CLUSTERING OF SPATIOTEMPORAL SIGNALS : APPLICATION TO THE ANALYSIS OF FMRI DATA

François G. Meyer, and Jatuporn Chinrungrueng

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

ABSTRACT

We explore a new paradigm for the analysis of event-related functional magnetic resonance images (fMRI). These datasets are large collections of spatiotemporal time series that are indexed by their spatial locations. The analysis of such datasets requires to partition the spatial domain into regions within which the time series are similar. We construct basis functions on which the projection of the fMRI time series reveals the organization of the fMRI dataset into "activated", and "non-activated" clusters. These basis functions maximally separate the time series into meaningful groups according to their time-scale behavior.

1. INTRODUCTION

Functional Magnetic Resonance Imaging (fMRI) can quantify hemodynamic changes induced by neuronal activity. FMRI studies can be performed with an event-related stimulus design. This paradigm consists in acquiring images while a subject is submitted to random (sensory or cognitive) stimulations. 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 stimulus. The analysis of event-related fMRI data commonly relies on the assumption that the fMRI signal $s_{\mathbf{x}}(t)$ can be modeled by the convolution between the stimulus waveform with some unknown hemodynamic response function [1]. Unfortunately, it has become clear that there is a nonlinear relationship between the fMRI signal and the stimulus time series [2]. In the absence of any detailed substantive understanding of the mechanism of the fMRI response, we advocate a non parametric data-driven approach that can better account for all important features present in the data. We regard the fMRI dataset as a set of spatiotemporal signals $s_{\mathbf{x}}(t)$, indexed by their position $\mathbf{x}$. The decision that a voxel $\mathbf{x}_0$ is activated is based not solely on the signal $s_{\mathbf{x}_0}(t)$, but rather on the comparison of all time series $s_{\mathbf{x}}(t)$ in a small neighborhood $W_{\mathbf{x}_0}$ around $\mathbf{x}_0$. The principle of the analysis is illustrated in Fig. 1. We consider the times series from the window $W_{\mathbf{x}_0}$ and we partition these time series into two clusters. If $W_{\mathbf{x}_0}$ 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 $W_{\mathbf{x}_0}$ is moved throughout the brain, local decision (activation/non activation) are computed for each voxel $\mathbf{x} \in W_{\mathbf{x}_0}$. This principle exploits the intrinsic spatial correlation that is present in the data : truly activated voxels tend to be spatially clustered, while falsely activated

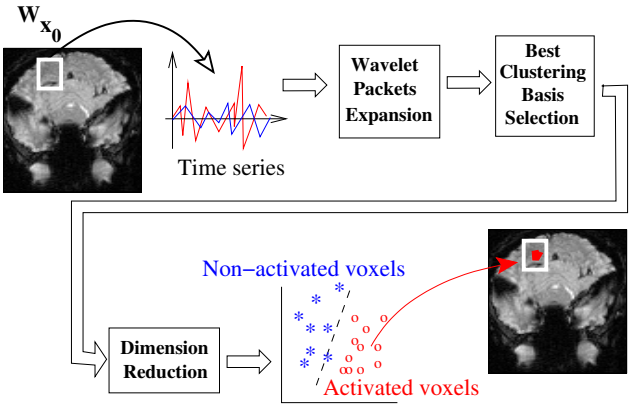

Fig. 1. Principle of the analysis of the fMRI time series $s_{\mathbf{x}}(t)$.

voxels will tend to be scattered. These local decisions are combined to generate a global activation map. Unlike global clustering methods [1], this local clustering approach only partitions the time series inside a small region of the brain. Furthermore, the clustering of the time series is not performed directly on the raw fMRI signal. Instead, the time series are projected on a set of basis functions conveniently chosen for their discriminating power. Several studies indicate that one finds dynamic changes of the fMRI signal in time, frequency, and in space. We favor therefore the use of wavelet packets that can efficiently represent the non-stationary structures present in event-related data. Instead of using a fixed basis we select from a very large and highly redundant dictionary of wavelet packets those wavelet packets on which the projection of event-related fMRI data reveals the organization of the time-series into "activated", and "non-activated" clusters. The construction of the "clustering basis functions" described in this work is applicable to a large category of problems where time series are indexed by a spatial variable.

2. WAVELET PACKETS, BEST-BASIS

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 ($J \leq J_0$).
- $k = 0, \ldots, 2^j - 1$ 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 $l\, 2^{-j}$.

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. 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. We define the index $\gamma = (j, k, l)$, and we consider $\boldsymbol{\psi}_\gamma = \boldsymbol{\psi}_{j,k,l}$ to be a $2^{J_0} \times 1$ vector. We define the $2^{J_0} \times 2^{J_0-j}$ matrix

$$\boldsymbol{\Psi}_{j,k} = \left[\boldsymbol{\psi}_{j,k,0} | \cdots | \boldsymbol{\psi}_{j,k,2^{J_0-j}-1}\right] \quad (1)$$

to be the collection of wavelet packets at the node (j, k) stacked together. The wavelet packet coefficients $\alpha(j, k, l)$ of the vector $\mathbf{s_x} = [s_\mathbf{x}(0), \cdots, s_\mathbf{x}(T-1)]^t$ at the node (j, k) are given by the projections of $\mathbf{s_x}$ on the 2^{J_0-j} wavelet packets stacked inside the matrix $\boldsymbol{\Psi}$. In matrix form we have :

$$\boldsymbol{\alpha}_\mathbf{x}(j, k) = \left[\alpha(j, k, 0), \cdots \alpha(j, k, 2^{J_0-j} - 1)\right]^t = \boldsymbol{\Psi}^t_{j,k}\mathbf{s_x} \quad (2)$$

Let us associate the dyadic frequency interval $[2^j k, 2^j(k+1))$ to the wavelet packet node $\boldsymbol{\Psi}_{j,k}$. If a collection of intervals $[2^j k, 2^j(k+1))$ provides a cover of the time-frequency plane, then the corresponding set of $\boldsymbol{\Psi}_{j,k}$ forms an orthonormal basis [3]. Coifman and Wickerhauser [3] suggested to use a dynamic programming algorithm with order $T \log(T)$ to explore the tree shown in Fig. 2, and 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 [4]. In this work we 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 non activated signals.

3. WHAT ARE THE INTERESTING PROJECTIONS ?

We provide in this section some experimental evidence that justify the principle of our approach. We consider an fMRI dataset that demonstrates activation of the visual cortex [5]. A visual stimulus composed of a flashing checkerboard was presented to a subject for 30s, and a blank image was presented for the next 30 seconds. Images were acquired every 3s, for a total of 80 images. We extracted the time series $s_\mathbf{x}(t)$ from all the voxels $\mathbf{x}$ inside a small window $W_{\mathbf{x}_0}$. We used SPM [6], a gold-standard method of analysis, to determine the status : activated/non activated of each time series $s_\mathbf{x}(t)$. In order to discover interesting projections of the time series, we used the library of wavelet packets shown in Fig. 2. One wavelet packet was of particular interest to us (the third from the right on the last row), because its frequency and phase matched the frequency and the phase of the stimulus. Because the stimulus was periodic we expected to see a large part of the energy of the fMRI signal to be at the stimulus frequency. For each voxel $\mathbf{x}$ we computed the inner product between the time series $\mathbf{s_x}$ and the wavelet packet basis function ψ_γ, $\alpha_\mathbf{x}(\gamma) =< \mathbf{s_x}, \psi_\gamma >$. We then moved the window $W_{\mathbf{x}_0}$ to a new position in the brain and repeated the computations. We considered the mixture of activated and background time series formed by all the time series inside the window $W_{\mathbf{x}_0}$, when it is placed inside an activated region, and we tested the hypothesis that the distribution of the coefficients $\alpha_\mathbf{x}(\gamma)$ was Gaussian. We used the W statistic provided by the Shapiro-Wilk test [7] for this purpose. Fig. 3 shows the histograms of the W statistic of the wavelet packet coefficients computed at each node for several different positions of $W_{\mathbf{x}_0}$. In order to quantify non-normality, we computed a χ^2 statistic between the empirical distribution of the W statistic and the distribution of W obtained

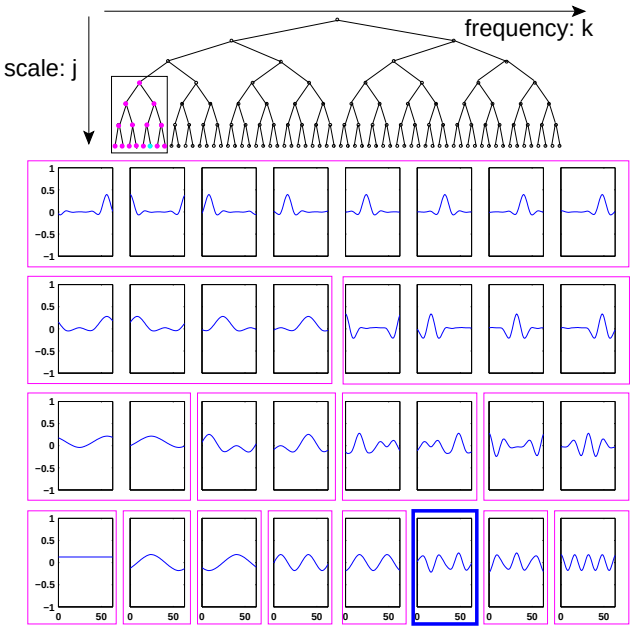

Fig. 2. Top : wavelet packet tree. Bottom : basis functions $\psi_{j,k,l}$ at each node (j, k) of the lower left subtree.

Fig. 3. Histograms of the W statistic of the wavelet packet coefficients at each node for the mixture of activated and background time series. The mean, variance and χ^2 statistic are shown for each plot.

under the normality assumption. A large χ^2 score indicates that the distribution of the wavelet packet coefficient is non Gaussian. While almost all projections of the mixture of time series appear to be Gaussian, there exists a few non Gaussian projections. The most non-Gaussian distribution is obtained for the wavelet packet whose frequency match exactly the frequency of the stimulus. In fact, a close inspection of Fig. 3 reveals that other nodes appear also (albeit relatively less) non Gaussian. As shown in Fig. 2 the wavelet packets at these nodes also oscillate at the frequency of the

stimulus, but have a smaller support (they have local oscillations). In summary, our data support the evidence that it is possible to discover the most interesting projections of the fMRI time series by finding the most non-Gaussian distributions of the wavelet packet coefficients.

4. LOCAL CLUSTERING BASES

We describe the principle of the search for a basis that best partitions the dataset into two clusters. One can think of one cluster as being composed of the activated time series, and the other cluster as being composed of distracting background signals of no interest. A good basis function ψ_γ is one through which we view the two classes as maximally separated as possible. Because most one dimensional projections of high dimensional data are approximatively Gaussian [8], bimodal or more generally non Gaussian, distributions of the wavelet packet coefficients are likely to reveal the presence of clusters. We consider the set of coefficients $\alpha_{\mathbf{x}}(\gamma)$ for all voxels $\mathbf{x} \in W_{\mathbf{x}_0}$, and we cluster the coefficients into two classes. The ability of the basis functions ψ_γ to split, or partition, the time series $\mathbf{s}_{\mathbf{x}}(t)$ into two groups can be estimated with the normalized distance between the two clusters, $\mathcal{D}(\gamma)$, defined by

$$\mathcal{D}(\gamma) = |\overline{\alpha}_\gamma(c_1) - \overline{\alpha}_\gamma(c_2)| \, / \sqrt{s^2_\gamma(c_1) s^2_\gamma(c_2)} \qquad (3)$$

where $\overline{\alpha}_\gamma(c)$ and $s^2_\gamma(c)$ are respectively the centroid and the in-class variance of the cluster c. A good separation will be achieved if $\mathcal{D}(\gamma)$ is large. Given a wavelet packet node (j,k), we consider the subspace spanned by $\mathbf{\Psi}_{j,k}$. In order to decide if we want to include $\mathbf{\Psi}_{j,k}$ in the best basis, we need to quantify the ability of the subspace (as a whole) to separate the dataset into two groups. One could consider the joint probability distribution of all the wavelet packet coefficients generated by the ψ_γ in $\mathbf{\Psi}_{j,k}$. A consequence of the curse of dimensionality is that the estimation of this joint distribution can become problematic as the dimension of the subspace $\mathbf{\Psi}_{j,k}$ becomes large. An alternate approach consists in working with the marginal distributions. In this case we consider the distribution of the wavelet packets generated by each basis function ψ_γ, and compute its discriminatory power $\mathcal{D}(\gamma)$. We consider that the subspace $\mathbf{\Psi}_{j,k}$ separates well the dataset if it concentrates most discriminatory power on a small number of basis functions ψ_γ. In other words, we favor a sparse distribution of the distances $\{\mathcal{D}(\gamma)\}_\gamma$. The sparsity of the distribution can be quantified by its entropy, defined by

$$\mathcal{M}(\mathbf{\Psi}_{j,k}) = - \sum_{l=0}^{2^{J_0-j}-1} \frac{\mathcal{D}^2(j,k,l)}{\|\mathcal{D}(j,k,\cdot)\|^2} \log \frac{\mathcal{D}^2(j,k,l)}{\|\mathcal{D}(j,k,\cdot)\|^2}, \quad (4)$$

where $\|\mathcal{D}(j,k,\cdot)\|^2 = \sum_{l=0}^{2^{J_0-j}-1} \mathcal{D}^2(j,k,l)$. We can now propose the following algorithm for selecting the best basis for unsupervised classification. The wavelet packet tree (see Fig. 2) is explored from the bottom up, and the optimal combination of the $\mathbf{\Psi}_{j,k}$ is kept. Given a set of N time-series vectors $\mathbf{s}_{\mathbf{x}} = [s_{\mathbf{x}}(0), \cdots, s_{\mathbf{x}}(T-1)]^t$, indexed by their position $\mathbf{x}$,

1. **Wavelet packet expansion**. For each vector $\mathbf{s}_{\mathbf{x}}$, compute the wavelet packet coefficients $\alpha_{\mathbf{x}}(\gamma)$ at each node (j,k) of the wavelet packet tree.
2. **Clustering**. For each wavelet packet index $\gamma = (j,k,l)$, cluster the set $\{\alpha_{\mathbf{x}}(\gamma), \mathbf{x} \in W_{\mathbf{x}_0}\}$ into two clusters, and compute the distance $\mathcal{D}(\gamma)$ with (3).
   For each wavelet packet node node (j,k), compute the cost function $\mathcal{M}(\mathbf{\Psi}_{j,k})$ with (4).
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, choose the best subspace $\mathbf{B}_{j,k} =$

$$\begin{cases} \mathbf{\Psi}_{j,k} & \text{if} \quad \mathcal{M}(\mathbf{\Psi}_{j,k}) \leq \mathcal{M}(\mathbf{B}_{j+1,2k}) + \mathcal{M}(\mathbf{B}_{j+1,2k+1}) \\ \mathbf{B}_{j+1,2k} \oplus \mathbf{B}_{j+1,2k+1} & \text{otherwise.} \end{cases}$$

The output of the algorithm is the best basis $\mathbf{B}_{0,0}$.

5. EXPERIMENTS

We first describe experiments conducted on artificial data sets of event-related fMRI time series. The fMRI signal was modeled using the parametric model proposed by Glover [9]. The fMRI signal $\mathbf{s}_{\mathbf{x}}(t)$ is zero before the stimulus t_s. For $t > t_s$, the signal is given by

$$a_1 (t-t_s)^{d_1} e^{-(t-t_s)/t_1} - 0.4 a_2 (t-t_s)^{d_2} e^{-(t-t_s)/t_2} \qquad (5)$$

where, the a_i are normalization constants. We take $t_s = 22.5s$, and we consider 32 time samples equispaced by 1.5 s. The parameters d_1, d_2, t_1, and t_2 are random variables normally distributed. We have chosen the mean of these random variables to be equal to the values estimated by Glover [9] for a motor response. Fig. 4(a) shows the mean realization of the event-related signals. We generate four different realizations of $\mathbf{s}_{\mathbf{x}}(t)$ according to (5). We add white Gaussian noise to these time series [1]. The variance of the noise is increased for each experiment. Finally, we generate 16 other time series that are realizations of white Gaussian noise. Six datasets with average SNR = 0.1, 0.2, 0.5, 0.8, 1, and 1.5 are generated. We apply the best clustering algorithm and obtain the best clustering basis for each dataset. Fig. 4(b) shows the first 2 best clustering basis vectors $\mathbf{b}_{\gamma_0}$, and $\mathbf{b}_{\gamma_1}$ obtained with the data with a SNR = 0.5. These two vectors capture 40% of the total wavelet packet coefficient variance. These vectors resemble the hemodynamic response function (see Fig. 4(a)) at different delay times and different positive response widths. Fig. 5 shows the scatter plots obtained by projecting a dataset onto the two best clustering basis vectors for SNR = 0.5. The two clusters are well separated. We performed a systematic comparison of the performance of our approach with two other standard methods: the t-test and the correlation analysis. Finally, we show some results with data, provided by Dr. Gregory McCarthy (Brain Imaging and Analysis Center, Duke University), that show prefrontal cortex activation in the presence of infrequent events [12]. 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. We extract 16-image segments consisting of the 8 images preceding and the 8 images following each target. We average several segments in order to increase the SNR. This average signal constitutes the time series $\mathbf{s}_{\mathbf{x}}(t)$. Fig. 6(a) shows the first two best clustering basis vectors that capture 40% of the total variance for a 4×4 window $W_{\mathbf{x}_0}$ that was placed in a region where we expected to see

[1] 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 [10]. This colored noise can be uncorrelated by the wavelet transform which 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 [11].

activation. We note that these two vectors, that were discovered automatically by the algorithm, behave as some delayed hemodynamic responses. A scatter plot obtained by projecting the 16 time series on the two best clustering vectors is shown in Fig. 7(a); the scatter plot is elongated and shows two well defined clusters. An activation map, shown in Fig. 7(b), coincides with an activation maps obtained with a standard t-test. The two time series detected as activated by the clustering algorithm are shown in Fig. 6(b). They have the shape expected for hemodynamic responses. We also notice that the two best clustering vectors shown in Fig. 6(a) perform as matched filters to detect these time series.

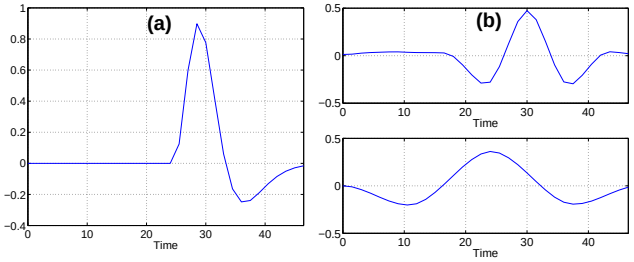

Fig. 4. (a) Mean hemodynamic response time series defined by (5). (b) The best two basis vectors obtained with an SNR=0.5.

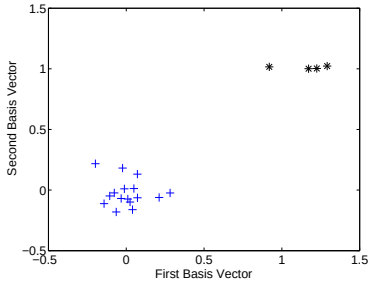

Fig. 5. Scatter plot obtained by projecting the activated time series (*) and the background time series (+) onto the first and the second basis vectors.

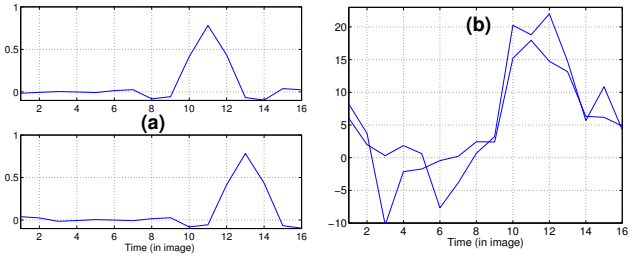

Fig. 6. (a) The two most discriminating basis vectors. (b) Time series extracted from the two voxels detected as activated.

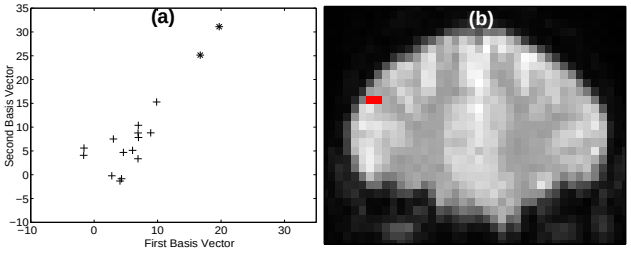

Fig. 7. (a) Scatter plot obtained by projecting the activated fMRI time series (*) and the background time series (+) onto the first and the second basis vectors. (b) Activation map obtained by clustering the wavelet packet coefficients of the two basis vectors.

6. 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. R. Soc. Lond. B*, , no. 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] R.R. Coifman and M.V. Wickerhauser, "Entropy-based algorithms for best basis selection," *IEEE Trans. Information Theory*, vol. 38, no. 2, pp. 713–718, March 1992.

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

[5] J.L. Tanabe, D. Miller, J. Tregellas, R. Freedman, and F.G. Meyer, "Comparison of detrending methods for optimal fMRI pre-processing," *NeuroImage*, vol. 15, pp. 902–907, 2002.

[6] 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.

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

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

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

[10] 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.

[11] F.G. Meyer, "Wavelet based estimation of a semi parametric generalized linear model of fMRI time-series," *IEEE Transactions on medical imaging*, vol. 22(3), pp. 315–322, 2003.

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