Wavelets, Fourier Transform, and Fractals

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

Abstract: Wavelets and Fourier analysis in digital signal processing are comparatively discussed. In data processing, the fundamental idea behind wavelets is to analyze according to scale, with the advantage over Fourier methods in terms of optimally processing signals that contain discontinuities and sharp spikes. Apart from the two major characteristics of wavelet analysis - multiresolution and adaptivity to nonstationarity or local features in data - there is a related application in functional neuroimaging on the basis of the fractal or scale invariant properties demonstrated by the brain imaging data. Wavelets provide an orthonormal basis for multiscale analysis and decorrelation of nonstionary time series and spatial processes. The discrete wavelet transform (DWT) has applications to statistical analysis in functional magnetic resonance imaging (fMRI) like time series resampling by wavestrapping, linear model estimation, and methods for multiple hypothesis testing in the wavelet domain.

Key–Words: Wavelets, Fourier transform, fractals, multiresolution analysis, functional magnetic resonance imaging.

1 Introduction

In practice, data are acquired in a raw form and, consequently, of little immediate usefulness. 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 the 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. A common task is to find an adequate somewhat reduced representation of a multivariate data set for discovery, analysis, and recognition of patterns. Transform domain processing is a versatile approach widely used in data analysis. In data compression, the energy compaction properties of some classes of linear transforms are employed [25]. In data processing, noise reduction is performed by generalized filtering based on the fact that the underlying signal tends to get concentrated into few coefficients while the noise is spread out more evenly. The overall effect of working in the transform domain is an increase in the signal-to-noise ratio (SNR).

The WT of an image produces a multiresolution representation where each wavelet coefficient represents the information content of the image at a certain resolution in a certain position. 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).

Wavelet methods approach the analysis of statistical fields by estimating the signal at any resolution among the random fluctuations. The appearance of explicit orthonormal bases entailed significant implications on fMRI data analysis. Unlike the traditional Fourier bases, wavelet bases offer a degree of localization in space as well as frequency. This enables the development of simple function estimates that respond effectively to discontinuities and spatially varying degrees of oscillations in a signal, even when the observations are corrupted by noise. 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. Therefore, statistical maps are transformed using the DWT, the resulting coefficients are thresholded, and then denoised statistical maps are reconstructed [24]. [20]. Since the control of false positives is stringent in this application, the Bonferroni approach to wavelet thresholding was initially suggested.

The computation of the variance wavelet transform (WT) is more straightforward for statistical maps than it is for images: (*i*) the noise variance of the image is 1; and (*ii*) the correlation function is computed because 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 [28]. Variances of wavelet levels are computed by the product of the power function of the field with the power function of the wavelet filters [23].

In general, the most attractive mathematical properties of wavelets in applications are: *compact support* (i.e., they vanish outside a finite interval) and *continuous differentiability*. By truncating the wavelet coefficients below a certain threshold, the data become *sparsely* represented, which makes wavelets appropriate for data compression and quite similar to the topographic independent component analysis (ICA) [10] [19]. The interest in wavelets is also due to computationally efficient implementation of multiresolution analysis (MRA). Just as Fast Fourier Transform (FFT) algorithms made the Fourier Transform (FT) a practical tool for spectral analysis, the MRA has made the DWT a viable tool for computational time-scale analysis.

The continuous wavelet transform (CWT) is commonly applied in Physics, whereas the discrete wavelet transform (DWT) is more common in numerical analysis, signal and image processing. The main three directions of wavelet applications in imaging are: (*i*) image compression, (*ii*) image denoising, and (*iii*) image/contrast enhancement. The following stands for that.

2 Continuous Function Spaces

The space of all functions $f(t), t \in \mathbb{R}$ that are square integrable

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

is denoted by $\mathrm{L}_2(\mathbb{R})$ or, simply L_2. The space $\mathrm{L}_2(\mathbb{R})$ may be endowed with some properties:

1. Scalar (inner) product:

$$\langle f(t), g(t) \rangle = \int_{-\infty}^{+\infty} f(t)\overline{g(t)} \, dt \tag{2}$$

   where the integral is taken in the Lebesgue's sense. A space where a scalar product is defined is called a *unitary* space. A unitary space with finite dimension is called an *Euclidian* space.

2. Norm:

$$\|f(t)\| = \int_{-\infty}^{+\infty} |f(t)|^2 \, dt < +\infty \tag{3}$$

   A norm always exists in a unitary space due to the scalar product defined therein. However, a norm may be defined in a space which is not endowed with a scalar product. The energy of a function $f(t)$ is given by its squared L_2-norm:

$$\|f\|_{\mathrm{L}_2}^2 = \langle f, f \rangle_{\mathrm{L}_2} \tag{4}$$

   thus, the notation $f \in \mathrm{L}_2$ is equivalent to the statement that $\|f\|_{\mathrm{L}_2}^2$ is finite.

3. Distance:

$$d(f(t), g(t)) = \|f(t) - g(t)\| \tag{5}$$

   Spaces endowed with a distance are called *metric* spaces. A unitary space is always a metric space since it is equipped with a norm However, a distance may be defined without a norm.

4. Completness: A *metric* space is called *complete* if all Cauchy sequences converge to an element in the space. That is, if $M \neq \emptyset$ is a metric space and $x_k \in M, k \in \mathbb{N}$ and $d(x, y)$ is the distance function, and if the following holds:

$$\forall \epsilon > 0, \exists N \in \mathbb{N} : n, m > N \Rightarrow d(x_n, x_m) < \epsilon \tag{6}$$

   then

$$\exists x \in M : \lim_{k \to \infty} x_k = x \tag{7}$$

A *Hilbert* space is a *complete unitary* space that corresponds to the inner product:

$$\langle f, g \rangle_{\mathrm{L}_2} = \int_{-\infty}^{+\infty} f(t) g(t) \, dt \tag{8}$$

where the integral is taken in Lebesgue's sense.

A *Banach* space is a *complete* space endowed with a *norm* of finite or infinite dimension.

More generally, one defines the L_p spaces for $1 \leq p \leq +\infty$ (where L stands for Lebesgue) as the set of functions whose L_p-norm:

$$\|f\|_{\mathrm{L}_p} = \left(\int_{-\infty}^{+\infty} |f(t)|^p \, dt \right)^{1/p} \tag{9}$$

is finite. These are Banach spaces; there is no corresponding inner product whatsoever (except for $p = 2$).

2.1 Fourier Transform

The Fourier transform (FT) of $f(t)$ is denoted hereafter by $f(\omega)$; if:

$$f \in \mathrm{L}_1 \Rightarrow \mathcal{F}(f(t)) = f(\omega) = \int_{-\infty}^{+\infty} f(t) \, e^{-j\omega t} \, dt \tag{10}$$

The definition can be extended for functions $f \in \mathrm{L}_p$, as well as for generalized functions $\mu \in S'$, where S' stands for Schwarz' space of temporal distributions on $\mathbb{R}$ [13].

2.2 Wavelet Transform

Wavelets appeared as the consequence of interest shifting from *frequency analysis* to *scale analysis*, which entailed the design of mathematical structures that vary in scale. In approximation using superposition of functions, Fourier basis functions (*sines* and *cosines*) are non-local and stretch out to infinity, rendering them unsuited to analyze choppy signals [4]. In contrast, wavelets were developed 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.

Wavelet analysis procedure is to adopt some two continuously-defined functions:

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

$$\phi(x) = \sqrt{2} \sum_{k \in \mathbb{Z}} h(k) \phi(2x - k) \qquad (11)$$

   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):

$$\psi(x) = \sqrt{2} \sum_{k \in \mathbb{Z}} g(k) \phi(2x - k) \qquad (12)$$

   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:

$$\int_{-\infty}^{+\infty} \psi(x)\, dx = 0 \qquad (13)$$

The term "small" 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.

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

$$f(x) = \sum_{j \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} d_j(k) \psi_{j,k}(x) \qquad (14)$$

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

$$d_j(k) = \langle f, \tilde{\psi}_{j,k} \rangle_{L_2},\ j, k \in \mathbb{Z} \qquad (15)$$

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:

$$\langle \tilde{\psi}_{j,k}, \psi_{i,l} \rangle = \delta_{j-i} \cdot \delta_{k-l},\ \ i, j, k, l \in \mathbb{Z} \qquad (16)$$

The *biorthogonal* basis is generated by a single template (*biorthogonal* wavelet) $\tilde{\psi}(t)$ [3].

The wavelet coefficients are efficiently calculated using Mallat's algorithm [15], which is based on a hierarchical application of the filterbank decomposition (Fig. 1). The only constraints on the choice of the filters are the perfect reconstruction (PR) conditions; in this case:

$$\begin{aligned} \tilde{H}(z^{-1})H(z) + \tilde{G}(z^{-1})G(z) &= 1 \\ \tilde{H}(z^{-1})H(-z) + \tilde{G}(z^{-1})G(-z) &= 0 \end{aligned} \qquad (17)$$

The underlying decomposition algorithm is best described by two complementary filters h and g. We consider hereafter non-redundant dyadic orthogonal wavelet transforms only. The algorithm consists of an iterated orthogonal filterbank with an *analysis* and a *synthesis* part. In a one-level WT, in the decomposition (analysis) step, the digital signal $c_i(k)$ is split in two half-size sequences: an approximation part $c_{i+1}(k)$ and a detail part $d_{i+1}(k)$ by filtering it with a conjugate pair of low-pass ($\widetilde{H}(z^{-1})$) and high-pass ($\widetilde{G}(z^{-1})$) filters, respectively (Fig. 1). The results are subsequently down-sampled. During the reconstruction (synthesis) step that follows, the signal is reconstructed by up-sampling, filtering, and summation of the components. The analysis and synthesis procedures are flow-graph transposes of each other. This kind of 2-channel processing is one-to-one and reversible if the z-transforms of the filters satisfy the PR conditions [26]. Different transforms correspond to different sets of PR filters. For a signal vector of length N_0, the operations required by the WT are

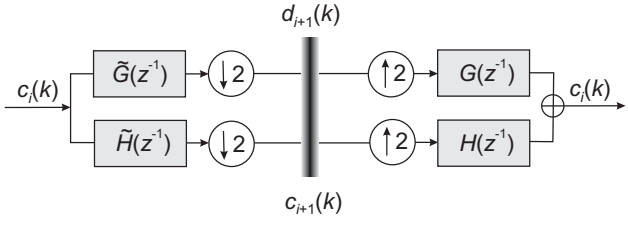

Figure 1: Decomposition and reconstruction quadrature mirror filters. Analysis part (left) and synthesis part (right) of the WT filterbank.

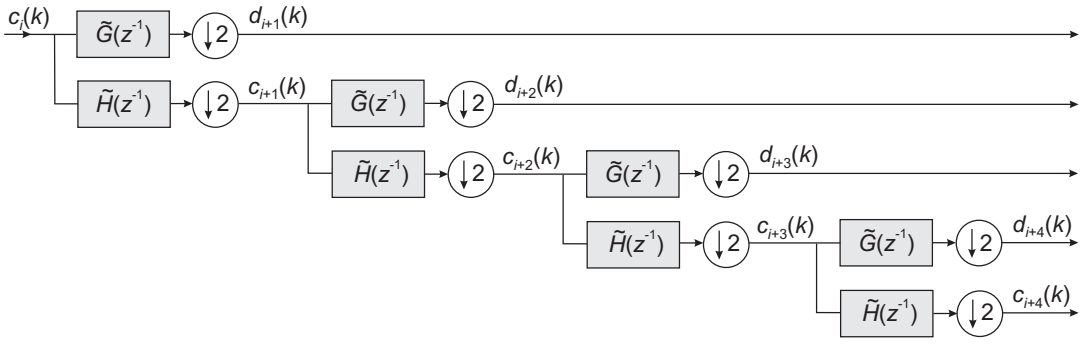

Figure 2: Blocks in recursive filterbank implementation of the wavelet transform.

$\mathcal{O}(N_0)$, as compared with the standard FFT complexity of $\mathcal{O}(N_0\, log N_0)$. In multilevel WT, rather than using a huge multichannel filterbank to encompass the full spectrum, the WT employs recursive 2-channel filterbanks; each subsequent c_i is split into an approximation c_{i+1} and detail d_{i+1} pair of coefficients. The inverse WT reconstructs each c_i from c_{i+1} and d_{i+1} (Fig. 2).

Spline bases posses the best approximation properties like the smallest L_2-error [25]. Due to their smoothness, splines are well localized in both time and frequency domains. Studies on wavelet application in neuroimaging data analysis emphasized the importance of symmetric wavelets and scaling functions that are free from phase distortions [20]. 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 [11]. As such, in wavelet analysis of fMRI time series, the preprocessed data are subject to spatial non-redundant DWT, rather than spatially convolved with a Gaussian kernel [17].

Orthogonal wavelet basis functions 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

$$\forall f \in L_2, f(x) = \sum_{j\in\mathbb{Z}}\sum_{k\in\mathbb{Z}} d_j(k)\psi_{j,k} + \sum_{j\in\mathbb{Z}}\sum_{k\in\mathbb{Z}} c_j(k)\phi_{j,k} \tag{18}$$

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:

$$d_j(k) = \langle f, \psi_{j,k}\rangle, \quad c_j(k) = \langle f, \phi_{j,k}\rangle, \tag{19}$$

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

$$f(x) = \sum_{j=1}^{J}\sum_{k\in\mathbb{Z}} d_j(k)\psi_{j,k} + \sum_{k\in\mathbb{Z}} c_J(k)\phi_{J,k} \tag{20}$$

Orthogonality imposes:

$$\tilde{H}(z) = H(z^{-1}) \quad \text{and} \quad \tilde{G}(z) = G(z^{-1}) \tag{21}$$

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:

$$G(z) = z \cdot H(-z^{-1}) \tag{22}$$

The filters must obey the quadrature mirror filter (QMF) conditions for PR (17), which, in terms of the low-pass filter h only, equate:

$$\begin{gathered} H(z)H(z^{-1}) + H(-z)H(-z^{-1}) = 2 \\ H(1) = \sqrt{2} \Leftrightarrow H(-1) = 0 \end{gathered} \tag{23}$$

The orthogonality property means that the inner product of the mother wavelet with itself is unity, and the inner products between the mother wavelet and the shifts and dilates of the mother wavelet are zero. The collection of shifted and dilated wavelet functions is called a *wavelet basis*. The grid in shift-scale space on which the wavelet basis functions are defined is called the *dyadic* grid. Any continuous function (e.g., signals or images) is uniquely projected onto the wavelet basis functions and expressed as a *linear combination* of the basis functions. The set of coefficients which weight the wavelet basis functions in representing any continuous function are referred to as the *wavelet transform* (WT) of the given function.

Remarkably, the WT yields an efficient representation for functions which have similar character to the functions in the wavelet basis. There are wavelet families of orthornormal basis functions like the biorthogonal wavelets or wavelet basis functions not orthogonal in any sense. The large number of known wavelet families and functions provides a rich space in which to search for a wavelet which optimally represent any function of interest.

The continuous-time interpretation of the wavelet decomposition is based on the fundamental concept of *scaling* (or *scale-varying*) function. 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. Decomposition of functions in terms of orthonormal basis functions is performed by a Fourier expansion as well. The wavelet basis functions have *compact support*, in contrast with the Fourier basis functions. That is, the wavelet basis functions are non-zero only on a finite interval, whereas the sinusoidal basis functions of the Fourier expansion are infinite in extent. The compact support allows the WT to adequately represent functions or signals which have *localized* features. This representation is suitable in data compression, signal detection and denoising. The point is that the structured component of a signal is well represented by a relatively few wavelet coefficients, whereas the unstructured component on the signal like noise projects quasi-equally onto all of the basis functions. Subsequently, the structured and unstructured parts of the signal are separated in the WT domain.

The decomposition (19) can be extended to *multiple dimensions* (e.g., *2D* or *3D*) by using tensor product 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:

$$\phi_{j,\mathbf{k}}(\mathbf{x}) = \prod_{i=1}^{q} \phi_{j,k_i}(x_i) \tag{24}$$

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 (24) 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:

$$b_i = \begin{cases} 1 & \text{if } \phi_{j,k_i} \text{ is replaced by } \psi_{j,k_i} \\ 0 & \text{otherwise} \end{cases}, \; i = 1, 2, ..., q \tag{25}$$

and

$$\varphi_{j,k_i} = \begin{cases} \psi_{j,k_i} & \text{if } b_i = 1 \\ \phi_{j,k_i} & \text{otherwise} \end{cases}, \; j \in \mathbb{Z}, \; i = 1, 2, ..., q \tag{26}$$

then the mixed tensor product wavelets can be rewritten [20]

$$w^m_{j,\mathbf{k}}(\mathbf{x}) = \prod_{i=1}^{q} \varphi_{j,k_i}(x_i), \quad m = 1, 2, ..., 2^q - 1 \tag{27}$$

with

$$m = \sum_{i=1}^{q} b_i 2^{i-1} \tag{28}$$

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. The corresponding multidimensional coefficients:

$$\begin{aligned} c_j(\mathbf{k}) &= \langle f, \phi_{j,\mathbf{k}} \rangle \\ d^m_j(\mathbf{k}) &= \langle f, w^m_{j,\mathbf{k}} \rangle \end{aligned} \tag{29}$$

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^1_{i+1}, d^2_{i+1}, and d^3_{i+1}, for horizontally, vertically, and diagonally oriented details, respectively. The biorthogonal scaling function and its corresponding wavelet function are also plotted (Fig. 3a). These wavelets are employed to run *2D* three-level WT of an axial MR brain slice (Fig 3b). Further on, the first three coarse approximation levels are separately shown in Fig. 4 for better visual comparison.

The time-frequency resolution difference between the FT and the WT is best revealed by looking at the basis function coverage of the time-frequency plane [27]. A windowed FT (WFT) is shown in Fig. 5a, where the window is a square wave that truncates the Fourier basis function (sine or cosine) to fit a particular window width. Because a single window size is used throughout, the resolution of the analysis is identical at all locations in the time-frequency plane. Contrarily, the window size varies in the WT, so that

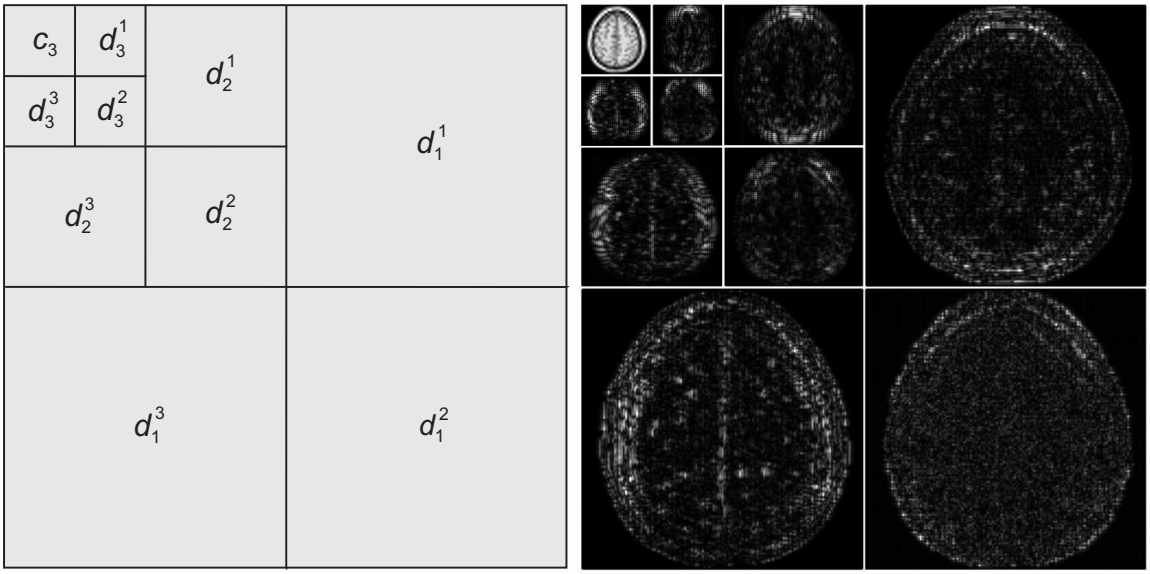

Figure 3: Approximation and detailed wavelet coefficients in a level-three *2D* WT (left); *2D* wavelet decomposition of a typical MR axial slice. A *coarse* (*approximation*) image at a resolution level L is represented by 2^L pixels in each direction. The *detail* images at a particular level L are produced by *horizontal*, *vertical*, and *diagonal* differences between successive levels. The set of coefficients consists of the lowest coarse level image and the higher level detail images. The original image is $2^8 \times 2^8$ pixels.

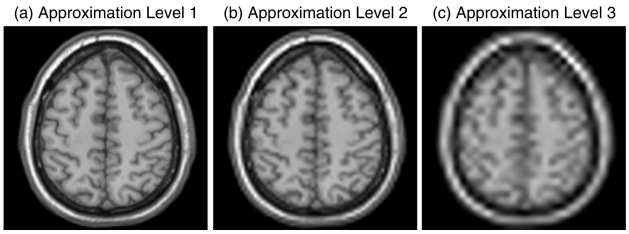

Figure 4: Coarse images of a typical MR axial slice at various approximation levels: (a) Approximation 1 (resolution level $L = 7$); (b) Approximation 2 ($L = 6$); (c) Approximation 3 ($L = 5$). All images are rescaled to the same size for better comparison.

sharp discontinuities are isolated by short high frequency basis functions and detailed frequency analysis is allowed by long low frequency ones (Fig. 5b). The FT employs sine and cosine basis functions only, whereas the set of wavelet transforms is infinite.

2.3 Families of Wavelets

There is a large variety of wavelet functions available. Various wavelet families differ by the compromise between their spatial compactness and smoothness. Some of particular interest in neuroimaging data analysis are presented hereafter (Fig. 6).

The Daubechies family of wavelets are only one of a number of wavelet families. Remarkably, the wavelet function (*mother wavelet*) is *orthogonal* to all functions which are obtained by *shifting* the mother right or left by an integer amount. Furthermore, the mother wavelet is orthogonal to all functions which are obtained by *dilating* or *stretching* the mother by a factor of 2^j and shifting by multiples of 2^j units. Due to the orthonormality of the Daubechies wavelets, any continuous function is uniquely projected onto the wavelet basis functions and expressed as a *linear combination* of the basis functions.

Symmlets are 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 [26], 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 [3]. 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 result-

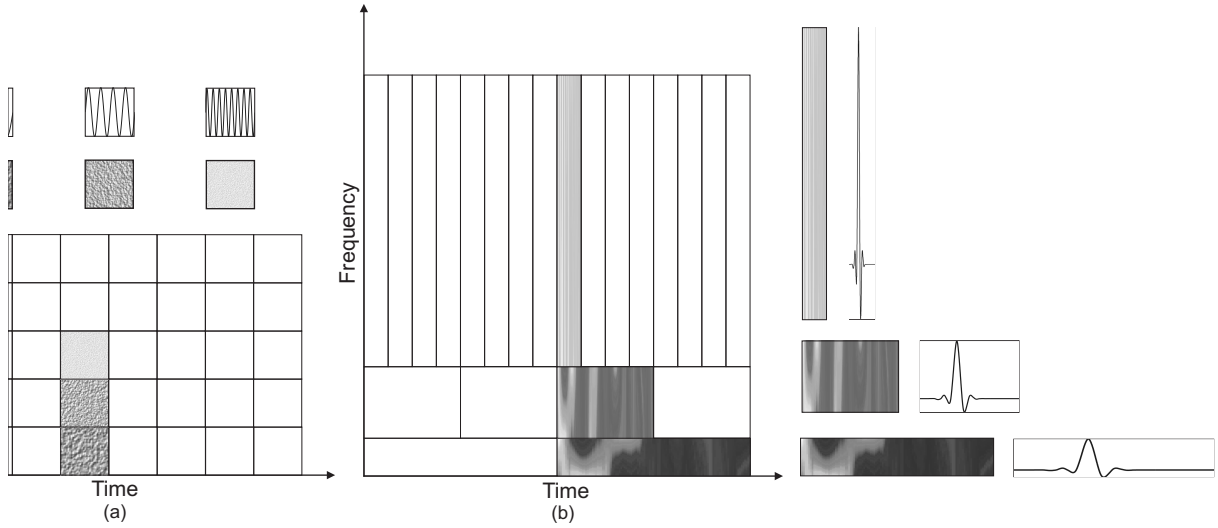

Figure 5: Tile coverage of the time-frequency plane: (a) by the Fourier basis functions; (b) by the S8 symmlet.

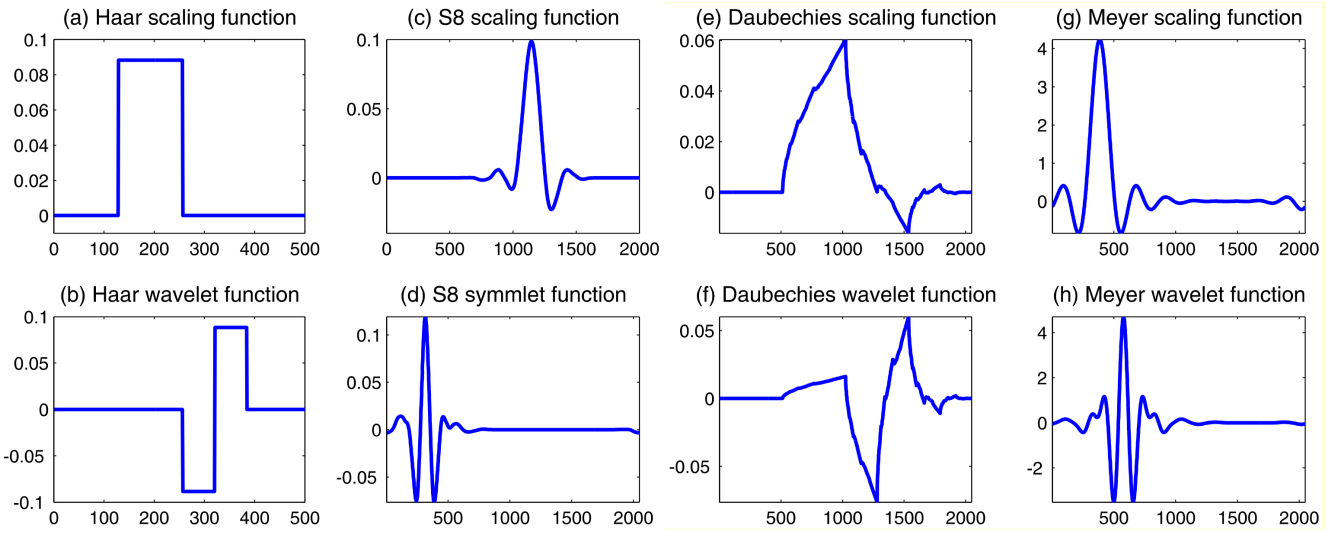

Figure 6: Various types of wavelets and their scaling functions: (a,b) Haar wavelet; (c,d) Nearly symmetric S8 symmlet with 8 vanishing moments, (e,f) Daubechies 4 wavelet, and (g,h) Meyer wavelet.

ing family of transforms use symmetric basis functions; (*iii*) the use of splines reduces spectral overlap between resolution channels by increasing the degree of spline n [20]. Nevertheless, small spectral overlap increases data decorrelation [21], 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 [8].

Donoho and Johnstone [7] proposed a method 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 the MSE estimation. The WT was performed using nearly symmetric wavelet with 8 vanishing moments (S8). The coefficients in an S8 wavelet expansion decay with scale more rapidly than the corresponding Haar coefficients. The SURE shrinkage car-

ried 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.

When applying the "MiniMax" thresholding rule to the noisy signal, the reconstruction suppresses almost entirely the noise, while preserving its sharp structural details . The soft thresholding nonlinearity has the threshold set to the magic number λ_n. In the case of "Visu" shrinkage thresholding , the reconstruction looks even better. The difference is due to the soft thresholding nonlinearity, which has the threshold set to $\sqrt{2\log(n)}$ [6].

Cycle-spinning is proposed by Coifman to remove artifacts from wavelet thresholding. Fully translation-invariant denoising, using the stationary WT is a method of removing artifacts from wavelet thresholding associated with discontinuities and singularities. It entails a hard thresholding nonlinearity, with threshold also set to $\sqrt{2log(n)}$.

3 Wavelet Basis as Fractals

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. Fractals [14] define a class of objects with the characteristic property of self-similarity (or self-affinity), meaning that the statistics describing the structure in time or space of a fractal process remain the same as the process is measured over a range of different scales (or *scale invariant*). Fractals are complex, patterned, statistically self-similar or self-affine, scaling or scale-invariant structures with non-integer dimensions, generated by simple iterative rules widespread in real and synthetic systems (Fig. 7). A wavelet *basis* is fractal (Fig. 8) and so a natural choice of basis for analysis of fractal data. Wavelet methods are particularly adequate in brain imaging data analysis due to broadly fractal properties exhibited by the brain in space and time. Hence wavelets may be more than just another basis for analysis of fMRI data (Fig. 11) [2].

Specific wavelet applications in neuroimaging include: (*i*) resampling (up to *4D*); (*ii*), smoothing by wavelet shrinkage [7] (which 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 [16]); (*iii*) estimation (robust and informative models); (*iv*) hypothesis testing (multiresolution and enhanced false detection rate control) [18]. Wavelet-based statistical

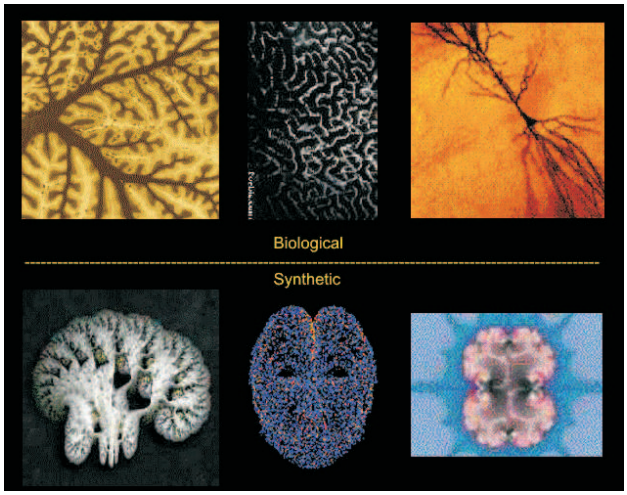

Figure 7: Brain images and fractal model counterparts.

analysis of fractal processes benefits are: (*i*) wavelet-based methods perform a multiresolution decomposition suitable for scale-invariant processes analysis; (*ii*) wavelet analysis is optimally whitening and provide Karhunen-Loève (KL) expansions [25] for long-memory ($1/f$-like) processes, which is the case in fMRI; and (*iii*) wavelets provide good estimators for the noise process parameters.

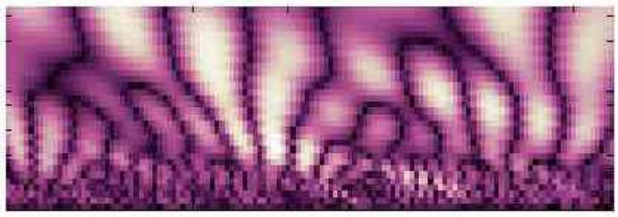

Figure 9: The CWT of fractal fMRI time series (after Bullmore [2]).

4 Conclusion

In image denoising, the aim is to suppress noise while preserving as much as possible of the image features. Thresholding in the wavelet domain is on the assumption of white Gaussian noise. In the orthonormal wavelet domain, most image information is contained in a few largest wavelet coefficients, while the noise is uniformly spread out across all coefficients. Thus thresholding mostly affects the noise rather than 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 by spatial Gaussian filtering with

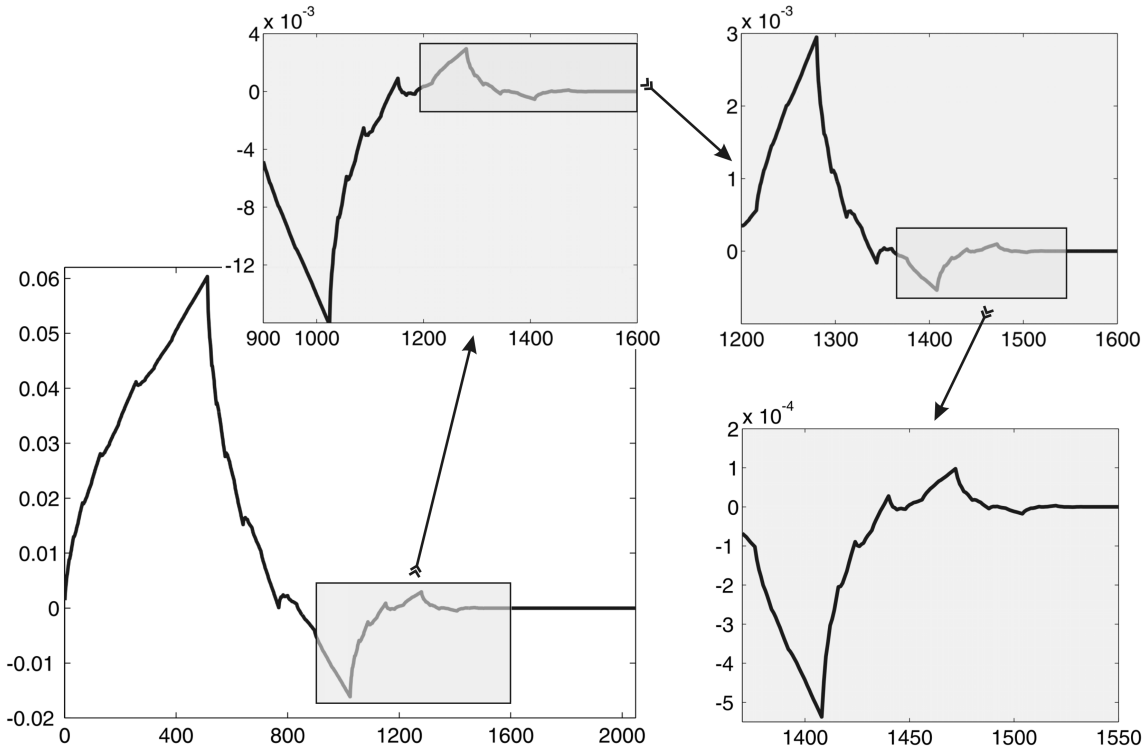

Figure 8: The fractal self-similarity of Daubechies 4 mother wavelet generated by WaveLab802 [1].

a single kernel. The risk of missing to detect spatial features of the smoothing kernel size or lower is alleviated by wavelets, which allow 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 [9]. 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. 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.

Ideal Fourier damping of noisy data performs image reconstruction by setting to zero in the Fourier domain all coefficients smaller than the noise level. Yet no means to determine which Fourier coefficients exceed the noise level exist. Nevertheless, ideal Fourier damping is not able to suppress noise while preserving image resolution.

Wavelet analysis is also computationally efficient: the DWT plays a similar role for computational time-scale analysis in MRA as the FFT does for spectral analysis.

Wavelet analysis is optimal in terms of detecting transients events in data and adapts well to conditions where responses change significantly in amplitude during experiments. Simple nonlinear wavelet methods outperform the traditional linear methods such as splines, Fourier series, and kernel-based smoothers.

Acknowledgements: The author is grateful to Dr. Dimitri Van De Ville, Biomedical Imaging Group (BIG), Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland for clearing up subtleties of the WSPM software.

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, 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.

[3] A. Cohen, I. Daubechies, and J.C. Feauveau, "Bi-orthogonal bases of compactly supported wavelets," *Commun. Pure Appl. Math.*, vol. 45, pp. 485-560, 1992.

[4] R. Crandall, Projects in Scientific Computation, Springer-Verlag, New York, 1994, pp. 197-198, 211-212.

[5] I. Daubechies, "Orthogonal bases for compactly supported wavelets," *Commun. Pure Appl. Math.*, vol. 41, pp. 909-996, 1992.

[6] D.L. Donoho, "Wavelet Shrinkage and W.V.D.: A Ten-Minute Tour," available at http://www-stat.stanford.edu/ donoho/Reports/1993/

[7] D.L. Donoho and I.M. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," *J. Am. Stat. Assoc.*, vol. 90, no. 432, pp. 1200-1224, 1995.

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

[9] M. Feilner, T. Blu, and M. Unser, "Optimizing wavelets for the analysis of fMRI data," *Proc. SPIE Conf. on Mathematical Imaging: Wavelet Applications in Signal and Image Processing VIII*, San Diego CA, USA, July 31-August 4, 2000, vol. 4119, pp. 626-637.

[10] A. Hyvärinen, P.O. Hoyer, and M. Inki, "Topographic independent component analysis," *Neural Comput.*, vol. 13, no. 7, pp. 1527-1558, 2001.

[11] 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.

[12] I.M. Johnstone and B.W. Silverman, "Wavelet threshold estimators for data with correlated noise," *J. Roy. Statist. Soc.*, vol. 59, pp. 319-351, 1997.

[13] Y. Katznelson, An Introduction to Harmonic Analysis. New-York, Dover, 1976.

[14] B.B. Mandelbrot, The Fractal Geometry of Nature, W.H. Freeman and Co., New York, 1977.

[15] S.G. Mallat, "Multiresolution approximations and wavelet orthogonal bases of $L^2(R)$," *Trans. Amer. Math. Soc.*, vol. 315, no. 1, pp. 6987, 1989.

[16] 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.

[17] R. Mutihac, "Wavelet-based statistical analysis of fMRI activation images," *International Society for Magnetic Resonance in Medicine, 14th Scientific Meeting and Exhibition* (ISMRM 2006) Seattle WA, USA, fMRI Data and Statistical Analysis Methods, p. 546, 2006.

[18] 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.

[19] R. Mutihac, "Sparse decomposition of brain imaging data - Wavelet transform and Independent Component Analysis," *12th Annual Meeting of the Organization for Human Brain Mapping* (OHBM 2006), Florence, Italy, 2360, 2006.

[20] U. 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.

[21] 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.

[22] J. Tian *et al.*, "Coifman wavelet systems: Approximation, smoothness, and computational algorithms," *Comput. Sci. 21st Century*, Tours, Paris, France, 1997.

[23] 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.

[24] 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.

[25] M. Unser, "On the approximation of the discrete Karhunen-Loève transform for stationary processes," *Signal Processing*, vol. 7, no. 3, pp. 231-249, 1984.

[26] M. Unser and T. Blu, "Wavelet theory demystified," *IEEE Trans. Signal Proc.*, vol. 51, no. 2, Feb. 2003.

[27] M. Vetterli and C. Herley, "Wavelets and filterbanks: Theory and design," *IEEE Transactions on Signal Processing*, vol. 40, pp. 2207-2232, 1992.

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