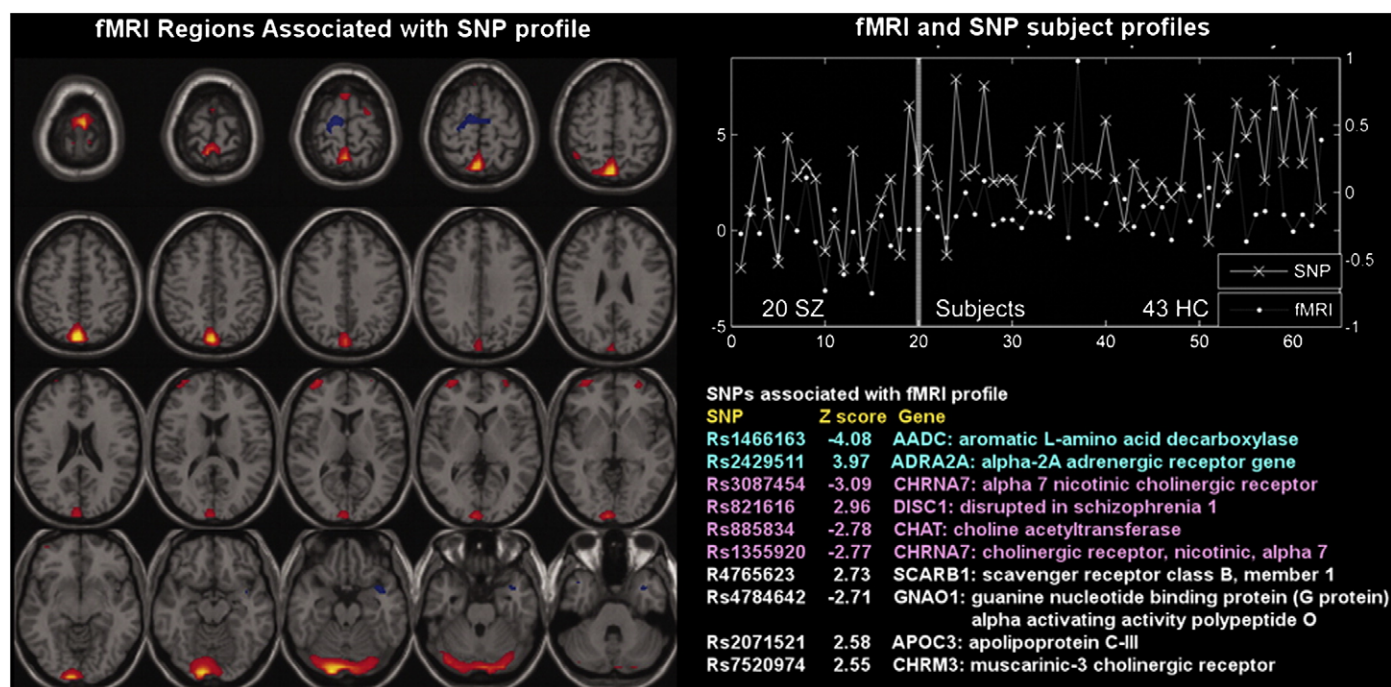


Fig. 9. Fusion of fMRI and genetic (SNP) data (from Liu et al., In Press): Parallel ICA provides an fMRI part (left) and a SNP part (bottom right) in addition to a correlated subject profile for both fMRI and SNP data (top right).

parameterize for use in a traditional GLM analysis. fMRI data from 15 subjects were collected during a 10 min paradigm with alternating 1 min blocks of fixation, simulated driving, and watching (Calhoun et al., 2002). ICA time courses were first analyzed to evaluate task-relatedness. Six components were identified and entered into a voxelwise one-sample *t*-test. A total of six components are presented showing different dynamics in response to simulated driving. In this case, ICA has proven to be a very powerful approach for analysis and enabled us to develop a model for the neural correlates of simulated driving which is nicely related to existing models based upon behavioral data (Calhoun et al., 2005; Calhoun et al., 2002; Calhoun et al., 2004b; Meda et al., In Press) (Fig. 6).

The second example we present is an analysis of fMRI data collected from an auditory oddball task for two patients groups as well as healthy controls. Back-reconstructed component maps were entered into two sample *t*-tests to evaluate pair-wise differences between the three groups. Results are presented for each group for two components, one in temporal lobe and also the default mode (Fig. 7; left). We performed a multiple regression including the target, novel, and standard stimuli and the mean of the estimated beta parameters is shown in Fig. 7 (right). We were also able to utilize these results to accurately differentiate healthy controls, schizophrenia patients, and patients with bipolar disorder. This example illustrates the ability of group ICA to differentiate groups and also shows both a comparison of the spatial maps and the time courses.

The next example involves data fusion of ERP and fMRI data using joint ICA. The fMRI data and the 64 channel ERP data are entered into a joint ICA analysis. This provides us with not only a temporal ERP profile and a spatial fMRI profile, but the topography from the ERP data provides additional information for interpretation (see Fig. 8). We developed a method for parallel spatial and temporal independent component analysis for concurrent multi-subject single-trial EEG-fMRI that addresses the mixing problem in both modalities, and integrates the data via correlation of the trial-to-trial modulation of the recovered fMRI maps and EEG time courses. The method affords extraction of a previously missed

spatiotemporal process corresponding to the auditory onset response and subsequent low-level orienting/change detection. For full details please see (Eichele et al., 2008).

Our final example shows results from a parallel ICA analysis of auditory oddball fMRI data and 367 SNPs from schizophrenia patients and healthy controls (Liu et al., In Press). When 43 healthy controls and 20 schizophrenia patients, all Caucasian, were studied, we found a correlation of 0.38 between one fMRI component and one SNP component. This fMRI component consisted of regions in parietal lobe, right temporal lobe, and bilateral frontal lobe. The relevant SNP component was contributed to significantly by 10 SNPs located in genes including those coding for the nicotinic alpha-7cholinergic receptor, aromatic amino acid decarboxylase, disrupted in schizophrenia 1, among others. Both fMRI and SNP components showed significant differences in loading parameters between the schizophrenia and control groups ($p=0.0006$ for the fMRI component; $p=0.001$ for the SNP component). The parallel ICA framework enabled us to identify interactions between brain functional and genetic information; our findings provide a proof-of-concept that genomic SNP factors can be investigated by using endophenotypic imaging findings in a multivariate format (Fig. 9).

Summary

ICA is a powerful data driven approach that can be used to analyze group fMRI data or to analyze multimodal data including fMRI, ERP, and genetic data. The examples demonstrate the utility and diversity of ICA-based approaches for the analysis of brain imaging data.

Conflict of interest

The authors declare that there are no conflicts of interest.

Acknowledgments

This research was supported in part by the National Institutes of Health (NIH), under grants 1 R01 EB 000840, 1 R01 EB 005846, and 1 R01 EB 006841.

References

- Adali, T., Novey, M., Cardoso, J.F., Adali, T., Novey, M., Cardoso, J.F., 2008. Complex ICA using nonlinear functions. *IEEE Trans. Signal Process.* 59, 4356–4544.
- Ardnt, C., 1996. Information gained by data fusion.
- Bach, F., Jordan, M., 2002. Kernel independent component analysis. *J. Mach. Learn. Res.* 3, 1–48.
- Beckmann, C.F., Smith, S.M., 2004. Probabilistic independent component analysis for functional magnetic resonance imaging. *IEEE Trans. Med. Imaging* 23 (2), 137–152.
- Beckmann, C.F., Smith, S.M., 2005. Tensorial extensions of independent component analysis for multisubject fMRI analysis. *NeuroImage* 25 (1), 294–311.
- Bell, A.J., Sejnowski, T.J., 1995. An information maximisation approach to blind separation and blind deconvolution. *Neural Comput.* 7 (6), 1129–1159.
- Boscolo, R.H., Pan, H., Roychowdhury, V.P., 2001. Non-parametric ICA. San Diego, California.
- Calhoun, V.D., Adali, T., 2006. 'Unmixing' functional magnetic resonance imaging with independent component analysis. *IEEE Eng. in Medicine and Biology* 25 (2), 79–90.
- Calhoun, V.D., Adali, T., McGinty, V., Pekar, J.J., Watson, T., Pearlson, G.D., 2001a. fMRI Activation In A Visual-Perception Task: Network Of Areas Detected Using The General Linear Model And Independent Component Analysis. *NeuroImage* 14 (5), 1080–1088.
- Calhoun, V.D., Adali, T., Pearlson, G.D., Pekar, J.J., 2001b. A method for making group inferences from functional MRI data using independent component analysis. *Hum. Brain Map.* 14 (3), 140–151.
- Calhoun, V.D., Adali, T., Pearlson, G.D., Pekar, J.J., 2001c. Spatial and temporal independent component analysis of functional MRI data containing a pair of task-related waveforms. *Hum. Brain Map.* 13 (1), 43–53.
- Calhoun, V.D., Pekar, J.J., McGinty, V.B., Adali, T., Watson, T.D., Pearlson, G.D., 2002. Different activation dynamics in multiple neural systems during simulated driving. *Hum. Brain Map.* 16 (3), 158–167.
- Calhoun, V.D., Adali, T., Pekar, J.J., 2004a. A method for testing conjunctive and subtractive hypotheses on group fMRI data using independent component analysis. *Mag. Res. Imag.* 22 (9), 1181–1191.
- Calhoun, V.D., Pekar, J.J., Pearlson, G.D., 2004b. Alcohol intoxication effects on simulated driving: exploring alcohol-dose effects on brain activation using functional MRI. *Neuropsychopharmacology* 29, 2097–2107.
- Calhoun, V.D., Carvalho, K., Astur, R.S., Pearlson, G.D., 2005. Using virtual reality to study alcohol intoxication effects on the neural correlates of simulated driving. *Appl. Psychophysiol. Biofeedback* 30 (3), 285–306.
- Calhoun, V.D., Adali, T., Giuliani, N., Pekar, J.J., Pearlson, G.D., Kiehl, K.A., 2006a. A method for multimodal analysis of independent source differences in schizophrenia: combining gray matter structural and auditory oddball functional data. *Hum. Brain Map.* 27 (1), 47–62.
- Calhoun, V.D., Adali, T., Kiehl, K.A., Astur, R.S., Pekar, J.J., Pearlson, G.D., 2006b. A method for multi-task fMRI data fusion applied to schizophrenia. *Hum. Brain Map.* 27 (7), 598–610.
- Calhoun, V.D., Pearlson, G.D., Kiehl, K.A., 2006c. Neuronal chronometry of target detection: fusion of hemodynamic and event-related potential data. *NeuroImage* 30 (2), 544–553.
- Calhoun, V.D., Pearlson, G.D., Maciejewski, P., Kiehl, K.A., Calhoun, V.D., Pearlson, G.D., Maciejewski, P., Kiehl, K.A., 2008. Temporal Lobe and 'Default' Hemodynamic Brain Modes Discriminate Between Schizophrenia and Bipolar Disorder. *Hum. Brain Map.* 29, 1265–1275.
- Cardoso, J.F., Souloumiac, A., 1993. Blind beamforming for non Gaussian signals. *IEE-Proceeding-F* 140 (6), 362–370.
- Celone, K.A., Calhoun, V.D., Dickerson, B.C., Atri, A., Chua, E.F., Miller, S., DePeau, K., Rentz, D.M., Selkoe, D., Albert, M.S., et al., 2006. Alterations in memory networks in mild cognitive impairment and alzheimer's disease: an independent component analysis. *J. Neurosci.* 26 (40), 10222–10231.
- Choi, S., Cichocki, A., Amari, S.I., 2000. Flexible independent component analysis. *J. VLSI Signal Process.* 26, 25–38.
- Correa, N., Adali, T., Li, Y., Calhoun, V.D., 2008. Examining associations between fMRI and EEG data using canonical correlation analysis. *Proc. ISBI, Washington, D.C.*
- Dawy, A., Sarkis, M., Hagenauer, J., Mueller, J., 2005. A Novel Gene Mapping Algorithm Based on Independent Component Analysis, pp. 381–384. Philadelphia, PA, USA.
- Eichele, T., Calhoun, V.D., Moosmann, M., Specht, K., Jongsma, M., Quiroga, R., Nordby, H., Hugdahl, K., 2008. Unmixing concurrent EEG-fMRI with parallel independent component analysis. *Int. J. Psych.* 67, 222–234.
- Esposito, F., Scarabino, T., Hyvarinen, A., Himberg, J., Formisano, E., Comani, S., Tedeschi, G., Goebel, R., Seifritz, E., Di, S.F., 2005. Independent component analysis of fMRI group studies by self-organizing clustering. *Neuroimage* 25 (1), 193–205.
- Friston, K., Poline, J.P., Strother, S., Holmes, A., Frith, C.D., Frackowiak, R.S., 1996. A multivariate analysis of PET activation studies. *Hum. Brain Map.* 4, 140–151.
- Garriy, A., Pearlson, G.D., McKiernan, K., Lloyd, D., Kiehl, K.A., Calhoun, V.D., 2007. Aberrant 'default mode' functional connectivity in schizophrenia. *Am. J. Psychiatry* 164 (3), 450–457.
- Guo, Y., Giuseppe, P., in press. A unified framework for group independent component analysis for multi-subject fMRI data. *NeuroImage*.
- Hong, B., Pearlson, G.D., Calhoun, V.D., 2005. Source-density driven independent component analysis approach for fMRI data. *Hum. Brain Map.* 25 (4), 297–307.
- Horwitz, B., Poeppel, D., 2002. How can EEG/MEG and fMRI/PET data be combined? *Hum. Brain Mapp* 17 (1), 1–3.
- Husain, F.T., Nandipati, G., Braun, A.R., Cohen, L.G., Tagamets, M.A., Horwitz, B., 2002. Simulating transcranial magnetic stimulation during PET with a large-scale neural network model of the prefrontal cortex and the visual system. *NeuroImage* 15 (1), 58–73.
- Hyvarinen, A., Oja, E., 1997. A fast fixed-point algorithm for independent component analysis. *Neural Comput.* 9 (7), 1483–1492.
- Jung, A., 2001. An introduction to a new data analysis tool: Independent Component Analysis, *Proc. Workshop GK "Nonlinearity"*. Regensburg.
- Lee, S.I., Batzoglou, S., 2003. Application of independent component analysis to microarrays. *Genome Biol.* 4 (11), R76.
- Lee, T.W., Girolami, M., Sejnowski, T.J., 1999. Independent component analysis using an extended Infomax algorithm for mixed subgaussian and supergaussian sources. *Neural Comput.* 11, 417–441.
- Liu, J., Calhoun, V.D., 2006. A novel approach to analyzing fMRI and SNP data via parallel independent component analysis. *Proc. SPIE*, pp. 1301–1311.
- Liu, J., Calhoun, V.D., 2007. Parallel independent component analysis for multimodal analysis: Application to fMRI and EEG Data, *Proc. ISBI, Washington, D. C.*
- Li, Y., Adali, T., Calhoun, V.D., 2007. Estimating the number of independent components for fMRI data. *Hum. Brain Map.* 28 (11), 1251–1266.
- Liu, J., Pearlson, G.D., Windemuth, A., Ruano, G., Perrone-Bizzozero, N.I., Calhoun, V.D., in press. Combining fMRI and SNP data to investigate connections between brain function and genetics using parallel ICA. *Hum. Brain Map.*
- Lukic, A.S., Wernick, M.N., Hansen, L.K., Strother, S.C., 2002. An ICA Algorithm For Analyzing Multiple Data Sets, *Int. Conf. on Image Processing (ICIP)*, Rochester, NY.
- McIntosh, A.R., Bookstein, F.L., Haxby, J.V., Grady, C.L., 1996. Spatial pattern analysis of functional brain images using partial least squares. *NeuroImage* 3 (3 Pt. 1), 143–157.
- McKeown, M.J., Makeig, S., Brown, G.G., Jung, T.P., Kindermann, S.S., Bell, A.J., Sejnowski, T.J., 1998. Analysis of fMRI data by blind separation into independent spatial components. *Hum. Brain Map.* 6, 160–188.
- McKeown, M.J., Hansen, L.K., Sejnowski, T.J., 2003. Independent component analysis of functional MRI: what is signal and what is noise? *Curr. Opin. Neurobiol.* 13 (5), 620–629.
- Meda, S., Calhoun, V.D., Astur, R., Turner, B., Ruopp, K., Pearlson, G.D., in press. Alcohol dose effects on brain circuits during simulated driving: An fMRI study. *Hum. Brain Mapp.*
- Moosmann, M., Eichele, T., Nordby, H., Hugdahl, K., Calhoun, V.D., 2008. Joint independent component analysis for simultaneous EEG-fMRI: principle and simulation. *Int. J. Psych.* 67, 212–221.
- Petersen, K.S., Hansen, L.K., Kolenda, T., Rostrup, E., Strother, S.C., 2000. On the Independent Components of Functional Neuroimages, pp. 615–620.
- Savopol, F., Armenakis, C., 2002. Mergine of heterogeneous data for emergency mapping: data integration or data fusion? *Proc. ISPRS, Buenos Aires, Argentina*.
- Schmithorst, V.J., Holland, S.K., 2004. Comparison of three methods for generating group statistical inferences from independent component analysis of functional magnetic resonance imaging data. *J. Magn. Reson. Imaging* 19 (3), 365–368.
- Seifritz, E., Esposito, F., Hennel, F., Mustovic, H., Neuhoef, J.G., Bilecen, D., Tedeschi, G., Scheffler, K., Salle, F.D., 2002. Spatiotemporal pattern of neural processing in the human auditory cortex. *Science* 297 (6), 1706–1708.
- Stevens, M., Kiehl, K.A., Pearlson, G.D., Calhoun, V.D., 2007. Functional neural circuits for mental timekeeping. *Hum. Brain Map.* 28 (5).
- Stone, J.V., Porri, J., Buchel, C., Friston, K., 1999. Spatial, Temporal, and Spatiotemporal Independent Component Analysis of fMRI Data, *Proc. Leeds Statistical Research Workshop, Leeds, UK*.
- Svensen, M., Kruggel, F., Benali, H., 2002. ICA of fMRI group study data. *NeuroImage* 16, 551–563.
- Vlassis, N., Motomura, Y., 2001. Efficient source adaptivity in independent in independent component analysis. *IEEE Trans. Neural Networks* 12, 559–566.