Analysis of Event-Related fMRI Data Using Best Clustering Bases

François G. Meyer*, *Member, IEEE*, and Jatuporn Chinrungrueng

***Abstract*—We explore a new paradigm for the analysis of event-related functional magnetic resonance images (fMRI) of brain activity. We regard the fMRI data as a very large set of time series $x_i(t)$, indexed by the position i of a voxel inside the brain. The decision that a voxel i_0 is activated is based not solely on the value of the fMRI signal at i_0, but rather on the comparison of all time series $x_i(t)$ in a small neighborhood W_{i_0} around i_0. We construct basis functions on which the projection of the fMRI data reveals the organization of the time series $x_i(t)$ into activated and nonactivated clusters. These clustering basis functions are selected from large libraries of wavelet packets according to their ability to separate the fMRI time series into the activated cluster and a nonactivated cluster. This principle exploits the intrinsic spatial correlation that is present in the data. The construction of the clustering basis functions described in this paper is applicable to a large category of problems where time series are indexed by a spatial variable.**

***Index Terms*—Best clustering basis, brain mapping, fMRI, functional magnetic resonance imaging, wavelets.**

I. Introduction

FUNCTIONAL magnetic resonance imaging (fMRI) utilizes the fact that oxygenated blood and deoxygenated blood have different magnetic susceptibility to create maps of changes in cerebral venous oxygen concentration that correlate with neuronal activity. There are currently two types of experimental designs: the periodic stimulus design (block design) and the event-related stimulus design (odd-ball design). The detection of activation using event-related fMRI data is the more difficult problem and will be the focus of this paper. The analysis of event-related fMRI data commonly relies on the assumption that the fMRI signal $\mathbf{x}_i = [x_i(0), \ldots, x_i(T-1)]^T$ at a voxel i inside the brain is obtained by convolving the stimulus waveform $\mathbf{s}_i = [s_i(0), \ldots, s_i(T-1)]^T$ with some unknown hemodynamic response function $\mathbf{h}_i$ [1]. Unfortunately, it has become clear that there is a nonlinear relationship between the variation in the fMRI signal and the stimulus presentation [2]. In the absence of any detailed substantive understanding of the mechanism of the fMRI response, we advocate a nonparametric data-driven approach that can better account for all important features present in the data. Fig. 1 shows a block diagram that illustrates the principle of our approach. We consider a group of voxels and the corresponding time series from a small window W_{i_0} inside the brain. We partition these time series into two clusters. If W_{i_0} is located in a part of the brain that is activated, then one of the two clusters encompasses the activated voxels or activated time series. The other cluster contains time series that correspond to background activity. If the window is in a part of the brain with no activity correlated to the stimulus, then all time series are considered to be background activity. As W_{i_0} is moved throughout the brain, local decisions (activation/nonactivation) are computed for each voxel $i \in W_{i_0}$. This principle exploits the intrinsic spatial correlation that is present in the data. Indeed, truly activated voxels tend to be spatially clustered, while falsely activated voxels tend to be scattered. One can then increase the sensitivity of the detection by using the fact that real activation should be more clustered than artifactual activation caused by noise. These local decisions are combined to generate a more robust global activation score, which can be presented in the form of an activation map. Unlike global clustering methods [3], our local clustering approach only partitions the time series in a small region of the brain. Furthermore, the clustering of the time series is not performed directly on the raw fMRI signal. Instead, the raw data are projected on a set of basis functions conveniently chosen for their discriminating power and their robustness to noise. If the stimulus is periodic, the Fourier basis provides the most interesting projection of the data [4]. If the stimulus is not periodic, then we need to replace the Fourier transform with a transform that can efficiently represent the nonlinear and nonstationary structures present in event-related data. Several studies indicate

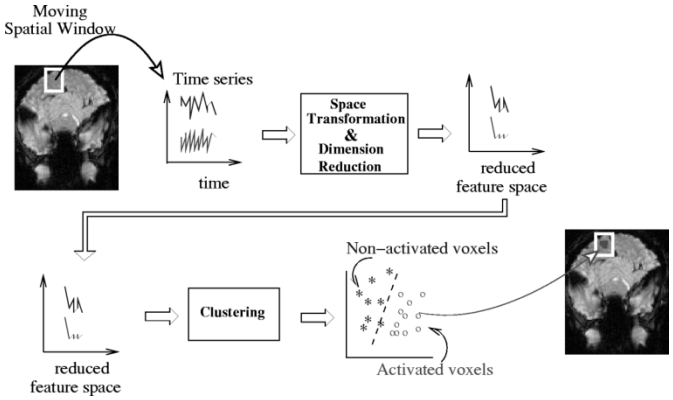

Fig. 1. We partition the time series inside a small window W_{i_0} into two clusters. The clustering is performed on the wavelet packet coefficients (feature space).

Manuscript received July 16, 2002; revised February 25, 2003. This work was supported by a Whitaker Foundation Biomedical Engineering Research Grant. The Associate Editor responsible for coordinating the review of this paper and recommending its publication was A. Aldroubi. *Asterisk indicates corresponding author.*

*F. G. Meyer is with the Department of Electrical Engineering, University of Colorado, Boulder, CO 80309 USA (e-mail: francois.meyer@colorado.edu).

J. Chinrungrueng is with the Department of Electrical Engineering, University of Colorado, Boulder, CO 80309 USA.

Digital Object Identifier 10.1109/TMI.2003.815869

that one finds dynamic changes of the fMRI signal in time, frequency, and space [5], [6]. Wavelet packets are time-frequency "atoms" that are localized in time and frequency. Wavelet packets have been used quite successfully to analyze transients in evoked potentials or electroencephalograms [7]. We favor, therefore, the use of wavelet packets to perform the analysis of the event-related hemodynamic response $\mathbf{x}_i$. Instead of using a fixed basis, we will select a basis from a very large and highly redundant dictionary of wavelet packets, $\mathcal{B} = \{\psi_\gamma\}$. Because the dictionary is highly redundant, one can select those wavelet packets $\psi_\gamma = [\psi_\gamma(0), \ldots, \psi_\gamma(T-1)]^T$ (where γ is a parameter indexing the functions) on which the projection of event-related fMRI data reveals the organization of the time series into activated and nonactivated clusters. The representation of an event-related signal $\mathbf{x}_i$ in such a basis is defined by $\mathbf{x}_i = \sum_\gamma \alpha_\gamma(\mathbf{x}_i)\psi_\gamma$. This approach does not assume any particular model of the hemodynamic response, but rather tries to "let the data speak for themselves."

This paper is organized as follows. In Section II, we provide a brief review of the wavelet packets and the best-basis algorithm [8]. In Section III, we describe the new best clustering basis algorithm. Results of experiments conducted on synthetic and *in vivo* fMRI data are presented in Section IV.

II. Wavelet Packets, Best Basis

A. Wavelet Packets

Let $\psi^0(t)$ be the scaling function and let $\psi^1(t)$ be the wavelet associated with a multiresolution analysis [9]. Let $\{h_n\}$ be the low-pass filter, and let $\{g_n\}$ be the high-pass filter associated with this wavelet transform. Let us define the basic wavelet packets by

$$\psi^{2n}(t) = \sum_k h_k \psi^n(2t-k)$$
$$\psi^{2n+1}(t) = \sum_k g_k \psi^n(2t-k). \tag{1}$$

A (multiscale) wavelet packet is given by

$$\psi_{j,k,l}(t) = \psi^k(2^j t - l). \tag{2}$$

A wavelet packet is identified by an index $\gamma = (j,k,l)$, where

- $j = 0, \ldots, J$ represents the scale: $\psi_{j,k,l}$ has a support of size 2^{-j}, and J is the maximum scale ($J \leq J_0$);
- $k = 0, \ldots, 2^j - 1$ represent the frequency index at a given scale j : $\psi_{j,k,l}$ has roughly k oscillations;
- $l = 0, \ldots, 2^{J_0 - j} - 1$ represents the translation index within a node (j,k) : $\psi_{j,k,l}$ is located at $l2^{-j}$.

As shown in Fig. 2, the library of wavelet packets organizes itself into a binary tree, where the nodes of the tree represent subspaces with different time-frequency localization characteristics. A word on notation: We consider ψ_γ, or equivalently $\psi_{j,k,l}$, to be a $2^{J_0} \times 1$ vector. We will abuse notations and define the $2^{J_0} \times 2^{J_0 - j}$ matrix

$$\mathbf{\Psi}_{j,k} = [\psi_{j,k,0} \,|\, \cdots \,|\, \psi_{j,k,2^{J_0-j}-1}] \tag{3}$$

to be the collection of wavelet packets at node (j,k) stacked together. The wavelet packet coefficients $\alpha_{j,k,l}, l = 0, \ldots, 2^{J_0-j}$ of the vector $\mathbf{x}$ at the node (j,k) are given by the projections of $\mathbf{x}$ on the 2^{J_0-j} wavelet packets stacked inside the matrix $\mathbf{\Psi}$. In matrix form we have

$$[\alpha_{j,k,0}, \ldots \alpha_{j,k,2^{J_0-j}-1}]^T = \mathbf{\Psi}_{j,k}^T \mathbf{x}. \tag{4}$$

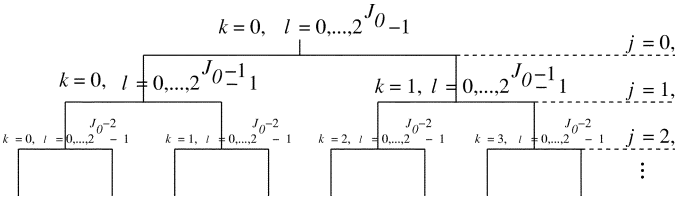

Fig. 2. Wavelet packets tree showing the index $\gamma = (j,k,l)$ at each node.

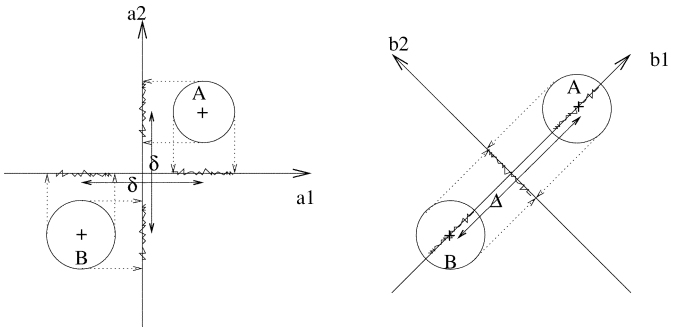

Fig. 3. Basis $(\mathbf{b}_1, \mathbf{b}_2)$ concentrates its discriminatory power on a single vector $\mathbf{b}_1$, whereas the basis $(\mathbf{a}_1, \mathbf{a}_2)$ distributes the power equally on the vectors $\mathbf{a}_1$ and $\mathbf{a}_2$.

Let us associate the dyadic frequency interval $[2^j k, 2^j(k+1))$ with the wavelet packet $\psi_{j,k,l}$. If a collection of intervals $[2^j k, 2^j(k+1))$ provides a cover of the time-frequency plane, then the set $\{\psi_{j,k,l}\}$ is an orthonormal basis [8]. If we perform a dyadic subdivision at each level, we get 2^{2^J} bases for J levels. Clearly, we have an extremely large amount of freedom for the construction of orthogonal bases from the wavelet packet library $\{\psi_{j,k,l}\}$. This greater flexibility can be exploited to increase the efficiency of the representation. Coifman and Wickerhauser [8] suggested the use of a dynamic programming algorithm with order $T\log(T)$ to search for that *best basis* which is optimal according to a given cost function $\mathcal{M}$. These dictionaries have been used recently in the context of supervised classification [10]. In this paper, we intend to address a more general and deeper problem, where one wants to determine the best clustering basis without having access to a training set containing activated and nonactivated signals.

III. Local Clustering Bases

In our problem, we need a cost functional $\mathcal{M}$ that measures how well a basis function ψ_γ can partition a set of signals into meaningful clusters. We illustrate our idea with a simple geometric example shown in Fig. 3. We consider two sets of points distributed respectively inside the two circles A and B. We define two different bases: $(\mathbf{a}_1, \mathbf{a}_2)$, and $(\mathbf{b}_1, \mathbf{b}_2)$. One could imagine that the bases $(\mathbf{a1}, \mathbf{a2})$ and $(\mathbf{b1}, \mathbf{b2})$ correspond to two different wavelet packet bases. In order to discover the two clusters A and B, we can partition the points using a clustering

algorithm. We can use either one of the two bases to perform the clustering. However, to simplify the clustering process and to avoid the curse of dimensionality, one should use the basis that concentrates its discriminatory power onto a small number of basis functions. Let us consider a vector $\mathbf{v}$, and let us project the dataset on this vector. If we cluster the projections into two partitions, we can define the discriminatory power of $\mathbf{v}$ as the distance between the two centroids. The vectors $\mathbf{a}_1$ and $\mathbf{a}_2$ have the same discriminatory power, δ (see Fig. 3). However, the vector $\mathbf{b}_1$ has a discriminatory power Δ that is larger than the discriminatory power of $\mathbf{a}_1$ or $\mathbf{a}_2$. The vector $\mathbf{b}_2$, on the other hand, does not discriminate one cluster from another. Therefore, if we use the basis $(\mathbf{b}_1, \mathbf{b}_2)$ the discriminatory power is only concentrated in the vector $\mathbf{b}_1$. We should choose the basis $(\mathbf{b}_1, \mathbf{b}_2)$ and perform the clustering using only the vector $\mathbf{b}_1$. Our clustering cost functional reflects this principle and selects the basis that concentrates the most discriminatory power on a small number of basis functions.

A. Normalized Distance

In order to define a cost functional adapted to our clustering problem, we need to define a measure of the separation between two clusters of projections along the basis vector ψ_γ. We partition the set of coefficients $\{\alpha_\gamma(\mathbf{x}_i), i = 0, \ldots, N-1\}$ into two clusters and define $\mu_{i,c}$ to be the membership value of the wavelet packet coefficient $\alpha_\gamma(\mathbf{x}_i)$ in the cluster c. $\mu_{i,c}$ is a real number in $[0,1]$ that measures the likelihood that $\alpha_\gamma(\mathbf{x}_i)$ belongs to the cluster $c, (\mu_{i,1} + \mu_{i,2} = 1)$. We first compute the centroid $\bar{\alpha}_\gamma(c) = \sum_{i=0}^{N-1} \mu_{i,c}\alpha_\gamma(\mathbf{x}_i)$ of the cluster c. We then calculate the in-class variance $s_\gamma^2(c) = \sum_{i=0}^{N-1} \mu_{i,c}(\alpha_\gamma(\mathbf{x}_i) - \bar{\alpha}_\gamma(c))^2/(N-1)$ of each cluster c. A good separation will be achieved if the distance between the cluster centroids is large relative to the in-class variance of each cluster. For a given wavelet packet $\psi_\gamma, \gamma = (j,k,l)$, the normalized distance $\mathcal{D}(\gamma)$ between the two clusters is defined by

$$\mathcal{D}(\gamma) = \frac{|\bar{\alpha}_\gamma(c_1) - \bar{\alpha}_\gamma(c_2)|}{\sqrt{s_\gamma^2(c_1)s_\gamma^2(c_2)}}. \tag{5}$$

Our definition of the normalized variance is similar to the definition of the Fisher linear discriminant. As explained in the previous section, we are interested in finding a basis where a small number of basis functions ψ_γ are capable of clustering the wavelet packets coefficients $\alpha_\gamma(\mathbf{x}_i)$ into two clusters with a large distance $\mathcal{D}(\gamma)$. In other words, we favor a sparse distribution of the distances $\{\mathcal{D}(\gamma)\}_\gamma$. One way to characterize the sparsity of the distribution is to compute its entropy. Given a wavelet packet node (j,k), we define the cost function for the subspace $\mathbf{\Psi}_{j,k}$ by

$$\mathcal{M}(\mathbf{\Psi}_{j,k}) = -\sum_{l=0}^{2^{J_0-j}-1} \frac{\mathcal{D}^2(j,k,l)}{\|\mathcal{D}(j,k,\cdot)\|^2} \log \frac{\mathcal{D}^2(j,k,l)}{\|\mathcal{D}(j,k,\cdot)\|^2} \tag{6}$$

where $\|\mathcal{D}(j,k,\cdot)\|^2 = \sum_{l=0}^{2^{J_0-j}-1} \mathcal{D}^2(j,k,l)$. $\mathcal{M}(j,k)$ will be maximum if all the wavelet packets $\psi_{j,k,l}$ at the node (j,k) have the same distance $\mathcal{D}(j,k,l)$. Such a subspace of wavelet packets will be of no interest for our purpose. Another way to measure the sparsity of the distribution of $\mathcal{D}(\gamma)$ is to compute the l^p norm $(p \leq 1)$ of the vector $[\mathcal{D}^2(j,k,0), \ldots, \mathcal{D}^2(j,k,2^{J_0-j}-1)]$.

B. Local Clustering Basis Selection

We are now ready to define the best clustering basis algorithm. This algorithm searches for the optimal basis according to the criterion defined in (6). The wavelet packet tree (see Fig. 2) is explored from the bottom up, and the optimal combination of the $\psi_{j,k,l}$ is kept. Given a set of N time series vectors $\{\mathbf{x}_i\}_{i=0}^{N-1}$, indexed by their position i, the search proceeds as follows.

1) *Wavelet Packet Expansion:* For each vector $\mathbf{x}_i$, compute the wavelet packet coefficients $\alpha_\gamma(\mathbf{x}_i)$ at each node (j,k) of the wavelet packet tree.
2) *Clustering:* For each wavelet packet index (j,k,l), cluster the set $\{\alpha_{j,k,l}(\mathbf{x}_i), i = 0, \ldots, N-1\}$ into two clusters, and compute the distance $\mathcal{D}(j,k,l)$.

   For each wavelet packet node node (j,k), compute the cost function $\mathcal{M}(\mathbf{\Psi}_{j,k})$.
3) *Divide and Conquer:* For the coarsest scale J, initialize the best basis with $\mathbf{B}_{J,k} = \mathbf{\Psi}_{J,k}, k = 0, \ldots, 2^J - 1$. For the scales $J-1$ until 0, determine the best subspace $\mathbf{B}_{j,k}$ as follows. If $\mathcal{M}(\mathbf{\Psi}_{j,k}) \leq \mathcal{M}(\mathbf{B}_{j+1,2k}) + \mathcal{M}(\mathbf{B}_{j+1,2k+1})$, then choose $\mathbf{B}_{j,k}$ to be the current node, $\mathbf{B}_{j,k} = \mathbf{\Psi}_{j,k}$. Otherwise, choose $\mathbf{B}_{j,k}$ to be the direct sum of $\mathbf{B}_{j+1,2k}$ and $\mathbf{B}_{j+1,2k+1}$, $\mathbf{B}_{j+1,2k} \oplus \mathbf{B}_{j+1,2k+1}$.

The output of the algorithm is the best basis $\mathbf{B}_{0,0} = \{\mathbf{b}_\gamma\}, \gamma \in \Gamma_0$. We have noticed in our experiments that the basis vectors that provide the best clustering power will generate a set of coefficients with a large variance. Conversely, vectors that are not useful for separating the time series give rise to coefficients with a very small variance. Because we want to retain only the basis vectors with the largest clustering power, we rank the basis vectors according to their variance $\sigma^2(\gamma)$, and we keep only the first T_r vectors $\mathbf{b}_{\gamma_l}$ that contribute to a percentage r of the total variance. These T_r basis vectors constitute the "clustering" space.

C. Interpretation of the Clusters

If the region W_{i_0} is located in a part of the brain that is activated, then one of the two clusters contains activated voxels, and the other cluster is composed of time series that describe background activity. According to our experimental findings the cluster with the largest in-class variance $s_\gamma^2(c)$ corresponds to the activated cluster. If the window is in a part of the brain with no activity correlated to the stimulus, then all time series correspond to background activity, and the partition of the time series into two clusters is artificial. In this latter case, both the normalized distance $\mathcal{D}_\gamma$ between the two clusters and the total variance of the wavelet packet coefficients are much smaller than when the region W_{i_0} contains a mixture of activated and background time series. We use this observation to discriminate between the two situations: 1) the two clusters come from background activity (unrelated to the stimulus) and 2) one cluster contains the activated time series and the other the background time series, by comparing the distance between the cluster centroids as well as the total variance within the window W_{i_0}.

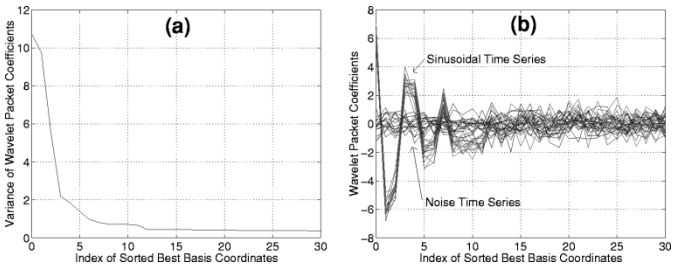

Fig. 4. (a) Variance $\sigma^2(\gamma_l)$ as a function of l. (b) Wavelet packet coefficients $\alpha_{\gamma_l}(\mathbf{x}_i)$ for each i. For both graphs, we only show the first 30 coefficients.

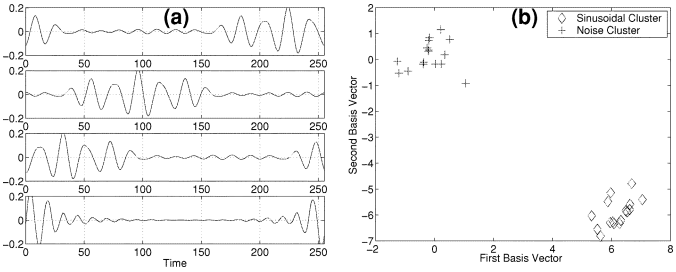

Fig. 5. (a) The four most discriminating basis vectors ($\gamma = (6,6,1), (6,6,3), (6,6,2)$, and $(5,4,6)$). These vectors locally oscillate at the frequency of the sinusoidal signal. (b) Scatter plot obtained by projecting the time series onto the top two vectors shown on the left.

IV. Experiments and Results

We present in this section experiments conducted on several artificial data sets and one event-related fMRI data set. For all experiments, we used the Coiflet filter with four vanishing moments. We implemented our local clustering basis algorithm using the Matlab Wavelab package.

A. Sinusoidal Signals

We now demonstrate that our algorithm can separate a mixture of periodic sinusoidal signals from white Gaussian noise signals. We generate 16 time series for each type (sinusoids or noise) of signals. The activated time series are given by $x_i(t) = \sin(2\pi t/20) + (1/2)n_1(t)$ and the noise time series are given by $x_i(t) = (1/2)n_2(t)$, where n_1 and $n_2 \sim N(0,1)$ are standard white Gaussian noises and $t = 0, \ldots, 255$. The signal-to-noise ratio (SNR) is 1. Fig. 4(a) shows the variance $\sigma^2(\gamma_l)$ of the first 30 wavelet packet coefficients of the best basis. We need $T_r = 12$ basis vectors to capture 40% ($r = 0.4$) of the variance. Fig. 4(b) shows the wavelet packet coefficients of all the time series projected onto the first 30 basis vectors $\{\mathbf{b}_{\gamma_l},\}_{l=0}^{29}$. As is clear from this figure, the wavelet coefficients of the sinusoidal signals cannot be distinguished from the coefficients of the noise as soon as $l > 12$. This visually confirms that only the first twelve vectors are useful to separate the two clusters. This reduces the dimensionality of the problem by a factor of 21. Fig. 5(a) shows the four most discriminating basis vectors, with indexes $\gamma = (6,6,1), (6,6,3), (6,6,2)$, and $(5,4,6)$. These four basis vectors clearly show localized oscillations at the frequency of the sinusoidal signals. Obviously, the best discriminating vector is the sinusoidal function $\sin(2\pi t/20)$. Because the wavelet packet dictionary does not contain this particular function, the local clustering algorithm cannot find it.

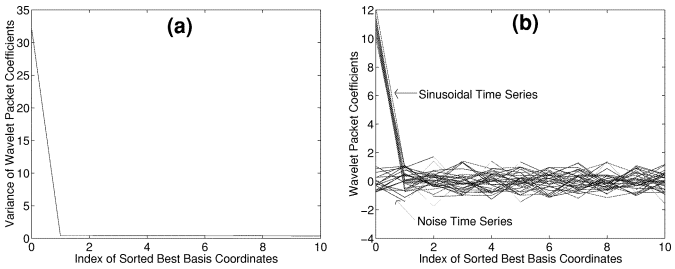

Fig. 6. (a) Variance $\sigma^2(\gamma_l)$ of the first ten wavelet packet coefficients of the best basis. (b) Wavelet packet coefficients projected onto the first ten basis vectors $\{\mathbf{b}_{\gamma_l},\}_{l=0}^{10}$.

Fig. 5(b) shows a scatter plot obtained by projecting all the time series onto the first two best basis vectors. The plot shows a very clear separation of the two clusters. Because wavelet packets can mimic sines and cosines, we generate a second data set similar to the previous one but with a new sinusoidal signal $\sin(\omega_0 t + \phi_0)$. The frequency ω_0 and the phase ϕ_0 are carefully chosen such that the signal $\sin(\omega_0 t + \phi_0)$ has the largest correlation with the wavelet packet that has the most sine-like structure. The SNR is 1. Fig. 6(a) shows the variance $\sigma^2(\gamma_l)$ of the first ten wavelet packet coefficients of the best basis. The first basis vector ($\gamma = (8,7,0)$) is the wavelet packet with the most sine-like structure; it is in fact visually indistinguishable from the original sinusoidal signal. It captures 40% of the total variance. Fig. 6(b) shows the wavelet packet coefficients projected onto the first ten basis vectors $\{\mathbf{b}_{\gamma_l},\}_{l=0}^{10}$. As is clear from this figure, the wavelet coefficients of the signals cannot be distinguished from the coefficients of the noise as soon as $l \geq 2$. This experiment indicates that our approach could be applied to periodic paradigm fMRI data; obviously, the Fourier transform would be better suited for this problem.

B. Artificial Event-Related Data

We now apply our method to an artificial data set of event-related fMRI time series. The fMRI time series rely on the parametric model of the hemodynamic response proposed by Glover [11]. The fMRI signal $y(t)$ is zero before the stimulus t_s. For $t > t_s$, the signal is given by

$$a_1(t-t_s)^{d_1} e^{-(t-t_s)/t_1} - 0.4a_2(t-t_s)^{d_2} e^{-(t-t_s)/t_2}. \quad (7)$$

The normalization constants are given by $a_i = \max((t-t_s)^{d_i} e^{-(t-t_s)/t_i})^{-1}$. We take $t_s = 22.5$ s. We consider 32 time samples: $t = [0, 1.5, 3, \ldots, 46.5]$. The parameters d_1, d_2, t_1, and t_2 are combined to control the delay and width of the positive response and the shape of the undershoot. We assume that these parameters are random variables normally distributed: $d_1 \sim N(5, 0.1), d_2 \sim N(12, 0.5), t_1 \sim N(1, 0.2)$, and $t_2 \sim N(0.9, 0.1)$. We have chosen the mean of the random variables to be equal to the values estimated by Glover [11] for a motor response. Fig. 7(a) shows the mean realization of the event-related signals. We generate four different realizations of $y(t)$ according to (7). We add white Gaussian noise to these time series. The variance of the noise is increased for each experiment. Finally, we generate 16 other time series that are realizations of white Gaussian noise. The ratio of the number of

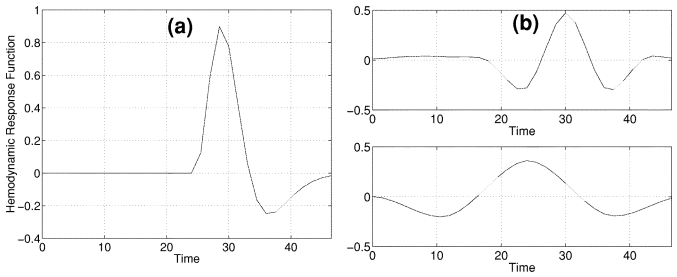

Fig. 7. (a) Mean hemodynamic response time series defined by (7). (b) The best two basis vectors obtained with an SNR $= 0.5$.

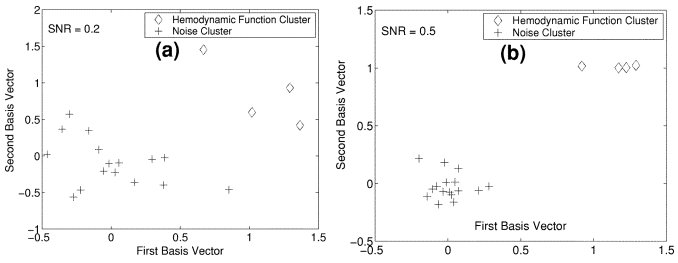

Fig. 8. Scatter plots obtained by projecting all the time series onto the first two best clustering basis vectors for (a) SNR $= 0.2$ and (b) SNR $= 0.5$.

nonactivated time series to the number of activated time series is 4 to 1 in accordance with the fact that in real datasets the number of nonactivated time series is higher than the number of activated time series. Six datasets with average SNR $= 0.1, 0.2, 0.5, 0.8, 1$, and 1.5 are generated. The average power of the hemodynamic function [see Fig. 7(a)] is 0.0678 (in arbitrary units). We apply the best clustering algorithm and obtain the best clustering basis for each dataset. Fig. 7(b) shows the first two best clustering basis vectors $\mathbf{b}_{\gamma_0}$ and $\mathbf{b}_{\gamma_1}$ obtained with the data with SNR $= 0.5$. These two vectors capture 40% of the total wavelet packet coefficient variance. These vectors resemble the hemodynamic response function at different delay times and different positive response widths. Fig. 8 shows the scatter plots obtained by projecting a dataset onto the two best clustering basis vectors for SNR $= 0.2$ and 0.5. The interpretation of the clusters is determined using *a priori* knowledge. Fig. 8 shows that the two clusters are well separated as soon as the SNR becomes greater than 0.2. We now compare the performance of our approach with two other standard methods: the t-test [4] and the correlation analysis [12]. The comparison will be based on the number of true and false positives for each value of the SNR. The true activation rate is the ratio between the number of true positives detected by the algorithm and the total number of true positives. The false activation rate is the ratio between the number of false positives detected by the algorithm and the total number of true negatives. For each value of the SNR, we generate ten independent datasets. We analyze each dataset and compute the average true and false activation rates. We describe in the following the details of each method of analysis.

1) Best Clustering Basis Algorithm: We apply the best clustering algorithm to the data and obtain the reduced feature space by retaining the first T_r vectors $\mathbf{b}_{\gamma_l}$ that capture $r = 40\%$ of the total variance of the wavelet packet coefficients. We then project

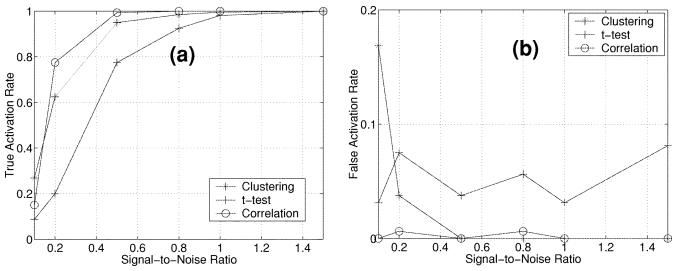

Fig. 9. (a) True activation rate and (b) false activation rate obtained with the best clustering basis algorithm $(*)$, the t-test $(+)$, and the correlation method $(\circ)$.

the time series onto this reduced subspace. Finally, we use the fuzzy C-means clustering algorithm to cluster the projected data into two clusters. The cluster c_a with the centroid with the higher energy will be assigned to be the activated cluster. The membership value μ_{i,c_a} measures the likelihood that the time series $\mathbf{x}_i$ is activated. An activation map can be obtained by thresholding μ_{i,c_a}. In our experiments, we use a threshold equal to 0.8.

2) T-Test: The t-test assumes that each fMRI time series corresponds to the realization of an identically independent stochastic process and divides the data into two groups, obtained during on (post-stimulus) and off (prestimulus) periods. We calculate a t statistic by computing the difference of the sample means of each group normalized by the pooled standard deviation [4]. We then threshold the p-value map at $p = 0.05$.

3) Correlation Analysis: We compute the correlation between the time series and the model (7) using the mean values of the random parameters: $d_1 = 5, d_2 = 12, t_1 = 1$, and $t_2 = 0.9$. The correlation threshold is 0.5. We note that the correlation analysis is equipped with an oracle: it has the perfect knowledge of the (average) true fMRI signal. Because this method represents an ideal situation that cannot be implemented in practice, it provides a challenging gold standard for our algorithm. Fig. 9 shows the average true and false activation rates as a function of the SNR. As expected, the best performance is obtained with the correlation analysis. When the SNR becomes greater than 0.4, the performance of our approach is comparable to the correlation analysis without requiring any knowledge about the hemodynamic response. Indeed, as shown in Fig. 7(b), our method can discover automatically the relevant structure of the hemodynamic response. Both the correlation analysis and the best clustering basis algorithm outperform the t-test by achieving a higher true activation rate and a lower false activation rate.

C. In Vivo Event-Related fMRI Data

We present here the results of experiments conducted with event-related fMRI data. The data, provided by Dr. G. McCarthy (Brain Imaging and Analysis Center, Duke University), demonstrate prefrontal cortex activation in the presence of infrequent events [13]. Visual stimuli were presented to the subjects; most of the images were squares. Infrequent events (targets) consisted in the appearance of circles at random time. Occasionally, images of everyday objects (novels) were also presented. A picture was displayed every 1.5 s. The subject was asked to mentally count the number of occurrences of the circles and

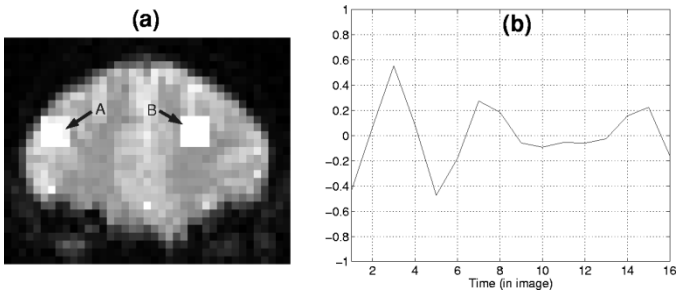

Fig. 10. (a) Two 4×4 windows are placed in region A where we expect activation and in region B where we expect no response related to the stimulus. (b) The most discriminating basis vector for region B.

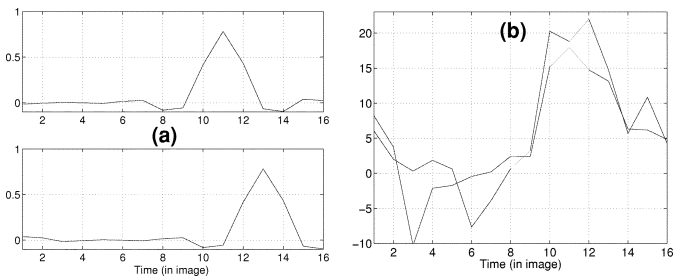

Fig. 11. (a) The two most discriminating basis vectors [$\gamma = (1,0,1), (1,0,2)$]. (b) Time series extracted from the two voxels detected as activated by both the best clustering algorithm and the t-test.

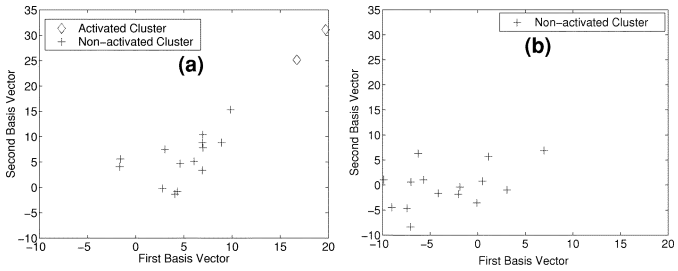

Fig. 12. Scatter plots obtained by projecting the fMRI time series extracted from (a) region A and (b) region B onto the first and the second basis vectors. The class membership is determined using the fuzzy C-mean clustering algorithm with the membership value threshold of 0.8.

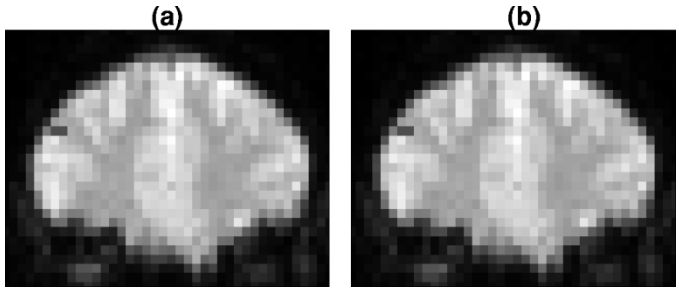

Fig. 13. (a) Activation map obtained by clustering the wavelet packet coefficients of the two basis vectors. The membership value threshold is 0.8. (b) Activation map obtained using the t-test with p-value threshold $= 0.005$.

report that number at the end of each run for a total of ten runs. The experiment was designed to study whether the processes that elicit P300, an event-related potential caused by infrequent target events whose amplitude is dependent on the preceding sequence of stimuli, could also be measured by fMRI [13]. The data was acquired with a gradient echoplanar EPI sequence ($\mathrm{TR} = 1500\,\mathrm{ms}$, $\mathrm{TE} = 45\,\mathrm{ms}$, $\mathrm{NEX} = 1$, $\mathrm{FOV} = 40 \times 20$ cm, slice thickness $= 7$ mm, and imaging matrix 128×64). More details about the experiments are available in [13]. In order to demonstrate the application of our best clustering algorithm to the study of activations by infrequent events, we extract 16 image segments consisting of the eight images preceding and the eight images following each target. We have in each run about five to six targets for a total of about 52 targets. We average the 52 segments in order to increase the SNR. This average signal constitutes the time series $\mathbf{x}_i$. We correct for baseline differences by subtracting from $\mathbf{x}_i$ the prestimulus mean value. Fig. 10 shows the first image of the averaged data. We place a 4×4 window W_{i_0} into two different regions: A, a region where we expect to see activation, and B, a region where we expect no activation. For each position, we compute the best clustering basis based on the 16 time series inside the window. Fig. 11(a) shows the first two basis vectors ($\gamma = (1,0,1)$ and $(1,0,2)$) that capture 40% of the total variance for region A. We note that these two vectors, which were discovered automatically by the algorithm, behave as some delayed hemodynamic responses. Fig. 11(b) shows the first best basis vector for region B. This function has no specific features. A scatter plot obtained by projecting the 16 time series on the two best clustering vectors is shown for each region A and B in Fig. 12. As expected, the scatter plot of the region with no activation (B) is fairly compact, whereas the scatter plot of region A is elongated and shows two well-defined clusters. We confirm this visual impression by measuring the distance between the cluster centroids for both regions. Obviously, the clusters in region B are not meaningful but the distance between the centroids provides a quantitative measurement of the spread of the coefficients. The distance between the centroid is 2.01 for region A and 0.9 for region B. The total variance of the wavelet packets coefficients is 55 for region A and 20.5 for region B. This clearly demonstrates that both the distance between the cluster centroids as well as the total variance within the window W_{i_0} can be used to discriminate between activated and nonactivated regions. The activated time series $\mathbf{x}_i$ were detected by thresholding the membership value μ_{i,c_a} at a level of 0.8. For comparison purposes, we also computed an activation map using the t-test. The p-value threshold was 0.005. The two activation maps shown in Fig. 13 are identical. The two time series detected as activated by the clustering algorithm and by the t-test are shown in Fig. 11(b). They have the shape expected for hemodynamic responses. We also notice that the two best clustering vectors shown in Fig. 11(a) perform as matched filters to detect these time series.

V. Conclusion

We have presented in this paper an algorithm that constructs a clustering basis that best separates activated time series from background noise using a small number of basis functions. Unlike most fMRI data analysis methods, our approach does not require any model of the hemodynamic response or any *a priori* information. Although we presented the local clustering basis

algorithm in the context of a clustering problem, the algorithm can be applied to classification problems when training sets are available. Our approach can also be applied to other classification and clustering problems. We are currently incorporating this algorithm within our local clustering framework [14]. We have shown with several experiments conducted on synthetic data that our approach was capable of finding the basis that best concentrated its discriminatory power on the fewest number of its vectors. The projection of the original data on the first few basis vectors always revealed the organization of the data. We also applied our method to an *in vivo* event-related fMRI dataset. The best clustering basis included a small number of vectors that had the characteristic features of an hemodynamic response. We have also shown that when the spatial window is placed inside a nonactivated region, the best basis could not separate the time series into meaningful clusters. In this case, the basis functions resembled noise and did not possess any relevant structures. We have assumed in all the synthetic experiments that the noise was white and Gaussian. This is clearly not the case for experimental fMRI data. It has been noticed by several authors that data collected under the null-hypothesis condition exhibit the $1/f$ spectrum associated with long memory processes [15], [16]. The colored noise can be uncorrelated by the wavelet transform. It has been shown [17] that the wavelet transform approximates the Karhunen–Loève transform for processes with $1/f$ spectrum. We can, therefore, assume that the noise in the wavelet domain is uncorrelated and Gaussian [18]. An interesting application of this research, suggested by one of the reviewers, concerns the study of networks of connectivity in the brain. The connectivity between regions may be studied with our approach by calculating the temporal correlation between the best clustering basis vectors computed from different regions.

ACKNOWLEDGMENT

The authors would like to thank Dr. G. McCarthy for making the fMRI data available for this work, and the reviewers for their comments and suggestions.

REFERENCES

[1] K. Magnus, T. E. Nichols, J.-B. Poline, and A. P. Holmes, "Statistical limitations in functional neuroimaging—II: Signal detection and statistical inference," *Phil. Trans. Roy. Soc. Lond. B*, vol. 354, pp. 1261–81, 1999.

[2] A. L. Vazquez and D. C. Noll, "Nonlinear aspects of the BOLD response in functional MRI," *Human Brain Mapping*, vol. 7, pp. 108–118, 1998.

[3] X. Golay, S. Kollias, G. Stoll, D. Meier, A. Valavanis, and P. Boesiger, "A new correlation-based fuzzy logic clustering algorithm for fMRI," *Magn. Res. Med.*, vol. 40, pp. 249–260, 1998.

[4] N. Lange and S. L. Zeger, "Non-linear Fourier time series analysis for human brain mapping by functional magnetic resonance imaging," *Appl. Statist.*, vol. 46, no. 1, pp. 1–29, 1997.

[5] P. P. Mitra and B. Pesaran, "Analysis of dynamic brain imaging data," *Biophys. J.*, vol. 76, pp. 691–708, 1999.

[6] E. Bullmore, C. Long, J. Suckling, J. Fadili, G. Calvert, F. Zelaya, T. A. Carpenter, and M. Brammer, "Colored noise and computational inference in neurophysiological (fMRI) time series analysis: Resampling methods in time and wavelet domain," *Human Brain Mapping*, vol. 78, pp. 61–78, 2001.

[7] J. Raz, L. Dickerson, and B. Turetsky, "A wavelet packet model of evoked potentials," *Brain Lang.*, vol. 66, no. 1, pp. 61–88, Jan. 1999.

[8] R. R. Coifman and M. V. Wickerhauser, "Entropy-based algorithms for best basis selection," *IEEE Trans. Inform. Theory*, vol. 38, pp. 713–718, Mar. 1992.

[9] S. Mallat, *A Wavelet Tour of Signal Processing*: Academic Press, 1999.

[10] N. Saito, "Classification of geophysical acoustic waveforms using time-frequency atoms," in *Proc. Ameer. Statist. Assoc. Statist. Computing*, 1996, pp. 322–7.

[11] G. H. Glover, "Deconvolution of impulse response in event-related bold fMRI," *NeuroImage*, no. 9, pp. 416–429, 1999.

[12] K. J. Friston, P. Jezzard, and R. Turner, "Analysis of functional MRI time series," *Human Brain Mapping*, vol. 1, pp. 153–171, 1994.

[13] G. McCarthy, M. Luby, J. Gore, and P. Goldman-Rakic, "Infrequent events transiently activate human prefrontal and parietal cortex as measured by functional MRI," *J. Neurophysiol.*, vol. 77, no. 3, pp. 1630–4, 1997.

[14] J. Chinrungrueng and F. G. Meyer, "Local clustering of functional magnetic resonance images in the frequency space," in *Proc. ISMRM*, 2001, p. 1739.

[15] E. Zarahn, G. K. Aguire, and M. D'Esposito, "Empirical analysis of BOLD fMRI statistics: I. Spatially unsmoothed data collected under null hypothesis conditons," *Neuroimage*, vol. 5, pp. 179–197, 1997.

[16] J. Fadili and E. Bullmore, "Wavelet-generalized least squares: A new BLU estimator of linear regression models with $1/f$ errors," *NeuroImage*, vol. 15, pp. 217–232, 2002.

[17] G. W. Wornell, "A Karhunen-Loève-like expansion for $1/f$ processes via wavelets," *IEEE Trans. Inform. Theory*, vol. 36, pp. 859–61, 1998.

[18] F. G. Meyer, "Wavelet-based estimation of a semi parametric generalized linear model of fMRI time series," *IEEE Trans. Med. Imag.*, vol. 22, 2003, to be published.