

## Research Article

# Activation Detection on fMRI Time Series Using Hidden Markov Model

Rong Duan<sup>1</sup> and Hong Man<sup>2</sup>

<sup>1</sup>AT&T Labs, Florham Park, NJ 07932, USA

<sup>2</sup>Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ 07030, USA

Correspondence should be addressed to Rong Duan, rongduan@research.att.com

Received 16 March 2012; Revised 23 June 2012; Accepted 23 June 2012

Academic Editor: Anke Meyer-Baese

Copyright © 2012 R. Duan and H. Man. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper introduces two unsupervised learning methods for analyzing functional magnetic resonance imaging (fMRI) data based on hidden Markov model (HMM). HMM approach is focused on capturing the first-order statistical evolution among the samples of a voxel time series, and it can provide a complimentary perspective of the BOLD signals. Two-state HMM is created for each voxel, and the model parameters are estimated from the voxel time series and the stimulus paradigm. Two different activation detection methods are presented in this paper. The first method is based on the likelihood and likelihood-ratio test, in which an additional Gaussian model is used to enhance the contrast of the HMM likelihood map. The second method is based on certain distance measures between the two state distributions, in which the most likely HMM state sequence is estimated through the Viterbi algorithm. The distance between the on-state and off-state distributions is measured either through a *t*-test, or using the Kullback-Leibler distance (KLD). Experimental results on both normal subject and brain tumor subject are presented. HMM approach appears to be more robust in detecting the supplemental active voxels comparing with SPM, especially for brain tumor subject.

## 1. Introduction

Functional magnetic resonance imaging (fMRI) is a well-established technique to monitor brain activities in the field of cognitive neuroscience. The temporal behavior of each fMRI voxel reflects the variations in the concentration of oxyhemoglobin and deoxyhemoglobin, measured through blood oxygen level-dependent (BOLD) contrast. BOLD signal is generally considered as an indirect indicator for brain activities, because neural activations may increase blood flow in certain regions of the brain.

*1.1. Characteristics of fMRI Data.* fMRI data are collected as a time series of 3D images. Each point in the 3D image volume is called a voxel. fMRI data have four important characteristics: (1) large data volume; (2) relatively low SNR; (3) hemodynamic delay and dispersion; (4) fractal properties. Typically, one fMRI data set includes over 100-K voxels from a whole brain scan and therefore has 100-K time series. The observed time sequences are combinations

of different types of signals, such as task-related, function-related, and transiently task-related (different kinds of transiently task-related signals coming from different regions of brain). These are the signals that convey brain activation information. There are also many types of noises, which can be physiology-related, motion-related, and scanning-related. The signal to noise ratio (SNR) in typical fMRI time series can be quite low, for example, around 0.2 to 0.5. For different regions and different trails, the SNR level also varies significantly. Such noise nature causes major difficulty in signal analysis. Hemodynamic delay and dispersion further increase the complexity of fMRI signal structure. Special efforts have been made to construct flexible hemodynamic response function (HRF) which can model the hemodynamic delay and dispersion through various regions and different subjects [1, 2]. fMRI data also have fractal properties, which means that a class of objects may have certain interesting properties in common. In other words, fMRI data are approximately scale invariant or scale free.

*1.2. Methodology of Analyzing fMRI Data.* Two areas of fMRI-based neural systems study have attracted lots of attention over the past two decades: functional activity detection and functional connectivity detection. Functional activity detection aims to locate the spatial areas that are associated with certain psychological tasks, commonly specified by a predefined paradigm. Functional connectivity detection focuses on finding spatially separated areas that have high temporal correlations [3]. Generally, functional connectivity detection is conducted under the resting-state condition. The differences in biophysical motivations and experiment designs of these two studies are reflected in their methodologies. Functional activity detection compares the temporal series of each voxel with the excitation paradigm and functional connectivity detection compares the voxel timeseries with other series in a predefined spatial region, that is, Region of Interesting (ROI), or “seed” region. Functional activity detection is more on the temporal correlation and functional connectivity detection is more on spatial correlation. Even though the two areas have some differences, they share many common statistical modelling strategies. Both of them attempt to build models to abstract the spatial and temporal relations from the observed fMRI data. To choose a model that can capture the properties of the time series accurately and efficiently is essential in both study areas.

A large number of methods have been proposed to analyze fMRI data. Most of them can be characterized into one of these two categories: modelling based approach and data driven approach. Reference [4] provided a detailed review. Even though the authors claimed the methodologies as functional connectivity detection, certain amount of the reviewed methods were commonly used in functional activity detection study too. This paper focuses on constructing a model to extract voxel temporal characteristics and test the voxel activities using functional activity detection as example. All the models referred following are for this task if no further clarification.

An established software called statistical parametric map (SPM) is a typical functional activity detection modelling package based on general linear model (GLM) [5], general linear model transforms a voxel time series into a space spanned by a set of basis vectors defined in the design matrix. These basis vectors include a set of paradigm waveforms convolved with hemodynamic response function (HRF), as well as several low-frequency DCT bases. The residual errors of this linear transform is modelled as Gaussian pdf. The key component of this method is how to constitute the design-matrix which can accurately model the brain activation effects and separate noises. Using GLM to analyze fMRI data has following intrinsic assumptions:

- (i) the activation patterns are spatially distributed in the same way for all subjects,
- (ii) the response between input stimulation and brain response is linear,
- (iii) the HRF function is the same for every voxel,

- (iv) the time series observation has a known Gaussian distribution,
- (v) the variance and covariance between repeated measurements are invariant,
- (vi) the time courses of different factors affecting the variance of fMRI signals can be reliably estimated in advance,
- (vii) the signals at different voxels are independent, and
- (viii) the intensity distribution of background (nonactive areas) is known whereas the distribution of active areas is not known.

Reference [6] substituted paradigm with the average temporal series of ROI as seed and applied SPM on functional connectivity detection. And the most recent release of SPM incorporates dynamic causal modelling to infer the interregional coupling, but it is designed more on EEG and MEG data [7].

There are some other methods which can be considered as special cases of GLM. For example, direct subtraction method subtracts the average of “off” period from the average of “on” period. Voxels with significant difference will then be identified as active. Students’  $t$ -test can be used to measure the difference in the means by the standard deviations in “off” and “on” periods. The larger the  $t$ -value is, the larger the “on-off” difference is, and the more active the voxel is. Correlation coefficient is another special case of GLM. It measures the correlation coefficient between a reference function waveform and each voxel temporal signal waveform. Voxels with large correlation coefficient are considered to be connected. If the reference function waveform is defined by paradigm [8], it is functional activity detection, and if the reference function waveform is defined by some seed time course, it is functional connectivity detection [9].

There exist some problems in GLM-based methods. For example, GLM assumes one HRF function for all voxels. The BOLD signal is only an indirect indicator of neural activity, and many nonneural changes in the body could also influence the BOLD signal. Different brain areas may have different hemodynamic responses, which would not be accurately reflected by the general linear model. Also, GLM method typically requires grouping or averaging data over several task/control blocks, which reduces sensitivity for detecting transient task-related changes, and make it insensitive to significant changes not consistently time-synched to the task block design. Low SNR makes it possible for nontask-relevant components overshadow task-relevant components and further reduces the sensitivity and specificity. GLM only considers the time series and ignores relationships between voxels, hindering the detection of brain regions acting as functional units during the experiment. Another problem is that the GLM method does not extract the intrinsic structure of the data, which may significantly weaken its effectiveness when the a priori of fMRI signal in response to the experimental events is not known or may not be constant across all voxels.

Besides GLM-based methods, a few methods have been proposed to improve the accuracy of modelling fMRI temporal signals. Reference [10] introduced a Bayesian modelling method which used a two-state HMM to infer an optimal state sequence through Markov Chain Monte Carlo sampling. The method assumed the observation is a linear combination of a two-state HMM, which infer hidden psychological states, plus constant and trend. The model is designed as the combination of offset, linear trend and a set of two-state HMMs state sequences start at different time points. MCMC is used to estimate the optimal state sequence for each voxel. This method is good in interpreting the dynamics in each voxel, but the computing complexity is big concern as mentioned by the authors, and also the totally paradigm-free approach might introduce noise that irrelevant with the experiment design. Reference [11] applied state-space model and Kalman filter to model the baseline and stimulus effect without any parametric constrain. Reference [12] employed multiple reference functions with 100 ms shift to find the highest correlation coefficient reference function for specific voxels, which avoids the common practice of using a single-reference function for all voxels. Reference [13] proposed Gaussian mixture models to describe the mutually exclusive fMRI time sequence. Reference [14] introduced an unsupervised learning method based on hidden semi-Markov event sequence models (HSMESMs) method which had the advantage of explicitly modelling the state occupancy duration. The method decomposed an observation into true positive events, false positive events, and missing observations. The “off-on” paradigm transitions were modelled as left to right HMM true positive states, and the other periods were considered as semi-Markov false positive states. The likelihood of HSMESM was calculated iteratively to detect activity base on predefined threshold. Reference [15] used first-order Markov chain to estimate the time series,  $t$ -test, and mutual information to detect the actions. All these methods are essentially two-stage approach. The temporal property is modelled at each voxel independently and then spatial modelling is performed based on the summarized statistics from temporal analysis. Fully Bayesian spatiotemporal modelling [16] considered spatial and temporal information together. This method decomposed the observation data into spatiotemporal signal and noise, and space-time simultaneously specified autoregressive model (STSAR) was employed to construct noise model. Half-cosine HRF model and activation height model were used to construct fMRI signal models.

Granger causality analysis [17] and dynamic causal modelling [18] are two popular methods in functional connectivity detection in recent years. A series comments and controversies [19–22] have been dedicated to comparing these two methods on model selection, causality, and deconvolution from biophysiological view. Reference [23] criticized dynamic causal modelling from computation complexity and model validation from mathematical perspective. The advantage of these two methods are that they consider the spatial and temporal information at the same time, but the disadvantage is the model complexity.

All the modelling-based approaches mentioned above are either too simple to capture the temporal or spatial dynamics for different voxels and subjects, or too complicated to estimate parameters accurately and inference easily.

In addition to these modelling-based methods, there are also data driven methods in analyzing fMRI data. One popular example of this approach is independent component analysis (ICA). ICA decomposes a 4D fMRI data volume (3D spatial and 1D temporal) into a set of maximum temporal or spatial independent components by minimizing the mutual information between these components. ICA does not require the knowledge of stimulus or paradigm in data decomposition, and similar voxel activation patterns will usually appear in the same component. ICA is also called blind source separation, because it does not need prior knowledge, and it is able to identify “transient task related” components that could not be easily identified by the paradigm. The first application of ICA in fMRI data was the spatial ICA (sICA) [24]. Temporal ICA (tICA) was introduced later by [25]. Reference [26] compared the sICA with tICA and reported that the beneficial of each method depends on the independence of the underlying spatial or temporal signal. sICA maximizes independence spatially and the corresponding temporal information might be highly correlated, and vice versa for tICA. To consider the mutual independence between space and time simultaneously, [27] proposed a spatiotemporal independent component analysis (stICA). Extended from entropy-based one-dimension ICA decomposition introduced in [28], the authors embedded the spatial and temporal components at the same time and also incorporated the spatial skewed probability density function to replace the kurtosis and symmetric probability density function in decomposing the independent signals. As pointed out in [29], the disadvantages of the Infomax and entropy-based stICA algorithms used in [27] are that the number of parameters needed to be estimated is large, the local minima and sensitive to noise characteristics of the gradient descent optimization methods. To improve the stability, robustness, and simplify the computing complexity, [29] adopted the generalized eigenvalue decomposition and joint diagonalization on both spatial and temporal autocorrelation to achieve spatial and temporal independent signals simultaneously. ICA methods have showed promising results in fMRI analysis, but similar as all other data-driven methods, it is hard to interpret the output and it usually requires special knowledge and human intervention. Also, ICA does not specify which component, among many output components, is the activation component, and there is no statistical confidence level of each components extracted.

Clustering is another well-developed data driven approach in brain activity and connectivity detection. It has been used to identify regions with similar patterns of activations. Common clustering algorithms include hierarchical clustering, crisp clustering, K-means, self-organizing maps (SOM), and fussy clustering. The major drawback of most clustering methods is that they make assumptions about cluster shapes and sizes, which may deviate in observed data structures. The optimization techniques used in clustering may also result in local maxima

and instable results. In addition, the number of clusters is frequently determined heuristically and randomly initialized, which makes the output inconsistent with each trial.

In this paper, we propose a simple dynamic state space model, which attempts to model the voxel time series as a random process driven by the experimental paradigm or some ROI area seed series. For a given voxel, its behavior is described by a two-state hidden Markov model with certain state distributions and state transitions. The HMM parameters are estimated from the prior statistics of the paradigm as well as from the testing time series. Two methods are introduced to detect the voxel activation based on the estimated HMM. The first method calculates the likelihood of each time series, given its HMM, and forms a likelihood map for all the voxels reside in a fMRI slice. A simple Gaussian model is also used to improve the contrast of this likelihood map. The second method uses the  $t$ -test or the Kullback-Leibler distance (KLD) to measure the distance between the on-state distribution and the off-state distribution. These distributions are estimated based on the most likely HMM state sequence, which is calculated through a Viterbi algorithm. The contribution of the method is that it unifies the robustness, stability, and reliability under the same framework in estimating paradigm driven fMRI study. First, it incorporates the dynamic characteristics of fMRI time series by adapting 2-state HMM model, which is robust in detecting active voxels with different delay and dispersion behaviors. Second, the proposed method utilizes paradigm prior knowledge in parameter estimation, which is not only to simplify the computing compared with the approach in [10], but also to improve the stability and reliability of the output due to the stability of paradigm.

The rest of this paper is organized as follows. In Section 2, we introduce the two-state hidden Markov model approach for fMRI data. In Section 3, we discuss activation detection methods base on the estimated HMM. In Section 4, we present the experimental results on two sets of fMRI data, one is normal subject and the other is brain tumor subject, and compare the results with GLM-based statistical parametric mapping package (SPM) [5].

## 2. Hidden Markov Model for fMRI Time Series

*2.1. Hidden Markov Model.* HMM is a very efficient stochastic method in modelling sequential data of which the distribution patterns tend to cluster and alternate among different clusters [30]. A hidden Markov model consists of a finite set of states. In a traditional Markov chain, the state is directly visible to the observer, and the state transition probabilities are the only parameters. In an HMM, only the observations influenced by the state are visible. Each of the hidden state is associated with a probability distribution. Transitions among the states are measured by transition probabilities. The most common first-order HMM implies that the state at any given time depends only on the state at the previous time step.

HMM is well developed in temporal pattern recognition applications. It was first applied to speech recognition [31].

Now it is widely used in multimedia [32], bioinformatics [33, 34], informational retrieval [35], and so forth.

An HMM can be described by the following elements [31]: (1) a set of observations  $O\{T\}$ , where  $T$  is the number of time samples; (2) a set of states  $Q\{N\}$ , where  $N$  is the number of states; (3) a state-transition probability distribution  $A = \{a_{ij}\}$ , where  $a_{ij} = P[q_{t+1} = S_j | q_t = S_i]$ ,  $1 \leq i, j \leq N$ ; (4) observation probability distribution for each state  $B = \{b_j(x)\}$ , where  $b_j(x) = P[o_t = x | q_t = S_j]$ ,  $1 \leq j \leq N$ ,  $x \in \mathfrak{X}$  is a possible observation value; (5) an initial state distribution  $\pi = \{\pi_j\}$ , where  $\pi_j = P[q_1 = S_j]$ ,  $1 \leq j \leq N$ .

An HMM is therefore denoted by  $\lambda = \{A, B, \pi\}$ . We further model each state distribution as a Gaussian pdf:

$$P[O_t = x | q_t = S_j] = \frac{1}{\sigma_j \sqrt{2\pi}} e^{-(x-\mu_j)^2/(2\sigma_j^2)}, \quad j \in \{0, 1\}. \quad (1)$$

Let  $Q = q_1, q_2, \dots, q_T$  be a possible state sequence and assume that the observation samples are independent, the likelihood of an observed sequence given this HMM can be calculated as:

$$\begin{aligned} P(O | \lambda) &= \sum_Q P(O | Q, \lambda) P(Q | \lambda) \\ &= \sum_Q \pi_{q_1} b_{q_1}(o_1) a_{q_1 q_2} b_{q_2}(o_2) \cdots a_{q_{T-1} q_T} b_{q_T}(o_T). \end{aligned} \quad (2)$$

Given the observation  $O$  and the HMM, the most likely state sequence  $Q = \{q_1, q_2, \dots, q_T\}$ , which maximizes the likelihood  $P(Q | O, \lambda)$ , can be calculated through the Viterbi algorithm [36]. The Viterbi path score function is defined as:

$$\delta_t(i) = \max_{q_1, q_2, \dots, q_{t-1}} P[q_1 q_2 \cdots q_t = i, o_1 o_2 \cdots o_t | \lambda], \quad (3)$$

where  $\delta_t(i)$  is the highest probable path ending in state  $i$  at time  $t$ . The induction can be expressed as:

$$\delta_{t+1}(j) = \max_{1 \leq i \leq N} [\delta_t(i) a_{ij}] b_j(o_{t+1}). \quad (4)$$

In an application of HMM, multiple HMMs are trained by different groups of labeled data. The HMM parameters are estimated based on these training data. The test data will be assigned to the one which has the maximum likelihood.

### 2.2. Brain Activation Detection

*2.2.1. HMM Likelihood Methods.* In our unsupervised learning methods, HMM parameters are estimated directly from the experimental paradigm or the voxel time series under examination. This is different from conventional HMM applications where HMM parameters are usually estimated from some training data. The attempt of avoiding training process is motivated by the fact that the true activation behavior varies from voxel to voxel and from patient to patient. Therefore, it is not advisable to use the parameters from certain set of voxels to characterize other voxels.

Since the simple block paradigm has only two levels, “on, off,” in this work we let the number of state  $N = 2$ , that is, on-state  $S_1$  and off-state  $S_0$ .

Because of the first-order Markov assumption, that is,  $P(q_t = j \mid q_{t-1} = i, q_{t-2} = k, \dots) = P(q_t = j \mid q_{t-1} = i)$ , the distribution of a state duration is exponential, and the expected value of a state duration can be expressed as:

$$\bar{d}_i = \frac{1}{1 - a_{ii}}. \quad (5)$$

Given an experimental paradigm, let the length (i.e., time samples) of the ON period be  $L_{\text{on}}$ , and the length of off period be  $L_{\text{off}}$ , the transition matrix  $A$  can be estimated as  $a_{00} = 1/(1 - L_{\text{off}})$ ,  $a_{01} = 1 - a_{00}$ ,  $a_{11} = 1/(1 - L_{\text{on}})$ ,  $a_{10} = 1 - a_{11}$ .

The parameters in  $B$  can be estimated from the voxel time series  $O$ . Assuming that the time samples are normalized, let  $p_{\text{on}}$  denote the paradigm ON periods, and  $p_{\text{off}}$  denote the paradigm off periods, the off-state  $S_0$  Gaussian parameters are

$$\mu_0 = \frac{1}{|p_{\text{off}}|} \sum_{t \in p_{\text{off}}} o_t, \quad \sigma_0 = \sqrt{\frac{1}{|p_{\text{off}}|} \sum_{t \in p_{\text{off}}} (o_t - \mu_0)^2}, \quad (6)$$

and the on-state  $S_1$  Gaussian parameters are

$$\mu_1 = \frac{1}{|p_{\text{on}}|} \sum_{t \in p_{\text{on}}} o_t, \quad \sigma_1 = \sqrt{\frac{1}{|p_{\text{on}}|} \sum_{t \in p_{\text{on}}} (o_t - \mu_1)^2}, \quad (7)$$

where  $|p_{\text{off}}|$  is the total number of time samples in the off periods, and  $|p_{\text{on}}|$  is the total number of time samples in the ON periods.

Because the paradigm always starts at the off state, the parameters in  $\pi$  are set as  $\pi_0 = 1$  and  $\pi_1 = 0$ .

Given a 2-state HMM as specified, if an observation sequence does have two distinguishable states in consistence with the paradigm states, the resulting  $\{\mu_0, \sigma_0\}$  will be clearly different from  $\{\mu_1, \sigma_1\}$ , and the likelihood of such sequence given this model will be relatively high. If an observation sequence does not have such clear 2-state characteristic, the corresponding state transition will be somehow random and will not fit well with the specified  $A$  matrix. In such situation, the likelihood of this sequence will be relatively low. Therefore, the value of voxel sequence likelihood can provide an indication about the activation of this voxel. A likelihood test on an fMRI slice will be able to produce a likelihood map with each point representing the likelihood of a voxel on this slice.

To enhance the contrast of this likelihood map, we introduce a simple Gaussian model for the  $p_{\text{off}}$  samples. This model is consistent with the  $S_0$  state distribution in the 2-state HMM. The likelihood of the entire sequence is calculated based on this model. The expectation is that if a voxel is non-active, its distribution in  $p_{\text{off}}$  periods and  $p_{\text{on}}$  periods should be similar, and therefore the likelihood to this model should be relatively high; on the other hand, if the voxel is active, its distribution in  $p_{\text{on}}$  periods will be quite

different from the distribution in  $p_{\text{off}}$  periods, and therefore the likelihood of the whole sequence on this model will be relatively low. The subtraction of the HMM log likelihood map and the Gaussian log likelihood map is equivalent to a general likelihood ratio test, and it provides an activation map with enhanced contrast.

**2.2.2. State Distribution Distance Methods.** If a voxel is active, its fMRI time series can be partitioned into segments associated with two states, and each state can be described by a distribution. The assumption is that if the on-state distribution is significantly different from the off-state distribution, we have high confidence to declare a voxel as active, and vice versa. Therefore, the second method we are investigating attempts to measure the distance between the presumed on-state and off-state distributions.

There are many techniques available for measuring the distance of two distributions. We study two of such measures in this work, one is the  $t$ -test, and the other is the Kullback-Leibler divergence. Both on-state and off-state distributions are models as simple Gaussian pdfs.

Given the Gaussian parameters,  $\{\mu_0, \sigma_0\}$  and  $\{\mu_1, \sigma_1\}$ , the  $t$ -test calculates the difference of two mean values corrected by their variance values

$$t = \frac{\mu_1 - \mu_0}{\sqrt{(\sigma_{\text{on}}^2/|p_{\text{on}}|) + (\sigma_{\text{off}}^2/|p_{\text{off}}|)}}. \quad (8)$$

A  $t$ -map is produced after the  $t$ -test is applied to all the voxels on an fMRI slice. High  $t$  values in the map usually indicate active voxels.

The Kullback-Leibler divergence [37] is frequently used as a distance measure for two probability densities, although in theory it is not a true distance measure because it is not symmetric. In general it is defined in the form of “relative entropy,”

$$D(p_i(x) \parallel p_j(x)) = - \int p_i(x) \log \frac{p_j(x)}{p_i(x)} dx. \quad (9)$$

For two Gaussian pdfs, a close form expression for KLD is available:

$$D(p_i(\cdot; \mu_i, \sigma_i) \parallel p_j(\cdot; \mu_j, \sigma_j)) = \frac{1}{2} \left[ \log \left( \frac{\sigma_j^2}{\sigma_i^2} \right) - 1 + \frac{\sigma_i^2}{\sigma_j^2} + \frac{(\mu_i - \mu_j)^2}{\sigma_j^2} \right]. \quad (10)$$

These are well-established methods. However a critical issue in fMRI analysis is how to estimate the correct on-state and off-state distributions. A simple assumption is to let all time samples in the paradigm ON periods be the on-state samples and let all samples in the paradigm off periods be the off-state samples. We refer to this approach as the “paradigm state” approach. The SPM takes a similar approach, except that the block paradigm is convolved with an HRF, which is normally a low-pass filter characterizing

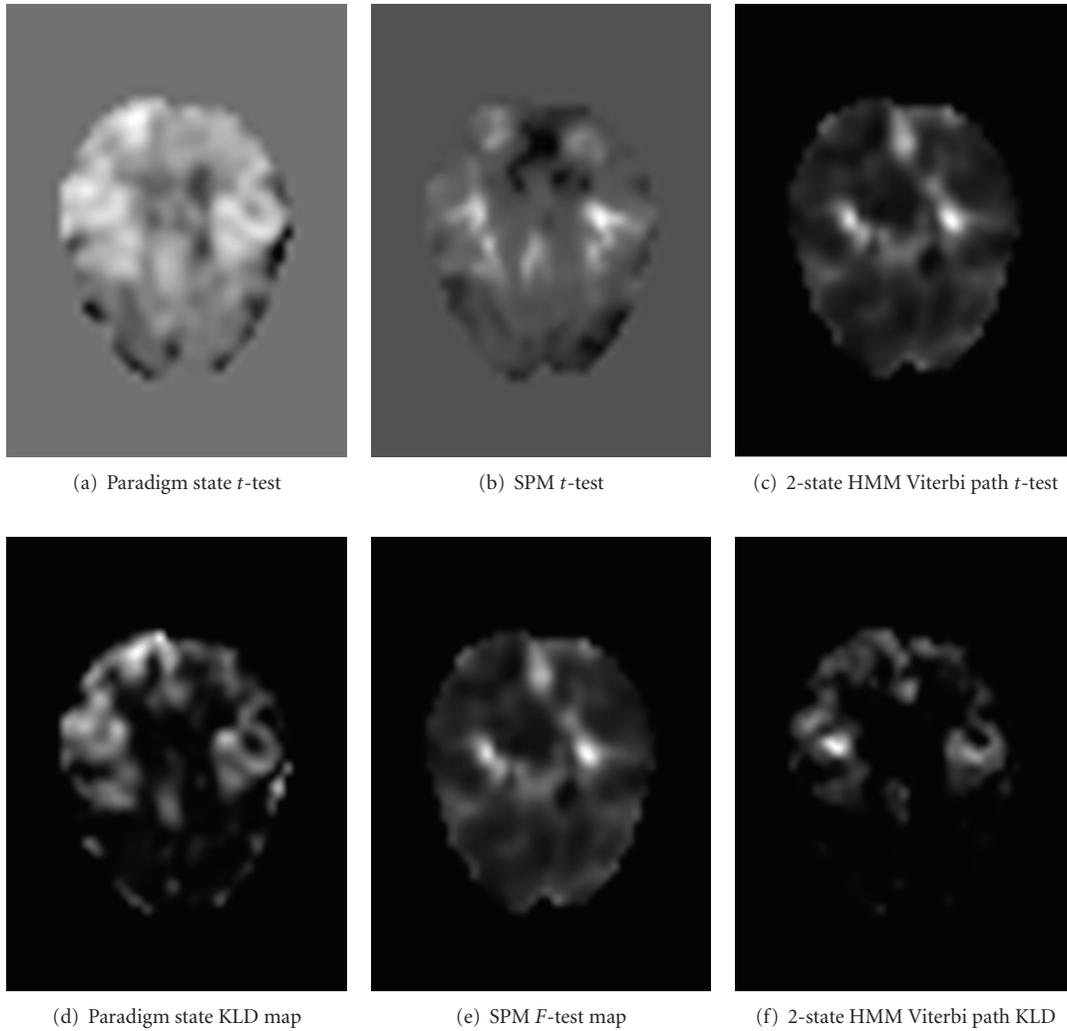


FIGURE 1: (a), (b), and (c) are  $t$ -test map for paradigm-defined state, SPM method, and 2-state HMM Viterbi path map respectively; (d), (e), and (f) are KLD map for paradigm defined state, SPM  $F$ -test, and 2-state HMM Viterbi path KLD map, respectively. The HMM Viterbi path methods (c) and (f) produce more compact and clearly highlighted regions than paradigm state estimation (a) and (d); HMM Viterbi path methods perform similarly to SPM with some minor differences.

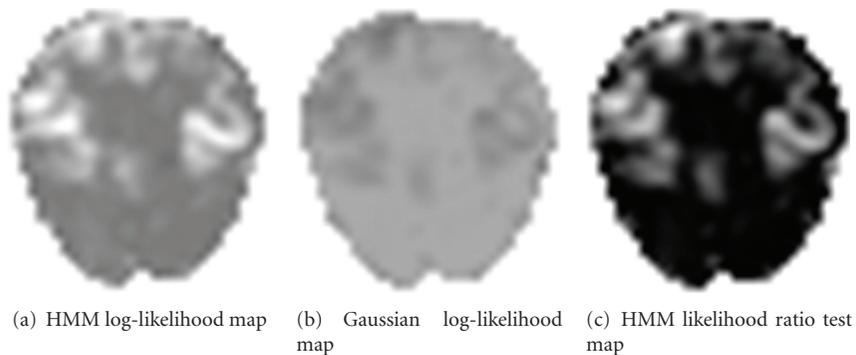


FIGURE 2: (a) HMM log-likelihood map—2-state HMM model is applied to all voxels. The active voxels with obvious 2-state on/off patterns will have relative high likelihood given this model; (b) Gaussian log-likelihood map—“off”-state Gaussian model is applied to all voxels. The nonactive voxels with obvious 2-state on/off patterns will have relative high likelihood given this model; (c) HMM likelihood ratio test map—subtraction of HMM log-likelihood map and Gaussian log-likelihood map equivalent to general likelihood ration test, and it enhances the contrast for activation map.

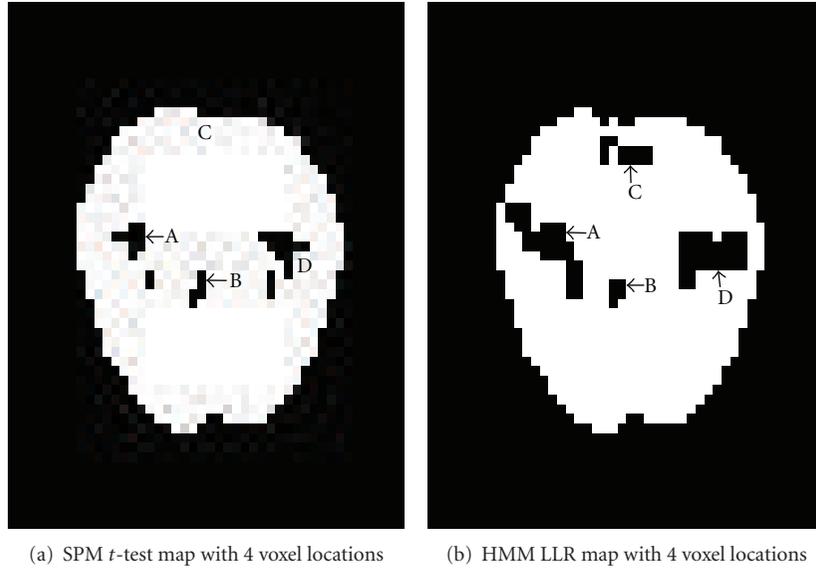


FIGURE 3: Glass map for SPM  $t$ -test and HMM likelihood ratio and 4 voxel time series. “A” and “B” are detected by both SPM and HMM methods, while “C” and “D” are only detected by the HMM log-likelihood ratio method.

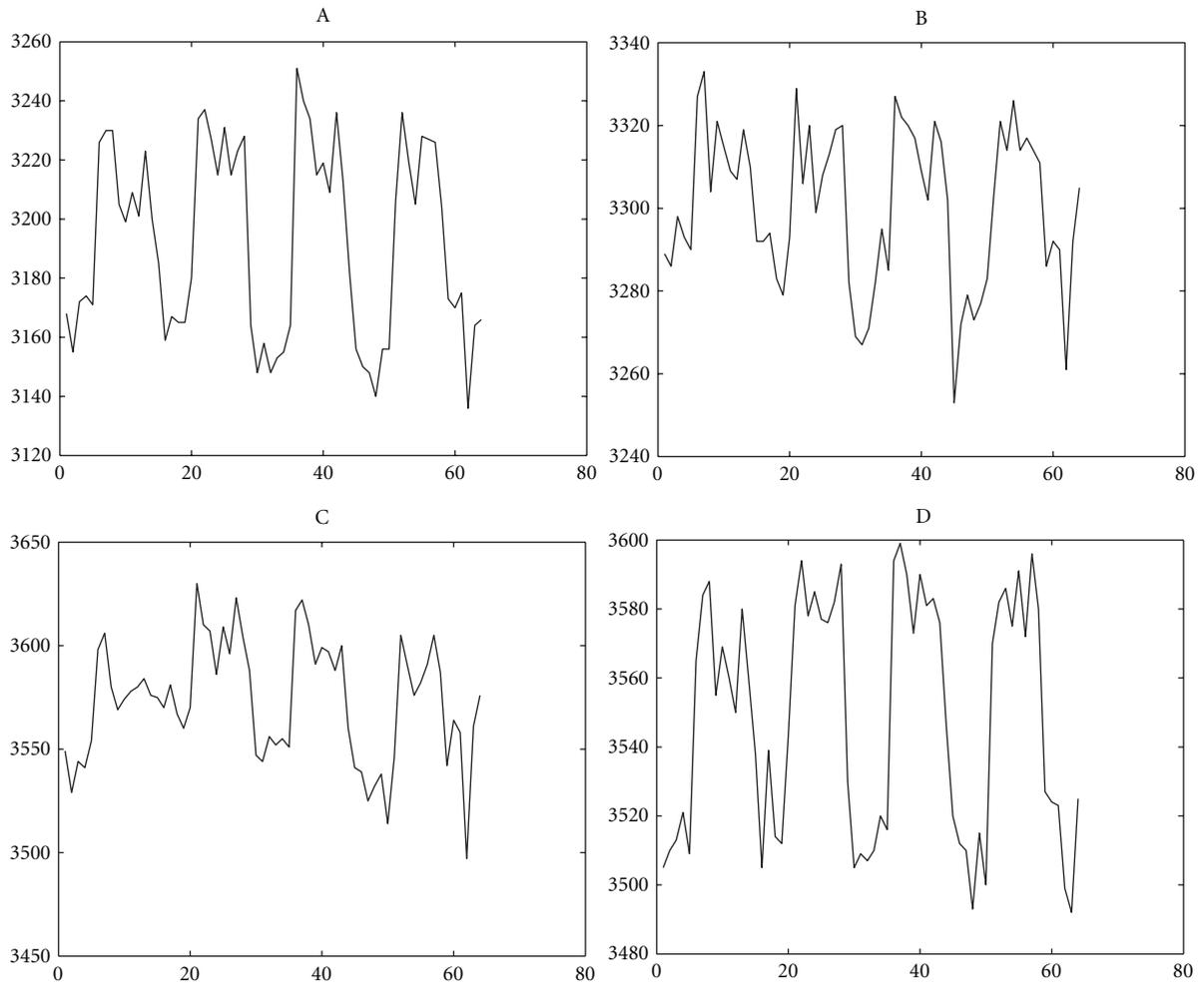


FIGURE 4: The time series of four active voxels. “A” and “B” are detected by both SPM and HMM methods, while “C” and “D” are only detected by the HMM log-likelihood ratio method.

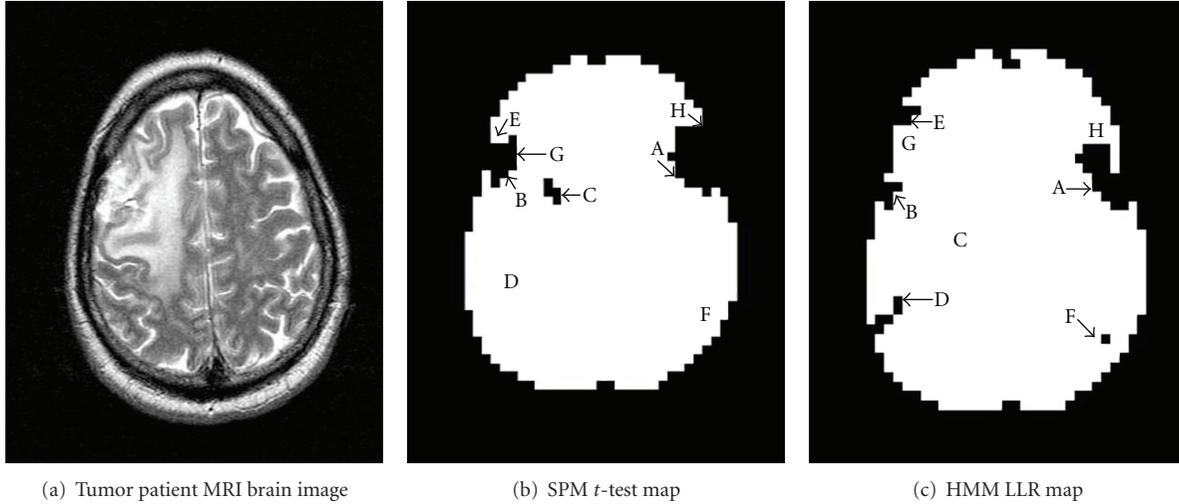


FIGURE 5: Tumor patient high-resolution image and glass maps for SPM test and 2-state HMM likelihood ratio test. The patient has a tumor on his left frontal lobe from (a). No supplemental voxels from SPM  $t$ -test (b). “D” and “F” are supplemental voxels detected from HMM LLR map in (c).

the nature voxel response to a stimulus. The  $\mu_0$  and  $\mu_1$  are obtained by projecting the time series to the HRF convolved paradigm waveform, and the  $\sigma_0$  and  $\sigma_1$  are set to be the same to model the residual error between the voxel time series and the weighted paradigm waveform.

We take a different approach by applying the 2-state HMM on each voxel series and calculate the most likely state sequence using the Viterbi algorithm. We refer to this approach as the “Viterbi path” approach. Then the on-state and off-state statistics are calculated according to the optimal state assignment for each time sample. The  $\{\mu_0, \sigma_0\}$  are obtained from all samples belonging to the off-state, and  $\{\mu_1, \sigma_1\}$  are obtained from all samples belonging to the on-state.

### 3. Experimental Results

**3.1. Normal Subject.** The data set is collected from a test with self-paced bilateral sequential thumb-to-digits opposition task. The task paradigm consists of a 32 sec baseline followed by 4 cycles of 30 sec ON and 30 sec OFF. The time series is sampled at 0.25 Hz, which produces 68 time samples for each voxel. The first four samples are ignored during analysis because of initial unstable measurement. The BOLD image is acquired in a 1.5 T GE echo speed horizon scanner with the following parameters: TR/TE = 4000/60, FOV = 24 cm,  $64 \times 64$  matrix, slice thickness 5 mm without gap, and 28 slices to cover the entire brain. Following acquisition of the functional data (with a resolution of  $3.75 \times 3.75 \times 5 \text{ mm}^3$ ), a set of 3 mm slice thickness, high resolution ( $256 \times 256$  matrix size), gadolinium-enhanced images are also obtained according to clinical imaging protocol. The data is aligned to remove the limited motion between data sets then smoothed with a Gaussian kernel before further processing [5]. We further normalize each time series with the mean and variance of its paradigm off period. In order to compensate DC drifting in

many voxels, each time series is partitioned into four equal-length segments, and normalization is performed separately on each of these segments. In the reported results, only one fMRI transverse slice is shown.

We first compare three methods based on two different distribution distance measures. These include the SPM with a  $t$ -test or an  $f$ -test, and our HMM Viterbi path method with a  $t$ -test or a KLD measure. The results are shown in Figure 1. From these results we can see that the primary motor and secondary motor areas are effectively highlighted by all these methods. We also have the following observations: (1) the HMM Viterbi path methods produce more compact and clearly highlighted regions, which indicates that Viterbi path estimation is more accurate than paradigm state estimation; (2) the HMM Viterbi path  $t$ -test method performs similarly to SPM  $t$ -test with some minor differences, mostly along the outer frontal regions; (3) the KLD methods have resemblance to SPM  $F$ -test in the sense that their results are pure positive, while  $t$ -test results are signed.

To test the effectiveness of our HMM likelihood ratio method, we compare its result with an SPM  $t$ -test result. In Figure 2, (a) shows the two-state HMM log-likelihood map; (b) shows the Gaussian log-likelihood map of the same slice; (c) shows the log-likelihood ratio test map. It can be seen that HMM log-likelihood map is almost the reverse of Gaussian log-likelihood map, which validates our expectation in Section 2.2.1. (c) is similar to (a), yet with enhanced contrast. This result resembles the SPM  $t$ -test result, although their magnitude scales are quite different.

We examine several active voxels detected by SPM and by HMM likelihood ratio test. In Figure 3, the SPM  $t$ -map and the HMM log likelihood ratio map are thresholded at certain level to yield similar number of active voxels. The corresponding voxel time series marked with “A,” “B,” “C,” and “D” are shown in Figure 4. The voxels “A” and “B” can be detected by both SPM and HMM likelihood ratio test.

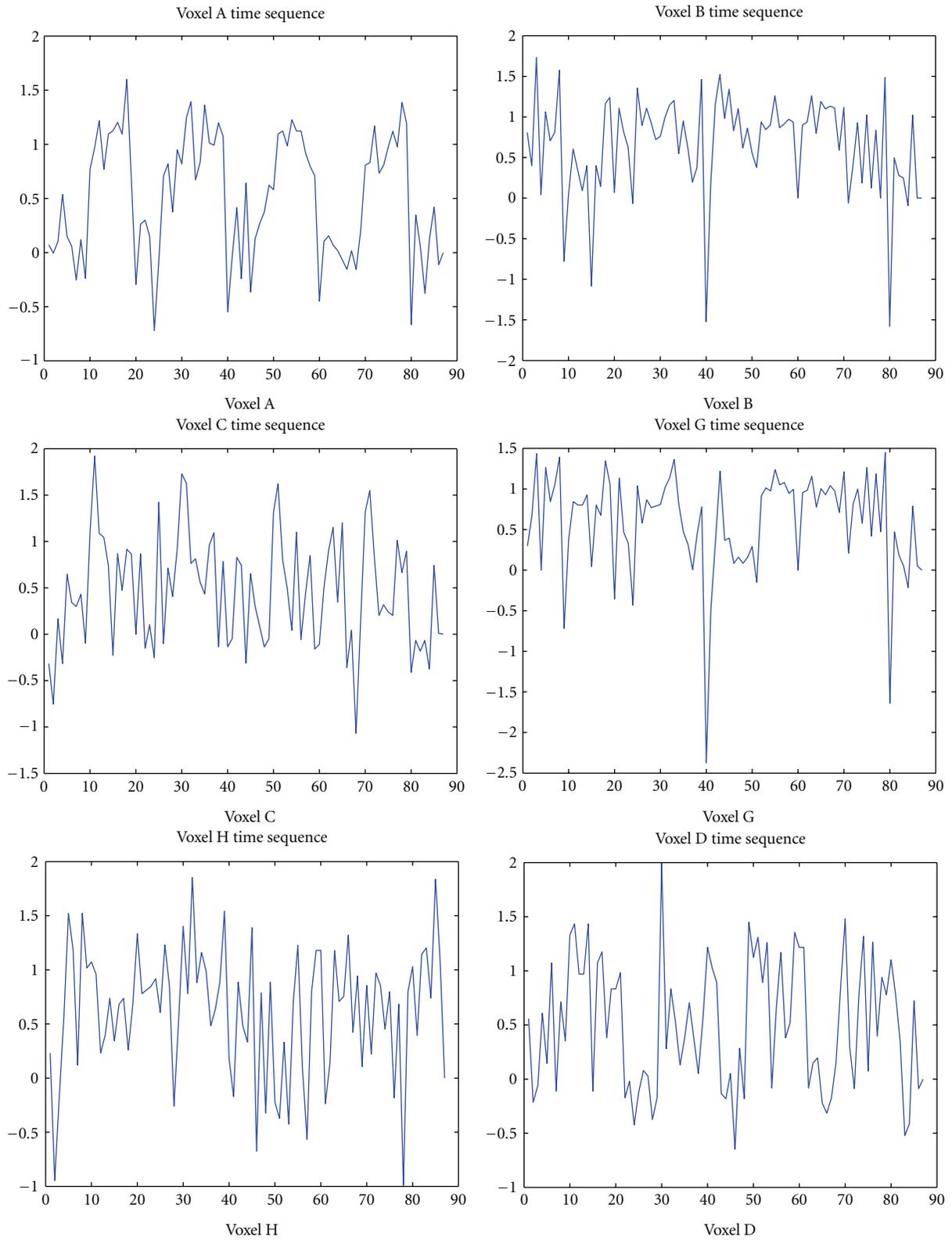


FIGURE 6: Continued.

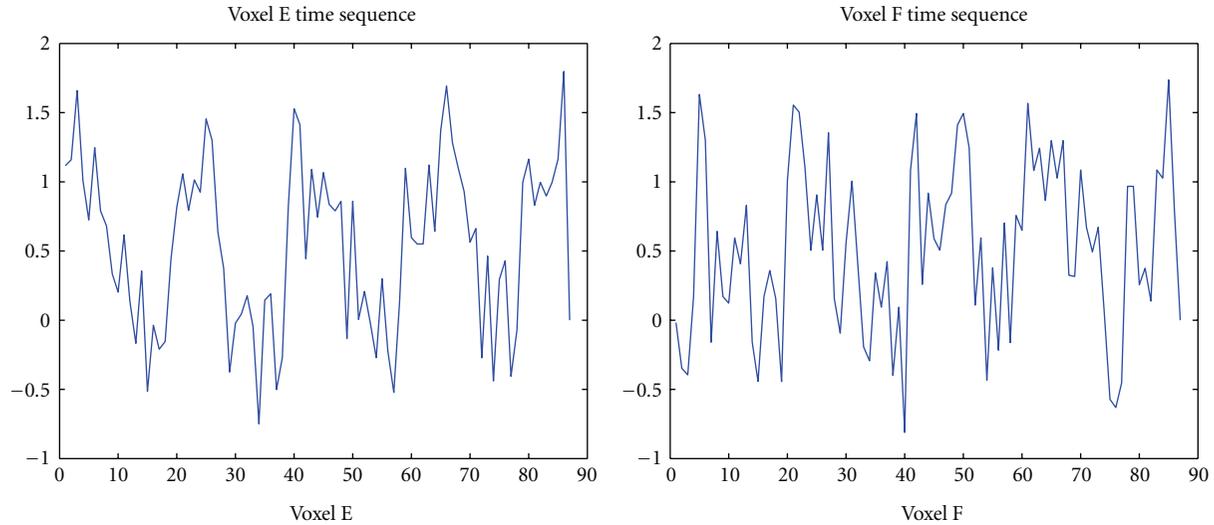


FIGURE 6: Voxel time series, voxels “A” and “B” are detected by both SPM  $t$ -test and 2-state HMM likelihood ratio test; voxels “C,” “G,” and “H” are detected by SPM  $t$ -test only. “D,” “E,” and “F” are detected by 2-state HMM likelihood ratio test only.

The voxels “C” and “D” are only highlighted by the HMM likelihood ratio test.

**3.2. Brain Tumor Subject.** Functional MRI is not only used for normal brain function mapping, it is also widely used for neurosurgical planning and neurologic risk assessment in the treatment of brain tumors. The growth of a tumor can cause functional areas to shift from their original locations. Large tumors can cause these critical regions to shift dramatically. Localizing the motor strip and coregistering the results to a surgical scan prior to a neurosurgical intervention can help guide the direct cortical stimulation during an awake craniotomy and possibly shorten operation time. In some cases, using fMRI to confirm the expected location of the motor strip may avoid awake neurosurgery altogether.

The HRF for brain tumor patient is more complicated than that for normal healthy subjects. We compare our unsupervised 2-state HMM model with GLM-based SPM on brain tumor patient and found 2-state HMM model is more robust to HRF and it is more sensitive in detecting supplemental motor activation.

The machine specification and the functional data acquisition for a tumor patient is the same as for the normal subject described above. The experiment design is different. The test is self-paced bilateral sequential thumb-to-digits opposition task. The task paradigm consists of a 20 sec baseline followed by 4 cycles of 20 sec ON and 20 sec off. Each point is 2 sec for a total of 3 min. The time series is totally 90 time samples for each voxel. The first 3 samples are ignored during analysis because of initial unstable measurement. The patient has a tumor on his left frontal lobe. As seen in the high resolution fMRI image in Figure 5(a).

Figure 5(b) is a thresholded SPM  $t$ -test map, both left and right motor areas are detected by SPM and there is no supplemental voxels. SPM  $t$ -test shows that the tumor does not impact the patient’s motor area. Thresholded HMM

likelihood ratio test result shown in Figure 5(c) indicates some weak motor activation in the left side motor area. In addition, there are some supplemental motor activation detected on surrounding areas. We further study several active voxels from Figures 5(b) and 5(c). The locations of selected active voxels are marked as “A,” “B,” “C,” “D,” “E,” “F,” “G,” and “H” in each figures, respectively. The corresponding voxel time series are shown from Figure 6. From these results, we can see that there are three types of voxels. The voxels “A” and “B” are detected by both SPM and HMM likelihood ratio test, which exhibit strong activation patterns. Voxels “C,” “G,” and “H” are detected only by SPM, and in fact they are either very weak activations or false positives. Voxels “D,” “E,” and “F” are detected only in HMM likelihood ratio test and their time series have strong activation patterns related to the paradigm but with different delay and dispersion. These results reaffirmed our understanding that SPM has difficulty in locating active voxels with unexpected delay and dispersion behaviors.

#### 4. Conclusion Remarks and Future Works

In this paper we have presented HMM-based method to detect active voxels in fMRI data. A 2-state HMM model is built based on paradigm on/off periods, and a 1-state HMM model is built based on paradigm off period. A log-likelihood ratio map is generated using the two log-likelihoods. Viterbi path is obtained for the 2-state HMM model. According to the Viterbi path,  $t$ -test map and KLD map are generated. From experiments we see that HMM methods are as effective as SPM method, and sometime HMM methods can detect supplemental active voxels that SPM may miss, especially in complicated cases with such as tumor patients. Overall we consider that the HMM methods are complementary to the SPM method, because SPM focuses on capturing fMRI signal waveform characteristics while HMM method attempts to

describe fMRI signal stochastic behaviors. In other words, the HMM methods can provide a second opinion to the SPM test results, which can be very helpful in practical situations.

## References

- [1] M. W. Woolrich, T. E. J. Behrens, and S. M. Smith, "Constrained linear basis sets for HRF modelling using Variational Bayes," *NeuroImage*, vol. 21, no. 4, pp. 1748–1761, 2004.
- [2] P. Ciuciu, J. B. Poline, G. Marrelec, J. Idier, C. Pallier, and H. Benali, "Unsupervised robust nonparametric estimation of the hemodynamic response function for any fMRI experiment," *IEEE Transactions on Medical Imaging*, vol. 22, no. 10, pp. 1235–1251, 2003.
- [3] L. Lee, L. M. Harrison, and A. Mechelli, "A report of the functional connectivity workshop, Dusseldorf 2002," *NeuroImage*, vol. 19, no. 2, pp. 457–465, 2003.
- [4] K. Li, L. Guo, J. Nie, G. Li, and T. Liu, "Review of methods for functional brain connectivity detection using fMRI," *Computerized Medical Imaging and Graphics*, vol. 33, no. 2, pp. 131–139, 2009.
- [5] K. J. Friston, A. P. Holmes, K. J. Worsley, J. P. Poline, C. D. Frith, and R. S. J. Frackowiak, "Statistical parametric maps in functional imaging: a general linear approach," *Human Brain Mapping*, vol. 2, no. 4, pp. 189–210, 1994.
- [6] M. D. Greicius, B. Krasnow, A. L. Reiss, and V. Menon, "Functional connectivity in the resting brain: a network analysis of the default mode hypothesis," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 1, pp. 253–258, 2003.
- [7] J. Ashburner, "Spm:a history," *NeuroImage*, vol. 62, no. 2, pp. 791–800, 2012.
- [8] G. K. Wood, "Visualization of subtle contrast-related intensity changes using temporal correlation," *Magnetic Resonance Imaging*, vol. 12, no. 7, pp. 1013–1020, 1994.
- [9] J. Cao and K. Worsley, "The geometry of correlation fields with an application to functional connectivity of the brain," *Annals of Applied Probability*, vol. 9, no. 4, pp. 1021–1057, 1999.
- [10] P. Hojen-Sorensen, L. Hansen, and C. Rasmussen, "Bayesian modelling of fmri time series," in *Proceedings of the 13th Annual Conference on Advances in Neural Information Processing Systems (NIPS '99)*, pp. 754–760, 2000.
- [11] C. Gossel, D. Auer, and L. Fahtmeir, "Dynamic models in fmri," *Magnetic Resonance in Medicine*, vol. 43, no. 1, pp. 72–81, 2000.
- [12] M. Singh, W. Sungkarat, J. W. Jeong, and Y. Zhou, "Extraction of temporal information in functional MRI," *IEEE Transactions on Nuclear Science*, vol. 49, no. 5, pp. 2284–2290, 2002.
- [13] V. Sanguineti, C. Parodi, S. Perissinotto et al., "Analysis of fMRI time series with mixtures of Gaussians," in *Proceedings of the International Joint Conference on Neural Networks (IJCNN '2000)*, vol. 1, pp. 331–335, July 2000.
- [14] S. Faisan, L. Thoraval, J. P. Armspach, and F. Heitz, "Unsupervised learning and mapping of brain fMRI signals based on hidden semi-Markov event sequence models," in *Proceedings of the 6th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI '03)*, pp. 75–82, November 2003.
- [15] B. Thirion and O. Faugeras, "Revisiting non-parametric activation detection on fMRI time series," in *Proceedings of the IEEE Workshop on Mathematical Methods in Biomedical Image Analysis*, vol. 1, pp. 121–128, December 2001.
- [16] M. W. Woolrich, M. Jenkinson, J. M. Brady, and S. M. Smith, "Fully bayesian spatio-temporal modeling of FMRI data," *IEEE Transactions on Medical Imaging*, vol. 23, no. 2, pp. 213–231, 2004.
- [17] A. Roebroeck, E. Formisano, and R. Goebel, "Mapping directed influence over the brain using Granger causality and fMRI," *NeuroImage*, vol. 25, no. 1, pp. 230–242, 2005.
- [18] K. J. Friston, L. Harrison, and W. Penny, "Dynamic causal modelling," *NeuroImage*, vol. 19, no. 4, pp. 1273–1302, 2003.
- [19] K. Friston, "Dynamic causal modeling and Granger causality comments on: the identification of interacting networks in the brain using fMRI: model selection, causality and deconvolution," *NeuroImage*, vol. 58, pp. 303–305, 2011.
- [20] O. David, "fMRI connectivity, meaning and empiricism. Comments on: Roebroeck et al. The identification of interacting networks in the brain using fMRI: model selection, causality and deconvolution," *NeuroImage*, vol. 58, pp. 306–309, 2011.
- [21] A. Roebroeck, E. Formisano, and R. Goebel, "Reply to Friston and David. After comments on: the identification of interacting networks in the brain using fMRI: model selection, causality and deconvolution," *NeuroImage*, vol. 58, pp. 296–302, 2011.
- [22] A. Roebroeck, E. Formisano, and R. Goebel, "Reply to Friston and David. After comments on: the identification of interacting networks in the brain using fMRI: model selection, causality and deconvolution," *NeuroImage*, vol. 58, pp. 310–311, 2011.
- [23] G. Lohmann, K. Erfurth, K. Mller, and R. Turner, "Critical comments on dynamic causal modelling," *NeuroImage*, vol. 59, pp. 2322–2329, 2012.
- [24] M. Mckeown, S. Makeig, G. Brown et al., "Analysis of fmri data by blind separation into independent spatial components," *Human Brain Mapping*, vol. 6, pp. 160–188, 1998.
- [25] B. B. Biswal and J. L. Ulmer, "Blind source separation of multiple signal sources of fMRI data sets using independent component analysis," *Journal of Computer Assisted Tomography*, vol. 23, no. 2, pp. 265–271, 1999.
- [26] V. D. Calhoun, T. Adali, G. D. Pearlson, and J. J. Pekar, "Spatial and temporal independent component analysis of functional MRI data containing a pair of task-related waveforms," *Human Brain Mapping*, vol. 13, no. 1, pp. 43–53, 2001.
- [27] J. V. Stone, J. Porrill, N. R. Porter, and I. D. Wilkinson, "Spatiotemporal independent component analysis of event-related fMRI data using skewed probability density functions," *NeuroImage*, vol. 15, no. 2, pp. 407–421, 2002.
- [28] A. J. Bell and T. J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural Computation*, vol. 7, no. 6, pp. 1129–1159, 1995.
- [29] F. J. Theis, P. Gruber, I. R. Keck, and E. W. Lang, "A robust model for spatiotemporal dependencies," *Neurocomputing*, vol. 71, no. 10–12, pp. 2209–2216, 2008.
- [30] G. McLachlan and D. Peel, *Finite Mixture Models*, John Wiley & Sons, New York, NY, USA, 2000.
- [31] L. R. Rabiner, "Tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, 1989.
- [32] G. Xu, Y. F. Ma, H. J. Zhang, and S. Q. Yang, "An HMM-based framework for video semantic analysis," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 15, no. 11, pp. 1422–1433, 2005.

- [33] A. Krogh, M. Brown, I. S. Mian, K. Sjolander, and D. Haussler, "Hidden Markov Models in computational biology applications to protein modeling," *Journal of Molecular Biology*, vol. 235, no. 5, pp. 1501–1531, 1994.
- [34] R. Durbin, S. Eddy, A. Krogh, and G. Mitchison, *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids*, Cambridge University Press, 1999.
- [35] D. Miller, T. Leek, and R. Schwartz, "A hidden markov model information retrieval system," in *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 214–221, ACM, 1999.
- [36] G. D. Forney, "The viterbi algorithm," *Proceedings of the IEEE*, vol. 61, no. 3, pp. 268–278, 1973.
- [37] S. Kullback and R. Leibler, "On information and sufficiency," *Annals of Mathematical Statistics*, vol. 22, no. 1, pp. 79–86, 1951.



**Hindawi**

Submit your manuscripts at  
<http://www.hindawi.com>

