Wavelet-Based Statistical Analysis in Functional Neuroimaging

RADU MUTIHAC
University of Bucharest
Department of Electricity and Biophysics
405 Atomistilor St., 077125 Bucharest
ROMANIA

Abstract: Wavelet-based analysis versus Gaussian smoothing in statistical parametric mapping (SPM) for detecting and analyzing brain activity from functional magnetic resonance imaging (fMRI) data is presented. Detection of activation in fMRI data can be performed in the wavelet domain by a coefficient-wise statistical *t*-test. The link between the wavelet analysis and SPM is based on two observations: (*i*) the low-pass analysis filter of the discrete wavelet transform (DWT) can be similarly shaped to a Gaussian filter in SPM, (*ii*) the subsampling scheme provides means to define the number of coefficients in the low-pass subband of the wavelet decomposition [52]. Analysis of an fMRI block-based visual stimulation paradigm was comparatively performed by wavelet analysis and statistical parametric mapping (SPM) [13] based on the Random Field Theory (RFT). The voxels were isotropic and the same general linear model (GLM) was employed in both the image space and the wavelet domain. Consequently, an equivalent spline degree for which the low-pass part of the wavelet analysis is basically equivalent to SPM can be computed. The processing of neuroimaging data in the wavelet domain was carried out by means of two different biorthogonal transforms: *3D* fractional-spline wavelets and *2D+Z* fractional quincunx wavelets [51], resulting in activation patterns similar to the activation maps obtained by linear regression analysis in SPM.

Key–Words: Wavelets, Fourier transform, multiresolution analysis, functional magnetic resonance imaging, general linear model, statistical parametric mapping.

1 Introduction

Analysis of functional magnetic resonance imaging (fMRI) data, the most complex biomedical time series, is a non-invasive method that allows to localize and study the dynamic brain processes in intact living brains. The promise of functional brain imaging is to provide a reliable answer to some of the fundamental open questions of understanding the brain, namely how microscopic and macroscopic organization in the brain relate and interact to produce brain function [42].

The index of neuronal activity widely spread in neuroimaging is the blood oxygenation level dependent (BOLD) contrast [35], which is based on the differing magnetic susceptibilities of oxygenated hemoglobin (diamagnetic) and deoxygenated hemoglobin (paramagnetic) in relation with the surrounding tissue. The basic assumption is that an increase in neuronal activity within a brain region leads to hyperoxemia, which attracts an increase in local blood flow and reduced concentrations of deoxyhemoglobin in the blood vessels. This is due to an overcompensation of the local tissue perfusion in response to the increased energy demand of the activated neurons [28]. Consequently, relative decreases in deoxyhemoglobin concentration attract a reduction in local field inhomogeneity and a slower decay of the MR signal, resulting in higher intensities in T2*-weighted images (Fig. 1). Most of neuroimaging data analysis relies on inferential hypothesis-driven analysis that employs spatially extended processes like SPM, which is a representation of functional activity distribution induced by a certain task. In a single-subject experiment, the analysis of statistical parametric maps consists in collecting results of univariate tests com-

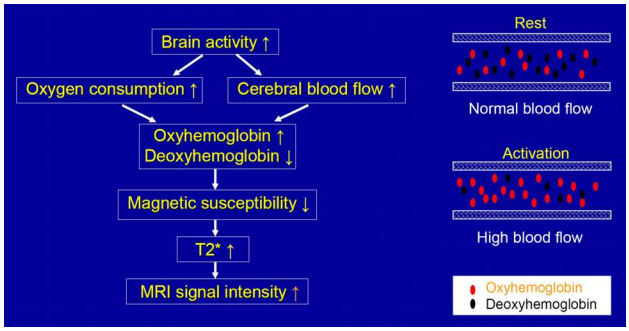

Figure 1: Brain blood flow responds to brain activation.

puted on each voxel of the brain. Changes in blood flow, which are closely coupled to functional activation, indicate brain areas involved in various tasks that subjects perform in the scanner (Fig. 2). However, these changes and oxygenation variability (vascular and hemodynamic) are temporally delayed relative to the neural firing (the hemodynamic *lag*). The nature of these regulatory processes and the BOLD signal change have not been completely elucidated so far. An additional difficulty in delineating functional correlates from spatiotemporal fMRI data sets stems from the relatively small effect sizes in blood-flow related phenomena ($1\% - 4\%$ in 1.5 T scanners), quite close to scanner-to-scanner variability, which is translated in low signal-to-noise ratio (SNR) of the BOLD signal. The difficulty in extracting information from raw data is supplementary increased by the possibility that functional correlates of brain activity may relate to given behavioral paradigms in complicated ways. Moreover, the interpretation of functional brain imaging data requires some assumptions on processing in the working brain that may not be entirely realistic and which preclude canonical methods of data analysis and experimental design.

Low SNR in fMRI and the inevitable presence of confounding effects often result in unsatisfactory analysis. Significant noise and artifacts present in fMRI time series, as well as their unknown structure, complicate the problem of activation detection in time domain. In functional neuroimaging, SNR can be improved by two means: (*i*) reducing the scan acquisition time (TR); and (*ii*) by stimulus repetition followed by scan averaging. The duration of BOLD-based fMRI experiments is primarily affected by the physiological sources of variability (cardiac, pulmonary, and other pulsations), subject and scanner movements, RF coil heating, and gradient drift. BOLD methods are sensitive to artifacts associated with head and/or vessel motion [17], vascular inflow [9], and drainage effects [25]. Quite often these artifacts are correlated with the signal of interest and, consequently, not discarded by simple image averaging. Still BOLD-based neuroimaging methods are generally superior to any imaging modalities in terms of: (*i*) equipment availability, (*ii*) no exogeneous contrast agents or exposing to ionizing radiation requirement, (*iii*) best spatial resolution, and (*iv*) activation images can be coregistered with anatomical images acquired on the same machine [37].

In both multisubject and/or multisession experiments, in order for a voxel to represent the same anatomical location for all subjects/sessions under every condition, raw data usually undergo a number of preprocessing steps and are mapped into a standardized coordinate space that accounts for differences

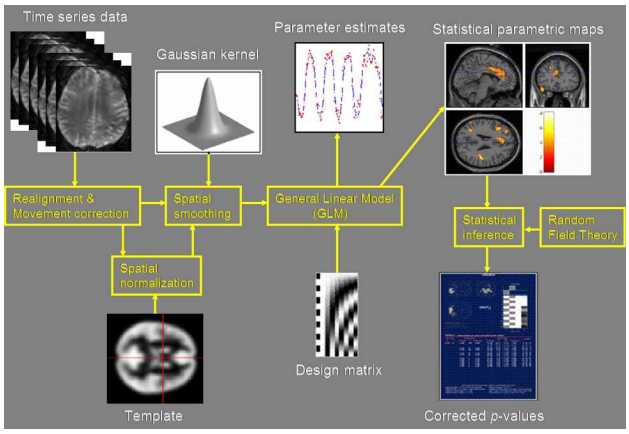

Figure 2: Hypothesis-driven statistical analysis in SPM resulting in parameter maps of task-correlated brain activity. Gaussian Random Field Theory is employed to handle multiple hypothesis testing.

in brain size and orientation [13], [54], such as the stereotaxic space [43]. Images of the same subject are realigned among themselves and than coregistered with the correspondent MRI volume. The latter is then used to compute the nonlinear transformation that warps images of all subjects in a standardized reference space. A test statistic (e.g., *t*-statistic) is computed at each anatomical site by application of a GLM that corrects for the effects of no interests. The result is a large set of locations that defines a *statistical field*.

Statistical fields are usually smoothed by application of a Gaussian filter. Spatial smoothing is applied in order to: (*i*) match the spatial scale of hemodynamic responses among subjects/sessions; (*ii*) increase the SNR (however, variation between subjects dominates the noise in multisubject data); (*iii*) better data match to the assumptions of the Gaussian RFT; (*iv*) normalize the error distribution to improve the validity of inferences based on parametric tests; (*v*) spatial smoothing of temporal autocorrelations decreases errors in the estimated SD's by increasing the effective degrees of freedom (*df*) and decreasing its sensitivity to underlying temporal correlation structure [57]. Less spatially variable *t*-statistic images and lower thresholds ($p < 0.05$) entail better detection of activations and improves the physiological relevance of statistical inference; (*vi*) in multisubject experiments, averaging is necessary to smooth the projected data down to a scale where homologies in functional anatomy are expressed across subjects. Running more subjects will, nevertheless, improve statistical power much more than moving to higher fields.

The final task in statistical analysis of neuroimaging data is to come out with a statistical map in that one has to identify the locations of the images affected

by the experimental paradigm. Commonly, these areas are identified by setting an appropriate threshold and retaining as significant those anatomical locations where the statistical field is higher than the chosen level. Values for the threshold are usually computed following criteria derived from the Gaussian RFT. This approach provides the probability of false positives for a certain threshold for a statistical field of a certain smoothness. The higher the smoothness, the higher is the correlation among adjacent locations, the lower the risk of a noise field crossing the selected threshold. By smoothing the field with a filter of fixed resolution, the probability of detecting signals of that particular size is maximized, yet sharper foci might be wiped out or signal of varying spatial scale could erroneously be detected. The latter is overcome by a multiresolution approach that employs a number of different smoothing levels and sets the threshold accordingly. However, thresholding the statistical fields equates to "modeling" the noise rather than the signal, since it comes out with a binary output: everything over the threshold is unlikely to be noise, everything under the threshold is likely to be noise [55]. The area of change can be detected, though nothing can be said on the shape, scale, and intensity of the signal because the values of the statistics over the threshold are obviously still subject to random (noise) fluctuations.

Based on some assumptions, various parametric and nonparametric denoising methods have been proposed, though these assumptions may not necessarily hold for fMRI data [29]. Wavelet methods approach the problem of the analysis of statistical fields by estimating the signal at any resolution among the random fluctuations. Recently, the appearance of explicit orthonormal bases in multiresolution analysis (MRA) entailed significant implications on fMRI data analysis. Wavelet application to statistical fields is similar to wavelet applications to images since statistical maps are just images with noise with variance equal to unity. Techniques that are able to identify small signal changes against a noisy background are employed in fMRI, yet many of them cannot deal with responses which can change amplitude in an unpredictable manner. Wavelets have successfully been applied to time-frequency analysis of nonstationary signals to detect hemodynamic responses to experimental stimuli. Wavelet shrinkage of statistic maps overlaps with the work on multiple hypothesis testing [30], yet wavelet shrinkage has not thoroughly been investigated for *2D* and *3D* denoising of fMRI statistic maps prior to statistical testing in the spatial domain. Gaussian spatial smoothing with single kernel is widely applied for this purpose at the risk of missing to detect spatial features of the smoothing kernel size or lower. Smoothing by wavelet shrinkage allows locally adaptive bandwidth so that the power to detect spatial features of varying extent is not constrained by the arbitrary choice of a single kernel size [3]. Many techniques have focused on improving the power of the statistical test in the wavelet domain by proposing alternative error rates (i.e., false discovery rate), by modifying the hypotheses to be tested (i.e., recursive testing), or by fine-tuning the statistical models (i.e., Bayesian framework) [51].

In summary, the wavelet-based statistical analysis provides: (*i*) multiresolution decomposition suitable for scale-invariant process analysis; (*ii*) sparse representation of typical brain activation maps; (*iii*) optimally whitening of data producing approximately decorrelated wavelet coefficients; (*iv*) Karhunen-Loève (KL) expansions for long-memory ($1/f$-like) processes, which is the case in fMRI; and (*v*) good estimators for the noise process parameters.

2 Functional Neuroimaging Data Analysis

Extracting information from raw data is aiming to reveal the structure in data and model the underlying processes that generated data. In practice, data are acquired in their rawest form and, consequently, are of little immediate use. It is only when the information is extracted via processing that data become meaningful. In data analysis, it is often desirable to reduce the dimension of feature space because there may be irrelevant or redundant features that complicate subsequent inferences and model design, increase the computational demand, and render the analysis suboptimal.

2.1 Specificity of Cerebral Activity

The general framework for the analysis of functional neuroimaging data was established in positron emission tomography (PET) and extended thereafter to fMRI. Such noninvasive neuroimaging techniques are widely used to study brain function and dysfunction by spatially localizing dynamic brain processes in intact living brain. Changes of neural activity associated with various stimulus conditions and behaviors are referred to as *functional correlates*. It is a common feature for all brain imaging methods to measure local neuronal activity by indirect means; PET and fMRI measure local properties of cerebral blood flow (CBF) [20]. PET signals are based on regional cerebral blood flow (rCBF), whereas fMRI signals are most commonly based on BOLD contrast, which can be employed to track blood-flow-related phenomena accompanying or following neuronal activations.

Two general principles of cerebral function have been derived from investigating brain lesions and recording signals from smaller or larger clusters of neurons: (*i*) *functional specialization* of brain regions, which means that different brain regions perform different tasks [58], and (*ii*) *functional integration*, which states that cerebral functions are carried out by networks of interacting regions and that different functions correspond to different networks [15]. There are consequently two main types of assumption underlying the interpretation of functional neuroimages, namely the *subtraction paradigm* and the *covariance paradigm* [19]. The subtraction paradigms assume that different brain regions are engaged in different brain functions (i.e., they rely on functional specialization). The covariance paradigms assess the temporal covariance between different brain regions during a particular task. Significant covariance between regions associated with a particular brain function is termed *functional connectivity*. The extraction of functional correlates from raw data sets is facilitated by using the subtraction or covariance paradigms for preprocessing. Due to their complementarity, it is often necessary to employ both of them in order to resolve all the functional components of a given cerebral process [41].

2.2 Inferential versus Exploratory Analysis

Data analysis methods employed in neuroimaging to reveal statistical regularities in data that can be associated with brain function can be loosely dichotomize in two large categories: hypothesis-driven (inferential) and data-driven (exploratory) analysis. Most of imaging neuroscience relies on inferential analysis, which makes use of spatially extended processes like statistical parametric mapping (SPM). The expected BOLD changes in fMRI are specified as regressors of interest in a multiple linear regression framework like the general linear model (GLM), and the estimated regression coefficients are tested against a null hypothesis. The voxel-wise test statistics form summary images known as statistical parametric maps that are representations of the spatial distribution of functional activity induced by the task, and which are subsequently assessed for statistical significance [36]. This approach is essentially confirmatory in nature and based on strong priors about the spatio-temporal characteristics of the signals in data. Consequently, the inferred spatial patterns of activation depend heavily on the accuracy of these assumptions.

The analysis of activation images in brain imaging relies on collecting results of univariate tests computed on each voxel of the brain under different experimental conditions in one subject, in different subjects or groups of subjects. Statistical analysis of image differences corresponding to different experimental conditions belongs to one of the two classes: (*i*) techniques which first partition the brain into regions of interest, based upon anatomical structure, and (*ii*) techniques in which one creates images of a statistical parameter at the level of full brain. In SPM one applies statistical hypothesis testing to each element of the difference image between two different activation states of the brain controlled by some experimental paradigm on the basis of modeling the difference image as a realization of a Gaussian random field. Because the images have poor SNR due to intrinsic biological heterogeneity, scanner-induced noise, and subject's movement, averaging over several experimental trials/subjects is a common practice that produces a mean difference image and its associated sample standard deviation (SD) image [37]. Their direct statistical analysis in the spatial domain is problematic because of: (*i*) low SNR signals, (*ii*) the large number of pixels subject of investigation, (*iii*) the inevitable spatial correlation among pixels caused by the acquisition process and the spatial preprocessing of data.

The main difficulties encountered by the the hypothesis-driven models in fMRI data analysis are the following:

1. Possible presence of unmodeled signals in data, particularly artifactual activity;

2. Structured noise, which is temporally non-orthogonal to an assumed regression model, will bias the parameter estimates in the design space, whereas noise orthogonal to the model subspace will inflate the residual errors, thus reducing the statistical significance and rendering the analysis suboptimal;

3. An increasingly number of models include spatial prior information, whereas the RFT-based inference deals with spatial properties *after* modelling has completed.

4. The standard GLM is essentially univariate [13]. The main limitations of univariate hypothesis-driven methods stem from their solely dependence on temporal predictability of phenomenon to be detected and neglecting information deriving from the *covariance* of the acquired voxel time series (though univariate approaches may be formulated independent of a temporal model [11]). Comparatively, multivariate fMRI data analysis relies on the covariance paradigm and is free of prior assumptions on activation functions.

In contrast with the inferential approach, the exploratory analysis makes no reference to prior knowl-

edge of the structure in data and provides models whose characteristics are determined by the statistical properties of data only. Within a well-defined hypothesis (model) space, competing models can be tested using Bayesian second level of inference, which estimates how probable a set of alternative models is [32]. Skilling *et al.* [40] defined the evidence of a model by integrating the normalization denominator in Bayes' theorem over its prior parameters. Bayesian ranking is carried out by evaluating and comparing the evidence for alternative models

Exploratory analysis of fMRI data is able to detect functional activity without any reference to the experimental protocol or any model space, rather it can reveal new components in data, which are difficult to *a priori* cast in a temporal model, even though multivariate models can be formalized in terms of temporal modelling. Moreover, multivariate analysis allows the study of neural phenomena that generate a nonzero mutual correlation among voxel time courses from different interconnected areas. The data-driven (model-free) methods include generally the (neuromorphic) unsupervised learning methods like self-organizing artificial neural networks (NNs), eigenimage analysis, temporal clustering analysis (CA) and fuzzy clustering analysis (FCA), factor analysis (FA), projection pursuit (PP), principal component analysis (PCA), and independent component analysis (ICA). Exploratory, data-driven techniques are complementary to hypothesis-led methods; the representative time courses they produce may be viewed as alternative hypotheses to the null hypothesis (i.e., no activation). A critical evaluation and comparison of the data-driven methods used in fMRI data analysis has not been published to date. Besides, as briefly stated by Huber [21], "... there are no panaceas in data analysis" whatsoever, so that an educated choice appears to be domain-dependent.

Tukey [44] argued that classical statistics leaning on analyzing small, homogeneous, stationary data by means of known distributional models and assumptions will prove inappropriate to deal with the problems raised by the analysis of large and complex data. A typical 3D multi-slice volume fMRI brain data set for a single subject ranges from 10^8 to 10^{11} bytes, and the values increase further in simultaneous inter-subject and/or multi-subject studies. Two features of fMRI data that characterize massive data sets, namely *nonstationarity* and *distributional heterogeneity*, if taken into account, are making the analysis even more complicated. It is claimed that the difference between real-life large data sets and smaller ones consist not only in size but in qualitative terms as well [22]. Consequently, the investigations of functional brain imaging data should primarily rely on critical consideration of methods that belong to data mining and exploratory data analysis (EDA) [31]. In this respect, multivariate techniques are often combined with data-driven techniques to provide more complex and insightful representations of data.

2.3 Wavelet Analysis

Wavelets are mathematical functions best suited to process signals that contain discontinuities and sharp spikes by cutting them up into different frequency components and subsequently analyzing each component with a resolution matched to its scale. Unlike the traditional Fourier bases, wavelet bases offer a degree of localization in space as well as in frequency. *Orthogonal* transforms project a signal onto a set of basis functions without alterations so that its key features can be detected in lower dimensional subspaces. This enables development of simple function estimates that respond effectively to discontinuities and spatially varying degrees of oscillations in a signal, even if noisy. Wavelets analysis procedure is to adopt a wavelet prototype function (*analyzing wavelet* or *mother wavelet*). Temporal analysis is performed with a contracted, high-frequency version of the prototype wavelet, while frequency analysis is performed with a dilated, low-frequency version of the same wavelet. Scale-varying basis functions render signal processing less sensitive to noise because it measures the average fluctuations of the signal at different scale. Since the original data can be represented in terms of wavelet expansion (i.e., a linear combination of the wavelet functions), any operations on data can be carried out using the corresponding wavelet coefficients only.

In the context of fMRI data analysis, the features of the brain signals are not well-known and the optimal basis functions cannot therefore be specified in advance. Wavelet methods approach the analysis of statistical fields by estimating the signal at any resolution among the random fluctuations. Wavelet application to statistical fields is similar to wavelet applications to images. The variance is computed more straightforward for statistical maps than for images because: (*i*) statistical maps are images with noise variance equal to unity; (*ii*) pure noise images (i.e., residual images) can be obtained by subtracting from the original scans the effects estimated through statistical analysis. Then the noise power of the field can be computed through Fourier techniques [56]. Variances of wavelet levels are computed by the product of the power function of the field with the power function of the wavelet filters [46]. Therefore, statistical maps are transformed using the DWT, the resulting coefficients are thresholded, and, finally, the denoised statistical

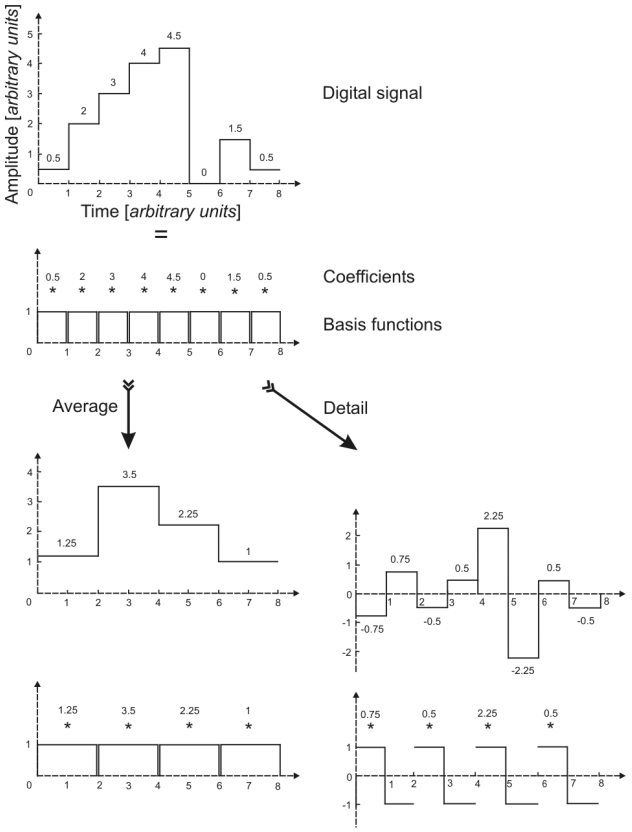

Figure 3: Multiresolution decomposition of a digital signal

maps are reconstructed by the inverse DTW [37] [47].

A multiresolution analysis (MRA) performed by the WT circumvents the problem by detecting and extracting the key signal features over many scales (Fig. 4). Wavelet methods are particularly attractive due to broadly fractal properties exhibited by the brain in space and time. Since the interest in fMRI data is focused on relatively spatially localized signals, wavelets are particularly suited to represent them by a small number of strong local coefficients, whereas the power of white noise is uniformly spread throughout the wavelet space. Some wavelet-space partitions may exhibit improved localized SNR conditions that can be exploited by an orthogonal wavelet decomposition. Then the statistical analysis can be restricted for the significant coefficients to these partitions only. The potential benefits are: (*i*) improved SNR conditions, and (*ii*) a decrease of the detection threshold due to the reduced number of statistical tests, both contributing to a higher detection sensitivity with no increase in type I errors (false positives) [37]. Therefore, the control of false positives is stringent and the Bonferroni approach to wavelet thresholding was initially suggested [45].

2.3.1 Mathematical Background

In order to consistently sketch the wavelet analysis framework, some basic mathematical concepts need to be reviewed. The space of all functions $f(x), x \in \mathbb{R}$ that are square integrable (e.g., finite energy) in the Lebesgue's sense

$$\int_{-\infty}^{+\infty} |f(x)|^2 \, dx < +\infty \tag{1}$$

is denoted by $L_2(\mathbb{R})$ or, simply, L_2. Wavelet analysis procedure is to adopt some two continuously-defined functions:

1. The *scaling function* (or *father function*) $\phi(x)$, The scaling function $\phi(x)$ is the solution of a *two-scale* equation:

$$\phi(x) = \sqrt{2} \sum_{k \in \mathbb{Z}} h(k) \phi(2x - k) \tag{2}$$

   where the sequence $\{h(k)\}_{k \in \mathbb{Z}}$ is the *refinement* filter.

2. Its associated *wavelet function* $\psi(x)$ (*prototype* or *mother wavelet* function):

$$\psi(x) = \sqrt{2} \sum_{k \in \mathbb{Z}} g(k) \phi(2x - k) \tag{3}$$

   where $\{g(k)\}_{k \in \mathbb{Z}}$ is a suitable weighting sequence.

Wavelets are "small" waves: they oscillate and their curves yield zero net area:

$$\int_{-\infty}^{+\infty} \psi(x) \, dx = 0 \tag{4}$$

The "smallness" refers to the fact that they are localized in time, in contrast to Fourier basis consisting of sines and cosines that are perfectly localized in frequency space but do not decay as a function of time (i.e., nonlocal support). Wavelets decay to zero as $x \to \pm\infty$ and exhibit good localization properties in frequency space.

In wavelet MRA, details at various levels of resolution are represented by the superposition of wavelets associated with the appropriate dilations (Fig. 4). Complex data can be inspected by dilation, which performs "zoom-in" on details. Reversely, details can be suppressed and wavelets are, therefore, candidates for data smoothing. Unlike the case of Fourier transforms, a large selection of wavelet families is available depending on the mother wavelet. Nevertheless, desirable properties like orthogonality, compactness of support, rapid decay, and smoothness impose several restrictions on the choice of mother wavelet.

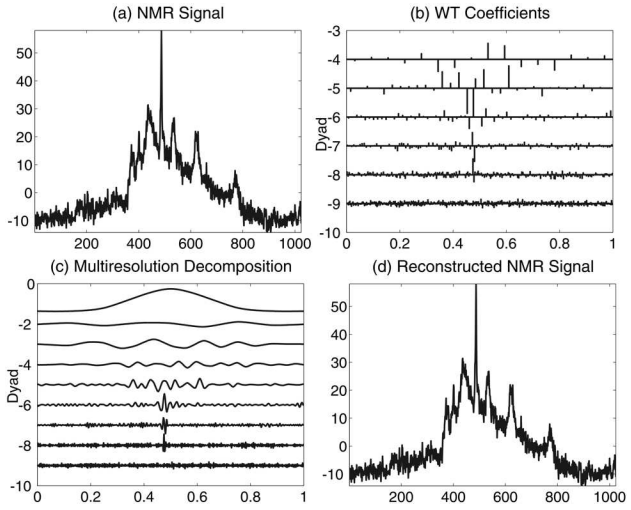

Figure 4: Wavelet MRA of a typical signal in fMRI: (a) the original MR signal, (b) forward periodized, biorthogonal, symmetric WT of the signal (wavelet coefficients), (c) multiresolution decomposition of the MR signal, (d) Inverse WT reconstruction of the MR signal from the wavelet coefficients.

2.3.2 Wavelet Basis Functions

Functional features of the brain signals are complex, largely unknown, and difficult to mathematical modeling, so that optimal basis functions cannot be specified in advance. Wavelet MRA circumvents this problem by detecting and extracting the key signal features over many scales. A wavelet *basis* is fractal and so a natural choice of the basis for analysis of fractal data. Hence wavelets may be more than just another basis for analysis of fMRI data [2].

Wavelet construction starts from the basic mother wavelet, $\psi(x)$, that generates the basis by (dyadic) dilation (index j) and translation (index k) in time $\psi(ax-b)$. For discrete wavelets, the parameter of translation, b, and dilation, a, are restricted to discrete sets, usually $a = 2^j$ and $b = k$ where $j, k \in \mathbb{Z}$. Dilation allows hierarchical representation of a data set. Temporal analysis is performed with a contracted, high-frequency version of the prototype wavelet, while frequency analysis is performed with a dilated, low-frequency version of the same wavelet.

$$\psi_{j,k} = 2^{-j/2}\psi(x/2^{-j} - k) \tag{5}$$

Any function $f \in L_2$ can be uniquely represented by the expansion:

$$f(x) = \sum_{j\in\mathbb{Z}}\sum_{k\in\mathbb{Z}} d_j(k)\psi_{j,k}(x) \tag{6}$$

The wavelet coefficients $\{d_j(k)\}_{j,k\in\mathbb{Z}}$ are obtained by forming the (double infinite) sequence of inner products:

$$d_j(k) = \langle f, \tilde{\psi}_{j,k}\rangle_{L_2},\ j,k \in \mathbb{Z} \tag{7}$$

where $\{\tilde{\psi}_{j,k}\}_{j,k\in\mathbb{Z}}$ is the *biorthogonal* basis of $\{\psi_{j,k}\}_{j,k\in\mathbb{Z}}$ such that:

$$\langle \tilde{\psi}_{j,k}, \psi_{i,l}\rangle = \delta_{j-i}\cdot\delta_{k-l},\ i,j,k,l \in \mathbb{Z} \tag{8}$$

Spline bases posses the best approximation properties like the smallest L_2-error [48]. Due to their smoothness, splines are well localized in both time and frequency domains. Studies on wavelet application in fMRI data analysis emphasized the importance of symmetric wavelets and scaling functions that are free from phase distortions [37]. Orthogonal bases are mainly recommended because of the following: (*i*) signal features not known beforehand can be detected and extracted in a multiresolution approach over many scales; (*ii*) transform of white noise into white noise [23]. As such, in wavelet analysis of fMRI time series, the preprocessed data are subject to spatial nonredundant DWT, rather than spatially convolved with a Gaussian kernel.

Orthogonal wavelet basis functions for fMRI time series can be found by appropriate choice of the sequences $\{h(k)\}_{k\in\mathbb{Z}}$ and $\{g(k)\}_{k\in\mathbb{Z}}$ or, equivalently, ϕ and ψ, such that $\{\psi_{j,k}\}_{j,k\in\mathbb{Z}}$ constitutes an orthonormal basis of L_2. Hence

$$\forall f \in L_2, f(x) = \sum_{j\in\mathbb{Z}}\sum_{k\in\mathbb{Z}} d_j(k)\psi_{j,k} + \sum_{j\in\mathbb{Z}}\sum_{k\in\mathbb{Z}} c_j(k)\phi_{j,k} \tag{9}$$

where the wavelet coefficients $\{d_j(k)\}_{j,k\in\mathbb{Z}}$ and the approximation coefficients $\{c_j(k)\}_{j,k\in\mathbb{Z}}$, due to orthogonality, are obtained by inner products with the corresponding basis functions:

$$d_j(k) = \langle f, \psi_{j,k}\rangle, \quad c_j(k) = \langle f, \phi_{j,k}\rangle, \tag{10}$$

The decomposition of any $f \in L_2$ is practically carried out on a finite number of scales only, say J, so that:

$$f(x) = \sum_{j=1}^{J}\sum_{k\in\mathbb{Z}} d_j(k)\psi_{j,k} + \sum_{k\in\mathbb{Z}} c_J(k)\phi_{J,k} \tag{11}$$

It is more convenient to describe the underlying decomposition algorithm that uses two complementary filters h and g. We consider hereafter nonredundant dyadic orthogonal wavelet transforms only. The algorithm consists of an iterated orthogonal filterbank with an *analysis* and a *synthesis* part (Fig. 5). Orthogonality imposes:

$$\tilde{H}(z) = H(z^{-1}) \quad \text{and} \quad \tilde{G}(z) = G(z^{-1}) \tag{12}$$

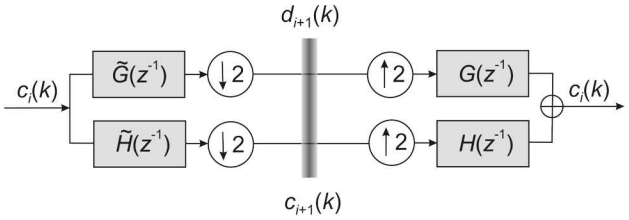

Figure 5: Decomposition and reconstruction quadrature mirror filters. Analysis part (left) and synthesis part (right) of the WT filterbank.

where $H(z)$ is the the synthesis scaling filter, that is, the transfer function (z-transform) of the low-pass refinement filter h, and $\tilde{H}(z)$ is the associated analysis scale filter. Likewise, $G(z)$ is the the synthesis wavelet filter, that is, the transfer function (z-transform) of the high-pass filter g, and $\tilde{G}(z)$ is the associated analysis wavelet filter. The high-pass filter g is the modulated version of h given by:

$$G(z) = z \cdot H(-z^{-1}) \tag{13}$$

For perfect reconstruction, the filters must obey the quadrature mirror filter (QMF) conditions:

$$\begin{aligned} \tilde{H}(z^{-1})H(z) + \tilde{G}(z^{-1})G(z) &= 1 \\ \tilde{H}(z^{-1})H(-z) + \tilde{G}(z^{-1})G(-z) &= 0 \end{aligned} \tag{14}$$

In terms of the low-pass filter h only, the QMF conditions equate:

$$\begin{aligned} H(z)H(z^{-1}) + H(-z)H(-z^{-1}) &= 2 \\ H(1) = \sqrt{2} \Leftrightarrow H(-1) &= 0 \end{aligned} \tag{15}$$

The wavelet decomposition is iteratively implemented using QMF filterbanks. For a signal vector of length N_0, the operations required by the WT are $\mathcal{O}(N_0)$, as compared with the standard FFT complexity of $\mathcal{O}(N_0\, log N_0)$.

Preprocessing in image analysis (like noise reduction, contrast enhancement, ...) can be carried out by making the operations frequency dependent (i.e., split signal/image into frequency subbands and apply different operations on each subband). Rather than using a huge multichannel filterbank to encompass the full spectrum, the WT employs recursive 2-channel filterbanks (a low-pass filter and a high-pass filter) (Fig. 6). In a one-level FWT, a signal c_0 is split into an approximation part c_1 and a detail part d_1. In multilevel FWT, each subsequent c_i is split into an approximation c_{i+1} and detail d_{i+1} pair of coefficients. The inverse FWT (IFWT) reconstructs each c_i from c_{i+1} and d_{i+1}. The decomposition (10) can be extended to *multiple dimensions* (e.g., *2D* or *3D*) by using tensor product

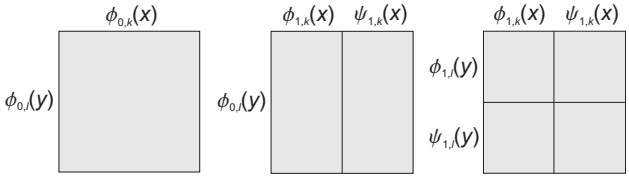

Figure 6: One iteration of the WT in *2D* space: the *1D* algorithm is first applied along the x-direction splitting the columns of data in two halves (top), which are subject to its subsequent application along the y-direction that splits the rows into two halves (bottom). The iteration continues on the upper left quadrant.

basis functions, which amounts to successively applying the *1D* decomposition algorithm along each dimension in multidimensional data. By iteration, 2^q different type of basis functions are generated in q dimensions. The corresponding qD separable scaling functions with $\mathbf{x} = (x_1, x_2, ..., x_q)$ are:

$$\phi_{j,\mathbf{k}}(\mathbf{x}) = \prod_{i=1}^{q} \phi_{j,k_i}(x_i) \tag{16}$$

where $\mathbf{k} = (k_1, ..., k_q)$ is the vector integer index. The rest of $2^q - 1$ types of wavelet basis functions are obtained by replacing one or more factors in (16) with wavelet terms of the form $\psi_{j,k_i}(x_i)$, $j \in \mathbb{Z}$, $i = 1, 2, ..., q$. Define $\mathbf{b} = (b_1, ..., b_q)$ a binary vector such as:

$$b_i = \begin{cases} 1 & \text{if } \phi_{j,k_i} \text{ is replaced by } \psi_{j,k_i} \\ 0 & \text{otherwise} \end{cases}, \; i = 1, 2, ..., q \tag{17}$$

and

$$\varphi_{j,k_i} = \begin{cases} \psi_{j,k_i} & \text{if } b_i = 1 \\ \phi_{j,k_i} & \text{otherwise} \end{cases}, \; j \in \mathbb{Z}, \; i = 1, 2, ..., q \tag{18}$$

then the mixed tensor product wavelets can be rewritten [37]

$$w^m_{j,\mathbf{k}}(\mathbf{x}) = \prod_{i=1}^{q} \varphi_{j,k_i}(x_i), \quad m = 1, 2, ..., 2^q - 1 \tag{19}$$

with

$$m = \sum_{i=1}^{q} b_i 2^{i-1} \tag{20}$$

Here, m indicates a preferential spatial orientation since ϕ is low-pass and ψ is high-pass. As for instance, in the *2D* case, $w^m_{j,k}(\mathbf{x})$ for $m = 1, 2, 3$ correspond to wavelets oriented along the horizontal, diagonal, and vertical directions, respectively (Fig. 7).

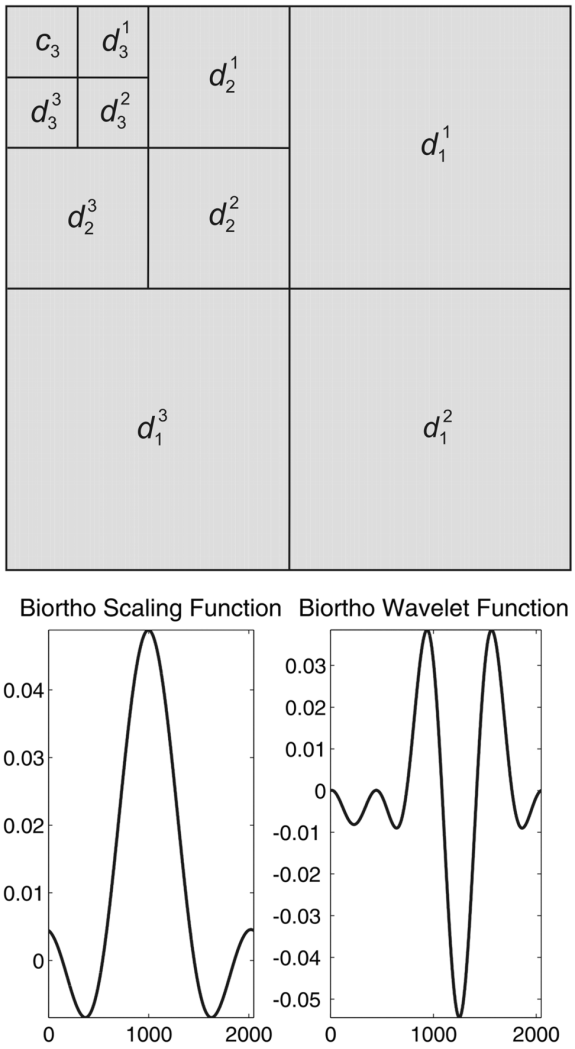

Figure 7: Approximation and detailed wavelet coefficients in a level-three *2D* WT (top); *1D* biorthogonal scaling and wavelet functions, respectively (bottom).

The corresponding multidimensional coefficients:

$$\begin{aligned} c_j(\mathbf{k}) &= \langle f, \phi_{j,\mathbf{k}} \rangle \\ d_j^m(\mathbf{k}) &= \langle f, w_{j,\mathbf{k}}^m \rangle \end{aligned} \tag{21}$$

are iteratively obtained by successive filtering and downsampling by a factor of two. In the case of multilevel FWT of *2D* images, each approximation coefficient c_i is split into an approximation coefficient c_{i+1} and three detail coefficients d_{i+1}^1, d_{i+1}^2, and d_{i+1}^3, for horizontally, vertically, and diagonally oriented details, respectively (Fig. 7). The biorthogonal scaling function ("father") and its corresponding wavelet function ("mother") are also showed in Fig. 7. These wavelets are employed to run *2D* three-level WT of an axial MR brain slice as presented in Fig. 8. Further on, the first three approximation levels are shown in Fig. 9 for visual comparison. Symmlets are

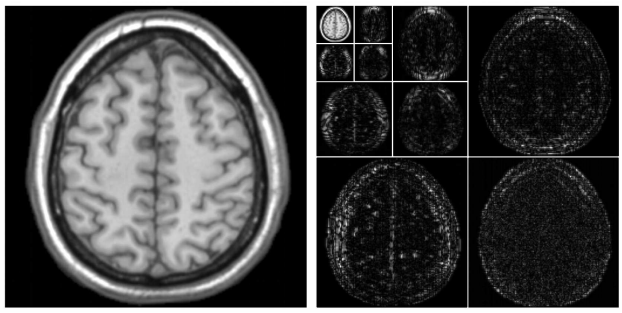

Figure 8: *2D* wavelet decomposition of a typical MR axial slice. A *coarse* (*approximation*) image at a resolution level L is represented by 2^L pixels in each direction. The *detail* images at a particular level L are produced by *horizontal*, *vertical*, and *diagonal* differences between successive levels. The set of coefficients produced by the WT consist of the lowest coarse level image and the higher level detail images. The original images is $2^8 \times 2^8$ pixels (left).

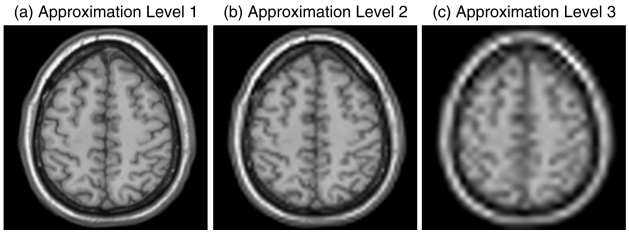

Figure 9: Coarse images of a typical MR axial slice at various approximation levels: (a) Approximation 1 (resolution level $L = 7$); (b) Approximation 2 ($L = 6$); (c) Approximation 3 ($L = 5$). All images were rescaled to the same size for better comparison.

wavelets within a minimum size support for a given number of vanishing moments, but they are as symmetrical as possible, as opposed to the Daubechies filters which are highly asymmetrical. They are indexed by the number of vanishing moments, which is equal to half the size of the support. Fractional splines of a real-valued degree were proposed to produce wavelet bases [49], such as symmetric and causal, orthogonal, and biorthogonal. A reasonable trade-off seems to lead to symmetric, orthonormal cubic spline wavelets. Though symmetric, orthonormal, smooth wavelet basis functions cannot have compact support, they exhibit exponential decay [4]. Symmetric basis functions do not introduce phase distortions, hence a better localization of the signal is achieved in the wavelet domain. Orthogonal spline wavelets were selected because of the following: (*i*) orthogonality is required by the subsequent statistical analysis; (*ii*) the resulting family of transforms use symmetric basis func-

tions; (*iii*) the use of splines reduces spectral overlap between resolution channels by increasing the degree of spline n [37]. Nevertheless, small spectral overlap increases data decorrelation [38], which raises the detection sensitivity. The decorrelation ability of orthogonal spline wavelets stems from the fact that splines with degree n yield $L = n + 1$ vanishing moments. The uncertainty principle limits the level of decorrelation across scale since the correlation suppression comes at the expense of a loss in spatial localization expressed in the decay rate of the filter coefficients. Besides, selecting the degree of splines depends to some extent on the assumed smoothness of the signal to be detected. Smooth wavelet bases are asymptotically near-optimal for estimating signals that may contain some points of discontinuity [6].

2.3.3 Denoising Images by Wavelet Shrinkage

Wavelet *shrinkage* refers to reconstructions obtained by WT of the original signal, followed by shrinking the empirical wavelet coefficients towards zero, followed by the inverse WT [8]. Wavelet-based shrinkage methods are nonparametric regression estimators that provide means of finding structure in a variety of data sets without imposing a parametric regression model. Wavelet-based denoising amounts to thresholding the detail coefficients in the wavelet domain. Gaussian spatial smoothing with a single kernel is widely applied in denoising neuroimaging data at the risk of missing to detect spatial features of the smoothing kernel size or lower. In contrast, smoothing by wavelet shrinkage allows locally adaptive bandwidth, so that the power to detect spatial features of varying extent is not constrained by the arbitrary choice of a single kernel size [3].

Thresholding in the wavelet domain was based on the assumption of white Gaussian noise. When the autocorrelation of the noise is unknown, a level-dependent threshold was suggested [24]. In the wavelet domain, the (white) noise is quite evenly spread out among all coefficients, while the signal is concentrated in a few coefficients only. Thus thresholding mostly affects the noise without disturbing the signal. This behavior is in contrast with traditional linear methods of smoothing, which perform noise suppression at the expense of significantly broadening the signal features. The WaveLab802 package [1] used for thresholding in the wavelet domain contains various shrinkage procedures, their discriminating characteristic being the amount of smoothing introduced in the denoised signal [18]. In fact, traditional methods of signal smoothing are just a bit more than diagonal projectors in the Fourier basis. Evidence exists that ideal diagonal projectors work better in the wavelet domain bases than (nonideal) projectors in the Fourier basis [8].

In order to select an optimal WT for wavelet shrinkage of brain signals, a sine with two clicks that mimics spatial inhomogeneities in fMRI data (Fig. 10a) as used by Donoho [5] was subject to several wavelet transforms and thresholding recipes. White noise normally distributed with variance 1 was additively added to this test signal (Fig. 10b). Both soft thresholding of the wavelet coefficients with an adaptively-chosen threshold and hard thresholding were initially compared. Soft thresholding is a spatially adaptive method for estimation of functions from noisy data, which is a nearly-ideal method of spatial adaptation in some well-defined theoretical sense [7]. Soft thresholding closely mimics an ideal diagonal projector in a wavelet basis [7]. The compression abilities of the wavelet bases are responsible for the mean-squared error advantages of the wavelet shrinkage. The results are presented for the case of wavelet shrinkage of the modified sine by the periodized symmetric biorthogonal wavelet. Both the wavelet coefficients of the original signal (Fig. 10c) and their noisy version (Fig. 10d) are presented stratified by scale and location. The significant coefficients occur in locations of significant spatial variability. For locally smooth functions, the coefficients tend to be small at fine scales. The noise in the empirical wavelet coefficients is a white noise and, therefore, about the same in every coefficient. The signal exceeds the noise level only in locations of significant spatial variability (Fig. 10d).

We applied a method proposed by Donoho and Johnstone [5] to find the threshold that minimizes the estimate of the mean squared error (MSE). The approach equates to applying a soft thresholding nonlinearity, with the threshold selected by the Stein's unbiased risk estimate (SURE) in the interval $[0, \sqrt{2log(n)}]$, where $n = 2^J$ is data number and J is the number of scales. This was proved to posses various optimality properties for MSE estimation. The SURE shrinkage carried out the best reconstruction of the original signal both in terms of noise suppression and sharp structure preservation in the neighborhood of the highly-variable spatial components (Fig. 10e,f).

2.4 Multiple Hypotheses Testing

Inferential analysis of functional neuroimaging data entails multiple testing in which a large number of correlated test statistics must be assessed. Consider a random variable with a large number of possible outcomes, say V. A statistical parametric map of V voxels consists of the *p*-values p_i, $i = 1, 2, ..., V$. Given a distribution of outcomes, a *p*-value is the probability

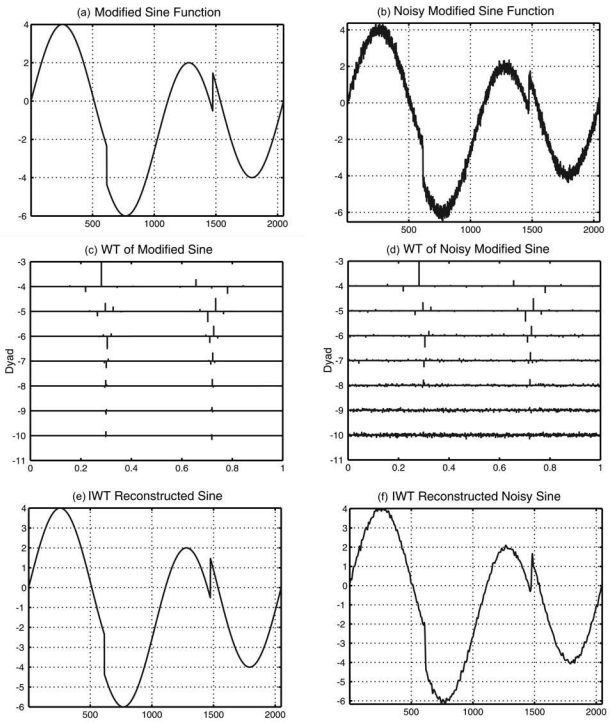

Figure 10: Wavelet shrinkage of the modified sine by the periodized symmetric biorthogonal wavelet; (a) A sinusoid with two clicks *1D* test signal; (b) White noise normally distributed with variance 1 superimposed; (c) Wavelet coefficients of the test signal stratified by scale and location; (d) Empirical wavelet coefficients of the noisy test signal; (e) Reconstruction of the modified sine function; (f) Reconstruction of the noisy modified sine function.

of getting an outcome at least as extreme as the one observed when the null hypothesis H_0 is correct. We hereafter discuss the one-sample *t*-test only, though SPM allows for many statistical tests. The temporal noise in fMRI data is assumed to be Gaussian distributed, $\mathcal{N}(\mu, \sigma^2)$. The null hypothesis states that $\mu = 0$. Testing for increased activation entails an one-side test: H_1 states that $\mu > 0$. Since σ^2 of the temporal noise distribution is unknown, it must be estimated via the sample variance s^2, which can be calculated from the residuals. Testing for increased activation is carried out via a *t*-test using these estimates. BOLD contrasts are constructed as linear combinations of the rows of the parameter matrix of the linear model ($\boldsymbol{\beta}$ in Fig. 13 and eq. 25), each of which is an image of V voxels, and their values are *t*-distributed. If a *t*-value in the BOLD contrast is in the upper $\alpha\%$ of the distribution, its *p*-value is bellow α. In other words, a small *p*-value provides strong evidence against the null hypothesis. Active voxels are those with *p*-values bellow a significance level α. In the case of one test, a common value is $\alpha = 0.05$, which is the probability of erroneously rejecting H_0. However, for V simultaneous tests with the same significance level α, approximately αV detected activations will be false positives since a large number of type I errors would be expected in null data. The problem of false positives with multiple statistical tests amounts to find an appropriate threshold for the *t*-values (or, equivalently, *z*-scores), so that we can be confident that the remaining suprathresholded *t*-values (or *z*-scores) are sufficiently high to be expected by chance.

2.4.1 Familywise Error

Consider image data on a *3D* regular or irregular lattice and, following a modelling process, assume an image of a test statistics $T = \{T_i\}$. Here T_i denotes the value of the statistic image at the spatial location i, $i \in \mathcal{V} = \{1, 2, ..., V\}$, where V is the number of voxels in the brain. Let $H = H_i$ be a hypothesis image such as $H_i = 0$ indicates that the null hypothesis holds at voxel i, and $H_i = 1$ indicates that the alternative hypothesis holds. The complete null case $H_i = 0, \forall i$, is indicated by H_0. A *decision* to reject the null for voxel i will be expressed as $\hat{H}_i = 1$ and not rejecting as $\hat{H}_i = 0$, the null distribution of T_i as F_{0,T_i}, and the image of *P*-values as $P = \{P_i\}$. The test is considered unbiased and all distributions are assumed continuous.

False positives must be controlled over all tests, but there is not a single measure of type I error in multiple hypotheses testing. The standard measure of type I errors in multiple testing is *familywise error rate* (FWE), which specifies the chance of false positives. The number of efficient FWE multiple testing procedures for fMRI is limited by the spatial dependence of functional neuroimaging data.

A valid test at α significance level at location i corresponds to a rejection threshold u where $\Pr\{T_i \geq u \mid H_i = 0\} \leq \alpha$. The central task in multiple hypotheses testing is to find a threshold u that controls some measure of false positives across the entire image. The *weak* control of FWE requires that false positives are controlled under the complete null hypothesis H_0 only:

$$\Pr\left(\bigcup_{i \in \mathcal{V}} \{T_i \geq u\} \mid H_0\right) \leq \alpha_0 \qquad (22)$$

where α_0 is the nominal FWE. The *strong* control of FWE requires that false positives are controlled for

any subset $\mathcal{V}_0 \subset \mathcal{V}$ where the null hypothesis holds:

$$\Pr\left(\bigcup_{i\in\mathcal{V}_0}\{T_i \geq u\} \mid H_i = 0,\, i \in \mathcal{V}_0\right) \leq \alpha_0 \quad (23)$$

Statistical significance of weak control implies that H_0 is false only and does not permit localization of individual significant voxels. In contrast, statistical significance of strong control allows rejection of individual H_i's while controlling the FWE at all non-significant voxels. Since localization is essential in neuroimaging, strong control of FWE is considered hereafter.

Basically, there are two broad classes of FWE control: (*i*) Bonferroni inequality-based methods, and (*ii*) maximum statistic distribution-based methods [33]. A statistical adjustment that is called Bonferroni correction effectively raises the standard of proof when a wide range of hypotheses are simultaneously considered. If V outcomes are tested rather than one, the significance level α is replaced by α/V, which will ensure that the proportions of false positives is still less than α in any subset $\mathcal{V}_0$ of the simultaneous tests $\mathcal{V}$. Consequently, Bonferroni correction has strong type I error (false positives) control, meaning that rejecting H_0 in a certain region of the brain is evidence for activation in that region. By reducing the probability of rejecting the null hypothesis, Bonferroni correction decreases the number of true positives, which introduces type II errors (false negatives).

2.4.2 False Discovery Rate

Another metric for measuring type I errors is the *false discovery rate* (FDR), which is defined as the expected proportion of rejected hypotheses that are false positives [14]. FDR controlling procedures are more powerful then FWE measures, yet still control false positives in a useful manner. FDR is an alternative to FWE, which does not require spatial smoothness. Instead of controlling the chance of *any* false positives (as Bonferroni or GRF do), FDR controls the expected proportion of false positives (i.e., type I errors) among suprathreshold voxels (i.e., rejected null hypotheses):

$$\mathrm{FDR} = \mathrm{E}\left\{\frac{\#false\,positives}{\#H_0\,rejected}\right\} \quad (24)$$

where E denotes expectation. $FDR = 0$ if the rejected null hypotheses (i.e., number of detections) = 0.

3 Results and Discussion

3.1 Data Acquisition and Preprocessing

One healthy right-handed subject was selected for single-shot MR FEEPI scanning at 1.5 T magnetic field while performing a block-based visual task during 12 identical sessions of 228 s each. Both the acquisition and the reconstruction matrices were $64 \times 64 \times 35$ and the voxel size was $3.8 \times 3.8 \times 3.75$ mm^3. In all sessions, 80 volumes were acquired at $TR = 3$ s, and the first 8 volumes were discarded to minimize the $T1$ saturation effects. A flashing checkerboard was presented in blocks of 24 s followed by 24 s of fixation, starting with activation.

All data were subject to some preprocessing steps: (*i*) acquisition time correction, (*ii*) realignment (i.e., movement correction) and coregistration, and (*iii*) spatial normalization to stereotaxic space [43]. For analysis in both spatial domain (SPM) and wavelet domain (WT), the design matrix was built up in SPM2, which also included a model for the hemodynamic response function (HRF).

3.2 Noise Models for fMRI Data

Most of the standard statistical tests in fMRI assume Gaussian distributed noise, though MR magnitude image data was found to obey a Rician distribution [16], [10], [39]. Unlike the additive Gaussian noise, the Rician noise is multiplicative (i.e., signal-dependent), which makes it difficult to separate from the signal. The Rician noise is especially problematic in low SNR regimes $(\mathrm{SNR} < 2)$ where it causes random fluctuations, and introduces a signal-dependent bias to data that reduces the image contrast [34]. The noise distribution is nearly Gaussian for $\mathrm{SNR} > 2$ [16]. Therefore, wavelet-based noise removal methods that adapt to variations of both signal and noise are attractive for filtering out the Rician noise. However, since the BOLD contrast is analyzed as the difference between two MR images (e.g., active minus baseline) both containing Rician distributed noise, the distribution of noise appears symmetric and closely approximates a Gaussian curve. Moreover, for very low signal intensities, based on Kolmogorov-Smirnoff (KS) test, a deviation from Gaussianity was evaluated statistically significant in very large images only [53].

The main task in image denoising is to maximally suppress noise along with preserving as much as possible of the image features. In a wavelet representation, wavelet coefficients carry both time and frequency information, as the basis functions vary in position and scale. The BOLD effect (due to spatial extent of neuronal events) and most preprocessing (particularly spatial smoothing) entail spatial au-

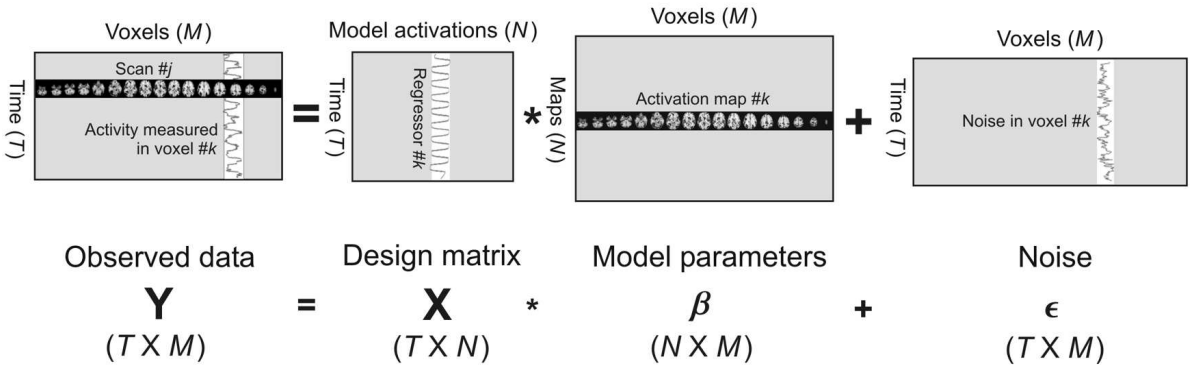

Figure 11: GLM of fMRI data.

tocorrelation. Accordingly, the optimal wavelet basis functions were selected by simulations of two types of spatial correlation: (*i*) white noise, and (*ii*) $1/f$ noise with $1/f$ power spectrum. In the orthonormal wavelet domain, most image information is contained in the largest wavelet coefficients, while the white noise is uniformly spread out across all coefficients. An efficient denoising method sets the smallest coefficients to zero and shrinks the remaining ones above a certain threshold.

3.3 SPM Analysis

In SPM, a convolution with an isotropic symmetric Gaussian kernel is applied to preprocessed data prior to statistical analysis. Apart form its benefits, Gaussian filtering degrades the image resolution and complicates the statistical analysis since the noise can no longer be considered independent. The inferential methods test voxelwise specific hypotheses about the expected changes in BOLD response. These changes are specified as regressors of interest in a (multiple) linear regression framework and their relative weights are given by the regression coefficients (i.e., model parameters). The general linear model (GLM) is applied voxelwise so that the observed fMRI time series at each voxel are linearly modeled as a superposition of model time courses of activation and Gaussian errors (Fig. 11).

Let the matrix $\mathbf{Y}_{[T\times V]}$ denote the fMRI data acquired in the experiment, where each matrix element y_{ij} denotes the observed value at time $i, i = 1, 2, ..., T$ and voxel location $j, j = 1, 2, ..., V$, V is the number of voxels in a volume, and T is the number time points (i.e., number of full volume scans). Then the linear model gives:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{25}$$

where $\mathbf{X}_{[T\times N]}$ is the design matrix having the regressors as N column vectors. The row vectors of the matrix $\boldsymbol{\beta}_{[N\times V]}$ are the model parameters of the effects of interest, and the elements in the matrix $\boldsymbol{\epsilon}_{[T\times V]}$ are the residuals (i.e., errors) of each voxel in each scan. A maximum likelihood (ML) estimate for the model parameters $\boldsymbol{\beta}$ is found by the least squares method (LSM) assuming no temporal correlations in the data:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} \tag{26}$$

The parameter selection is done by a contrast vector $\mathbf{c}$, which compares one or multiple parameter values. If a model of the residuals $\boldsymbol{\epsilon}$ exists, then the statistical significance of the regression coefficients and, implicitly, of the modeled hemodynamic changes can be calculated in each voxel via hypothesis testing. The activation image reconstruction is carried out using a voxel-specific statistic that tests hypotheses on dynamics. The statistical inference relies on continuous Gaussian RFT and takes into account the dependencies introduced by the Gaussian spatial smoothing. The average number of resels (resolution elements) available in the data after smoothing is defined as [50]:

$$\#resels = \frac{V}{\prod_{i=1}^{3} FWHM_i} \tag{27}$$

where $FWHM_i$ is the Gaussian full width at half maximum of the *i*-th dimension, $i = 1, 2, 3$. The voxelwise test statistics form summary images known as *statistical parametric maps*, which are commonly assessed for statistical significance against the null hypothesis (e.g., no activation). The resulting map of a statistic is a representation of the spatial distribution of functional activity induced by the experimental task.

Denoising by spatial Gaussian filtering was carried out with smoothing kernels of $4 \times 4 \times 3.75$ mm^3, $8 \times 8 \times 3.75$ mm^3, and $12 \times 12 \times 3.75$ mm^3. The shape of the detected activation areas after denoising should not significantly differ from that obtained

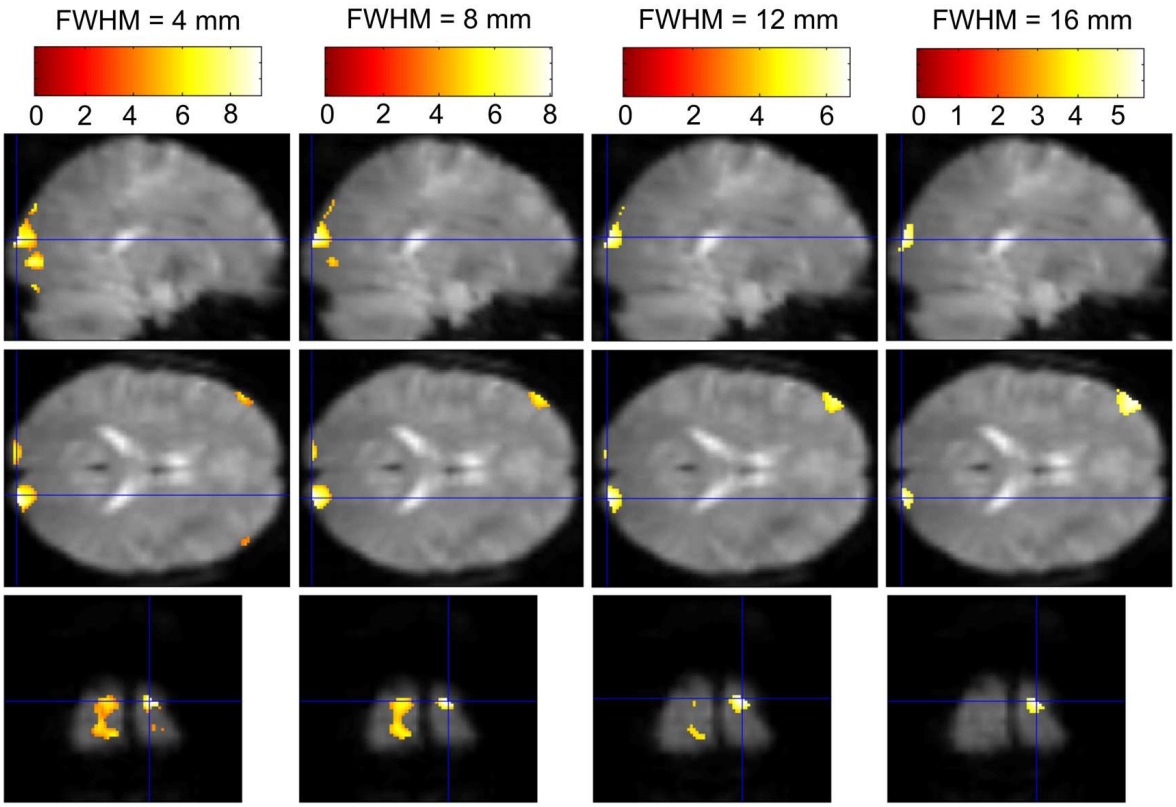

Figure 12: Activation parametric maps obtained by SPM of spatially Gaussian smoothed fMRI data after FDR thresholding with $q = 0.05$.

from the original time series. Similarly to artificially generated fMRI series, the active regions detected by SPM exhibited elliptic shapes with increasingly larger FWHMs of Gaussian kernels (Fig. 12). The denoised activation images were compared in all cases with the activation map of the original raw data with minimum preprocessing (i.e., acquisition and movement corrections only) as a reference (Fig. 13-left).

3.4 Wavelet Analysis

A large variety of non-redundant orthogonal WTs were tested and several denoising methods were applied in each wavelet subband to denoise both synthetic and real-life MR images corrupted by Rician noise. In agreement with [26], for a wide range of input noise levels, the orthogonal fractional (α, τ)-B-splines gave the best peak SNR (PSNR), as compared to some standard wavelet bases (e.g., Daubechies, symmlets, and coiflets).

The DWT was implemented through an iterated filterbank [27] by means of two types of wavelet families: (*i*) the separable *3D* fractional-spline wavelets, and (*ii*) $2D + Z$ quincunx wavelets [50]. The high frequency information contained in the data was preserved in the wavelet subbands, contrarily to spatial smoothing with a Gaussian kernel in SPM. Then the GLM (as in SPM) was applied to the time series of each wavelet coefficient. Significantly non-zero clusters of wavelet coefficients were identified using Kolmogorov-Smirnoff statistics. The activation pattern was spatially localized by the IDWT of the thresholded coefficient map and directly compared with the activation images obtained by statistical inference in the spatial domain (Fig. 16). Less spatially variable *t*-statistic images and lower thresholds ($p < 0.05$) enforced better detection of activation and improved the physiological relevance of statistical inference.

Following [50], the functions employed throughout as symmetric scaling functions in the DWT were the separable *3D* fractional-splines that depend on two parameters: (*i*) the degree, which is a continuous-varying parameter $\alpha > -1/2$, and (*ii*) the shift parameter τ. Their expression in the Fourier domain is:

$$\hat{\beta}^{\alpha} = \left[\frac{1 + z^{-1}}{2}\right]^{\alpha+1} \tag{28}$$

For the symmetric functions $\tau = 0$, hence their two-

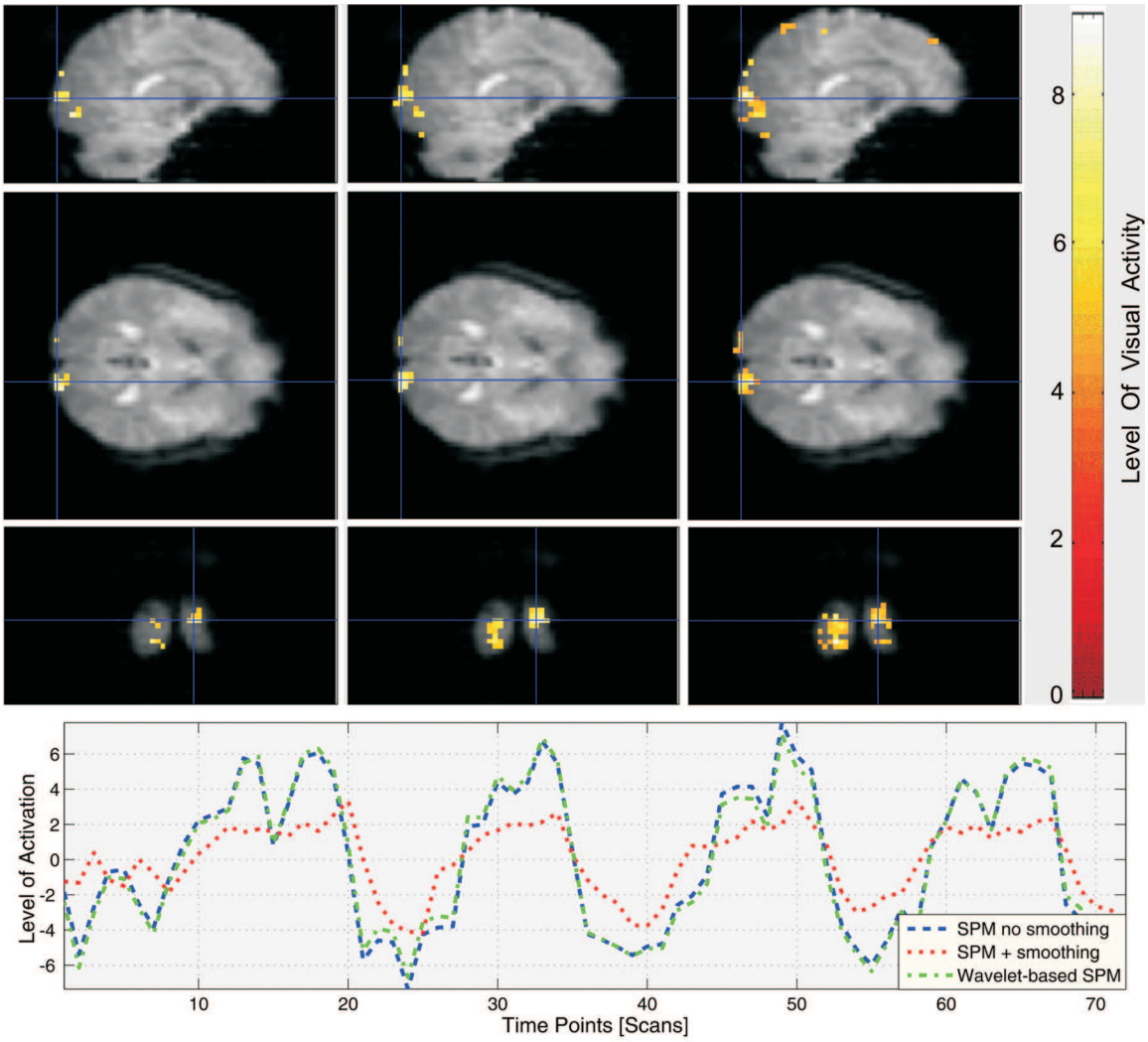

Figure 13: (a) Activation parametric maps generated by: minimally preprocessed raw fMRI data (left); SPM analysis with Gaussian smaoothing (mid); wavelet-based SPM analysis; and their corresponding time courses of activation (bottom). In all cases, the multiple hypothesis testing was controled by FDR thresholding ($q = 0.05$).

scale relationship is given by the scaling filter:

$$B^{\alpha}(z) = \sqrt{2}\left[\frac{1-\exp\left(-j\omega\right)}{j\omega}\right]^{\alpha+1} \qquad (29)$$

Both orthogonal and biorthogonal transforms can be constructed, though of special interest are the *dual* filters. The analysis part consists of the pure B-spline filter, whereas the synthesis part consists of the dual part, which is not so well localized. The fractional-spline wavelet transforms were implemented using the FFT. The use of non-separable *2D* quincunx scheme for the XY-planes [12], combined with a separable transform along the *Z*-direction, entails a quincunx subsampling matrix:

$$\mathbf{D} = \begin{Vmatrix} 1 & 1 \\ 1 & -1 \end{Vmatrix} \qquad (30)$$

Since $det\mathbf{D} = -2$, each *2D* iteration introduces only 2 subbands, as opposed to 4 in the case of a *2D* separable transform.

The *3D* volumes, each consisting of 35 axial planes of 64×64 voxels were transformed plane-by-plane to the wavelet domain. The wavelet decomposition level was set to 4. The wavelet analysis lead to similar activation patterns as in SPM based on Gaussian random fields. We found bilateral activation within a network of visual responsive regions including the inferior occipital gyrus, fusiform gyrus, superior temporal sulcus, amygdala, inferior frontal gyrus, and orbitofrontal cortex (Fig. 13-right). A rather ad-hoc threshold was required in the wavelet domain after the reconstruction of the coefficient map in order to put apart the activated and non-activated voxels. Though the back-projected results were de-

ployed of a precise statistical meaning in the spatial domain, the wavelet analysis yielded activation maps of higher resolution when using the coefficients from the high-pass subband too.

For synthetic fMRI-like time series, both the heavy Gaussian spatial smoothing and the more smoothing wavelet-based denoising schemes introduced severe deformations and blurred the edges of the activated regions introducing false negatives (type II errors). In contrast, for low SNR's, the less smoothing methods, both Gaussian filtering and wavelets, generated false positives (type I errors). In the mid SNR range, wavelet-based denoising methods led to less errors comparatively with Gaussian smoothing.

For real fMRI data, only the smallest Gaussian smoothing kernel yielded reliable results. The wider smoothing kernels yielded much larger detected areas (meaning more type I errors) or completely missed the active regions smaller than the kernel, in contrast to those obtained via less smoothing wavelet denoising methods. The bottom line is that wavelet-based denoising methods by introducing relatively little smoothness are generally preferably over Gaussian *spatial* smoothing for denoising fMRI time series. We may speculate that wavelet denoising may prove a good alternative for *temporally* smoothing as well.

4 Conclusion

Signal denoising methods that introduce heavy smoothness are better suited for low SNR input signals yielding the highest gain in SNR. For images of reasonable quality (i.e., with relatively high SNR input), the denoising schemes that produce heavy smoothness are not recommended because of introducing serious deformations of the objects in the image. Clearly, the more smoothing applied, the larger the deformations, no matter the denoising scheme applied. Wavelet methods perform as well as Gaussian smoothing for low SNR's, and better than Gaussian smoothing for higher SNR's. Wavelet-based denoising methods, by introducing less smoothing, preserve the sharpness of images and retain the original shapes of the active regions. Virtually for all wavelet-based denoising methods, the output SNR is a linear function of the input SNR, that is, the wavelet methods, contrarily to Gaussian smoothing, improve the SNR of the input images that already have a high SNR.

For both Gaussian and wavelet-based smoothing schemes, the difference in performance is relatively smaller for $1/f$ noise than for white noise. In the case of white noise and low SNR's, the less smoothing wavelet methods yield relatively lower output SNR's, which indicate that smoothness (i.e., discarding image features) is required to improve images with very low SNR's.

Previously reported data indicated that wavelet-based methods are preferable for denoising images with SNR $>$ 2 dB and exhibit maximum efficiency for SNR $>$ 10 dB [53]. The preliminary findings so far pointed out that the methods producing smooth images introduce more false positives. The less smoothing wavelet-based methods, though generating more false negatives, produce a smaller total number of errors than (spatial) Gaussian smoothing.

Wavelet-based methods provide a naturally multiscale alternative to single scale Gaussian spatial smoothing as widely used before hypothesis testing. Scale-varying wavelet-based methods for hypothesis testing of brain activation maps circumvent the need to specify a priori the size of signals expected and, therefore, the optimal choice of the smoothing kernel required by Gaussian filtering. Due to the smoothness of the wavelet representation, the estimated statistical parameter maps reveal more compact regions of activation than their counterparts obtained by statistic testing in the spatial domain.

Wavelet analysis is optimal in terms of detecting transients events in fMRI time series and adapts well to local or nonstationary features in data within scales of the decomposition. Wavelet-based methods are likely to provide an overall richer characterization of distributed brain activation.

Acknowledgements: The author is grateful for brain storming with Dr. Vince Calhoun (Associate Prof., Department of Psychiatry, Yale University and Director, Medical Image Analysis Laboratory, Olin Neuropsychiatry Research Center, Institute of Living, New Haven, CT, USA) and Dr. John Ashburner (Wellcome Department of Imaging Neuroscience, Institute of Neurology, University College London, UK) for clearing up the subtleties and commenting the latest updates in SPM5.

References:

[1] J.B. Buckheit and D.L. Donoho, "Wavelab and Reproducible Research." Dept. Statist., Stanford Univ., Stanford, CA, 1995. Available at: http://www-stat.stanford.edu/ wavelab

[2] E. Bullmore, "Wavelets and functional MRI," in *Mathematics in Brain Imaging*, IPAM, UCLA, July 21, 2004, available on-line.

[3] E. Bullmore, J. Fadili, M. Breakspear, R. Salvador, J. Suckling, M. Brammer, "Wavelets and statistical analysis of functional magnetic resonance images of the human brain," *Statistical*

Methods in Medical Research, vol. 12, no. 5, pp. 375-399, Oct. 2003.

[4] A. Cohen, I. Daubechies, and J. C. Feauveau, "Biorthogonal bases of compactly supported wavelets," *Commun. Pure Appl. Math.*, vol. 45, pp. 485560, 1992.

[5] D.L. Donoho and I.M. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," *Journal of the American Statistical Association*, vol. 90, no. 432, pp. 12001224, 1995.

[6] D. L. Donoho, "Unconditional bases are optimal bases for data compression," *Appl. Comput. Harmonica Anal.*, vol. 1, pp. 100-115, 1993.

[7] D.L. Donoho and I.M. Johnstone, "Ideal spatial adaptation via wavelet shrinkage," Tech. Report, Statistics, Univ. Stanford, 1992.

[8] D.L. Donoho,"Nonlinear wavelet methods for recovery of signals, densities, and spectra from indirect and noisy data," *Proc. Symposia in Applied Mathematics*, vol. 00, 1993.

[9] J.H. Duyn, C.T.W. Moonen, R.W. de Boer, G.H. van Yperen, and P.R. Luyton, "Influx versus deoxyhemoglobin effects in BOLD functiona MRI using gradient echoes at 1.5 T", *NMR Biomed.*, vol. 7, pp. 83-88, 1994.

[10] W.A. Edelstein, P.A. Bottomley, and P.M. Pfeifer, "A signal-to-noise calibration procedure for NMR imaging systems," *Med. Phys.*, vol. 11, pp. 180185, 1983.

[11] F. Esposito, E. Seifritz, E. Formisano, R. Morrone, T. Scarabino, G. Tedeschi, S. Cirillo, R. Goebel, and F. Di Sallee, "Real-time independent component analysis of fMRI time-series," *Neuroimaging*, vol. 20, pp. 2209-2224, 2003.

[12] M. Feilner, M. Jacob, and M. Unser, "Orthogonal quincunx wavelets with fractional orders," *Proc. of the 2001 IEEE Intl. Conf. on Image Processing* (ICIP'01), vol. I, pp. 606-609, Thessaloniki, Greece, 2001.

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

[14] C.R. Genovese, N.A. Lazar, and T.E. Nichols, "Thresholding of statistical maps in functional neuroimaging using the false discovery rate," *NeuroImage*, vol. 15, pp. 772-786, 2002.

[15] G.L. Gernstein, P. Bedenbaugh, and A.M.H.J. Aertsen, "Neuronal assemblies," *IEEE Trans. on Biomedical Engineering*, vol. 36, pp. 4-14, 1989.

[16] H. Gudbjartsson and S. Patz, "The Rician distribution of noisy MRI data," *Magn. Reson. Med.*, vol. 34, pp. 910-914, 1995.

[17] J. V. Hajnak, R. Myers, A. Oatridge, J. E. Schwieso, I. R. Young, and G. M. Byder, "Artifacts due to stimulus correlated motion in functional imaging of the brain", *Magn. Reson. Med.*, vol. 31, pp. 283-291, 1994.

[18] M. Hilton, T. Ogden, D. Hattery, G. Eden, and B. Jawerth, "Wavelet processing of functional MRI data," in *Wavelets in Biology and Medicine* (A. Aldroubi and M. Unser, Eds.) CRC Press, Boca Raton, FL, 1996.

[19] B. Horwitz and O. Sporns, "Neural modeling and functional neuroimaging," *Hum. Brain Map.*, vol. 1, pp. 269-283, 1994.

[20] B. Horwitz, K. Frinston, and J. G. Taylor, "Neural modeling and functional brain imaging: An overview," *Neural Networks*, vol. 13, nos. 8-9, pp. 829-846, 2000.

[21] P.J. Huber, "Projection pursuit," *Annals of Statistics*, vol. 13, pp. 435-475, Feb. 1985.

[22] P.J. Huber, "Huge data sets," *Proceedings, Compstat 1994* (R. Dutter and W. Grossman, Eds.), pp. 3-13, Physica Verlag, Heidelberg, 1994.

[23] M. Jansen and A. Bultheel, "Empirical bayes approach to improve wavelet thresholding for image noise reduction," *J. Amer. Statist. Assoc.*, vol. 96, pp. 629639, 2001.

[24] I.M. Johnstone and B.W. Silverman, "Wavelet threshold estimators for data with correlated noise," *J. Roy. Statist. Soc.*, vol. 59, pp. 319351, 1997.

[25] S. Kim, K. Hendrich, X. Hu, H Merkle, and K. Ugurbil, "Potential pitfalls of functional MRI using conventional gradient-recalled echo techniques", *NMR Biomed.*, vol. 7, pp. 69-74, 1994.

[26] F. Luisier, T. Blu, B. Forster, and M. Unser, "Which wavelet bases are the best for image denoising ?" Biomedical Imaging Group (BIG), Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland.

[27] S. Mallat, "A theory for multiresolution signal decomposition: The wavelet decomposition," *IEEE Trans. Patter. Anal. Mach. Intell.*, vol. 11, pp. 674-693, 1989.

[28] C.T.W. Moonen, "Imaging of human brain activation with functional MRI", *Biological Psych.*, vol. 37, pp. 141-143, 1995.

[29] R. Mutihac, "Wavelet denoising versus Gaussian spatial smoothing of fMRI data," *European Society for Magnetic Resonance in Medicine*

and Biology, 22nd Annual Meeting (ESMRMB 2005), Basel, Switzerland, Functional MRI, S204, 2005.

[30] R. Mutihac, "Thresholding brain activation maps - Multiple hypotheses testing versus wavelet shrinkage," *12th Annual Meeting of the Organization for Human Brain Mapping* (OHBM 2006), Florence, Italy, 984, 2006.

[31] R. Mutihac, "Exploratory analysis of functional MRI data," The Abdus Salam ICTP, IC/2004/69, pp. 1-27, 2004.

[32] R. Mutihac, A. Cicuttin, K. Jansen, and R.C. Mutihac, "An essay on Bayesian inference and Maximum Entropy," *Romanian Biotechnology Letters*, vol. 5, no. 2, pp. 83-114, 2000.

[33] T. Nichols, S. Hayasaka, and T. Wager, "Controlling the familywise error rate in functional neuroimaging: A comparative review," *Statistical Methods in Medical Research*, vol. 12, no. 5, pp. 419-446, 2003.

[34] R.D. Nowak, "Wavelet-based Rician noise removal for magnetic resonance imaging," *IEEE Transactions on Image Processing*, vol. 8, no. 10, pp. 1408-1419, Oct. 1999.

[35] S. Ogawa, T.M. Lee, A.S. Nayak, and P. Glynn, "Oxygenation-sensitive contrast in magnetic resonance image of rodent brain at high magnetic fields," *J. Magnetic Resonance in Medicine*, vol. 14, pp. 68-78, 1990.

[36] J.-B. Poline, K.J. Worsley, A.C. Evans, and K. Frinston, "Combining spatial extent and peak intensity to test for activations in functional imaging," *Neuroimage*, vol. 5, pp. 83-96, 1997.

[37] U.E. Ruttimann, M. Unser, R. Rawlings, D. Rio, N. Ramsey, V. Mattay, D. Hommer, J. Frank, and D. Weinberger, "Statistical analysis of fMRI data in the wavelet domain," *IEEE Trans. Med. Imaging*, vol. 17, no. 2, pp. 142-154, Feb. 1998.

[38] U.E. Ruttimann, M. Unser, D. Rio, and R.R. Rawlings, "Use of the wavelet transform to investigate differences in brain PET images between patients," *Proc. SPIE*, vol. 2035, *Mathematical Methods in Medical Imaging II*, pp. 192-203, San Diego, CA, 1993.

[39] J. Sijbers, A.J. den Dekker, J. Van Audekerke, M. Verhoye, and D. Van Dyck, "Estimation of the noise in magnitude in MR images," *Magn. Reson. Imag.*, vol. 16, pp. 8790, 1998.

[40] J. Skilling, D.R.T. Robinson, and S.F. Gull, "Probabilistic displays," in *Maximum Entropy and Bayesian Methods in Science and Engineering* (W.T. Grandy and L.H. Schick, Eds.), pp. 365-368, Kluwer Academic Publishers, Laramie, Wyoming, 1991.

[41] F.T. Sommer, J.A. Hirsch, and A. Wichert, "Theories, data analysis, and simulation models in neuroimaging - An overview," in E*xploratory Analysis and Data Modeling in Functional Neuroimaging*(F. T. Sommer and A. Wichert, Eds.), pp. 1-13, Neural Information Processing Series, The MIT Press, Cambridge, 2003.

[42] M. Spitzer, "Forward," in *Exploratory Analysis and Data Modeling in Functional Neuroimaging*, (F.T. Sommer and A. Wichert, Eds.), Neural Information Processing Series, The MIT Press, Cambridge, 2003.

[43] J. Talairach and P. Tournoux, *Co-Planar Stereotaxic Atlas Of The Human Brain*, Thieme Medical Publishers, New York, 1988.

[44] J.W. Tukey, "The future of data analysis." *Annals of Statistics*, vol. 33, pp. 1-67, 1962.

[45] F.E. Turkheimer, M. Brett, D. Visvikis, and V. J. Cunningham, "Multiresolution analysis of emission tomography images in the wavelet domain", *Journal of Cerebral Blood Flow and Metabolism*, vol. 19, no. 11, pp. 1189-1208, 1999.

[46] F.E. Turkheimer, J.A.D Aston, and V.J. Cunningham, "On the logic of hypothesis testing in functional imaging," *European Journal of Nuclear Medicine*, vol. 31, no. 5, pp. 725-732, 2004.

[47] M. Unser, P. Thevenaz, C. Lee, and U.E. Ruttiman, "Registration and statistical analysis of PET images using the wavelet transform," *IEEE Eng. Med. Biol. Mag.*, vol. 14, pp. 603-611, 1995.

[48] M. Unser, "Splines: A perfect fit for signal and image processing," *IEEE Signal Processing Mag.*, vol. 16, pp. 2238, Nov. 1999.

[49] M. Unser and T. Blu, "Fractional splines and wavelets," *SIAM Rev.*, vol. 42, pp. 4367, 2000.

[50] D. Van De Ville, T. Blu, and M. Unser, "Surfing the brain: An overview of wavelet-based techniques for fMRI data analysis," *IEEE Engineering in Medicine and Biology Magazine*, vol. 25, no. 2, pp 65-78, Mar.-Apr. 2006.

[51] D. Van De Ville, T. Blu, and M. Unser, "On the multi-dimensional extension of the quincunx subsampling matrix," *IEEE Signal Processing Letters*, vol. 12, no.2, pp. 112-115, 2005.

[52] D. Van De Ville, T. Blu, and M. Unser, "Integrated wavelet processing and spatial statistical testing of fMRI data," *NeuroImage*, vol. 23, no. 4, pp. 1472-1485, Dec. 2004.

[53] A.M. Wink and J.B.T.M. Roerdink, "Denoising functional MR images: A comparison of wavelet denoising and Gaussian smoothing," *IEEE Trans. Med. Imaging*, vol. 23, no. 3, pp. 374-387, March 2004.

[54] K.J. Worsley, A.C. Evans, S. Marrett, P. Neelin, "A three-dimensional statistical analysis for CBF activation studies in human brain," *Journal of Cerebral Blood Flow and Metabolism*, vol. 12, no. 6, pp. 900-918, 1992.

[55] K.J. Worsley, "Discussion of the paper by Lange and Zeger." *Appl. Statist.*, vol. 46, p. 25, 1997.

[56] K.J. Worsley, S. Marrett, P. Neelin, and A.C. Evans, "Searching scale space for activation in PET images," *Human Brain Map.*, vol. 4, pp. 74-90, 1996.

[57] K.J. Worsley, "Spatial smoothing of autocorrelations to control the degrees of freedom in fMRI analysis", OHBM, Toronto, Canada, 2005.

[58] S. Zeki, "Functional specialization in the visual cortex: the generation of separate constructs and their multistage integration," in *Signal and Sense* (G.M. Edelman, W.E. Gall, and W.M. Cowan, Eds.), pp. 85-130, John Wiley, New York, 1990.