Available online at www.sciencedirect.com
SCIENCE DIRECT®
Medical Image Analysis 9 (2005) 51–68
MEDICAL IMAGE ANALYSIS
www.elsevier.com/locate/media

Spatiotemporal clustering of fMRI time series in the spectral domain

Francois G. Meyer *, Jatuporn Chinrungrueng

Department of Electrical Engineering, University of Colorado at Boulder, Boulder, CO 80309, USA

Received 19 July 2002; received in revised form 10 December 2003; accepted 28 July 2004
Available online 15 September 2004

Abstract

We propose a new method for the analysis of functional magnetic resonance images (fMRI). The decision that a voxel v_0 is activated is based not solely on the value of the fMRI signal at v_0, but rather on the comparison of all time series $s_v(t)$ in a small neighborhood $\mathcal{N}(v_0)$ around v_0. Our approach explicitly takes into account the intrinsic spatiotemporal correlations that exist in the data. We focus on experimental designs with periodic stimuli, and therefore we can capture most of the features of the BOLD signal with a low dimensional subspace in the frequency domain. The presence of activated time series can be detected by partitioning the time series in this low dimensional space. Experiments with simulated data, and experimental fMRI data, demonstrate that our approach can outperform standard methods of analysis, such as the t-test.

Keywords: Functional magnetic resonance imaging; Clustering; Spectral analysis; fMRI; Brain mapping

1. Introduction

Functional magnetic resonance imaging (fMRI) is a non-invasive imaging technique that can produce maps of local changes in cerebral blood flow and oxygenation level. Blood oxygenation level-dependent (BOLD) fMRI uses deoxyhemoglobin as a contrast agent: deoxygenated hemoglobin induces a difference in magnetic susceptibility relative to the surroundings. This local modification of susceptibility can be measured, and maps of changes in cerebral venous oxygen concentration can be obtained. Unfortunately changes in the BOLD signal that occur during brain activation are very small (1–5%) and are often contaminated by noise (created by the imaging system hardware or physiological processes). Statistical techniques that handle the stochastic nature of the data are commonly used for the detection of activated voxels (Petersson et al., 1999a,b). Unfortunately, these techniques rely heavily on oversimplified assumptions about the fMRI time series. For instance, it is often assumed that the fMRI time series correspond to the realization of an identically independent stochastic process. In fact, correlations in time and in space do exist (Friston et al., 1993, 1999). The temporal correlation can be taken into account by using a parametric model of the hemodynamic response (Friston et al., 1994). The parameters of such a model can be estimated from the data with a maximum likelihood approach (Petersson et al., 1999a,b). Even though this approach offers a more sophisticated treatment of the temporal correlation, it requires a priori knowledge of the hemodynamic response function, which in many cases is difficult to obtain (Gössl et al., 2000; Baumgartner et al., 1998). In addition, this approach does not yet take into account the existence of the spatial correlation between neighboring voxels, since it detects activation on a voxel per voxel basis. In order to take advantage of existing correlations between time series, an alternative approach considers each time series as a vector in a large dimensional space (the dimension is given by

* Corresponding author.
E-mail address: francois.meyer@colorado.edu (F.G. Meyer).
URL: http://ece-www.colorado.edu/~fmeyer.

1361-8415/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.media.2004.07.002

the number of time samples) and assumes that time series can be grouped together by their structural similarities. Classification and clustering techniques can be used for separating activated voxels from non-activated voxels (Petersson et al., 1999a,b). Classification methods all require that a training set be available. In fMRI analysis it would be very difficult to use previously collected fMRI data sets as a training set since the fMRI response depends on a great number of factors such as image acquisition parameters, paradigm design, subject, and region of the brain activated (Aguirre et al., 1998; McGonigle et al., 2000). To overcome the need for a training set, clustering techniques (Fischer et al., 1997; Baumgartner et al., 1998; Golay et al., 1998; Goutte et al., 1999) have been proposed. The clustering analysis approach partitions time series by optimizing a criterion based on similarity indexes such as the Euclidean distance or a correlation index.

All methods based on clustering that have been proposed in the literature so far suffer from the following problems. First, the clustering analysis usually partitions all the time series from the entire brain and therefore requires some knowledge about the number of clusters (Somorjai et al., 2000). Although experiments are generally designed to activate only one specific region, artifactual activations are common in other regions of the brain and make the "global" clustering analysis unnecessarily complicated. The number of activated voxels is usually a very small fraction of the total number of voxels, and therefore the "activated" cluster will be difficult to detect. Second, most of the clustering techniques perform the clustering in the original high dimensional time domain. The curse of dimensionality (Scott, 1992) forces us to use a number of time series that increase as an exponential function of the number of time samples to achieve good classification error rates. The clustering would be greatly simplified if we could project the time series on a low dimensional subspace wherein we would perform the clustering. Finally, the most widely used similarity indexes are the Euclidean distance and the correlation indexes. These two indexes, as will be shown later in this work, are not adapted to fMRI data, which are very noisy and where correlations between time series exist. In order to address the first problem, we introduce in this work a new framework for "locally" detecting activations in fMRI data using clustering methods. Since it has been shown that the human cerebral cortex can be partitioned into distinct functional areas (Zeki, 1990), we perform the analysis of the fMRI time series locally in space. The second problem, the curse of dimensionality, can be efficiently addressed with an appropriate change of coordinates in such a way that, in the new coordinate system, only a small number of coordinates are required to describe most of the variation in the data. Principle Component Analysis (PCA) (Gabbay et al., 2000) is an example of such methods. Although our framework is not restricted to a certain type of fMRI experiments, we focus on experimental designs with periodic stimuli. It has been shown that the BOLD response to periodic stimuli can be well described by a small number of Fourier coefficients (Bullmore et al., 1993; Lange and Zeger, 1997; Marchini and Ripley, 2000). Replacing a Fourier analysis by a wavelet analysis should allow us to handle non-periodic paradigms within the same framework (Meyer and Chinrungrueng, 2003a). Finally, we address the third problem and propose a new similarity index based on both the Euclidean distance and a correlation index to improve the performance of our clustering method. The original contributions of this work include: (1) a new local clustering framework that integrates spectral estimation and dimension reduction; and (2) a new similarity index based that measures the proximity of time series in the frequency domain. The emphasis of this work has been placed on developing a method which can rely on very little a priori information about fMRI data and is general enough to be applicable to any periodic fMRI experiment. We performed an extensive study of the method in order to evaluate the influence of the parameters of the method on the overall performance.

The paper is organized as follows. In Section 2 we provide a general description of the method. We review in Section 3 spectral estimation techniques and describe how the dimensionality of the problem can be reduced. Section 4 briefly discusses clustering algorithms, presents an fMRI model, and describes a new similarity index. In Section 5 we describe the method for combining local decisions to generate a global activation map. Experiments conducted on artificial data and real fMRI data are presented in Section 6.

2. General description of the method

Fig. 1 shows a block diagram that illustrates the principle of our approach. We consider a group of voxels,

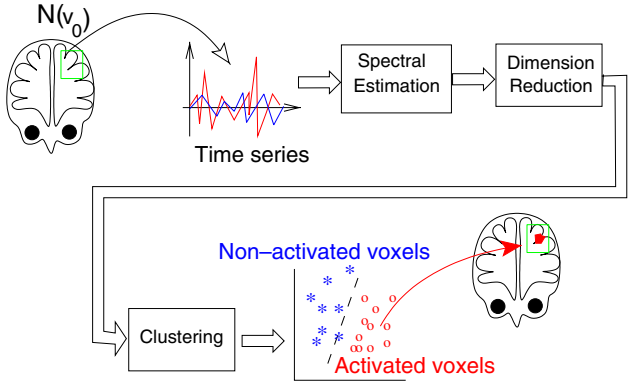

Fig. 1. Principle of the local frequency-based clustering algorithm.

and the corresponding time series, from a small neighborhood $\mathcal{N}(v_0)$ inside the brain. We partition these time series into two clusters. If $\mathcal{N}(v_0)$ 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 neighborhood 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 $\mathcal{N}(v_0)$ is moved throughout the brain, local decision (activation/non-activation) are computed for each voxel $v \in \mathcal{N}(v_0)$. 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 will 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, 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. We expect the energy of the time series to be distributed over only a small number of coefficients in the Fourier domain (Lange and Zeger, 1997; Marchini and Ripley, 2000; Mitra and Pesaran, 1999). One can then reduce the dimensionality of the problem by keeping only those Fourier coefficients that explain most of the variance in the signal. Finally, a clustering algorithm partitions the time series in the reduced frequency space into "activated" and "non-activated" time series.

3. Spectral estimation and dimensionality reduction

The majority of fMRI experiments rely on periodic stimuli. The periodic structure of the stimulus is exploited in many techniques such as the t-test (Petersson et al., 1999a,b) and correlation analysis (Bandettini et al., 1993) to detect activation. However, these methods require that the stimulus time course (or hemodynamic function) and the delay of the response be known (Bandettini et al., 1993; Friston et al., 1994). The existing variation in delay (Müller et al., 2001; Saad et al., 2001) is often overlooked and assumed to be constant over all activated voxels (Bandettini et al., 1993; Friston et al., 1994), yielding results that are not robust. In the spectral (Fourier) domain the delay creates a shift in the phase of the spectrum of the time series, and the magnitude of the time series is left unchanged.

Let $s_v(t)$ be the fMRI signal measured at the voxel v, at time $t = 0, \ldots, T-1$. We assume that $s_v(t)$ is a stationary stochastic process indexed by the position v. The stochastic process is characterized by its power spectrum $f_{s_v s_v}(\omega)$. A number of spectral-based techniques have been proposed to analyze fMRI data (McCarthy et al., 1996; Bullmore et al., 1993; Lange and Zeger, 1997; Marchini and Ripley, 2000; Mitra and Pesaran, 1999).

3.1. Spectral estimation

We consider a tapered estimate of the power spectrum (Brillinger, 2001),

$$I_{s_v s_v}(\omega_k) = \left(2\pi \sum_{t=0}^{T-1} h^2(t)\right)^{-1} \left| \sum_{t=0}^{T-1} h(t) s_v(t) \exp\{-\mathrm{i}\omega_k t\} \right|^2, \tag{1}$$

where $\omega_k = 2\pi k/T$. When h is the Dirichlet taper ($h(t) = 1$), $I_{s_v s_v}$ is the periodogram. The periodogram suffers from being a biased and inconsistent estimate of the power spectrum. The bias can be reduced by using a smoother taper with a faster decay in the Fourier domain. Combining several independent estimates of $I_{s_v s_v}$ obtained with different tapers can lead to a reduction of the variance of the estimator. Thomson (1982) proposed to estimate the power spectrum with several orthogonal tapers h_n that have a Fourier transform with a fast decay. These tapers are constructed using the discrete prolate spheroidal sequences (DPSS) (Thomson, 1982). Each DPSS h_n satisfies the homogeneous integral equation

$$\int_{-B}^{B} \frac{\sin T\pi(\omega - \xi)}{\sin \pi(\omega - \xi)} h_n(\xi)\, \mathrm{d}\xi = \lambda_n h_n(\omega), \tag{2}$$

where $B \in (0,1/2]$ is the bandwidth of the main lobe of the Fourier transform $H_n(\omega)$ of h_n. The eigenvalues λ_n can be ranked in decreasing order, $1 > \lambda_0 > \lambda_1 > \cdots$. Each λ_n measures the fraction of the energy of $H_n(\omega)$ in the main lobe $(-B,B)$. The first 2TB eigenvalues are very close to 1, and therefore most of the energy of $H_n(\omega)$ is localized in the interval $(-B,B)$. Thomson (1982) proposed to estimate the power spectrum by averaging several $I^n_{s_v s_v}(\omega_k)$ computed with the tappers h_n, $n = 0,\ldots,TB-1$. Because the tapers are orthogonal to one another (they are eigenvectors), averaging reduces the variance of the final estimate. This method is called the multi-taper spectral estimation. The variance of the final spectral estimate decreases as more $I^n_{s_v s_v}(\omega_k)$ are included in the average. Unfortunately this comes at a price: one needs to increase B (the spread of $H_n(\omega)$) in order to obtain a larger number of h_n from Eq. (2). As a result each h_n will have a broader Fourier

transform, and each individual power spectrum estimate $I^n_{s_v,s_v}(\omega_k)$ will have a larger variance.

3.2. Reduction of dimensionality

We perform the clustering of the data in the Fourier domain. For periodic stimuli, it is possible to perform the clustering on a smaller subset of frequencies, and therefore reduce significantly the complexity of the clustering algorithm. We use the following experimental observations to choose the set of frequencies that will be used for clustering.

(1) If v is an activated voxel, then the maximum of the power spectrum f_{s_v,s_v} will be at the stimulus frequency ω_s (Bullmore et al., 1993; Lange and Zeger, 1997; Marchini and Ripley, 2000; Mitra and Pesaran, 1999). Fig. 2 (left) illustrates this finding. The figure shows the estimate of the power spectrum for the centroids of the "activated" and "non-activated" time series. The time series were extracted from a small neighborhood in the visual cortex (details about the experiment are provided in Section 6.2.1).
(2) The estimate of the variance of $I_{s_v,s_v}(\omega_s)$ computed over a set of activated voxels v is significantly larger than the same variance computed from a set of non-activated voxels (Baumgartner et al., 1998).

If a neighborhood $\mathcal{N}(v_0)$ contains a mixture of activated and non-activated voxels, then we expect the variance of $I_{s_v,s_v}(\omega_s)$ computed over the group of activated voxels to be larger than the variance computed using the background (non-activated) voxels only. This observation suggests that we should keep only the frequencies with the largest variance. In order to select this subset of frequencies, we estimate the variance of $I_{s_v,s_v}(\omega_k)$ over the N voxels inside the neighborhood $\mathcal{N}(v_0)$,

$$\sigma^2_{v_0}(\omega_k) = \frac{1}{N-1}\sum_{v\in\mathcal{N}(v_0)} (I_{s_v,s_v}(\omega_k) - \overline{I}_{v_0}(\omega_k))^2, \tag{3}$$

where

$$\overline{I}_{v_0}(\omega_k) = \frac{1}{N}\sum_{v\in\mathcal{N}(v_0)} I_{s_v,s_v}(\omega_k)$$

is the average power spectrum estimate in the neighborhood $\mathcal{N}(v_0)$ at the frequency ω_k. We define the frequency indexes k_l, $l = 0,\ldots,T-1$ such that

$$\sigma^2_{v0}(\omega_{k_0}) \geqslant \cdots \geqslant \sigma^2_{v_0}(\omega_{k_l}) \geqslant \sigma^2_{v_0}(\omega_{k_{l+1}}) \geqslant \cdots \sigma^2_{v_0}(\omega_{k_{T-1}}). \tag{4}$$

In order to reduce the dimension of the problem, we keep only the first τ frequencies ω_{k_l}, where τ is chosen in such a way that the relative partial sum of $\sigma^2_{v_0}(\omega_{k_l})$,

$$\frac{\sum_{l=0}^{\tau-1}\sigma^2_{v_0}(\omega_{k_l})}{\sum_{l=0}^{T-1}\sigma^2_{v_0}(\omega_{k_l})}, \tag{5}$$

is larger than a specified threshold $\gamma \in [0,1]$. In the rest of the paper, the set of frequencies indexes $k_0,\ldots,k_{\tau-1}$ is referred to as a the "reduced frequency set." Fig. 2 (right) shows a scatter plot of 75 time series. Our algorithm identified two classes: activated and non-activated time series. We consider the centroid of each class, and construct the axis that passes through the two centroid in the frequency space (see Fig. 3). We then choose a random direction orthogonal to this axis, and define a second axis. Finally, we project all the time series on the plane defined by these two axes (see Fig. 3) to create the scatter plot shown in Fig. 2 (right). The axis defined by the two centroids is in fact very close to the axis defined by the frequency ω_k that is the closest to the stimulus frequency. We observe that the variation of the data along the horizontal axis is much larger than the

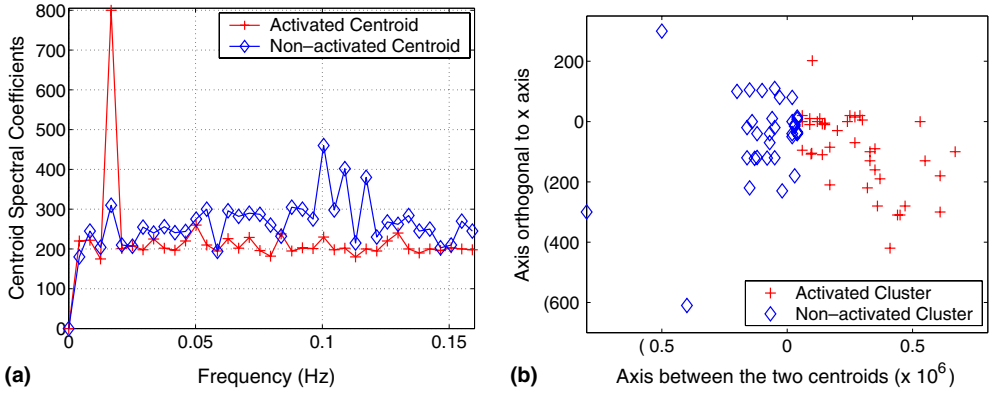

Fig. 2. (a) Spectral coefficients of the two cluster centroids obtained by clustering in the frequency space 75 time series extracted from a $5\times5\times3$ neighborhood located in the visual cortex (see Section 6.2.1 for details about the experiment). The activated centroid shows a sharp peak at the stimulus frequency (1.67×10^{-2} Hz). (b) Scatter plot obtained by projecting all the 75 time series onto the axis between the two centroids and an orthogonal axis. The variation of the data along the x axis is greater than that along the y axis as expected.

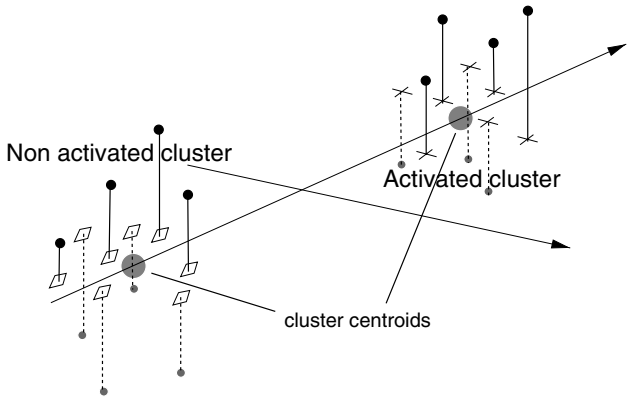

Fig. 3. Projection of the time series in a plane defined by the centroids and a random orthogonal axis.

variation along the vertical axis (note the difference in scale), and therefore the vertical axis is not useful for separating the clusters. This confirms the fact that it is possible to cluster the time series using only a reduced set of frequencies.

4. Local clustering

We consider in this section a small neighborhood $\mathcal{N}(v_0)$ placed around a voxel v_0 inside the brain. Our goal is to partition the N times series $s_v(t)$ from $\mathcal{N}(v_0)$ into two clusters.

We first rapidly review the general principles of clustering algorithms. We then describe a statistical model for the distribution of the estimate of the power spectrum I_{s_v,s_v} for activated and background voxels. This leads us to the definition of an index for measuring similarities between two voxels u and v based on I_{s_u,s_u} and I_{s_v,s_v}. Finally, we explain how we determine the labels "activated" and "non-activated" for each cluster.

A clustering algorithm is a method for partitioning a set of vectors into groups of similar vectors. Given a cluster c, a soft clustering algorithm estimates the likelihood $\mu_c(v)$ that each voxel v belongs to the cluster c. The membership value can be normalized between 0 (v does not belong to the cluster c) and 1 (v belongs to the cluster c). The fuzzy K-means algorithm (Bezdek, 1984) is an example of soft clustering algorithms. When the clustering algorithm is only allowed to take hard decisions ($\mu_c(v) = 0$, or 1), then it is called a hard clustering algorithm. Hierarchical clustering algorithms and the K-means clustering algorithm (Hartigan, 1975) are examples of hard clustering algorithms. A description of the fuzzy K-means and K-means clustering algorithms can be found in Appendix A and B respectively. All clustering algorithms require a similarity index to measure how likely a vector belongs to a particular cluster. We describe in Section 4.1 the probability distribution of the tapered periodogram. Section 4.2 describes the choice of similarity indexes.

4.1. Distribution of the spectral estimate I_{s_v,s_v}

The most common method for acquiring an MR image is the phase encoding method (Liang and Lauterbur, 1999) that effectively samples the two dimensional Fourier transform of the MR image. The measured MR complex signal at a voxel v is obtained by performing an inverse Fourier transform. A reasonable model for the noise of the complex signal (in the Fourier and spatial domain) is an additive zero mean complex Gaussian noise. In this work, we work with the magnitude s_v of the reconstructed complex signal. The distribution of s_v can then be shown to be given by a Rician distribution (Gudbjartsson and Patz, 1995; Kisner et al., 2002).

Our method relies on a spectral analysis of the time series $s_v(t)$, and therefore we need to find the distribution of the estimates of the power spectrum I_{s_v,s_v}. The estimation of the distribution of the tapered periodogram I_{s_v,s_v} is difficult, and in our case we only have access to the asymptotic (when the number of time samples goes to infinity) distribution of I_{s_v,s_v} (Brillinger, 2001).

Theorem 1. *$I_{s_v,s_v}(\omega_k)$ are asymptotically independent random variables distributed with a $f_{s_v,s_v}(\omega_k)\chi^2_2/2$ distribution,*

$$p(x) = \frac{f_{s_v,s_v}(\omega_k)}{4} \mathrm{e}^{-x/2}. \tag{6}$$

Also,

$$\mathrm{EI}_{s_v,s_v}(\omega_k) = f_{s_v,s_v}(\omega_k) \quad \text{and}$$
$$\mathrm{var}(I_{s_v,s_v})(\omega_k) = f^2_{s_v,s_v}(\omega_k). \tag{7}$$

We note that this asymptotic result is very general and requires only very weak assumptions on the stochastic process $s_v(t)$. The sampling property of the tapered periodogram is actually independent of the distribution of $s_v(t)$. The fact that the variance of the tapered periodogram is equal to the power spectrum at the corresponding frequency casts a new light on the experimental observations that we mention in Section 3.2. Indeed, these observations can be interpreted as a property of the estimator I_{s_v,s_v}, and not a property of the data. At the frequency of the stimulus ω_s we expect that $f_{s_v,s_v}(\omega_s)$ will be greater for activated voxels than for background voxels, and consequently the variance of the estimator, given by (7), will be larger for the activated voxels than for the background voxels. This fact was observed experimentally, as explained in Section 3.2.

We propose the following model for the power spectrum of an activated voxel v:

$$f_{s_v,s_v}(\omega_k) = \begin{cases} A_v & \text{if } \omega_k = \omega_s, \text{the stimulus frequency,} \\ \varepsilon_v & \text{otherwise,} \end{cases} \tag{8}$$

where $A_v \gg \varepsilon_v$ are two positive values. Obviously this model does not take into account the presence of large values of f_{s_v,s_v} at multiples of ω_s (harmonics). We have found this assumption to be true in our experiments (see also (Crellin et al., 1999) and (Marchini and Ripley, 2000) for a discussion on the lack of large power spectrum at harmonics of the stimulus frequency).

4.2. Choice of similarity indexes

We first describe in this section two indexes that are frequently used for clustering: the Euclidean distance and the correlation index. We explain their limitations, and propose a new index.

The Euclidean distance index d_E is given by the standard Euclidean distance. This is the optimal index when each cluster corresponds to the realization of a Gaussian random variable with identity covariance matrix (no correlations in the data). The Euclidean distance index is therefore not appropriate for our purpose. In order to handle the possible large variations of the power spectrum estimates at the stimulus frequency, we could use a correlation index d_C defined by

$$d_C(u,v) = g(\rho(u,v)), \tag{9}$$

where g is a continuously and monotonically decreasing function, and ρ is the Pearson's correlation factor computed over the set of reduced frequencies,

$$\rho(u,v) = \frac{\sum_{k=0}^{\tau-1} (I_{s_u,s_u}(\omega_k) - \mathrm{ave}(I_{s_u,s_u}))(I_{s_v,s_v}(\omega_k) - \mathrm{ave}(I_{s_v,s_v}))}{\mathrm{var}(I_{s_u,s_u})\mathrm{var}(I_{s_v,s_v})}, \tag{10}$$

where

$$\mathrm{ave}(I_{s_u,s_u}) = \sum_{j=0}^{\tau-1} I_{s_u,s_u}(\omega_j) \tag{11}$$

and

$$\mathrm{var}(I_{s_u,s_u}) = \sum_{k=0}^{\tau-1} (I_{s_u,s_u}(\omega_k) - \mathrm{ave}(I_{s_u,s_u}))^2. \tag{12}$$

The role of g is to translate a large correlation ($\rho \approx 1$) to a small distance ($d_c \approx 0$). In order to penalize the situations where $\rho = -1$, we use a non-linear function (Golay et al., 1998),

$$g(x) = \left(\frac{1-x}{1+x}\right)^{\beta}, \quad \beta > 0. \tag{13}$$

Because the correlation index measures only the angle between the two vectors I_{s_u,s_u} and I_{s_v,s_v} in $\mathbb{R}^\tau$, the activated cluster will not only contain activated voxels with varying strength, but also noisy voxels. Such noisy voxels v will be included if their power spectrum is maximum at the stimulus frequency ω_s, even if $I_{s_v,s_v}(\omega_s)$ is quite small. This will create false positives.

In order to lower the number of false positives created by the correlation index, we propose a new index that combines the Euclidean distance and the correlation index, and is defined by

$$d_M = d_E^a d_C^{1-a}, \quad \text{with} \quad 0 \leqslant a \leqslant 1. \tag{14}$$

A clustering algorithm equipped with this index will group together time series that have the same frequency contents (large Pearson's index) and similar amplitudes (low Euclidean distance). With this index, weakly activated voxels (larger Euclidean distance) will be separated from the noise only if they are strongly correlated (very small correlation index) with the centroid of the activated cluster. Because we expect the variance of the power spectrum at the stimulus frequency to be large for all activated time series (as explained in Section 4.1), we need to relax the constraint on the Euclidean distance (we penalize less the Euclidean index), hence we choose $a < 1/2$.

4.3. Activated cluster identification

For each position of the neighborhood $\mathcal{N}(v)$, the clustering algorithm provides us with two clusters. We need to decide if one of the two clusters corresponds to an activated cluster. If the neighborhood $\mathcal{N}(v)$ 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. If the neighborhood 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. According to our model of the power spectrum (8), the time series in the activated cluster should exhibit a sharp peak at the stimulus frequency ω_s. We decide that a cluster is an activated cluster if the estimate of the power spectrum of the cluster centroid contains a sharp peak. Because we are not using the knowledge of the stimulus frequency ω_s, we will sometimes detect large pulsating veins as "activated". The idea of detecting a peak in the power spectrum was originally suggested in (Bandettini et al., 1993). A technique based on the similar principle is also proposed in (Marchini and Ripley, 2000). We note that the centroid is a virtual time series (with no physical existence) that represents the average of group of time series. Most Fourier-based techniques in fMRI detect the presence of activation based on the power spectrum of each individual voxel. Working with the power spectrum of the centroid allows us to combat the presence of random noise in the data. We perform the peak detection as follows. For each cluster c, we compute the average of the power spectrum of the centroid I_{s_c,s_c} over the reduced set of frequencies with Eq. (11). We decide that cluster c is a cluster of activated voxels if there exists only one frequency ω_{k_0} such

that $I_{s_c,s_c}(\omega_{k_0}) \geqslant \alpha$ ave(I_{s_c,s_c}). The value of α is chosen experimentally; in our experiments we use $\alpha = 1.5$. This peak detection algorithm is simple and works well in practice.

4.4. What should be the size of $\mathcal{N}(V)$?

The neighborhood $\mathcal{N}(V)$ wherein the clustering is performed should be large enough to yield reliable clustering results. As $\mathcal{N}(V)$ becomes larger (the limiting case being that $\mathcal{N}(V)$ is the entire brain), then the cluster of activated voxels will be harder to detect. On the other hand, if $\mathcal{N}(V)$ is too small it may contain only activated voxels. Because we always split $\mathcal{N}(V)$ into two clusters, the most reliable results will be obtained when the number of activated voxels in $\mathcal{N}(v)$ is roughly equal to the number of background voxels in $\mathcal{N}(V)$. Intuitively, the neighborhood size should be about the same size as the activated region. Indeed, in this case the number of activated voxels will be approximately the same as the number of non-activated voxels as soon as $\mathcal{N}(v)$ overlaps significantly with the activated region (see Fig. 4).

5. Global activation map with local clustering

In the previous section, we described a method to decide if one of the two clusters identified by the clustering algorithm corresponds to an activated cluster or not. We now describe the computation of the "global" membership value using these "local" clustering results. Let $\mathcal{N}(v)$ be a small neighborhood of radius l that contains v_0. We cluster all the time series in $\mathcal{N}(v)$, and test for the presence of an activated cluster. If one of the two cluster is detected as activated, then we consider $\mu_v(v_0,a)$, the membership of v_0 to the activated cluster a. Activated regions should not overlap, and therefore if the same voxel v_0 belongs to an activated cluster a in a neighborhood $\mathcal{N}(v)$, and also belongs to an activated cluster a' in a neighborhood $\mathcal{N}(v')$, then the clusters a and a' should be composed of voxels from the same activated region A.

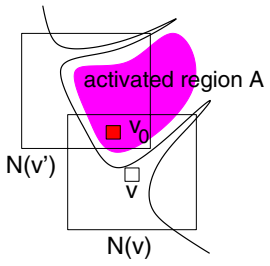

Fig. 4. A voxel v_0 is belongs to the activated region A for several different position of the neighborhood $\mathcal{N}(v)$.

We can define the membership of the voxel v_0 to an activated region A to be

$$\mu_v(v_0, A) = \begin{cases} \mu_v(v_0, a) & \text{if } a \text{ is an activated cluster,} \\ 0 & \text{otherwise.} \end{cases} \tag{15}$$

Now, $\mu_v(v_0,A)$ still depends on the neighborhood $\mathcal{N}(v)$ that contains v_0. If we consider all the neighborhoods $\mathcal{N}(v)$ that contain v_0 (see Fig. 4), we can define the average membership value of v_0 to an activated region A to be

$$\overline{\mu}(v_0, A) = \sum_v K(\|v - v_0\|)\mu_v(v_0, A). \tag{16}$$

K is a kernel that gives more importance to the outcome of the clustering decision when the neighborhood $\mathcal{N}(v)$ is centered around the voxel v_0. In this work, we consider two kernels defined on $[-1,1]$,

- the Epanechnikov kernel, defined by

$$K(x) = \begin{cases} \frac{3}{4}(1 - x^2) & \text{if } |x| \leqslant 1, \\ 0 & \text{otherwise;} \end{cases} \tag{17}$$

- the triweight kernel, defined by

$$\begin{cases} \frac{35}{32}(1 - y^2)^3 & \text{if } |x| \leqslant 1, \\ 0 & \text{otherwise.} \end{cases} \tag{18}$$

As the membership value $\overline{\mu}(v,A)$ gets closer to 1, the probability that v belongs to the activated region A increases. We can threshold the membership maps to construct activation maps. For a given threshold μ_0, we decide that a voxel v is activated if $\overline{\mu}(v,A) \geqslant \mu_0$.

6. Experiments and results

Our method relies on two components: the spectral analysis of the time series, and the clustering of the power spectra. In this section we compare two clustering algorithms and two spectral estimation techniques and analyze their influences on the performance of our algorithm. We use synthetic data in order to have access to the ground truth. We also study the influence of the other parameters for the algorithm: the similarity indexes, the neighborhood size, and the kernels. Finally, we apply our algorithm to the analysis of in vivo fMRI data. We provide comparisons with the t-test for both the artificial data and fMRI data. In order to analyze the performance of our approach, we define the following terms:

True activation rate is the ratio between the number of time series correctly identified as activated and the total number of truly activated time series.

False activation rate is the ratio between the number of time series incorrectly identified as activated and the total number of truly non-activated time series.

In order to study the influence of the key parameters of our method on the overall performance, we generate plots of true and false activation rates (number of true and false positives) at different membership value thresholds. The definition of the false activation rate is similar to the definition of the p-value, which is the probability of identifying a false positive. We can therefore relate the membership value to the p-value through the empirical false positive rate. We can find the membership value that yields a false activation rate that is equal to a given p-value. In general, a membership value larger than 0.8 corresponds to a small p-value ($\sim 10^{-3}$). We derive in Appendix C a non-linear relationship between the membership value and the p-value, when the clustering is based on the Pearson's correlation index. We also plot Receiver Operating Characteristic (ROC) curves. The ROC analysis (Sorenson and Wang, 1996) was developed as a tool to provide standardized quantitatively and statistically meaningful comparisons of detection accuracy between different analysis techniques. The essence of ROC analysis is the comparison of the true activation rates obtained with different analysis techniques for a given false activation rate. The comparison can be performed regardless of the type of rating scales (p-value or membership value) that is used to detect activation in the first place. A plot of true activation rate versus false activation rate for different threshold values of a rating scale is called an ROC curve. Under the assumption that the underlying data for truly positives and truly negative trials form a binormal distribution, the area under the ROC curve can be shown to be the probability that the corresponding analysis technique will correctly identify the true positives (Sorenson and Wang, 1996). In this work it is sufficient to provide a visual comparison of the ROC curves obtained with different analysis techniques, as we do not intend to explore the full capabilities of the ROC analysis. An fMRI study by Constable et al. (1995) also relied on the visual comparisons of ROC curves.

6.1. Experiments with artificial data

6.1.1. Artificial data

We generate a synthetic three-dimensional fMRI dataset in which one small region (≈3 voxels in diameter) and one large region (≈5–6 voxels in diameter) are activated. The dataset contains six slices. Slices 2 and 3, which are shown in Fig. 5, contain significant activation. The MR signal at a voxel v and at time t is the sum of a complex signal $A_v(t)e^{i\theta_v(t)}$ and a complex Gaussian white noise $n^c(t)$,

$$A_v(t)e^{i\theta_v(t)} + n_v^c(t). \tag{19}$$

The amplitude A_v is a sinusoidal function of time,

$$A_v(t) = M_v \sin(\omega_s t + \varphi_v), \quad t = 0, \ldots, T-1. \tag{20}$$

The phase of the MR signal is not used, and we only consider the magnitude

$$s_v(t) = |A_v(t)e^{i\theta_v(t)} + n_v^c(t)|. \tag{21}$$

If v is a non-activated voxel, we have $A_v = 0$. We consider two levels of activation, $M_v = 500$ or $M_v = 510$. The complex Gaussian noise n_v^c is centered with unit variance. If v is activated, then the signal $s_v(t)$ is distributed according to a Rician distribution. If v is non-activated, then the signal $s_v(t)$ is distributed with a Rayleigh distribution (Gudbjartsson and Patz, 1995). Without loss of generality we choose $\theta_v(t) = \pi/4$ (the real and imaginary channels play a symmetric role). We consider $T = 80$ discrete time samples. The frequency of the signal is $\omega_s = \pi/10$. The random delay φ_v in (20) is Gaussian distributed with zero mean and unit variance.

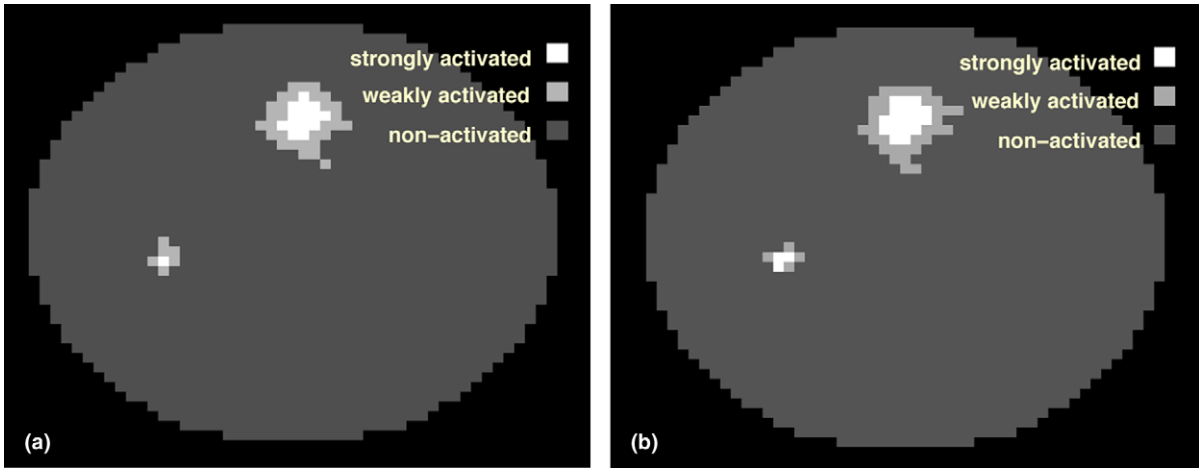

Fig. 5. Slice 2 (a) and slice 3 (b) of the artificial dataset with two levels of activation.

6.1.2. Choice of the clustering algorithms

We first study the influence of the clustering algorithm on the performance of the approach. Specifically, we compare two different types of clustering methods: soft and hard thresholding. We compare the fuzzy K-means clustering algorithm (Bezdek, 1984), a soft clustering method, to the K-means clustering algorithm (Hartigan, 1975), a hard-type clustering algorithm. Both algorithms rely on the Euclidean distance to measure proximity in the frequency space. In the following experiments, we use the periodogram to estimate the power spectrum, and we keep only the frequencies that can explain $\Gamma = 50\%$ of the total variance.

Fig. 6 shows the number of true and false positives as a function of the membership value using the two clustering algorithms. This experiment shows that the fuzzy K-means clustering algorithm outperforms the K-means clustering algorithm in terms of identifying more true positives over a wide range of the membership values, over which false positives obtained with both clustering algorithms are comparatively low (see also the ROC curves in Fig. 7). The number of true positives tends to be lower with the K-means clustering algorithm. We believe that a hard decision (0 and 1) is not adapted to fMRI data. We performed the same experiment using the correlation index and the modified index, and the results (not shown here) are similar to the result obtained with the Euclidean distance. We therefore use the fuzzy K-means clustering algorithm for all experiments presented in the rest of the paper.

6.1.3. Similarity indexes

We then compare the effect of the similarity indexes: the Euclidean distance, the correlation index, and the modified index. Fig. 8 shows the number of true positives and false positives with the three similarity indexes. Fig. 9 shows the ROC curves for the different indexes. As expected, the Euclidean distance yields few true positives with extremely few false positives since it discovers only strongly activated time series, but conservatively identifies weakly activated time series as non-activated. The correlation index can identify these weakly activated voxels as activated. Unfortunately, it also includes non-activated noisy time series that happen to be correlated with the cluster centroid. The effect is an increase in true positives at the expense of a significant increase in false positives. The modified index yields approximately the same number of true positives as the correlation index without significantly increasing the number of false positives. We conclude that the modified index yields the best performance with our artificial data in terms of the number of true positives obtained under the condition that the number of false positives be the same. Hence, we use the fuzzy K-means clustering algorithm with the modified index in the rest of our experiments.

6.1.4. Neighborhood size

Another important parameter of our algorithm is the size of the neighborhood $\mathcal{N}(v)$ (defined by the length l

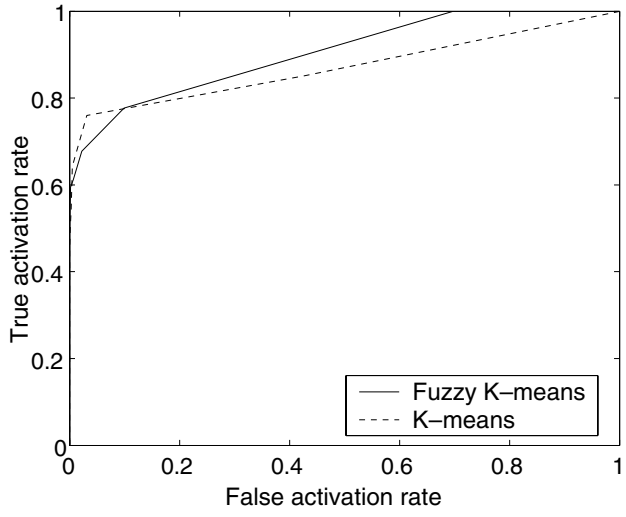

Fig. 7. ROC curves for the comparison of the soft and hard K-means clustering algorithms.

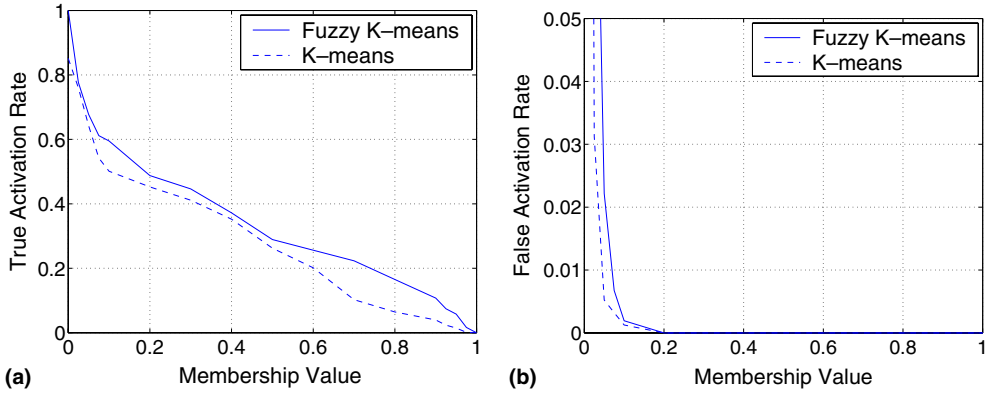

Fig. 6. Number of true positives (a) and false positives (b) using a soft and a hard clustering algorithm.

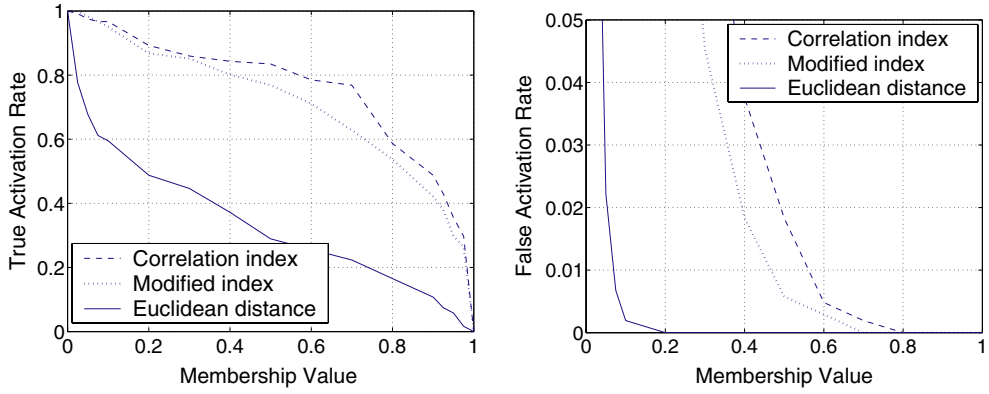

Fig. 8. Number of true positives and false positives obtained with the local frequency clustering algorithm using the fuzzy K-means clustering algorithm with three different similarity indexes.

of each side) within which we perform the local clustering. In this section, we study experimentally the effect of the neighborhood size on the clustering. We evaluate the performance of the local frequency clustering algorithm with three different lengths for the neighborhood $\mathcal{N}(v)$: 4–6 voxels. The true and false activation rates are plotted in Fig. 10 (see also Fig. 11 for the ROC curves). While the true activation rates obtained with these three neighborhoods are comparable to one another, the optimal false activation rate is obtained with the size-5 neighborhood. This optimal neighborhood size coincides with the average size of the activated regions (Fig. 5), as is suggested in Section 4.4.

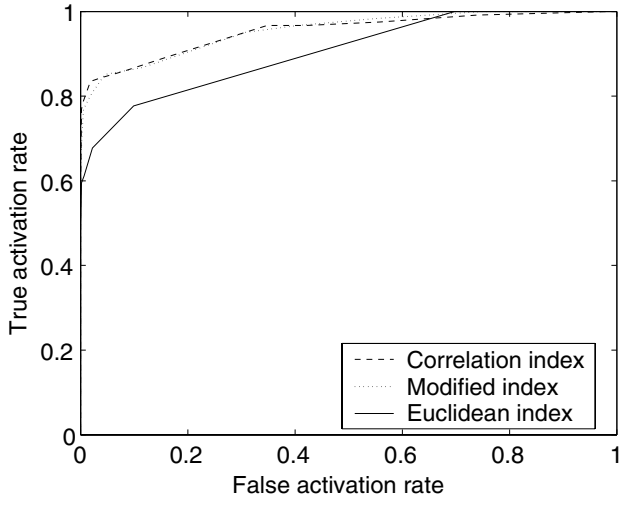

Fig. 9. ROC curves for the comparison of clustering indexes.

6.1.5. Averaging effect and Kernel design

In this section we show that we can significantly decrease the false activation rate by combining local clustering decisions. We applied the local frequency clustering algorithm to the simulated dataset using a neighborhood of size 5. Each voxel v_0 is visited $5^3 = 125$ times, and 125 "local" membership values $\mu_v(v_0, A)$ are computed for each position of the neighborhood $\mathcal{N}(v)$. Fig. 12 shows the true and false activation rates obtained by thresholding the average membership $\overline{\mu}(v_0, A)$ computed with a uniform kernel K using 1 voxel (v_0 only), 25 voxels $v \in \mathcal{N}(v_0)$, ($\mathcal{N}(v_0)$ is in the same

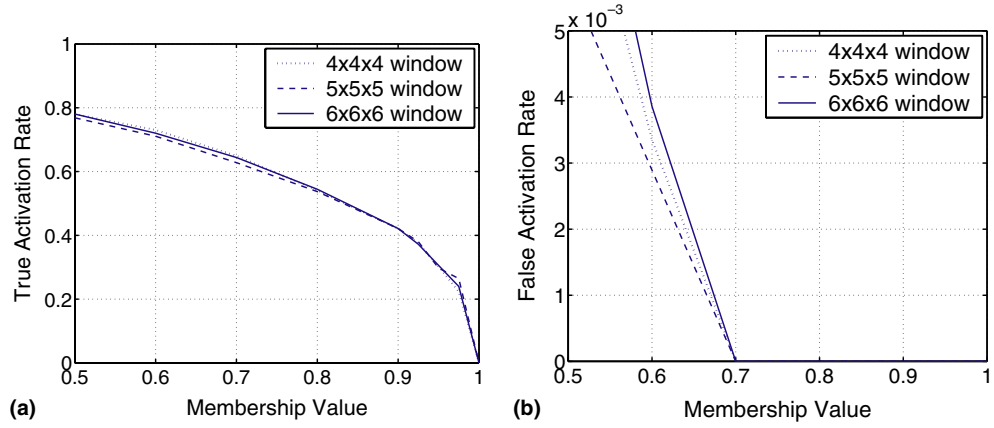

Fig. 10. True and false activation rates obtained with the local clustering algorithm using different window sizes.

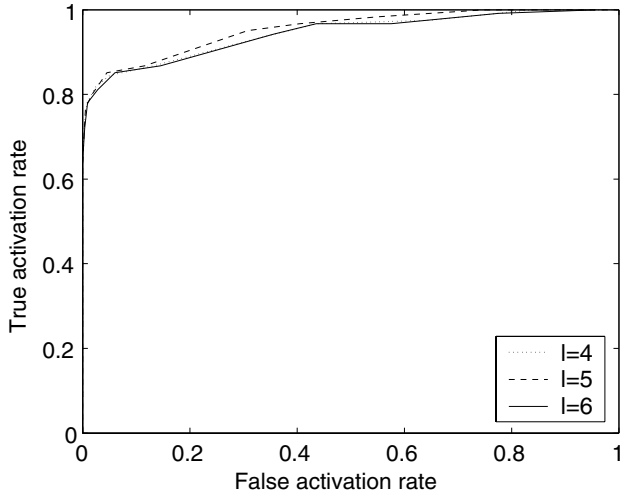

Fig. 11. ROC curves for several neighborhood of size $l \times l \times l$.

slice as v_0); and 125 voxels $v \in \mathcal{N}(v_0)$, ($\mathcal{N}(v_0)$ spans three slices). Fig. 13 shows the same results in the form of ROC curves. As expected, averaging "local" decisions lowers the false activation rate and yields a more reliable "global" decision. However, giving equal importance to all the "local" membership values will significantly lower the true activation rate. Fig. 12 shows that the true activation rate obtained by computing the average of 125 membership values is much lower than that obtained by computing the average of only 25 membership values without significantly improving the false activation rate when the membership value is higher than 0.8.

In the computation of the weighted average $\overline{\mu}(v_0, A)$ (see (16)) we penalize the membership values $\mu_v(v_0, A)$ when the neighborhood $\mathcal{N}(v)$ is far away from the voxel v_0. We compare here three different weighting kernels K: the uniform kernel, the Epanechnikov kernel and the triweight kernel. The discrete versions (we consider only five samples here) of the Epanechnikov and triweight kernels are displayed in Fig. 14.

The Epanechnikov kernel is relatively flatter than the triweight kernel, which is a fast decreasing kernel that gives much higher weights to the voxels in the middle of the neighborhood. Fig. 15 compares the performance of the uniform kernel, Epanechnikov kernel, and triweight kernel. The triweight kernel results in the highest true activation rate with only a marginal increase in the false activation rate for a membership value greater than 0.8 (see also Fig. 16).

6.1.6. Spectral estimation techniques

We now compare two spectral estimation techniques: the periodogram and the multi-taper spectral estimation. The multi-taper was used with $2TB = 2$ and 4 respectively. For such values of the parameter, the estimate of the power spectrum is computed by averaging two and four eigenspectra $I^n_{s_v s_v}$ respectively. The main lobe bandwidth of the DPSS taper with $2TB = 4$ is twice as large as the main lobe bandwidth of the DPSS taper with $2TB = 2$. In this experiment, we used a spatial neighborhood of size $l = 5$ and the uniform kernel. As expected the multi-taper method results in more reliable decisions in terms of lower false activation rates by making a more accurate spectral estimate with lower bias and variance (see Figs. 17 and 18). In this experiment the signal-to-noise ratio is large and the multi-taper approach results in true activation rates lower than with the periodogram approach. The false activation rate is not significantly decreased by the DPSS method. The effect of the DPSS tapers is to blur the original spectrum, without any significant improvement in variance reduction. In general, if the signal-to-noise ratio is high, then the periodogram may be used to achieve a higher true activation rate. However, if the data are very noisy, the multitaper method will yield a lower false activation rate.

6.1.7. Comparison with the t-test

We finally compare the results obtained with the frequency clustering algorithm to the results obtained with

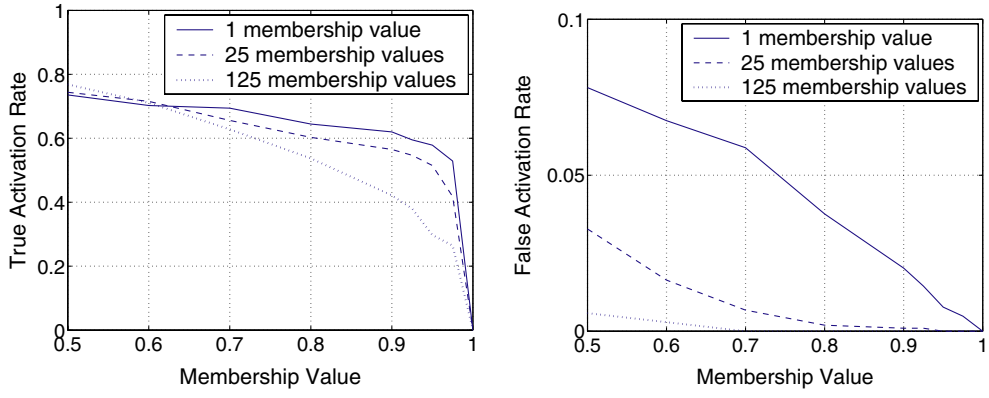

Fig. 12. True and false activation rates obtained by uniformly averaging 1, 25 and 125 local membership values.

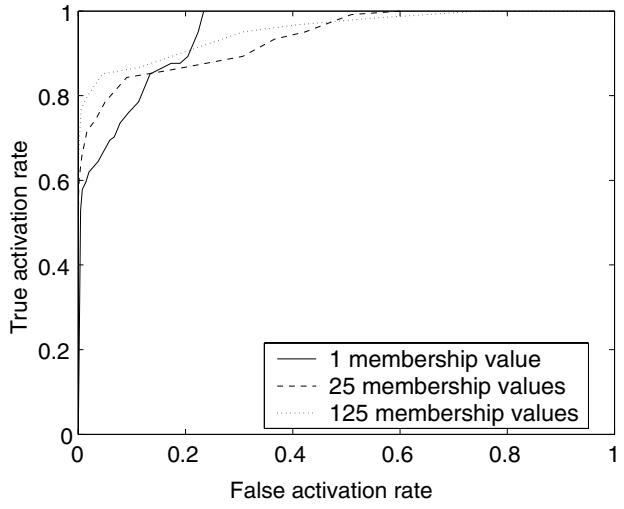

Fig. 13. ROC curves computed from uniformly averaging the membership values of 1, 25 and 125 voxels respectively.

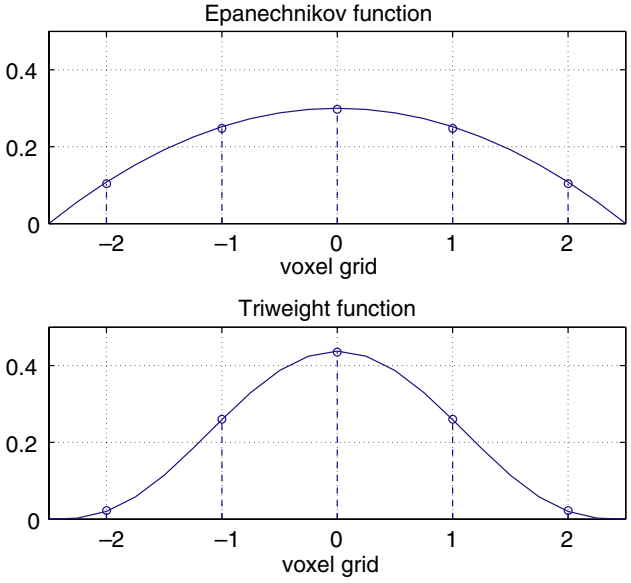

Fig. 14. Discrete kernels sampled at voxel grids.

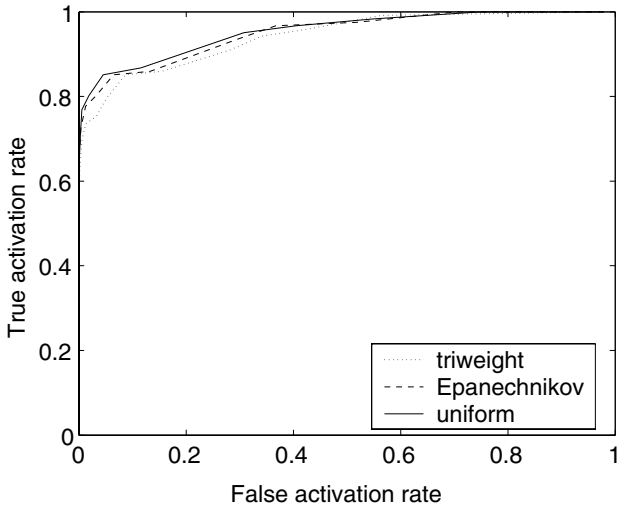

Fig. 16. ROC curves computed with three different kernels.

a two-sided Student's t-test. The t-test assumes that fMRI time series correspond to the realization of an identically independent stochastic process and divides the data into two groups, obtained during on (stimulus) and off (no stimulus) periods. It then computes a t statistic, which follows the t distribution with $T-2$ degree of freedom, by dividing the difference of the sample means of each group by the pooled standard deviation (Petersson et al., 1999a,b). In this experiment we used the following parameters for the local clustering algorithm: a neighborhood of size 5, the triweight kernel, and the power spectrum was estimated with the periodogram. Fig. 19 shows the ROC curves obtained with our frequency clustering algorithm and the t-test applied to the artificial data. The visual comparison reveals that our frequency clustering algorithm outperforms the t-test on the artificial fMRI data by yielding a larger area under the ROC curve.

Fig. 20 shows the activation maps for slice 3 of the artificial data that were generated using the t-test and our local clustering algorithm respectively. The activa-

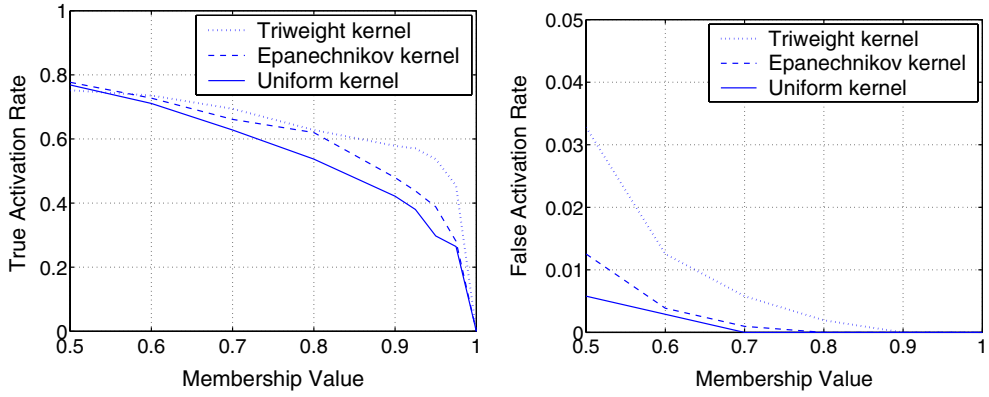

Fig. 15. True and false activation rates obtained with three different kernels.

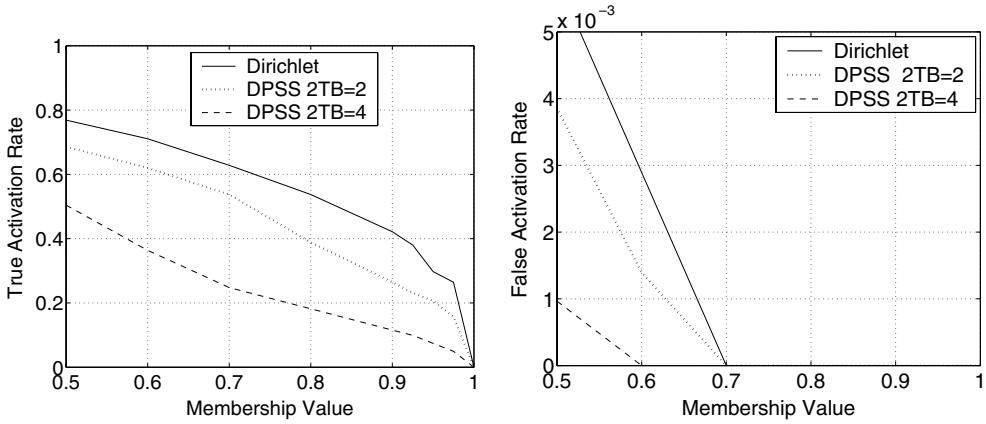

Fig. 17. True and false activation rates obtained with three different tapers.

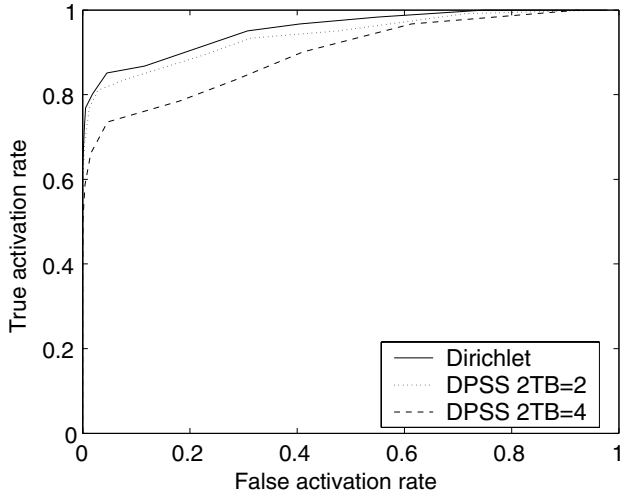

Fig. 18. ROC curves computed with three different tapers.

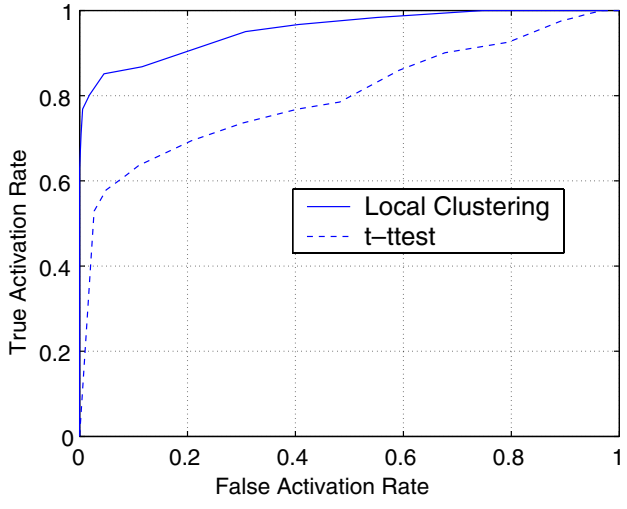

Fig. 19. ROC curves comparing the performance of our local clustering algorithm and the *t*-test applied to the artificial dataset.

tion map obtained with the local clustering algorithm shows that the activated voxels tend to cluster with one another, while with the *t*-test the activated voxels do not form clusters and show more scattered false positives. The numbers of true positives and false positives obtained in slice 3 are summarized in Table 1. Our approach detects more true positives with a lower number of false positives.

6.2. Experiment with fMRI data

We report in this section the results of experiments conducted with two different fMRI data sets: one with visual stimuli and another with auditory stimuli. In these experiments, we apply the local clustering algorithm using the following parameters:

(1) the fuzzy K-means clustering algorithm with the modified index;
(2) the number of features is chosen in order to capture 50% of the variance;
(3) the threshold of the membership value is 0.8.

6.2.1. Experiment with visual stimuli

This dataset (collected by Dr. Jody Tanabe, University of Colorado) demonstrates activation of the visual cortex. A flashing checkerboard image was presented to a subject for 30 s, and a blank image was presented for the next 30 s. This alternating cycle was repeated four times.

Images were acquired with a 1.5 T Siemens MAGNETOM Vision equipped with a standard quadrature head coil and an echoplanar subsystem. Functional images were acquired using a gradient-echo echoplanar sequence (TR = 3,000, FOV = 24×24 cm, slice thickness = 3 mm, imaging matrix = 128×128, voxel size = $1.88 \times 1.88 \times 3$ mm). 80 images were obtained from each axial slice for a total of 12 contiguous slices. The frequency of the stimulus is 1.67×10^{-2} Hz.

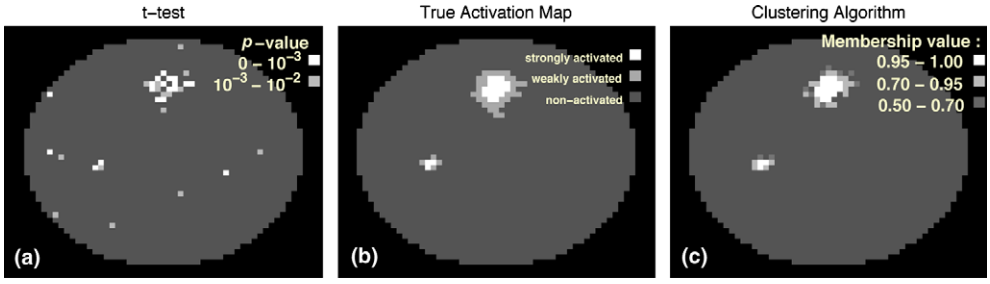

Fig. 20. Activation maps of the artificial data (slice 3) obtained with the *t*-test (a) and the clustering algorithm (c). The true activation map (b) is provided again for a convenient comparison.

Table 1
Number of true positives and false positives obtained by applying the *t*-test and the local clustering algorithm to the artificial data set

	t-Test		Local clustering	
Threshold	0.01	0.001	0.5	0.95
True positives	22	14	44	21
False positives	12	5	5	0

The total number of true positives is 58.

An activation map generated by the local frequency clustering algorithm is shown in Fig. 21 (left). We use a neighborhood of size $5 \times 5 \times 3$ that covers a brain region that is almost isotropic ($9.4 \times 9.4 \times 9$ mm^3). Because a strong activation is obtained, the periodogram is used for the spectral estimation, and the kernel used is the triweight function. The strongest activation was detected in slice 9. We notice that the activated voxels form a cluster in the visual cortex. In Fig. 21 (right) an activation map obtained with the *t*-test threshold at $p = 10^{-3}$ shows scattered false positive voxels inside the brain and in the eyes. The false positives in the eyes detected by the *t*-test do not appear with the local frequency clustering algorithm.

6.2.2. Experiment with auditory stimuli

The second data set was provided by Dr. Gregory McCarthy, Duke University. This data set demonstrates left posterior temporal lobe activation during auditory comprehension (Schlosser et al., 1998). A native English speaker, who could not understand Turkish, was presented with four runs of an alternating series of 28 English and Turkish auditory segments. Two runs started with English (E–T–E–T...), while the other two started with Turkish (T–E–T–E...).

Images were acquired with a 1.5 T General Electric (Milwaukee, WI) Signa scanner equipped with a standard quadrature head coil and an ANMR (Wilmington, MA) echoplanar subsystem. Functional images were acquired using a gradient-echo echoplanar sequence (TR = 1,500, TE = 45, $\alpha = 60°$, NEX = 1, FOV = 40×20 cm, slice thickness = 9 mm, skip = 2 mm, imaging matrix = 128×64, voxel size = $3.2 \times 3.2 \times 9$ mm). The images for each of the seven slices were acquired

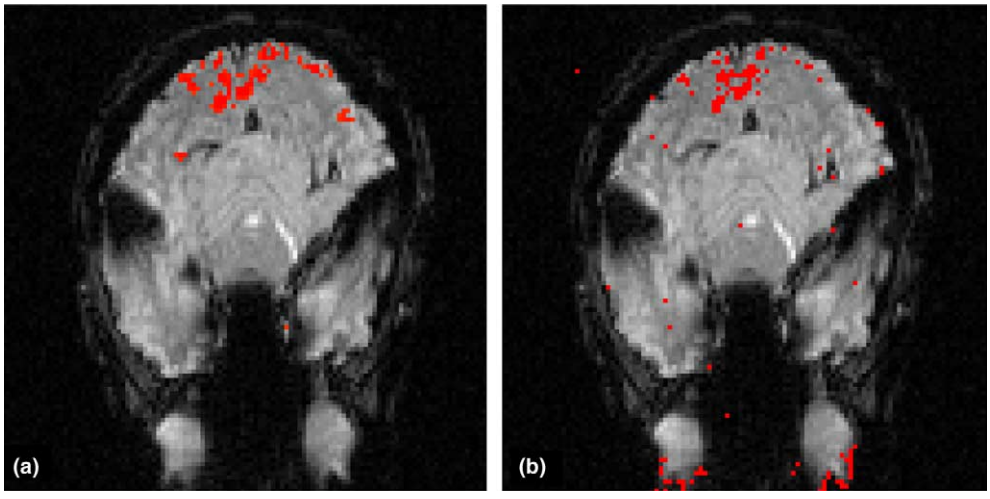

Fig. 21. Visual checkerboard experiment. (a) Activation maps obtained with the local frequency clustering algorithm with a membership value threshold = 0.8. (b) Activation map obtained with the *t*-test with a *p*-value threshold = 10^{-3}.

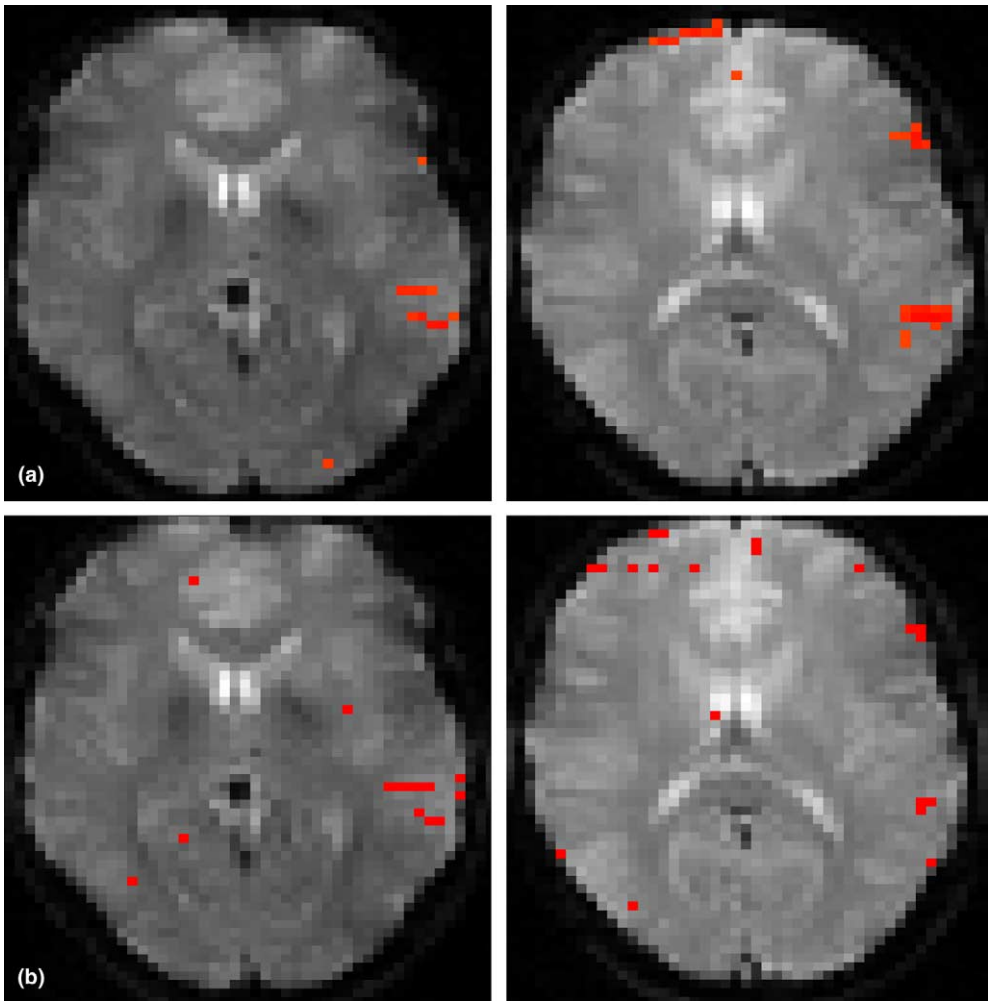

Fig. 22. Auditory comprehension experiment. (a) Activation maps obtained with the local frequency clustering with a membership value threshold = 0.8. (b) Activation maps obtained with the t-test with a p-value threshold = 10^{-3}.

in equally spaced time intervals over the 1.5-s TR in the slice order 1–3–5–7–2–4–6. Each of the four imaging runs consisted of 128 images per slice (196-s scan time) preceded by four radio frequency (RF) excitations to achieve steady-state transverse magnetization. In our analysis, these four-run images were averaged with time-shift correction. The runs started with Turkish were circular shifted so that they started with the second

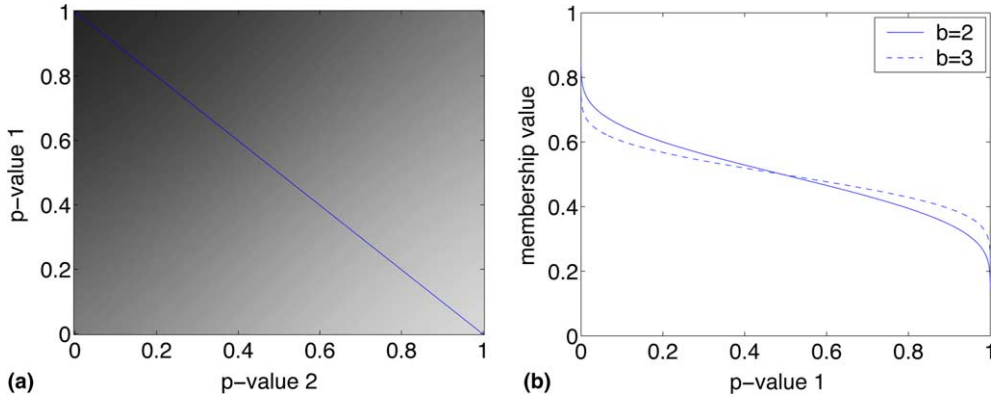

Fig. 23. (a) Membership value as a function of p-value 1 and p-value 2. (b) Membership value as a function of p-value 1 along the diagonal line.

segment (English) and ended with the first segment (Turkish). The frequency of the stimulus is 8.33×10^{-2} Hz.

Activation maps of two slices, obtained with our local frequency clustering algorithm, are shown in Fig. 22 (top). A $3 \times 3 \times 1$ neighborhood was used so that it covers a brain region of size $9.6 \times 9.6 \times 9$ mm^3. Since the signal-to-noise ratio of this data is low, we use the multitaper spectral estimation technique with $2TB = 2$ and apply the Epanechnikov kernel to average more membership values. This choice of parameters should reduce the noise in the data. The activation maps are consistent with the results shown by Schlosser et al. (1998). Significant activations in the left posterior temporal lobe and left inferior frontal lobe, shown in their paper, are found with our method. In Fig. 22 (bottom), activation maps of the same slices obtained with the *t*-test show more scattered noisy voxels within the brain.

7. Conclusion

We have proposed a new method to detect activation in periodic fMRI data. The approach is based on unsupervised classification (clustering) methods and takes into account the spatial correlations existing in fMRI data that conventional statistical methods often fail to exploit. The idea of this approach is to perform several "local" clusterings of fMRI data in a reduced frequency space. These "local" decisions are then combined to yield a more robust "global" decision. We have provided an extensive performance analysis of our algorithm with artificial data sets that were generated according to a simple yet realistic model of fMRI data that accounts for the variation of the response strength among activated voxels. Although we focus on analyzing periodic fMRI data, the approach could be extended to analyze non-periodic fMRI data (event-related fMRI) by replacing the spectral analysis with a wavelet analysis. A wavelet packet, that provides localized features, could be used to obtain an appropriate set of vectors that are well adapted to non-stationary features of event-related fMRI (Meyer and Chinrungrueng, 2003a,b). Our method outperformed the *t*-test on the artificial data sets both in terms of visual appearance and true and false activation rates (ROC analysis). The activation map obtained with in vivo fMRI data showed that activated voxels tend to cluster with one another.

Acknowledgements

JC was supported by a Whitaker Foundation Biomedical Engineering Research Grant.

The authors thank Dr. Jody Tanabe, University of Colorado Health Sciences Center, and Dr. Gregory McCarthy, Director of the Brain Imaging and Analysis Center, Duke University, for making the fMRI data available for this work. The authors thank the reviewers for their comments that helped improve the paper.

Appendix A. Fuzzy K-means Clustering Algorithm (FKM)

The FKM (Bezdek, 1984) is a clustering method that partitions a group of vectors $\{\mathbf{x}_v\}_{i=1}^N$ into C clusters. The algorithm is based on the optimization of the generalized least square function:

$$J_m = \sum_{i=1}^{N} \sum_{c=1}^{C} u^m(\mathbf{x}_v, \mathbf{v}_c) \cdot d^2(\mathbf{x}_v, \mathbf{v}_c) \text{ with } C \leqslant N, \qquad (A.1)$$

where

- $\mathbf{v}_c$ is the centroid vector for cluster c.
- $u(\mathbf{x}_v,\mathbf{v}_c)$ is the membership value of the vector $\mathbf{x}_v$ in the centroid $\mathbf{v}_c$.
- $d(\mathbf{x}_v,\mathbf{v}_c)$ is a similarity index between vectors $\mathbf{x}_v$ and $\mathbf{v}_c$.
- m, $(m \geqslant 1)$ is the fuzzy membership degree.

The fuzzy K-means theorem (Bezdek, 1984) states that this function is minimum if and only if

$$u(i,c) = \frac{1}{\sum_{c'=1}^{C} \left(\frac{d(\mathbf{x}_v,\mathbf{v}_c)}{d(\mathbf{v}_{c'},\mathbf{v}_c)} \right)^{\frac{2}{m-1}}} \qquad (A.2)$$

and

$$\mathbf{v}_c = \frac{\sum_{i=1}^{N} u^m(\mathbf{x}_v, \mathbf{v}_c)\mathbf{x}_v}{\sum_{i=1}^{N} u^m(\mathbf{x}_v, \mathbf{v}_c)}. \qquad (A.3)$$

The algorithm starts with an initial guess for the number of clusters, C, and some random membership values (with the constraint that $\sum_c u(\mathbf{x}_v, \mathbf{v}_c) = 1$). The algorithm then iterates the following steps until $u(\mathbf{x}_v,\mathbf{v}_c)$ or $\mathbf{v}_c$ converges: (1) computing the new centroids, and updating the similarity indexes between each spectral estimate and the centroids; (2) updating the membership values $u(\mathbf{x}_v,\mathbf{v}_c)$ with the new similarity index. (The following parameters are used in this paper: $C = 2$, $m = 1.5$.)

Appendix B. K-means clustering algorithm

K-means clustering algorithm (Hartigan, 1975) is a hard clustering algorithm that is based on minimizing the sum of squared distances from spectral estimate to

its cluster centroid. The algorithm starts with a random partition of vectors $\mathbf{x}$ into C groups and then computes the cluster centroids for this partition. The vectors $\mathbf{x}$ are then reassigned to their nearest cluster centroid, and the cluster centroids are recomputed with this new assignment of the vectors $\mathbf{x}$. The process is repeated until no change can be made in reassigning the vectors $\mathbf{x}$.

Appendix C. Relationship between membership value and *p*-value

A membership value is a number between 0 and 1 specifying how likely a vector belongs to a set and is utilized in fuzzy set theory as the decision-making criterion. This membership value can be arbitrarily assigned without obeying the law of probability theory. As a result, the membership value needs not be related to the p-value in any obvious way. Here, we shall attempt to provide a relationship between these two quantities in order to give the researcher an intuitive idea how to make a proper decision using the membership value. We formulate this relationship by considering the membership value obtained with the fuzzy K-means clustering algorithm using the correlation index.

The probability that a vector $\mathbf{x}_v$ of size T is correlated with any cluster center $\mathbf{v}_c$ can be derived from the Z score (Golay et al., 1998) using the cross-correlation factor $\rho(\mathbf{x}_v,\mathbf{v}_c)$ by the following:

$$Z(\mathbf{x}_v,\mathbf{v}_c)=\frac{\rho(\mathbf{x}_v,\mathbf{v}_c)\sqrt{T-2}}{\sqrt{1-\rho^2(\mathbf{x}_v,\mathbf{v}_c)}}. \tag{C.1}$$

Assume that the correlation index $d(\mathbf{x}_v,\mathbf{v}_c)$ between the vectors $\mathbf{x}_v$ and $\mathbf{v}_c$ is defined by $g(x)=((1-x)/(1+x))^{\beta}$. We can then derive the Z score in terms of $d(\mathbf{x}_v,\mathbf{v}_c)$ as follows:

$$Z(\mathbf{x}_v,\mathbf{v}_c)=\frac{b(1-d(\mathbf{x}_v,\mathbf{v}_c)^{\frac{2}{\beta}})}{d(\mathbf{x}_v,\mathbf{v}_c)^{\frac{1}{\beta}}} \tag{C.2}$$

and

$$d(\mathbf{x}_v,\mathbf{v}_c)^{\frac{1}{\beta}}=\frac{-Z(\mathbf{x}_v,\mathbf{v}_c)+\sqrt{Z(\mathbf{x}_v,\mathbf{v}_c)^2+4b^2}}{2b}, \tag{C.3}$$

where $b=(\sqrt{T-2})/2$. On the other hand, we derive the membership value $u(\mathbf{x}_v,\mathbf{v}_c)$ in terms of the correlation index $d(\mathbf{x}_v,\mathbf{v}_c)$ using (A.2) under the condition that the cost function in (A.1) is minimized. In this work, since we cluster the vectors $\mathbf{x}$ into two clusters, the membership values specifying these vectors in these two clusters can thus be expressed as

$$\begin{aligned} u(\mathbf{x}_v,\mathbf{v}_1)&=\frac{1}{1+\left(\frac{d(\mathbf{x}_v,\mathbf{v}_1)}{d(\mathbf{x}_v,\mathbf{v}_2)}\right)^{\frac{2}{m-1}}} \\ u(\mathbf{x}_v,\mathbf{v}_2)&=1-u(\mathbf{x}_v,\mathbf{v}_1). \end{aligned} \tag{C.4}$$

To keep the comparison simple, we assume that $m-1/2=\beta$. Substituting $d(\mathbf{x}_v,\mathbf{v}_c)$ from (C.3) into (C.4), we obtain the membership value as a function of Z scores as follows:

$$u(\mathbf{x}_v,\mathbf{v}_1)=\frac{1}{1+\frac{-Z(\mathbf{x}_v,\mathbf{v}_1)+\sqrt{Z(\mathbf{x}_v,\mathbf{v}_1)^2+4b^2}}{-Z(\mathbf{x}_v,\mathbf{v}_2)+\sqrt{Z(\mathbf{x}_v,\mathbf{v}_2)^2+4b^2}}}. \tag{C.5}$$

The comparison between (C.2) and (C.4) reveals that the Z score (p-value) is an absolute concept that depends on the correlation between a time series to its representative, whereas the membership value is a function calculated based on the relative correlations between the time series to both cluster centroids. In this sense, the concept of membership value is a non-parametric counterpart of the concept of likelihood ratio test, in which the probability that a time series belongs to each of the two clusters are compared. Fig. 23 (left) shows the membership value obtained numerically using (C.5) as a function of two p-values (Z scores). The membership values extracted along the diagonal line represents the case where two cluster centroids are 180° out of phase and are plotted in Fig. 23 (right) with two different values of the parameter b. The comparison between the membership value and the p-value shows that, if the size of the time series (parameter b) is large enough, membership values larger than 0.8 should yield a result with small p-values ($\sim$0). Although it is non-trivial to show an analytical relationship between the membership value with the Euclidean distance and the p-value, our empirical results with the artificial data sets consistently showed that large membership values correspond to small p-values.

References

Aguirre, G., Zarahn, E., D'Esposito, M., 1998. The variability of human, BOLD hemodynamic responses. NeuroImage 8, 360–369.

Bandettini, P., Jesmanowicz, A., Wong, E., Hyde, J., 1993. Processing strategies for time-course data sets in functional MRI of the human brain. Magnetic Resonance in Medicine 30, 161–173.

Baumgartner, R., Windischberger, C., Moser, E., 1998. Quantification in functional magnetic resonance imaging: fuzzy clustering vs. correlation analysis. Magnetic Resonance Imaging 16, 115–125.

Bezdek, J., Full, W., Ehrlich, R, 1984. FCM: the fuzzy C-means clustering algorithm. Computers & Geosciences 10, 191–203.

Brillinger, D., 2001. Time Series: Data Analysis and Theory. SIAM.

Bullmore, E., Brammer, M., et al., 1993. Statistical methods of estimation and inference for functional MR image analysis. Magnetic Resonance in Medicine 35, 261–277.

Constable, R., Skudlarski, P., Gore, J., 1995. An ROC approach for evaluating functional brain MR imaging and post-processing protocols. Magnetic Resonance in Medicine 34, 57–64.

Crellin, N., Hastie, T., Johnstone, I., July 1999. Statistical models for image sequences. Technical Report 205, Department of Statistics, Stanford University.

Fischer, H., Buechert, M., Henning, J., 1997. Assessing the dynamics of fMRI data using self-organizing map clustering. In: Proceedings of the Fifth SMR Meeting.

Friston, K., Frith, C., Liddle, P., Frackowiak, R., 1993. Functional connectivity: The principal component analysis of large data sets. Journal of Cerebral Blood Flow and Metabolism 13, 5–14.

Friston, K., Jezzard, P., Turner, R., 1994. Analysis of functional MRI time-series. Human Brain Mapping 1, 153–171.

Friston, K., Phillips, J., Chawla, D., Buchel, C., 1999. Revealing interactions among brain systems with nonlinear PCA. Human Brain Mapping 8, 92–97.

Gabbay, M., Brennan, C., Kaplan, E., Sirovich, L., 2000. A principal components-based method for the detection of neuronal activity maps: application to optical imaging. NeuroImage 11, 313–325.

Golay, X., Kollias, S., Stoll, G., Meier, D., Valavanis, A., Boesiger, P., 1998. A new correlation-based fuzzy logic clustering algorithm for fMRI. Magnetic Resonance in Medicine 40, 249–260.

Gössl, C., Auer, D., Fahrmeir, L., 2000. Dynamic models in fMRI. MRM 43, 72–81.

Goutte, C., Toft, P., Rostrup, E., Nielsen, F., Hansen, L., 1999. On clustering times series. NeuroImage 9, 298–310.

Gudbjartsson, H., Patz, S., 1995. The rician distribution of noisy MRI data. Magnetic Resonance in Medicine 34, 910–914.

Hartigan, J., 1975. Clustering Algorithms. John Wiley and Sons.

Kisner, S., Talavage, T., Ulmer, J., 2002. Testing a model for MR imager noise. In: Proceedings of the EMBS/BME Conference, pp. 1086–1087.

Lange, N., Zeger, S., 1997. Non-linear Fourier time series analysis for human brain mapping by functional magnetic resonance imaging. Applied Statististics 46 (1), 1–29.

Liang, Z., Lauterbur, P., 1999. Principles of Magnetic Resonance Imaging. Wiley, IEEE Press.

Marchini, J.L., Ripley, B.D., 2000. A new statistical approach to detecting significant activation in functional MRI. NeuroImage 12, 366–380.

McCarthy, G., Puce, A., Luby, M., Belger, A., Allison, T., 1996. Magnetic resonance imaging studies of functional brain activation: analysis and interpretation. Electroencephalography & Clinical Neurophysiology – Supplement 47, 15–31.

McGonigle, D., Howseman, A., Athwal, B., Friston, K., Frackowiak, R., Holmes, A., 2000. Variability in fMRI: An examination of intersession differences. NeuroImage 11, 708–734.

Meyer, F., Chinrungrueng, J., 2003a. Analysis of event-related fMRI data using best clustering bases. IEEE Transactions on Medical Imaging 22 (8), 933–939.

Meyer, F., Chinrungrueng, J., 2003b. Analysis of event-related fMRI data using best clustering bases. In: Taylor, C., Noble, J. (Eds.), Information Processing in Medical Imaging. Springer-Verlag, Berlin, pp. 623–634.

Mitra, P., Pesaran, B., 1999. Analysis of dynamic brain imaging data. Biophysical Journal 76, 691–708.

Müller, K., Lohmann, G., Bosch, V., Cramon, D., 2001. On multivariate spectral analysis of fMRI time series. NeuroImage 14, 347–356.

Petersson, K.M., Nichols, T., Poline, J.-B., Holmes, A., 1999a. Statistical limitations in functional neuroimaging I. Non-inferential methods and statistical models. Philososphical Transactions of the Royal Society of London B (354), 1240–1260.

Petersson, K.M., Nichols, T., Poline, J.-B., Holmes, A., 1999b. Statistical limitations in functional neuroimaging II. Signal detection and statistical inference. Philososphical Transactions of the Royal Society of London B (354), 1261–1281.

Saad, Z., Ropella, K., Cox, R., DeYoe, E., 2001. Analysis and use of fMRI response delays. Human Brain Mapping 13, 74–93.

Schlosser, M., Aoyagi, N., Fullbright, R., Gore, J., McCarthy, G., 1998. Functional MRI studies of auditory comprehension. Human Brain Mapping 6, 1–13.

Scott, D., 1992. Multivariate Density Estimation. Wiley, New York.

Somorjai, R., Dolenko, B., Baumgartner, R., Jarmasz, M., 2000. Validating clusters in fMRI data, derived by fuzzy clustering analysis: A supervised approach. In: Proceedings of the ISMRM, p. 841.

Sorenson, J., Wang, X., 1996. ROC methods for evaluation of fMRI techniques. MRM 36, 737–744.

Thomson, D., 1982. Spectrum estimation and harmonic analysis. Proceedings of the IEEE 70, 1055–1096.

Zeki, S., 1990. Parallelism and functional specialization in human visual cortex. In: Cold Spring Harbor Symposium Quantitative Biology, vol. 55, pp. 651–661.