LA-UR 95-1072

Conf-9504137--1

Los Alamos National Laboratory is operated by the University of California for the United States Department of Energy under contract W-7405-ENG-36

TITLE: **SPACE-BASED RF SIGNAL CLASSIFICATION USING ADAPTIVE WAVELET FEATURES**

AUTHOR(S): Michael Caffrey, NIS-3
Scott Briles, NIS-3

SUBMITTED TO: SPIE - Aero Sense Conference, April 17-21, 1995.

DISCLAIMER

Los Alamos

Los Alamos National Laboratory
Los Alamos, New Mexico 87545

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

Space-based RF signal classification using adaptive wavelet features

Michael Caffrey, Scott Briles

Los Alamos National Laboratory
NIS-3/D440
Los Alamos, NM 87545

ABSTRACT

RF signals are dispersed in frequency as they propagate through the ionosphere. For wide-band signals, this results in nonlinearly-chirped-frequency, transient signals in the VHF portion of the spectrum. This ionospheric dispersion provide a means of discriminating wide-band transients from other signals (e.g., continuous-wave carriers, burst communications, chirped-radar signals, etc.). The transient nature of these dispersed signals makes them candidates for wavelet feature selection. Rather than choosing a wavelet ad hoc, we adaptively compute an optimal mother wavelet via a neural network. Gaussian weighted, linear frequency modulate (GLFM) wavelets are linearly combined by the network to generate our application specific mother wavelet, which is optimized for its capacity to select features that discriminate between the dispersed signals and clutter (e.g., multiple continuous-wave carriers), not for its ability to represent the dispersed signal. The resulting mother wavelet is then used to extract features for a neural network classifier. The performance of the adaptive wavelet classifier is then compared to an FFT based neural network classifier.

Keywords: Wavelets; classification; feature selection; neural networks.

INTRODUCTION

The detection and classification of impulsive radio-frequency (RF) events occurring in the earth's atmosphere, such as signals from illegal nuclear tests, are crucial in the nonproliferation effort. These signals are dispersed in frequency due to the ionosphere when observed in orbit; the frequency dispersion varies widely in accordance with the state of the ionosphere. A single matched filter cannot match the large variations in the signal duration; however, a wavelet transform can be viewed as a bank of matched filters[1] and can take advantage of the a priori knowledge of the signal of interest.

Taking advantage of *a priori* knowledge of the signal requires that the basis function of the wavelet transform be designed for the signal of interest. Selection of a common basis function or mother wavelet is not as efficient as constructing a mother wavelet for the specific task. Previous work[2] has indicated that adaptive construction is a method for the determination of the proper mother wavelet. This paper presents the construction, using a neural network, of a mother wavelet for feature selection. This mother wavelet is then used to extract features for a neural network classifier. For comparison, an FFT based neural network classifier[3] will be used to contrast the generalization and robustness characteristics of the adaptive wavelet classifier.

SIGNAL MODEL

Signals encounter frequency-dependent dispersion as they move through the ionosphere.[4,5] The total electron content (TEC) of the ionosphere is the major factor in the dispersion profile of the received signal. The total electron content is the integration of the electron density along the path the signal travels through the atmosphere. TEC varies in time and observation location; using the symbol, N_e, for TEC of the ionosphere, the TEC can assume values in the range

$$1\times10^{16}\left(e^-/\mathrm{m}^2\right)\leq N_e\leq 5\times10^{18}\left(e^-/\mathrm{m}^2\right). \tag{1}$$

The time taken by an RF signal to traverse the ionosphere depends on the frequency of the signal. There is a transit time difference, Δt_d, between the transit times of two frequencies, f_o and $f_o-\Delta f$, where $f_o \geq f_o-\Delta f$ and $\Delta t_d \geq 0$. The relationship between the transit time difference and the two frequencies is

$$\Delta t_d = \frac{5.3\times10^{-6}N_e}{4\pi^2}\left(\frac{1}{(f_o-\Delta f)^2}-\frac{1}{f_o^2}\right). \tag{2}$$

Solving Equation 2 for Δf, the frequency difference from the upper frequency of f_o, yields

$$\Delta f = f_o\left(1-\sqrt{\frac{N_e'}{N_e'+f_o^2\,\Delta t}}\right), \tag{3}$$

where,

$$N_e' = \frac{5.3\times10^{-6}N_e}{4\pi^2}. \tag{4}$$

Integrating Equation 3, as the instantaneous frequencies of the time-domain signal, will yield the argument of a sinusoid. Thus the time-domain signal can be expressed as

$$s(t)=\mathrm{rect}\left(\frac{\mathrm{t}}{\mathrm{T}}-\frac{1}{2}\right)\cos\left(f_o t-2\left(\frac{\sqrt{N_e'}}{f_o}\right)\left(\sqrt{N_e'+f_o^2 t}\right)+\theta\right), \tag{5}$$

where the rect function is used to bound the signal in time between $0\leq t\leq T$. Equation 5 expresses the real part of the analytic form of the signal, and includes a fixed phase term, θ. However, in receiver systems a fixed bandwidth will constrain Equation 5. Thus the instantaneous frequency of Equation 5 must be limited to the receiver bandwidth, B, such that $\Delta f \leq B$. This implies that the upper limit of the receiver bandwidth is f_o. The signal of interest used in this paper is expressed in Equation 5 with an added bandwidth constraint.

METHODOLOGY

FINDING THE 'MOTHER' WAVELET

The objective is to use wavelets to extract features from a signal of interest. The signal of interest has an instantaneous frequency that is functionally quadratic. This prompted the use of the Gaussian weighted , linear frequency modulated, (GLFM) sinusoid as the fundamental wavelet. The intent is to stretch, compress, and shift a number of these GLFM wavelets, then add them into a mother wavelet that will be useful for feature extraction. The scale and shift are chosen adaptively with a neural network. The GLFM function, h(t), is

$$h(t) = \exp\left(-\alpha t^2\right)\cos\left(\omega_1 t + \omega_2 t^2\right) \tag{6}$$

where α is the damping factor of the Gaussian and ω_1 and ω_2 determine the slope of the instantaneous frequency. The wavelet is scaled and shifted by substituting t' for t where t' is defined as

$$t'_k = \left(\frac{t - b_k}{a_k}\right) \tag{7}$$

where b represents the shift and a the scale. Two examples of scaled and shifted GLFMs are shown in figure (1) below.

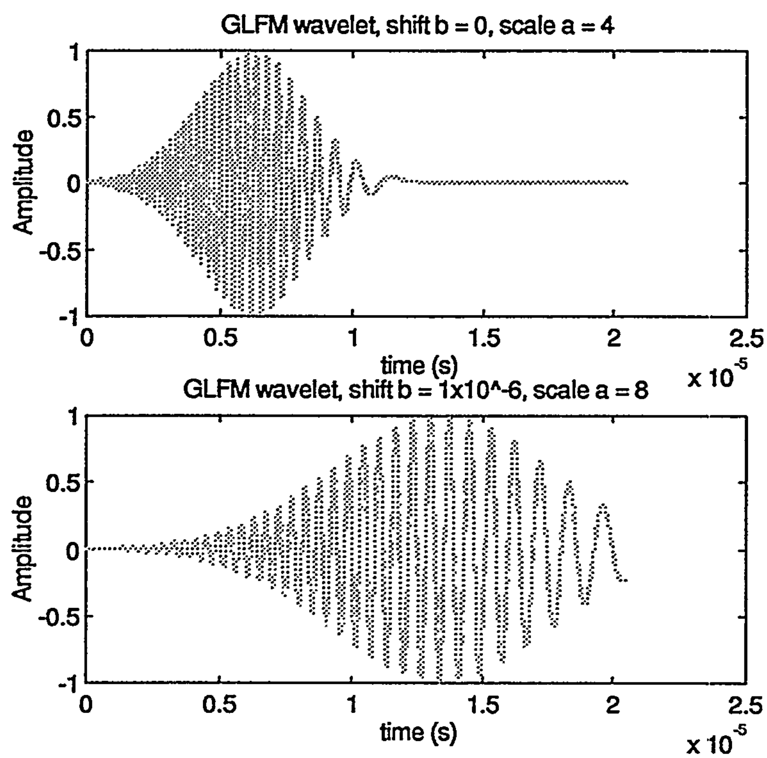


Figure 1: Two examples of the GLFM wavelet with different scale (a) and shift (b).

The architecture of the neural network used for choosing the weight (w), shift and scale parameters is shown below in figure (2).

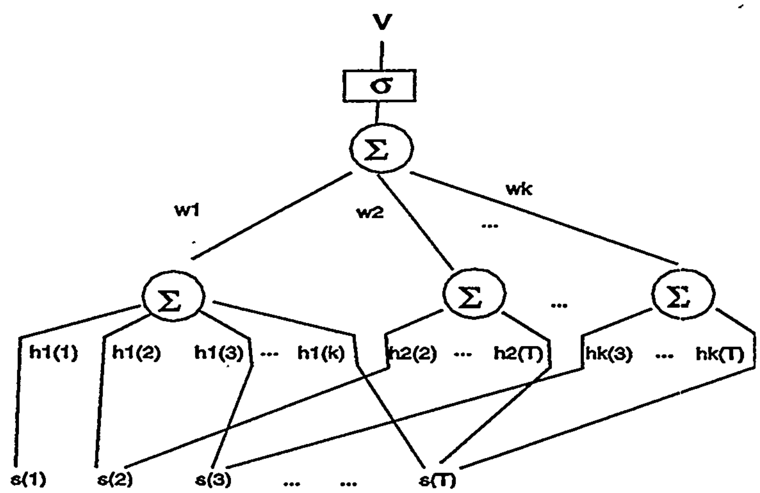


Figure 2: Neural network used for adaptively finding the mother wavelet.

The time domain signal is projected onto nine GLFM wavelets, each with a and b chosen adaptively. Each projection is weighted then summed. The sum is applied to the sigmoidal transfer function and the output determines the class membership of the signal of interest. The training set consists of two classes: transients with additive white Gaussian noise and added carriers with random frequencies (clutter), and noise with clutter but lacking the transient. The network minimizes the error function

$$E = \frac{1}{2}\sum_{n=1}^{N}(d_n - v_n)^2 \tag{8}$$

where v_n is the output of the network for pattern s_n and d_n is the desired network output for pattern s_n . N is the number of vectors in the training set. v_n can be expressed as

$$v_n = \sigma\left(\sum_{k=1}^{K} w_k \sum_{t=1}^{T} h_{kt} s_{nt}\right) \tag{9}$$

where K is the number of GLFM daughter wavelets, $h_k(t)$. T is the number of samples in the pattern s times the sample period. The summation over t is simply the vector inner product of the vectors h and s. The sigmoidal transfer function has the form

$$\sigma(x) = \left(\frac{1}{1+e^{-x}}\right). \tag{10}$$

This transfer function squashes the output, limiting it to the interval (0,1). The conjugate gradient descent method updates the scales (a), shifts (b), and weights(w). Using the functional form of h(t) from Equation (6), the GLFM wavelet is

$$h(t) = \exp(-\alpha t^2)\cos(\omega_1 t + \omega_2 t^2)\ . \tag{11}$$

Time, t, can be replaced by a scaled and shifted version t' such that

$$t'_k = \left(\frac{t - b_k}{a_k}\right). \tag{12}$$

Now, the gradients of the error function, E, can be written as

$$\nabla E_{w_k} = -\sum_{n=1}^{N}(d_n - v_n)\sigma'(x)\mathbf{h}_k \bullet \mathbf{s}_n \tag{13}$$

$$\nabla E_{a_k} = -\sum_{n=1}^{N}(d_n - v_n)\sigma'(x) w_k \mathbf{s}_n \bullet \frac{d\mathbf{h}_k}{da_k} \tag{14}$$

$$\nabla E_{b_k} = -\sum_{n=1}^{N}(d_n - v_n)\sigma'(x) w_k \mathbf{s}_n \bullet \frac{d\mathbf{h}_k}{db_k} \tag{15}$$

where the derivative of the sigmoidal function is

$$\sigma'(x) = \sigma(x)(1-\sigma(x)),$$
$$x = \left(\sum_{k=1}^{K} w_k \sum_{t=1}^{T} h_{kt} s_{nt}\right) \tag{16}$$

The partial derivative of h(t') with respect to shift b can be written as

$$\frac{d\mathbf{h}_k}{db_k} = \exp(-\alpha t'^2)\cos(\omega_1 t' + \omega_2 t'^2)\left(\frac{2\alpha t'}{a_k}\right) + \exp(-\alpha t'^2)\sin(\omega_1 t' + \omega_2 t'^2)\left(\frac{\omega_1 + 2\omega_2 t'}{a_k}\right). \tag{17}$$

Similarly, the partial derivative of h_k with respect to scale a can be written as

$$\begin{aligned}\frac{d\mathbf{h}_k}{da_k} &= \exp(-\alpha t'^2)\cos(\omega_1 t' + \omega_2 t'^2)\left(\frac{2\alpha t'^2}{a_k}\right) + \exp(-\alpha t'^2)\sin(\omega_1 t' + \omega_2 t'^2)\left(\frac{\omega_1 t' + 2\omega_2 t'^2}{a_k}\right). \\ &= \frac{d\mathbf{h}_k}{db_k} t'\end{aligned} \tag{18}$$

The search directions for updating the weights, shifts, and scales can be found with the Fletcher-Reeves formula[6]. The new search direction can be written

$$s_{w_k} = -\nabla E_{w_k} + \frac{\nabla E_{w_k}^T \nabla E_{w_k}}{\nabla E_{w_{k-1}}^T \nabla E_{w_{k-1}}} s_{w_{k-1}} \tag{19}$$

for the weights. The shift and scale update directions have similar expressions. The update to the weight vectors has the form

$$w_{k+1} = w_k + \mu_k s_{w_k}. \tag{20}$$

A line search[6] is used to find the optimal learning factor, μ, at each iteration. The result of this neural network configuration is a weight vector, w, and K scaled and shifted GLFM daughter wavelets. A new mother wavelet is formed by the superposition of these daughter wavelets as shown:

$$\mathbf{h}_{new}(t) = \sum_{k=1}^{K} w_k \mathbf{h}\left(\frac{t-b_k}{a_k}\right). \tag{21}$$

This is the mother wavelet that best extracts the features of the training data. The next step uses this wavelet to extract features for a two layer feed-forward neural network classifier.

NEURAL NETWORK CLASSIFIER BASED ON DAUGHTER WAVELETS

Using the mother wavelet found above, four daughter wavelets can be formed by scaling h_{new} by four octaves. Let $a_k \in \{.5, 1, 2, 4\}$, so the daughter wavelets can be given by $h'_k(t) = h_{new}(t/a_k)$. We can formulate a two class neural network classifier that operates on the vector inner products of signals and daughter wavelets. Essentially, the signal is being projected onto four basis vectors as a dimensionality reduction and feature extraction measure. The signal is reduced from a great many samples in sample space to a four dimensional pattern vector. These pattern vectors are used as input to a two layer feed-forward neural network utilizing the backpropagation learning rule, as shown below in figure 3.

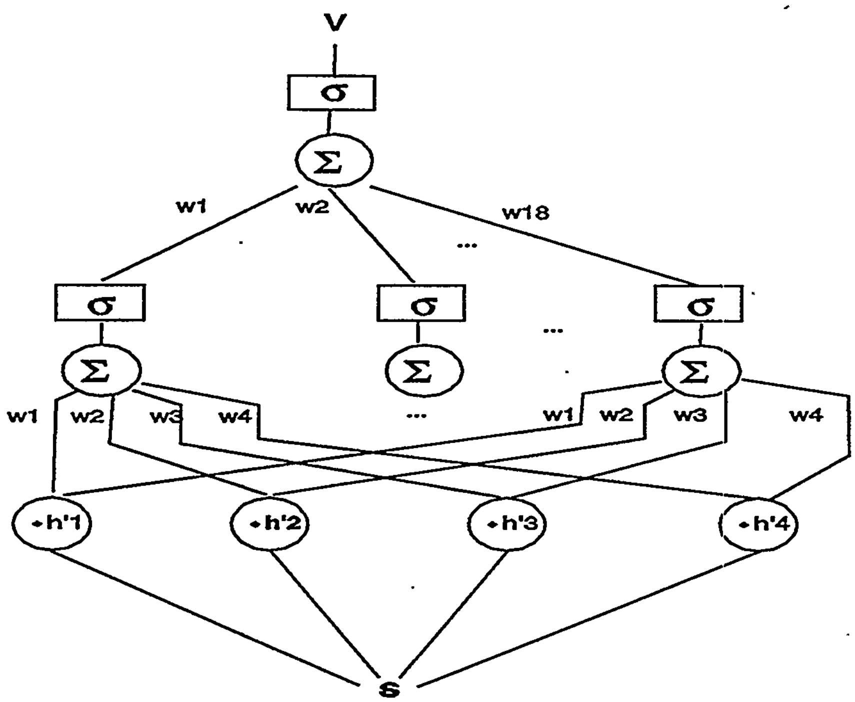


Figure 3: Adaptive wavelet neural network classifier.

A training set of signals was projected onto the daughter wavelets, yielding a training set of reduced dimensionality patterns. The network was trained for patterns containing the transient and not containing the transient with the expectation that the two classes would be separable based on their projections onto these daughter wavelets. Both patterns contained additive white Gaussian noise and added sinusoids at random frequencies for clutter.

RESULTS

TRAINING

The data consisted of synthesized vectors, 1024 points long, containing transients of randomly (uniform) varying TEC. The signals represent 25 MHz of bandwidth. White Gaussian noise (10dB SNR) was added to the transient signal vector. Sinusoids were then added to simulate clutter in the signal. The power of the sinusoids varied between (0, -5) dB. For the two class problem, vectors with the same noise and clutter power but lacking the transient were used. Spectrograms of samples from the two classes are shown below in figure (4).

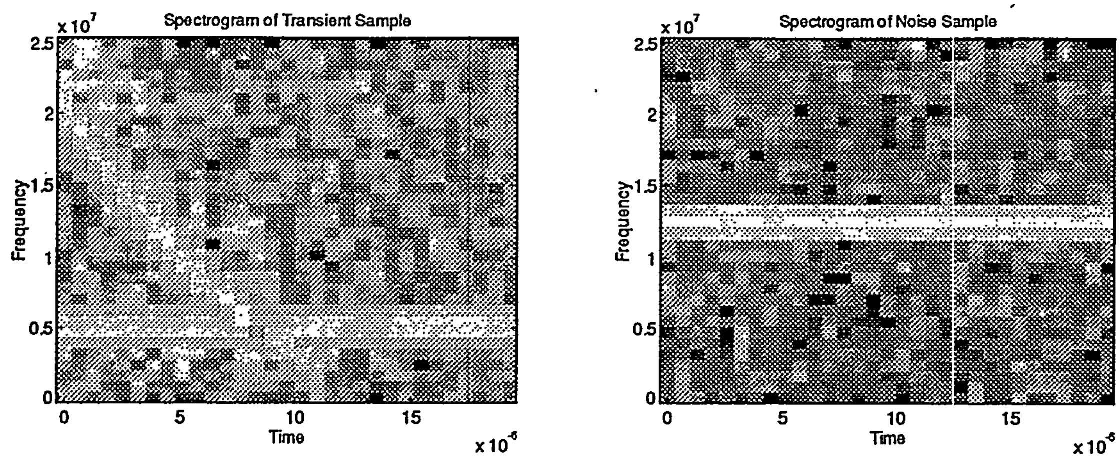


Figure 4: Two examples, one from each class, of the training data. The spectrogram shows frequency on the left and time on the independent axis; this shows the behavior of frequency over time.

To find the 'mother' wavelet, the weight vectors were initialized to small (0-.3) random numbers. The shifts, b, were initially set to zero. The scale, a, was chosen to vary linearly (e.g. 1, 2, 3 ... , 9). Nine GLFM daughter wavelets used in the synthesis of the mother wavelet. 60 pattern vectors were used, 30 from class one (transients) and 30 from class two (noise and clutter) for training data. Training was done in batch mode (weights, shifts, and scales updated after each epoch). A line search for the optimum learning parameters was employed. After 100 epochs, there where zero errors in the training set. A spectrogram of the mother wavelet determined by the network and scaled by .5 is shown in figure (5) below.

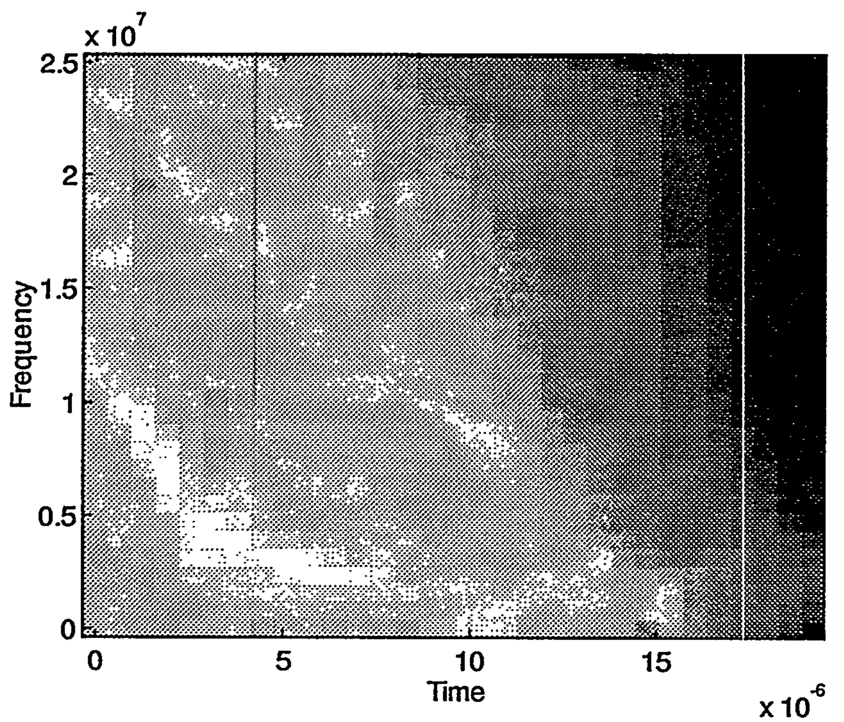


Figure 5: Spectrogram of the optimized 'mother' wavelet.

Notice the "quadratic" behavior of the instantaneous frequency, an artifact of the signal of interest, which is reassuring to find in the synthesized mother wavelet.

The neural network classifier based on the daughter wavelets of the above was trained using backpropagation. A second (different) training set of 60 patterns was projected on to the daughter wavelets yielding a training set of 60 four dimensional patterns. A two layer feed-forward neural network with 4 input nodes, 18 hidden nodes and one output node was used. The network was trained in batch mode for 100 iterations; the network properly classified all the training patterns.

Neural Network Classifier Based on FFT

For comparison, an FFT based neural network classifier was used; a more complete description of the technique can be found in previous work.[3] The input signals were divided into sequences of non overlapping windows of 128 points. The FFT was applied to the first two windows, the rest were discarded. The magnitude of the positive frequency component of each of the two windows (64 points each) were concatenated into a 128 point pattern vector as shown in figure (6) below.

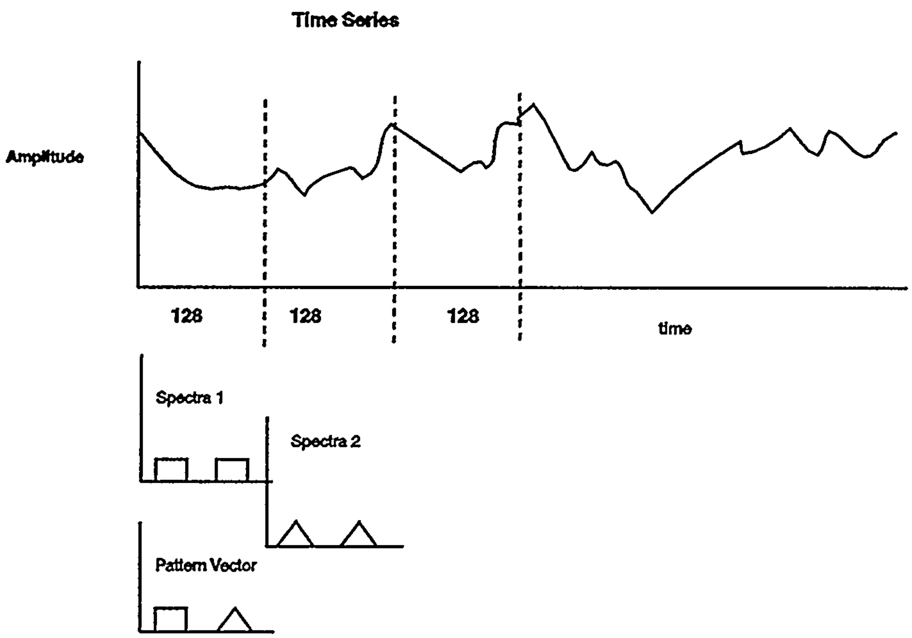


Figure 6: Obtaining the pattern vector from subsequent FFTs on segments of the time-series.

The pattern vectors form the training set for a two layer feed-forward neural network which also uses backpropagation. The first class contained a transient with additive white Gaussian noise and added sinusoids, as clutter. The second class contained noise and clutter, but no transient. The network used 18 hidden nodes and one output node. All the transfer functions were sigmoidal. The desired output for class one was a (1) and a (0) for class two. Forcing the output to a 1 or 0 by thresholding at .5 insured no unclassified patterns. The FFT based neural network classifier required 400 iterations in batch mode to train. This network made no errors on the training set.

TESTING

Applying test sets to the trained adaptive wavelet classifier yielded disappointing results. Test sets with the same level of noise and clutter as in the training data resulted in error rates around 40%. The complete failure of the network to generalize prompted another attempt in training. Utilizing a validation set, we monitored the classifiers ability to generalize as it learned the training set. The plot in figure (7), below, clearly indicates that the classifier only functions on the training set because it "memorizes" the patterns. At no time was the mean squared error (MSE) reduced in the validation set of patterns.

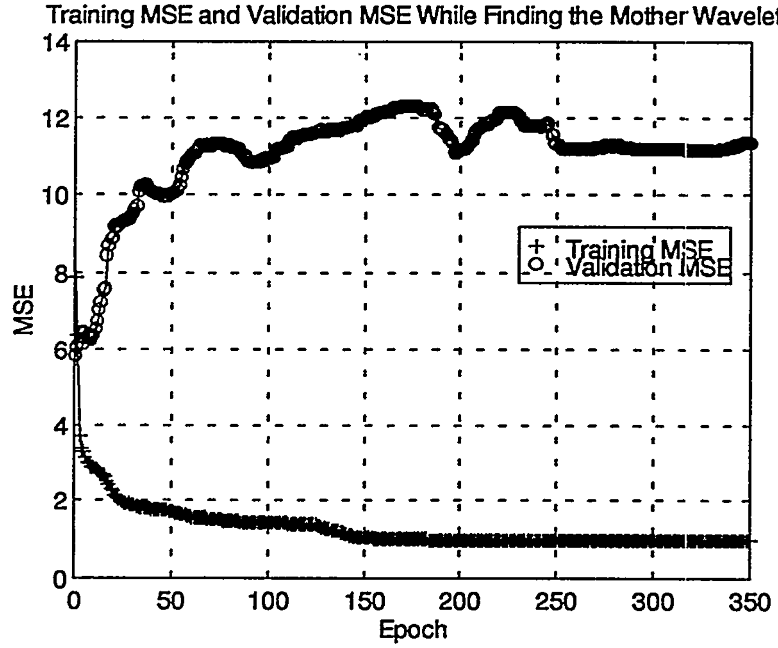
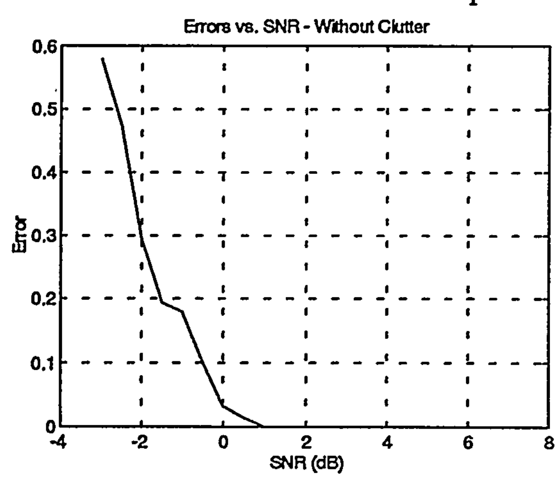


Figure 7: The classifier memorizes the training data; at no time is validation error reduced.

The test results from the FFT based neural network classifier were more encouraging. The two plots below indicate the percentage of errors in classification for a test set with no clutter and a test set with added sinusoids as clutter. Both plots show encouraging results for synthetic data.



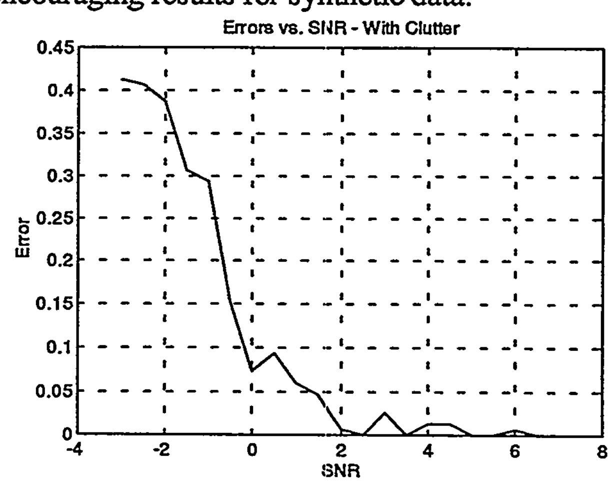


Figure 8: The ratio of errors vs. white noise SNR for test data both with and without clutter for the FFT based classifier.

The FFT based classifier makes no errors on the test data at the 10 dB SNR that the network was trained with which indicates excellent generalization characteristics.

DISCUSSION

The failure of the adaptive wavelet feature selection in test data indicates that the wavelets found are ineffective in extracting the frequency dispersion of the signal of interest. They are also ineffective at extracting the wide bandwidth characteristics of the signals. Two potential reasons seem most likely. First, scaling the mother wavelet as we did ($a_k \in \{.5, 1, 2, 4\}$) reduces the bandwidth of the daughter wavelet for scale greater than one. The frequency of the signal of interest always spans the bandwidth of the receiver, so the daughter wavelets, with reduced bandwidth, and the signal are mismatched. The result is that the four dimensional pattern vector obtained with the daughter wavelets contains perhaps one dimension of useful information.

The second likely problem is that the classifier fails to capture phase information. Using the vector inner product evaluates the convolution at one instant which is not guaranteed to have the optimal phase relationship between the wavelet and signal. Clearly, more work is required to further investigate the actual significance of this problem.

The results of the FFT based classifier indicate that the frequency spectra is a very effective method of extracting the features of the signal, and that it clearly outperforms the wavelet method in this application. The wavelet method should not be discarded, but the choice of signal for its application should be made carefully. It may be suitable for signals that vary in scale as shown in other work.[2]

ACKNOWLEDGMENTS

The authors wish to thank Kurt Moore and the Deployable Adaptive Hardware Project for their support. This work was carried out under the auspices of the United States Department of Energy.

REFERENCES

1. H. Szu, Y. Sheng, and J. Chen, "Wavelet transform as a bank of matched filters," *Applied Optics*, Vol. 31 No. 17, pp. 3267 - 3277, June 1992.

2. H. Szu, B. Telfer and S. Kadambe, "Neural network adaptive wavelets for signal representation and classification," *Optical Engineering*, Vol. 31 No. 9, pp. 1907 - 1916, September 1992.

3. K. R. Moore, et. al. , "A space -based classification for RF transients," Proceedings of the International Workshop on Artificial Intelligence Applications in Solar-Terrestrial Physics, Lund, Sweden, 1993, pp. 205, pub. by NOAA Space Environment Laboratory, Boulder, CO.

4. D. Enemark and M. Shipley, "The FORTE Receiver and Sub-Band Triggering Unit," Los Alamos National Laboratory Report LA-UR-94-2665, September 1994.

5. K. G. Budden, The Propagation of Radio Waves, Cambridge University Press, 1985.

6. A. Cichocki and R. Unbehauen, Neural Networks for Optimization and Signal Processing, pp. 90, 147, Wiley, New York, 1993.

7. S. Haykin, Neural Networks - A Comprehensive Foundation, pp. 215, Macmillan College Publishing Co., New York, 1994.

8. L. Weiss, "Wavelets and Wideband Correlation Processing," *IEEE Signal Processing Magazine*, pp. 13 - 31, January 1994.