MULTISCALE ANALYSIS OF FMRI DATA WITH MIXTURE OF GAUSSIAN DENSITIES

François G. Meyer and Xilin Shen

University of Colorado at Boulder, Department of Electrical Engineering,
E-mail: francois.meyer@colorado.edu.

ABSTRACT

We describe in this work an exploratory analysis of fMRI data. We regard the fMRI dataset as a set of spatiotemporal signals $\mathbf{s_x}$ indexed by their position $\mathbf{x}$. The analysis is performed on the wavelet packet coefficients of the fMRI signals $\mathbf{s_x}$, and we show that we can characterize the coefficients in terms of a mixture model of multivariate Gaussian distributions.

1. INTRODUCTION

Functional Magnetic Resonance Imaging (fMRI) can quantify hemodynamic changes induced by neuronal activity. The goal of the analysis is to detect the "activated" voxels $\mathbf{x}$ where the dynamic changes in the fMRI signal $s_{\mathbf{x}}(t), t = 0, \cdots, T-1$ can be considered to be triggered by the (sensory or cognitive) stimulus. Statistical techniques are commonly used for the detection of activated voxels. Unfortunately, these techniques often rely on oversimplified assumptions. We describe in this work an exploratory analysis of the fMRI data. We regard the fMRI dataset as a set of spatiotemporal signals $\mathbf{s_x}$, indexed by their position $\mathbf{x}$. Our analysis 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 ability to reveal the structure of the dataset. Several studies indicate that one finds dynamic changes of the fMRI signal in time and in frequency [1]. Wavelet packets are time-frequency "atoms" that are localized in time and in frequency. We favor therefore the use of wavelet packets to perform the analysis of fMRI data.

2. MULTISCALE ANALYSIS AND WAVELET PACKETS

A wavelet packet is given by $\psi_{j,k,l}(t) = \psi^k(2^j t - l)$, where

- $j = 0, \ldots,$ J represents the scale : $\psi_{j,k,l}$ has a support of size 2^{-j}. J is the maximum scale ($2^J \leq T$).
- $k = 0, \ldots, 2^j - 1$ 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 $l\, 2^{-j}$.

The library of wavelet packets can be constructed iteratively starting from the scaling function ψ^0, and the wavelet ψ^1. As shown in Fig. 1 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. We define the index

This work was supported by a Whitaker Foundation Biomedical Engineering Research Grant.

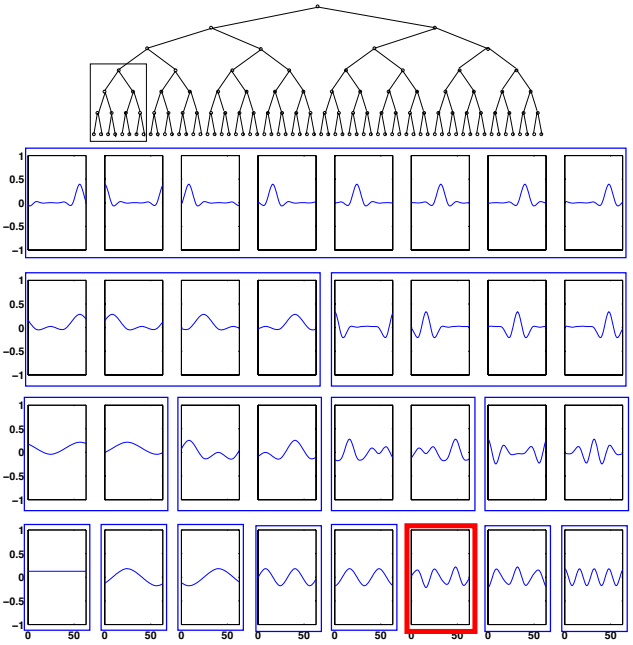

Fig. 1. Top: wavelet packet tree. Bottom : wavelet packet basis functions $\psi_\gamma = \psi_{j,k,l}$ of the subset of nodes $\mathcal{W}_l$. The wavelet packets at each node of the tree are shown on the corresponding row in an enclosed box.

$\gamma = (j,k,l)$, and we consider $\boldsymbol{\psi}_\gamma = \boldsymbol{\psi}_{j,k,l}$ to be a $T \times 1$ vector. The wavelet packet coefficient

$$\alpha_{\mathbf{x}}(\gamma) = \boldsymbol{\psi}_\gamma^T \mathbf{s_x} \tag{1}$$

can be computed with a fast algorithm at each node of the tree. We denote $\mathcal{W}_l$ the set of basis functions from each node of the subtree that is enclosed in a box shown in the top of Fig. 1. These basis functions have lost their temporal localization, but are more precisely located in the frequency domain : they roughly behave as sinusoidal functions oscillating at low frequencies. One wavelet packet, $\boldsymbol{\psi}_{\gamma_0}$ is of particular interest to us (the third from the right on the last row, with a bold frame), because its frequency and phase match the frequency and the phase of the periodic stimulus that will be described in the next section.

3. GLOBAL STATISTICS OF THE COEFFICIENTS

We consider an fMRI dataset that demonstrates activation of the visual cortex [2]. A visual stimulus composed of a flashing checkerboard was presented to a subject for 30s, and a blank image was

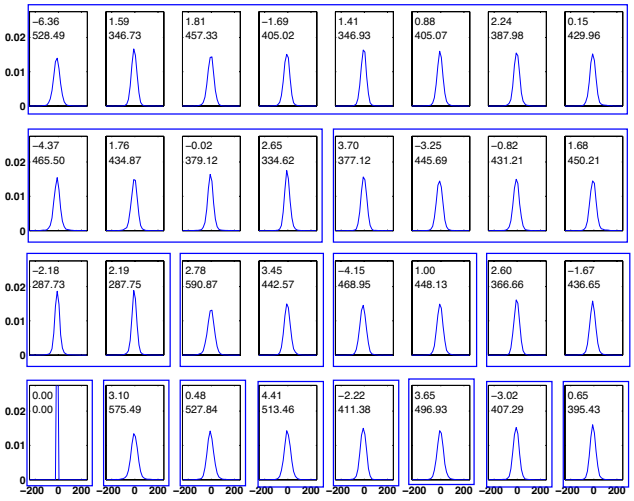

Fig. 2. Histogram of the wavelet packet coefficients (for $\psi_\gamma \in \mathcal{W}_l$) of the background time series. The mean (top) and the variance (bottom) is shown for each distribution.

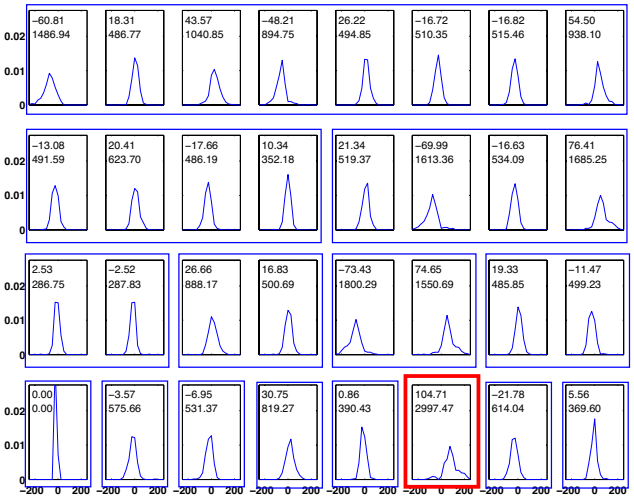

Fig. 3. Histogram of the wavelet packet coefficients (for $\psi_\gamma \in \mathcal{W}_l$) of the activated time series. The mean (top) and the variance (bottom) is shown for each distribution.

presented for the next 30 seconds. Images were acquired every 3s, for a total of 80 images. The voxel size was $1.88 \times 1.88 \times 3$mm, and the image size was 128×128. We used SPM [3] to determine the status : activated/non activated of each time series $\mathbf{s_x}$. The background time series had a p-value greater than 0.1, the activated time series had a p-value smaller than 10^{-3}. We computed the wavelet packet coefficient $\alpha_{\mathbf{x}}(\gamma)$ of each time series $\mathbf{s_x}$. We want to estimate the probability distribution of the coefficients $\alpha_{\mathbf{x}}(\gamma)$, for different γ. We first observe that the stochastic process $\alpha_{\mathbf{x}}$ is not stationary (with regard to shifts in $\mathbf{x}$). Indeed, we know that the activated voxels are not randomly scattered throughout the whole brain (activated voxels tend to be clustered spatially) and therefore the distribution of the coefficients $\alpha_{\mathbf{x}}$ is not translation invariant. A more detailed analysis of the distribution of the coefficients requires to partition the time series into two classes, (1) activated and (2) background (non activated). We can assume that within each class, the distribution of the time series will be the same. We can then use all the values of $\alpha_{\mathbf{x}}$ for all the voxels within each class to estimate the distribution of $\alpha_{\mathbf{x}}$ (ergodicity argument). Fig. 2 shows the histogram of the wavelet packets (for $\psi_\gamma \in \mathcal{W}_l$) of the background time series. The empirical distributions appear to be zero-mean Gaussian distributions, with a standard deviation of about 20. Figure 3 shows the histogram of the wavelet packets (for $\psi_\gamma \in \mathcal{W}_l$) of the activated time series. Because there are much fewer activated time series than background time series (from five to ten percent of the whole brain is activated) the histograms are coarser than the corresponding histograms for the background time series. The stimulus is periodic, and we expect the energy of the activated time series to be distributed over only a small number of coefficients in the Fourier domain [1,4]. Most of the energy of the activated time series will be at the stimulus frequency. Unfortunately, the wavelet packet library does not include true sinusoidal functions. However, as noted in the previous section there exists one basis function, ψ_{γ_0}, with a frequency and a phase that match the frequency and the phase of the periodic stimulus. Most distributions have a small mean. The distribution with the largest mean is obtained with the wavelet packet whose frequency match exactly the frequency of the stimulus. In fact, a close inspection of

Fig. 3 reveals that other nodes also have a large mean (albeit relatively smaller). As shown in Fig. 1 the wavelet packets at these nodes also oscillate at the frequency of the stimulus, but have a smaller support (they have local oscillations), and thus contribute to a smaller correlation with $\mathbf{s_x}$. The variance at these nodes is much larger (1500-3000) than at the other nodes (400). Also, the distribution at these nodes appears to be skewed and less Gaussian than at the other nodes. The empirical distributions appear to be zero-mean Gaussian distributions at all other nodes. As indicated in [4], there is very little energy at harmonics of the stimulus frequency, and therefore the histograms of the wavelet packet coefficients for the other nodes of the tree (not shown here) also have a zero mean. We finally considered a three dimensional neighborhood $\mathcal{N}(\mathbf{x}_0)$ that was placed around an activated region and included a mixture of activated and background time series.

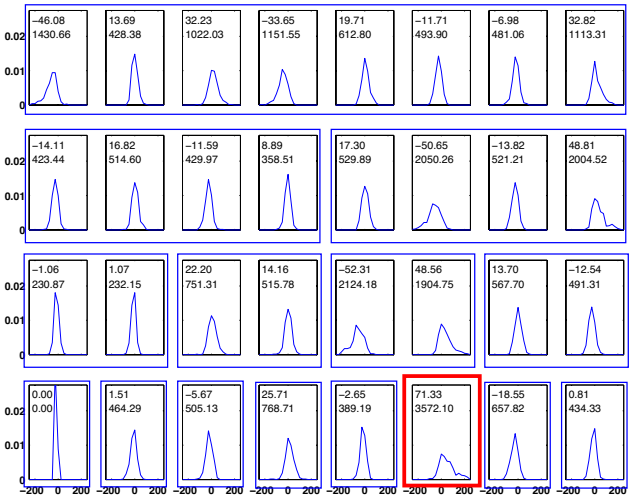

Fig. 4. Histogram of the wavelet packet coefficients (for $\psi_\gamma \in \mathcal{W}_l$) of the mixture time series. The mean (top) and the variance (bottom) is shown for each distribution.

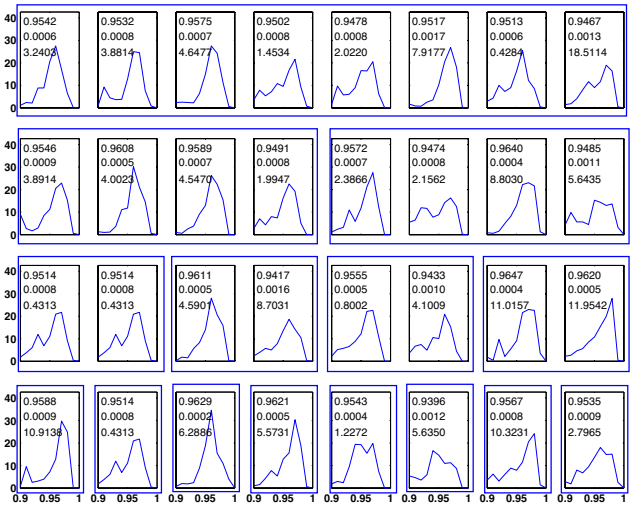

Fig. 5. Histogram of the W statistic computed from activated time series only. The mean (top), the variance (center), and the χ^2 score (bottom) are displayed for each distribution.

Fig. 6. Histogram of the W statistic computed from a mixture of activated and background time series. The mean (top), the variance (center), and the χ^2 score (bottom) are displayed for each distribution.

Figure 4 shows the histogram of the coefficients (for $\psi_\gamma \in \mathcal{W}_l$) of all the time series in $\mathcal{N}(\mathbf{x}_0)$. Most projections appear to be Gaussian with a small mean and a variance around 400. The same wavelet packet ψ_{γ_0} gives rise to a multimodal distribution that is clearly non Gaussian.

4. LOCAL STATISTICS OF THE COEFFICIENTS

In the previous section we studied the empirical distribution of the wavelet packet coefficients computed over the entire brain. While we expect the background signal to be constant (except for a possible slowly varying drift), the strength of the signal $\mathbf{s_x}$ may vary from one activated voxel $\mathbf{x}$ to another. We can interpret the empirical distribution of $\alpha_{\mathbf{x}}(\gamma_0)$ in Fig. 4 as follows : the activated region is composed of a small number of subregions wherein $\alpha_{\mathbf{x}}(\gamma)$ is Gaussian distributed ; and the mean is different for each subregion. The distribution of $\alpha_{\mathbf{x}}(\gamma_0)$ can be thus described by a mixture of Gaussian distributions. We can test this hypothesis by performing a local analysis of the distribution of $\alpha_{\mathbf{x}}(\gamma)$. For each position of the neighborhood $\mathcal{N}(\mathbf{x}_0)$ we test the hypothesis that the distribution of the coefficients $\{\alpha_{\mathbf{x}}(\gamma), \mathbf{x} \in \mathcal{N}(\mathbf{x}_0)\}$ is Gaussian. Several test for normality exist, and we use the W Shapiro-Wilk test [5] because of its ability to detect non Gaussian distributions. As $\mathbf{x}_0$ is moved throughout the brain we collect many samples for the W statistic. We can compute the empirical distribution of W (for all values of $\mathbf{x}_0$), as is shown in Fig. 5 and Fig. 6. In order to quantify non-normality, we compare this empirical distribution with the distribution of W obtained under the normality assumption. The comparison is performed using the χ^2 distance. A large χ^2 score indicates that the distribution of the wavelet packet coefficient is non Gaussian. Fig. 5 shows the empirical distribution of the W statistic computed locally from activated time series only (we use $\alpha_{\mathbf{x}}$ in the computation of W only if $\mathbf{x}$ is activated). Because all the χ^2 distances are small, the hypothesis that $\alpha_{\mathbf{x}}$ is Gaussian can be accepted.

A similar experiment was conducted, and the results (not shown here) indicate that the distribution of $\alpha_{\mathbf{x}}(\gamma)$ is Gaussian for all γ if $\mathbf{s_x}$ is a background time-series. Fig. 6 shows the empirical distribution of the W statistic computed locally when $\mathcal{N}(\mathbf{x}_0)$ contains a part of an activated region as well as background voxels. While almost all distributions appear to be Gaussian (small χ^2), there exists a few non Gaussian distributions. The most non-Gaussian distribution is obtained again for the wavelet packet whose frequency match exactly the frequency of the stimulus.

5. MIXTURE OF GAUSSIAN DENSITIES

Our experimental results indicate that it is possible to find a small number of interesting projections that reveal the presence of activated time series. Because most one dimensional projections of high dimensional data are approximatively Gaussian [6], bimodal or more generally non Gaussian, distributions of the wavelet packet coefficients are likely to reveal the presence of non background time series. A natural model for the joint distribution of the wavelet packet coefficients $\boldsymbol{\alpha}_{\mathbf{x}} = [\alpha_{\mathbf{x}}(\gamma_1), \cdots, \alpha_{\mathbf{x}}(\gamma_T)]^T$ is a finite mixture of multivariate Gaussian densities,

$$p(\boldsymbol{\alpha}) = \sum_{m=1}^{M} \pi_m \phi(\boldsymbol{\alpha}, \mu_m, \Sigma_m). \tag{2}$$

The mixing parameters are positive weights that add up to 1. The density ϕ is the normal density function. This assumption is in perfect agreement with our experimental findings. Indeed, since the marginals of a mixture of multivariate Gaussian densities are mixture of Gaussian densities, we expect the distribution of the $\alpha_{\mathbf{x}}(\gamma)$ to be mixtures of Gaussian densities. This is exactly what we observed in our experiments. We note that the model (2) is global, and we need more than one activated component in order to describe different levels of activation. This model can be interpreted in the temporal domain. The signal $\mathbf{s_x}$ can be decomposed as follows

$$s_{\mathbf{x}}(t) = \theta_{\mathbf{x}}(t) + a_{\mathbf{x}}(t) + \nu_{\mathbf{x}}(t) \tag{3}$$

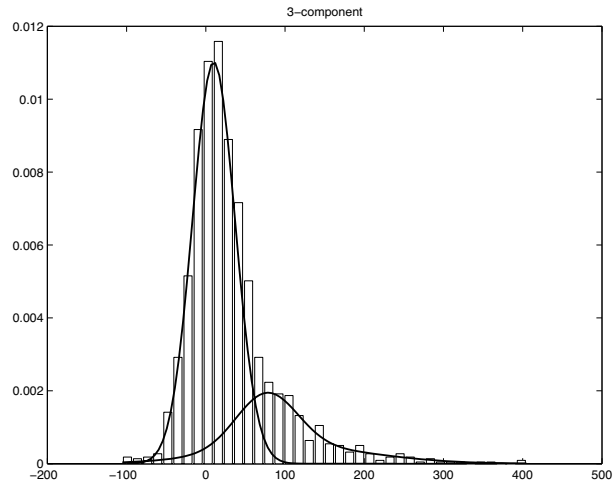

Fig. 7. Mixture model with three components superimposed on the empirical distribution of the wavelet packets.

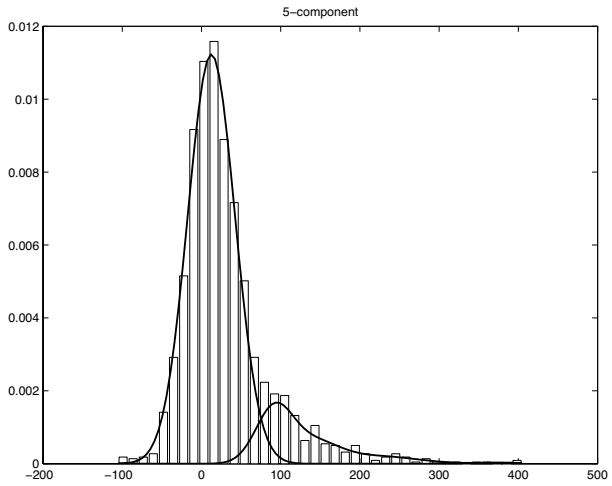

Fig. 8. Mixture model with five components superimposed on the empirical distribution of the wavelet packets.

where $\theta_{\mathbf{x}}(t)$ is a baseline drift, and $a_{\mathbf{x}}(t)$ is the stimulus-induced response. The noise $\nu_{\mathbf{x}}(t)$ is correlated with a $1/f$ spectral behavior associated with long memory processes [7–9]. The wavelet coefficients of the drift $\theta(t)$ correspond to low frequencies, and should be zero for $\gamma = \gamma_0$. The wavelet coefficients of the noise will be uncorrelated [8, 9], and will be Gaussian distributed. The noise coefficients will contribute to the zero-mean component of the mixture. The wavelet coefficients of the activated component will be non zero for $\gamma = \gamma_0$, and Gaussian distributed. The mean of the distribution will vary with the strength of the activation at the voxel $\mathbf{x}$. Voxels with similar activation strength will contribute to a common component in the mixture. The maximum likelihood estimates of μ_m and Σ_m given the observations can be computed with the Expectation Minimization (EM) algorithm. Figures 7 and 8 show the mixture model superimposed on the empirical distribution of the wavelet packets $\alpha_{\mathbf{x}}(\gamma_0)$, where $\mathbf{x}$ are taken in a neighborhood $\mathcal{N}(\mathbf{x}_0)$ that overlap with an activated region. In order to capture the tail on the right side of the distribution we used several activated components. Tables 1 and 2 show the mixture parameters for all the components.

m	1	2	3
π_m	0.74	0.15	0.11
μ_m	10	77	114
σ_m	27	38	100

Table 1. Parameters of the mixture for a 3 component model.

m	1	2	3	4	5
π_m	0.84	0.08	0.05	0.02	0.002
μ_m	13	91	139	224	377
σ_m	30	23	33	48	26

Table 2. Parameters of the mixture for a 5 component model.

6. REFERENCES

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

[2] F.G. Meyer and J. Chinrungrueng, "Spatiotemporal clustering of fMRI time series in the spectral domain," Accepted for publication in *Medical Image Analysis*, 2003.

[3] K.J. Friston, A.P. Holmes, K.J. Worsley, J.P. Poline, C.D. Frith, and R.S.J. Frackowiak, "Statistical parametric maps in functional imaging: A general linear approach," *Human Brain Mapping*, vol. 2, pp. 189–210, 1995.

[4] J. L. Marchini and B. D. Ripley, "A new statistical approach to detecting significant activationin functional MRI," *NeuroImage*, vol. 12, pp. 366–380, 2000.

[5] S.S. Shapiro and M.B. Wilk, "An analysis of variance test for normality (complete samples)," *Biometrika*, vol. 52, pp. 591–611, 1965.

[6] P. Diaconis and D. Freedman, "Asymptotics of graphical projection pursuit," *Annals of Statistics*, vol. 12, no. 3, pp. 793–815, 1984.

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

[8] 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.

[9] F.G. Meyer and J. Chinrungrueng, "Analysis of event-related fMRI data using best clustering bases," *IEEE Transactions on medical imaging*, vol. 22(8), pp. 933–939, 2003.