1997

Functional Magnetic Resonance Imaging (fMRI) Data Interpreted As Spatial Mixtures*

Martin J. McKeown* MD, Scott Makeig PhD, Greg G. Brown PhD, Tzyy-Ping Jung PhD, Sandra S. Kindermann PhD, Terrence J. Sejnowski PhD.

Computational Neurobiology Laboratory, Salk Institute for Biological Studies,
La Jolla, CA 92186-5800 {martin, scott, jung, terry}@salk.edu and Dept. of Psychiatry,
School of Medicine University of California, San Diego La Jolla, CA 92093
* Author to whom correspondence should be addressed.

ABSTRACT

Functional Magnetic Resonance Imaging (fMRI) is a non-invasive method for examining brain hemodynamics, and indirectly, neural activity during performance of psychomotor tasks. Most analytical techniques being applied to fMRI data require averaging over the full time course of an experiment to make inferences about spatial distributions of activity. However, the gross anatomy of the brain and the cerebral vasculature are fixed throughout an fMRI experiment, suggesting a meaningful interpretation to decomposing the measured signal into stationary maps of 3-dimensional activity that are variably activated through time. Here we use the statistical theory of Independent Components Analysis (ICA) to decompose the fMRI data sets from three normal subjects performing two trials of Stroop color-naming and control tasks into virtually spatially independent fMRI "components". Each ICA component consists of a spatially stationary 3-dimensional component map and an associated time course of activation. The ICA method extracted more than 140 components for each trial. In all cases, ICA found one component with a corresponding time course closely matching the experimental block design. Time courses of other ICA components were transiently task-related, periodic, or slowly varying. The application of ICA to fMRI is the mathematical dual of its application to evoked related potential (ERP) and electroencephalographic (EEG) data. ICA can detect and separate non-task related signal components, movements, and other artifacts, as well as both transient and sustained task-related changes in fMRI data, without *a priori* assumptions about their time course or distribution. Thus, considering the fMRI signal to be the sum of spatial mixtures appears to be a highly promising way to interpret the data from fMRI studies done in normal and clinical populations.

*The Heart & Stroke Foundation of Ontario, the Office of Naval Research and the Howard Hughes Medical Institute supported this research. Part of this work was presented at the 49th annual meeting of the American Academy of Neurology, April, 1997.

1 Introduction

Functional Magnetic Resonance Imaging (fMRI) is a technique for the non-invasive monitoring of brain activation based on the fact that cerebral neural activity and local blood flow are coupled. Normally, an increase in the neuronal activity in a given area of the brain causes local dilatation of the cerebral vessels, increasing regional blood flow. The resultant increase in local oxygenated hemoglobin is in excess of metabolic need, thus reducing the proportion of deoxyhemoglobin in the vessels. The different paramagnetic susceptibilities of oxygenated and deoxyhemoglobin provide the basis for the completely noninvasive BOLD (Blood Oxygenation Level Dependent) contrast techniques, most used in fMRI studies. fMRI's excellent spatial resolution (typically two to three mm), temporal resolution of about one second, coupled with the fact that the method allows repeated scans on a given subject make it a highly promising technique for observing higher brain function both in normal and neurological patient populations. Specifically, the ability to monitor evolving cerebral function during brain development and recovery from pathological insults promises to be of great use to clinicians.

Many current fMRI experiments use a block design in which the subject is instructed to sequentially perform experimental and control tasks in an alternating sequence of 20-40 second control and experimental blocks. The resultant time series recorded from each voxel are complicated mixtures of high and low frequency activity presenting a formidable challenge for analytical methods attempting to tease apart task-related changes in the time courses of 5,000 - 25,000 voxels.

Correlation techniques [1] are based on the assumption that task-related brain regions should show different fMRI signal levels during task performance. A *reference function,* created by convolving the block design of the behavioral experiment with an estimate of the hemodynamic impulse response function, is correlated with the time series recorded from each voxel. However, even in areas of activation, the task-related signal changes are typically small (< 5%), so other time-varying phenomena must produce the bulk of the measured signals. These phenomena can be conceptualized as multiple concurrent "component processes", each having a separate time course and spatial extent, and each producing simultaneous changes in the fMRI signals of many voxels. If the non-task relevant component processes are monotonic, simple linear detrending [1] can be expected to enhance the accuracy of correlational analysis. However, the time courses of processes related to changes in arousal, task strategy, head position, machine artifact or other endogenous processes occurring during a trial may not resemble simple linear functions.

More general ANOVA-like approaches [2] test the signal at every voxel using univariate measures (e.g., t-tests, or f-tests) under the null hypothesis that the values are distributed under a known probability distribution (typically Gaussian). Voxels in which the signal difference between the task and control conditions exceeds a pre-defined level of significance are selected, resulting in a distributed spatial image giving anatomical areas of significant task-related activation differences. However, ANOVA-like methods are based on the tenuous assumptions that the observations are Gaussian distributed and the time courses of different factors affecting the variance of the fMRI signal can be reliably estimated in advance.

Another drawback of the ANOVA-based and correlational measures is that they typically require grouping or averaging over several task/control blocks. This reduces their sensitivity for detecting transient task-related changes in the fMRI signal, and makes them insensitive to significant changes not time locked to the task block design.

Principal Component Analysis (PCA) has been proposed as a way to look for structure in functional imaging data [3]. However, if task-related fMRI changes are only a small part of the total signal variance, finding the orthogonal eigenimages capturing the greatest variance in the data may give little information about task-related activations or other processes of interest producing the observed signals.

Here we describe a new technique for the analysis of fMRI data based on the statistical method of Independent Component Analysis (ICA) [4].

2 Independent Component Analysis

We propose that the brain regions that combine to form the signals recorded during an fMRI experiment may be represented by independent components, each associated with a single time course of enhancement and/or suppression and a component map. We suppose the component maps, each specified by a spatial distribution of fixed values (one at each voxel), are spatially independent. This means that if $p(C_k)$ specifies the probability distribution of the voxel values C_k in the k^{th} component map, then the joint probability distribution of all n components factorizes:

$$p(C_1, C_2, ..., C_n) = \prod_{k=1}^{n} p(C_k)$$

We further assume that there are relatively few highly active voxels in each map, and that the observed fMRI signals are the sum of the contributions of the individual component processes. With these assumptions, the fMRI signals recorded during the performance of psychomotor tasks can be decomposed into a number of component maps and their associated component activation waveforms.

The Independent Component Analysis (ICA) algorithm of Bell & Sejnowski [4] is an iterative unsupervised learning algorithm, that can perform blind separation of input data into the linear sum of time-varying modulations of maximally independent component maps. The ICA algorithm uses 'infomax' to iteratively determine the unknown unmixing matrix W from which the component maps and time courses of activation can be computed.

The matrix of component maps is given by multiplying the observed data matrix by W,

$$C = WX \qquad (1)$$

where X is the fMRI signal data matrix, W is the unmixing matrix, and C is the matrix of component map voxel values. Note that W is a square matrix of full rank so its inverse, W^{-1}, is well defined.

The columns of the inverse weight matrix, W^{-1}, give the time course of modulations of the individual maps. We constrain W to be square, so the number of independent components extracted is the same as the number of time points in the data.

To find and display voxels contributing significantly to a particular component map, the values in each map can be scaled to z-scores. Voxels whose absolute z-scores are greater than some threshold (e.g., 2) can be considered to be the 'active' voxels for that component. *Negative* z scores indicate voxels whose BOLD signals are modulated *opposite* to the time course of activation of that component.

2.1 The ICA Algorithm

Given the assumptions that component processes can be represented by differentially activated sparse and independent maps whose linear sum equals the observed data, an unmixing matrix W can be determined using the ICA algorithm [4].

Specifically, the algorithm initializes *W* to the identity matrix (*I*), then iteratively attempts to maximize H(y), where,
y = g(C); *C* = *WX*$_s$, and g() is a specified nonlinear function. Here, *X*$_s$ is a 'sphered' version of the data matrix defined by, *X*$_s$ = *PX*, where *P* = $2<X^TX>^{-1/2}$.
In our computations, we use the logistic function for g(): g(C*i*) = (1 + exp(C*i*))-1

The elements of *W* are updated using small batches of data vectors drawn randomly from {*Xs*} without substitution, according to, $\Delta W = \varepsilon\left(\frac{\partial H(y)}{\partial W}\right)W^TW = \varepsilon\left(I + \hat{y}C^T\right)W$ where ε is the learning rate and the vector $\hat{y}$ is defined as: $\hat{y}_i = \frac{\partial}{\partial C_i}\ln\left(\frac{\partial y}{\partial C_i}\right)$ The W^TW term avoids matrix inversion and speeds convergence. During training, the learning rate is reduced gradually until the weight matrix *W* stops changing appreciably.

3. Methods

A total of three normal subject volunteers participated in two 6-minute trials of a Stroop color-naming task.

BOLD signal brain activity was scanned in a 1.5T GE Signa MRI system. Eight-to-ten (5 mm thick, 1 mm inter-slice gap) 64 x 64 echo planar, gradient recalled (TR = 2500ms, TE=40 ms) axial images with a 24 cm field of view were collected for each trial. For each slice, 146 images were collected at 2.5-s sampling intervals. Stimuli were presented one at a time by overhead projector onto a screen placed at the foot of the magnet. In control blocks, the subjects were simply required to covertly name the color of a displayed rectangle (red, blue or green). During experimental Stroop-task blocks, subjects were required to name the color of the script used to print a color name. For example, when the word "red" was presented in blue script, the subject was to think (but not speak) the word "blue." Each trial involved four cycles consisting of a 40-s control block and a 40-s experimental block, followed by a final 40-s control block. The six-minute trial was repeated about 15 minutes after its initial presentation.

For each time point in a trial, temporally smoothed BOLD signals from all brain voxels were placed into subsequent rows of a data matrix. The mean of each row was then subtracted. The ICA algorithm was applied separately to data from two 6-minute trials for each subject.

4. Results

Some maps contained multi-focal groupings of active voxels, while others (usually explaining a small amount of variance of the original data) had diffuse or "speckled" spatial distributions.

4.1 Sustained, task-related components

In all trials, exactly one ICA component out of 140 had a time course highly correlated (r = 0.6-0.8) with the task-block design of the behavioral experiment, but explaining < 1% variance of the data (Figure 1). Maps of active voxels for these task-related components contained areas of activation similar to those detected by standard correlation methods (not shown). Figure 2 superimposes the four 80-s task cycles of the square-wave ICA component in each of the Stroop trials. The fine temporal structure of the activation was stereotyped within subjects. The right side of the figure shows the mean of the 8 ICA component task activations in the two trials from each subject, superimposed on the expected response (one cycle of the reference function used in the correlational analysis). Note that the mean time courses (right column) for each subject were not reliably estimated by the reference function, suggesting the true hemodynamic activation during Stroop task performance was not constant, but tended to decline during the course of the experimental blocks. All three subjects showed greater activation near the beginning of the trial. Subjects also differed in the rise-time of activation.

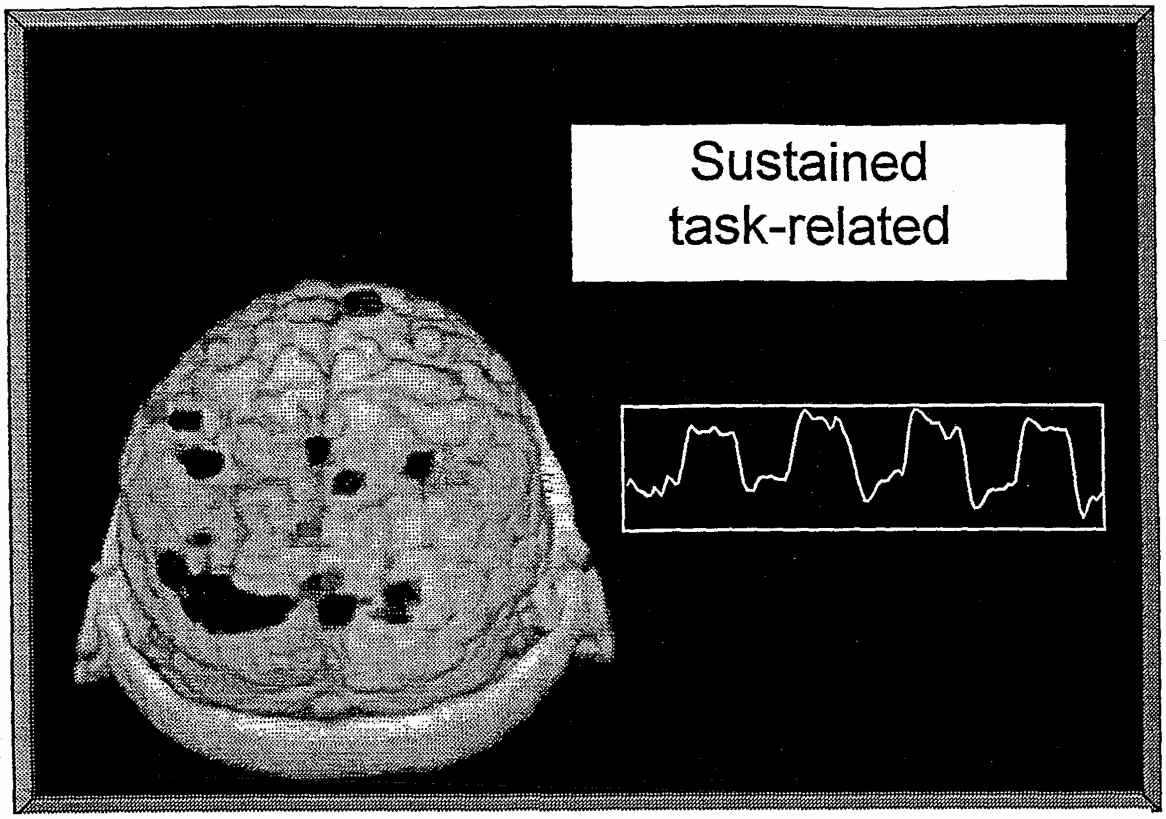

Figure 1. One of the components had a temporal course that was task-related throughout the entire behavioral experiment. The volume rendered brain image depicts the active areas (|z| > 2) in the component map of this component for subject 2, trial 1. The activation time course for this component during the 6-min trial is depicted on the right.

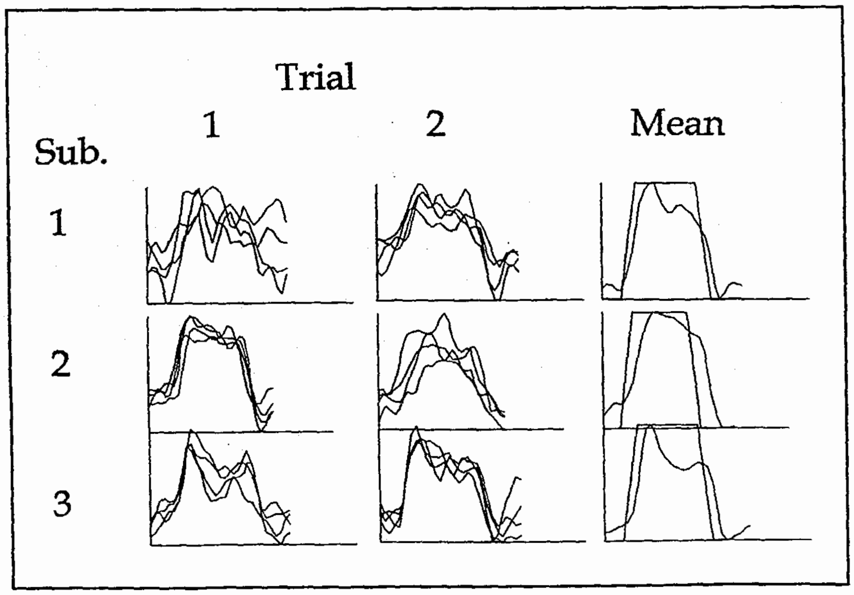

Figure 2: The time courses of the sustained task-related component for each trial of each subject was detrended, separated into the four task/control block pairs, and superimposed. Note the stereotyped activation within subjects.

4.2 Other components

The other 140 or more components for each trial could be grouped empirically into several broad classes, according to their spatio-temporal structure. Some components appeared to be time-locked to the task block design during **part** of their time courses. Such transient task-related activity may not be detected by a correlational analysis that averaged over all the task cycles in a trial. In most trials, there were also slowly varying components. Possibly, the gradual signal changes detected by ICA might reflect temperature or other drift effects in the fMRI recording apparatus. Several components had oscillating time courses with periods near 14 and 40 seconds. Quasi-periodic fMRI signal fluctuations might be caused by aliased cardiac (~1/sec) and respiratory (~1/4 sec) rhythms [5]. Some components either had abrupt changes in their time course (Figure 3) and/or ring-like spatial distributions, suggesting they represented sudden or slow head movements. Simulations we performed (not shown) tended to support this hypothesis. The smallest ICA components had diffuse or "noisy" spatial and temporal patterns and most probably represented noise in the data.

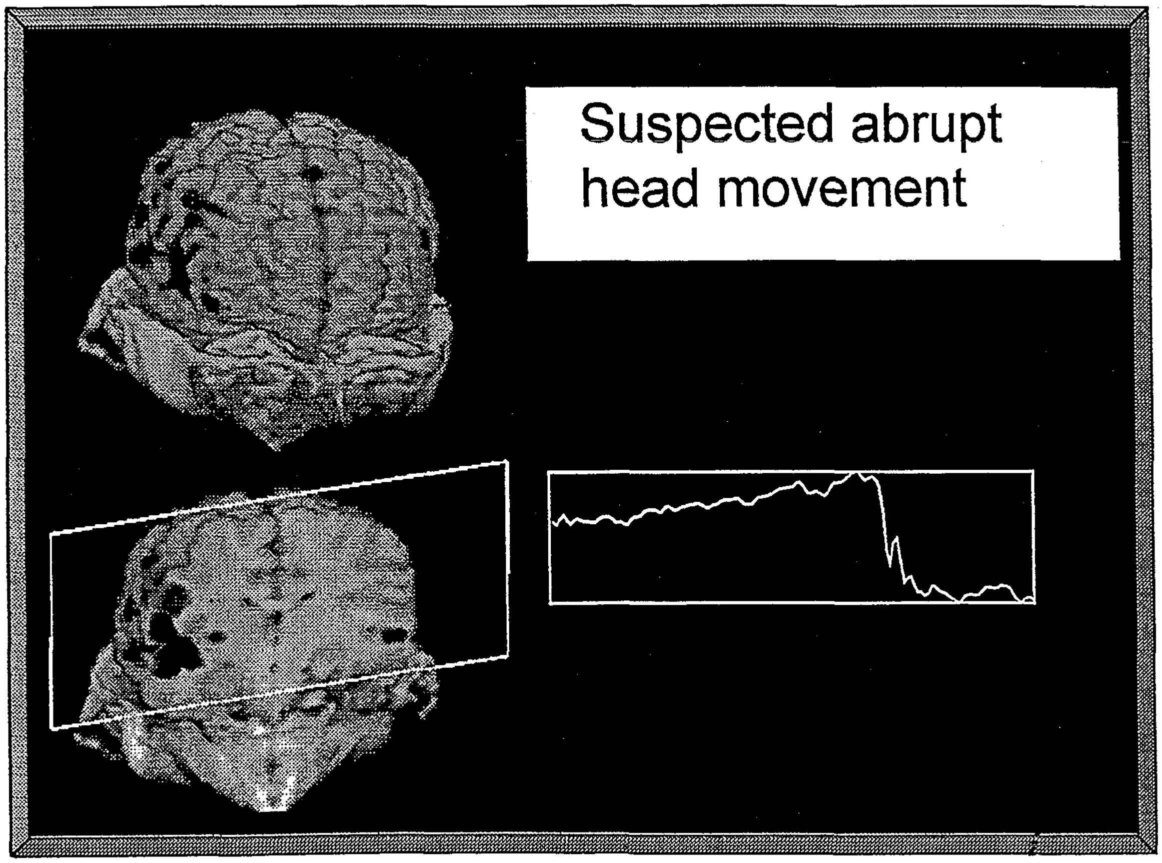

Figure 3. Some components had abrupt time courses and spatial distributions with regions of positive z-values abutting regions of negative z-values. We interpreted these to be the result of head movements during the trial.

5 Discussion

These results indicate that the ICA method can reliably separate sustained and transient task-related and non-task related physiological phenomena as well as machine and movement artifacts that are mixed together in observed fMRI signals. The ICA method can be viewed as a version of the "general linear model" [2] currently used in functional neuroimaging and given by,

$$X = G\beta + e$$

where X is a data matrix, and G is the "design matrix" specifying the time courses of all the factors hypothesized to be present in the observed data (e.g., the task reference function, a linear trend, etc.), β is a matrix of map voxel values for each hypothesized factor, and e is a matrix of noise or residual modeling errors. In contrast, the ICA method extracts intrinsic spatially independent components of the observed data and determines explicitly their time courses, rather than relying on *a priori* hypotheses as to what they should be.

There are several questions about the ICA decomposition of fMRI data that still need to be addressed: As yet, we do not know what proportion of a given component is physiological signal or identifiable artifact, and what is noise. Methods for testing the statistically reliability of ICA component time courses and areas of activation need to be developed.

The ICA algorithm is equally or more sensitive than correlation in finding task-related activations, and can potentially separate artifact, as well as other significant phenomena in the data using only minimal assumptions about the spatiotemporal structure of the signals. Thus, ICA appears to provide a powerful method for exploratory analysis of fMRI data obtained from both clinical and normal subject populations.

REFERENCES

[1] Bandettini PA, Jesmanowicz A, Wong EC, Hyde JS, (1993): Processing strategies for time-course data sets in functional MRI of the human brain. Magn Reson Med 30:161-73.

[2] Friston KJ (1996) : Statistical Parametric Mapping and Other Analyses of Functional Imaging Data. In: A. W. Toga, J. C. Mazziotta eds. Brain Mapping, The Methods. San Diego: Academic Press, 1996:363-396.

[3] 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.

[4] Bell AJ , Sejnowski TJ, (1995): An information-maximization approach to blind separation and blind deconvolution. Neural Comput 7:1129-59.

[5] Biswal B, DeYoe AE, Hyde JS, (1996): Reduction of physiological fluctuations in fMRI using digital filters. Magn Reson Med 35:107-13.

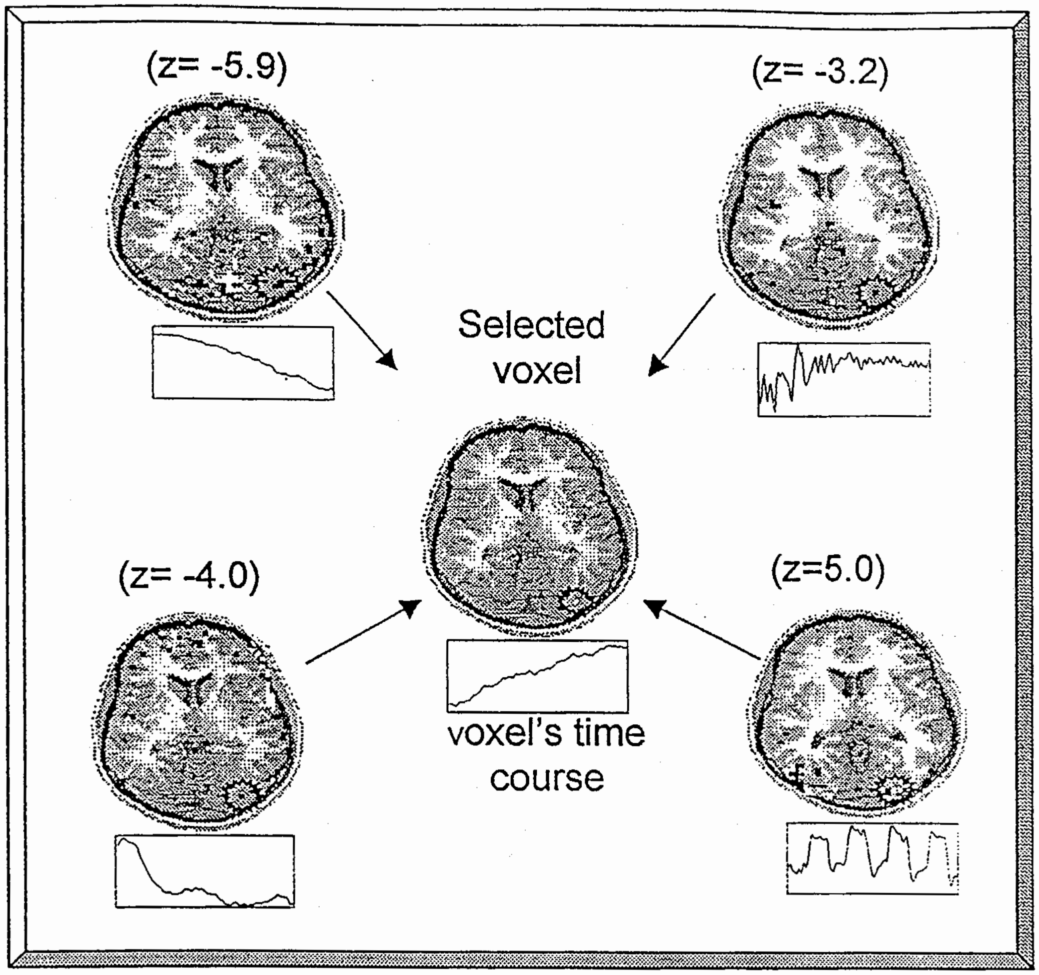

Figure 4. Most voxels analyzed were active in one to six ICA components (i.e., with map values |z| > 2). The figure shows one voxel in a posterior visual association area that participates strongly (z = 5.0) in the sustained task-related component (lower right) as well as in two other larger (ie explaining a larger amount of the variance) and one smaller component.

Figure 4 demonstrates that a single voxel could participate significantly in several ICA components of more than one of the types listed above. The time course of the BOLD signal of the voxel highlighted in the center image is shown below it. This voxel was highly weighted (z=5.0) in the sustained task-related component (lower right), but was also active in three other components of various types, two of which explained a greater percentage of the original variance of the data. Calculations showed that most voxels were active in 1-6 components.