INDEPENDENT COMPONENT ANALYSIS WITH FEATURE SELECTIVE FILTERING

Yi-Ou Li[1], Tülay Adalı[1], and Vince D. Calhoun[2,3]

[1]Department of Computer Science and Electrical Engineering
University of Maryland Baltimore County, Baltimore, MD 21250
[2]Olin Neuropsychiatry Research Center
Institute of Living, Hartford, CT 06106
[3]Yale University School of Medicine
New Haven, CT 06520

Abstract. In this contribution, we propose a feature selective filtering scheme for independent component analysis (ICA) to improve the estimation of the sources of interest (SOI), *i.e.*, sources that have certain desired features in their sample space. As an example, we show that ICA with a smooth filtering scheme can improve the estimation of the smooth image sources from a mixture of images, as well as the estimation of a smooth visual activation map in a hybrid functional magnetic resonance imaging (fMRI) data set. Hence, the technique can potentially be used in the analysis of fMRI data to improve the ICA estimation of functional activation regions that are expected to be smooth.

INTRODUCTION

Independent component analysis is an exploratory data analysis approach that can interpret the data as a combination of statistically independent sources. It has been successfully applied in a variety of fields in signal processing such as the analysis of fMRI data.

McKeown *et al.* [14] applied ICA to fMRI data analysis based on the assumption of spatial independence among regions of brain activations and artifacts. This model is the so-called spatial ICA (SICA) model [5] and it is consistent with the two principles of the functional organization of brain: localization and connectionism [15]. Based on this model, the practical ICA algorithm gives brain maps whose spatial distributions achieve maximal independence and the spatial distribution here is interpreted as the distribution of map voxel values. However, the spatial feature of the brain map as an image is discarded in this treatment. According to the localization principle of brain activation, the nearby activated voxels have similar intensity.

Specifically, the convolutive effect of the vascular point spread function on the hemodynamic sources implies that smoothness can be a valid assumption for the independent activations of our interest [2]. It is desirable to incorporate constraints such as these into the estimation of SOIs in fMRI data analysis.

Different ways of incorporating priors into ICA estimation have been studied. Bayesian estimation is a systematic framework for estimation of signals with a priori information, *e.g.*, probability distributions, correlation, etc. ICA under this framework has been studied by, *e.g.*, [7, 8, 13, 16]. However the demanding computational load for Bayesian estimation is a major concern given the large volume of fMRI data. Based on the sparsity assumption of the mixing matrix, Hyvarinen [10] proposed ways of incorporating the sparsity constraint in the form of conjugate priors into the ICA estimation. For the case of fMRI data analysis, this kind of prior is only proper for the temporal ICA (TICA) model [5]. For SICA model on fMRI data analysis, Calhoun *et al*. [3] proposed a method of imposing regularization on the time courses based on the experimental paradigm.

In general, it is not straightforward to include priors about the spatial features of functional activation map such as the smoothness of the activated region into SICA for fMRI data. Most ICA algorithms update the demixing vectors in an iterative manner to satisfy the optimality criterion and the source data is not explicitly processed. One possible way to incorporate the spatial feature constraint is to perform a selective filtering in the source space and project the filtering effect back to the space of the demixing vectors. Therefore, the influence of the priors takes place in the estimation through this filtering-projection process. For ICA algorithms employing iterative methods to optimize certain independence measure, the filtering could introduce a proper variance on the convergence of the ICA algorithm leading to a local extremum that gives better estimation of the SOI.

In the next section, the general ICA model is introduced, followed by the description of feature selective filtering scheme within the ICA framework. We then express the filtering scheme in matrix form and explain the effect of the filtering process on ICA estimation. After that, we introduce a controlling mechanism to improve the overall performance of the scheme. The third section contains simulations of this filtering scheme with two ICA algorithms on synthetic image source separation and hybrid fMRI data analysis. The estimation result is compared with that from the original ICA algorithm. Finally, we present a discussion of this feature selective filtering scheme.

ICA WITH FEATURE SELECTIVE FILTERING

ICA model

In the ICA model, it is assumed that a set of statistically independent sources $s_i, i = 1, 2, ..., m$, are mixed linearly by a mixing matrix $\mathbf{A}$, resulting in a set of interrelated observations $x_i, i = 1, 2, ..., m$. The estimation task can be stated as finding a demixing matrix $\mathbf{W}$, such that the independent sources can be recovered from those observations by this demixing transformation.

Various algorithms have been developed based on different principles of independence and different optimization methods. Infomax [1] and FastICA [12] are two of the most widely used algorithms to perform ICA.

For the subsequent discussion, we introduce the notation of ICA model that takes into account the data samples in certain domain, *e.g.*, time sequence, signal strength from spatial locations, etc. We define $\mathbf{S}$ as an $m \times n$ source data matrix where each row of $\mathbf{S}$ is a vector of data samples from one source; and $\mathbf{X}$ as the observed data matrix where each row is a vector of one observed data set. Therefore, the ICA model for data analysis becomes

$$\mathbf{X} = \mathbf{AS} \tag{1}$$

and the estimation task is restated as finding the demixing matrix $\mathbf{W}$ such that

$$\hat{\mathbf{S}} = \mathbf{WX} \tag{2}$$

recovers those independent sources in each row of $\hat{\mathbf{S}}$. Taking example in image data analysis, the source data is a set of image pixels and the feature selective filtering is applied spatially on those image pixels.

For application of SICA on fMRI data, because the number of independent sources of practical interest is typically much smaller than the number of time samples from the observations, dimension reduction is performed as a preprocessing step before the data is provided for ICA estimation. Typically, principal component analysis (PCA) is used as the preprocessing step because it projects the data into a decorrelated set of components with significant variance. Because of the intrinsic ambiguity of the ICA model [12], we assume unit variance on all the independent sources and assume that the mixing vectors have unit norm. As a result, the observations should have unit variance as well. To comply with this condition, we normalize each principal component while performing PCA. Therefore, the resulting data components are whitened, *i.e.*, uncorrelated and of unit variance.

ICA with feature selective filtering

When a priori knowledge of the SOI is expressed as certain feature in its sample space, we can design a filter in that space with the same characteristics, *i.e.*, a feature selective filter and apply it to the intermediate estimates during the iterative ICA estimation process. To project the filtering effect from the sample space to the demixing vector space, one way is to solve the least squares problem: $\mathbf{w}'_i = \arg\min_{\mathbf{w}_i} \|\hat{\mathbf{s}}'_i - \mathbf{w}_i\mathbf{X}\|^2$, where the solution is

$$\mathbf{w}'_i = \hat{\mathbf{s}}'_i\mathbf{X}^T(\mathbf{XX}^T)^{-1}. \tag{3}$$

To study the effect of feature selective filtering on ICA estimation, note that $\mathbf{X}$ is preprocessed by PCA and is prewhitened, so we have $\mathbf{XX}^T =$

$(\mathbf{X}\mathbf{X}^T)^{-1} = \mathbf{I}_{(m\times m)}$. The filtering process on $\hat{\mathbf{s}}_i$ can be represented in matrix form as $\hat{\mathbf{s}}'_i = \hat{\mathbf{s}}_i\mathbf{H}$, where $\mathbf{H}$ is the convolution matrix of the feature selective filter. Using the whitening condition and the ICA model given in (1), we can write

$$\mathbf{w}'_i = \hat{\mathbf{s}}_i\mathbf{H}\mathbf{S}^T\mathbf{A}^T. \tag{4}$$

Now, we express the intermediate source estimates $\hat{\mathbf{s}}_i$ in the form of a weighted combination of the true sources as

$$\hat{\mathbf{s}}_i = \boldsymbol{\alpha}_i\mathbf{S} \tag{5}$$

where $\boldsymbol{\alpha}_i = [\alpha_{i1}, \alpha_{i2}, ..., \alpha_{im}]$ is a vector with the weights of each source component in the ith estimated source. Substituting (5) into (4), the demixing vector $\mathbf{w}'_i$ becomes

$$\mathbf{w}'_i = \boldsymbol{\alpha}_i\mathbf{S}\mathbf{H}\mathbf{S}^T\mathbf{A}^T. \tag{6}$$

At this point, we compare two cases to explain the effect of filtering on the estimated demixing vectors:

1. When no filtering is applied, *i.e.*, $\mathbf{H}$ is an identity matrix, the term $\mathbf{S}\mathbf{H}\mathbf{S}^T$ is also identity by the assumption that all the true sources are independent and of unit variance. The ICA algorithm forces the weight vector converge to an m-dimensional unit vector with only one nonzero element. Correspondingly, the demixing vector converges to one column of the mixing matrix $\mathbf{A}$, which means the condition of estimation of one true source is achieved (Theorem 1 [11]).

2. When filter is applied to each intermediate estimate $\hat{\mathbf{s}}_i$, we define

$$\mathbf{B} \equiv \mathbf{S}\mathbf{H}\mathbf{S}^T = \mathbf{S}'\mathbf{S}^T \tag{7}$$

as the correlation matrix between the filtered sources $\mathbf{S}'$ and the true sources $\mathbf{S}$. Substituting (7) in (6), we have

$$\mathbf{w}'_i = \boldsymbol{\alpha}_i\mathbf{B}\mathbf{A}^T \tag{8}$$

Because each source is filtered individually, no correlation will be introduced across different sources by this filtering process. In other words, if $\mathbf{S}\mathbf{S}^T$ is identity, $\mathbf{B}$ will be a diagonal matrix with the correlation coefficients of the original sources and their filtered counterparts on the main diagonal, *i.e.*, will be given by

$$\mathbf{B} = \begin{bmatrix} \mathbf{s}'_1{\mathbf{s}_1}^T & 0 & .. & 0 \\ 0 & \mathbf{s}'_2{\mathbf{s}_2}^T & .. & 0 \\ \\ 0 & 0 & .. & \mathbf{s}'_m{\mathbf{s}_m}^T \end{bmatrix}. \tag{9}$$

From (8) we can see that, vector $\boldsymbol{\alpha}_i$ is weighted by this correlation matrix. For an SOI, because its feature complies with the characteristics of the filter, the correlation coefficient will be high. Therefore, the element in $\boldsymbol{\alpha}_i$ representing that source component is relatively amplified. Meanwhile, the demixing vector will produce an estimation closer to that SOI.

Implementation of the filtering scheme

Equation (8) suggests that the filtering effect is the same for all the demixing vectors. As a result, it will degrade the estimation of the sources not of interest when improving the estimation of SOIs. To avoid this negative effect, the filtering should be carried out in a controlled manner: (i) the filtering should be applied after a certain level of convergence of the original ICA algorithm so that the estimated source images start to show the right trend in their spatial pattern; (ii) before filtering is applied, the a priori feature of each intermediate estimate needs to be evaluated in order to identify those converging to SOIs. The identified ones are then subject to the filtering process. Motivated by (9), we define the measure of the desired feature as

$$\beta_i = \hat{\mathbf{s}}_i' \hat{\mathbf{s}}_i^T. \tag{10}$$

In practice, this measure is taken on all the intermediate estimates and a selection is made to pick out a subset of them for feature selective filtering.

For an ICA algorithm adopting iterative method, a pseudo code of implementing feature selective filtering is shown below.

Algorithm 2.1

while {
  Check for termination condition;
  $\mathbf{W}_{(k+1)} = \text{ICA update}(\mathbf{W}_{(k)})$;
  $\hat{\mathbf{S}} = \mathbf{W}_{(k+1)}\mathbf{X}$;
  for i = 1 to m
  {
  $\hat{\mathbf{s}}_i' = \text{filter}(\hat{\mathbf{s}}_i)$;
  if($\hat{\mathbf{s}}_i' \hat{\mathbf{s}}_i^T >$ threshold) {
  $\mathbf{w}_i' = \hat{\mathbf{s}}_i' \mathbf{X}^T (\mathbf{X}\mathbf{X}^T)^{-1}$;
  }end **if**
  }end **for**
  }end **while**

SIMULATIONS

We choose 2D smoothing filter as the feature selective filter in our simulations, *i.e*, smooth sources define our SOI. The filtering is applied after the original ICA algorithm reaches a moderate convergence. Specifically, we apply the filtering if the change of demixing vectors between two adjacent iterations is less than an error tolerance $\epsilon_1 > \epsilon_0$ where ϵ_0 is the final convergence error tolerance of the entire ICA algorithm. PCA and whitening are performed as the preprocessing steps of ICA. All the results are based on 20 independent runs of the algorithm with random initial condition for each run. The algorithm is written in Matlab programming language and executed on Dell PC with Intel Pentium-4 CPU and 512M RAM.

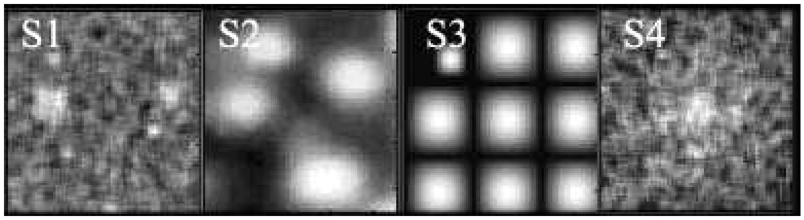

Figure 1: Synthetic image set

TABLE 1: STATISTICS OF THE SYNTHETIC SOURCE IMAGES

Source	S1	S2	S3	S4
Kurtosis	2.24	−0.69	−1.20	0.41
Smoothness	0.94	0.95	0.99	0.94

Synthetic Image Source Separation

In the synthetic image source separation test, we generate four 90×90 pixel gray level images as the independent sources. The mixing matrix contains three gamma function waveforms of different periods resembling the temporal dynamics and one random waveform resembling dynamics of a noise source.

Based on the smoothness measure of the true sources, we choose a moving average filter with a 3×3 kernel. Such a kernel size is proper for introducing a reasonably small variation when filtering the image data.

The four image sources are displayed in Figure 1 and two relevant statistics of each source are tabulated in Table 1. The measure of smoothness is defined in (10). Two ICA algorithms, Infomax and FastICA, are used in the simulation. For a performance comparison, both the original algorithm and the one with feature selective filtering are tested. We use the correlation between true and estimated sources as the measure of estimation performance. The estimation results are listed in Table 2, together with the ratio of the two error tolerances used for convergence test.

For both algorithms, the incorporation of feature selective filtering improved the estimation of image source 2 and 3, both of which have high smoothness measure as shown in Table 1. Since the filtering scheme changes the convergence direction of the ICA algorithm during iterations, it is reasonable to observe that the total iteration steps and cpu time are increased for both algorithms when filtering is incorporated.

fMRI Data Analysis

In this simulation, we perform ICA on a hybrid fMRI data set consisting of three true fMRI artifacts and a spatially smoothed visual activation map associated with a smoothed time course.

To generate the hybrid fMRI data, we first perform ICA on a true fMRI data set from a visual activation experiment where the visual stimulus was

Table 2: Estimation result of synthetic images

	ICA w/o smooth filtering		ICA w/ smooth filtering	
	Infomax	FastICA	Infomax	FastICA
ϵ_1/ϵ_0	-	-	10	500
Source1	0.96±0.01	0.96±0.01	0.96±0.01	0.97±0.01
Source2	0.92±0.01	0.81±0.06	0.97±0.01	0.88±0.02
Source3	0.92±0.01	0.98±0.02	0.99±0.01	0.99±0.01
Source4	0.97±0.01	0.89±0.04	0.97±0.01	0.95±0.01
Iterations	50±2	34±6	167±3	46±23
cpu time (sec.)	5.0±0.2	1.9±1.2	15.0±1.0	3.1±1.5

applied in an on-off pattern with a period of 60 seconds for 4 minutes [6]. After the visual activation map is estimated, we perform spatial smooth filtering on the visual activation map and temporal smooth filtering on its associated time course. We substitute the corresponding activation map and the time course in the estimated data matrices with the smoothed ones. The new brain maps are then mixed by the new time courses to form the hybrid fMRI data set.

We apply Infomax and FastICA with and without feature selective filtering to this hybrid fMRI data. Since we have the prior knowledge that there is one smooth activation map included in the data set, we incorporate a moving average filter with a 5×5 pixel kernel as the feature selective filter.

Because the visual activation is task related, we measure the correlation between the estimated time course and the experimental paradigm for the evaluation of the estimation result. Similar smoothness measure as in the synthetic image simulation is used for the estimated activation map.

Table 3 shows the correlation and the smoothness measure of the estimated visual activation maps for both the FastICA and Infomax algorithms. The result without feature selective filtering is also listed for comparison. For FastICA, the incorporation of the filtering increases the estimation of the smooth visual activation map, correspondingly, the associated time course is better aligned to the experimental paradigm. For Infomax, because the original algorithm already provides a good estimation of the activation map and the time course, the improvement is not so evident. The performance difference of the two algorithms for fMRI analysis is consistent with the observations in [4]. Figure 2 shows the estimated visual activation map and time course of FastICA algorithm without and with filtering respectively. Each activation map is converted to Z-scores and thresholded at $|Z| > 2.5$. Because of Z-score conversion and thresholding, the spatial smoothness of the activation map is no longer preserved. However, filtering leads to a smaller Z-score range because of higher degree of spatial smoothness for this case.

Table 3: Estimation result of hybrid fMRI data

	ICA w/o smooth filtering		ICA w/ smooth filtering	
	Infomax	FastICA	Infomax	FastICA
ϵ_1/ϵ_0	-	-	100	5000
Correlation w/ paradigm	0.81±0.01	0.68±0.01	0.82±0.01	0.79±0.01
Smoothness	0.95±0.01	0.89±0.06	0.92±0.02	0.94±0.01
Iterations	39±1	8±1	141±4	8±3
cpu time (sec.)	2.5±0.3	1.2±0.1	5.4±1.5	1.5±0.2

DISCUSSION

In this work, we define and study a feature selective filtering scheme within a general ICA framework. We show that this kind of filtering can improve the estimation of SOIs having the same feature as that of the filter used. In the simulations, we use a 2D smoothing filter as the feature selective filter and show that the estimation result of the spatially smooth image sources is improved.

We show that the effect of feature selective filtering in the ICA estimation process is to improve the expression of SOI in the estimation through the filtering correlation matrix defined in (9). Conversely, this suggests that the characteristics of the filter should be designed according to the feature of the

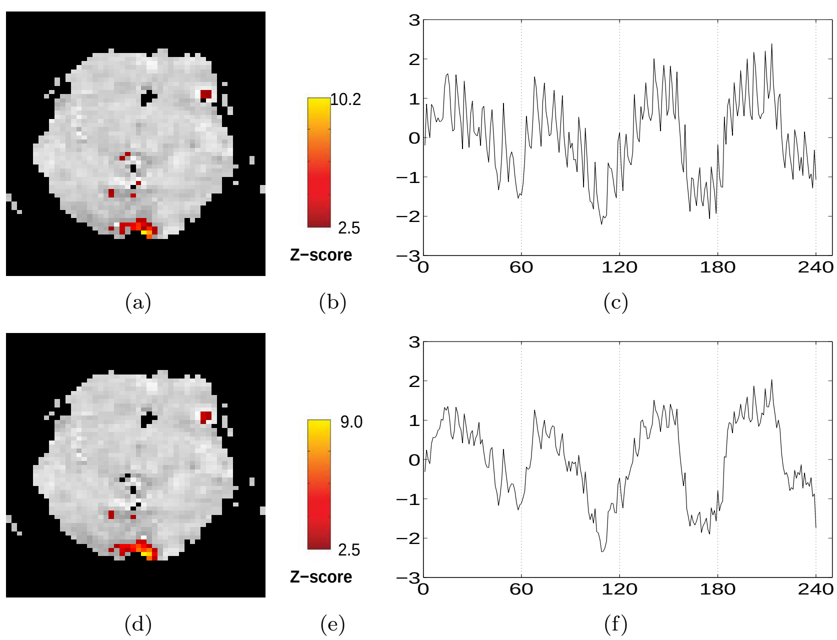

Figure 2: Estimated visual activation map, Z-score range and time course of FastICA without filtering (a, b and c) and FastICA with filtering (d, e and f)

SOI. For the same feature, the selectivity of the filter can be adjusted to influence the estimation at different levels. As in the case of fMRI data analysis, prior knowledge such as the properties of the vascular point spread function on the hemodynamic sources and the resolution of the imaging technique can be used to determine the effective kernel size of the smoothing filter. Moreover, the filtering is only applied to the estimations of SOIs by specifying a threshold on the measure of desired features and selecting estimates surviving the threshold for filtering. Therefore, the estimation of other independent components is not affected. Another approach to select the SOIs would be to rank the estimates with respect to a selected measure and to base the selection on this ranking.

Since the feature selective filtering is not related to the goal of achieving independence, it should be introduced into the estimation process after a preliminary convergence of the original ICA algorithm. The convergence behavior when filtering is incorporated is algorithm dependent. For Infomax that uses natural gradients, convergence time varies significantly depending on when the filtering is introduced. In our observation, the later the filtering is applied, the longer it takes for the learning algorithm to converge. For FastICA that performs fixed point iterations within orthogonal subspaces, the convergence rate is not observed to be sensitive to the point at which the filtering is introduced. This might be explained by the high convergence rate of the FastICA algorithm [9]. The convergence behavior of ICA algorithms with feature selective filtering is under further investigation.

The feature selective filtering introduces additional computational load into the algorithm. This includes (i) extra steps required for the learning algorithms to converge because of the perturbation due to the filtering process, and (ii) computations needed due to the feature-selective filtering operations. For the latter, suppose that there are m independent components that are being estimated where k of which are SOIs. For a single iteration, feature selective filtering requires m filtering operations and the calculation of the correlation factors defined in (10), as well as k matrix multiplications for the projection. If the mixture data are prewhitened, the matrix inversion in (3) can be neglected.

Although our motivation of feature selective filtering has been in improving the estimation of the spatially smooth sources, the proposed filtering scheme is not restricted to this specific application scenario. Filters can be designed on different sample spaces, *e.g.*, a bandpass filter can be designed for ICA of periodic signal sources or for reduction of out-of-band noise in ICA estimation for given applications. For example, in TICA of fMRI data, a bandpass filter can be designed based on the experimental paradigms to extract task related time courses. Furthermore, a multi-filter scheme can be implemented to extract SOIs with different features. In general, if the feature of an SOI can be selected by filtering in its sample space, the proposed filtering scheme will be able to improve the estimation of such SOI when properly incorporated into an iterative ICA algorithm.

REFERENCES

[1] A. Bell and T. Sejnowski, “An information-maximization approach to blind separation and blind deconvolution,” **Neural Computation**, vol. 7, no. 6, pp. 1004–1034, 1995.

[2] V. D. Calhoun, **Independent Component Analysis for Functional Magnetic Resonance Imaging**, Ph.D. thesis, University of Maryland Baltimore County, Baltimore, MD, 2002.

[3] V. D. Calhoun and T. Adali, “Semi-blind ICA of fMRI: a method for utilizing hypothesis-derived time courses in a spatial ICA analysis,” submitted to **IEEE Workshop on Machine Learning and Signal Processing (MLSP2004)**, Sao Luis, Brazil.

[4] V. D. Calhoun, T. Adali, and G. D. Pearlson, “Independent components analysis applied to fMRI data: a generative model for validating results,” to appear, **Journal of VLSI Signal Processing Systems**, 2004.

[5] V. D. Calhoun, T. Adali, L. K. Hansen, J. Larsen, and J. J. Pekar, “ICA of functional MRI data an overview,” in **Proc. Int. Symp. Indep. Comp. Analysis (ICA 2003)**, Nara, Japan.

[6] V. D. Calhoun, T. Adali, G. D. Pearlson, P. C. M. V. Zijl, and J. J. Pekar, “Independent components analysis of fMRI data in the complex domain,” **Magnetic Resonance in Medicine**, vol. 48, pp. 180–192, 2002.

[7] R. A. Choudrey, **Variational Methods for Bayesian Independent Component Analysis**, Ph.D. thesis, University of Oxford, Oxford, UK, 2002.

[8] P. A. d. F. R. Hojen-Sorensen, L. K. Hansen, and C. E. Rasmussen, “Bayesian modeling of fMRI time series,” in **Advances in Neural Information Processing Systems**, MIT Press, no. 12, 2000.

[9] A. Hyvarinen, “Fast and robust fixed-point algorithms for independent component analysis,” **Neural Networks**, vol. 10, pp. 626–634, 1999.

[10] A. Hyvarinen and R. Karthikesh, “Sparse priors on the mixing matrix in independent component analysis,” in **Proc. Int. Symp. Indep. Comp. Analysis (ICA 2000)**, Helsinki, Finland.

[11] A. Hyvarinen and E. Oja, “Independent component analysis by general nonlinear Hebbbian-like learning rules,” **Signal Processing**, vol. 3, pp. 301–313, 1998.

[12] A. Hyvarinen and E. Oja, “Independent component analysis: algorithms and applications,” **Neural Networks**, vol. 13, pp. 411–430, 2000.

[13] C. Marrelec, H. Benali, P. CiuCiu, and J.-B. Poline, “Bayesian estimation of the hemodynamic response function in functional MRI,” **Bayesian Inference and Maximum Entropy Methods in Science and Engineering, AIP Conference Proceedings**, vol. 617, pp. 229–247, 2001.

[14] M. J. McKeown, S. Makeig, G. G. Brown, T.-P. Jung, S. S. Kindermann, A. J. Bell, and T. J. Sejnowski, “Analysis of fMRI data by blind separation into independent components,” **Human Brian Mapping**, vol. 6, pp. 160–188, 1998.

[15] C. G. Phillips, S. Zeki, and H. B. Barlow, “Localization of function in the cerebral cortex. Past, present and future,” **Brain**, vol. 107, pp. 327–361, 1984.

[16] D. B. Rowe, **Multivariate Bayesian Statistics**, Chapman and Hall/CRC, 2000.