Review of Wavelet Transforms for Pattern Recognitions

Harold H. Szu

Center for Advanced Computer Studies

University of Southwestern Louisiana, Lafayette LA

Abstract

After relating the adaptive wavelet transform to the human visual and hearing systems, we exploit the synergism between such a smart sensor processing with brain-style neural network computing.

The freedom of choosing an appropriate kernel of a linear transform, which is given to us by the recent mathematical foundation of the wavelet transform, is exploited fully and is generally called the adaptive wavelet transform (WT). However, there are several levels of adaptivity: (i) Optimum Coefficients: adjustable transform coefficients chosen with respect to a fixed mother kernel for better invariant signal representation, (ii) Super-Mother: grouping different scales of daughter wavelets of same or different mother wavelets at different shift location into a new family called a superposition mother kernel for better speech signal classification, (iii) Variational Calculus to determine ab initio a constraint optimization mother for a specific task. The tradeoff between the mathematical rigor of the complete orthonormality and the speed of order (N) with the adaptive flexibility is finally up to the user's needs. Then, to illustrate (i), a new invariant optoelectronic architecture of a wedge-shape filter in the WT domain is given for scale-invariant signal classification by neural networks.

Keywords: Adaptive Wavelet Transform, Smart Sensors, Pattern Recognition, Neural Network,

1. Introduction

Wavelets have received significant attention as both a computational tool for analyzing data at multiple scales and as a model for early processing in mammalian visual and hearing systems. Progress in both of these areas makes wavelets attractive as a preprocessing method for artificial neural networks. Artificial neural networks (ANNs) also are an attractive method for adaptively tuning wavelets to best fit particular applications.

The question of why the wavelet transform (WT) was answered by two major points[1]. A real world application is often contaminated with noise. If such a noisy signal happens to have a wideband transient (WT) nature, then the signal-to-noise ratio can be enhanced by using a localized WT basis, but not by a Fourier basis. This is because the noise which is global everywhere will be picked up by the inner product with a Fourier transform (FT) which has also a global sinusoidal basis. On the other hand, the localized signal can only contribute to the inner product where the signal is nonzero and is matching well with a localized WT basis.

The other reason of using the WT is that often an interesting application is nonlinear in nature, and the mathematical freedom to choose the transform kernel

0-8194-2143-X/96/$6.00

other than the sinusoidal FT allows us to pay the nonlinear price up front. For example, an envelope soliton has been suggested by Szu[1] as the mother wavelet which is better suited for the nonlinear ocean wave (NOW) than the traditional FT followed by a mode-mode nonlinear coupling. These nonlinear phenomena occasionally possess exact solutions in somewhat idealized special cases, which can be nevertheless exploited as the zeroth order solution. And then the departure from the nonideal case may be accommodated by the linear WT expansion coefficients.

Before we present why adaptive WT[14], we wish to review why the Complete Orthonormal (CON) Discrete WT, with its order O(N), is faster by Log N than FFT, which is (by using the harmonic redundancy $\cos(120^o) = -\cos(60^o)$) of order O(N Log N). The reason is the simple mathematical truth , 0 x anything= 0, which is due to the localized WT basis that gives rise to more and more zero value as the WT goes to higher and higher resolution, which requires less and less multiplication and addition operations. According to Strang, this is indeed the "holy grail" of the complexity theory in terms of any matrix -vector linear transform, i.e. from the standard full matrix-vector multiplication $O(N^2)$, reduced to O(N Log N) by the FFT butterfly symmetry, which is furthermore reduced to O(N) by the systematically sparse matrices in the multiresolution pyramid. Then, the tradeoff of the computational speed must be finally compared with the computational cost of about O(N Log N) of an Adaptive WT and the advantage of Adaptive WT.

2. Introduction to Wavelets

We begin with a brief comparison between the WT and FT as follows. The WT generalizes the 1-D Fourier basis, $e_f(t)=\exp(2\pi ift)$ and harmonics, to a wideband transient 2-D basis, generated by an affine group of scale and shift operations,

$$h_{ab}(t) \equiv \quad h((t-b)/a) \,/\, a. \tag{1}$$

Usually, digital implementation of the WT uses a discrete basis, $a/a_o=2^{\pm I}$ and $b/b_o=\pm I$ (integer I) for a constant resolution for different scales. In Eq(1), the normalization constant 1/a is often preferred.[9] (cf.Eq(31) of Ref.[10] for the conventional wavelet normalization of inverse square-root a.) The admissible condition of a basis is that a square-integrable kernel h(t) must have zero dc component and a sufficiently fast decay at high frequencies[10]. The WT is defined in the square integrable Hilbert space:

$$W T_{ab}\{s(t)\} \equiv (h_{ab}(t), s(t)) = \int_{-\infty}^{+\infty} h^*_{ab}(t)\, s(t)\, dt \equiv W(a,b). \tag{2}$$

For the discrete version of the WT, a simple example is given as follows:

Discrete Haar Wavelet Matrix Expansion

Given a piecewise-constant approximation of a staircase function f(x)=(2,4,6,8) in 4 quarter steps, one can compute by the inner product of f(x) with 4 Haar wavelet bases to obtain the following:

$$f(x) = 5\,\Phi(x) - 2W(x) - W(2x) - W(2x-1)$$

which corresponds to the following column vector equation:

$$\begin{matrix} 2 \\ 4 \\ 6 \\ 8 \end{matrix} = 5 \begin{matrix} + \\ + \\ + \\ + \end{matrix} -2 \begin{matrix} + \\ + \\ - \\ - \end{matrix} - \begin{matrix} + \\ + \\ 0 \\ 0 \end{matrix} - \begin{matrix} 0 \\ 0 \\ + \\ - \end{matrix}$$

Note that such a local base is sparse, meaning zero entries!

$$y = \begin{matrix} 2 \\ 4 \\ 6 \\ 8 \end{matrix} = \begin{matrix} + & + & + & 0 \\ + & + & - & 0 \\ + & - & 0 & + \\ + & - & 0 & - \end{matrix} \begin{matrix} 5 \\ -2 \\ -1 \\ -1 \end{matrix} = [W_4]b$$

which defines a real matrix$[W_4]$ consisting of 4 orthogonal bases, which is easily verified. Consequently, the inverse DWT is straightforward, just the transpose denoted by the superscript T:

$$b = [W_4]^{-1} y = [W_4]^T y,$$

This simple inverse is a major reason for choosing CON DWT. Now we will define the family tree of Haar wavelets as shown in Figure 1.

From the Haar family tree follows dilation and wavelet equations:

$$\phi(x) = \phi(2x) + \phi(2x-1)$$

$$W(x) = \phi(2x) - \phi(2x-1)$$

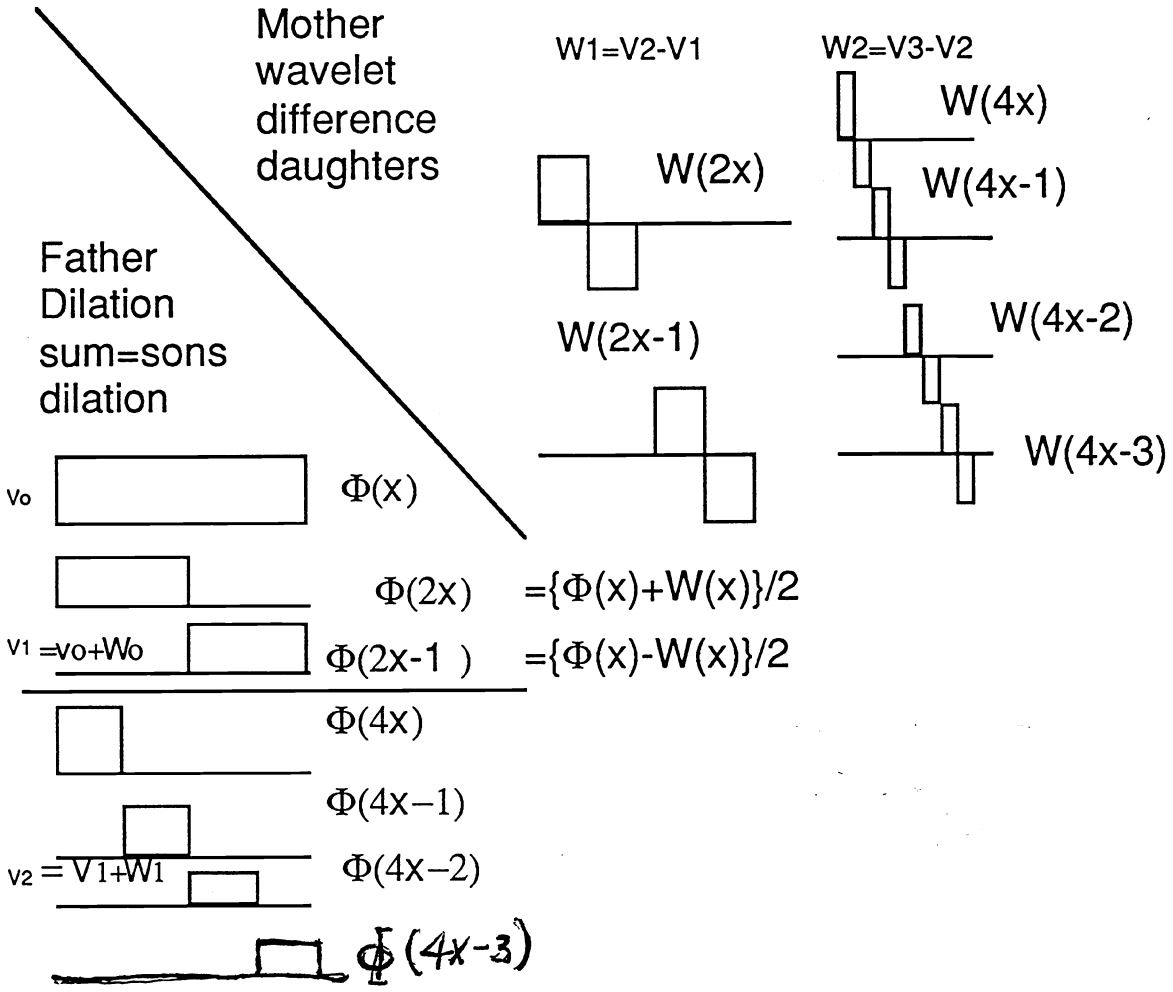

Figure 1: Family tree of Haar wavelets.

for which Daubechies succeeded in generalizing the piecewise constant Haar wavelet to piecewise linear function, which gives better functional approximation than constant Haar.

•Why CON DWT is computational O(N) rather than CON FFT O(N Log N) ?

The computational saving is due to the zero entries that require no multiplication with the local bases forming sparse column vectors. The following pyramid algorithm: $O(N^2)$->O(N), developed independently by Burt-Adelson-Mallat, show a local base of sum & difference giving sparse matrix $N=2^j$

$$j=1\ [W2] = \begin{matrix} ++ \\ +- \end{matrix} \qquad j=2\ [W4] = \begin{matrix} + & + & + & 0 \\ + & + & - & 0 \\ + & - & 0 & + \\ + & - & 0 & - \end{matrix}$$

The Holy Grail of Complexity Theorem alluded to earlier early is the O(N) reduction from $O(N^2)$:

$$[W4] = \begin{matrix} [W2] & & +000 & [W2] & \\ & & 00+0 & & \\ & & 0+00 & & \\ & [W2] & 000+ & & [I2] \end{matrix}$$

•N=4, N-1 =3 smaller matrix [W2]
•Non-zero =Nx(j+1) = $Nx(Log_2N + 1)$

• Major Contribution of Multiresolution (MR)

Multi-Resolution Paradigm					Mallat			
levels=j	Proj.=P_j Scale=Φ_j				Proj.=Q_j Wavelet=W_j			
j=0		sum=s (=son)			difference=d (=daughter)			
j=1	s	s	d	s	s	d	d	d
j=2	sss	dss	sds	dds	ssd	dsd	sdd	ddd

FT: Early generations, low freq. parents, win over children, high freq., because of Gibb's overshooting

W T: M.R. overcomes the FT dominant parenthood, puts all on an equal footing.

• Companion between Discrete WT Vs. Continuous WT

- Discrete = Fast WT if C.O.N. O(N)
  - •Inner Product

$$y = \Sigma_{jk} \langle f(x) \mid W(2^j - k) \rangle W(2^j - k)$$

$$= \Sigma_{jk} b_{jk} W_{jk}$$

  - •Easy Inverse WT, if C.O.N.
- •Continuous WT
- •Completeness = Inverse WT
- •FFT speedups by O(N Log_2 N).

3. Evidence for Wavelets in the Brain

Evidence for wavelets has been found in both the visual and hearing systems. As described elsewhere this handbook (Daugman), Gabor wavelets have been shown to be excellent models for the 2-D receptive fields in simple cells in the cat visual cortex.

In the hearing system, sound waves striking the eardrum cause vibrations that pass through the middle ear to the fluid-filled spiral-shaped cochlea. The pressure waves in cause vibrations in the basilar membrane of the cochlea, which are transduced to electrical signals by bending cilia that line the cochlea. The basilar membrane can be viewed as a bank of bandpass filters, with the filter center frequencies increasing with increasing distance into the cochlea. For human hearing above 800 Hz, the filter frequency responses are dilated versions of each other, meaning they are wavelets[22]! See Figure 2. Other ways to say the same thing are that the filters are constant-Q, (the Q factor is center frequency divided by bandwidth), or that the frequency responses appear nearly identical translated versions along a logarithmic frequency axis. This interpretation ignores nonlinear phenomena that play important roles at low sound levels, but are less important for normal speech volumes[22].

4. Combining Adaptive Wavelets with Neural Networks

There are several levels of adaptivity:

(i) Optimum Coefficients: adjustable transform coefficients chosen with respect to a fixed mother kernel for better representation,

(ii) Super-Mother: grouping different scales of daughter wavelets of same or different mother wavelets at different shift locations into a new family called a

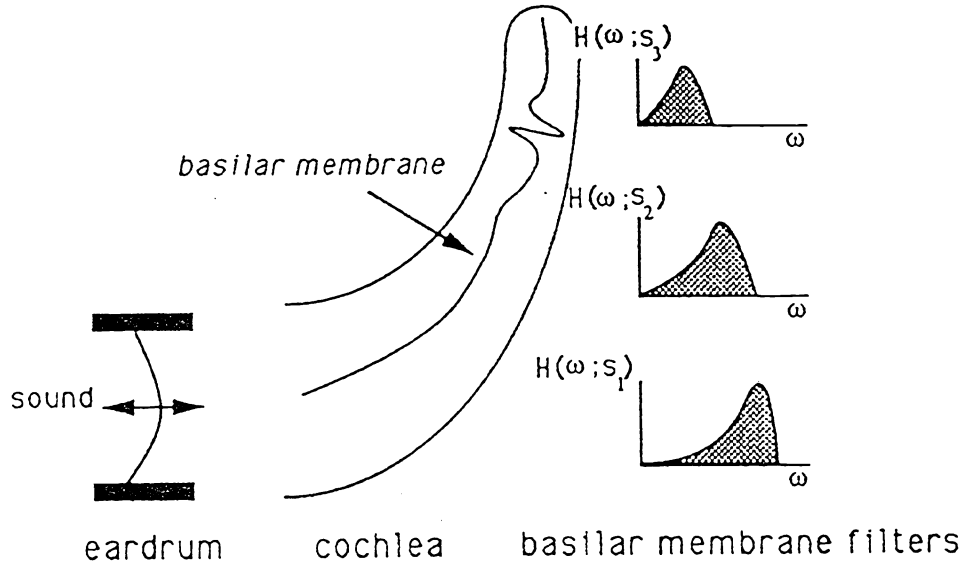

Figure 2: Interpretation of cochlear basilar membrane response as bank of wavelet filters[22].

superposition mother kernel for better classification,

(iii) Variational Calculus to determine ab initio a constraint optimization mother for a specific task. The tradeoff between the mathematical rigor of the complete orthonormality and the speed of order (N) with the adaptive flexibility is finally up to the user's needs.

In terms of functionality, representation versus classification is an important issue. A signal may be compressed for its faithful representation or feature extraction for the best discrimination capability against other signal's features. The former works on the probability distribution of the majority data, while the latter is the decision boundary with the overlapping minority data.

In this review, we emphasize neural network adaptive wavelets for signal classification rather than representation for the following reasons. Adaptively computing wavelet parameters (coefficients, dilation and shifts) with neural networks for the purpose of signal representation is not competitive in speed with the fast wavelet transform on serial hardware. The fast wavelet transform can be implemented in O(N), and methods exist for quickly deciding which wavelet coefficients should be retained for the best compression. Implementing the neural network algorithms on available high-speed parallel neural chips will greatly increase their speed, but chips are also available for computing the fast wavelet transform. For classification, on the other hand, the fast wavelet transform can compute coefficients quickly, but it is not a trivial process to determine the best coefficients to retain to best separate the training data into classes, rather than represent it.

It is not so important that the wavelet features be orthogonal, as it is that they separate the classes of training data. Since classifiers normally require relatively lengthy off-line training, adapting wavelet features together with the classifier during training is an attractive approach to minimize the misclassification rate.

For both neural and non-neural approaches, representation has received a large amount of attention while classification has been rarely considered. Daugman uses a neural network to learn the best set of coefficients for approximating an image with a set of Gabor wavelets[16]. A similar aproach has also been taken with a constructive algorithm that allocates additional wavelet functions in areas as needed[17]. Neural networks have also been used to adaptively compute wavelet shift and dilation parameters in addition to the coefficients[18]. All of these approaches employed nonorthogonal wavelets. An attractive alternative of using discrete orthogonal wavelets has also been advanced[19]. In this method, neural network weights can be computed very quickly by exploiting the wavelets' orthogonality, and making gradient descent unnecessary. Although not proposed

in terms of neural networks, an efficient method has been proposed for computing the best orthogonal wavelet from a scaling function for signal representation[20]. Another interesting alternative is to adatively compute wavelet waveforms in the short-time Fourier time-frequency domain of a particular signal[21]. This avoids some of the local minima problems that may arise in the time domain if the wavelets are initialized poorly. However, extension to higher dimensions may be more difficult, and computing wavelets from a Fourier-based time-frequency space has drawbacks[17].

For representation, a signal s(t) can be approximated by daughters of a mother wavelet h(t) according to

$$s(t) = \Sigma_k \, w_k \, h[(t-b_k)/a_k)],$$

where the w_k, b_k and a_k are the weight coefficients, shifts, and dilations for each daughter wavelet. This approximation can be expressed by the neural network of Figure 3, which contains wavelet nonlinearities in the artificial neurons. This architecture is similar to a radial basis function (RBF) neural network because symmetric wavelets form a family of RBFs. The network parameters can be optimized by minimizing an objective function, such as mean squared error. To demonstrate how adaptive wavelets can approximate functions, we consider three phonemes, "a," "e," and "i" (long vowels spoken in isolation) shown in Figure 4. Figure 5 shows a single period extracted from each together with wavelet approximations, showing good results with a small number of wavelets. The resulting approximations can then be treated as super-wavelets, in that their dilations can be used to approximate other voiced sounds with the same periodic shape but different frequencies [adaptivity level (ii)]. More details are provided elsewhere[8].

For representation, neural network learning does not appear to be competitive with the fast wavelet transform. However, neural networks are well suited for learning wavelet features in combination with a classifier. As an example, we consider a combined classifier and wavelet feature detector given by

$$v_n = \sigma(u_n) = \sigma\{\Sigma_k \, w_k \, \Sigma_t \, s_n(t) \; h[(t-b_k)/ak]\},$$

where v_n is the output for the n-th training vector $s_n(t)$ and $\sigma(z) = 1/[1+\exp(-z)]$, a sigmoidal function. This classifier can be depicted as the neural network of Figure 6, which uses wavelet weights rather than the wavelet nonlinearities of the representation network of Figure 3. The lower part of Figure 6 produces inner

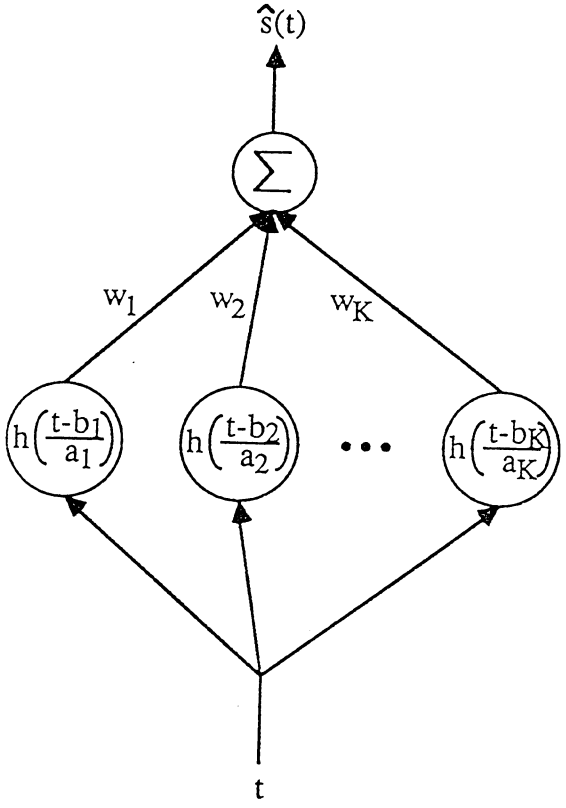

Figure 3: Example neural network architecture for wavelet signal approximation, where the time value t feeds into the K nodes with wavelet nonlinearities.

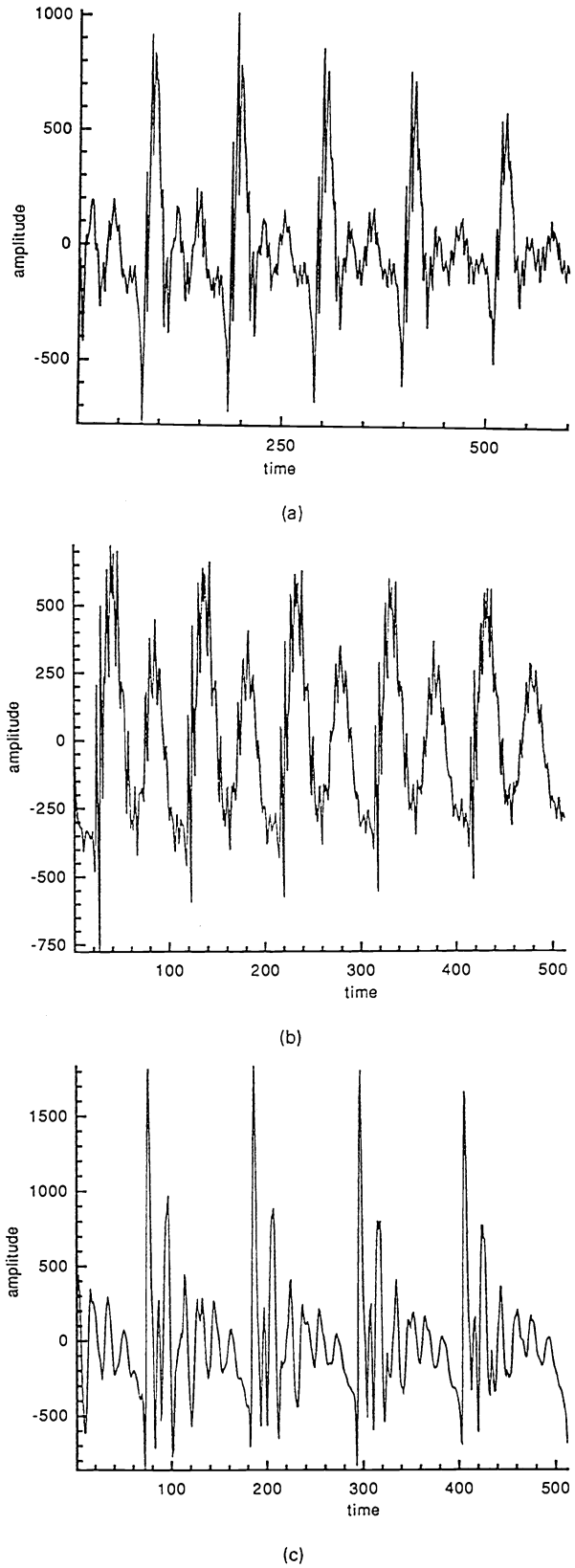

Figure 4: Phonemes: (a) "a," (b) "e," (c) "i."

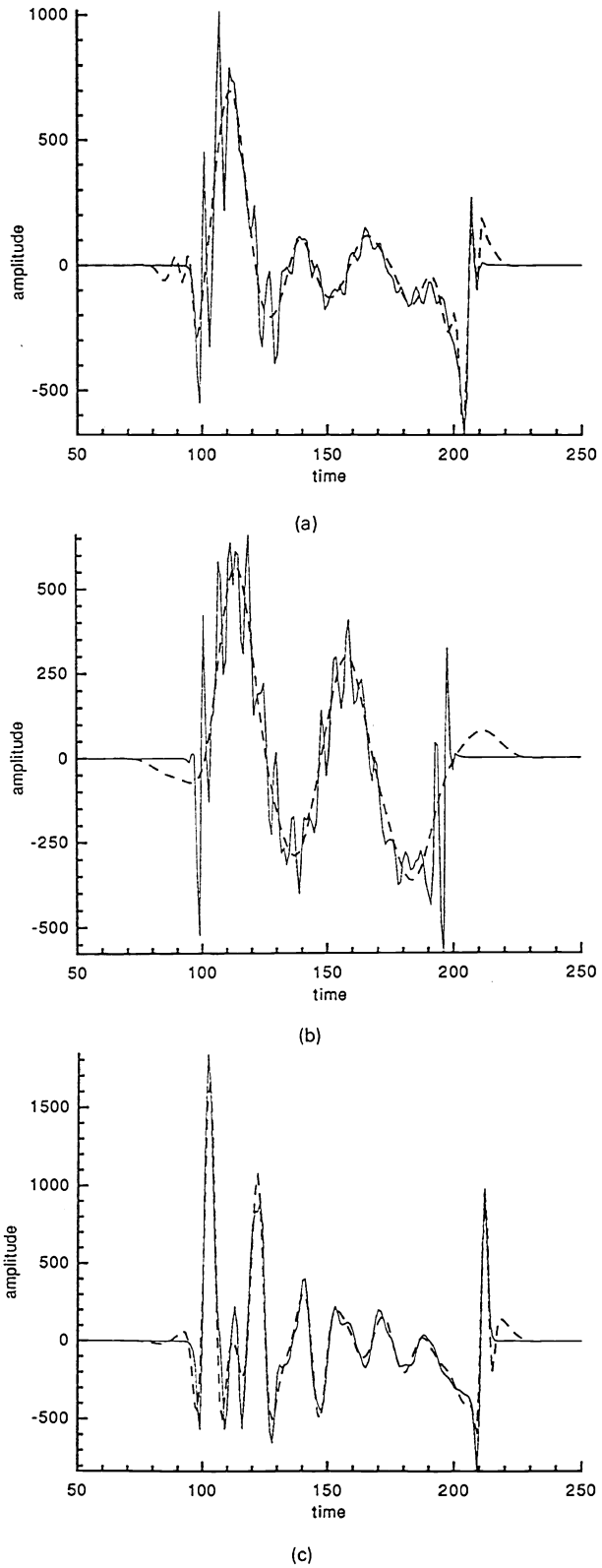

Figure 5: Single periods of phonemes extracted from Fig. 4 signals (solid lines) and adaptive wavelet approximations (dashed lines): (a) "a," (b) "e," (c) "i."

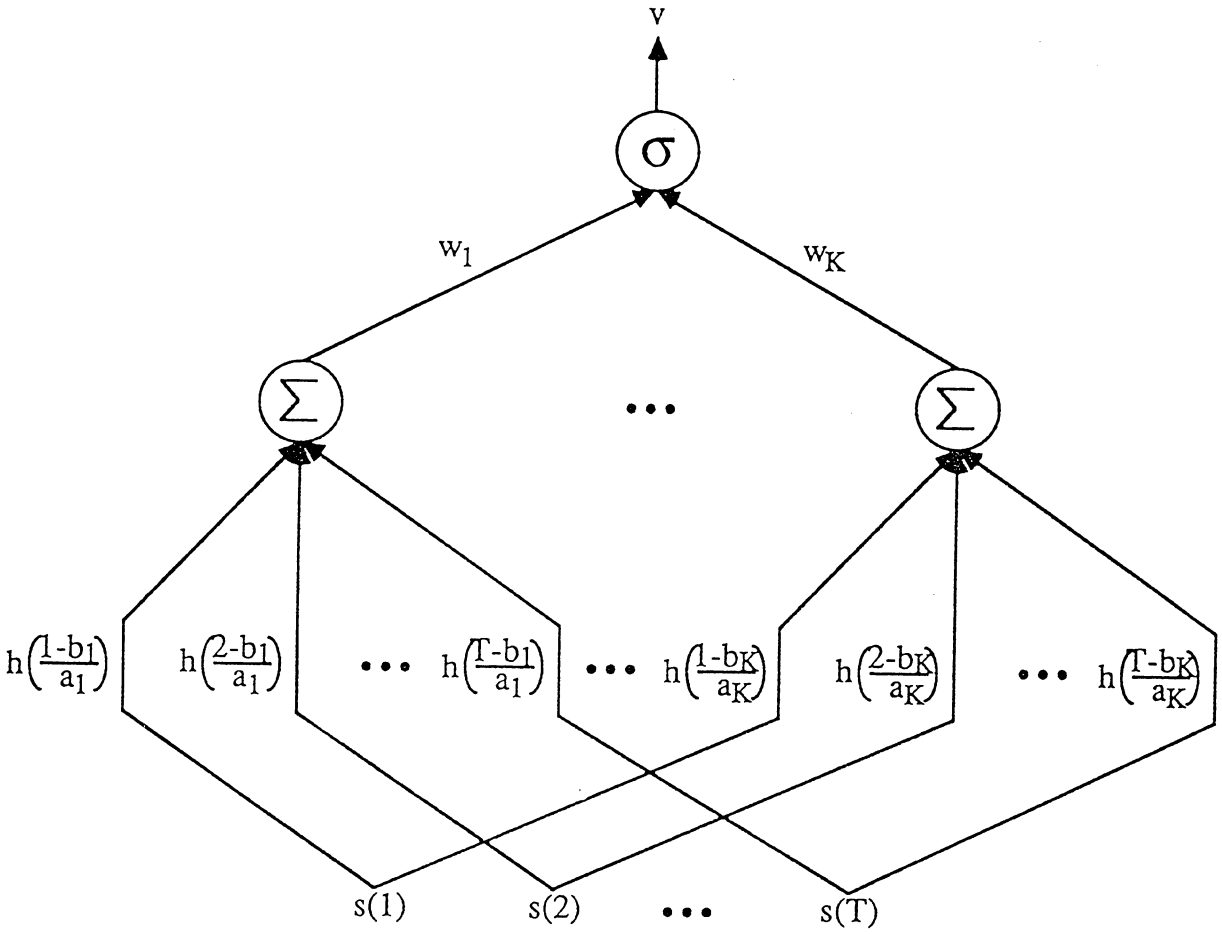

Figure 6: Example neural network architecture for classifier with wavelet features (after synthesis all weights compress to single layer because of nonlinearities).

products of the signal and wavelets, with the first wavelet on the left and the K-th wavelet on the right. Figure 6 shows two layers of weights, but for this simple example architecture, once the network is synthesized, the two layers collapse into one because a nonlinearity does not exist between layers. A simple classification test case was generated from the speech signals in Figure 4. When only allowing the weights to change during training and keeping the wavelet features fixed to their initial values, 13% error resulted, while allowing the wavelet shifts and dilations to adapt as well as the weights reulted in only 2.5%. This demonstrates the value of adapting the parameters of wavelet features as well as the classifier weights. More details are provided elsewhere[8].

5. Invariant WT

As an example of the first methodology of adaptivity, we consider an interesting application of invariant WT. There are many interests in applying neural network technology to experimental measurements as an automation tool. Similarly, the usefulness of the wavelet transform (WT) to replace the Fourier transform (FT) has been demonstrated in data compression as well. This chapter contributes toward the synergism of both technologies. One of the challenges in experimental diagnoses is to achieve distortion invariant classification. Traditional approaches are based on time-frequency joint representations (TFJR) such as the E. Wigner distribution based on the quantum mechanical uncertainty principle and the P. Woodward ambiguity function based the Doppler shift uncertainty. Unfortunately, both involve a second order convolution and correlation integral of quadratic order, such that multiple pulses produce a double amount of spectrum pulses that complicate the identification task for both TFJRs. Recently, a first-order time-scale joint representation (TSJR), called the Wavelet Transform[2-4], was developed to replace the traditional FT by computing the TSJR for noisy wideband transients. Fractal aggregate[5] and turbulence data[6,7] have been analyzed by WT giving efficient interpretations. Motivated by multimedium satellite communication (Broadband-ISDN), as well as the need for lossless fingerprint compression, to be used in the National Crime Information Center, the wavelet transform is being widely adopted for signal and image processing.
A WT-reduced set of data may be called a wavelet feature set, which can be used to train and implement electronically a smaller size of the ANN for pattern classification. We begin by comparing the FT and WT.

The FT is known to be an angle-preserving, or conformal mapping, which has suggested the design of wedge-ring detectors in the Fourier plane[11]. This device works for scale-rotation invariance based on shift invariance because, despite of the motion of the object, the square-law detectors guarantee the object centroid alignment at the origin of the Fourier domain. This shift invariance is based on the

modulus-invariant phase information in the following straightforward mathematics:

$$|FT_f\{g(t)\}|^2=|(e_f(t),g(t))|^2=|G(f)|^2=|G(f)\exp(-2\pi jfb)|^2=|\int_{-\infty}^{+\infty}\exp(-2\pi jft)g(t-b)dt|^2 \quad (3)$$

On the other hand, the invariant property of the WT is based on the intrinsic scale property in the TSJR domain (a,b). The idea is simple. To investigate the invariant WT is to compute WT of various scales of the identical signal. Hopefully, those scale-related WT coefficients organize themselves in such a fashion that can be easily collected to produce the scale invariant features. Given a generic signal under additive white noise,

$$s'_i(t) = s_i(\alpha_i t) + n(t); \qquad i=1,2,.., \quad (4)$$

where the unknown scales α's (suppressing class index i) are equivalent to unknown frequency compaction or hopping of similar waveforms s(t)'s. The associated WT coefficient denoted by the prime is computed

$$W'(a,b) \equiv \int_{-\infty}^{+\infty} dt\, s'(t)h^*((t-b)/a)/a = \int_{-\infty}^{+\infty} dt'\, s'(t')h^*((t'-b')/a')/a' \quad (5)$$

Use is made of the change of variables: $t'=\alpha t$, $a'=\alpha a$, and $b'=\alpha b$, and Eq(5) becomes exactly equal to that original W(a,b) Eq(2) located radially by a factor of α in both a and b plane.

$$W'(a,b) = W(\alpha a,\alpha b) + \text{noise.} \quad (6)$$

Example:

If $\alpha=2$, then signal s(2x) is shrunk by half (where for example the peak value of $s(x_o)$ at x_o is shrunk to $x=x_o/2$ location). Therefore, the WT coefficient $W'(a,b)=W(2a,2b)$ which is shifted toward the origin by a factor of 2, so that for a compacted size signal the WT locates toward the center by the factor 2 along both the a and b axes.

This observation makes it clear the geometric meaning of Eq(6) and our design of a wedge detector on the TSJR plane follows. The wedge filter bank consists of N radial wedges that have N equally spaced angles in the upper half plane where $a\geq 0$.

$$b_n \equiv a\tan(n\theta_o) \quad (7)$$

$$b_{n+1} \equiv a\tan((n+1)\theta_o), \quad (8)$$

where $\theta_o \equiv 2\pi/N$ for a total of N wedges. The total output value collected through the wedge-shaped filter denoted by w_n' for s'(t) is integrated

$$w_n' \equiv \int_0^{\infty} da \sum_{bn}^{bn+1} W'(a,b) = w_n/\alpha \quad (9)$$

to give its value

$1/\alpha$ multiplying with that w_n for s(t). Since this formula (9) is true for every wedge extracted value, a simple normalization of all values gives us a set of scale-invariant feature values. This fact is obvious, as often is the case in hindsight, in that the size information of the object is mapped to the value of some area integral and then a simple threshold can ignore the size information.

The present technique based on wavelets and neurons seems to be an ideal tool to solve real time physics application problems that require scale-invariance. For example, wavelet transforms have been applied to the onset of turbulence flow by Frisch and Orzag[7], and the Kolmogorov turbulence cascade by Meneveau[6] who have demonstrated the self-similar nature of the turbulent flow. Thus, the present technique might be used to enhance the turbulence onset signal giving better diagnosis in real fluids, because it collects scale invariant information at multiple resolution scales. A similar application by Freysz et.al.[5] that can be extended by invariant wavelet technique was to investigate the scale-invariant fractal aggregation in diffusion-limited cases. Another application of invariant wavelet transform is the bubble chamber particle track recognition problem with a large throughput rate in the superconductor supercollider experiments. The challenge there is perhaps similar to the star track minutiae associated with the task of compressing FBI finger print files by a lossless wavelet transform which is approved by FBI to be better than the traditional discrete cosine transform used in the high definition TV and data compression. However, the particle tracking is not stationary and is further compounded by the time-dependent event under electromagnetic field fluctuations, for which a scale-invariant wavelet transform automation may prove to be useful for minor changes in the measurement setup. In the interest of simplicity to bring forward the use of this novel synergism between wavelets and neurons we consider a one dimensional example. An important measurement problem in astrophysics is detecting a weak signal under an unknown Doppler shift. This type of frequency modulation by environmental perturbations is modeled as follows.

As a generic illustration, we consider two classes of pulse signals that have arbitrary carrier frequency and different frequency modulations and hopping. One class has high, low and medium frequency modulated pulses, denoted as HLM, and the other class has three modulated pulses of equal frequency, denoted as MMM for a medium carrier frequency. A natural wavelet to analyze the change of pulses is the bipolar Haar wavelet h(t) ={-1, +1}[12]. The HLM and MMM signal templates are shown inRef. 13. The first and third HLM segments in Fig. 1a[13] have an amplitude of 2 while the middle segment has 1.75 amplitude, and altogether there are 512 data points. The MMM signal template has one medium frequency in three pulses. Training and test vectors with differing scale and noise are normalized so each feature has a zero mean and 0.75 standard deviation. The training set consisted of

$s_1(t)$ and $s_2(t)$ (subscripts denote the two classes), while the test set consisted of $s_1(2t)$, $s_2(2t)$, $s_1(2t)+n(t)$ and $s_2(2t)+n(t)$, where n(t) is 20 dB noise. Ref. 13 shows the two noisy test signals, and simulated wedge outputs for the original HLM signal $s_1(t)$ and scaled and shifted versions. The wedges of the scaled signal are essentially identical to the original. The wedges of the shifted signal show graceful degradation for shifts as well. To capture shift invariance, one can likewise design a horizontal bar filter bank (space does not permit us to discuss this further).

This invariant feature set can be furthermore fed into a two-layer feedforward ANN for both the interpolation of non-integer scale value and subsequent classification as demonstrated in the righthand side of Figure 7. Each signal is wavelet transformed by the Haar wavelet, and collected through 32 wedge detectors and fed into 32 input neurons of the input layer of ANN. An MxN two-layer feedforward ANN has M=1 single output layer neuron for two classes (target or clutter) and N=32 neurons at the input layer, one for each wedge detector output. The ANN learning is described as follows.

A performance or energy function E(v) is a function of the output neuron activity function v. The usual choice is the square difference of desired output d (during the training, say 1 for class 1 and zero for class 2) and the actual output v.

$$E(v) = (v - d)^2/2 \tag{10}$$

The first layer has N=32 neurons whose input u_j and output v_j are denoted by the subscript j (note that the i-index of second layer neuron is suppressed for a single neuron at the second layer considered here and thus the traditional double indices of the interconnect weight w_{ij} is also reduced to one index w_j for a single neuron at the second layer). The net input to the second layer neuron is a w_j weighted sum from all output from the first layer:

$$u = \Sigma^{32}_{j=1} w_j v_j - q \tag{11}$$

And the q-threshold logic is defined by the logistic function:

$$v = 1/[1 + \exp(-u)]; \; (dv/du) = v(1-v) \geq 0 \tag{12a,b}$$

which has a nonnegative logic with more input u implying more output v. The learning of the interconnect weight value is achieved by a local gradient descent.

$$\partial w_j/\partial t = -\partial E/\partial w_j \tag{13}$$

Then, the standard gradient descent algorithm is used to determine the weights

$$\begin{aligned} \Delta w_j \equiv (\partial w_j/\partial t)\Delta t &= -(\partial E/\partial v)\,(dv/du)\,(\partial u/\partial w_j)\,\Delta t \\ &= (d - v)\; v(1-v)\; v_j\, \tau \end{aligned} \tag{14}$$

where use is made of the gradient descent to compute the chain rule of

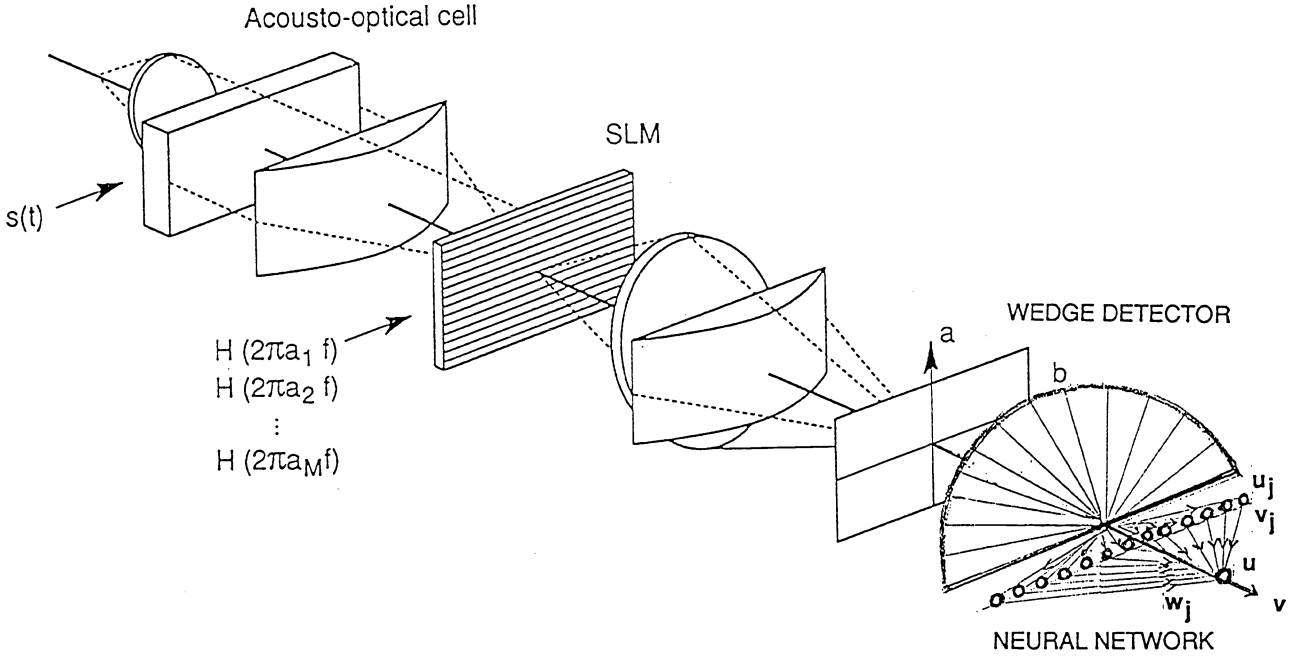

Figure 7: Two dimensional optical correlator with cylindrical FT lenses and a bank of the wavelet filters (located at SLM) for the WT of a one-dimensional signal (Ref[10]) followed by the wedge detector, in the time-b and scale-a domain, and a two-layer feedforward neural network for the noisy and distorted signal classification.

differentiations from Eqs(10-12), and $\Delta t = \tau$. Given signal templates $s_i(t)$, i=1,2,for the training by Eq(14), we have achieved perfect classification of two noisy contracted signals Eq(4) with unknown compaction scale a_i. The network outputs in Table 1 can be seen to give the desired results:

Training		Testing without noise		Testing with noise	
$s_1(t)$=HLM	0	$s_1(2t)$	$5X10^{-6}$	$s_1(2t)+n(t)$	0.023
$s_2(t)$=MMM	1	$s_2(2t)$	0.999	$s_2(2t)+n(t)$	0.990

Table 1: Network outputs for training and test set.

An optoelectronic architecture for optical implementation[10] of the WT preprocessing and the neural network classification is given in Ref.13. Input data is represented by an acousto-optical transducer (with high GHz speed). Daughter wavelets are encoded in the form of a film mask, or holographic matched filters called the spatial light modulator (SLM). The output is a two dimensional plane. The horizontal direction corresponds to the shift b (continuously distributed), whereas the vertical direction corresponds to the dilation factor a (discrete sampled). An integer discrete scale is easy for optical wedge detector layout and the noninteger scale is taken care of by the fault tolerance of perceptron neural network interpolation. Furthermore, we assume that a small angular separation $\theta_o \equiv 2\pi/N$ between the width between $b_n \equiv a \tan(n\theta_o)$ and $b_{n+1} \equiv a \tan((n+1)\theta_o)$ is approximated by a constant width, a θ_o, for a large N detector bank, as depicted in Ref.13. This is easier for the layout of the optoelectronics device so that each row of detectors has an equal spacing proportional to the location a of the row. A more expensive monolithic design of the wedge detectors (where each wedge is a single large detector) can also be used but the integration time will be longer to slow down the real-time operation. Then, a simple neural network is used to classify the yes and no decision about the detection of a specific event under arbitrary distortion and noise in the domain of time-scale joint representation.

6. Conclusion

In this review, we have shown that wavelets are attractive both for their computational properties (e.g., fast O(N) wavelet transform) and for their usage in biological preprocessing. This has led us and others to employ ANNs to give wavelets varying degrees of adaptivity, with the most adaptivity coming from synthesizing wavelets best suited to particular applications [adaptivity levels (ii) and (iii)]. This ANN approach is better suited to determining classification features than to representation.

As an example of scale-invariance, we have also demonstrated that adaptivity methodology (i) can be used with a wedge-shape combination of wavelet

coefficients to lead to a scale-invariant WT using ANN automation.

Acknowledgement: The support of the NSWCDD Independent Research Fund is gratefully appreciated.

References

1. H. Szu, "Why the soliton wavelet transform is useful for nonlinear dynamic Phenomena," SPIE Proceeding Vol. 1705 (Visual Info Proc. by Huck & Juday), pp. 280-288, Orlando 22 April 1992

H. Szu, "Two dimensional optical processing of one dimensional acoustic data," Opt. Eng. 21, 804-813, 1982.

2. Special Issue of Opt. Eng. on Wavelet Transform, Ed. by Szu & Caulfield,September 1992

3. Special Issue of IEEE Trans. Inf. Theory on Wavelet Transform, Ed. by Daubechies & Mallat, March 1992.

4. Special Issue of IEEE Trans. Sig. Proc. on Wavelet Transform, Ed.by Duhamel, Flandrin, Nishitani,Tewfik, Vetterli, Summer 1993.

5. E. Freysz, B. Pouligny, F. Argoul, A. Arneodo, Optical Wavelet Transform of Fractal Aggregates," Phys. Rev. Lett. 64, p. 745, 1990.

6. C. Meneveau, "Dual Spectra and mixed energy cascade of turbulence in the wavelet representation," Phys. Rev. Lett., 66, p.1450, 1991.

7. U. Frisch, S. A. Orzag, "Turbulence: Challenges for Theory and Experiment," Physics Today, pp. 24-32, Jan. 1990.

8. H. Szu, B. Telfer, and S. Kadambe, "Neural Network Adaptive Wavelets for Signal Representation and Classification," Optical Engineering, Vol. 31, pp. 1907-1916, Sept. 1993.

9. B. Telfer & H. Szu, "New Wavelet Transform Normalization to Remove Frequency Bias," Optical Engineering, vol. 31, pp. 1830-1834, Sept. 1993.

10. H. Szu, Y. Sheng, J. Chen, "Wavelet Transform as a bank of matched filters," Appl. Opt. 31, pp. 3267-3277, Jun. 1992.

11. N. George, A. L. Kasdan, Proc. of Electr, Opt. Sys. Des. Conf. pp.494-503, 1975.

12. A. Haar,"Zur Theorie der orthogonalen Funktionen-systeme," Math. Annal. 69, pp. 331-371, 1910.

13. H. Szu, X-Y. Yang, B. Telfer, Y. Sheng, "Neural Network and Wavelet Transform for Scale-Invariant Data Classification, " to appear in Phys. Rev. E, 1993.

14. H. Szu, "Why Adaptive Wavelet Transform?" SPIE V.1961, Orlando 22 April 1993.

15. Y. Sheng, D. Roberge, H. Szu, T. Lu, "Optical wavelet matched filters for shift-invariant pattern recognition," Opt. Lett. Vol. 18, pp. 209-301, 15 Feb 1993.

16. J.G. Daugman, "Complete Discrete 2-D Gabor Transforms by Neural Networks for Image Analysis and Compression," IEEE Trans. ASSP, vol. 36, pp. 1169-1179, July 1988.

17. Y.C. Pati and P.S. Krishnaprasad, "Analysis and Synthesis of Feedforward

Neural Networks using Discrete Affine Wavelet Transforms," IEEE Trans. NN, vol. 4, pp. 73-85, Jan. 1993.
18. Q. Zhang and A. Benveniste, "Wavelet Networks," IEEE Trans. NN, vol. 3, pp. 889-898, Nov. 1992.
19. T.I Boubez and R.L. Peskin, "Wavelet Neural Networks and Receptive Field Partitioning," IEEE Inter. Conf. Neural Networks, pp. 1544-1549, Mar. 1993.
20. A. Tewfik, D. Singha, P. Jorgensen, "On the Optimal Choice of a Wavelet for Signal Representation," IEEE Trans. IT, vol. 38, pp. 719-746, Mar. 1992.
21. S. Mann and S. Haykin, "Adaptive 'Chirplet' Transform: an Adaptive Generalization of the Wavelet Transform," Optical Engineering, vol. 31, pp. 1243-1256, June 1992.
22. X. Yang, K. Wang, S.A. Shamma, "Auditory Representations of Acoustic Signals," IEEE Trans. IT, vol. 38, pp. 824-839, Mar. 1992.