ReMake by the authors of 1998 SPIE original

A Wavelet neural network for detection of signals in communications

Raquel Gómez-Sánchez[a], Diego Andina[a]

[a] Grupo de Automatización en Señal y Comunicaciones
Universidad Politécnica de Madrid
Ciudad Universitaria s/n, 28040 Madrid[*]

ABSTRACT

Our objective is the design and simulation of an efficient system for detection of signals in communications in terms of speed and computational complexity. The proposed scheme takes advantage of two powerful frameworks in signal processing: Wavelets and Neural Networks. The decision system will take a decision based on the computation of the a priori probabilities of the input signal. For the estimation of such probability density functions, a Wavelet Neural Network (WNN) has been chosen. The election has arosen under the following considerations: (a) neural networks have been established as a general approximation tool for fitting nonlinear models from input/output data and (b) the increasing popularity of the wavelet decomposition as a powerful tool for approximation. The integration of the above factors leads to the wavelet neural network concept. This network preserve the universal approximation property of wavelet series, with the advantage of the speed and efficient computation of a neural network architecture. The topology and learning algorithm of the network will provide an efficient approximation to the required probability density functions.

Keywords: detection, communications, wavelets, neural networks, a priori probabilities, nonlinear models.

1. INTRODUCTION

Wavelet theory has emerged as a powerful tool in many areas of engineering and applied mathematics, as signal processing and numerical analysis. Its application range from constituting relevant theoretical models[1] to providing efficient mechanisms for implementations in one and two dimensions, through Quadrature Mirror Filter banks[2] (QMF banks structures).

However, the filter bank implementations of wavelet bases suffers from the lack of adaptability. QMF banks provide decompositions for all functions in the associated subspace[3]. When the objective is to approximate a single function, many wavelets in the subspace basis could be probably eliminated for a more efficient representation. Another disadvantage is that QMF banks provide decompositions just over orthogonal wavelet bases, while nonorthogonal bases, easier to implement, could be sufficient.

An alternative for constructing wavelet basis arises from the *neural network* concept[4]. Many studies have shown the ability of neural networks for approximating functions[5]. In its analytical expression, a neural network transfer function appears as a signal decomposition in terms of the activation function. When a wavelet is used for the nodes, the resulting network has been denominated *Wavelet Neural Network*[6].

Wavelet Neural Networks (WNN) constitute a recent topic in signal processing and neural networks areas. His goal consists on combining wavelet properties and neural network structures to determine the net parameters which best approximate a given function. Several ways have been proposed to approach the design of wavelet neural networks. In [6] wavelet network is introduced as a class of feedforward networks composed of wavelets, in [7] the Discrete Wavelet Transform is used for analyzing and synthesizing feedforward neural networks, and in [8] orthonormal wavelet bases are used for constructing wavelet-based neural networks.

[*] Further author information:
D. A. : Email: andina@gc.ssr.upm.es;

This paper is based on one of the most recent works about the topic[9], where algorithms for wavelet network construction are proposed. In the following section, a brief introduction to wavelet theory is presented. Results of wavelet neural networks implementation, when applied to the problem of function approximation, are reported in Section 3. In section 4, these implementations are applied to the problem of signal detection in communications. Finally, conclusions will be drawn in Section 5.

2. WAVELET BASES FOR APPROXIMATION

The problem of approximating a function in the Hilbert space $L^2(\Re)$ has been classically solved by orthonormal Fourier bases of the type $\{a_n\, e^{j\omega_n t}, \omega_n \in \Re, a_n \in C\}$, where the function decomposition coefficients, a_n, supply information about frequency content. But in many cases it would be desirable that the coefficients supply information about both, time and frequency content. It is well known[10] Fourier series only appears non-localized frequency information. In order to obtain time-frequency localization, alternative transformations have been developed, such the Short-Time-Fourier-Transform[11], and Continuous and Discrete Wavelet Transforms[1]. Below a brief resume of the elemental theory for approximating functions with wavelets is presented.

1) *Unidimensional case*. It can be shown[12] that, adequately selecting α and β, the family of functions:

$$\Psi = \left\{ \psi_{m,n}(x) = \alpha^{m/2} \psi(\alpha^m x - \beta n) : m, n \in Z,\ \alpha \in \Re^+,\ \beta \in \Re \right\} \tag{1}$$

under the admisibility condition for a function in $L^2(\Re)$, $\psi: \Re \to \Re$:

$$C_\psi = \int_0^\infty \frac{\Psi(\omega)}{\omega} d\omega < \infty \tag{2}$$

constitutes a *frame*, that is, a natural generalization of orthonormal bases for Hilbert spaces, which leads to decompositions of signals in terms of functions that have less requirements than orthonormal bases. Moreover, if ψ satisfies further conditions[12], the family (2) constitutes an orthogonal basis.

When the condition (2) is satisfied, the function ψ is said to be a *mother wavelet*. If the further conditions for orthogonality are required, the mother wavelet constitutes an *orthogonal mother wavelet*. For the problem of function approximation, a frame suffices, so that the conditions on the mother wavelet can be relaxed and much more freedom on the choice of the wavelet function is gained. Nevertheless, the fast algorithms associated with orthonormal wavelet bases are lost.

Then, given any $f \in L^2(\Re)$, it can be decomposed in term of the frame elements (1) as:

$$f = \sum_{m,n} c_{m,n} \psi_{m,n} \tag{3}$$

Under certain conditions on the parameters α and β, the family (1) may be a tight frame[7], so that the coefficients $c_{m,n}$ can be computed as the usual Fourier coefficients. Then the problem of representing a function f lies on computing the frame coefficients in (3).

2) *Extension to multidimensional wavelets*. In the Hilbert space $L^2(\Re^n)$, a mother wavelet can be obtained as follows:

$$\psi(\bar{x}) = \psi_s(x_1) \ldots \psi_s(x_n), \bar{x} \in \Re^n \tag{4}$$

where ψ_s is a mother wavelet in $L^2(\Re)$. The correspondent frame appears as:

$$\Psi = \left\{ \det(D_k)^{1/2} \psi(D_k \bar{x} - \bar{t}_k) : \bar{t}_k \in \Re^n, D_k = diag(\bar{d}_k), \bar{d}_k \in \Re^n, k \in Z \right\} \tag{5}$$

which is a direct extension of the unidimensional case.

For a simpler structure in the multidimensional case, usually radial wavelets functions[13] are used. A radial wavelet in $L^2(\Re^n)$ satisfies:

$$\psi(\bar{x}) = g(\|\bar{x}\|) \tag{6}$$

where g: $\Re \rightarrow \Re$ is a single variable function and ||.|| represents the vector norm. The radial wavelets frames are naturally single-scaling, so then, structurally simpler than multiscaling wavelet frames, as (5).

A function $f \in L^2(\Re^n)$ can be decomposed in terms of a multidimensional frame in the way it was done for the unidimensional case, shown in (3). Now, the computation of the scalar products is significatively more expensive, due to the higher dimensionality of the problem.

3. WAVELET NETWORKS

Neural networks are a class of computational architectures which are composed of interconnected, simple processing nodes with weighted interconnections. Figure 1 shows a simple neural network, known as (1+1/2)-layer neural network[6].

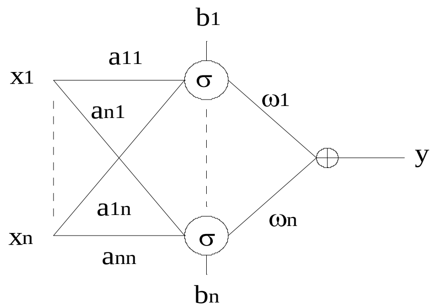


Figure 1: The (1+1/2)-layer neural network.

This network only differs to the traditional 2-layer neural network in the absence of the output node. The transfer function of the network in the figure can be analytically described as follows:

$$y = f(\bar{x}) = \sum_{i=1}^{n} w_i \sigma(\bar{a}_i^{\,T} \bar{x} + b_i), \quad \bar{a}_i, \bar{x} \in \Re^n, \; b_i, w_i \in \Re \tag{7}$$

which reminds of an expansion in terms of the functions:

$$\left\{\sigma_i : \sigma_i(\bar{x}) = \sigma(\bar{a}_i^{\,T} x + \bar{b}_i)\right\} \tag{8}$$

For the unidimensional case and being σ a mother wavelet, expression (7) results identical to a decomposition in terms of a truncated wavelet frame. In such case, network in Figure 1 is called wavelet neural network.

For the unidimensional case, the wavelet neural network implements a truncated version of the frame in (1), being the translation parameters $\beta n = b_i$ the offsets and the scale parameters $\alpha^k = a_k$ the weights from the input to the neurons[7]. In the wavelet frame, for each translation parameter, infinite scales parameters (resolution levels) are considered. In the truncated frame, a number J of resolution levels is selected. This means that the number of wavelets in the truncated frame (nodes in the network) corresponds to J×P, being P the number of the selected translation parameters. As a result, the offsets of the nodes are repeated J times.

Several training algorithms have been proposed and developed for networks like Figure 1. The *backpropagation* algorithm[14] with a sigmoidal activation function have been established as a general tool for approximation. A number of rigorous mathematical proofs have been provided to explain the ability of feedforward neural networks to approximate maps under these training algorithms. However, these methods do not naturally give rise to systematic synthesis procedures for these networks.

With the interpretation in (7) as a wavelet frame decomposition, an explicit link between the network coefficients and some appropriate transform is provided. This may be extremely useful for achieving the optimal structure (number of nodes), and guaranties the universal approximation property.

Several works have appeared on the topic of constructing wavelet neural networks. One of the most recent works[9] in which this paper is based, propose algorithms to approach the problem of *nonparametric regression estimation*, and will be summarized below.

Two random variables $x \in \Re^n$ and $y \in \Re$ satisfy a regression model, when they are related as follows:

$$y = f(x) + e \tag{9}$$

where f is a nonlinear function belonging to some functional space, and e is a white noise independent of x. The variables x and y are referred as the input and output model respectively.

A training data set is defined as a sample set of the input-output pair (x,y):

$$\begin{aligned} X &= \{x_1, x_2 \ldots, x_n\} \\ Y &= \{y_1, y_2, \ldots, y_n\} \end{aligned} \tag{10}$$

The problem to be solved is to find a nonparametric estimator f_e of f based on the data sample (X,Y). With this information, a truncated wavelet frame must be selected and the correspondent decomposition coefficients obtained, in order to determine a wavelet neural network structure which fits f.

For the unidimensional case, the first problem addressed above consists on selecting the finite number of elements in the family (1) which best approximate the data. The starting point is a regular wavelet lattice, as shown in Figure 2.

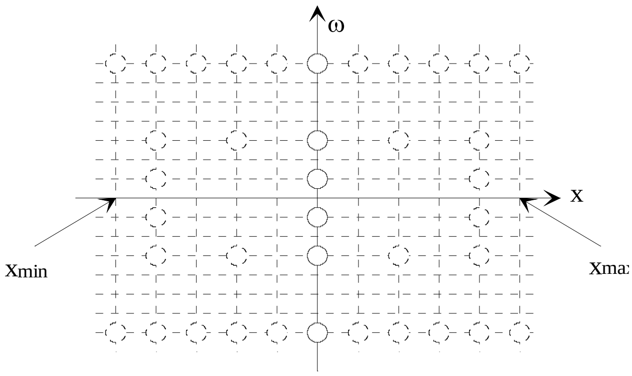


Figure 2: A regular wavelet lattice in the time-frequency plane.

Many wavelets in this regular lattice will probably not contain any data point in their supports. The training data do not provide any information for determining the coefficients of these “empty” wavelets. These wavelets are thus superfluous for the regression estimation problem and should be eliminated.

Once selected the coefficients b_i which must appear in the decomposition, attention must be paid to coefficients a_k. Due to the scaling property, increasing the number of parameters a_k results in increasing the resolution of the approximation for a given temporal coefficient b_i . In most cases 4 or 5 resolution levels have shown to be sufficient for fitting the training data set.

At this point, a truncated wavelet frame have been achieved:

$$S = \{\psi_i(x), i = 1,2,\dots, L\} \tag{11}$$

being $L \leq n$ (some nodes can be eliminated in the training network algorithm). The remaining problem is to find the coefficients w_i which define the resultant wavelet neural network:

$$f_e(x) = \sum_{i=1}^{L} w_i \psi_i(x) \tag{12}$$

One of the algorithms proposed to find these coefficients is known as *stepwise selection by orthogonalization*[9]. In the first step selects the wavelet (node) in S that best fits the observed data, then repeatedly selects the wavelet in the remainder of S that best fits the data while combining with the previously selected wavelets. For computational efficiency, later selected wavelets are orthonormalized to earlier selected ones.

The following vectorial notations are defined:

$$v_j = k \begin{bmatrix} \psi_j(x_1) \\ \dots \\ \psi_j(x_n) \end{bmatrix} \tag{13}$$

where $\psi_j \in S$, j=1,...L, $x_k \in X$. The k constant is a scalar so that v_j has unity norm,

$$y = \begin{bmatrix} y_1 \\ \dots \\ y_n \end{bmatrix} \tag{14}$$

being $y_k \in Y$. At iteration i, l_i denote the index of the selected wavelet from S. If i is the current iteration number, then v_{l1}, ... v_{li-1} have been selected in the previous iterations. Let $l_1 = arg\ max_i\ (v_i^T y)$, denote $q_{l1} = v_{l1}$. The Gram-Schmidt procedure is used for the following orthogonalization:

$$\begin{aligned} p_j &= v_j - (v_j^T q_{l1}) q_{l1}, \quad j \neq l_1 \\ q_j &= \frac{p_j}{\|p_j\|} \end{aligned} \tag{15}$$

Figure 3 shows that $\|r_j\|^2$ will be minimized if $u_j^2 = (y^T q_j)^2$ is maximized. This leads to the criterion for the selection of the best wavelet in each step of the algorithm.

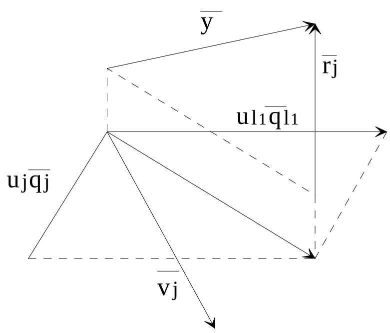


Figure 3

The resulting wavelet network is:

$$f_e(x) = \sum_{i=1}^{s} w_{li} \psi_{li}(x) \tag{16}$$

where the coefficients w_{li} are determined by

$$A \begin{bmatrix} \omega_{l1} \\ \dots \\ \omega_{ls} \end{bmatrix} = \begin{bmatrix} u_{l1} \\ \dots \\ u_{ls} \end{bmatrix} \qquad A = \begin{bmatrix} \alpha_{11} & \alpha_{12} & \dots & \alpha_{1s} \\ 0 & \alpha_{22} & \dots & \alpha_{2s} \\ 0 & 0 & \dots & \dots \\ 0 & 0 & \dots & \alpha_{ss} \end{bmatrix} \tag{17}$$

Orthogonalization algorithm:

I={1, 2, ..., L};
$p_j = v_j$ for all $j \in I$;
$l_0 = 0$, $q_{l0} = 0$;
Begin loop
For i = 1:L
$p_j = p_j - (v_j^T q_{li-1}) q_{li-1}$ for all $j \in I$;
$I = I - \{ j : p_j = 0 \}$;
If I is empty, set s = i-1 and break the loop;
$l_i = \arg\max_j (p_j^T y)^2 / p_j^T p_j$;
$\alpha_{ji} = v_{li}^T q_{lj}$, for j =1, ..., i-1;
$\alpha_{ii} = (p_j^T p_j)^{1/2}$;
$q_{lj} = \alpha_{ii}^{-1} p_{li}$;
$u_{li} = q_{lj}^T y$;
$I = I - \{l_i\}$;
End loop.

In the figures below several results for different functions approximation using the described algorithm are depicted.

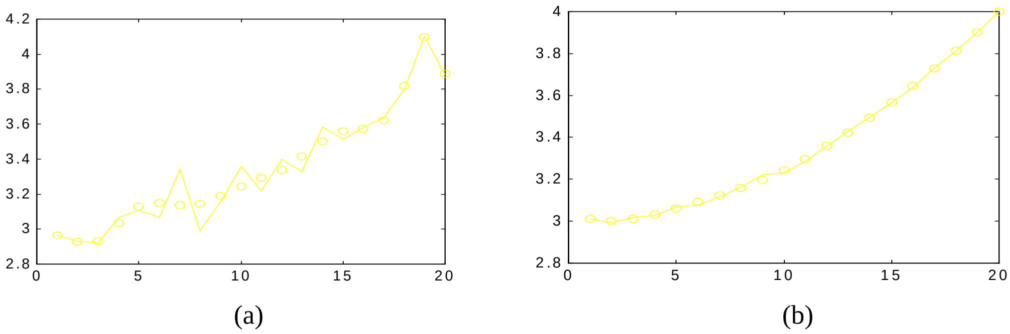
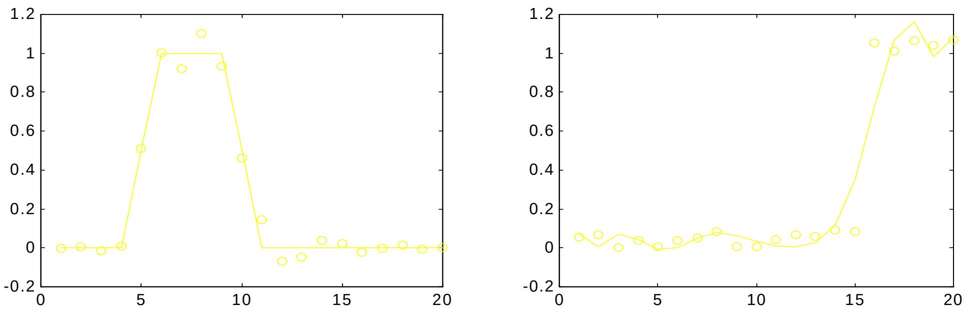


Figure 4: Approximation curves. Noisy approximated function (solid line) and wavelet neural network output (dot line) for gaussian noise and variances (a) $\sigma^2 = 0.1$ and (b) $\sigma^2 = 0.01$.



(a) (b)

Figure 5: Approximation curves. Noisy approximated function (dot line) and wavelet neural network output (solid line) for uniform noise and variances (a) σ^2 = 0.1 and (b) σ^2 = 0.01.

4. SIGNAL DETECTION IN COMMUNICATIONS

This section concerns the problem of signal detection in digital communication systems. The interest will be focused in the approximation to the demodulator decision problem by means of a wavelet neural network. A conventional digital communication system[15] is shown in Figure 6:

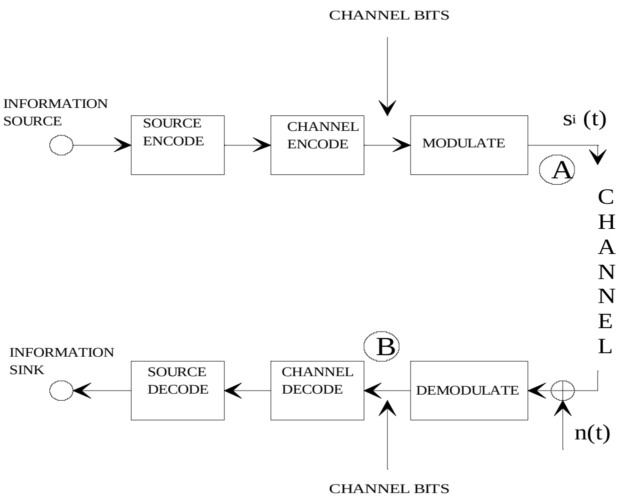


Figure 6: Digital communication system scheme.

The subject of study is constituted by the portion between **A** and **B** points. One element of a set of 2^k symbols, {$s_i(t)$, i=1, ..., M=2^k }, is being transmitted during a symbol interval T_s. The channel attenuates the signal amplitude, and a white noise is added at the input to the demodulator. Therefore, supposed s_i to be the transmitted symbol, the received signal is modeled as:

$$b_i(t) = \alpha s_i(t) + n(t) \tag{18}$$

being α the attenuation factor. This simple model does not take into consideration other problems existing in a communication channel, as *intersymbol interference*[15] (IES). Classical developments of demodulators are based on this model, so that, in order to establish comparisons and to obtain conclusions, our work will assume it too.

The received signal passes through a cascade of adapted filters and samplers, so that a vectorial representation, a scalar x for the unidimensional case, is provided. The decision is taken based on the MAP (*Maximum A Posteriori*) criterion:

$$max_i \quad p(s_i(t) / x) \qquad \mathrm{i} = 1, \; \ldots, \mathrm{M} \tag{19}$$

which selects the symbol satisfying that, assumed x received, is the most probable to have been transmitted. Under the constraint of symbol equiprobability, the MAP criterion can be expressed as follows:

$$max_i \quad p(x \; / \; s_i(t)) \qquad \mathrm{i} = 1, , \ldots, \mathrm{M} \tag{20}$$

where $p(x/s_i(t))$ represents the probability density function of x supposed $s_i(t)$ transmitted. Then, the decision block must implement the M functions following:

$$\begin{array}{c} p(x / s_1(t)) \\ \ldots \\ p(x / s_M(t)) \end{array} \tag{21}$$

which are nonlinears functions of the form $f : \Re^n \rightarrow \Re$, being n the dimension of the modulation system (n =1 for binary signals, n =2 for PSK signals, for example).

To accomplish the problem of approximating (21) several solutions may be adopted. When the mean, variance and type of distribution function of the additive noise is known, the problem can be analytically easily solved[14]. But when the statics and type of noise is unknown, an adaptive and efficient approximation tool is required. We propose a wavelet neural network for the task.

As it has been shown in Section 3, in order to approximate a function f, a training data set, as in (10), is needed. For obtaining this data set, we first obtain a set of a large number of output vectors x for each one of the transmitted symbols s_i is transmitted. Let denominate the set $X(s_i)$. The samples for the probability density function $p(x/s_i)$ are then obtained in the following way:

$$p(v / s_i) = \frac{\text{number of elements in } \mathrm{X}(\mathrm{s_i}) \; \in \left[v - \Delta, v + \Delta \right]}{\text{number of elements in } \mathrm{X}(\mathrm{s_i})} \tag{22}$$

Obviously, these are noisy samples of the real probability density functions in (21) (due to quantization effects among others), so the training data set for approximate each function in (21) results:

$$\begin{array}{c} X_i = \left\{ x_1, \ldots, x_n \right\} \\ Y_i = \left\{ p(x_1 \; / \; s_i(t)) + e_1, \ldots, p(x_n \; / \; s_i(t)) + e_n \right\} \end{array} \tag{23}$$

The above expression establishes an identical problem to the one addressed in Section 3. So, a wavelet neural network can be applied to approximate the M probability density functions in (21). The Figure 7 schematizes the demodulator:

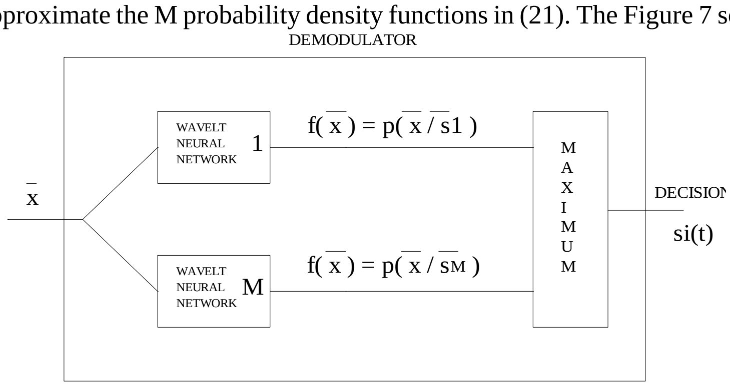


Figure 7

The main characteristic of this approach to the detection problem is that it is not needed to know the statistics and distribution of the noise. The only information needed is the periodical actualization of the $X(s_i)$ values in order to construct the training data set which guaranties the adaptability of the demodulator. The main advantage of this WNN compared to other neural network structures is the fast speed adaptability algorithm.

The proposed demodulator has been tested for several types of modulation systems and different kinds of noise. For binary signals the probability of error has resulted sligthly higher to the obtained in the classical MAP-based models[15], for equal conditions on the SNR per bit. Nevertheless, as the system modulation dimension, *n*, increases, the results have been more similar. This leads us to think of the better properties of the wavelet neural networks compared to other nets, when high dimensions are used.

5. CONCLUSIONS

A solution for the implementation of the MAP criterion in demodulators for digital communication systems have been proposed. The approach has been based on the wavelet neural network concept. These networks are inspired by both neural networks and the wavelet decomposition. The basic idea is to replace the neurons by wavelets, i.e., computing units obtained by cascading an affine transform and a multidimensional wavelet. Then these affine transforms and the weights have to be computed for a given noise corrupted input/output data.

The main characteristics of these networks consists on the following. First, as a direct byproduct of the wavelet decomposition, the “universal approximation” is guaranteed. Second, the availability of a direct and closed form formula for computing the decomposition is useful for designing the wavelet network coefficients. It is possible to obtain an algorithm that automatically determine the network size and estimate the network coefficients in a reasonable number of iterations.

The MAP criterion may be implemented based on the approximation properties of wavelet networks, and the correspondent model has been implemented. The system has been simulated and, for multidimensional signals, has achieved results on the probability error close to the real system values. Due to the adaptability to the changing distributions, the proposed system appears as one possible solution for the signal detection problem. Much work is left on the topic, but this prior result obtained are promising.

6. REFERENCES

1. C. K. Chui, Ed., *Wavelets: A Tutorial in Theory and Applications*. New York: Academic, 1992.
2. Vetterly M., Herley C. “Wavelets and Filter Banks: Theory and Design”, *IEEE Trans. on Signal Processing*, Vol. 40, pp. Sept. 1992.
3. Mallat S. G. “A Theory for Multirresolution Signal Decomposition: The Wavelet Representation” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 11, No. 7, pp. 674-693, July 1989.
4. Zurada, J.M. "Introduction to Artificial Neural Systems" West Publishing Company, 1992.
5. K. Hornik, “Multilayer feedforward networks are universal approximators”, *Neural Networks*, Vol. 2, 1989.
6. Qinghua Zhang, Albert Benveniste, “Wavelet Networks”, *IEEE Trans. on Neural Networks*, Vol. 3, No. 6, November 1992.
7. Y. C. Pati, P. S. Krishnaprasad, “Analysis and synthesis of feedforward neural networks using discrete affine wavelet transformations”, *IEEE Trans. on Neural Networks*, Vol. 4, pp. 73-85, Jan. 1993.
8. J. Hong, “Identification of stable systems by wavelet transform and artificial neural networks”, Ph. D. dissertation, Univ. Pittsburg, PA, 1992.
9. Qingua Zhang, “Using wavelet network in nonparametric estimation”, *IEEE Trans. on Neural Networks*, Vol. 8, No. 2, March 1997.
10. A. V. Oppenheim, *Signal and Systems*, Ed. Prentice-Hall, 1992.
11. J. B. Allen, L. R. Rabiner, “A unified approach to Short-Time-Fourier analysis and synthesis”, *Proc. of the IEEE*, Vol. 65, No. 11, November 1977.
12. I. Daubechies, *Ten Lecturas on Wavelets*, CBMS-NSF regional series in applied mathematics, Philadelphia, PA: SIAM, 1992.
13. T. Kugarajah, Q. Zhang, “Multidimensional wavelet frames”, *IEEE Trans. On Neural Networks*, Vol. 6, pp. 1552-1556, November 1995.
14. R. Hecht-Nielsen, “Backpropagation error surfaces can have local minima”, *IEEE-INNS Int. Joint Conference on Neural Networks*, Washington, D. C., June 1989.
15. B. Sklar, *Digital Communications*, Ed. Prentice-Hall, 1992.