◆ Human Brain Mapping 6:368–372(1998) ◆

Independent Component Analysis of fMRI Data: Examining the Assumptions

Martin J. McKeown[1]* and Terrence J. Sejnowski[1,2]

[1]*Computational Neurobiology Laboratory, Salk Institute for Biological Studies, La Jolla, California*
[2]*Department of Biology, University of California at San Diego, La Jolla, California*

Abstract: Independent component analysis (ICA), which separates fMRI data into spatially independent patterns of activity, has recently been shown to be a suitable method for exploratory fMRI analysis. The validity of the assumptions of ICA, mainly that the underlying components are spatially independent and add linearly, was explored with a representative fMRI data set by calculating the log-likelihood of observing each voxel's time course conditioned on the ICA model. The probability of observing the time courses from white-matter voxels was higher compared to other observed brain regions. Regions containing blood vessels had the lowest probabilities. The statistical distribution of probabilities over all voxels did not resemble that expected for a small number of independent components mixed with Gaussian noise. These results suggest the ICA model may more accurately represent the data in specific regions of the brain, and that both the activity-dependent sources of blood flow and noise are non-Gaussian. *Hum. Brain Mapping 6:368–372, 1998.*

Key words: fMRI; independent component analysis; statistical analysis

INTRODUCTION

Functional magnetic resonance imaging (fMRI) data sets contain mixtures of many different sources of variability. Physiological signals, including brain activations and cardiac and respiratory pulsations, may overlap spatially and temporally with fluctuations due to subtle head movements and machine or environmental noise. Extracting the small task-related changes in the fMRI signal, typically ~10–15% of the variance at 1.5 T, is therefore difficult because the relationship between the sources of variability may vary across space and over time. Univariate techniques approach this problem by examining each voxel individually, to determine if a given voxel is deemed task-related by a specified criterion, such as a predefined level of significance of a Student t-statistic under the null hypothesis that the distribution of voxel values are identical during control and experimental conditions. Voxels considered task-related are then assembled to form a spatially distributed map of task-related activation.

Frequently, fMRI experiments reveal coactivation of spatially disparate brain regions, which cannot be rigorously investigated with univariate techniques, because they ignore the relationships between voxels. The time courses from two voxels, for example, may both be individually correlated with the task reference function (an estimate of expected task-related changes seen in a voxel) above a certain threshold, yet be uncorrelated with one another.

In contrast, multivariate techniques separate the data into a set of spatial patterns or maps that together

Contract grant sponsor: Heart and Stroke Foundation of Ontario, Canada.

*Correspondence to: Dr. M.J. McKeown, Computational Neurobiology Laboratory, Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, CA 92037–1099. E-mail: martin@salk.edu

Received for publication 13 February 1998; accepted 30 June 1998

compose the data, enabling the analysis of coactivation in spatially divergent areas within a given map. In a linear decomposition of fMRI data, the data matrix can be transformed into a set of volume maps, $\boldsymbol{C}$, by taking linear combinations, defined by an n by n matrix $\boldsymbol{W}$, of the volumes recorded at each time point:

$$\boldsymbol{C} = \boldsymbol{W}\boldsymbol{X} \quad (1)$$

where $\boldsymbol{C}$ is an $n \times v$ matrix of component maps (where n is the number of time points in the experiment and v is the number of brain voxels), $\boldsymbol{X}$ is an $n \times v$ row mean-zero data matrix with each row representing the entire volume recorded at each given time point, and $\boldsymbol{W}$ is an $n \times n$ matrix containing combinations of volumes. In principal component analysis (PCA) [Friston et al., 1993], $\boldsymbol{W}$ in Eq. (1) is selected so that the resultant component maps $\boldsymbol{C}$ are uncorrelated and summarize the variability in the data in as few maps as possible. Independent component analysis (ICA) [Comon, 1994; Bell and Sejnowski, 1995] is a generalization of PCA that selects $\boldsymbol{W}$ in Eq. (1) so that the rows in $\boldsymbol{C}$ are made maximally statistically independent. The stricter criteria for spatial independence used by ICA appear to improve estimates for the temporal and spatial extent of task-related activity, and provides a practical means for exploratory analysis of fMRI data [McKeown et al., 1998b,c].

If $\boldsymbol{W}$ in Eq. (1) is an invertible matrix, then the data can be recovered from the components:

$$\boldsymbol{X} = \boldsymbol{W}^{-1}\boldsymbol{C}. \quad (2)$$

The columns in $\boldsymbol{W}^{-1}$ give the time course of activation for the spatial maps. Unlike PCA, ICA allows that the time courses be nonorthogonal.

Eq. (2) implies that the recorded data can be accurately modeled as component maps, $\boldsymbol{C}$, linearly combined as specified in the matrix $\boldsymbol{W}^{-1}$. As the fMRI data change through time, Eq. (2) assumes that this is a result of changes in the relative contributions from each of the component maps rather than of changes in the component maps themselves, i.e., the maps are assumed to be fixed throughout the fMRI experiment. Eq. (2) also implies that the relative contribution from each component map at a given time point in the experiment is the same throughout the head.

When ICA is used to determine $\boldsymbol{W}$ in Eq. (1), the additional assumption is made that fMRI data are composed of the linear sum of spatially independent patterns of activity. Task-related activations which vary in space and time can then be modeled as a consistently task-related map, and as several spatially independent transiently task-related maps, each with unique time courses, so that the sum of all task-related components provides a measure of the full spatiotemporal extent of task-related activity. In the ICA implementation used for fMRI analysis [McKeown et al., 1998b,c], there was no explicit noise model; rather, the noise was assumed to be distributed among one or more of the components.

If any of the above assumptions are not valid, then the ICA algorithm will be less able to separate out statistically independent component maps. The estimated probability of observing the data under the null hypothesis that the ICA assumptions are valid will therefore be reduced.

More formally, we can relatively easily estimate the probability of observing the i^{th} voxel's time course under the model specified by Eq. (2), i.e., $P(X_i \mid W)$, by exploiting the fact that ICA attempts to separate the data into spatially independent components (see Methods). The minus log-likelihood of the data, given the model for each voxel i, defines a function $u(v_i)$, a dimensionless sequence of numbers. The $u(v_i)$ function can be smoothed using a spatial filter to create a smoothed function, $u_s(v_i)$, that quantifies the degree to which the ICA model fits differing regions of the brain.

In this report, we calculate $u_s(v_i)$ from a representative data set to examine the validity of the following ICA assumptions: 1) constant mixing of components throughout the brain, 2) linear mixing of components, and 3) number of components contained in the data being the same as the number of time points in the experiment.

METHODS

FMRI data recorded from one subject performing a 6-min trial of a Stroop color-naming task were used for exploratory analysis. During 40-sec control blocks, the subject was simply required to covertly name the color of a displayed rectangle. During 40-sec experimental Stroop-task blocks, the subject was asked to name the color of the script used to print a color name (i.e., ''red,'' ''green,'' or ''blue''). Each color name was displayed in a different color from the one it was named. The data were collected from 8 slices consisting of 64×64 voxels, with TR = 2.5 sec. The data were then temporally smoothed and the ICA weight matrix in Eq. (1) was determined using methods reported elsewhere [McKeown et al., 1998b,c].

The $u(v_i)$ function for the data set was calculated by first determining the likelihood of observing the i^{th} voxel's time course, $\boldsymbol{X_i}$, under the model specified in Eq. (2) by $\boldsymbol{W^{-1}}$ and $\boldsymbol{C}$:

$$P(X_i|W) = \det(W)P(C_i) \tag{3}$$

where $\boldsymbol{C_i}$ is the i^{th} column of the matrix in Eqs. (1,2). Since ICA selects $\boldsymbol{W}$ in Eq. (1) such that $\boldsymbol{C}$ is separated into rows that are approximately statistically independent, then computation of Eq. (3) can be estimated by:

$$P(X_i|W) = \det(W)P(C_i) \approx \det(W)\prod_{k=1}^{n} P_k(C_{ki}) \tag{4}$$

or equivalently:

$$-\log(P(Xi|W)) \approx -\log(\det(W)) - \sum_{k=1}^{n} \log(P_k(C_{ki})) = u(v_i) \tag{5}$$

which defines the dimensionless sequences of numbers, $u(v_i)$, with v_i ranging over all voxels.

An estimate of the probability distribution function for each of the $\boldsymbol{P_k}$ $\boldsymbol{(C_k)}$ was obtained by taking a smoothed histogram of each row of $\boldsymbol{C}$

(over the whole volume), from which the probability of the i^{th} point in the row, $\boldsymbol{P_k(C_{ki})}$, was determined [McKeown et al., 1998a].

A smoothed function, $u_s(v_i)$, was determined by spatially smoothing the $u(v_i)$ function with a three-dimensional 6-mm full-width-at-half-maximum Gaussian kernel.

In order to estimate the number of spatially independent components contained in fMRI data, a combined PCA / ICA approach was used to separate the data. The PCA matrix was first partitioned:

$$V = [V_p|V_{n-p}] \tag{6}$$

where V is an $n \times n$ matrix whose columns are the eigenvectors of the data covariance matrix, $\boldsymbol{XX^T}$, Vp is the $n \times p$ submatrix whose columns are the eigenvectors corresponding to the p largest eigenvalues, and V_{n-p} is the submatrix composed of the eigenvectors corresponding to the remaining $n - p$ eigenvalues.

The dimensionality of the data was reduced by:

$$X_p = V_p^T X \tag{7}$$

where X_p is a $p \times v$ reduced-dimension data matrix composed of eigenimages. Blind separation of X_p by ICA was performed, yielding:

$$C_p = W_p X_p \tag{8}$$

where C_p is a $p \times v$ matrix of components and W_p is a $p \times p$ square ICA unmixing matrix. Substituting for X_p from Eq. (7) gives:

$$C_p = W_p V_p^T X; \tag{9}$$

The combined, partitioned PCA / ICA weight matrix was then created by:

$$W_c = \left[\frac{W_p V_p^T}{V_{n-p}^T}\right] \tag{10}$$

Spatial components were then separated with the combined weight matrix:

$$C = W_c X \tag{11}$$

The data were separated into spatial components using a combined weight matrix, with the value of p in Eq. (11) ranging from 10–140, the total number of time points in the experiment. After each separation by the combined ICA / PCA matrix, the mean $u(v_i)$ function, calculated over all brain voxels, was determined.

A simulation was performed to determine the ability of the combined ICA / PCA matrix to estimate the number of sources in an artificial data set. A reduced set of eigenimages was created, $\boldsymbol{X_p}$, using Eq. (8) and $p = 50$. A random 50×140 mixing matrix, $\boldsymbol{M}$, was used to create a deficient-rank matrix, $\boldsymbol{X}_{sim} = \boldsymbol{M}\ \boldsymbol{X_p}$. The rank-deficient matrix, $\boldsymbol{X}_{sim}$, was made full-rank by adding noise, sampled from a Gaussian distribution, to each element of $\boldsymbol{X}_{sim}$. This simulated data set with added noise was separated by combined weight matrices, using a range of values for p in Eq. (9), and the mean $u(v_i)$ was calculated.

RESULTS

The ICA algorithm separated the data into 140 components, one of which was consistently task-related, and several of which were transiently task-related or quasiperiodic, or had ring-like spatial distributions suggesting subtle head movements [McKeown et al., 1998b,c]. The $u_s(v_i)$ function for the data set revealed a fairly clear distinction between white and gray matter in the brain and blood vessels (Fig. 1a).

To determine if the ICA model fit less well due to the violation of the assumption of constant linear mixing

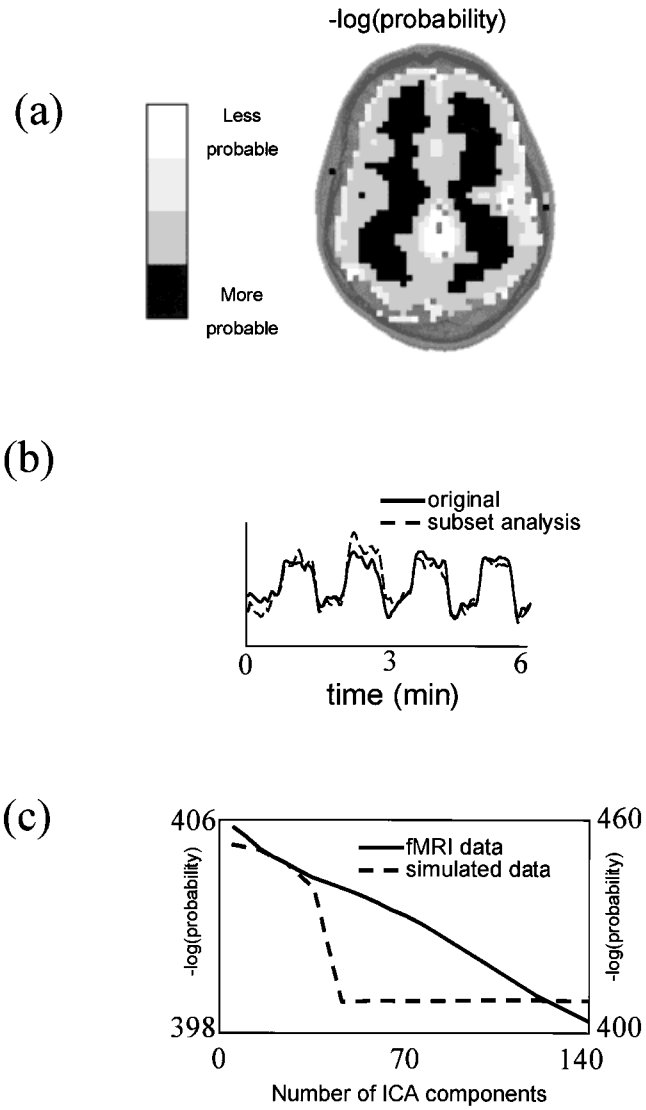

Figure 1.

"The $u_s(v)$ function for the fMRI data. Voxels whose time courses are least likely to be observed under an ICA model are white, more probable voxels are darker. Note the approximate differentiation between cortical and sub-cortical white matter regions." b) The consistently task-related component was similar between brain regions. An ICA decomposition restricted to the subset of voxels fitting less-well in (a) revealed a consistently task-related component (dotted line) very similar to the original decomposition (solid line). c) Estimating the number of components. Combined PCA/ICA weight matrices separated the data into spatial components. The mean u(v) is plotted as a function of *p*, the position of the PCA/ICA partition. For actual data (solid line, left axis), there was a progressive decline as more and more components were separated. In simulated data consisting of 50 components and added gaussian noise, (dotted line, right axis), there was an abrupt change at $p = 50$.

of the components throughout the head, those voxels fitting less well (mostly cortical voxels, n = 4,294) were selected and a new unmixing matrix was determined by the ICA algorithm, using only these voxels. The time course of the consistently task-related component in this new decomposition was very similar (Fig. 1b), and when the new unmixing matrix was applied to the whole data set, the resultant $u(v_i)$ correlated highly ($r = 0.995$) with the original $u(v_i)$ when the previously-determined unmixing matrix was used (not shown).

Increasing the relative amounts of ICA relative to PCA in the combined weight matrix and applying this to the fMRI data set resulted in progressive decreases in the mean $u_s(v_i)$ across all voxels in the head (Fig. 1c). The simulated data set, based on the first 50 principal components of the data and added, purely Gaussian noise, demonstrated an abrupt change in the slope of the mean $u_s(v_i)$ vs. *p* curve after the first 50 components (Fig. 1c).

DISCUSSION

ICA provides a method to "blindly" separate the data into spatially independent components, enabling exploratory analysis on fMRI data [McKeown et al., 1998b,c]. The key assumptions that ICA makes are that the data set consists of *p* spatially independent components, which are linearly mixed and spatially fixed. The number of components extracted, *p*, can be reduced by first preprocessing the data with PCA. As higher-order statistics are used to enforce stricter criteria for spatial independence between maps, better estimates for the consistently task-related components have been obtained [McKeown et al., 1998b,c], suggesting that spatial independence is a reasonable assumption. However, spatial *dependence* between consistently task-related and transiently task-related components can be inferred by the changes in the transiently task-related maps when the consistently task-related component is removed [McKeown et al., 1998b].

Figure 1a demonstrates that for this data set, there is a distinct spatial structure to regions where the ICA model fits less well, with white matter being better modeled than either cortex or regions around blood vessels. Restricting the analysis to mostly cortical and vessel voxels revealed a similar time course for the consistently task-related component (Fig. 1b) and resulted in essentially the same spatial pattern for $u_s(v_i)$, suggesting that different linear mixing of assumed underlying components is not the reason for the ICA model fitting less well in cortical areas. The spatial

variances in the likelihood of the ICA model were not due to difference in mean values between gray- and white-matter voxels, because making all voxels zero-mean before separating by ICA did not affect the spatial pattern of $u_s(v_i)$ (not shown). Several possibilities to explain the difference are, first, that the number of spatially independent components differed between the more metabolically active gray matter and the white matter, and second, that there may have been nonlinear mixing between spatial components, suggesting a limitation for linear models and a potential role for models incorporating nonlinear mixing [Lee et al., 1997].

The simulation results indicate that separation based on a combined ICA / PCA matrix is capable of detecting the number of sources in artificially created data sets with additive, purely Gaussian noise. The fact that the mean $u_s(v)$ steadily declined without a steep falloff as ICA was used to separate greater number of eigenimages from the actual fMRI data suggests that even the eigenimages explaining the smallest variance in our data still had a statistical structure unlike Gaussian noise. This may have implications for analysis techniques that assume fMRI data to consist of underlying components and additive, purely Gaussian noise [Friston, 1996].

These exploratory results suggest that advances in the application of ICA to fMRI data may require addressing possible nonlinear interactions between components, and / or the performance of separate analyses on different subsets of the brain, such as cortical and white matter.

ACKNOWLEDGMENTS

M.J.McKeown. is supported by the Heart and Stroke Foundation of Ontario, Canada. The authors are grateful for many useful discussions with Drs. Michael Lewicki, Bruno Olshausen, Scott Makeig, Tzyy-Ring Jung, Tony Bell, and Te-Won Lee. We are indebted to Drs. Sandy Kindermann and Greg Brown for providing us with their fMRI data.

REFERENCES

Bell AJ, Sejnowski TJ (1995): An information-maximization approach to blind separation and blind deconvolution. Neural Comput 7:1129–1159.

Comon P (1994): Independent component analysis: A new concept? Signal Processing 36:11–20.

Friston KJ (1996): Statistical parametric mapping and other analyses of functional imaging data. In: Toga AW, Mazziotta JC (eds): Brain Mapping: The Methods. San Diego: Academic Press, pp 363–396.

Friston KJ, Frith CD, Liddle PF, Frackowiak RS (1993): Functional connectivity: The principal-component analysis of large (PET) data sets. J Cereb Blood Flow Metab 13:5–14.

Lee T-W, Koehler BU, Orglmeister R (1997): Blind source separation of nonlinear mixing models. In: J. Principe (ed.) Proceedings of "IEEE International Workshop on Neural Networks for Signal Processing," pp 406-415 New York, IEEE, Inc. (Florida, 1997).

McKeown MJ, Humphries C, Achermann P, Borbely A, Sejnowski TJ (1998a): A new method for detecting state changes in the EEG: Exploratory application to sleep data. J Sleep Res, 7 (suppl. 1), 48–56. (in press).

McKeown MJ, Jung T-P, Makeig S, Brown GG, Kindermann SS, Lee T-W, Sejnowski TJ (1998b): Spatially independent activity patterns in functional magnetic resonance imaging data during the Stroop color-naming task. Proc Natl Acad Sci USA 95:803–810.

McKeown MJ, Makeig S, Brown GG, Jung T-P, Kindermann SS, Bell AJ, Sejnowski TJ (1998c): Analysis of fMRI data by decomposition into independent spatial components. Hum Brain Mapp 6:160–188.