

## THRESHOLDING WAVELET NETWORKS FOR SIGNAL CLASSIFICATION

NEMUEL D. PAH\* and DINESH KANT KUMAR

*RMIT University, Melbourne, Australia*

*\*s98102262@student.rmit.edu.au*

Received 25 October 2002

This paper reports a new signal classification tool, a modified wavelet network called Thresholding Wavelet Networks (TWN). The network is designed for the purposes of classifying signals. The philosophy of the technique is that often the difference between signals may not lie in the spectral or temporal region where the signal strength is high. Unlike other wavelet networks, this network does not concentrate necessarily on the high-energy region of the input signals. The network iteratively identifies the suitable wavelet coefficients (scale and translation) that best differentiate the different signals provided during training, irrespective of the ability of these coefficients to represent the signals. The network is not limited to the changes in temporal location of the signal identifiers. This paper also reports the testing of the network using simulated signals.

*Keywords:* Wavelet networks; neural networks.

### 1. Introduction

Pattern recognition and classification of signal and data are essential for our computerised society for automating the process of identifying a person, an object or an event. Pattern recognition consists of at least three phases: feature extraction, feature selection and classification.<sup>1</sup>

Feature extraction gives condensed representation of the input signal. Appropriate representation of the signal must contain all the important features that can be used to discriminate the signal pattern from the other signals. The purpose of representation is to reduce the complexities of the signal, while highlighting the properties that are suitable to recognize the pattern. Feature extraction involves transforming the signal while maintaining its completeness. The transformed data contain discriminative and other information. Selection of the transformation to be used is determined by initial knowledge of the discriminative pattern and space of the transformation.

To make the process of recognition more efficient, it is necessary to select suitable features that best identify the different patterns. Feature selection phase is necessary to separate the meaningful features that can best represent the input pattern with

minimum redundancy. The optimisation of the feature selection is controlled by its cost function.

The classification phase is the process to modify and map the selected features to the desired output format.

The development of Wavelet Transform has opened more opportunities in signal classification. Wavelet transform provides the temporal and spectral perspectives. It also provides the flexibility of choice of a wide variety of wavelet functions with properties well matched with the signal.

Wavelet transform has been widely used in the computation of many signal properties such as instantaneous frequency, detection of singularities, regularity measurement and fractal structures.<sup>2–11</sup> All these properties are significant in order to determine patterns of a class of signal.

The flexibility offered by the wavelet transform comes at the cost of having to select or adjust several number of input parameters. For suitable representation, the wavelet function, range of scale and translation have to be selected properly. This requires either prior knowledge of the signal and wavelet properties or experimentally determining the most suitable parameters. The latter is difficult due to computational complexity.

To overcome the problem, the efficacy of neural networks has been combined with wavelet transforms to generate wavelet networks.<sup>1</sup> The wavelet networks can generate wavelet coefficients iteratively to represent and classify the input signal. In the past few years, many types of wavelet networks have been introduced. Most of the wavelet networks are designed for the purpose of function approximation.<sup>12–24</sup>

Some researchers<sup>1,25–32</sup> have attempted to use the combination of neural networks with wavelet transforms to directly classify the signal. The input signal is classified by iteratively locating the positions in wavelet's time-scale space that best discriminate the input patterns. The inconsistent temporal alignment of the important events in the input signals often creates problem for this classification technique.

This paper reports our efforts to develop a new wavelet network that can overcome the limitations of previous wavelet networks. This wavelet network introduces an upper-lower thresholding process that allows the network to select the suitable wavelet coefficients based on signal properties. The selection is not depending on the temporal alignment of the coefficients.

This paper is organised into six sections. Section 2 reviews the basic concept of wavelet transform while Sec. 3 reviews the different types of wavelet networks. Section 4 presents the proposed thresholding wavelet network. Section 5 presents the simulation of the network, while Sec. 6 concludes the paper.

## 2. Wavelet Transform

Wavelet transform of a function  $f \in L^2(R)$  is defined as the correlation between  $f$  and a dilated wavelet function  $\psi$ .

$$Wf(u, s) = \langle f, \psi_{u,s}^* \rangle, \quad (1)$$

$$Wf(u, s) = \int_{-\infty}^{+\infty} f(t) \frac{1}{\sqrt{2}} \psi^* \left( \frac{t-u}{s} \right) dt. \quad (2)$$

The magnitude of wavelet coefficient  $|Wf(u, s)|$  will be higher if the properties of the function  $f$  in the neighbourhood of  $t = u$  match the properties of the wavelet function  $\psi_{u,s}^*$  at scale  $s$ . By selecting a wavelet function with a certain properties (temporal or spectral), one can extract the portion of the analysed signal that match the properties of the wavelet function being used. The magnitude of wavelet coefficients contains information related to certain properties in the analysed signal such as the instantaneous frequency and regularity.<sup>2,4</sup> The maxima of wavelet transform can also be used to approximate signal as well as removing noises.<sup>2,33,34</sup>

### 2.1. Instantaneous frequency

Wavelet transform can measure time-frequency variation of spectral components with a non-constant time frequency resolution.<sup>2</sup> The time-frequency atom of a wavelet function varies according to scale  $s$ . A wavelet atom

$$\psi_{u,s}(t) = \frac{1}{\sqrt{2}} \psi \left( \frac{t-u}{s} \right) \quad (3)$$

is centred at  $t = u$  with time spreading proportional to scale  $s$ . In frequency domain, the atom is centred at  $\eta/s$  with a size proportional to  $1/s$ . The centre frequency at scale  $s = 1$  is defined as:

$$\eta = \frac{1}{2\pi} \int_0^{+\infty} \omega |\hat{\psi}(\omega)|^2 d\omega. \quad (4)$$

The ability of wavelet atom to increase its time resolution in high frequency, and increase its frequency resolution in low frequency makes it able to follow the rapid change in spectral component with good resolution.

The instantaneous frequency can be detected from the ridges<sup>2</sup> of normalised scalogram:

$$\frac{\xi}{\eta} P_W f(u, \xi) = \frac{|Wf(u, s)|^2}{s} \quad \text{for } \xi = \frac{\eta}{s}. \quad (5)$$

The ridges  $(u, \xi(u))$  are the maxima points at each translation  $u$ . The magnitude of the normalised scalogram (5) corresponds to the amplitude of the particular spectral components. If the coefficients were calculated using analytic wavelet, the phase and amplitude information of the analysed signal are well separated.

### 2.2. Regularity measurement

Most signals are characterised by their regularity pattern.<sup>2</sup> Wavelet coefficients can be used to calculate pointwise regularity of a signal as well as singularity spectrum. A Lipschitz exponent (can be estimated from the slope of wavelet coefficients maxima towards finest scale.<sup>2</sup>

A function  $f$  is pointwise Lipschitz  $\alpha \geq 0$  at  $t = \nu$  if there exist  $K > 0$ , and a polynomial  $p_m$  of degree  $m = \lceil \alpha \rceil$  such that:

$$|f(t) - p_m(t)| \leq K|t - \nu|^\alpha. \quad (6)$$

Wavelet transform estimates the Lipschitz exponent  $\alpha$  by using wavelet with  $n = \lceil \alpha \rceil$  vanishing moments. The wavelet transform of the function  $f = p_m + e_\nu$  only depends on the singularity  $e_\nu$  since a wavelet with  $n$  vanishing moments is orthogonal to any polynomial of degree  $m \leq n - 1$ .

$$Wf(u, s) = We_\nu(u, s). \quad (7)$$

Lipschitz regularity is related to the decay of the magnitude of wavelet coefficients  $|Wf(u, s)|$  and the singularity point is located by finding the abscissa  $\nu$  where wavelet modulus maxima converge to at the finest scale.

The global singularity of a signal with many non-isolated singularities can also be measured from the decay of wavelet modulus maxima.

### 2.3. Maxima of wavelet coefficients

Sections 2.1 and 2.2 show that the location and the magnitude of wavelet transform maxima carry information related to the spectral component of the signal as well as its singularities. Wavelet modulus maxima at scale  $s_0$  is defined as any points  $(u_0, s_0)$  such that  $|Wf(u_0, s_0)|$  satisfy:

$$\frac{\partial Wf(u_0, s_0)}{\partial u} = 0. \quad (8)$$

By modifying wavelet maxima, one can approximate a signal at different level of approximation, remove noises or extract part of the signal that relates to specific category. Thresholding wavelet coefficients can effectively select the way a signal is going to be approximated for a specific purpose.

## 3. Wavelet Networks

A wavelet network is a class of neural networks that combines wavelet transform with neural network algorithm.<sup>1</sup> This combination provides a tool to perform wavelet transform analysis in a parallel, adjustable and able-to-learn fashion.

Wavelet transform calculates the correlations between  $f$  and wavelet function  $\psi$  at various different combination of translation and scale  $(u, s)$ . The calculation can be faster if it is done in parallel computation, where each unit calculates only one wavelet coefficient.

In wavelet analysis, there are many parameters to be adjusted such as the wavelet function, scale, thresholding level, as well as maxima tracking. Wavelet network gives the flexibility of adjusting these parameters in parallel computation, and based on patterns that are not *a priori* defined. Wavelet networks also remove the uncertainty by providing an iterative learning based approach where these parameters can be determined by learning from examples.

There are many different types of wavelet networks. In general they can be grouped into two major streams<sup>1</sup>: approximation wavelet networks and classification wavelet networks.

### 3.1. Approximation wavelet networks

Approximation Wavelet Network (AWN)<sup>12–24</sup> is based on nonlinear wavelet approximation. The nonlinear approximation of a function  $f$  from  $M$  wavelet coefficients indexed with  $I_M$  is defined as<sup>2</sup>:

$$f_M = \sum_{(j,n) \in I_M} \langle f, \psi(j,n) \rangle \psi(j,n). \quad (9)$$

The selection of the set of wavelet coefficients  $\{(j,n)_m\}_{m \in I_M}$  is generally based on the effort to minimise the approximation error:

$$\epsilon[M] = \|f - f_M\|^2. \quad (10)$$

A small approximation error can be achieved with a minimal number of wavelet coefficients if the transformation uses orthogonal wavelets. However, if the network can maintain the reduction of error and redundancy, it is not necessary to use orthogonal wavelets. Non-orthogonal wavelets are needed especially in analysing signals with many non-isolated singularities. It is also possible to use different type of wavelets in one network to approximate the function  $f$ . In this case the wavelet functions are most likely to be non-orthogonal.

An AWN is commonly implemented as three layers neural network. The output layer has only one node with a linear transfer function. The hidden layer has  $M$  wavelet nodes with the wavelet function as their transfer function.

$$\begin{aligned} f_M &= \sum_{(j,n) \in I_M}^M W_{j,n} \psi_{j,n} + b_0 = \sum_{(j,n) \in I_M}^M W_{j,n} \psi\left(\frac{t - 2^j n}{2^j}\right) + b_0 \\ &= \sum_{(j,n) \in I_M}^M W_{j,n} \psi(2^{-j}t - n) + b_0. \end{aligned} \quad (11)$$

The wavelet coefficients are represented by weighting factors  $W_{j,n}$  between hidden and output layer.

$$W_{j,n} \approx \langle f, \psi_{j,n} \rangle. \quad (12)$$

The weighting factor of the hidden nodes  $w_h$  determines the scale of the wavelet function while its bias  $b_h$  determines the time translation.

$$W_h = \{2^{j_h}\}_{h=1,2,\dots,M} \quad \text{and} \quad b_h = \{n_h\}_{h=1,2,\dots,M}. \quad (13)$$

Bias of output node  $b_0$  is necessary to approximate function with non-zero average. Figure 1 shows the basic architecture of wavelet networks for approximation.

During the networks learning process, the value of the weights and biases are optimised to determine the scale and translation of each wavelet function that best

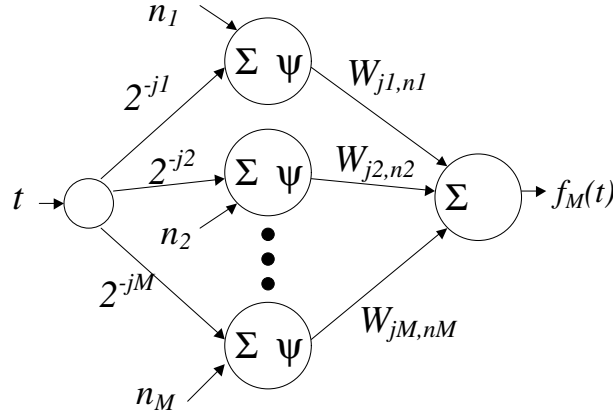


Fig. 1. The architecture of approximation wavelet networks.

approximates the function  $f$ . Some of the learning algorithms that have been used in the optimisation of the parameters are gradient descent,<sup>18,19,23,24,34</sup> competitive learning,<sup>35</sup> dyadic selection,<sup>12</sup> evolutionary computation<sup>36</sup> and genetic algorithm.<sup>37</sup> The number of wavelet nodes  $M$  can be optimised by using an algorithm such as competitive learning.

Like all approximation processes, the wavelet network optimises its parameters by minimising the energy of its error  $\varepsilon[M]$ . The selection of wavelet coefficient is therefore, based on the coefficients magnitude. The network tends to prioritise high-energy regions of  $f$ . If the approximation  $f_M$  is to be used in classification purposes, it is only effective if the deterministic factors are immersed in the high-energy regions of the signal. There are number of natural signals where this may not be the case.

### 3.2. Classification wavelet networks

Classification Wavelet Network (CWN)<sup>1,25–32</sup> uses neural networks algorithm to select the locations of wavelet coefficient  $(u, s)$  that are reliable for categorising the input signals. The CWN, as shown in Fig. 2, can be seen as a neural network, which the first layer consists of wavelet nodes followed by some neural network layers. The wavelet layer acts as a feature extraction layer while the neural network layers are for feature selection and classifier.

The output of each wavelet node  $\varphi^W$  is defined as the inner product of the input signal  $f$  and a wavelet function at a specific time-scale  $\psi_{u_0, s_0}$ . The output  $\varphi^W$  is actually a wavelet coefficient at scale  $s_0$  and translation  $u_0$ .

$$\varphi_0^W = \langle f, \psi_{u_0, s_0} \rangle = Wf(u_0, s_0). \quad (14)$$

The subsequent layers work as the standard neural network layers. The output of the neural network  $\varphi^N$  is defined as the response of the transfer function of the neural network,  $f_T$ , to the correlation between the output of the wavelet layer,  $\varphi^W$

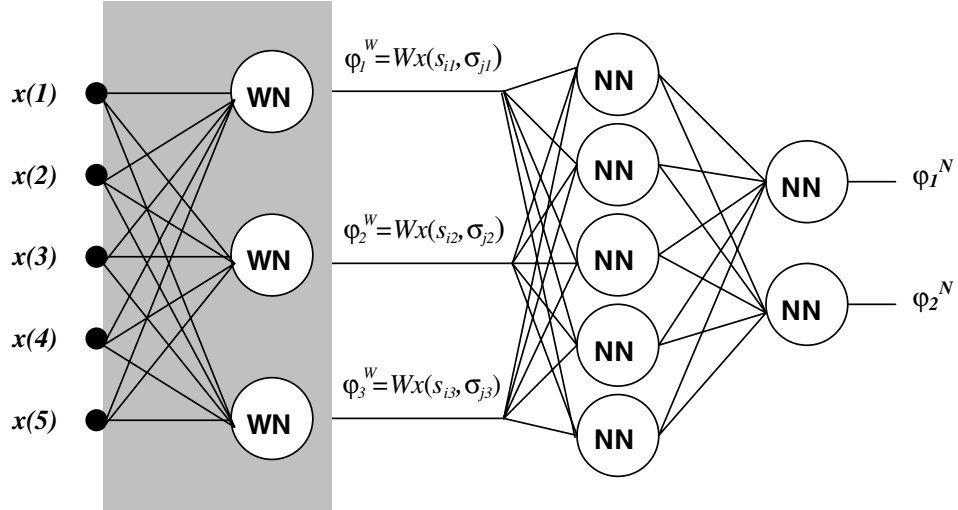


Fig. 2. The architecture of classification wavelet networks. The first layer consists of wavelet nodes which calculates wavelet coefficients of the input  $x(n)$ . The subsequent layers are neural network layers.

and weighting factors  $w$  and bias  $b$ .

$$\varphi^N = f_T \left( b + \sum_{n=1}^K \varphi_n^W w_n \right). \quad (15)$$

During the learning phase, the parameters of wavelet nodes  $u$  and  $s$  and the weights and biases of the neural network's nodes are adjusted by backpropagating the error of the neural network. The learning process of each wavelet node can be seen as the movement of a wavelet's Heisenberg box in time-scale space from its initial position  $(u_i, s_i)$  to its optimum position  $(u_0, s_0)$  as shown in Fig. 3.

The number of nodes in wavelet layer  $K$  determines the number of wavelet Heisenberg boxes. The number of nodes  $K$  as well as the initial value of the parameters  $(u_i, s_i)$  have to be selected carefully for a specific classification problem. The selection of initial value is significant to ensure that the nodes are not overlapping each other during the learning phase.

There are two main disadvantages of this wavelet networks; temporal dependence of classification and lack of optimisation for classification. The resultant network have fixed spots in time-scale space  $\{(u_n, s_n)\}_{n=1,2,3,\dots,K}$ . Consequently, the input signal must be precisely aligned to ensure the consistent temporal location of the signals important events. For signals such as biosignals, this is a major hurdle as the signal is non-stationary and localisation of events is difficult. The effect of this disadvantage can be observed from the fact that the resultant will have a large number of wavelet nodes  $K$  covering nearly the whole time duration of the input signal with weak links to the subsequent layer (small  $|w_n|$ ).

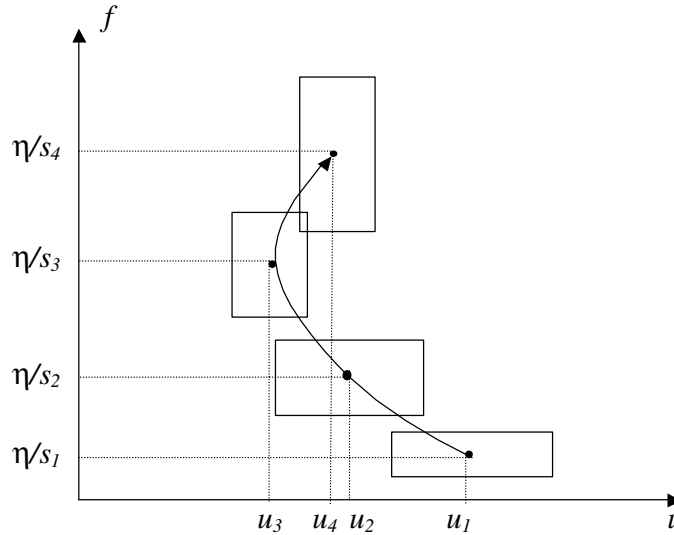


Fig. 3. Learning process in each wavelet nodes is illustrated as the movement of a wavelet Heisenberg box from initial position  $(u_i, s_i)$  to its optimum position  $(u_o, s_o)$ .

The second disadvantage is that this network forces the subsequent neural network layers to classify input signal based only on the values of the wavelet coefficients  $Wf(u, s)$ . Processing of these coefficients before classification may help optimise the classification process and overcome these problems.

#### 4. Thresholding Wavelet Networks

To overcome the above disadvantages of wavelet networks in classifying non-stationary, semi-random signals, a new type of wavelet network called Thresholding Wavelet Network (TWN) has been designed. Wavelet thresholding has been widely used in the past in wavelet signal estimation and de-noising.<sup>33,38</sup>

The TWN, reported in this paper, uses wavelet thresholding in a different approach. The TWN was designed for classification purpose. This network selects the wavelet nodes based on the thresholding of the local maxima in each of the scale bands of the Continuous Wavelet Transform (CWT).

Wavelet maxima are the wavelet coefficients that satisfy Eq. (8). In discrete domain this zero derivative is nearly impossible. For that reason, wavelet maxima are selected where the partial derivation changes sign from positive to negative.

The wavelet maxima are thresholded at the thresholding layer. This process selects the wavelet maxima based on their magnitude in relation to the thresholding levels. Maxima with magnitude less than the thresholding level are removed while maxima with higher magnitude are propagated.

The most commonly used thresholding functions are soft-thresholding function and hard-thresholding function.<sup>33</sup> The soft-thresholding of  $x$  is defined as:



$$x_s = \eta_s(x, \theta) \operatorname{sgn}(x) (|x| - \theta)_+, \quad (16)$$

where  $\theta > 0$  is the threshold, while hard thresholding of  $x$  is defined as:

$$x_h = \eta_h(x, \theta) = \begin{cases} x & \text{for } |x| > \theta, \\ 0 & \text{for } |x| \leq \theta. \end{cases} \quad (17)$$

#### 4.1. The significance of thresholding

Most signals can be characterised by the occurrence of transients such as QRS-complex in electrocardiography and motor unit action potential (MUAP) in electromyography.<sup>39</sup> These transients can be characterised by their properties such as magnitude, frequency distribution, regularity and polynomial degree. It is important to use these features for the classification of such signals. Wavelet transform can be used to measure the spectral content, transients, regularity and polynomial degree of a signal while thresholding extracts these features with respect to the specific magnitudes and thus are an obvious choice.

To classify the signals, the occurrence of these transients must be detected and analysed. The temporal locations of the transients are often not necessary in the classification. By thresholding the wavelet coefficients, the temporal location may be lost but the features related to the occurrence of the events are extracted.

It is known that the position of wavelet coefficients maxima stores information about the frequency and singularity point. The question will be, what information is stored in the magnitude of these maxima?

The magnitude of wavelet coefficients  $|Wf(u, s)|$  are related to the magnitude of spectral components in the analysed signal  $f$ . Suppose:

$$f(t) = a(t) \cos \phi(t), \quad (18)$$

where  $a(t)$  is the amplitude and  $\omega(t) = \phi'(t)$  is instantaneous frequency.

The amplitude  $a(t)$  recovered from wavelet coefficients when using a nearly analytic wavelet  $\psi(t) = g(t)e^{int}$  is calculated as

$$a(u) = \frac{2\sqrt{s^{-1}|Wf(u, s)|^2}}{|\hat{g}(0)|}, \quad (19)$$

where  $u = t - t_0$  is the time shift and  $\hat{g}(0)$  is the Fourier transform of window  $g(t)$  at  $\omega = 0$ . If the coefficients are not calculated with analytic wavelet the magnitude and phase are not well separated. However, the magnitude of wavelet coefficients are still related to the spectral amplitude  $a(t)$ .

$$a(t - t_0) \propto |Wf(u, s)|. \quad (20)$$

By thresholding the wavelet coefficient, TWN can classify signals based on the occurrence and magnitude of events or transients that have specific spectral properties.

The magnitude of wavelet coefficients is also related to the energy of signal with specific singularity. Equations (21)–(24) examine this relationship. Regularity  $\alpha$  is measured from the decay of wavelet coefficients towards finest scale.

$$\alpha \propto \lim_{s \rightarrow 0} \left( -\frac{\partial |Wf(u, s)|}{\partial s} \right). \quad (21)$$

It is proven in Ref. 2 that if  $f$  has Lipschitz regularity  $\alpha$  at  $t = \nu$  and  $u$  is in the neighbourhood of the point  $t = \nu$ , then

$$|Wf(u, s)| \leq K s^{\alpha+1/2}. \quad (22)$$

The magnitude of wavelet coefficient at  $t = u$  at a specific scale  $s_0$ , is related to the energy of the signal in the temporal region of the singularity. Suppose a function  $f$  has regularity  $\alpha$  at  $t = \nu$ , the function  $f$  can be approximated with a polynomial of degree  $n = \lfloor \alpha \rfloor$  and an error function  $\varepsilon$ :  $f = p_\nu + \varepsilon_\nu$ . If the function is then amplified by a factor of  $A$ , its wavelet coefficient at  $t = \nu$  will also be amplified by the same amount.

$$Af = Ap_\nu + A\varepsilon_\nu, \quad (23)$$

$$\begin{aligned} WAf(\nu, s) &= 0 + WA\varepsilon(\nu, s) \\ &= \frac{1}{\sqrt{2}} \int_{-\infty}^{+\infty} A\varepsilon(t) \psi \left( \frac{t - \nu}{s} \right) dt, \\ WAf(\nu, s) &= AW\varepsilon(\nu, s). \end{aligned} \quad (24)$$

By thresholding the coefficient, the portion of the input signal with a specific singularity at certain amplification can be selected.

Thresholding of wavelet coefficients has numerous applications. One of the applications is signal denoising.<sup>33</sup> Denoising is a process of estimating a class of function from its corrupted version. It is based on the assumption that the energy of the uncorrupted function is concentrated in a small number of coefficients while the noise is spreading over a large number of wavelet coefficients.<sup>34</sup> If  $d$  represents the corrupted version of the original signal  $f \in [0, 1]$ , it is expressed as:

$$d_i = f(t_i) + \sigma z_i \quad \text{for } i = 1, 2, 3, \dots, n - 1. \quad (25)$$

The noise  $z_i \stackrel{iid}{\sim} N(0, 1)$  is approximated by an independent and identically distributed Gaussian function of level  $\sigma$ .

The other assumption is that the energy of noise is less than the energy of the uncorrupted function. For this reason, de-noising process is often interpreted as optimising the mean squared error  $(E\|d_i - f_i\|^2)/n$  of the estimation.

When thresholding is used for classification, the above assumptions are not always true. Noise in the classification perspective is the section of the signal or function that does not contain information suitable to categorise the signal. This type of noise can often have higher energy content compared to the signal. As the consequence, the single cutoff thresholding process is not suitable. Thresholding in

a classification process must include upper–lower threshold to select the coefficients belonging to the middle range energy. The upper–lower thresholding function is defined as:

$$x_{\text{th}} = \eta_{\text{th}}(x, \theta_l, \theta_h) = \begin{cases} x & \text{for } \theta_l < x < \theta_h \\ 0 & \text{for } x \leq \theta_l \text{ or } x \geq \theta_h. \end{cases} \quad (26)$$

#### 4.2. Architecture

Thresholding wavelet networks consist of four blocks of network layer: a wavelet layer, maxima layer, thresholding layer and neural networks layers (Fig. 4). The input signal is applied to the wavelet layer. The input to the thresholding layer is the wavelet maxima at each scale of interest. The thresholds  $(\theta_l, \theta_h)$  for all maxima corresponding to one scale are the same. The scale-dependent-threshold allows the network to distinguish between scales based on magnitude. The output of wavelet thresholding node  $\varphi$  is the number of wavelet maxima with magnitude between  $\theta_l$  and  $\theta_h$ .

$$\varphi = \sum_{i=1}^N \eta_T(|Wf(u_i, s)|, \theta_l, \theta_h). \quad (27)$$

$N$  is the number of wavelet maxima for that particular scale  $s$ , while  $\eta_T$  is the thresholding function. The output of the thresholding node is the number of wavelet maxima with magnitude between  $\theta_l$  and  $\theta_h$ . The output of threshold function  $\eta_T$  in Eq. (27) equals to one if its input is between  $\theta_l$  and  $\theta_h$  and equals to zero otherwise.

$$x_T = \eta_T(x, \theta_l, \theta_h) = \begin{cases} 1 & \text{for } \theta_l < x < \theta_h, \\ 0 & \text{for } x \leq \theta_l \text{ or } x \geq \theta_h. \end{cases} \quad (28)$$

This thresholding function also has to be a continuous function so that the derivative of the function exists. Thus, a Gaussian function  $g(x) = e^{-x^2}$  was used as the thresholding function as below:

$$x_T = \eta_T(x) = e^{-\alpha^2(x-\beta)^2}. \quad (29)$$

The output of the thresholding node is:

$$\varphi = \sum_{i=1}^N e^{-(\alpha(|Wf(u_i, s)|-\beta))^2}. \quad (30)$$

The bias  $\beta$  determines the centre of  $g(x)$  while  $\alpha$  determines the width of the function. The combination of  $\alpha$  and  $\beta$  determines the upper and lower thresholds. Figure 5 illustrates the thresholding wavelet node.

Depending on the application and the complexities of the input signals, each scale may have several thresholding nodes as shown in Fig. 4.

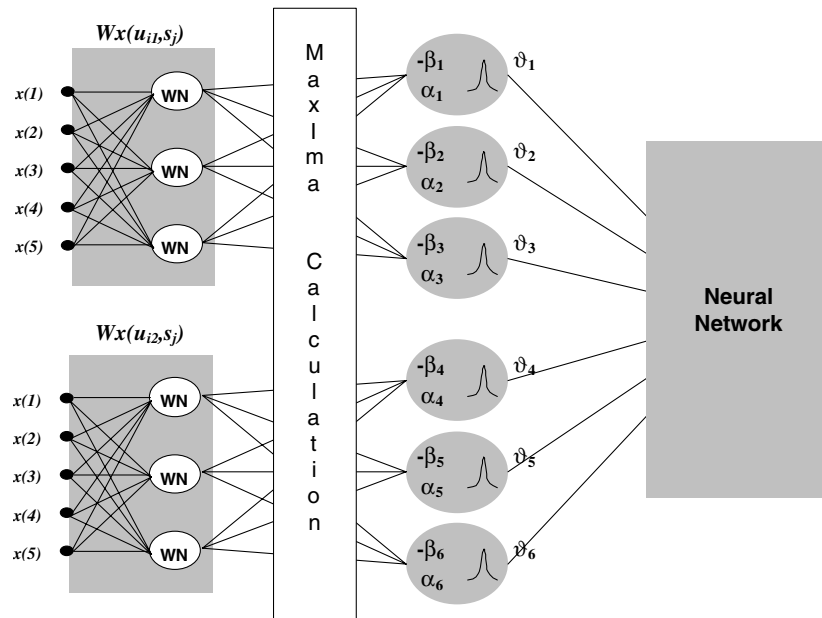


Fig. 4. A thresholding wavelet network. The network has six thresholding nodes (three for each scale).

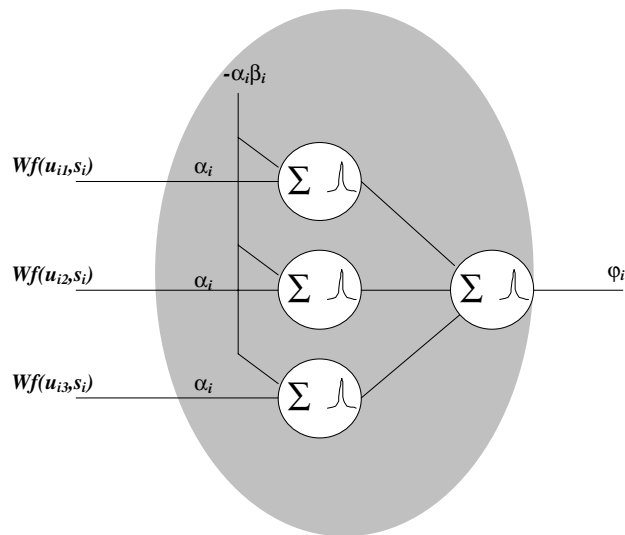


Fig. 5. The wavelet thresholding node.

### 4.3. Learning process

The parameters of this network ( $\alpha$ ,  $\beta$  and neural network weights and biases) were initialised with random value. During the iterative learning process the value of

these parameters were changed to reduce the classification error. The network was using a supervised learning algorithm. In the supervised learning algorithm, the network's parameters are adjusted to properly relate the inputs  $f(t)$  and target output  $\varphi_T$  provided in the training set  $\{(f_1(t), \varphi_{T1}), (f_2(t), \varphi_{T2}), \dots, (f_n(t), \varphi_{Tn})\}$ .

In each thresholding node, the learning is the process to locate the lower and upper threshold levels ( $\theta_l$  and  $\theta_h$ ) of each node. Determining the optimum value of the threshold parameters ensures number of wavelet coefficients with magnitude  $\theta_l \leq |Wf(u, s)| \leq \theta_h$  can best categorise the class of input signals.

There are many algorithms currently available that can be used to optimising these parameters. The TWN reported in this paper was using the gradient descent algorithm. In gradient descent algorithm, the parameters are optimised to the negative direction of partial derivation of its cost function  $SSE$  to each parameter.

$$\alpha_{\text{new}} = \alpha_{\text{old}} - \rho \frac{\partial SSE}{\partial \alpha_{\text{old}}}, \quad (31)$$

$$\beta_{\text{new}} = \beta_{\text{old}} - \rho \frac{\partial SSE}{\partial \beta_{\text{old}}}. \quad (32)$$

The cost function  $SSE$  is defined as the sum-squared of the difference between target output  $\varphi_T$  and the actual output  $\varphi$ .

$$E = \frac{1}{2}(\varphi_T - \varphi)^2. \quad (33)$$

The amount of change in each iteration is determined by the partial derivation of the cost function and a learning rate coefficient  $\rho$ .

The partial derivations of the cost function are:

$$\frac{\partial SSE}{\partial \alpha} = 2(\varphi_T - \varphi) \sum_{u,s} \alpha (|Wf(u, s)| - \beta)^2 e^{-(\alpha|Wf(u,s)| - \beta)^2}, \quad (34)$$

$$\frac{\partial SSE}{\partial \beta} = -2(\varphi_T - \varphi) \sum_{u,s} \alpha^2 (|Wf(u, s)| - \beta) e^{-(\alpha|Wf(u,s)| - \beta)^2}. \quad (35)$$

The learning process is repeated until the sum-squared-error  $SSE$  falls below a predefined maximum error  $E_T$ . At this stage the network is considered as able to classify the training pattern with an error less than  $E_T$ .

## 5. Network Simulation

An experiment was conducted to determine the ability of this network to classify simulated signals. There were 160 simulated signals created for the experiments. The length of each signal was 1,000 samples. These signals had the following properties:

- (1) Consists of number of transients representing the current distribution of muscle fibre action potential (MUAP),  $I_m$ .<sup>40,41</sup>

$$I_m = CA\lambda^2(\lambda z)[6 - 6\lambda z + (\lambda z)^2]e^{-\lambda z} \quad (36)$$

with  $A = 96$  mV,  $C$  and  $\lambda$  are both 1.

- (2) The signals represent two classes each of 80 signals.
- (3) Signals in class A (Fig. 6(a)) consist of 100 transients of magnitude 0.1 and 100 transients of magnitude of 0.04. These transients were located randomly.
- (4) Signals in class B (Fig. 6(b)) were very similar to signals belonging to class A. These signals consist of 100 transients of magnitude 0.1, 100 transients of magnitude of 0.04 and 10 transients of magnitude 0.07. These transients were located randomly.

The difference between the two classes was not observable in the plot of the signals (Fig. 6) as well as in their spectrum (Fig. 7) since the difference was determined by a small (10) number of transient with a low amplification.

The proposed thresholding wavelet network was simulated using graphic user interface of Matlab 6.1 as shown in Fig. 8. 40 signals of each class were used to train the network. These signals were scaled by a factor of 100 before being applied to the network to avoid quantisation error in the network calculation.

The training process was recorded in Fig. 8. It was observed that after 200 epochs the *SSE* of the network drops from about 10 to 0.1. The graphs on top-left

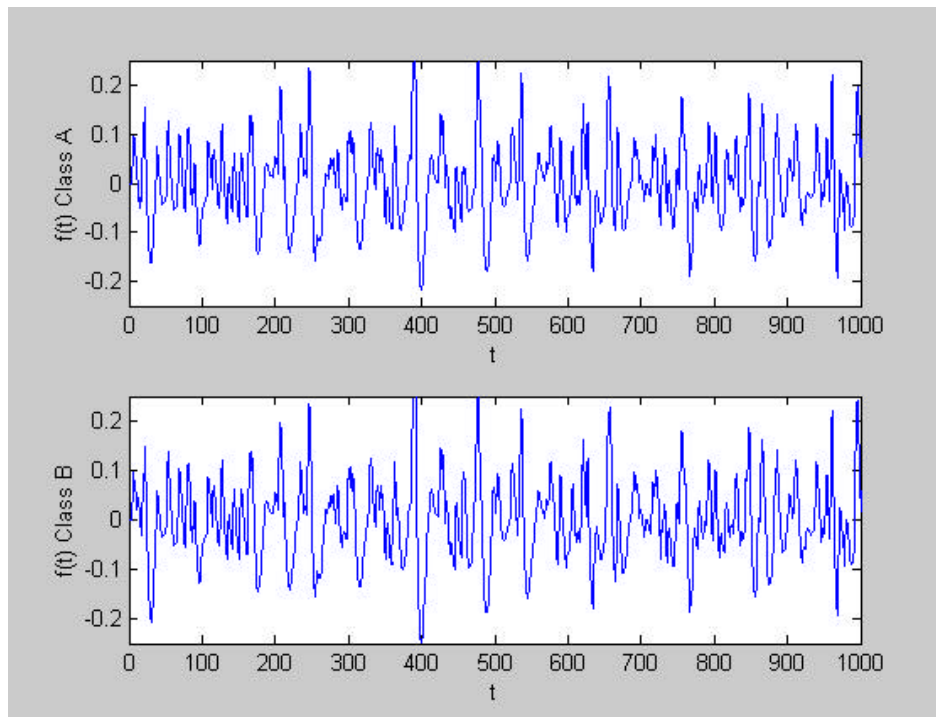


Fig. 6. (a) A signal of Class A consists of 100 wave  $I_m$  with gain of 0.1 and 100 wave  $I_m$  with gain of 0.04. (b) A signal of Class B, similar to that of Class A with extra 10 waves with gain of 0.07.

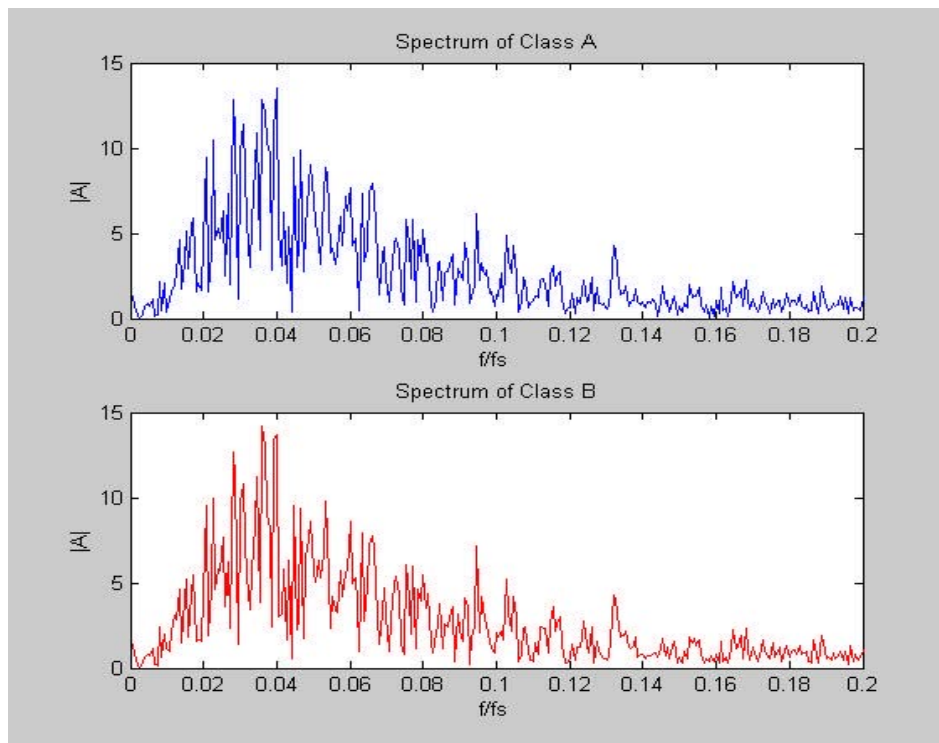


Fig. 7. The spectrum of simulated signals from Class A and B.

and bottom-left corner of Figure 11 show that the biases ( $\beta$ ) of thresholding wavelet nodes converged to four values (or ranges): 0.3, 0.59–0.66, 0.91–0.92, and 1.3 and four of these converged to one point 0.91–0.92. It can also be observed that the four nodes that converge to a common value have higher weights ( $\alpha$ ) and thus have narrower thresholding function.

After completing the training process, the network was tested using the 40 test signals. The results are presented in Fig. 9. The signals that belongs to group A are marked with “o” while that of group B are marked with “\*”. From the figure it is observed that the wavelet network is able to converge the thresholding values to best classify the signals and in this example, 100% classification is achieved.

## 6. Conclusion

This paper presents a new type of wavelet network, the Thresholding Wavelet Network. It is designed for the purpose of classifying signals based on the magnitude of wavelet maxima.

The network is a combination of wavelet transform, thresholding process and neural network. It adaptively adjusts its thresholding level during its learning

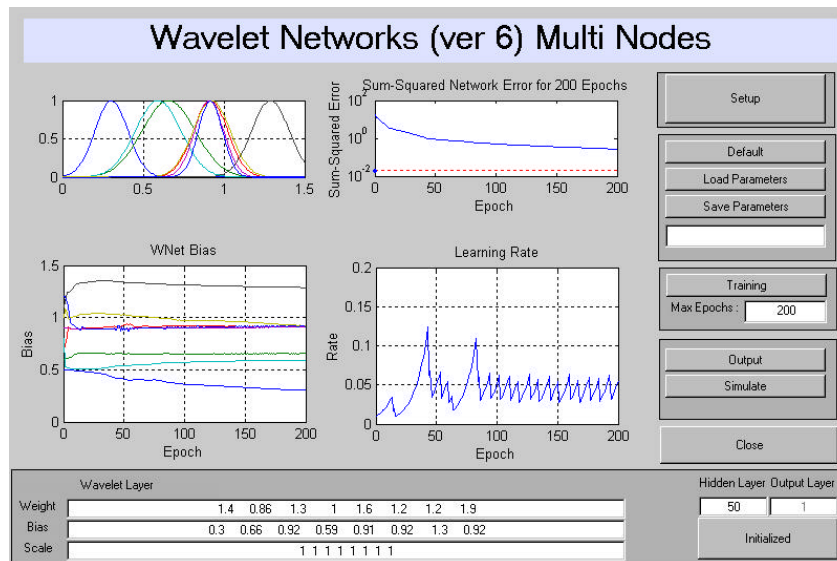


Fig. 8. The learning process of thresholding wavelet network. The network is trained to classify 2 groups of signals (see Fig. 9) each consists of 20 signals. From top left corner clockwise: The thresholding functions of eight nodes after 200 epochs; the recording of SSE of the entire network; the recording of learning rate; the recording of bias of each thresholding nodes which indicates the centre of the thresholding function.

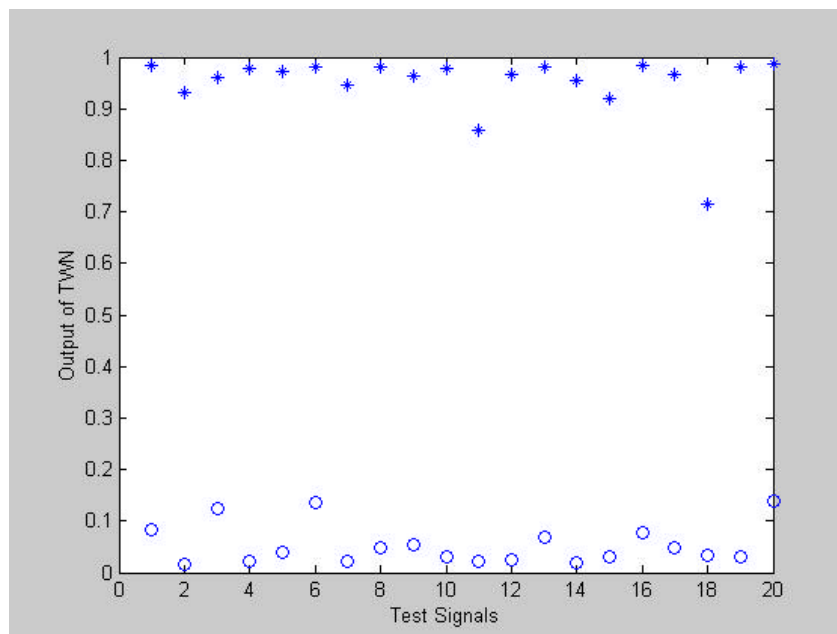


Fig. 9. The testing result of the Thresholding Wavelet Network. The signals that belong to Class A are marked with "o" while that of Class B are marked with "\*".



process to select wavelet maxima with certain magnitude that characterised the input signals.

The paper reports the simulation and testing of the network and the results are extremely promising. The results demonstrate several advantages of thresholding wavelet networks.

- (1) It classifies the signals irrespective of the difference being in the low or high-energy region of the signal.
- (2) It can extract the features of transients occurs in a signal.
- (3) It is dynamic and classification is not affected by the change in temporal location of the event.

This network has to be tested with real signals to determine the efficacy of the network. One shortcoming of this network is that the scale and wavelet basis function are invariant. The authors intend to include selection of scale and wavelet basis in its learning process.

## References

1. H. Dickhaus and H. Heinrich, *Classifying biosignals with wavelet networks*, in *IEEE Engineering in Medicine and Biology* (IEEE, 1996), pp. 103–111.
2. S. Mallat, *A Wavelet Tour of Signal Processing. 1999, London* (Academic Press, 1999).
3. W.-L. Hwang and S. Mallat, *Characterization of Self-Similar Multifractals with Wavelet Maxima* 1993, Courant Institute of Mathematical Sciences, Computer Science Department, New York University, pp. 1–25.
4. S. Mallat and W. L. Hwang, *Singularity detection and processing with wavelets*, in *IEEE Transaction on Information Theory* **38**(2) (1992) 617–643.
5. A. Arneodo *et al.*, *Singularity spectrum of multifractal function involving oscillating singularities*, in *J. Fourier Anal. Appl.* **4**(2) (1998) 159–174.
6. P. Goncalves, R. Reidi and R. Baraniuk, *A simple statistical analysis of wavelet-based multifractal spectrum estimation*, INRIA Rocquencourt (1998) pp. 1–5.
7. J. F. Muzy, E. Bacry and A. Arneodo, *The multifractal formalism revisited with wavelets*, in *Int. J. Bifurcation Chaos* **4**(2) (1994) 245–302.
8. F. H. Y. Chan *et al.*, *Multiscale characterization of chronobiological signals based on discrete wavelet transform*, in *IEEE Trans. Biomedical Engrg.* **47**(1) (2000) 88–95.
9. J. Y. Tham *et al.*, *A general approach for analysis and application of discrete multi-wavelet transform*, in *IEEE Trans. Signal Processing* **48**(2) (2000) 457–464.
10. I. W. Selesnick, *Balanced multiwavelet bases based on symmetric FIR filters*, in *IEEE Trans. Signal Processing* **48**(1) (2000) 184–191.
11. S. Karlsson, J. Yu and M. Akay, *Time-frequency analysis of myoelectric signals during dynamic contractions: A comparative study*, in *IEEE Trans. Biomedical Engrg.* **47**(2) (2000) 228–238.
12. K. Hwang *et al.*, *Characterisation of gas pipeline inspection signals using wavelet basis function neural networks*, in *NDT&E Int.* **33** (2000) 531–545.
13. S. Dohler and L. Ruschendorf, *An approximation result for nets in functional estimation*, in *Statist. Probab. Lett.* **52** (2001) 373–380.
14. D. H. C. Ho, P. A. Zhang and J. Xu, *Fuzzy wavelet networks for function learning*, in *IEEE Trans. Fuzzy Systems* **9**(1) (2001) 200–211.

15. Y. Fang and T. W. S. Chow, *Orthogonal wavelet neural networks applying to identification of Wiener model*, in *IEEE Trans. Circuit Systems* **47**(4) (2000) 591–593.
16. C. C. Holmes and B. K. Mallick, *Bayesian wavelet networks for nonparametric regression*, in *IEEE Trans. Neural Networks* **11**(1) (2000) 27–35.
17. C. Bernard, S. Mallat and J. J. Slotine, *Wavelet Interpolation Network*.
18. Q. Zhang and A. Benveniste, *Wavelet networks*, in *IEEE Trans. Neural Networks* **3**(6) (1992) 889–898.
19. L. Jiao, J. Pan and Y. Feng, *Multiwavelet neural network and its approximation properties*, in *IEEE Trans. Neural Networks* **12**(5) (2001) 1060–1066.
20. Q. Zhang, *Using wavelet network in nonparametric estimation*, in *IEEE Trans. Neural Networks* **8**(2) (1997) 227–236.
21. D. Clancy and U. Ozguner, *Wavelet neural networks: A design perspective*, in 1994 *IEEE Int. Symp. Intelligent Control*, Columbus, Ohio, USA (IEEE, 1994).
22. J. Zhang *et al.*, *Wavelet neural networks for function learning*, in *IEEE Trans. Signal Processing* **43**(6) (1995) 1485–1497.
23. K. Kobayashi, *Self-organizing wavelet-based neural networks*, in *Time Frequency and Wavelets in Biomedical Signal Processing*, ed. M. Akay (IEEE Press, 1998), pp. 685–702.
24. H. Heinrich and H. Dickhaus, *Analysis of evoked potentials using wavelet networks*, in *Time Frequency and Wavelets in Biomedical Signal Processing*, ed. M. Akay (IEEE Press, 1998), pp. 669–684.
25. T. Kalayci and O. Ozdamar, *Wavelet preprocessing for automated neural network detection of EEG spike*, in *IEEE Engrg. Medicine Biology* **14** (1995) 160–166.
26. A. J. Hoffman and C. J. A. Tollig, *The Application of Classification Wavelet Networks to the Recognition of Transient Signals* (IEEE, 1999), pp. 407–410.
27. S. C. Woo, C. P. Lim and R. Osman, *Development of a speaker recognition system using wavelets and artificial neural networks*, in *Proc. of 2001 Int. Symp. on Intelligent Multimedia*, Video and Speech Recognition, 2001, Hong Kong.
28. L. Angrisani, P. Daponte and M. D’Apuzzo, *A method based on wavelet networks for the detection and classification of transient*, in *IEEE Instrumentation and Measurement Technology Conference*, Minnesota, USA (IEEE, 1998).
29. S. George *et al.*, *Speaker Recognition using Dynamic Synapse Based Neural Networks with Wavelet Preprocessing* (IEEE, 2001), pp. 1122–1125.
30. F. Phan, E. Micheli-Tzanakou and S. Sideman, *Speaker identification using neural networks and wavelets*, in *IEEE Engrg. Medicine Biology* **9** (2000) 92–101.
31. S. Santoso *et al.*, *Power quality disturbance waveform recognition using wavelet-based neural classifier — Part 1: Theoretical Foundation*, in *IEEE Trans. Power Delivery* **15**(1) (2000) 222–228.
32. H. H. Szu, B. Telfer and S. Kadambe, *Neural network adaptive wavelets for signal representation and classification*, in *Optical Engrg.* **31**(9) (1992) 2907–2916.
33. D. L. Donoho, *De-noising by soft-thresholding*, in *IEEE Trans. Infor. Theory* **41**(3) (1995) 613–627.
34. X. P. Zhang, *Thresholding neural network for adaptive noise reduction*, in *IEEE Trans. Neural Networks* **12**(3) (2001) 567–584.
35. L. Wang, K. K. Teo and Z. Lin, *Predicting time series with wavelet packet neural networks*, in *Int. Joint Conf. on Neural Networks*, 2001.
36. C. M. Huang and H. T. Yang, *Evolving wavelet-based networks for short-term load forecasting*, in *IEE Proc. on Gener. Transm. Distrib.* **148**(3) (2001) 222–228.
37. A. Prochazka and V. Sys, *Time series prediction using genetically trained wavelet networks*, in *Proc. of the 1994 IEEE Workshop, Neural Network for Signal Processing*, Ermioni, Greece (IEEE, 1994).

38. D. L. Donoho *et al.*, *Density estimation by wavelet thresholding*, *Ann. Statist.* **24** (1993) 508–539.
39. J. R. Cram, G. S. Kasman and J. Holtz, *Introduction to Surface Electromyography* (Aspen, 1998).