Published in final edited form as:

Magn Reson Imaging. 2007 June ; 25(5): 684–694.

Performance of Blind Source Separation Algorithms for fMRI Analysis using a Group ICA Method

Nicolle Correa^{*}, Tülay Adalı^{*}, and Vince D. Calhoun^{§,∇}

^{*}University of Maryland, Baltimore County, Dept. of CSEE, Baltimore, MD 21250

[§]Olin Neuropsychiatry Research Center, Institute of Living, Hartford, CT 06520

[∇]Yale University School of Medicine, New Haven, CT 06520

Abstract

Independent component analysis (ICA) is a popular blind source separation (BSS) technique that has proven to be promising for the analysis of functional magnetic resonance imaging (fMRI) data. A number of ICA approaches have been used for fMRI data analysis, and even more ICA algorithms exist, however the impact of using different algorithms on the results is largely unexplored. In this paper, we study the performance of four major classes of algorithms for spatial ICA, namely information maximization, maximization of non-gaussianity, joint diagonalization of cross-cumulant matrices, and second-order correlation based methods when they are applied to fMRI data from subjects performing a visuo-motor task. We use a group ICA method to study the variability among different ICA algorithms and propose several analysis techniques to evaluate their performance. We compare how different ICA algorithms estimate activations in expected neuronal areas. The results demonstrate that the ICA algorithms using higher-order statistical information prove to be quite consistent for fMRI data analysis. Infomax, FastICA, and JADE all yield reliable results; each having their strengths in specific areas. EVD, an algorithm using second-order statistics, does not perform reliably for fMRI data. Additionally, for the iterative ICA algorithms, it is important to investigate the variability of the estimates from different runs. We test the consistency of the iterative algorithms, Infomax and FastICA, by running the algorithm a number of times with different initializations and note that they yield consistent results over these multiple runs. Our results greatly improve our confidence in the consistency of ICA for fMRI data analysis.

Keywords

fMRI; functional; independent component analysis; ICA; visuo-motor

Introduction

Independent component analysis (ICA) is a popular blind source separation (BSS) technique that has been successfully applied to analysis of functional Magnetic Resonance Imaging (fMRI) data [1]. The majority of applications of ICA to fMRI use the Infomax [2] algorithm or FastICA [3] algorithm. A few comparative studies [4], [5], [6] have compared Infomax to FastICA and/or a BSS approach using second-order statistics, the Molgedey-Schuster

Correspondence: Nicolle Correa, Dept. of Comp. Sci. and Electrical Engineering, University of Maryland Baltimore County, 1000 Hilltop Circle, Baltimore, MD 21250 Phone: 410-455-1631 E-mail: nicolle1@umbc.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

algorithm [7]. The Statistical Parametric Mapping software package (SPM2) [8] has been used extensively to study fMRI data, however it requires that the model time courses be parameterized *a priori* and hence it can be used primarily in the study of task-related components. ICA, on the other hand, extracts both spatially independent brain maps and their activation time courses from the fMRI data without *a priori* specification of the time courses and is shown to be promising for the study of components whose time courses are not known or cannot be easily modeled (see e.g. [9]).

Our purpose in this paper is to investigate the performance of different ICA algorithms that are commonly used to analyze fMRI data along with SPM. We use fMRI data from 12 subjects performing a visuo-motor task and instead of entering each subjects' data into a separate ICA analysis, we use a group ICA technique [10] to estimate one set of components and then back-reconstruct from the aggregate mixing matrix to obtain the individual subject maps. This method has the advantage of ordering the components in different subjects in the same way, which is a tedious task if individual ICA analyses are performed because unlike the general linear model (GLM) used by SPM and other univariate methods that can easily be generalized to group analysis, subjects as processed by ICA do not share common regressors thus making it difficult to match components across subjects. In [10], it is shown that the back-reconstructed ICA results using our group approach are similar to those performed for each individual subject, which shows that the group ICA technique allows for cross-subject variability.

To date there have been several approaches presented to perform group ICA [11], [12], [10]. Schmithorst and Holland [13], compared three group ICA methods using the FastICA algorithm on simulated group data and noted that the group ICA technique proposed in [10] provided the best overall performance in terms of accurate estimation of the sources and associated time courses. In our study, we used this group ICA approach that has been implemented in the user-friendly environment of the *group* ICA of fMRI toolbox (GIFT) [14]. We use this toolbox to study the performances of four BSS/ICA algorithms when applied to fMRI data. Thus far nine algorithms have been implemented in GIFT, the algorithms we have used for the comparative study in this paper are: Infomax (maximum likelihood) [2], FastICA [3], joint approximate diagonalization of eigenmatrices (JADE) [15], and eigenvalue decomposition (EVD) [16]. This selection includes an algorithm from each of the major approaches to performing ICA namely, information maximization, maximization of non-gaussianity, joint diagonalization of cross-cumulant matrices, and second-order correlation based methods.

Infomax and FastICA are iterative algorithms that require proper initialization and setting of parameters. To enable the efficient use of ICA for fMRI data, the issue of consistency of the estimates needs to be addressed. ICASSO [17] is a toolbox developed to evaluate the similarity among multiple estimates of the FastICA algorithm. We incorporate ICASSO in GIFT to evaluate the consistency in estimates of Infomax and FastICA.

The algorithms we examine use different methodologies to achieve separation of the sources causing the estimates from different algorithms to differ. We propose a number of evaluation techniques to compare the estimation results among the ICA algorithms. One of the techniques we use involves comparison of the estimates with masks that include expected areas of activation. Another comparison method uses Infomax as a reference and calculates the difference images of Infomax and the other algorithms. This method enables us to easily evaluate the differences in estimation among the algorithms. We also compare the performance of the ICA algorithms based on correlations of the time courses to the model paradigms. For the time courses, we also check for significant departures from the Infomax results using t-tests. Finally, we enter the results from the four algorithms into ICASSO and report of how the estimates from the different algorithms cluster based on correlation.

Our results show that the ICA and GLM methods generally yield similar estimates for the task-related components. The iterative ICA algorithms yield consistent results over multiple runs for the components examined. The two iterative algorithms that utilize higher-order statistical information; namely Infomax and FastICA give the best overall performance. JADE performs comparably well; however, its estimated area of activation is usually smaller than that of Infomax and FastICA. The EVD approach, though very fast, yields results that are quite different from those given by the other algorithms and it is determined to be a useful analysis tool only when the spectra of the components are different.

Materials and Methods

Participants and Experimental paradigms

Twelve right-handed participants with normal vision [six females, six males, average age 30 years] participated in the study. Participants provided written informed consent for a protocol approved by the Hartford Hospital Institutional Review Board. The subjects performed a visuomotor task involving two identical but spatially off-set, periodic, visual stimulus, shifted by 20 s from one another. The visual stimulus was presented alternately to the right and left visual hemifield. The motor task consisted of participants touching their thumb to each of their four fingers sequentially, back and forth, at a self-paced rate using the hand on the same side on which the visual stimulus is presented. For more details on the paradigm, please see [18].

Imaging parameters

Scans were acquired at the Olin Neuropsychiatry Research Center at the Institute of the Living on a Siemens Allegra 3T dedicated head scanner. The functional scans were acquired using gradient-echo echo planar imaging with the following parameters (repeat time (TR) = 1.50 s, echo time (TE) = 27 ms, field of view = 24 cm, acquisition matrix = 64×64 , flip angle = 60 degrees, slice thickness = 4 mm, gap = 1 mm, 28 slices, ascending acquisition). Six “dummy” scans were performed at the beginning to allow for longitudinal equilibrium, after which the paradigm was automatically triggered to start by the scanner.

Data analysis

Preprocessing—Data were preprocessed using SPM2. Images were realigned using INRIalign - a motion correction algorithm unbiased by the local signal changes [19], [20]. Data were spatially normalized into the standard Montreal Neurological Institute space [21], spatially smoothed with a $8 \times 8 \times 8$ mm full width at half-maximum Gaussian kernel. The data (originally acquired at $3.75 \times 3.75 \times 4$ mm) were slightly resampled to $3 \times 3 \times 5$ mm, resulting in $53 \times 63 \times 28$ voxels. For display, slices 2-26 are presented.

SPM—Each subject’s data are entered in the GLM analysis framework using SPM2. Regressors consisted of the two time courses (left and right visuo-motor stimulation) convolved with an estimate of the standard SPM2 hemodynamic response function. Data are high pass filtered (drift removal) by entering sinusoidal functions into the model up to a frequency of 1/128 s as covariates. The single subject’s results from the first level analysis were thresholded at $p < 0.0005$ ($t = 4.5$) to obtain voxels that are significantly task related. These twelve individual analysis are then entered together into a second level “random effects” analysis by performing a one-sample t-test to identify the voxels that are significantly task related across the group. Results are compared at a number of thresholds: $p < 0.0005$ ($t = 4.4$), $p < 0.005$ ($t = 3.11$), and $p < 0.05$ ($t = 1.6$). The results for SPM at $p < 0.0005$ ($t = 4.4$) for the left task-related component and at $p < 0.05$ ($t = 1.6$) for the right task-related component are displayed in Fig. 1

ICA—GIFT is a MATLAB-based ICA/BSS tool, that includes a number of analysis and visualization techniques in a user friendly graphical interface. In [22], a comparison of a number

of algorithms was reported for a simulated set of fMRI-like data and actual fMRI data from a single slice. The results suggested that Infomax, FastICA, and JADE performed well. SIMBEC (simultaneous blind extraction using cumulants) and AMUSE (algorithm for multiple unknown signal extraction) were not robust. SIMBEC, however, performed well for sub-gaussian sources representing artifacts. In this paper, we use GIFT to perform data reduction and ICA analysis on fMRI data from 12 subjects. Data reduction is performed on each of the 12 subjects to decrease computational complexity while retaining most of the variance. In the first reduction step, each subject's data is reduced from 119 to 60 components and then the reduced data is placed into three groups, each group containing data from four subjects. In each group, the data is concatenated in the temporal dimension and further reduced to 25 components. The 75 components (25 from each group) are then concatenated and a final reduction step is performed to reduce the data to 25 components. We then perform ICA on this final set. The resulting mixing matrix is then used to back-reconstruct spatial maps and time courses for each individual subject and each component. The results are scaled from arbitrary units to percentage signal change so as to be able to compare results across algorithms.

As discussed in detail in [10], the data reduction, sub-grouping, and concatenation in the temporal dimension, is performed primarily for computational efficiency and easier interpretability. For the work presented in this paper, its effect is deterministic since the processing at this stage is kept the same for all ICA algorithms. It is our experience that changing the number of components retained in the PCA data reduction stages does not make a significant difference to the outcome if the number of components in the earlier reduction stages are not considerably smaller than the number of components in the later stages so as to be able to retain most of the variance in the data. Order selection is an important issue to be taken into account while performing ICA (see e.g. [23], [24]) since estimating a much greater number of components than are actually present in the mixture may cause the independent components to split whereas estimating a smaller number will give erroneous results. In this paper, the number of components estimated is chosen such that it is approximately one standard deviation away from the average number of components estimated using the minimum description length (MDL) [25] criterion for each subject.

The algorithms evaluated in the paper belong to four different classes of BSS/ICA techniques. Infomax [2] uses a global measure of minimizing the mutual information among the source estimates by maximizing the information transfer from the input to the output of a network using a non-linear function and natural (relative) gradient updates. FastICA [3] maximizes the higher-order statistics or the negentropy of the output to maximize the non-gaussianity of the estimated sources using fixed-point iterations. Using even non-linearities $\tanh = (1/a)\log \cosh(ay)$, $\text{pow3} = (1/4)y^4$, and $\text{gauss} = -\exp(-y^2/2)$ to approximate the negentropy reduces the computational complexity of this method. The pow3 non-linearity essentially uses the kurtosis-based approximation that has a higher growth rate than the other non-linearities. The $\tanh$ and gauss non-linearities address the robustness issue by their slower growth rate than the pow3 non-linearity, and also provide a computationally efficient way to approximate negentropy.

JADE [15] and EVD [16] are computationally efficient and explicitly use fourth and second-order statistics respectively neglecting statistics of other orders. JADE uses the Jacobi technique, to perform joint approximate diagonalization on fourth-order cumulant matrices to achieve spatial independence among sources. EVD is based on second-order statistics and uses higher-order correlation functions to achieve separation. It performs singular value decomposition on the sum of a number of shifted cross-variance matrices and uses the structure within the data to achieve source separation. This algorithm, in contrast to the other three, relaxes the condition of non-gaussianity and independence. It is computationally simple; however it is not able to differentiate well sources that have very similar spectra.

Postprocessing—Using GIFT, back-reconstruction is performed on the ICA results to obtain results for the individual subjects. Back-reconstruction involves computing the individual subject maps and time courses from the group ICA mixing matrix and the data reduction matrices from the first data reduction stage. An average over ten ICA estimations was calculated for the non-deterministic algorithms. For each component, we then enter the results of the individual subjects into voxel-wise one sample t-test and threshold at a t-score of $t = 4.5$ ($p < 0.0005$) to find voxels that are significantly active across the 12 subjects. We also apply a contiguity filter to remove activations which are less than six voxels in size. We generate masks corresponding to five components of interest, which we expect from previous results and used the masks to select five out of the 25 components, using the spatial correlation sorting option in GIFT. For 10% of the cases we chose the component based on visual evaluation since the component had not been correctly estimated by the algorithm and hence the sorting technique selected the wrong component. The five components we analyze are

1. Right task related (RTR) component: a right visual and right motor area corresponding to the task [26].
2. Left task related (LTR) component: a left visual and left motor area corresponding to the task [26].
3. Transiently task related (TTR) component: an area in visual cortex that is activated due to both right and left tasks [26].
4. Temporal component: an area activated in bilateral temporal lobe [27].
5. Default mode component: areas that tend to decrease when the subject performs a task [28], [29], [30].

The results of each of the algorithms, averaged over the 12 subjects for the five selected components are displayed in Fig. 2(a) through Fig. 2(d).

Analysis methods

Infomax has been one of the most commonly used algorithms for fMRI data analysis and has proven to be quite reliable [4]. Also, for the task-related components, Infomax shows similar results to SPM as seen in [26] and as shown later in this paper. Based on these results, in some of our analysis methods we use Infomax as a reference and compare the performance of the other ICA algorithms against Infomax. We use six different methods to evaluate the performance of the ICA algorithms, which we explain next.

Consistency analysis using ICASSO—ICASSO [31] presents a visualization technique that enables analysis of the differences among estimates obtained from multiple runs due to the iterative nature of FastICA. In [32], ICASSO was used to analyze ICA of fMRI data using FastICA. We include ICASSO in GIFT to check for consistency of the iterative ICA algorithms we used, namely FastICA and Infomax. ICASSO runs the ICA algorithm a number of times (we use ten runs), each run produces different estimates of the components. ICASSO collects the components estimated from all the runs and then matches components across runs by clustering components based on their absolute value of the correlation between the squared source estimates. We set the number of clusters to be equal to 25 (number of components estimated). If the algorithm gives consistent results across runs, each cluster will contain the estimates of one of the components obtained over multiple runs. The cluster will be compact if the estimates correlate highly with each other. The toolbox also outputs the correlation between clusters. This gives us an indication of how independent each component is from the other. The “compactness” and “isolation” of the clusters is given by a measure R which is the difference between the average intra-cluster correlations and the average inter-cluster correlations for each cluster. This measure is useful to quantify the consistency in the

performance of the algorithms. The location of the clusters in the plot is such that the proximity of clusters is directly proportional to their similarity.

Analysis using masks—Using WFU_PickAtlas [33], [34], [35] a tool that allows the user to create masks by selecting different areas of the brain, both appropriate Brodmann areas (BAs) and functional areas; we create masks corresponding to each of the five components. These masks include neuronal areas that are expected to be activated during the task. This prior knowledge was obtained from published experimental results (see section on postprocessing) on subjects performing similar visuo-motor tasks. These are the same masks that were used to select the five interesting components based on the correlation of the estimated activation maps with the masks.

The areas included in each of the masks are as follows:

1. Task related: BAs 1, 2, 3, 4, 6, 17, 18, and 19 that include somatosensory, motor, and visual areas. The right and left hemispheres containing the above regions were chosen for the right and left task-related masks respectively.
2. Transiently task related: BAs 17, 18, 19, 23, 30, and 31 that include visual areas and posterior cingulate cortex.
3. Bilateral temporal lobe: BAs 20, 21, 22, 37, 38, and 42 that include areas responsible for visual, auditory, and sensory perception as well as language function.
4. Default mode: precuneus, posterior cingulate as well as BAs 7, 10, and 39 that include areas from posterior parietal cortex, frontal pole and occipitoparietal junction.

The SPM and ICA results are compared to these masks to investigate the overlap between the results and the masks. In addition we examined the results in several other ways as follows.

Analysis using voxel-wise paired t-test—Infomax performs quite reliably for fMRI data analysis and hence, we take Infomax as a reference and use the following procedure to calculate the difference in estimation between algorithm, say X, and Infomax. Let I_{ijk} and X_{ijk} denote the k^{th} voxel in the j^{th} component of the i^{th} subject as estimated by Infomax and algorithm X respectively.

1. In the postprocessing step, t-tests are performed voxel-wise on each component to check for voxels that are significantly active across the 12 subjects. The t-test of the k^{th} voxel of the j^{th} component has 12 samples given by X_{ijk} , $i = 1, \dots, 12$. A t-image is formed containing only the t-values of the significantly active voxels. The t-image is thresholded at $t = 4.5$ to generate an image that contains voxels that are significantly active at $p < 0.0005$. Only the voxels that are above this threshold are included in the subsequent steps. The signal decreases have been ignored for the analysis.
2. Difference images are calculated for each component for each subject. The k^{th} voxel of the difference image of the j^{th} component of the i^{th} subject is given by $D_{ijk} = I_{ijk} - X_{ijk}$.
3. A t-test is calculated, one voxel at a time, on the difference images to check for significant differences. Each t-test has 12 samples given by D_{ijk} , $i = 1, 2, \dots, 12$.
4. The difference images are thresholded at $|t| = 4.5$ to obtain voxels that are estimated significantly differently by the two algorithms at $p < 0.0005$.

This method of calculating difference images provides a simple, but useful approach to compare the areas of activation estimated by the different algorithms.

Analysis using model time courses—We model the task-related components using SPM2. The right and left task-related time courses are generated by convolving a temporal model of the on-off task with a canonical hemodynamic response function.

The average time courses as estimated by the algorithms are displayed in Fig. 2(a) through Fig. 2(d). The correlation of these estimated time courses with the model time courses is calculated for the task-related components for each of the algorithms.

Analysis using difference t-tests on averaged time cycle—The task-related time courses are periodic with four cycles. We average them over these cycles and calculate time point by time point t-tests on the difference between the cycle-averaged time courses of Infomax and the other algorithms. The temporal and default mode components also appear to have periodic time courses however; due to lack of prior knowledge of the structure of these two time courses we have performed point-by-point t-tests on the differences between the entire time course of Infomax and the other algorithms instead of the cycle-averaged time courses.

Results

Comparison of SPM and ICA results

At $p < 0.0005$ ($t = 4.4$), for the left task-related component the results of ICA (Infomax) and SPM are quite similar. However, at this threshold the right task-related component gets cut-off below the threshold. When the threshold is lowered for SPM to $p < 0.005$ ($t = 3.11$), the right task-related area barely passes the threshold and when it is further lowered to $p < 0.05$ ($t = 1.6$) the SPM result is comparable to the ICA results. Hence, we can conclude that the two components are estimated similarly by both methods with respect to the areas detected; with the ICA approach providing much higher t-values or contrast to noise ratio.

ICASSO consistency evaluation

The clusters obtained for Infomax (Fig. 3(a)) are very tight especially for the five interesting components we use. FastICA using the *tanh* and the *pow3* non-linearities also estimate consistent components corresponding to the tight clusters. FastICA using the *gauss* non-linearity (Fig. 3(b)) has more random estimates than those obtained by the other two non-linearities, however the components of interest are still quite consistent. The maximum value that R can take is 1, which will occur when the average intra-cluster correlations is 1 and the average inter-cluster correlations is 0. We obtain $R > 0.9$ for the components of interest for all algorithms; except for the default component, as estimated by FastICA using the *pow3* non-linearity where the value is 0.87 which is also quite high which indicates that the estimates are consistent.

Overlap with masks

The number of voxels in the estimated components that overlap with the corresponding masks is calculated for each algorithm and displayed in Table 1

SPM, Infomax, and FastICA have very similar number of voxels coinciding with the masks. JADE underestimates the RTR component and the temporal component (Fig. 2(b)). For the RTR, EVD estimates an activation area in both the right and left visual cortex, which is evident from the poorly estimated time course and the area is very widely spread (Fig. 2(d)). EVD estimates a very small number of voxels for the LTR component since it places most of the left visual neuronal activation along with the right visual activation. The default mode is estimated similarly by the higher-order algorithms, however EVD groups many additional regions in with the typical default mode areas thus estimating a much larger area for this component.

Voxel-wise paired t-test

The difference images reveal specific dissimilarities among the algorithms. JADE consistently estimates smaller activation areas as compared to Infomax. This is especially evident for the temporal component. FastICA and Infomax estimate mostly overlapping activation regions with Infomax estimating a slightly larger area. Overall the three non-linearities perform almost identically for FastICA. The performance of EVD differs from that of Infomax as discussed next. For the right task-related component, EVD fails to include the right motor area in its estimation. It includes much of the left visual area along with the right visual area. The left task-related component is hence poorly estimated. For the transiently task-related component, the areas estimated by EVD and Infomax are completely different. For the temporal component EVD, estimates a smaller area than Infomax while for the default mode component it estimates a much larger area than Infomax.

Model time courses

The time courses estimated by JADE have the highest correlation with the model time courses. Infomax and FastICA yield results that are quite similar results to one another. EVD has a very low correlation with Infomax for the right and left task-related time courses. The correlation coefficients are displayed in Table 2.

Difference t-tests on averaged time cycle

The results of the task-related time courses are displayed in Fig. 4 We observe that the time courses of JADE have almost no time points significantly different from Infomax. FastICA using the three different non-linearities have few significantly different time points. FastICA using *pow3* appears to be closest to Infomax among the three non-linearities. EVD shows a larger number of significant differences as compared to the other algorithms.

Clustering of all ICA algorithms in ICASSO

We also use ICASSO to cluster the components of interest as estimated by all six ICA algorithms we used, to check how well the estimates from different algorithms correlate with each other. Initially we get four clusters since the transiently task-related estimates and default mode estimates get clustered together. We then set the number of clusters to be five and once again the results remain the same except that the temporal component in EVD forms a separate cluster by itself. We observe that excluding EVD from the analysis gives us the expected five clusters. We then set the number of clusters to be six and a value of $R > 0.55$ for all the clusters. We get the five clusters corresponding to the five interesting components and a sixth cluster containing only the temporal component as estimated by EVD, see Fig. 5 Also the transient component in EVD gets clustered with the default mode estimates of the rest of the algorithms and the default mode component in EVD gets clustered with the transient estimates of the other algorithms. These results increase our confidence in ICA for fMRI data as the results from the different algorithms correlate with each other and also emphasize the poor performance of EVD as well as the similar performance of the other three algorithms for the five components examined.

Computational cost

Infomax takes slightly longer than the other algorithms. Infomax takes approximately 70 s, while JADE takes 62 s on a Pentium 4, 1.6 GHz, 512 GB, Windows XP computer using Matlab v7.0. FastICA using the *tanh* non-linearity takes 35 s, using the *gauss* non-linearity takes 38 s and using the *pow3* non-linearity takes 22 s. EVD is the fastest taking approximately 3 s. Infomax and FastICA use iterative updates whereas EVD and JADE use batch procedures. JADE is memory and computation intensive since it needs to store all the cumulant matrices. In our study we estimate 25 components and in this case JADE did not exhibit any noticeable

difference in estimation time, which is likely to be an issue when the number of estimated components increases further.

Discussion

We experimentally studied the estimation results of four spatial ICA/BSS algorithms when applied to fMRI analysis. We compared the estimation results of ICA against those obtained using SPM for the task-related components. Using Infomax, we obtained regions that were similar to those obtained using SPM, however Infomax estimates components with a higher contrast to noise ratio which may be because ICA algorithms do not constrain the time courses. SPM on the other hand allows for explicit tests on the shapes on the time courses which is required for certain fMRI experiments. Recent methods that combine both SPM and ICA [36] or constrained-ICA approaches [37] are useful to achieve a balance between this tradeoff.

Infomax, FastICA, and JADE, which all implicitly or explicitly use higher order statistics, yield results that are consistent with each other. Among the ICA algorithms we have tested, Infomax yielded results with highest t-values and JADE yielded the lowest. Infomax estimates a component set that is maximally independent and thus Infomax provides a global measure of separation whereas FastICA maximizes the non-gaussianity of the estimated components. Using Infomax as a reference to compare the other algorithms proved to be reasonable since visually it is seen that Infomax yields results that appear to be the best. In their comparisons of spatial and temporal ICA algorithms, Petersen et al. [6] found that temporal Infomax had problems because the version of the Infomax algorithm they use provides a better match to super-gaussian sources, which is violated when estimating time courses that are typically sub-gaussian. We use spatial ICA and since our five components of interest are all super-gaussian, the results using sigmoid-type non-linearity that is matched to super-gaussian sources yielded good results. Esposito et al. [5] found Infomax to be superior to the fixed-point algorithm (FastICA) in terms of estimation of the ICA model and noise reduction capabilities. We found that Infomax performed slightly better than FastICA due to higher contrast to noise ratio.

JADE on an average estimates smaller areas of activation in comparison to Infomax and FastICA. Lowering the threshold for JADE, does increase the size of the activation areas however this also causes inclusion of a large amount of irrelevant areas, likely false positives, thus making it evident that JADE yields estimates with a lower contrast to noise ratio. JADE uses fourth-order statistics and minimizes the cross-cumulants to achieve independence among the estimated components in contrast to approaches such as Infomax and FastICA that implicitly incorporate all higher order statistics through a nonlinear function.

EVD is based on second-order statistics and results suggest that it shows poor selectivity for fMRI analysis. It cannot differentiate between sources with similar spectra as in the case of the right and left task-related sources in our study. It places both of these sources in the same component and also estimates the time courses of these components poorly. Petersen et al. [6] also observed that the second-order spatial Molgedey-Schuster algorithm showed poor performance due to poor differentiability of sources with similar spectra. Overall the performance of EVD is lower than that of the higher-order algorithms we have tested.

The iterative algorithms Infomax and FastICA yield consistent results over runs and the initialization of parameters do not cause large changes in results. For FastICA, non-linearities *tanh* and *pow3*, yield more consistent results than the *gauss* non-linearity. In terms of computational time, among the four algorithms, Infomax takes slightly longer and EVD is the fastest.

Since we have used the group ICA technique to make the comparison process more convenient, a question arose regarding the dependence of the results on this technique. We thus ran a subset

of the analysis on ICA results obtained from individual subjects instead of the back-reconstructed results from the group ICA technique and found that though the numerical values of the tests varied slightly, the conclusions about the comparative performance of the algorithms remained the same.

Our work shows that the most commonly used ICA approaches yield consistent results for fMRI data analysis. Infomax, FastICA and JADE yield reliable results for fMRI data and thus are useful algorithms for studying components about which we have no prior information. Also the iterative ICA algorithms yield results that are quite consistent over runs for the components of interest. These results greatly increase our confidence in the use of ICA for fMRI analysis.

REFERENCES

- [1]. Mckeown MJ, Makeig S, Brown GG, Jung T, Kindermann SS, Bell AJ, Sejnowski TJ. Analysis of fMRI by blind separation into independent spatial components. *Hum. Brain Mapp* 1998;6:160–188. [PubMed: 9673671]
- [2]. Bell AJ, Sejnowski TJ. An information maximization approach to blind separation and blind deconvolution. *Neural Comput* 1995;7:1129–1159. [PubMed: 7584893]
- [3]. Hyvärinen A, Oja E. A fast fixed-point algorithm for independent component analysis. *Neural Comput* 1997;9:1483–1492.
- [4]. Calhoun VD, Adali T, Pearlson GD. Independent component analysis applied to fMRI data: a generative model for validating results. *J. VLSI Signal Process. Sys* 2004;37:281–291.
- [5]. Esposito F, Formisano E, Seifritz E, Goebel R, Morrone R, Tedeschi G, Di Salle F. Spatial independent component analysis of functional MRI time-series: to what extent do results depend on the algorithm used? *Hum. Brain Mapp* 2002;16:146–157. [PubMed: 12112768]
- [6]. Petersen, KS.; Hansen, LK.; Kolenda, T.; Rostrup, E.; Strother, SC. On the independent components of functional neuroimages; *Proc. ICA.*; Finland. 2000; p. 615-620.
- [7]. Molgedey L, Schuster H. Separation of independent signals using time-delayed correlations. *Phys. Rev. Letters* 1994;72:3634–37.
- [8]. SPM: <http://www.fil.ion.ucl.ac.uk/spm2>
- [9]. Calhoun VD, Pekar JJ, McGinty VB, Adali T, Watson TD, Pearlson GD. Different activation dynamics in multiple neural systems during simulated driving. *Hum. Brain Mapp* 2002;16:158–167. [PubMed: 12112769]
- [10]. 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;14:140–151. [PubMed: 11559959]
- [11]. Beckmann CF, Smith SM. Tensorial extensions of independent component analysis for group fMRI data analysis. *NeuroImage* 2005;25:294–311. [PubMed: 15734364]
- [12]. Svensen M, Kruggel F, Benali H. ICA of fMRI group study data. *NeuroImage* 2002;16:551–563. [PubMed: 12169242]
- [13]. Schmithorst VJ, Holland SK. Comparison of three methods for generating group statistical inferences from independent component analysis of functional Magnetic Resonance Imaging data. *J. Magn. Reson. Imag* 2004;19:365–368.
- [14]. GIFT: <http://icatb.sourceforge.net>
- [15]. Cardoso JF, Souloumiac A. Blind beamforming for non gaussian signals. *IEE Proc* 1993;140:362–370.
- [16]. Georgiev, P.; Cichocki, A. Blind source separation via symmetric eigenvalue decomposition; *Proc. 6th Symp. Signal Process. and its Applications.*; Kuala Lumpur, Malaysia. 2001; p. 17-20.
- [17]. ICASSO: <http://www.cis.hut.fi/projects/ica/icasso/about+download.shtml>
- [18]. Calhoun VD, Adali T, Pekar JJ, Pearlson GD. Latency (in)sensitive ICA: group independent component analysis of fMRI data in the temporal frequency domain. *NeuroImage* 2003;20:1661–1669. [PubMed: 14642476]

- [19]. Freire L, Mangin JF. Motion correction algorithms may create spurious activations in the absence of subject motion. *NeuroImage* 2001;14:709–722. [PubMed: 11506543]
- [20]. Freire L, Roche A, Mangin JF. What is the best similarity measure for motion correction in fMRI time series? *IEEE Trans. Med. Imag* 2002;21:470–484.
- [21]. Friston K, Ashburner J, Frith CD, Poline JP, Heather JD, Frackowiak RS. Spatial registration and normalization of images. *Hum. Brain Mapp* 1995;2:165–189.
- [22]. Correa, N.; Adali, T.; Li, Y-O.; Calhoun, VD. Comparison of blind source separation algorithms for fMRI using a new Matlab toolbox: GIFT; *Proc. ICASSP.*; Philadelphia, PA, USA. 2005;
- [23]. Beckmann CF, Smith SM. Probabilistic independent component analysis for functional magnetic resonance imaging. *IEEE Trans. Med. Imag* 2004;23:137–152.
- [24]. McKeown MJ. Detection of consistently task-related activations in fMRI data with hybrid independent component analysis. *NeuroImage* 2000;11:24–35. [PubMed: 10686114]
- [25]. Rissanen J. Modeling by the shortest data description. *Automatica* 1978;14:465–471.
- [26]. Calhoun VD, Adali T, McGinty VB, Pekar JJ, Watson TD, Pearlson GD. FMRI activation in a visual-perception task: network areas detected using the general linear model and independent component analysis. *NeuroImage* 2001;14:1080–1088. [PubMed: 11697939]
- [27]. Calhoun VD, Kiehl KA, Liddle PF, Pearlson GD. Aberrant localization of synchronous hemodynamic activity in auditory cortex reliably characterizes Schizophrenia. *Biol. Psychiatry* 2004;55:842–849. [PubMed: 15050866]
- [28]. Beckmann CF, Luca MDe, Devlin JT, Smith SM. Investigations into resting-state connectivity using independent component analysis. *Phil. Trans. Roy. Soc. of Lon., Ser. B* 2005;360:1001–1013. [PubMed: 16087444]
- [29]. Raichle ME, MacLeod AM, Snyder AZ, Powers WJ, Gusnard DA, Shulman GL. A default mode of brain function. *Proc. Natl. Acad. Sci* 2001;98:676–682. [PubMed: 11209064]
- [30]. McKiernan KA, Kaufman JN, Kucera-Thompson J, Binder JR. A parametric manipulation of factors affecting task-induced deactivation in functional neuroimaging. *J. Cogn. Neuroscience* 2003;15:394–408.
- [31]. Himberg, J.; Hyvärinen, A. Icasso: software for investigating the reliability of ICA estimates by clustering and visualization; *Proc. IEEE Workshop on Neural Networks for Signal Process (NNSP)*; Toulouse, France. 2003; p. 259-268.
- [32]. Himberg J, Hyvärinen A, Esposito A. Validating the independent components of neuroimaging time-series via clustering and visualization. *NeuroImage* 2004;22:1214–1222. [PubMed: 15219593]
- [33]. Lancaster LJ, Summerlin JL, Rainey L, Freitas CS, Fox PT. The Talairach Daemon, a database server for Talairach atlas labels. *NeuroImage* 1997;5:S633.
- [34]. Lancaster JL, Woldorff MG, Parsons LM, Liotti M, Freitas CS, Rainey L, Kochunov PV, Nickerson D, Mikiten SA, Fox PT. Automated Talairach atlas labels for functional brain mapping. *Hum. Brain Mapp* 2000;10:120–131. [PubMed: 10912591]
- [35]. Maldjian JA, Laurienti PJ, Kraft RA, Burdette JH. An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. *NeuroImage* 2003;19:1233–1239. [PubMed: 12880848] (WFU Pickatlas, version 1.04).
- [36]. Hu D, Yan L, Liu Y, Zhou Z, Friston KJ, Tan C, Wu D. Unified SPM-ICA for fMRI analysis. *NeuroImage* 2005;25:746–755. [PubMed: 15808976]
- [37]. Calhoun VD, Adali T, Stevens M, Kiehl KA, Pekar JJ. Semi-blind ICA of fMRI: a method for utilizing hypothesis-derived time courses in a spatial ICA analysis. *NeuroImage* 2005;25:527–538. [PubMed: 15784432]

Fig. 1.
Components as estimated by SPM. Yellow indicates right task-related component and blue left task-related component.

Fig. 2. Components as estimated by a: Infomax, b: JADE, c: FastICA using tanh nonlinearity, d: EVD: The color code for the components is: Yellow: Right task related; Blue: Left task related; Red: Transiently task related; Orange: Temporal; Green: Default mode.

(a)

(b)

Fig. 3.

Cluster plots for a: Infomax (cluster numbers 2, 3, 8, 10, and 16 correspond to the five components selected) and b: FastICA using *gauss* non-linearity (cluster numbers 1, 2, 4, 7, and 13 correspond to the five components selected), generated using ICASSO. S_{in} denotes the correlation between estimates within a cluster and s_{ij} denotes the correlation between estimates occurring in other clusters.

Fig. 4.

The columns show the RTR, LTR and TTR cycle-averaged time courses (from left to right). The time courses in blue are of Infomax and those in red are of JADE in a, b, and c; of FastICA using tanh nonlinearity in d, e, and f; of FastICA using pow3 nonlinearity in g, h, and i; of FastICA using gauss nonlinearity j, k, and l; and of EVD in m, n, and o. The vertical lines indicate time points at which the time courses significantly differ.

Fig. 5.

Clustering results for all algorithms for six clusters: Cluster 1: Temporal component as estimated by EVD; Cluster 2: Left task-related components; Cluster 3: Transiently task-related components; Cluster 4: Right task-related components; Cluster 5: Default components; Cluster 6: Temporal components.

Table 1
Number of voxels in the estimated components that overlap with the corresponding masks ($p = 0.0005$, except for (*) where $p = 0.05$ as explained in the text).

Algorithm	Components			Temporal	Default mode
	Right task	Left task	Transiently task		
SPM	345(*)	298	-	-	-
Infomax	336	214	1781	1648	1097
JADE	194	289	1492	518	1147
FastICA tanh	356	211	1533	1579	976
FastICA pow3	389	247	1668	1412	1090
FastICA gauss	398	191	1742	1376	1083
EVD	1140	48	475	301	2659

Table 2

Correlations of the estimated time courses with the model time courses.

Algorithm	Time courses	
	<i>RTR</i>	<i>LTR</i>
<i>Infomax</i>	0.8148	0.9032
<i>JADE</i>	0.8577	0.9189
<i>FastICA tanh</i>	0.7878	0.9003
<i>FastICA pow3</i>	0.8161	0.9063
<i>FastICA gauss</i>	0.8008	0.9147
<i>EVD</i>	0.3776	0.1483