

INDEPENDENT COMPONENT ANALYSIS APPLIED TO FMRI DATA: A NATURAL MODEL AND ORDER SELECTION

V. Calhoun^{1,2}, T. Adali², G. Pearson¹

¹Johns Hopkins University Div. of Psychiatric Neuro-Imaging, 600 N. Wolfe St., Baltimore, MD 21205

²University of Maryland Baltimore County, Dept. of CSEE, Baltimore, MD 21250

ABSTRACT

We introduce a framework for the application of independent component analysis (ICA) to functional magnetic resonance (fMRI) data. We present a model for the task with two main sections: data generation (synthesis) and data processing (analysis) and give examples of how such a model can be utilized in fMRI analysis. We assume a generative model for the data involving 1) the signal being measured (hemodynamic response to neuronal activation) and 2) the instrument being used to measure this data (the fMRI scanner). Such a structure, as we show, can be useful in designing and evaluating the performance of the data analysis methods. In the analysis section, we incorporate a data reduction stage within the model such that information theoretic criteria can be used to estimate the number of effective brain sources. After describing the components of the model in detail, we give examples of its potential use. We compare principle component analysis to clustering and demonstrate that clustering may be a more natural approach than principle components analysis for data reduction in fMRI analysis and can improve source localization. We demonstrate simulated results and results from an fMRI experiment in which two task related waveforms are present.

1. INTRODUCTION

Functional MRI (fMRI) is a tool which measures changes in blood flow and blood oxygenation. When brain neurons are activated, there is a resultant localized change in blood flow and oxygenation. These hemodynamic changes, occur several seconds following the electrical activity in the brain and tend to last 10-15 seconds in duration [1]. fMRI scanners are capable of both a fast sampling rate (relative to the hemodynamic response, e.g. 1-3 seconds) as well the ability to scan subjects over long periods of time.

fMRI analysis approaches range from model-based to exploratory [2]. We propose a hybrid of the two approaches, informed by information such as spatial (3D) and temporal (1D) correlations. We assume that during a given fMRI experiment there are a number of brain regions (networks) which are spatially independent from one another (sources) and are *mixed* together via a network specific hemodynamic time course. This is a reasonable assumption given that (a) brain networks tend to be distributed differently and (b) the changes in blood oxygenation due to neuronal firing often cover large areas (due to draining veins) and appear linear to first approximation at least within sources [3].

ICA has only recently been applied to fMRI data [4]. It has proved promising, but there is a need to study the properties of ICA as applied to fMRI data. There is also a need to develop ICA methods to address specific issues involved in fMRI. Additionally, new methods for evaluating the performance of ICA as applied to fMRI data are needed. When applied to fMRI, ICA can be used to separate sources that are independent in either time or space [5]. For clarity, the following discussion will be limited to spatial-ICA although extension to temporal-ICA is straightforward.

The dimensionality of the data in fMRI (how many image volumes are acquired) is determined by the repeat time (TR) parameter. This

can be changed from scan to scan and has no relationship to the number of sources in the brain. We assume that more time points are acquired than the number of brain sources. This assumption is justified for many fMRI experiments. Because of the high dimensionality of fMRI data, a data reduction scheme is typically applied prior to ICA. The model we introduce (see Figure 1) provides a structured approach for the problem and for studying some of the key assumptions that have to be made, such as independence of the sources, normal distribution of mixtures, etc.

The contribution we provide is not the introduction of a two-stage model *per se*, for such models have been previously proposed for the analysis of fMRI data [6], rather we develop a formalization to understand the interactions between the various parts of the model in the context of fMRI data. Additionally, we are interested in the extension of ICA to analyze groups of subjects, and the formalization we propose is extendable to provide for such studies.

Our goal is to understand the interaction of the assumptions made at all stages including a) preprocessing, b) data reduction (clustering and principal component analysis (PCA)), and c) ICA estimation (in particular infomax [7], maximum likelihood [8], negentropy [9]). In this paper we demonstrate how the choice of data reduction can affect the results.

2. MODELING FOR FMRI

Our model is pictured in Figure 1. In the *data generation block* we assume that there are a set of statistically independent hemodynamic source locations in the brain (indicated by $s_i(v)$ at location v for the i^{th} source). The sources

$$\mathbf{s}(v) \triangleq [s_1(v), s_2(v), \dots, s_N(v)] \quad (1)$$

have weights that specify the contribution of each source to each voxel; these weights are multiplied by each source's hemodynamic time course. Finally, it is assumed that each of the N sources are added together so that a given voxel contains a mixture of the sources, each of which fluctuates according to its weighted hemodynamic time course. This *linear mixing* is represented by the matrix \mathbf{A} and yields

$$\mathbf{u}(v) \triangleq [u_1(v), u_2(v), \dots, u_N(v)]. \quad (2)$$

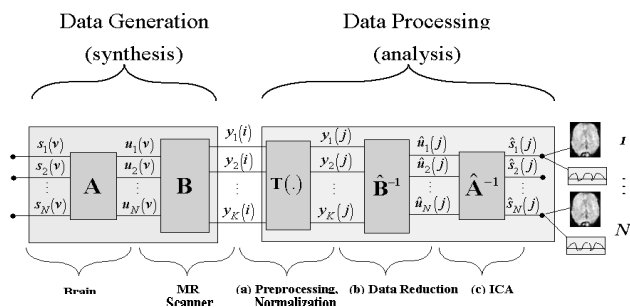


Figure 1: Model for fMRI data in which spatially independent sources are mixed linearly by the hemodynamic response and then overdetermined by the data acquired by the fMRI scanner.

The first portion of the data generation block takes place within the brain. The second portion of the data generation block involves the fMRI scanner. We assume that K time points were acquired with the scanner and that there are more time points acquired than there are sources in the brain. Note that during the fMRI acquisition, the data is discretely sampled (at locations indicated by $i=1,2,\dots,M$ where M is the number of voxels), after which the data is preprocessed (motion correction and smoothing) and spatially normalized into a standard space (at location indicated by $j=1,2,\dots,M$). The sampling of the brain's hemodynamics with the fMRI scanner results in

$$\mathbf{y}(i) \triangleq [y_1(i), y_2(i), \dots, y_N(i)] \quad (3)$$

Hence by introducing the systems, \mathbf{A} and \mathbf{B} , and the signals, $\mathbf{s}(v)$, $\mathbf{u}(v)$, and $\mathbf{y}(i)$, we can develop a formalization which can account for the interaction between the various parts of the model. Each of the systems presented can then be modeled and their suitability to such a task evaluated.

2.1. Data Generation System Models

The matrix \mathbf{A} represents the linear mixing of the statistically independent hemodynamic sources, $\mathbf{s}(v)$, by their temporal hemodynamic responses. Initially, we assume a linear mixing. The signals are processed by the system in parallel and thus it is a memoryless system. It may be represented as:

$$u_j(v) = \sum_i^N a_{ij} s_i(v) \quad (4)$$

where a_{ij} are the elements of \mathbf{A} . System \mathbf{A} is time-invariant, causal, and bounded-input-bounded-output (BIBO) stable. Additionally, there may be feedback occurring at the linear mixing stage. This may be useful for extracting the interaction between different sources. It may also be useful to extend this system to account for the fact that the hemodynamic signal being measured has a point spread function (this can be estimated from optical studies).

System \mathbf{B} represents the sampling of the mixed hemodynamic sources, $u_i(v)$, by the fMRI scanner, and is assumed to be linear. This system is time-invariant, causal, and BIBO stable as well. The signals at this stage are also processed by the system in parallel and thus it is a memoryless system represented by

$$y_k(j) = \sum_i^N b_{ik} u_i(v_j) \quad (5)$$

where $u_i(v_j)$ indicates $u_i(v)$ sampled at $j=1,2,\dots,M$ positions, v_j , and b_{ik} represents the elements of the matrix \mathbf{B} .

The fMRI scanner also has a point spread function that can be estimated and incorporated. For example one can estimate the point spread function, $h_{\text{MRI}}(v)$, over a neighborhood of size L and the linear system becomes

$$y_k(j) = \sum_i^L h_{\text{MRI}}(v_j - i) \left(\sum_j^N b_{ik} u_j(v_i) \right) \quad (6)$$

a delay of L would then be required.

In the *data processing block* we have a transformation, $\mathbf{T}(\cdot)$, representing a number of possible preprocessing stages, including slice phase correction, motion correction, spatial normalization and smoothing. This stage may change the spatial locations, and thus following this stage they are indexed by j

The matrices $\hat{\mathbf{B}}^{-1}$ and $\hat{\mathbf{A}}^{-1}$ indicate systems designed to estimate signals $\hat{\mathbf{u}}(j)$ and $\hat{\mathbf{s}}(j)$, respectively. The resultant source

estimate, $\hat{\mathbf{s}}(j)$, along with the unmixing matrix, $\hat{\mathbf{A}}^{-1}$, can then be presented as fMRI activation images and fMRI time courses, respectively.

2.2. Data Generation Signal Models

The weights of the hemodynamic sources at each voxel location, v , serve as the inputs to our model. These sources, $\mathbf{s}(v)$, are assumed to be spatially independent, that is, the joint source distribution can be represented as

$$p_{\mathbf{s}}(v) \triangleq p_{s_1, s_2, \dots, s_N}(v_1, v_2, \dots, v_N) = \prod_{i=1}^N p_{s_i}(v_i) \quad (7)$$

This is a reasonable assumption which imposes the condition that a different set of areas (perhaps overlapping) are involved for each signal. The results should always be interpreted in light of this assumption as well as the definition of our source signals. Additionally, we assume that each source, $s_i(v)$, is spatially correlated (smooth) in v and can be described as

$$s_i(v) = \int_{u \in P} h_{\text{vas},v}(u) x_i(v-u) du \quad (8)$$

where $h_{\text{vas},v}(u)$ is the vascular point spread function at position v which is nonzero for values of u in a local neighborhood, P , and $x_i(u)$ is the generating "neuronal" source [10].

We do not make any assumptions about the point spread function other than the resulting smoothness of $s_i(v)$. We assume that a given signal is *stationary* across subjects such that

$$p(v, i) = p(v, i+k) \quad \forall k \quad (9)$$

for subject, i , with distribution, $p(v, i)$, at spatial location v . We also assume that the vector, $\mathbf{s}(v)$, consists of signals of interest that are not normally distributed. This is a reasonable assumption for most fMRI signals [11]. Additionally, the mixtures of the spatial source observations may be temporally correlated due to the hemodynamic response function, $h_{\text{hrf}}(t)$ [12]. This can be written

$$b_i(t) = \int_{u \in P} a_i(u) h_{\text{hrf},i}(u-t) du \quad (10)$$

where $h_{\text{hrf},i}(t)$ is the hemodynamic response function for source i and is nonzero for values of t in a local neighborhood, P , and $a_i(u)$ and $b_i(u)$ are source mixing functions. Finally, we assume $\mathbf{s}(v)$ is a continuous valued function and continuous in v .

The signals $\mathbf{u}(v)$, representing the linearly mixed hemodynamic sources in the brain, is a linear combination of the signals in $\mathbf{s}(v)$, which were assumed to be non-Gaussian. As implied by the central limit theorem $\mathbf{u}(v)$, a sum of independent random processes with finite variance, i.e.

$$u(v) = \sum_{i=1}^N \alpha_i s_i(v), \quad (11)$$

will tend to be (more) normally distributed than $\mathbf{s}(v)$. Note that the number of sources, N , need not be large for this argument to hold. Additionally, the elements of the vector, $\mathbf{u}(v)$, will not be independent of one another. This seems reasonable if one considers that the vascular signals are mixed linearly because they are (a) generated from overlapping neural sources and (b) further spread out in space from the original neural sources due to the vasculature [10]. The signals, $\mathbf{u}(v)$, are temporally and spatially correlated due to these properties of $\mathbf{s}(v)$. We assume $\mathbf{u}(v)$ is a continuous valued function, continuous in v .

The signal $\mathbf{y}(i)$ represents the sampling of $\mathbf{u}(v)$ by the fMRI scanner. We know $\mathbf{y}(i)$ is temporally correlated due to properties of $\mathbf{s}(v)$ and $\mathbf{u}(v)$. Depending upon the scan parameters, $\mathbf{y}(i)$ may also be correlated due to magnetic equilibrium effects. This has

been modeled successfully using an autoregressive model. We know that $\mathbf{y}(i)$ is spatially correlated due to the reconstruction algorithm (zero-filled FFT) as well as to the point-spread function of the scanner. It is also known that $\mathbf{y}(i), i=1,2,\dots,N$ are not independent of one another. After sampling by the scanner, the signal is discretely sampled at spatial locations, i .

2.3. Data Reduction Stage

There are many methods for data reduction that will affect the properties of $\hat{\mathbf{u}}(v)$ differently. The most widely used method is principal component analysis (PCA). Another possibility is data clustering using the well-known k-means algorithm. We now outline these two methods, describe methods for estimating how much reduction is needed, and discuss the properties of the reduced signal $\hat{\mathbf{u}}(v)$.

2.3.1. Principal Component Analysis

PCA is a data reduction technique that decomposes the data into a lower dimensional set of orthogonal vectors. The PCA decomposition has been shown to uniquely satisfy the minimum sum of the squares of the errors (and also minimum variance) criterion. This transformation can be regarded as a simple rotation of the coordinate system from the original \mathbf{y}_i 's to a new set of coordinates given by the $\hat{\mathbf{u}}_j$'s. If the data can be completely characterized by second order statistics (such as in the Gaussian case), PCA will provide an optimal decomposition.

Methods for estimating the amount of reduction needed have been developed by using standard information theoretic methods. These methods make a decision by trading off the error in the model with the complexity of the model. Two basic approaches to the problem are the use of Akaike's information criterion (AIC) [13] or the minimum description length criterion (MDL) [14]. Both assume a penalized likelihood form with different model complexity terms, MDL incorporating the number of sample hence enabling consistency of the model order estimation while AIC may perform better at lower signal-to-noise ratios. For PCA decomposition, assuming that the signals, $\mathbf{y}(v)$, have a Gaussian distribution they take the following form [15]:

$$AIC(N) = -2M(K-N)\mathcal{L}(\hat{\theta}_N) + 2N(2K-N), \text{ and} \quad (12)$$

$$MDL(N) = -M(K-N)\mathcal{L}(\hat{\theta}_N) + \frac{N(2K-N)}{2}. \quad (13)$$

$$\mathcal{L}(\hat{\theta}_N) = \frac{(\lambda_{N+1}\dots\lambda_K)^{\frac{1}{K-N}}}{\frac{1}{K-N}(\lambda_{N+1} + \dots + \lambda_K)}. \quad (14)$$

where M is the number of voxels, K is the number of time points and N is the number of sources. The number of sources is determined from the minimum of the above functions. The expression for $\mathcal{L}(\hat{\theta}_N)$ is the ratio of the geometric mean of the $K-N$ smallest eigenvalues to their arithmetic mean. Note that we have written the number of parameters as a function of the number of sources N .

In the context of fMRI data, λ_k represents the variance associated with the best-fit line through the fMRI data in a multi-dimensional least squares sense (consider a graph with K axes, one per time point, and with one point plotted per voxel). The variance associated with the next best-fit line (constrained to be orthogonal to the first line) is represented by λ_{k-1} continuing down to λ_1 . The number of eigenvalues will be equal to the number of time points in the fMRI data set, although many of the smaller eigenvalues will be

very close to zero. The $K-N$ eigenvectors with the smallest eigenvalues are discarded, and the $K-by-K$ matrix of eigenvectors now becomes $N-by-K$ and is the basis upon which the data, $\mathbf{y}(j)$, is projected.

The estimates $\hat{\mathbf{u}}(j)$ after data reduction by PCA will still be linear mixtures of the sources, however important higher order information might have been lost. These signals will not be independent, however the PCA method will force the $\hat{\mathbf{u}}(j)$ to be orthogonal which has implications for the source separation algorithm utilized. Additionally, $\hat{\mathbf{u}}(j)$ will be spatially correlated.

2.3.2. Clustering

Another data reduction approach is to cluster the amplitudes of the time data and use the mean of each cluster as the reduced data. Since a second order statistic is not used as the criteria for reducing the data, such as in PCA, clustering may perform better in certain situations. For clustering, AIC and MDL assume the following forms:

$$AIC(N) = -2\mathcal{L}(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}) + 2G(N), \text{ and} \quad (15)$$

$$MDL(N) = -\mathcal{L}(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}) + \frac{G(N)\ln(K)}{2} \quad (16)$$

where $G(N)$ is the number of parameters in the model, $\mathcal{L}(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})$ is the log likelihood of the ML model parameters, $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$, and N is the number of sources. We formulate this problem as a multivariate Gaussian mixture with the dimensionality equal to the number of spatial data points. Equal prior probabilities for the clusters are assumed, thus for N clusters and M spatial voxels, we have

$$G(N) = N\left(\frac{M^2}{2} + \frac{5M}{2}\right). \quad (17)$$

The likelihood equation then has the form:

$$\mathcal{L}(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}) = \sum_{i=1}^N \ln\left(\pi_i \exp\left(\frac{1}{2}(\mathbf{y} - \hat{\boldsymbol{\mu}}_i)^T \hat{\boldsymbol{\Sigma}}_i^{-1}(\mathbf{y} - \hat{\boldsymbol{\mu}}_i)\right)\right) \quad (18)$$

where the cluster statistics for cluster i are represented by $\boldsymbol{\Sigma}_i$, a diagonal $M-by-M$ covariance matrix and $\boldsymbol{\mu}_i$, the $M-by-1$ cluster mean where M is the number of voxels. If we assume uncorrelated samples such that

$$\boldsymbol{\Sigma} = \frac{1}{\sigma^2} \mathbf{I}, \quad (19)$$

we simplify the calculation considerably and reduce the number of parameters. This is a more attractive approach given the high dimensionality of fMRI data.

After clustering, the reduced signals, $\hat{\mathbf{u}}(j)$, will still be linear mixtures of the sources. These signals will not be independent, and they will also not be orthogonal to one another. Additionally, $\hat{\mathbf{u}}(j)$ will be spatially correlated.

2.4. Mixing Stage

ICA is an application of blind source separation that attempts to decompose a data set into statistically independent components and is proposed by Herault and Jutten [16] and further developed in [17]. ICA is most commonly utilized for source separation, although it can also be used for other applications such as feature extraction, blind deconvolution, and others. ICA can be related to classical methods such as redundancy reduction, projection pursuit, factor analysis, principal component analysis and blind deconvolution [18].

In spatial-ICA (SICA) of fMRI data, suppose \mathbf{S} is an $N \times M$ matrix (where N is the number of sources and M is the number of voxels). The “signals” are the M spatial voxels, flattened to a 1-D vector, and there are thus N different instances of these signals (whereas temporal-ICA (TICA) would consider the “signals” the M individual time courses of which there are N instances). The SICA decomposition can then be described as $\mathbf{S} = \hat{\mathbf{A}}^{-1}\mathbf{Y}$, where $\hat{\mathbf{A}}$ is the $N \times N$ estimated mixing matrix found using ICA, and \mathbf{S} is an $N \times M$ matrix containing the N independent components. We can then write $\mathbf{Y} = \hat{\mathbf{A}}\mathbf{S}$ where the spatially independent components (images) are located in the rows of \mathbf{S} and the associated spatially independent time courses are found in the columns of $\hat{\mathbf{A}}$.

There are several approaches to finding the independent components, a number of which are closely related ways of looking at the same problem from similar information theoretic views. In the simulations to follow, we have utilized a negentropy-based algorithm [9]. We explore the differences between two data reduction schemes.

3. EXPERIMENTS

3.1. Simulation

We generated three sources similar to those encountered in fMRI data. The first distribution is Gaussian whereas the other distributions (modeling fMRI “activations”) are a sum of two Gaussian distributions such that the mean was zero and the kurtosis was positive. The distributions used are presented below (the v variable is suppressed for clarity).

$$\begin{aligned} p_{s_1}(s) &\sim N(0,1) \\ p_{s_2}(s) &\sim N(-1/3,2) + N(1/3,1/32) \\ p_{s_3}(s) &\sim N(-1/2,3) + N(1/2,1/16) \end{aligned} \quad (20)$$

where $N(\mu, \sigma)$ indicates a Gaussian distribution with mean μ and standard deviation σ . These distributions are graphed in Figure 2. For clarity, in this simple example the signals are not mixed (the mixing matrix is diagonal). A plot of the distributions is presented in Figure 2.

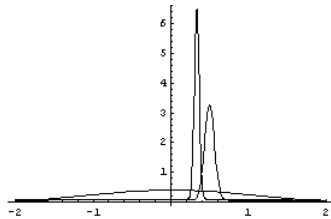


Figure 2: Independent source distributions

When applying methods for estimating the number of sources, the sensitivity of the data reduction method to the number of sources is an important factor. Suppose we incorrectly estimated two sources for this example. We would then reduce these three signals down to two signals using either PCA or clustering. In our example the signals are independent, thus the PCA decomposition will project onto signals u_1 and u_2 since this captures the largest signal variance. The information from source s_3 (which is important as it represents the fMRI “activation”) has been completely lost since all of its information was contained in signal u_3 .

The clustering results will however preserve such information by grouping u_3 into 1) it’s own cluster or 2) into a cluster including one of the other two mixed signals. In either case, information in

$p(s_3)$ is preserved. For example, if u_3 is grouped with another source the resulting distribution will be a weighted combination of the s ’s

$$y_2 = as_2 + bs_3 \quad (21)$$

and can be calculated from

$$p(y_2) = \int_{-\infty}^{\infty} \frac{F'_{s_2}(w/a) F'_{s_3}(y_1 - w/b)}{a b} dw \quad (22)$$

where F indicates the cumulative distribution functions of the sources, and a and b represent the weights. The resulting distribution will preserve information about the fMRI “activation” sources and can then be unmixed by the ensuing ICA analysis. We have thus demonstrated a situation in which clustering provides improved data reduction to PCA. In situations where the signal-to-noise ratio is higher, PCA may provide a better decomposition of the signals than clustering. However for fMRI data we are usually dealing with very small signals and large noise.

In another simulation, we generated two activation sources and mixed them with an fMRI-like hemodynamic function as depicted in Figure 3. Gaussian noise was added to the mixed images as well to more closely simulate an fMRI experiment.

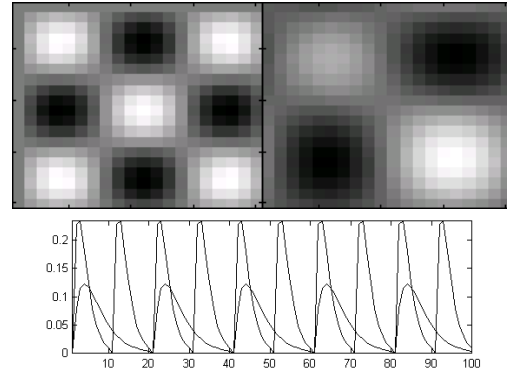


Figure 3: Simulated 20x20 Source locations, top, and simulated hemodynamic mixing functions (100 time points), bottom.

The data were reduced using either PCA or clustering and then unmixed using ICA. The resulting source distributions are depicted below. Note that clustering produced a source distribution more closely resembling the original.

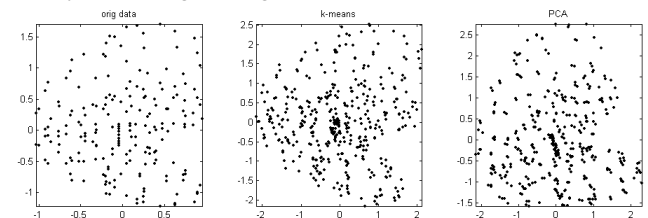


Figure 4: Data after source separation following reduction to 2 dimensions and using clustering (middle) or PCA (right). Note that the clustering approach produces a source projection that matches well with the original sources (left).

3.2. fMRI Experiment

fMRI data were acquired on a Philips 1.5T Scanner. Functional scans were acquired with an echo planar sequence (64x64, flip angle=90, TR=1s, TE=39ms) over a 6-minute period for a total of 360 time points. A visual paradigm was presented in which there was a period during which an 8Hz reversing black and white

checkerboard was presented to either the left or the right visual field (see Figure 5). Activations are expected in the right or left visual cortex, respectively.

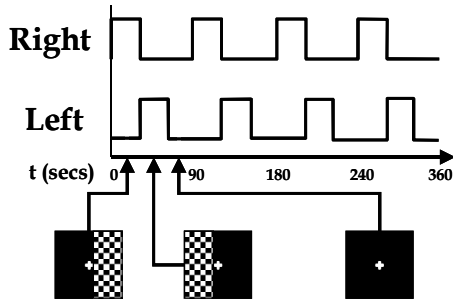


Figure 5: Paradigm for the fMRI experiment. Subjects were alternately presented an 8 Hz flashing checkerboard to the Right eye or the Left eye over a six minute run.

The fMRI data were analyzed using the two-stage model presented above. The number of sources in the data was estimated by both AIC and MDL criterion to be ten (see Figure 6). Results demonstrate successful separation of sources in left and right visual cortices.

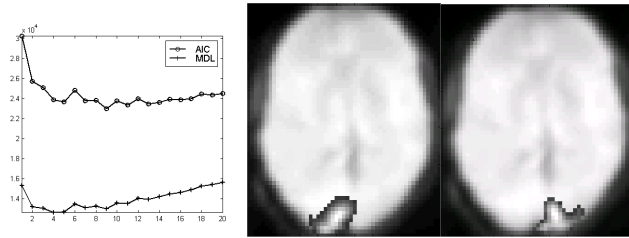


Figure 6: Estimation of number of sources present in fMRI data (right), and thresholded sources demonstrating successful identification of left and right visual cortex.

4. CONCLUSIONS

We have introduced an extendable model for understanding ICA as applied to fMRI data. We have outlined the properties of our generative model and the signals as they propagate through the model. We demonstrate differences between two data reduction schemes when operating on fMRI-like sources. Future work will incorporate each of these properties and examine how each affects the ICA results when utilizing different objective functions and algorithms.

5. REFERENCE LIST

- [1] K.J.Worsley and K.J.Friston, Analysis of FMRI Time-Series Revisited--Again *Neuroimage.*, vol. 2 , pp. 173-181, 1995.
- [2] N.Lange, S.C.Strother, *et al.*, Plurality and Resemblance in FMRI Data Analysis *Neuroimage.*, vol. 10, pp. 282-303, 1999.
- [3] G.M.Boynton, S.A.Engel, *et al.*, Linear Systems Analysis of Functional Magnetic Resonance Imaging in Human V1 *J.Neurosci.*, vol. 16, pp. 4207-4221, 1996.
- [4] M.J.McKeown, S.Makeig, *et al.*, Analysis of FMRI Data by Blind Separation Into Independent Spatial Components *Hum.Brain Map.*, vol. 6, pp. 160-188, 1998.
- [5] V.Calhoun and J.Pekar, Where and Where Are Components Independent? On the Applicability of Spatial- and Temporal- ICA to Functional MRI Data *Neuroimage.*, vol. 11, p. S682, 2000.

- [6] M.J.McKeown, Detection of Consistently Task-Related Activations in FMRI Data With Hybrid Independent Component Analysis *Neuroimage.*, vol. 11, pp. 24-35, 2000.
- [7] A.J.Bell and T.J.Sejnowski, An Information Maximisation Approach to Blind Separation and Blind Deconvolution *Neural Computation*, vol. 7, pp. 1129-1159, 1995.
- [8] J.F.Cardoso, Infomax and Maximum Likelihood for Source Separation *IEEE Letters on Signal Processing*, vol. 4, pp. 112-114, 1997.
- [9] A.Hyvarinen and E.Oja, A Fast Fixed-Point Algorithm for Independent Component Analysis *Neural Computation*, vol. 9, pp. 1483-1492, 1997.
- [10] D.Malonek, U.Dirnagl, *et al.*, Vascular Imprints of Neuronal Activity: Relationships Between the Dynamics of Cortical Blood Flow, Oxygenation, and Volume Changes Following Sensory Stimulation *Proc.Natl.Acad.Sci.U.S.A.*, vol. 94, pp. 14826-14831, 1997.
- [11] M.J.McKeown and T.J.Sejnowski, Independent Component Analysis of FMRI Data: Examining the Assumptions *Hum.Brain Map.*, vol. 6, pp. 368-372, 1998.
- [12] R.L.Buckner, Event-Related FMRI and the Hemodynamic Response *Hum.Brain Map.*, vol. 6, pp. 373-377, 1998.
- [13] H.Akaike, A New Look at Statistical Model Identification *IEEE Trans.on Automatic Control*, vol. 19, pp. 716-723, 1974.
- [14] J.Rissanen, A Universal Prior for Integers and Estimation by Minimum Description Length *Ann.of Statistics*, vol. 11, pp. 416-431, 1983.
- [15] M.Wax and T.Kailath, Detection of Signals by Information Theoretic Criteria *IEEE Trans.Acous.Speech, and Sig.Proc.*, vol. 33, pp. 387-392, 1985.
- [16] J.Herault and J.Jutten, Space or Time Adaptive Signal Processing by Neural Network Models *Neural Networks for computing: AIP conference proceedings*, vol. 151, 1986.
- [17] C.Jutten and J.Herault, Blind Separation of Sources, Part I: An Adaptive Algorithm Based on Neuromimetic Architecture *Signal Proc.*, vol. 24, pp. 1-10, 1991.
- [18] A.Hyvarinen, Survey on Independent Component Analysis *Neural Computing Surveys*, vol. 1, pp. 94-128, 1999.