Probabilistic Framework for Brain Connectivity From Functional MR Images

Jagath C. Rajapakse*, *Senior Member, IEEE*, Yang Wang, Xuebin Zheng, and Juan Zhou

Abstract—This paper unifies our earlier work on detection of brain activation (Rajapakse and Piyaratna, 2001) and connectivity (Rajapakse and Zhou, 2007) in a probabilistic framework for analyzing effective connectivity among activated brain regions from functional magnetic resonance imaging (fMRI) data. Interactions among brain regions are expressed by a dynamic Bayesian network (DBN) while contextual dependencies within functional images are formulated by a Markov random field. The approach simultaneously considers both the detection of brain activation and the estimation of effective connectivity and does not require *a priori* model of connectivity. Experimental results show that the present approach outperforms earlier fMRI analysis techniques on synthetic functional images and robustly derives brain connectivity from real fMRI data.

Index Terms—Conditional random fields, dynamic Bayesian networks (DBNs), effective connectivity, functional magnetic resonance imaging (fMRI), graphical models, Markov random field (MRF).

I. INTRODUCTION

IN RECENT years, there has been increasing interest in using rapid developments of medical imaging in the study of human brain function. In particular, functional magnetic resonance imaging (fMRI) has become a popular technique for noninvasively locating brain functions under various cognitive and behavioral tasks. Functional brain studies acquire a time-series of brain scans while the subject is alternatively performing an experimental task and a baseline task. Brain regions of interest are then detected through measuring the oxygenation level variations in blood vessels near the neurons activated by the input stimulus, i.e., the blood oxygenation level dependent (BOLD) contrast. The detection of brain activation provides functional maps showing which brain regions are specialized for specific sensory or cognitive functions.

More recently, functional integration studies describing how functionally specialized areas interact and how these interactions lead the brain to perform a specific task have attracted more and more attention in human brain research [5], [9], [11], [13], [22], [32], [34]. In fMRI experiments, the BOLD signal (or the hemodynamic response) changes resulted from neural activities are usually close to the noise level. Hence, it is important to develop analysis methods that can robustly detect activated brain regions and derive functional connectivity from noisy fMRI time-series. Given fMRI data, the significance of activation can be assessed using various statistical models including correlation analysis [1], t-test or F-test [10], and mixture models [7], [33].

Besides local information measured at individual brain voxels, the strategy to effectively fuse contextual dependencies within functional imaging data is a key issue for the detection of brain activation [14], [30], [33]. Because of the interconnection within brain areas, the time-series of a brain voxel is highly dependent on those of its neighbors. Spatial smoothing or filtering can be used in the preprocessing of fMRI data to enhance the overall signal-to-noise ratio (SNR) in activated regions since neighboring sites are likely to belong to the same class that is activated or nonactivated. The notion of Markov random field (MRF) has also been introduced to encourage contiguous results of activation detection by defining pairwise potentials between neighboring activation (activated/inactive) labels in a generative framework [6], [28], [29]. In these MRF approaches, spatial regularization and activation detection are simultaneously handled to enhance the performance of fMRI data analysis. Recently, we introduced conditional random fields (CRF) to take into account the conditional independencies of observed fMRI data [34].

Numerous techniques have been proposed to use functional imaging data to characterize effective connectivity of the brain, i.e., the influence that one region exerts on another. Given brain activation map, the activity of an activated region is represented by the average of the time courses of hemodynamic responses of the neurons in the brain region and the fMRI experiment is represented by the dataset containing activities of all activated brain regions. The structural equation modeling (SEM) is a commonly used method that analyzes the functional connectivity among brain regions by finding the maximum likelihood (ML) parameters of connectivity [22], [25], [31]. The information about functional interactions is extracted by decomposing interregional covariances among detected region activities. The multivariate autoregressive (MAR) model characterizes effective connectivity by modeling particular brain regions as variables in a causal, dynamical, and linear system [11], [13]. Fully connected models, in which every region is connected to every other region, are usually used [5]. More recently, the dynamic causal model (DCM) has been introduced to describe the func-

*J. C. Rajapakse is with the School of Computer Engineering and the BioInformatics Research Centre, Nanyang Technological University, 50 Nanyang Avenue, 639798 Singapore, and also with the Biological Engineering Division, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139 USA (e-mail: asjagath@ntu.edu.sg).

Y. Wang, X. Zheng, and J. Zhou are with the BioInformatics Research Centre, Nanyang Technological University, 639798 Singapore.

tional interactions at the neuronal level [9], [26]. The DCM comprises of a bilinear model for neurodynamics and an extended balloon model for hemodynamics. SEM, MAR, and DCM all are confirmatory models in the sense that they usually require an initial interaction model. The prior model is often under anatomical constraints that are mainly obtained in the studies of animals. It is not always certain which areas to be included in the study, especially for brain functions unique to human, such as language and cognition. Moreover, the detection of brain activation and the analysis of effective connectivity are usually performed as two separate processes; then, the error in activation detection will further influence the accuracy of brain connectivity analysis. From this point of view, it is relatively comprehensive to simultaneously consider activation detection and effective connectivity during the analysis of fMRI data.

On the other hand, graphical models provide a natural tool for handing uncertainty and complexity through a general formalism for compact representation of joint probability distribution [17]. In particular, Bayesian networks (BN) and MRF attract more and more interests in the study of brain functional imaging [21], [34]. A technique based on BN was proposed to derive effective connectivity of the brain from fMRI data in an exploratory manner [35]. More recently, since BN does not reflect the temporal characteristics of image time-series, we proposed dynamic Bayesian networks (DBN) to explicitly take into account temporal characteristics in a stationary Markov chain [32]. We have earlier shown that DBN are better suited for modeling brain connectivity than BN [32]. This paper presents a probabilistic framework based on DBN and MRF for learning effective connectivity among activated brain regions from fMRI data. The proposed method jointly estimates brain activation and effective connectivity in a unified probabilistic framework. The interactions among brain regions are modeled by DBN and the contextual dependencies within functional imaging data are formulated by MRF. The method relaxes the constraint of requiring a prior connectivity model and effectively fuses contextual constraints within functional images. Experimental results show that the proposed method outperforms previous fMRI analysis methods with synthetic functional data and robustly derives brain connectivity from real fMRI data.

The rest of the paper is arranged as follows. Section II presents the probabilistic framework for the modeling of effective connectivity from functional imaging data. Section III describes the implementation details. Section IV discusses the experimental results. Then, our technique is concluded in Section V.

II. METHOD

In an fMRI experiment of a specific sensory or cognitive task, consider a neural system consisting of a set of n activated brain regions that are capable of collectively performing the particular task. For a pixel (or voxel) x within a 2-D (or 3-D) image scan, the observed data and activation label (i.e., activated or not) of the site are denoted by d_x and l_x respectively. Here, $x \in X$, and X is the spatial domain of the scene. The observation d_x consists of measured activity (or BOLD signal) at the site x. For a fMRI time-series from total T image scans, the

observation is expressed as $d_x = (d_{x,1}, d_{x,2}, \ldots, d_{x,T})$. The label $l_x \in \{0, 1, \ldots, n\}$ equals zero if point x is inactive, or it assigns site x to one of n activated regions (or classes) composing the neural system. The entire activation pattern and observed data over the scene are compactly expressed as $L = (l_x : x \in X)$ and $D = (d_x : x \in X)$, respectively. The activity of an activated brain region is represented by the average of the time-series responses of the region, which is denoted as $r_i = (r_{i,1}, r_{i,2}, \ldots, r_{i,T})$ for the ith region. The region activities for the entire neural system are compactly expressed by the set of activated regions as $R = \{r_1, r_2, \ldots, r_n\}$. The influence that the ith region exerts on the jth region is denoted by $c_{i,j}$, and the set for effective connectivity among all the activated regions is represented by C.

Given brain imaging data D, we hope to jointly estimate the activation labels L, region activities R, and effective connectivity C for the neural system. To simplify the computation, it is assumed that the activation pattern is independent of the effective connectivity and the effective connectivity can be derived from region activities. The joint posterior probability is estimated as

$$p(C, L, R | D) \propto p(C, L, R, D)$$
$$= p(C)p(L)p(R|C)p(D|L, R) \quad (1)$$

where $p(C)$ and $p(L)$, respectively, are the prior probabilities of effective connectivity and activation pattern, $p(R|C)$ is the conditional probability of region activities given the effective connectivity, and $p(D|L, R)$ is the data likelihood given activation labels and region activities.

The probability $p(C)$ reflects the prior information for the effective connectivity among brain regions. To prevent the trivial solution of a fully connected network, the prior probability of functional connectivity is expressed as

$$\ln p(C) \propto \sum_{i,j} \delta(c_{i,j}), \quad (2)$$

where

$$\delta(z) = \begin{cases} 1, & \text{if } z = 0 \\ 0, & \text{otherwise} \end{cases}$$

is the Kronecker delta function, and $c_{i,j}$ is the strength of effective connectivity from the ith region to the jth region. The prior probability will become small with increasing number of nonzero connections. Thus, sparsity of functional interactions among brain regions is encouraged.

The prior probability of the activation pattern $p(L)$ is formulated by a MRF to model contextual dependencies among activation labels. Here, the conditional distribution of an activation label at site x is totally determined by the labels from its neighborhood, that is, $p(l_x | l_y, y \neq x) = p(l_x | l_y, y \in N_x)$, where N_x denotes the neighborhood of the point (e.g., see Fig. 1). Using the Hammersley–Clifford theorem and considering up to pairwise potentials [2], [28], [29], the prior probability $p(l)$ is given by a Gibbs distribution with the following form:

$$p(L) \propto \exp \left\{ -\sum_{x \in X} \left[V_x(l_x) + \sum_{y \in N_x} V_{x,y}(l_x, l_y) \right] \right\}. \quad (3)$$
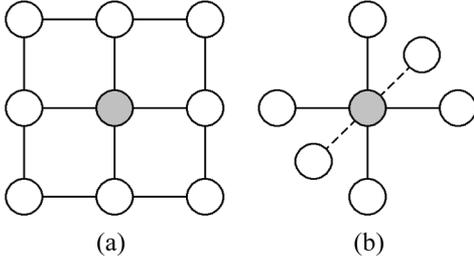
Fig. 1. (a) The 8-pixel neighborhood. (b) The 6-voxel neighborhood.

Fig. 2. Illustration of the network of DBN representing a neural system with five activated regions: the nodes represent activities of the regions; the strengths of the links characterize interactions among the regions. Two layers represent adjacent scans.

The one-pixel potential $V_x(l_x)$ reflects the prior knowledge of different label classes. The prior information of individual classes can be expressed by the following one-pixel potential,

$$V_x(l_x = i) = -\ln p(l_x = i) = \alpha_i. \qquad (4)$$

The smaller α_i is, the more likely a site x will be labeled as the ith class. Meanwhile, the two-pixel potential $V_{x,y}(l_x, l_y)$ models the contextual (or pairwise) constraint between neighboring sites. Since the regions of interest usually consist of a number of contiguous brain voxels, the connectivity constraint is imposed by the following pairwise potential to encourage contiguous activation pattern:

$$V_{x,y}(l_x, l_y) \propto 1 - \delta(l_x - l_y). \qquad (5)$$

The pairwise smoothness constraint is imposed only when the two activation labels are different. Thus, neighboring pixels are more likely to belong to the same class than to different classes.

The conditional probability $p(R|C)$ measures how well the estimated effective connectivity fits the obtained region activities of the neural system. The interactions among activated regions are represented by a BN [16], [35], in which each brain region has a set of parents that will directly impact on its activity. Using the chain rule of probability, the posterior of region activities for the neural system can be given by a compact Bayesian network representation

$$p(R|C) = \prod_i p(r_i | r_{P_i}, C_i) \qquad (6)$$

where P_i is the set of regions that directly impact on the ith region, $r_{P_i} = \{r_j | j \in P_i\}$, and $C_i = \{c_{i,j} | j \in P_i\}$. The form of brain connectivity is specified by linear regression that describes how the activity in one region is related to the activities of other regions via a set of path coefficients. In a DBN, the connectivity is considered at each time instant in a transition network modeled as a BN (see Fig. 2). With the stationary assumption, the transition network is unfolded into a DBN representing the effective connectivity among the regions. At each time instant t, the conditional probability is assumed to be normal

$$p(R|C) = \prod_i p(r_i | r_{P_i}, C_i)$$

$$= \prod_i \prod_t N \left(r_{i,t}; \sum_{j \in P_i} c_{j,i} r_{j,t}, \sigma_r^2 \right) \qquad (7)$$
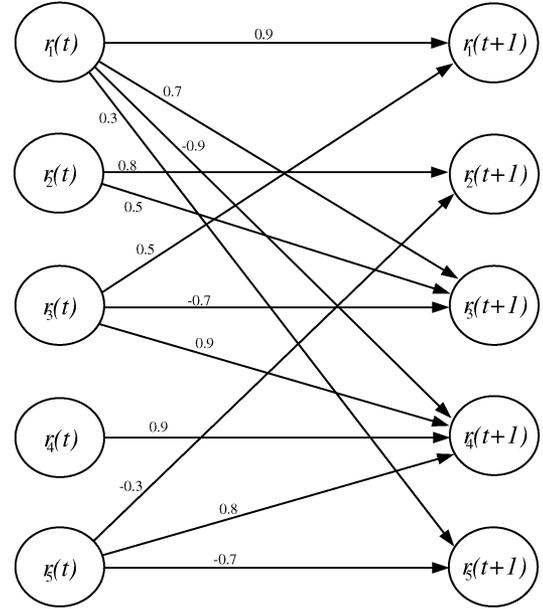
where σ_r^2 is the variance of system noises. The system noise at region i is given by $n_{i,t} = r_{i,t} - \sum_j c_{j,i} r_{j,t}$. This is a linear general model of connectivity among the regions, corrupted by Gaussian noise. The linear and Gaussian assumptions seem valid as the results in our experiments with synthetic and real fMR image data are convincing. The Gaussian noise is easily tractable in the Bayesian framework. Multinomial distribution is preferred when modeling nonlinear connectivity in the probabilistic framework [32].

Given the activation label, the local likelihood of the observation at a site is approximated by a Gaussian distribution centering at the corresponding region activity. Hence, the data likelihood is factorized as

$$p(D|L, R) = \prod_{x \in X} p(d_x | r_{l_x})$$

$$= \prod_{x \in X} \prod_t p(d_{x,t} | r_{l_x,t})$$

$$= \prod_{x \in X} \prod_t N \left(d_{x,t}; r_{l_x,t}, \sigma_d^2 \right) \qquad (8)$$

where σ_d^2 is the variance of observation noises (where the observation noise can be estimated as $v_{x,t} = d_{x,t} - r_{l_x,t}$).

Combing (1)–(8), the maximum *a posteriori* (MAP) estimate of the effective connectivity, activation labels, and region activities become

$$(\hat{C}, \hat{L}, \hat{R}) = \underset{C,L,R}{\arg\max}\, p(C, L, R|D)$$

$$= \underset{C,L,R}{\arg\max}\, p(C)p(L)p(R|C)p(D|L, R). \qquad (9)$$

III. IMPLEMENTATION

A. Optimization

To directly estimate brain connectivity, activation labels, and region activities by maximizing the joint posterior probability in (9) is ultimately difficult. In this work, the optimization is performed by iterating over the following two steps.

First, region activities are estimated given both activation pattern and brain connectivity

$$\hat{R} = \arg\max_{R} p(R|D, \hat{L}, \hat{C}). \tag{10}$$

Second, activation pattern and effective connectivity are estimated given region activities

$$(\hat{C}, \hat{L}) = \arg\max_{C,L} p(C, L|D, \hat{R}). \tag{11}$$

And, the MAP estimate becomes

$$\hat{R} = \arg\max_{R} p(R|D, \hat{L}, \hat{C})$$
$$= \arg\max_{R} p(R|\hat{C})p(D|\hat{L}, R) \tag{12a}$$

$$(\hat{C}, \hat{L}) = \arg\max_{C,L} p(C, L|D, \hat{R})$$
$$= \arg\max_{C,L} p(C)p(L)p(\hat{R}|C)p(D|L, \hat{R})$$
$$= (\arg\max_{L} p(L)p(D|L, \hat{R}), \arg\max_{C} p(C)p(\hat{R}|C)). \tag{12b}$$

Hence, the Bayesian MAP estimate can be obtained by optimizing the following objective functions:

$$\hat{R} = \arg\max_{R} \prod_{i} \prod_{t} N\left(r_{i,t}; \sum_{j \in P_i} \hat{c}_{j,i} r_{j,t}, \sigma_r^2\right)$$
$$\times \prod_{x \in X} \prod_{t} N\left(d_{x,t}; r_{\hat{l}_x,t}, \sigma_d^2\right), \tag{13a}$$

$$\hat{L} = \arg\max_{l} \exp\left\{-\sum_{x \in X}\left[\alpha_{l_x} - \sum_{y \in N_x} \beta\delta(l_x - l_y)\right]\right\}$$
$$\times \prod_{x \in X} \prod_{t} N\left(d_{x,t}; \hat{r}_{l_x,t}, \sigma_d^2\right), \tag{13b}$$

$$\hat{C} = \arg\max_{C} \gamma^{\sum_{i,j} \delta(c_{i,j})}$$
$$\times \prod_{i} \prod_{t} N\left(\hat{r}_{i,t}; \sum_{j \in P_i} c_{j,i} \hat{r}_{j,t}, \sigma_r^2\right) \tag{13c}$$

where the set $\{\alpha_i\}$ controls the sensitivity of activation detection, it is simply assumed that $\alpha_i = \alpha$ for all the activated regions and $\alpha_0 = 0$ for inactive points. Positive parameters β and γ, respectively, weigh the importance of contextual constraint in functional images and interaction sparsity for functional connectivity.

The optimization of the activation pattern is generally difficult due to the contextual interactions among neighboring sites. Here, the mean field approximation scheme is employed to get the suboptimal estimate of the activation pattern [20]. The mean field algorithm suggests that when estimating the label mean at a single site, the influence from neighboring sites can be approximated by that of their means. Moreover, the Metropolis–Hastings algorithm is employed to search the space of effective connectivity [17], with the Bayesian information criterion (BIC) as score function to find optimal connectivity model. Bayes Net Toolbox[1] is adopted for implementation.

B. Initialization and Parameters

The parameters α and β reflect the influences of prior information for brain activation and contextual constraint from neighboring sites. The higher the value of α, the more easily activated regions will be detected. Meanwhile, the higher the value of β, the stronger contextual constraints are utilized. In our experiments, it is found that $2 \leq \alpha \leq 3$ and $1 \leq \beta \leq 2$ produce the visually optimal activation pattern for the analysis of fMRI data. The parameter γ is set as $\sqrt{T}$, where T is the size of time-series. Thus, given the region activities, the effective connectivity is estimated by the BIC

$$\hat{C} = \arg\max_{C} p(C)p(\hat{R}|C)$$
$$= \arg\max_{C} \ln p(\hat{R}|C) - \frac{1}{2}s \ln T$$

where $s = n(n-1)/2 - \sum_{i,j} \delta(c_{i,j})$ is the number of nonzero connections within the neural system.

Significantly activated regions are initialized by the popular statistical parametric mapping (SPM) software[2] [10]. All the fMRI data are corrected for motion artifacts and spatial smoothing is performed with a Gaussian filter of 6-mm full-width at half-maximum (FHWM) using the SPM. In the initialization, the activity of an activated region is extracted by taking the average of the time-series responses in the region. The region activity for inactive areas is set as zero. Noise variances σ_r^2 and σ_d^2 are then estimated from the initialization. The 24-pixel neighborhood for 2-D images and 124-voxel neighborhood for three-dimensional images are utilized in this work.

IV. RESULTS

The proposed algorithm was quantitatively tested on synthetic functional imaging data. Real fMRI time-series gathered on silent reading task were also employed for qualitative evaluation.

A. Synthetic Data

Synthetic fMRI data was generated to test the feasibility and robustness of the proposed method for detecting the underlying neural system. A neural system was simulated with synthetic time-series where interactions among the brain regions are represented by a general linear model describing how activity of one region is related to the activities of other regions with a set of linear coefficients: $r_t = Cr_t + e_t$, where $r_t(r_{1,t}, r_{2,t}, \ldots, r_{n,t})^{\mathrm{T}}$ denotes the vector of region activities at time t, and e_t is the

[1]http://www.ai.mit.edu/murphyk/Software/BNT/

[2]http://www.fil.ion.ucl.ac.uk/spm/
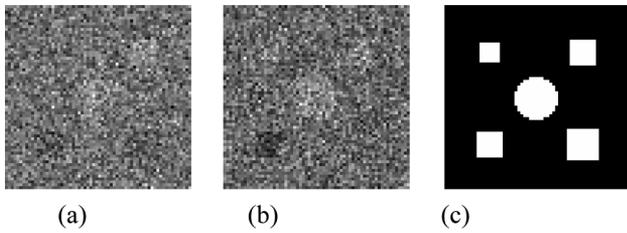
Fig. 3. Synthetic functional images with five activated regions: (a) the 30th image scan, (b) the 60th image scan, and (c) the activation pattern. Central activation represents region 1 of the connectivity model in Fig. 2, the top-left activation region 2, the bottom-left activation region 3, the bottom-right activation region 4, and the top-right activation region 5.
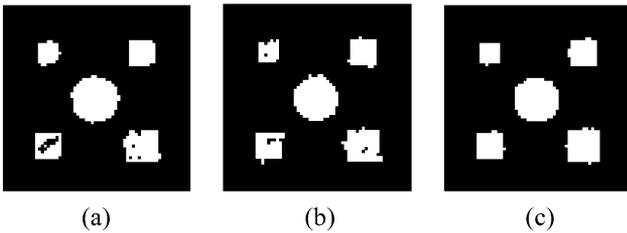
Fig. 4. Detected activation from synthetic functional image data by: (a) SPM approach, (b) MRF approach, and (c) the proposed algorithm.

vector of zero mean Gaussian innovation. The nonzero elements of the linear coefficient matrix C for a five region synthetic network are set to be $c_{1,2} = 0.8, c_{1,5} = 1.1, c_{2,4} = -0.9, c_{4,3} = 1.2, c_{5,2} = -1.0$, and $c_{5,4} = 0.7$. The region activity r_1 was taken from a normalized time-series from an activated brain voxel in a real fMRI experiment (TR $= 2$ s, in each cycle the stimulation lasted for 16 s and was followed by a resting period of 16 s).

Gaussian variants were randomly generated to simulate the rest region activities. Two-dimensional dataset of 96 scans with 64×64 pixels per image scan were generated by using the simulated region time-series. Region time-series were designed for activated pixels, while inactive pixels remained unchanged over time. Gaussian random noises were then added to the time-series of both activated and inactive pixels. Pixel intensities of image scans are shown in Fig. 3. The SNR is defined as the ratio of the standard deviation of region activities over the standard deviation of image noises.

Fig. 4 shows activation detection results by SPM, MRF approach [34], and the proposed approach for the synthetic functional data with five activated regions. The results are similar in SPM and MRF approaches. Although most activated points are detected by SPM and MRF approaches, details (or high-frequency information) of activated regions are blurred. As illustrated, the proposed approach generates relatively accurate boundaries and contiguous results by simultaneously performing activation detection and spatial regularization. By incorporating contextual interactions, the accuracy of activation detection is significantly improved by the proposed approach for functional imaging data under noisy environments.

The detection results were also evaluated quantitatively by comparing to the ground-truth image. Several synthetic datasets were generated to simulate brain systems with different number
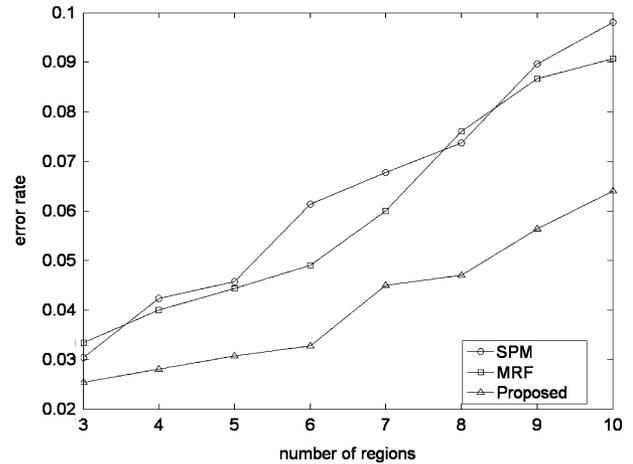
Fig. 5. Error rates of activation detection by the SPM approach, the MRF approach, and the proposed approach for synthetic functional image data with different number of regions.
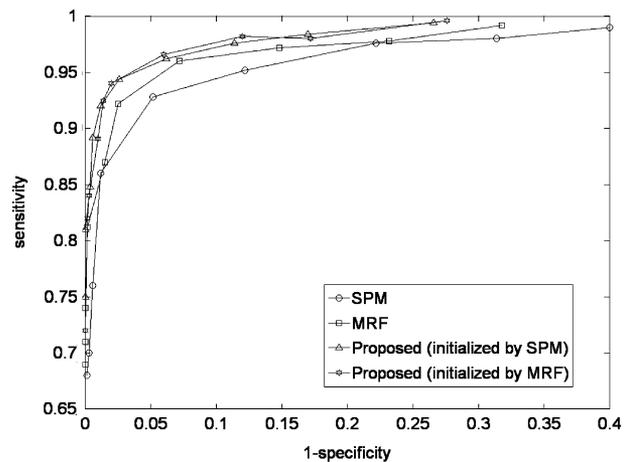
Fig. 6. ROC curves of SPM, MRF, and the proposed approach initialized with SPM and MRF.

of regions. The corresponding error rates (the portion of misclassified points over the activation regions) of the three algorithms are shown in Fig. 5. Compared to the SPM and MRF techniques, the proposed approach shows lower error rates. The corresponding receiver operating characteristic (ROC) curves for a neural systems with five regions $(\text{SNR} = 1.2)$ across three algorithms are shown in Fig. 6.

Compared to SPM and MRF approaches, the proposed approaches show better performance. Initialization with MRF and SPM gave similar results for the present approach. The substantial increase of the detection accuracy accompanies enhanced estimation of regional activities and effective connectivity.

The likelihood-ratio (LR) measure was used to assess the matching between the learned structure and the known structure of the effective connectivity. Given region activities, LR is defined as the ratio of the likelihood of the estimated structure over the likelihood of the known structure. The DBN approach [32] was used to derive neural systems of synthetic datasets and compared with the present technique. The affects of image noise, the number of scans, and the number of regions on derivation of brain connectivity were investigated.
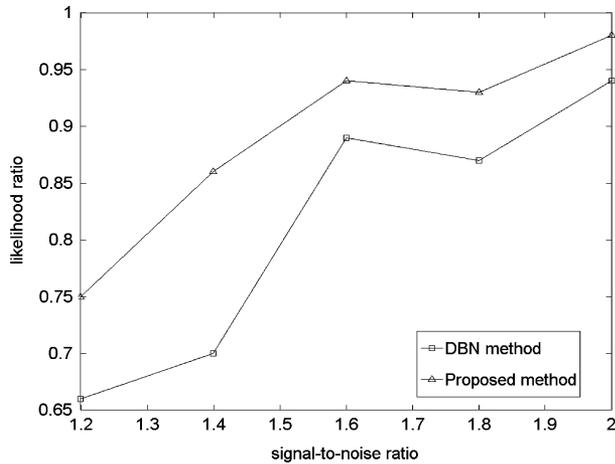
Fig. 7. The accuracy of the detection of connectivity at different noise levels of time-series from activated brain regions.
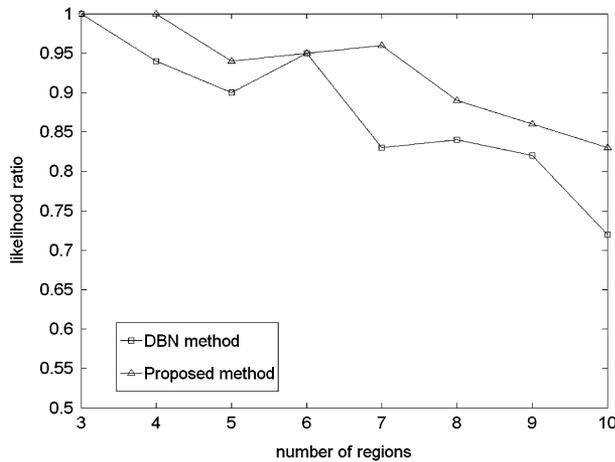
Fig. 8. Likelihood ratio of effective connectivity found by the DBN approach and the present method on synthetic functional images against different numbers of interacting regions.

As seen, the accuracy of the detection of connectivity improves with SNR while the present approach showed better performance than DBN, as shown in Fig. 7. The LR values against different numbers of brain regions at SNR value of 1.6 are plotted in Fig. 8. Both methods degrade in their performances as the number of regions grows. This could be due to the difficulty in finding the optimal connectivity pattern in (11). Fig. 9 shows the effect on the number of scans. The present method showed over 90% accuracy after 60 scans. As seen, in all above scenarios, the present method shows better performance than when activation detection and connectivity analysis are done in two separate steps.

B. fMRI Data

The fMRI data gathered on silent reading task (from the fMRI Data Center[3] access number 2–2000–11189) was used to test the proposed algorithm. In this experiment, for each subject 360 brain scans with 35 slices ($64 \times 64 \times 35$ voxels per scan) were acquired using an EPI sequence ($TR = 3.15s, TE = 40$ ms). The experimental task involved alternative reading of words and pseudo-words with variable presenting frequencies, and the

[3]http://www.fmridc.org
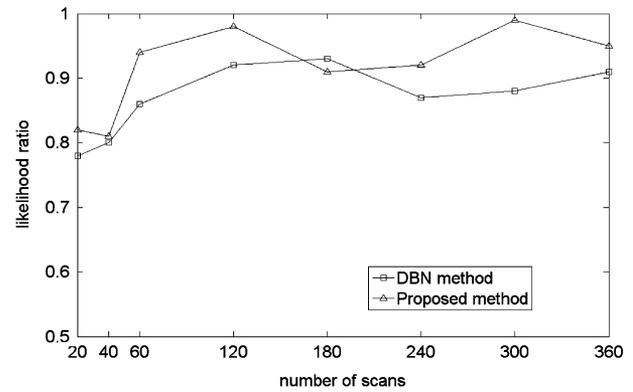
Fig. 9. Likelihood ratio of connectivity pattern by previous DBN approach and the present method on synthetic functional images having different numbers of scans.

resting condition involved fixating a cross in the middle of the screen. Each trial lasted 21 s and was followed by a resting period of 16 s. More details of the experiments can be found in [24].

Figs. 10 and 11 show the results of activation detection from fMRI data gathered on the silent reading task by the proposed approach and SPM (the initialized pattern), respectively. Expected activation in related brain regions including bilateral ventral extrastriate cortex (VEC, BA 18/19), superior parietal lobule, (SPL, BA 7), middle temporal cortex (MTC, BA 21/22), inferior frontal gyrus (IFG, BA 44/45), and middle frontal gyrus (MFG, BA 9/46) were found by both methods. The present method is improved upon the detected activation from the connectivity patterns. As seen from the figures, the present method improved upon localization and significance of activation from initially detected regions. In this experiment, the regions of activation did not change over the iterations in the connectivity analysis.

The step (10) is performed at individual level [34]. In order to perform connectivity study at the group level in (11), the probabilities of activations across subjects are averaged and the peak-activated voxels are detected. The time-course representing each significantly activated region is extracted by taking the eigen time-series of a neighborhood (radius $= 3$ mm) surrounding the peak-activated voxel. Two steps (10) and (11) are run iteratively until no change in connectivity is seen. Fig. 12 shows the neural systems derived using the proposed approach and DBN approach on the same silent reading task data. The structures learned by both methods are largely similar and consistent with the previous literature [34]. However, the proposed approach seems to discover more important connections than DBN approach by simultaneously optimizing the detection of activation and the estimation of connectivity.

The extrastriate cortex (EC: BA18, BA19) in the visual cortex plays the important role of visual representation in word processing [18]. The superior parietal lobe (SPL: BA7) plays the role of visual analysis and mainly makes efferent connections to the prefrontal cortex including middle frontal gyrus (MFG: BA46, BA9) and inferior frontal gyrus (IFG: BA44, BA45), providing more elaborate information [18]. As seen, the functional links among EC, SPL, and prefrontal cortex form the dorsal visual pathway of language processing [23]. Both approaches
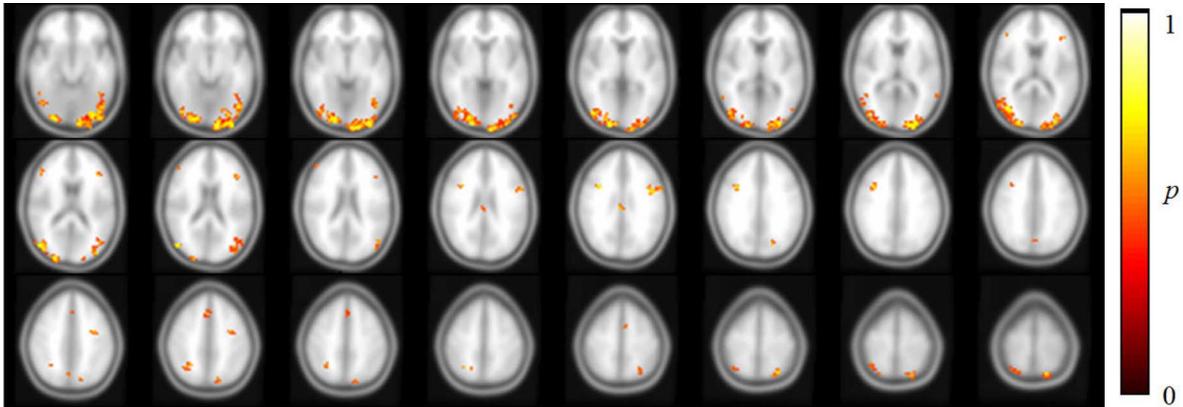
Fig. 10. Detected brain activation from fMRI data of a representative subject performing the silent reading task by using the probabilistic framework.
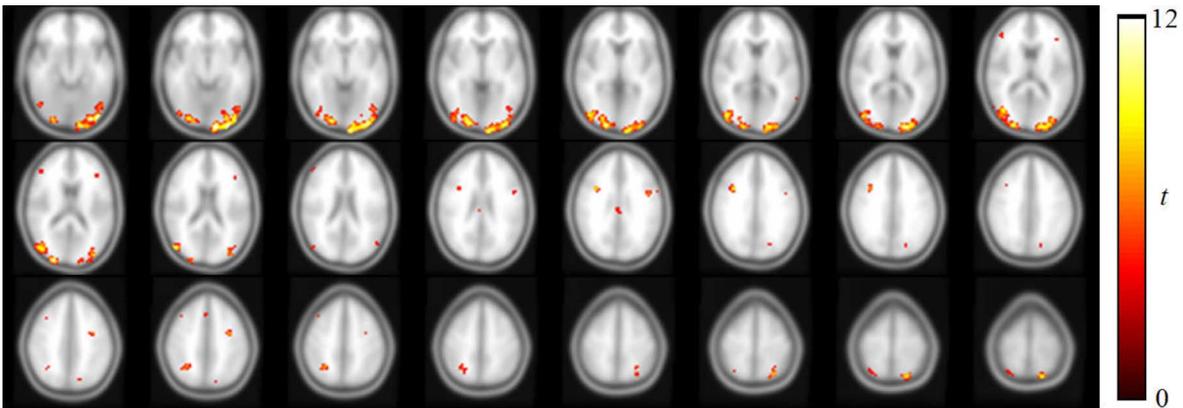
Fig. 11. Detected brain activation from fMRI data of a representative subject performing the silent reading task by using SPM.

t+1 (a)

t \	LEC	LSPL	LMTC	LIFG	LMFG	REC	RSPL	RMTC	RIFG	RMFG
LEC	1		0.4			0.7				
LSPL	0.6	1				0.6	0.5		0.6	0.6
LMTC			1		1		1	1	1	1
LIFG	1		0.3	1		0.9	0.8		1	
LMFG					1					1
REC					0.8	1	0.8			0.7
RSPL			0.6				1			
RMTC				0.4	0.4		0.6	1	0.7	0.6
RIFG		0.5				0.3			1	
RMFG										1

t+1 (b)

t \	LEC	LSPL	LMTC	LIFG	LMFG	REC	RSPL	RMTC	RIFG	RMFG
LEC	1		0.5			0.8				
LSPL	1	1		0.4			0.8	0.6	0.3	0.5
LMTC		0.7	1	1		1			0.7	1
LIFG	0.6		0.8	1		0.9			0.4	
LMFG					1				0.9	1
REC						1				
RSPL	0.5		0.6				1		0.8	0.5
RMTC				0.9		0.5		1		0.5
RIFG		0.5				0.8			1	
RMFG										1
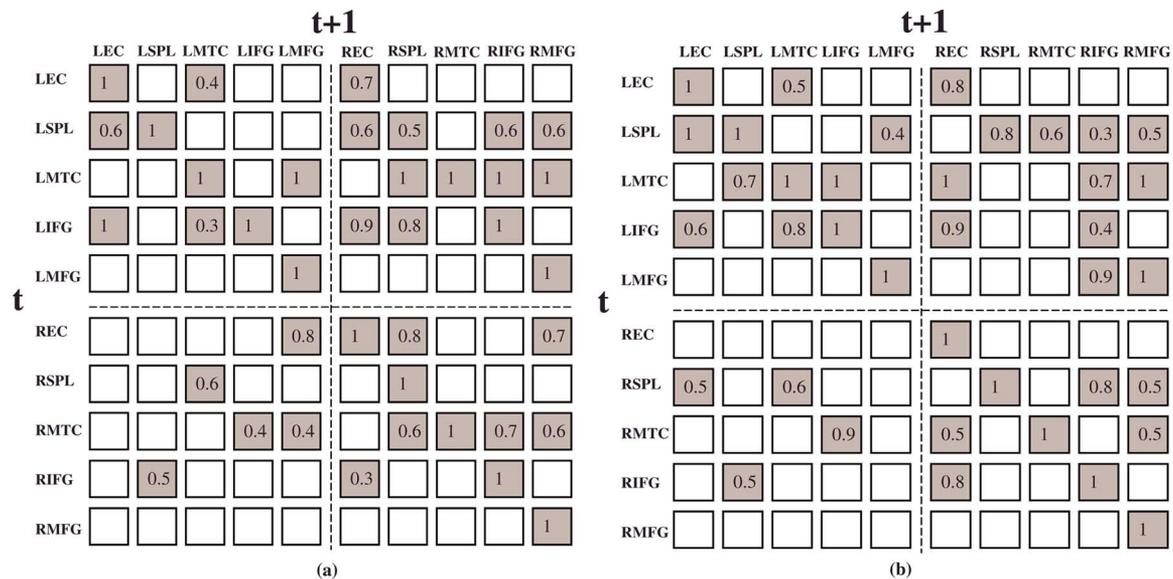
Fig. 12. Neural systems derived from fMRI data of the silent reading task using: (a) DBN method, and (b) the present method. A shaded square in the transition diagram indicates the existence of a directed connection from a region in the row to a region in the column, with the connection strength indicated.

found one dorsal pathway ($\text{LSPL} \rightarrow \text{LEC}, \text{LSPL} \rightarrow \text{RIFG}, \text{LSPL} \rightarrow \text{RMFG}$), but only the proposed iterative approach found the dorsal pathway in the opposite hemisphere $\text{RSPL} \rightarrow \text{LEC}$ and corresponding connections with prefrontal cortex ($\text{RSPL} \rightarrow \text{RIFG}, \text{RSPL} \rightarrow \text{RMFG}$). Direct connections between EC and prefrontal regions were also found ($\text{LIFG} \rightarrow \text{LEC}, \text{LIFG} \rightarrow \text{REC}, \text{RIFG} \rightarrow \text{REC}$) for semantic decision and analysis [19]. In addition, a homologous inter-hemispheric connection between the ECs of both sides ($\text{LEC} \rightarrow \text{REC}$) was found by both approaches, which may be due to the transcallosal inferences between two hemispheres [23]. A homologous interhemispheric connection was also found between the SPLs ($\text{LSPL} \rightarrow \text{RSPL}$).

The middle temporal cortex (MTC: BA21, BA22) involved in the model is the general association cortex that integrates the input from the lower level auditory and visual areas for

TABLE I
MEAN SQUARE ERRORS OF ESTIMATES OF REGIONAL ACTIVATIONS PREDICTED BY THE NETWORKS DERIVED
BY DBN AND PRESENT APPROACHES

Region	LFC	LSPL	LMTC	LIFC	LMFC	REC	RSPL	RMTC	RIFC	RMFG
DBN	1.91	2.64	2.22	6.86	2.41	2.31	1.63	1.19	7.14	0.85
Present method	1.12	2.15	2.32	2.81	2.03	0.63	1.87	1.13	3.21	0.79

retaining in the memory. The connections between EC and MTC are associated with the retaining and recalling of words from the memory [18]. The connection LEC $\rightarrow$ LMTC was found by both approaches while only the proposed iterative approach found LMTC $\rightarrow$ REC and RMTC $\rightarrow$ REC for memory retention [23], [27]. Middle temporal cortex is supposed to have connections with LSPL for movement control (LMTC $\rightarrow$ LSPL $\rightarrow$ RMTC) [12], which was correctly found by the proposed approach but not by DBN approach. The common connections between MTC and prefrontal cortex in the results of both approaches includes LMTC $\rightarrow$ RIFG, L(R)MTC $\rightarrow$ RMFG, LIFG $\rightarrow$ LMTC, and RMTC $\rightarrow$ LIFG for semantic phonologic retrieval and semantic processing [15]. Again, the proposed approach was able to discover a new connection LMTC $\rightarrow$ LIFG.

The MFG is involved in tasks that require executive control, such as the selection of behavior based on short term memory [19]. It receives inputs from the posterior parietal and superior temporal sulci. The IFG is most active for phonemic decisions and receives inputs from temporal lobes and parietal lobes [3], [27]. As seen, except for the connections that have been mentioned above, there are interhemispheric connections between the prefrontal regions (LIFG $\rightarrow$ RIFG, LMFG $\rightarrow$ RMFG), which may be involved in semantic processing during inner speech. The connection (LMFG $\rightarrow$ RIFG) was found by the proposed approach but not by DBN approach. It has been earlier reported in an experiment demanding semantic categorization and subvocal rehearsal [4].

We used the goodness of fix index (GFI) [4] to measure the overall model fit of the dataset, i.e., how the derived structure matches with the data. A better fit was realized by the present method (GFI $= 0.92$) than the DBN (GFI $= 0.84$). Using the linear model of connectivity, the activation (the average time-series response) of one region was predicted from the activation of the other activated regions, and thereby mean square errors of estimates of regional activations were measured and given in Table I. Reduced errors of estimates indicate the present method better fit the data.

V. CONCLUSION

We proposed a probabilistic framework of deriving effective connectivity among brain regions from fMRI data. Unlike earlier approaches, both brain activation detection and effective connectivity analysis are simultaneously considered in a unified framework. We take a probabilistic approach based on graphical models including dynamic Bayesian networks and Markov random fields. Furthermore, the proposed method does not require a prior connectivity model to begin with. As seen in the experiments, the present approach gives more refined activity and connectivity patterns by jointly estimating both brain activation and effective connectivity to enhance the accuracy of fMRI analysis.

Experimental results show that the proposed approach outperforms earlier methods with synthetic functional images and robustly learns brain connectivity from real fMRI data. The connectivity pattern derived on silent reading data was close to the networks earlier derived by DBN, and better explained the reading task. However, the present method demands more computational resources and time for iterative improvement. In our experiments, the algorithm converged in less than five iterations because initial connectivity based on SPM activation seemed close to the optimal pattern.

We presumed linearly connected networks; nonlinear models of connectivity increases the complexity but might shed more insight into brain connectivity. We assumed a stationary model where the connectivity pattern remains unchanged during fMRI experimentation. However, learning and habituation effects are likely to exist during a typical fMRI experiment. How this research is extended to handle such scenarios would be a future direction of this research. Also, theoretical work into the effects of the number of scans, the number of regions, and scanner and hemodynamic noise on derivation of connectivity is worth pursuing.

REFERENCES

[1] P. A. Bandettini, A. Jesmanowicz, E. C. Wong, and J. S. Hyde, "Processing strategies for time-course data sets in functional MRI of the human brain," *Magn. Reson. Med.*, vol. 30, pp. 161–173, 1993.

[2] J. Besag, "On the statistical analysis of dirty pictures," *J. Roy. Statist. Soc. B*, vol. 48, pp. 259–302, 1986.

[3] Brain systems, functions and problems [Online]. Available: http://www.brainplace.com/bp/brainsystem/

[4] E. Bullmore, B. Horwitz, G. Honey, M. Brammer, S. Williams, and T. Sharma, "How good is good enough in path analysis of fmri data?," *NeuroImage*, vol. 11, no. 4, pp. 289–301, Apr. 2000.

[5] E. R. Cosman and W. M. Wells, "Bayesian population modeling of effective connectivity," in *Proc. Inf. Process. Med. Imag. Conf.*, 2005, pp. 39–51.

[6] X. Descombes, F. Kruggel, and D. Y. von Cramon, "Spatio-temporal fmri analysis using markov random fields," *IEEE Trans. Med. Imag.*, vol. 17, no. 6, pp. 1028–1039, Dec. 1998.

[7] B. S. Everitt and E. T. Bullmore, "Mixture model mapping of brain activation in functional magnetic resonance images," *Hum. Brain Mapp.*, vol. 7, pp. 1–14, 1999.

[8] C. J. Fiebach, A. D. Friederici, K. Muller, and D. Y. V. Cramon, "FMRI evidence for dual routes to the mental lexicon in visual word recognition," *J. Cogn. Neurosci.*, vol. 14, no. 1, pp. 11–23, 2000.

[9] K. J. Friston, "Dynamic causal modelling," *NeuroImage*, vol. 19, no. 4, pp. 1273–1302, Aug. 2003.

[10] K. J. Friston, A. P. Holmes, J.-B. Poline, P. J. Grasby, S. C. R. Williams, R. S. J. Frackowiak, and R. Turner, "Analysis of fmri time-series revisited," *NeuroImage*, vol. 2, no. 1, pp. 45–53, Mar. 1995.

[11] R. Goebel, A. Roebroeck, D.-S. Kim, and E. Formisano, "Investigating directed cortical interactions in time-resolved fmri data using vector autoregressive modeling and granger causality mapping," *Magn. Reson. Imag.*, vol. 21, pp. 1251–1261, 2003.

[12] M. Hampson, B. S. Peterson, P. Skudlarski, J. C. Gatenby, and J. C. Gore, "Detection of functional connectivity using temporal correlations in MR images," *Hum. Brain Mapp.*, vol. 15, pp. 247–262, 2002.

[13] L. Harrison, W. D. Penny, and K. Friston, "Variational bayesian inference for fmri time series," *NeuroImage*, vol. 19, no. 3, pp. 727–741, Jul. 2003.

[14] N. V. Hartvig and J. L. Jensen, "Spatial mixture modeling of fmri data," *Hum. Brain Mapp.*, vol. 11, pp. 233–248, 2000.

[15] B. Horwitz and A. R. Braun, "Brain network interactions in auditory, visual and linguistic processing," *Brain Language*, vol. 89, no. 2, pp. 377–384, 2004.

[16] F. Jenson, *Bayesian Networks and Decision Graphs*. New York: Springer-Verlag, 2001.

[17] M. I. Jordan, *Learning in Graphical Models*. Cambridge, MA: MIT Press, 1999.

[18] B. Kolb and I. Q. Whishaw, "Fundamantal of Human," in *Neuropsychology*. San Francisco, CA: W. H. Freman, 1996.

[19] B. J. Krause, B. Horwitz, J. G. Taylor, D. Schmidt, F. M. Mottaghy, H. Herzog, U. Halsband, and H. W. Muller-Gartner, "Network analysis in episodia encoding and retrieval of word-pair associates: A PET study," *J. Neurosci.*, vol. 11, pp. 3293–3301, 1999.

[20] S. Z. Li, *Markov Random Field Modeling in Image Analysis*. Berlin, Germany: Springer-Verlag, 2001.

[21] G. Marrelec, P. Ciuciu, M. Pelegrini-Issac, and H. Benali, "Estimation of the hemodynamic response function in event-related functional MRI: Directed acyclic graphs for a general bayesian framework," in *Proc. Inf. Process. Med. Imag. Conf.*, 2003, pp. 635–646.

[22] A. R. McIntosh and F. Gonzalez-Lima, "Structural equation modeling and its application to network analysis in functional brain imaging," *Hum. Brain Mapp.*, vol. 2, pp. 2–22, 1994.

[23] B. R. McIntosh, C. L. Grady, L. G. Ungerieider, J. V. Haxby, S. I. Rapoport, and B. Horwitz, "Network analysis of cortical visual pathways mapped with PET," *J. Neurosci.*, vol. 14, no. 2, pp. 655–666, 1994.

[24] K. J. Mechelli Friston and C. J. Price, "The effects of presentation rate during word and pseudoword reading: A comparison of PET and fmri," *J. Cogn. Neurosci.*, vol. 12, no. , pp. 145–156, 2000.

[25] W. D. Penny, K. E. Stephan, A. Mechelli, and K. J. Friston, "Modelling functional integration: A comparison of structural equation and dynamic causal models," *Neuroimage*, vol. 23, pp. s264–s274, 2004.

[26] W. D. Penny, K. E. Stephan, A. Mechelli, and K. J. Friston, "Comparing dynamic causal models," *NeuroImage*, vol. 22, pp. 1157–1172, 2004.

[27] J. Price, "The anatomy of language: Contributions from functional neuroimaging," *J. Anatomy*, vol. 179, pp. 335–359, 2000.

[28] J. C. Rajapakse, J. N. Giedd, and J. L. Rapoport, "Statistical approach to segmentation of single-channel cerebral MR images," *IEEE Trans. Med. Imag.*, vol. 16, no. 2, pp. 176–186, Apr. 1997.

[29] J. C. Rajapakse and F. Kruggel, "Segmentation of MR images with intensity inhomogeneities," *Image Vision Comput.*, vol. 16, no. 3, pp. 165–180, 1998.

[30] J. C. Rajapakse and J. Piyaratna, "Bayesian approach to segmentation of statistical parametric maps," *IEEE Trans. Biomed. Eng.*, vol. 48, no. 10, pp. 1186–1194, Oct. 2001.

[31] J. C. Rajapakse, C. L. Tan, X. Zheng, S. Mukhopadyay, and K. Yang, "Exploratory analysis of brain connectivity with ICA," *IEEE Eng. Med. Biol. Mag.*, vol. 25, no. 2, pp. 102–111, Mar./Apr. 2006.

[32] J. C. Rajapakse and J. Zhou, "Learning effective brain connectivity with dynamic Bayesian networks," *NeuroImage*, vol. 37, pp. 749–760, 2007.

[33] E. Salli, H. J. Aronen, S. Savolainen, A. Korvenoja, and A. Visa, "Contextual clustering for analysis of functional MRI data," *IEEE Trans. Med. Imag.*, vol. 20, no. 5, pp. 403–414, May 2001.

[34] Y. Wang and J. C. Rajapakse, "Contextual modeling of functional MR images with conditional random fields," *IEEE Trans. Med. Imag.*, vol. 25, no. 6, pp. 804–812, Jun. 2006.

[35] X. Zheng and J. C. Rajapakse, "Learning functional structure from fmr images," *NeuroImage*, vol. 31, pp. 1601–1613, 2006.