A WAVELET TRANSFORM FREQUENCY CLASSIFIER FOR STOCHASTIC TRANSIENT SIGNALS

Francisco M. Garcia *Isabel M. G. Lourtie*

ISR - Instituto de Sistemas e Robótica, IST - Instituto Superior Técnico
Torre Norte, Av. Rovisco Pais, P-1096 Lisboa Codex, Portugal
E-mail : fmg@isr.isr.ist.utl.pt

ABSTRACT

The paper presents a wavelet transform based frequency classifier for stochastic bandpass transient signals. Using compactly supported bases of wavelets, continuous nonstationary bandpass processes can be described by reduced dimension discrete time signal representations. We develop an optimization procedure to adapt both the sampling frequency and the discrete time wavelet filters to the continuous classes of signals involved. When compared to the classical Bayesian structures, the proposed receiver strongly reduces the required computational load. The receiver performance is accessed by Monte Carlo simulation, the probabilities of detection for each signal class, and the probability of false alarm, being computed.

1. INTRODUCTION

Detection and classification of stochastic transient signals is an important problem in areas such as underwater acoustics [1, 2] and seismology [3]. Stochastic transient signals are short-time duration processes, nonstationary in nature. Consequently, traditional signal processing techniques based on, for example, the Fourier series, are not well suited for processing transient signals.

The conventional approach to optimal statistical signal classification is the Bayes classifier [4]. It consists on a M-ary detection scheme that can be implemented as a binary classification tree: the likelihood ratio (LR) between each pair of hypotheses is compared with a threshold that depends on the signal classes *a priori* known probabilities.

Underlying the development of the optimal receiver, the decomposition of signal and noise processes under a set of orthonormal basis functions is often used. For binary signal detection, one classical approach is the Karhunen-Loève expansion (KLE) that provides a representation of both signal and white noise processes in terms of a discrete number of random uncorrelated signal coefficients. In general, for M-ary signal detection, unless the several signal classes autocorrelation functions have the same eigenfunctions [5], a KLE of the observations cannot be obtained.

In this paper, we develop a frequency signal classifier suited for stochastic bandpass transient signals. The design of the classifier relies on properties of the wavelet transform (WT), namely the capability of nearly describing a bandpass transient signal by a small set of wavelet decomposition scales [6], giving rise to a computationally efficient classification scheme.

2. PROBLEM FORMULATION

Let $x(t)$ be a sample function of the observed process, modeled by

$$x(t) = s_i(t) + w(t), \tag{1}$$

where t is the time variable. Signals $s_i(t)$, $i = 1, 2, ..$, belong to a set of strictly bandpass Gaussian stochastic transient signals. The observation noise $w(t)$ is a zero mean Gaussian white noise process. Signals and noise are mutually independent stochastic processes with known autocorrelation function.

Let us denote by $S_{s_i}(t,\omega)$ the time-frequency power spectral density of each signal $s_i(t)$, defined as the Fourier transform of the autocorrelation function $R_{s_i}(t+\tau/2, t-\tau/2)$, computed over the lag-variable τ. Signals $s_i(t)$ are, thus, characterized by exhibiting a limited bandwidth in the frequency domain, i.e.,

$$\forall t \quad S_{s_i}(t,\omega) = 0 \ \text{ if } \ |\omega| \notin [\omega_{\min_i}, \omega_{\max_i}], \tag{2}$$

and a nearly finite duration in the time domain, i.e.,

$$\forall \omega \quad S_{s_i}(t,\omega) \simeq 0 \ \text{ if } \ t \notin [0, t_{s_i}]. \tag{3}$$

In the above equations, $[\omega_{\min_i}, \omega_{\max_i}]$ stand for the maximum frequency range, and t_{s_i} for the time duration, of the transient signal $s_i(t)$.

Given the observation process $\{x(t), t \in [0,T]\}$, where T stands for the observation time interval, our goal is to decide about the presence (or absence) of a transient signal and simultaneously characterize to which one of the *a priori* given classes the arriving signal belongs.

By the Nyquist theorem, it is possible to reconstruct any bandpass process $s(t)$ from its samples $s(nT_s)$, $n = 1, 2, ...$, provided that the sampling time interval $T_s \leq \pi/\omega_{\max}$. In the above context, we are going to deal with this problem in the discrete time domain, that is, we consider that we first sample the continuous observation process $x(t)$, giving rise to the observed discrete time sequence

$$x(nT_s) = s_i(nT_s) + w(nT_s). \tag{4}$$

Afterwards, the decomposition of the signal under a discrete wavelet orthonormal basis is considered.

In the next section, the discrete wavelet transform decomposition of a bandpass transient signal is discussed and related with its continuous time characterization as a function of both the sampling interval and the wavelet functions of the basis.

3. WAVELET TRANSFORM REPRESENTATION OF BANDPASS TRANSIENTS

Using Mallat's recursive algorithm for image decomposition [7, 8] in its one-dimensional form,

$$\begin{aligned} c_k^j &= \sum_n h(n-2k)c_n^{j-1} \\ d_k^j &= \sum_n g(n-2k)c_n^{j-1} \end{aligned}, \tag{5}$$

0-7803-3192-3/96 $5.00©1996 IEEE

where $h(n)$ and $g(n)$ are Daubechies [8] finite length filters, a discrete time sequence $c_n^0 = s_i(nT_s)$ is decomposed in the subsequences d^1, d^2, ..., d^m and c^m. The recursive filtering equations (5) can also be expressed in terms of the original sampled signal s_i as the internal product

$$\begin{array}{l} c_k^j = < s_i, h_k^j > \\ d_k^j = < s_i, g_k^j > \end{array} . \qquad (6)$$

Filters $h(n)$ and $g(n)$ are, respectively, lowpass and highpass filters. While c^m represents a smoothed version of s_i, sequences d^j stand for filtered versions of s_i in different frequency bands [6]. The internal product $< s_i, g_k^j >$, is relevant only when both signal s_i and filter g_k^j exhibit overlapping frequency bands. Consequently, performing the WT decomposition of s_i until the lowpass residue corresponding to $< s_i, h_k^m >$ is close to zero, and neglecting those sequences d^j for which $< s_i, g_k^j >$ is also close to zero, one gets an approximate representation of s_i based on a smaller set of frequency scales.

The elements of the wavelet transform coefficients (WTC) covariance matrix representing the filtered versions of the sampled transient signal $s_i(nT_s)$ are

$$\begin{aligned} \mathrm{E}[d_n^m d_k^j] &= \mathrm{E}[< s_i, g_n^m >< s_i, g_k^j >] \qquad (7) \\ &= \sum_{\ell=-\infty}^{+\infty} \sum_{\Delta=-\infty}^{+\infty} g_n^m(\ell+\Delta) R_{s_i}^{T_s}(\ell+\Delta, \ell) g_k^j(\ell), \end{aligned}$$

where $R_{s_i}^{T_s}(\ell_1, \ell_2)$ represents the sampled version of the continuous signal autocorrelation function $R_{s_i}(\ell_1 T_s, \ell_2 T_s)$. Filters g_n^m are finite length filters ($g_n^m(\ell) = 0$ for $\ell \notin [L_{n_0}^m, L_{n_1}^m]$). Consequently, expression (7) is non-zero only when the time domain compact support of both g_n^m and g_k^j overlap, giving rise to the sparsity structure of the WTC covariance matrix.

Let us assume that the signal autocorrelation function is nearly stationary over the compact support of each finite length filter g_n^m, that is,

$$R_{s_i}^{T_s}(\ell+\Delta, \ell) \simeq R_{s_i}^{T_s}(L_n^m+\Delta, L_n^m) \text{ if } L_n^m \in [L_{n_0}^m, L_{n_1}^m]. \quad (8)$$

In other words, the time variation of the autocorrelation function is slow when compared to the time duration of the wavelet function (WF) of the basis.

In the above context, from (7), the diagonal terms of the WTC covariance matrix become

$$\mathrm{E}[d_n^{m^2}] \simeq \sum_{\Delta=-\infty}^{+\infty} R_{s_i}^{T_s}(L_n^m+\Delta, L_n^m) \sum_{\ell=-\infty}^{+\infty} g_n^m(\ell+\Delta) g_n^m(\ell). \qquad (9)$$

Using Parseval's theorem, and taking into account the symmetric structure of the autocorrelation function, it results after manipulation

$$\mathrm{E}[d_n^{m^2}] \simeq \frac{1}{2\pi} \int_{-\pi}^{+\pi} S_{s_i}^{\mathrm{disc}}(L_n^m, \Omega) |G_n^m(\Omega)|^2 d\Omega, \qquad (10)$$

where $S_{s_i}^{\mathrm{disc}}(L_n^m, \Omega)$ stands for the discrete time Fourier transform of the sampled autocorrelation function $R_{s_i}^{T_s}(L_n^m+\Delta, L_n^m)$ computed on the integer lag-variable Δ. An interesting issue regarding the wavelet functions of the basis is that the amplitude characteristic $|G_n^m(\Omega)|$ does not depend on n [6], that is,

$$\forall n \ \ |G_n^m(\Omega)| = |G^m(\Omega)|, \qquad (11)$$

being the same at every n for each scale m.

From the sampling theorem,

$$S_{s_i}^{\mathrm{disc}}(L_n^m, \Omega) = \frac{1}{T_s} S_{s_i}(L_n^m T_s, \frac{\Omega}{T_s}), \ \ \Omega \in [-\pi, \pi], \qquad (12)$$

where $S_{s_i}(t, \omega)$ is the continuous time-frequency power spectral density of the transient signal $s_i(t)$, equation (10) becomes

$$\mathrm{E}[d_n^{m^2}] \simeq \frac{1}{2\pi} \int_{-\frac{\pi}{T_s}}^{+\frac{\pi}{T_s}} S_{s_i}(L_n^m T_s, \omega) |G^m(\omega T_s)|^2 d\omega. \qquad (13)$$

Due to the transient signal nearly finite duration, condition (3), independently of the frequency bands of both signal and WF g_n^m,

$$S_{s_i}(L_n^m T_s, \omega) |G^m(\omega T_s)|^2 \simeq 0, \ \text{ for } \ L_n^m T_s \notin [0, t_s]. \qquad (14)$$

Since two consecutive vectors g_n^m and g_{n+1}^m have their non-zero terms shifted by 2^m [6], only a finite number of non-zero coefficients $\mathrm{E}[d_n^{m^2}]$ can be found for each scale m. Simultaneously, the bandpass power spectral density of both the transient signal and the frequency scaled wavelet function $G^m(\omega T_s)$ need to overlap in the frequency domain. Moreover, since each scale m corresponds to different frequency bands, only a finite number of scales m gives rise to non-zero $\mathrm{E}[d_n^{m^2}]$ coefficients. Finally, let us point out that the frequency scaling introduced by the sampling interval T_s in $|G^m(\omega T_s)|$ allows us to define the location of its non-zero frequency range. Therefore, with an appropriate choice of both the sampling interval and the wavelet basis, it is possible to tune the decomposition of process $s_i(t)$ so that most of its energy fall in a small set of coefficients. In the next section, we present the classifier structure, including the tuning procedure to optimize the energy distribution among the several classes of transient signals.

4. CLASSIFICATION SCHEME

The classification problem is herein addressed as a M-ary detection problem. Each hypothesis H_i, $i = 0, 1, ...$, corresponds to receiving, under the presence of white noise, a stochastic transient signal $s_i(t)$ from class i. Hypothesis H_0 stands for the situation where no signal is present.

Let us denote by X the observation vector obtained by sampling and decomposing, under a set of orthonormal functions that span every possible class of signals, the observed continuous time waveform $x(t)$. In the Bayes classifier, implemented as a binary classification tree, the decision about which class the received signal belongs results from sequentially comparing the *a posteriori* probabilities

$$P(\mathrm{H}_j|X) \underset{\mathrm{H}_k}{\overset{\mathrm{H}_j}{\gtrless}} P(\mathrm{H}_k|X), \quad \forall k \neq j, \qquad (15)$$

for every pair of hypotheses, as shown in Figure 1. For Gaussian signal and noise processes, the binary test (15) becomes

$$\mathcal{L} = X^\top(\Lambda_k^{-1} - \Lambda_j^{-1})X \underset{\mathrm{H}_k}{\overset{\mathrm{H}_j}{\gtrless}} \ln\left(\frac{|\Lambda_j|}{|\Lambda_k|}\right) + 2\ln\left(\frac{P(\mathrm{H}_k)}{P(\mathrm{H}_j)}\right), \qquad (16)$$

where $P(\mathrm{H}_i)$ is the assumed known *a priori* probability of hypothesis i, Λ_i represents the covariance matrix of the observation vector X under hypothesis i, and $|\cdot|$ stands for the matrix determinant.

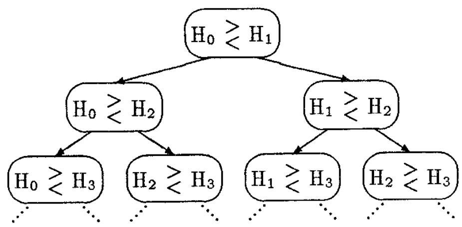

Figure 1. Bayesian classification tree.

When the dimension of the observation vector X is large, the evaluation of the likelihood ratio $\mathcal{L}$, left side of inequality (16), is a heavy computational procedure to be taken at every processing step along time. To reduce this computational load, it is convenient to decompose the observation signal $x(t)$ under the smallest possible set of coefficients. For this purpose, we are going to define, for each binary test, the signal decomposition that better adapts to the two hypotheses under analysis.

Given two distinct decompositions, X and Y, of the same observations, the binary tests $P(\mathrm{H}_j|X) \gtrless P(\mathrm{H}_k|X)$ and $P(\mathrm{H}_j|Y) \gtrless P(\mathrm{H}_k|Y)$ are equivalent if and only if X and Y are one-to-one mappings. This is not the case if X and Y correspond to different samplings of the continuous waveform $x(t)$, even if both sampling rates are under the transient signal Nyquist condition. Nevertheless, both tests reflect, with similar accuracy, the comparison between "signal classes j and k *a posteriori* probabilities given the continuous time observation $\{x(t),\ t \in [0,T]\}$". For this reason, we assume that using different sampling rates at each pair of binary tests in the classification tree is harmless.

Our goal is to define, for each pair of binary tests (15), the "best" signal decomposition, in the sense that it minimizes the number of wavelet coefficients required to describe, up to a given amount of error, the transient signals under classification.

This problem is solved in two steps: First, given the two classes of transients, j and k, find T_s and a filter g (corresponding to a wavelet family (WF) orthonormal basis of compact support) such that there exists two different scales m_j and m_k that maximize the functional

$$J = \sum_{n=N_0(m_j)}^{N_1(m_j)} E[d_n^{m_j 2}|\mathrm{H}_j] + \sum_{n=N_0(m_k)}^{N_1(m_k)} E[d_n^{m_k 2}|\mathrm{H}_k]. \quad (17)$$

The maximization of the above functional leads to the choice of the two scales that "better" represent the classes of signals under classification, in the sense that they include the most part of the energy of signal classes j and k, respectively. Remark that, in (17), we are assuming that $E[d_n^{m_i 2}] \simeq 0$ for $n \notin [N_0(m_i); N_1(m_i)]$. These limits are established based on the transient signals time duration t_s. Secondly, and once defined the most important scale to describe each signal class, the other scales to be included are obtained in order to guarantee that the mismatch between the complete and the approximate signal representations is smaller than an *a priori* given maximum allowable error.

From (13), and defining

$$W_i(\omega) = \sum_{n=N_0(m_i)}^{N_1(m_i)} S_{s_i}(L_n^{m_i} T_s, \omega) \ , \quad (18)$$

one concludes that maximizing (17) is equivalent to maximize

$$J = \int_{-\frac{\pi}{T_s}}^{+\frac{\pi}{T_s}} W_j(\omega)|G^{m_j}(\omega T_s)|^2 + W_k(\omega)|G^{m_j}(\omega T_s)|^2 d\omega. \quad (19)$$

The above functional has several local maximum pairs (m_j, m_k). For example, if the time average power spectral densities, $W_i(\omega)$, of two distinct signal classes are centered around two frequencies that are not separated from each other by more than one octave, then a maximum is obtained for $m_j = m_{k\pm 1}$. For fixed WF, if $f_s = 1/T_s$ is the sampling frequency corresponding to a maximum, then other maxima are reached at sampling frequencies close to $2^l f_s,\ l \in N$. Typically, we choose the lowest sampling frequency that corresponds to a maximum, in order to reduce the computational load of the receiver. Care must be taken, however, to ensure that for each pair of classes, the sampling frequency is higher than the Nyquist frequency for all classes, to avoid aliasing effects that may reduce the performance of the classifier.

The binary tests (16) are, then, implemented based on different signal decompositions for each pair of hypothesis j and k, and taking only into account the dominant scales for each signal class, that is,

$$X_k^\mathsf{T} \Lambda_k^{-1} X_k - X_j^\mathsf{T} \Lambda_j^{-1} X_j \underset{\mathrm{H}_k}{\overset{\mathrm{H}_j}{\gtrless}} \ln\left(\frac{|\Lambda_j|}{|\Lambda_k|}\right) + 2\ln\left(\frac{P(\mathrm{H}_k)}{P(\mathrm{H}_j)}\right). \quad (20)$$

In (20), X_i is the decomposition of the observed waveform $\{x(t),\ t \in [0,T]\}$ under the "smallest" decomposition achieved for class i, the corresponding covariance matrix Λ_i being built accordingly.

5. SIMULATIONS

The performance of the herein proposed receiver is evaluated based on a 20,000 run Monte Carlo experiment, consisting of four hypothesis H_i, $i = 0,1,2,3$, for classification. Hypothesis H_0 considers that only noise is observed, while the other three hypotheses, assumed to be equally likely, correspond to the case where a stochastic transient signal is embedded in white noise. The observation noise spectral height is 16.

The three classes of transient signals are generated accordingly to the block diagram shown in Figure 2, where

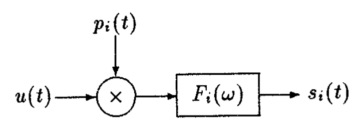

Figure 2. Signal structure.

$p_i(t)$ is a gating function in the time interval $[0,1]$ s., $F_i(\omega)$ is a strictly bandpass filter, and $u(t)$ is a Gaussian zero-mean white noise process with spectral height 1. Filters $F_i(\omega)$ are obtained by truncating, outside a pre-specified bandwidth, the frequency response of fourth order linear filters characterized by having unitary maximum gain, a double zero at the origin and a double pair of complex conjugated poles. Table 1 summarizes the gating and filtering characteristics in the generation of the three classes of signals.

The received waveform $x(t)$ is decomposed using Daubechies' D14 family of wavelets [8]. The tuning of the signal decomposition to each binary test in the classification tree is performed by maximizing the functional (19). Table 2 shows the sampling time and the dominant scales

Table 1. **Signal classes.**

Signal	gating	$F(\omega)$	
		poles	band (rad/s)
1	$16.4e^{-50(t-0.5)^2/2}$	$-20 \pm \jmath 80$	$[66, 99]$
2	$11.9e^{-1.5t}$	$-30 \pm \jmath 120$	$[99, 148]$
3	54.6	$-45 \pm \jmath 180$	$[148, 223]$

Table 2. Dominant scales and sampling times for classification tree.

	Test $i - j \Leftrightarrow P(\mathrm{H}_i\|X) \gtrless P(\mathrm{H}_j\|X)$					
	1-2	1-3	2-3	1-0	2-0	3-0
T_s (ms)	7.8	13.5	10.4	13.2	8.8	5.9
Scale i	3	2	2	2	2	2
Scale j	2	3	3	-	-	-

obtained for each binary hypothesis test. Remark that, for the decision tests in-between any signal class and the noise only situation, the sampling interval is tuned taking only into account the signal characterization.

Two experiments are conducted: In the first one, the complete (exact) WT decomposition of the received process is considered. In the second one, only the dominant terms for each signal class decomposition are used, corresponding to about 88 % of the total energy. The number of dominant terms is in-between 11 and 56, corresponding to an efficient improvement regarding signal decomposition.

Denoting by P_{D_i} the probability of detecting a signal from class s_i under hypothesis H_i, that is, of saying that a transient signal of class s_i is present when it is, and by P_{Fa} the probability of false alarm, i.e., the probability of saying that a transient is present when it is not, the classifier probability of error is given by

$$P(\epsilon) = \sum_{i=1}^{3} [1 - P_{D_i}] P(\mathrm{H}_i) + P_{Fa} P(\mathrm{H}_0). \qquad (21)$$

To access the performance of the receiver under both case studies, we plot in Figure 3, for each signal class, the P_{D_i} *vs.* P_{Fa} curves, i.e., the receiver operating characteristics (ROC) curves. Table 3 shows, in percentage, the *a posteriori* probabilities, $P(s_i|\mathrm{H}_j)$, of detecting a signal of class i under hypothesis H_j, assuming that all hypotheses are *a priori* equally likely, that is, when $P(\mathrm{H}_i) = 0.25$, $\forall i$. First and second values correspond, respectively, to the complete and approximate signal descriptions.

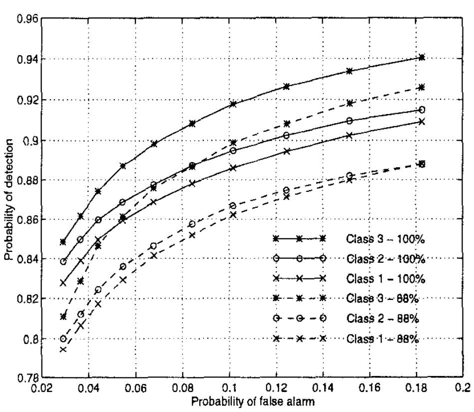

Figure 3. Receiver operating characteristics.

Table 3. Classification results.

j	$P(s_1\|\mathrm{H}_j)$	$P(s_2\|\mathrm{H}_j)$	$P(s_3\|\mathrm{H}_j)$	$P(\text{noise}\|\mathrm{H}_j)$
1	85.9-82.9	0.1-0.2	0.2-0.4	13.8-16.5
2	0.1-0.2	86.9-83.6	0.2-0.4	12.9-15.8
3	0.1-0.1	0.1-0.1	88.7-86.1	11.1-13.6
0	1.9-2.1	1.0-1.3	2.6-3.1	94.5-93.5

The receiver that uses the complete WT decomposition of the processes is the best Bayes classifier for the given signal classes. The achieved ROCs represent, for each class of signals, an upper bound for the receiver performance. As expected, when only the dominant terms for each signal class decomposition are used, a small performance degradation is observed, due to the mismatch between assumed and actual signal characterization. Remark, however, that the computational efficiency of the receiving scheme regarding the second experiment is greatly improved.

6. CONCLUSION

The paper presents a classifier for Gaussian stochastic transients, based on the wavelet transform. The main contribution of the present work consists on a procedure to adapt both the sampling interval and the wavelet function of the basis to the continuous classes of signals involved.

Other interesting properties of the wavelet transform, that motivated this work but are not highlighted herein, point out to the possibility of designing receiving schemes well suited to operate in real-time [6], leading to an integrated procedure to simultaneously perform the transient signal segmentation and classification tasks.

The Monte Carlo simulation study, carried out with synthetic data, leads to the conclusion that only a small set of features are required to describe bandpass stochastic transient signals. Although exhibiting a small degradation in performance with respect to the optimal solution, the achieved structure shows to be computationally efficient.

REFERENCES

[1] B. Boashash, P. O'Shea, "A Methodology for Detection and Classification of Some Underwater Acoustic Signals Using Time-Frequency Analysis Techniques," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 38, no. 11, pp. 1829- 1841, November 1990.

[2] J.-P. Hermand, P. Nicolas, "Adaptive Classification of Underwater Transients," in *Proc. 1989 IEEE Int. Conf. Acoust., Speech, Signal Proc.*, Glasgow, UK, vol. 4, pp. 2712-2715, May 1989.

[3] G. Olmo, L. Lo Presti, "Applications of Wavelet Transform for Seismic Activity Monitoring," in *Wavelets: Theory, Algorithms, and Applications*, eds. C. K. Chui, L. Montefusco, L. Puccio, Academic Press, Inc, 1994.

[4] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, Academic Press, Inc., 1972.

[5] H. V. Poor, *An Introduction to Signal Detection and Estimation*, 2nd edition, Springer-Verlag, 1994.

[6] F. M. Garcia, I. M. G. Lourtie, "An Arrival Time Estimator for Transient Stochastic Signals using the Wavelet Transform," *38th Midwest Symposium on Circuits and Systems*, Rio de Janeiro, Brazil, August 1995.

[7] S. Mallat, "A Theory For Multiresolution Signal Decomposition : The Wavelet Representation," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 674-693, July 1989.

[8] I. Daubechies, "Orthonormal Bases of Compactly Supported Wavelets," *Comm. on Pure and Applied Math.*, vol. XLI, pp. 909-996, 1988.