Statistics and Computing (1995) **5**, 265–273

Application of wavelets to the pre-processing of underwater sounds

K. J. POWELL, T. SAPATINAS, T. C. BAILEY and W. J. KRZANOWSKI

Department of Mathematical Statistics and Operational Research, University of Exeter, Laver Building, North Park Road, Exeter, EX4 4QE, UK

Received October 1994 and accepted January 1995

In this paper we consider data on underwater sounds of differing types. Our objective is to filter background noise and achieve an acceptable level of reduction in the raw data, whilst at the same time maintaining the main features of the original signal. In particular, we consider data compression through the use of wavelet analysis followed by a thresholding of small coefficients in the resulting multiresolution decomposition. Various methods to threshold the wavelet representation are discussed and compared using recordings of dolphin sounds. An empirical modification to one of them is also proposed which shows promise in better preserving certain structures in our particular sound data.

Keywords: Underwater sounds, signal processing, wavelet decomposition, thresholding, discrimination.

1. Introduction

This paper reports some comparisons of various wavelet decomposition and thresholding methods to data on underwater sounds. Our objective is to filter the original 'noisy' recordings to produce a significant reduction in the amount of raw data, whilst preserving the main features of the original signal. The future intention is to use the resulting representation for discriminating between such sounds, either by employing multivariate discrimination methods, or by using a neural network (Smith *et al.*, 1993). However, this ultimate objective is only briefly referred to here, where the focus is mainly on the pre-processing stage.

The available data relate to underwater sounds of different types, including noises produced by dolphins, shrimps, seals, whales, ice breaking and others. They consist of selected extracts from lengthy recordings taken in real life conditions in the ocean, using a hydrophone placed several metres below sea level. The recording apparatus sampled the signal at 40 kHz with 16 bit resolution (near CD quality). The volume of raw data is therefore considerable. A typical example is shown in Figure 1, where a dolphin's 'whistle' is depicted over a period of 0.8 s and involves in excess of 30 000 data values, amplitude being between −32768 and +32767 (16 bit). The sampling rate of 40 kHz is 2.56 times the maximum frequency of sounds which we wish to identify and discriminate between, so there are few anti-aliasing problems. For this example we have deliberately selected a section of data that is clearly audible as a whistle, although there is also a series of short term 'clicks' towards the end of the portrayed signal.

The analysis of underwater sounds has similarities with human speech recognition. However, due to the special recording circumstance and the nature of the sounds themselves, there are unique features not found with the more commonly analysed speech data. For example, underwater data have a larger amount of background noise than is generally the case in speech recognition work. Speech data are also biased towards low frequency levels, whereas the underwater sounds are spread more evenly across the frequency range between 0 and 16 kHz. Finally, the *a priori* knowledge that can be brought to bear in the analysis of human speech is not available for sounds from dolphins or shrimps; for example, we cannot in general identify precisely when sounds start or for how long they last.

In the remainder of this paper we consider the results of applying wavelet methods to reduce background noise and enhance signal in this type of raw data. However, our discussion may also be applicable to data from other sources; for example, that obtained from spectroscopy, medical scanning, or tomography of various descriptions.

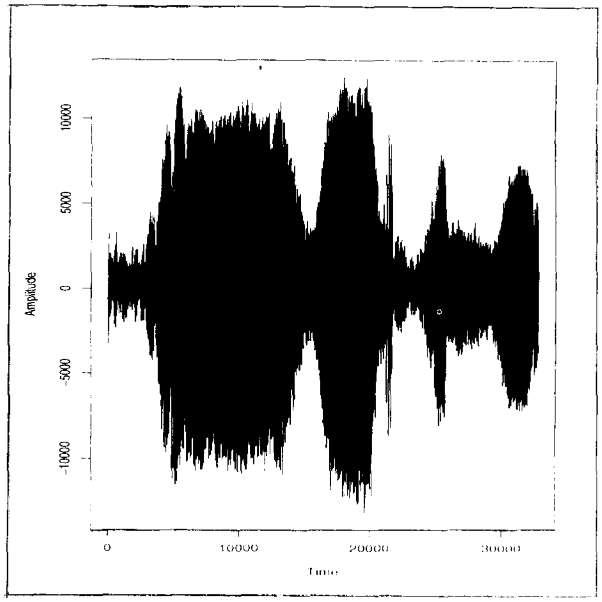

Fig. 1. *An example of a dophin whistle*

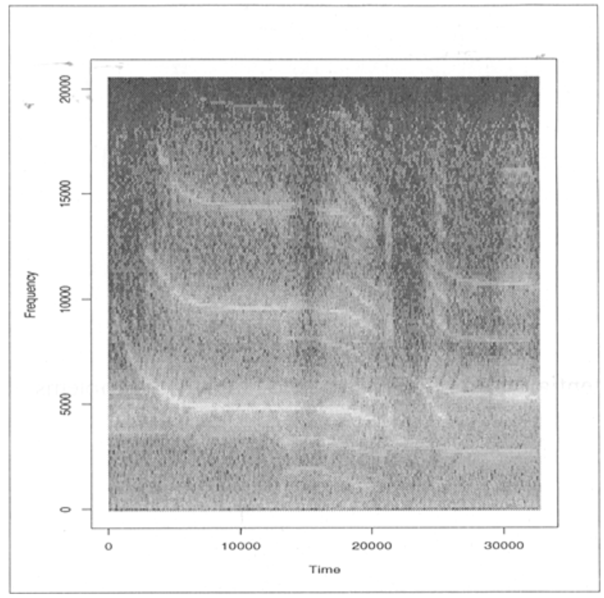

Fig. 2. *A spectrogram of the whistle shown in Fig.* 1

In Section 2, we first present arguments for preferring wavelet decomposition to more conventional spectral analysis based upon the short-time Fourier transform and briefly summarize some relevant aspects of wavelet theory. In Section 3, we discuss available techniques for thresholding the coefficients of a wavelet decomposition in order to reduce noise and enhance signal and suggest an empirically derived modification to one of these thresholding methods for use on underwater sounds. In Section 4, we compare the various thresholding methods using recordings of dolphin sounds. Finally, in Section 5, we discuss the implications of these results in the wider context of our ultimate objective to discriminate between different types of sounds.

2. The wavelet decomposition

One of the most common ways of pre-processing sound signals is to use a spectral analysis, based on the short time Fourier transform (STFT) (see, for example, Bloomfield, 1976). This presupposes that it is generally easier to analyse sound by frequency components rather than by using the amplitude at a given time point. In practice the infinite series of the general Fourier expansion is approximated for finite values in the STFT, a fast Fourier transform being employed on small successive 'windows' of the data. (A variety of different types of windows have been suggested, with various relative advantages.) A typical method of summarizing the results from such successive time windows is to display them as a 'spectrogram'. This has time on the horizontal axis and frequency on the vertical axis, and the value of the Fourier coefficient at each point is displayed as a greyscale (usually black for a low value through to white for a high value). Figure 2 shows a spectrogram of the dolphin 'whistle' from Fig. 1, based on a window size of 256 points.

Although such a representation is useful, there are a number of problems with using the STFT with sound data. Firstly, the Fourier transform inherently loses time information since it integrates over time. If the time window employed is too large, a STFT may provide a good broad description, but lose individual events of short duration. Such events occur frequently in underwater sound data. Window size therefore becomes an important factor in the analysis and results may be heavily dependent on the choice of this parameter. In general, it will not be possible to determine one window size which performs well across the whole range of sounds of interest, so different sizes are needed in different sections. Moreover, for each different window size a complete STFT computation is required, since one cannot construct the STFT for smaller windows directly from that using a larger window or vice versa. A second problem with Fourier analysis, particularly relevant in the discrimination context, is that power spectra of recordings made in different circumstances must first be levelled before their Fourier spectral decompositions can be compared. Without this levelling of spectra results are biased towards certain frequencies only, and since the background characteristics of different types of underwater sounds are quite different—each having been recorded in different circumstances—serious comparison of the spectral decompositions of different sounds cannot sensibly be undertaken.

For these reasons we consider an alternative decomposition using wavelets. This has the theoretical potential for better temporal resolution. It is also invariant to choice of the size of the successive time windows which are decomposed; one can immediately obtain the wavelet decomposition of any required sub-window of time from that of a larger window without the need for further computation. Furthermore, when it comes to comparison of the decompositions of sound recordings made in different circumstances, there is no requirement to employ additional (and usually *ad hoc*) methods to level the different power spectra; this is effectively taken care of automatically within the wavelet decomposition. All the above objections to Fourier analysis are thus overcome.

Wavelet analysis has been used recently as a data compression technique and as a way of smoothing time series such as digitized speech (Donoho, 1993). Some of the potential uses of wavelets for statistical problems have been recently discussed and developed by Donoho and Johnstone (1994, 1995) and Donoho *et al.* (1995). The discrete wavelet transform (DWT) is in some ways similar to the STFT, but is better able to zoom in on very short-lived high frequency phenomena, such as transients in signals or singularities in functions. The fast wavelet transform used in the DWT (as described by Mallat (1989)) is in principle faster than the fast Fourier transform, both of them being applied to 2^n observations, for some n. The DWT for this case is $O(n)$ whilst the FFT is $O(n \log n)$. For a Fourier series expansion the basis functions are $\sin(\cdot)$ and $\cos(\cdot)$ functions at different frequencies. With a wavelet series expansion the basis functions are all dilations and translations of a single function referred to as the mother wavelet and denoted by ψ. The dilation and translations of the mother wavelet are:

$$\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k), \quad j, k \in \mathbb{Z}, \tag{1}$$

where j is the dilation factor, and k is the translation factor. For certain choices of ψ the set of functions $\psi_{j,k}$ form an orthonormal basis for all functions in $L^2(\mathbb{R})$, and it is therefore possible to use the wavelets $\{\psi_{j,k}\}$ as the basis for an expansion of a function of interest. The wavelet series representation of such a function $f(\cdot) \in L^2(\mathbb{R})$ is then:

$$f(x) = \sum_{j,k} d_{j,k}\psi_{j,k}(x),$$

where the wavelet coefficients $d_{j,k}$ are given by:

$$d_{j,k} = \int_{-\infty}^{\infty} f(x)\psi_{j,k}(x)dx = \langle f, \psi_{j,k}\rangle,$$

where $\langle \cdot, \cdot \rangle$ denotes inner product.

Of course, the signals that we deal with are unlikely to reside in $L^2(\mathbb{R})$ but for certain choices of ψ, the wavelets can form bases for other spaces (e.g. Besov).

The oldest example of a function ψ for which the $\psi_{j,k}$ defined above constitute an orthonormal basis for a function $f(\cdot) \in L^2(\mathbb{R})$ is the Haar function defined by $\psi(x) = 1$ on $[0, 0.5)$, $\psi(x) = -1$ on $[0.5, 1]$ and $\psi(x) = 0$ elsewhere. However, a major advance in this area was the realization that there are many other families, based on mother wavelets other than the Haar function. In particular Daubechies (1988) suggested families that have better time–frequency localization than that based on the Haar function. Daubechie's orthonormal wavelet bases can be constructed via a multiresolution analysis as developed by Mallat (1989). Therein, the wavelet coefficients $d_{j,k}$ of a function $f(\cdot)$ for a fixed j, describe the difference between two approximations of $f(\cdot)$, one with resolution 2^j and one with coarser resolution 2^{j-1}.

Daubechies (1988) constructed two families of wavelets which are referred to as the *extremal phase* or *least asymmetric* wavelets. These families possess properties such as regularity, compact support and vanishing moments. (Daubechies (1992) provides important implications of these various properties for specific applications.) In particular, the mother wavelets, indexed by an integer N and denoted by $\psi_N(x)$, have regularity proportional to N. Regularity enables the wavelet to be selected according to the smoothness of the signal to be detected. In later sections we focus on the cases $N = 2$ and $N = 6$.

3. Thresholding of wavelet coefficients

Whilst a wavelet decomposition of a raw signal (that is the set of coefficients resulting from the analysis) can be of value in understanding the structure of the signal and comparing different types of signals, it does not in itself reduce noise or enhance the signal in any way. Indeed the original raw signal can be reconstructed precisely and uniquely from the wavelet decomposition; there are as many coefficients as original data points. A further step, called 'thresholding', is thus employed to reduce the number of coefficients by removing small coefficients (considered to be noise) and leaving only the those values deemed to be significant according to some chosen criteria; the retained coefficients can then be used in subsequent analysis in place of the full decomposition of the raw data. Only an approximation to the original data can be reconstructed after thresholding; the hope is that a cleaner version of the original signal will result when only the significant components are retained. Such reconstructions are unique, and in many cases the compression in the number of data points can be very large whilst the reconstruction still approximates the original data fairly well. Such an approach is attractive in the current application, particularly given the ultimate aim of discrimination between different sound types.

Donoho and Johnstone (1994, 1995) discuss thresholds appropriate for recovery of a function of unknown smoothness from noisy, sampled data. Suppose data $y_1, y_2, \ldots, y_n$ $(n = 2^{J+1})$ can be modelled by

$$y_i = f(x_i) + \epsilon_i, \quad i = 1, 2, \ldots, n, \tag{2}$$

where the $\{\epsilon_i\}_{i=1}^n$ are independent and identically distributed (i.i.d.) $N(0, \sigma^2)$, and $f(\cdot)$ is the function we would like to estimate. Suppose further that $f = \{f(x_i)\}_{i=1}^n$ and

$\hat{f} = \{\hat{f}(x_i)\}_{i=1}^n$ are the vectors of true and estimated sample values, and that we wish to minimize:

$$R(\hat{f}, f) = \frac{E\|\hat{f} - f\|}{n}.$$

Let $\mathcal{W}$ denote the DWT, emphasizing that the transform is implemented by a sequence of special finite-length filtering steps and not by matrix manipulation. Its orthogonality has a fundamental statistical consequence: it transforms white noise into white noise. Hence, the above model (2) is equivalent to

$$y_{j,k} = w_{j,k} + \epsilon_{j,k}, \quad j = 0, 1, \ldots, J; \quad k = 0, 1, \ldots, 2^j - 1,$$

where $\{y_{j,k}\}$ are the wavelet coefficients of $\{y_i\}$, $\{w_{j,k}\}$ are the wavelet coefficients of $\{f(x_i)\}$, and $\{\epsilon_{j,k}\}$ is an i.i.d. $N(0, \sigma^2)$ noise sequence. Consequently, if $\hat{w}_{j,k}$ are estimates of the coefficients (depending on $y_{j,k}$), then there is an estimate $\hat{f}$ of f obtained by $\hat{f} = \mathcal{W}^{-1}\hat{w}$, where $\mathcal{W}^{-1}$ denotes the inverse of the DWT. Thus, by thresholding the wavelet coefficients $w_{j,k}$ to produce $w^*_{j,k}$, we can reconstruct an estimate of f by $\hat{f}^* = \mathcal{W}^{-1}\hat{w}^*$.

The choices of appropriate wavelet transform and thresholding method are interrelated. In general, given a threshold value, the production of the $w^*_{j,k}$ can be either by 'hard' or 'soft' thresholding. Hard thresholding simply compares a coefficient with the threshold, if it is larger in absolute magnitude it is left alone, otherwise it is set to zero. Soft thresholding modifies all coefficients $w_{j,k}$ by the formula $w^*_{j,k} = \text{sgn}(w_{j,k})\ \ (|w_{j,k}| - \lambda)I(|w_{j,k}| > \lambda)$, where sgn is the sign of $w_{j,k}$, λ is the threshold, and I is the usual indicator function.

Choice of the threshold value will depend on the noise level σ in the data. In practice σ will need to be estimated from the data, since the true noise level is unlikely to be known. The standard deviation of the $w_{j,k}$ provides one such estimate but, as mentioned in Donoho and Johnstone (1994, 1995), it is important to use a robust estimator. They proposed the median absolute deviation of the wavelet coefficients at the level of the decomposition corresponding to the maximum dilation, divided by 0.6745, because the wavelet coefficients at this level are, with a few exceptions, essentially pure noise. Given an estimate of σ, one then needs to decide on how to use this to set an appropriate threshold value to 'shrink' the empirical wavelet coefficients. Donoho and Johnstone (1994, 1995) proposed the following three methods (see therein for more details and statistical properties):

1. A spatially adaptive method which they refer to as *RiskShrink*. *RiskShrink* mimics the performance of an 'oracle' for selective wavelet reconstruction as well as it is possible to do so. Based on a new inequality in multivariate normal decision theory (which they call its *oracle inequality*), they have shown that attained performance differs from ideal performance by at most a factor $\sim 2\log(n)$, where n is the sample size.

2. Use of the threshold $\lambda = \sigma\sqrt{2\log(n)}$, which they refer to as *VisuShrink* (also known as *universal thresholding*). This is based on the result that when $\{\epsilon_i\}$ is a white noise sequence of independent and identically distributed N(0,1) errors, then $P\{\max_i|\epsilon_i| > \sqrt{2\log(n)}\} \to 0$ as $n \to \infty$. *VisuShrink* is easy to compute, and is asymptotically optimal.

3. An estimated threshold value which is in a sense optimally smoothness-adaptive and which they refer to as *SureShrink*. Here a threshold value is assigned to each level of the wavelet decomposition by the principle of minimizing the Stein unbiased estimate (Stein, 1981) of risk for thresholded estimates. They show that *SureShrink* is asymptotically near-minimax and the computational effort of the overall procedure is of order $n\log(n)$.

We now propose a fourth method, which is an empirically derived modification of *VisuShrink* and which we have found to perform particularly well with our underwater sounds. The method is motivated by requiring the thresholding method to retain some of the temporal structure present in the coefficients in any level of the decomposition, as well as taking into account their relative magnitudes. It is therefore essentially in the spirit of a 'position dependent' threshold, which Donoho *et al.* (1995) suggested might be of value in some applications.

4. The method is based on the simple idea of performing multiple non-parametric runs tests on groups of coefficients within each level of the decomposition. More specifically, each level of the wavelet decomposition is split into equal-sized groups of coefficients, where we arrange the number of groups to be the same on each level. This means a group corresponds to the same time period on all levels. If the number of coefficients that this produces in a group at any level is less than 32 then the runs test is not applied and the *VisuShrink* alone is used. Otherwise, significance of runs within each group at any level is assessed on the basis of the observed number of runs of either positive or negative coefficients, using the usual approximate test statistic (see, for example, Siegel, 1956). If the two-tailed test is significant at 5%, then the new threshold value is taken as a pre-specific percentage of the overall *VisuShrink* value (currently we use a value of 10%). Otherwise, the overall *VisuShrink* is used. We have implemented this procedure in the *WaveThresh* package developed within S-Plus by Nason and Silverman (1994) as a new threshold function called *RunShrink* for the wavelet decomposition object. The syntax is the same as for *VisuShrink* with optional parameters which allow control of the runs test applied.

4. Comparison of different thresholding methods

In this section we apply and compare the thresholding

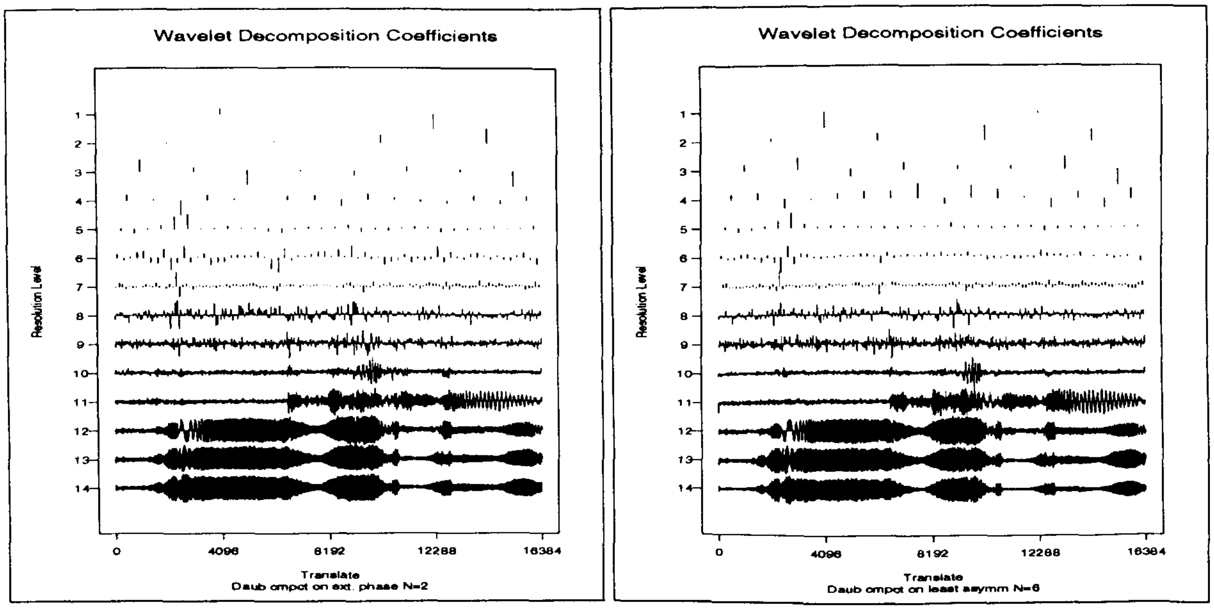

Fig. 3. *Wavelet decompositions of the whistle using* extremal phase *for $N = 2$ and* least asymmetric *for $N = 6$, respectively*

methods given above using the underwater sound data described earlier. For simplicity, we present here results relating to the example of dolphin sound given in Fig. 1. In fact we have analysed a far more extensive range of examples, both in terms of background noise level and type of sound, and the results obtained have been broadly similar to those presented here.

We have used the *WaveThresh* package, referred to earlier, for the *VisuShrink* and *RunShrink* thresholding. For *RiskShrink* and *SureShrink* we have used Matlab functions developed by Donoho and Johnstone and then transferred results into S-Plus for graphical presentation; all the figures that follow are therefore from *WaveThresh*. In the wavelet decomposition we employed two different wavelets:

(i) The *extremal phase* compactly supported wavelet for $N = 2$ (these are defined through a set of four non-zero coefficients whose numerical values may be found in Daubechies (1992, Table 6.1, p. 195)).

(ii) The *least asymmetric* compactly supported wavelet for $N = 6$ (these are defined through a set of 12 non-zero coefficients whose numerical values may be found in Daubechies (1992, Table 6.3, p. 198)).

Figure 3 shows the wavelet decomposition of the complete

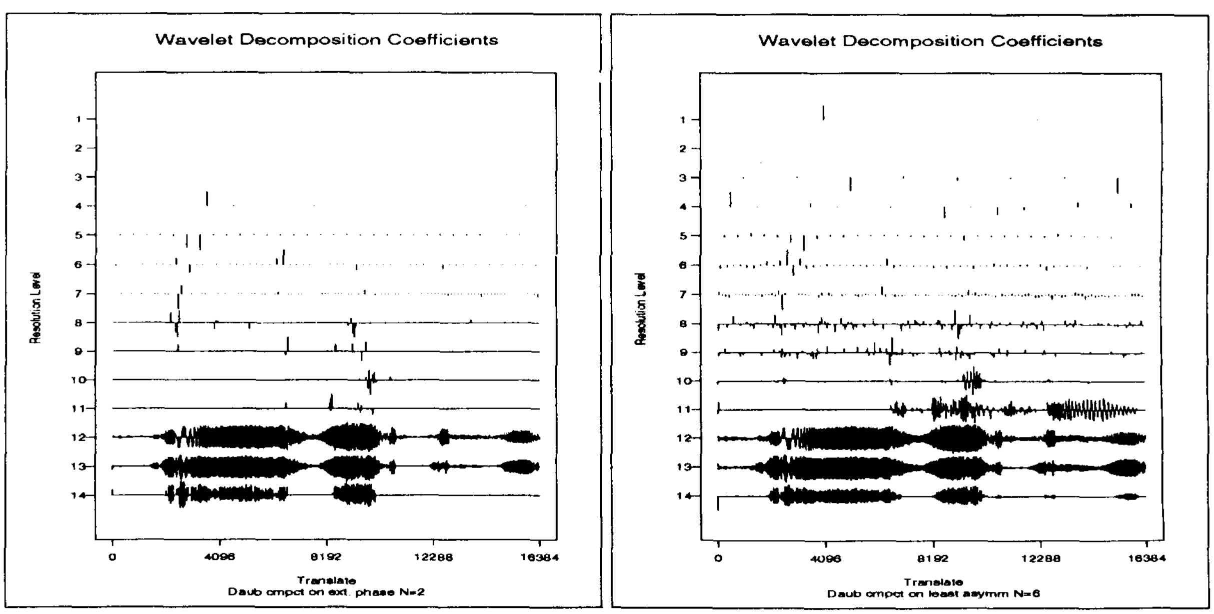

Fig. 4. *Thresholding the wavelet decompositions of the whistle using* RiskShrink

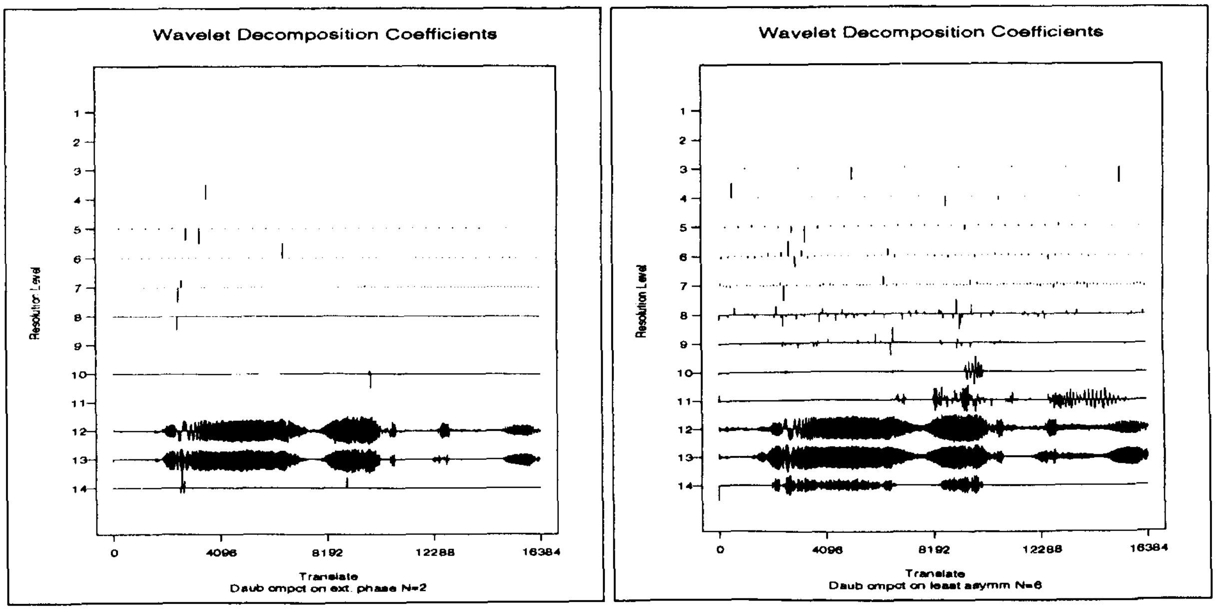

Fig. 5. *Thresholding the wavelet decompositions of the whistle using* VisuShrink

dolphin whistle shown in Fig. 1. Figures 4, 5, 6 and 7 show the results of the *RiskShrink*, *VisuShrink*, *SureShrink* and *RunShrink* of the previous wavelet analyses on the original dolphin whistle. Table 1 summarizes the number of non-zero coefficients and the corresponding compressions of these figures. In all thresholding methods we have used soft thresholding, and the median absolute deviation of the wavelet coefficients at the final level J, divided by 0.6745, has been taken as an estimate of the noise level. Donoho and Johnstone (1994, 1995) have discussed various statistical advantages of using soft thresholding. In fact, similar results held when we used hard thresholding instead of soft, although we do not reproduce the results here.

In measuring how well the various thresholding methods work, previous studies, which have mostly used simulated data, have been able to compare reconstructions based on the thresholded coefficients with the underlying, known, signal. In our case, as we are using real data and the underlying true signal is unknown (and is doubtless very complex) this is not a viable method. However, with these particular data a natural way to test the results presents itself; that of listening to the reconstructed signals and comparing them with the original. Although this may appear

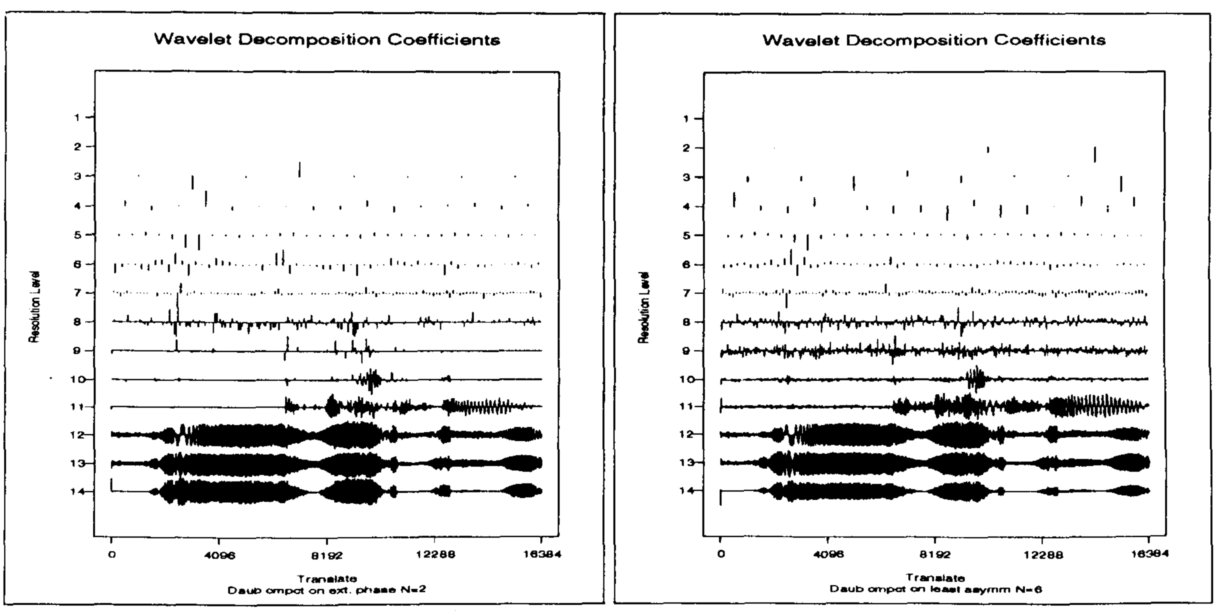

Fig. 6. *Thresholding the wavelet decompositions of the whistle using* SureShrink

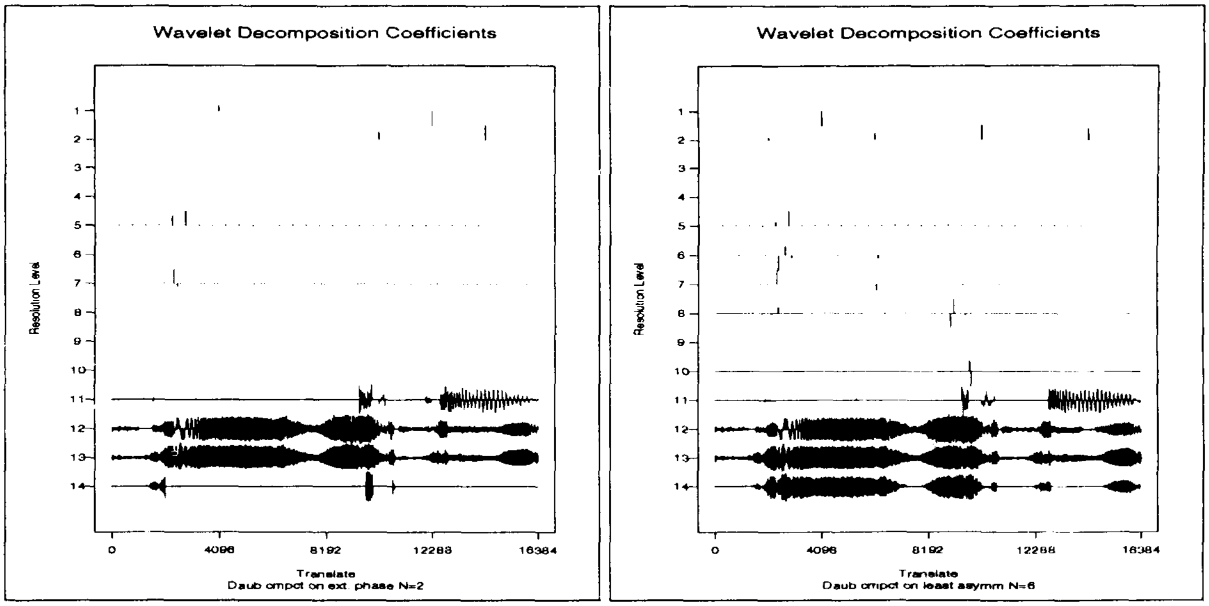

Fig. 7. *Thresholding the wavelet decompositions of the whistle using* RunShrink

subjective, there is really no quantitative alternative available given that the true signal is unknown. In any case, even if such a quantitative comparison were possible, it is highly doubtful that it would be more sensitive than the human ear at detecting differences in the quality of sound. In fact when the sounds are actually heard, there are very clear and easily discernible audible differences in the quality of the reconstructions obtained from the various thresholding methods.

To obtain an approximate quantitative measure of the reproduction quality of the different thresholding methods, a panel of four musicians was asked to rank the quality of the eight reconstructions in terms of closeness to the original whistle. The panel were played the reconstructions in different random orders but the rankings were remarkably consistent—see Table 2. The two

Table 1. *Numbers of non-zero coefficients and compressions for the* RiskShrink, VisuShrink, SureShrink *and* RunShrink *thresholdings*

	#Non-zero coefficients	Compression
RiskShrink		
extremal phase $N = 2$	8111	75%
least asymmetric $N = 6$	13353	59%
VisuShrink		
extremal phase $N = 2$	5157	84%
least asymmetric $N = 6$	10763	67%
SureShrink		
extremal phase $N = 2$	19292	41%
least asymmetric $N = 6$	23336	29%
RunShrink		
extremal phase $N = 2$	11214	66%
least asymmetric $N = 6$	17242	47%

RunShrink reconstructions were rated as the closest to the original by all the panel whilst the $N = 2$ *RiskShrink*, *SureShrink* and *VisuShrink* reconstructions were always rated as the least like the original. For the $N = 6$ results, *RiskShrink* and *SureShrink* were rated as slightly better than *VisuShrink*.

Overall, therefore, considering the relative compression ratios and quality of reconstructed sound, *RunShrink* with $N = 2$ provided the most acceptable results; the reconstructed sound in this case was dramatically clearer than the original noisy signal and involved a data compression of 66%.

5. Conclusions

Wavelet analysis has been applied to underwater sounds, with the objective of filtering the original 'noisy' recordings to produce a significant reduction in the amount of raw data, whilst preserving the main features of the original signal. It has been demonstrated that some of the common methods employed to threshold the wavelet coefficients do not work well with these particular kind of data. A modification of the *VisuShrink* method has been proposed which shows promise in better preserving certain structures in our particular underwater sound data.

Reflecting on these results, one of the reasons that some of the common thresholdings fail with these kind of data may be the fact that background noise is unlikely to be normal. For example, there are arguments to suggest that errors in the underwater noise waveform may be more likely to exhibit a Rayleigh distribution (Creasey *et al.*, 1989a, b). The common thresholding methods we have

Table 2. *The panels rankings for the* RiskShrink, VisuShrink, SureShrink *and* RunShrink *thresholdings*

	Subject 1	Subject 2	Subject 3	Subject 4	Average Rank
RiskShrink $N = 2$	7	7	8	6	7
RiskShrink $N = 6$	4	3	4	3	4
VisuShrink $N = 2$	8	7	7	8	8
VisuShrink $N = 6$	4	3	5	4	5
SureShrink $N = 2$	6	6	6	6	6
SureShrink $N = 6$	3	3	3	4	3
RunShrink $N = 2$	2	1	2	1	1
RunShrink $N = 6$	1	2	1	2	1

presented are statistically justified on the basis of noise that is normal, or close enough to normality based on central limit theorem considerations. Moreover, (linear) orthogonal wavelet analysis is itself theoretically tied to the assumption of normality. For example, Donoho (1993) shows that if one considers Cauchy noise then the common techniques of Section 3 are not in any sense optimal. Currently, there is little theory or methodology developed in the case of non-normal errors. Donoho (1993) has proposed using non-linear multiresolution analysis instead of linear analysis, and further theoretical work in this direction is still in progress. Nason (1994) has suggested a thresholding method based on cross-validation which makes no assumption about noise distribution. However, this method would not necessarily be applicable to our data where the true signal is non-sparse, since it depends upon a linear interpolation between every other time point. Indeed, Nason observes that, in general, the cross-validation thresholding method seems to work better when estimating a smooth signal, while *SureShrink* tends to do better with non-sparse signals. At present, then, the question of appropriate thresholding of wavelet decompositions remains open and ultimately continues to be empirical and data dependent.

We continue to consider improved thresholding methods for our data. For example, we are looking at the possibility of basing threshold choice on the criterion of maximizing certain differences in the reconstructions of different types of sound. In other words of creating a direct link between thresholding and our ultimate discrimination objective. In the meantime, *RunShrink* seems to perform reasonably well with our data. Certainly, it provides an acceptable balance between suitable signal reconstruction and data compression and allows us to consider using the resulting thresholded wavelet decomposition as a basis for signal discrimination. In this respect, we are currently investigating summarizing the thresholded wavelet coefficients within small temporal windows in terms of several characteristics, such as mean and variability within and across levels, as well as various aspects of the sequential pattern within level. These summary variables then effectively provide us with a representation of the raw signal as a multivariate time series observed at equally spaced time points; so forming a basis for the application of discrimination techniques, either of a statistical nature, or using neural networks. This work is still in progress, but initial results are encouraging.

References

Bloomfield, P. (1976) *Fourier Analysis of Time Series: An Introduction.* Wiley, New York.

Creasey, D. J., Smith, B. V. and Gazey, B. K. (1989a) *Underwater Acoustics and Sonar Systems—Acoustics, Transducers and Propagation*, Vol. 1. Lecture Notes, Birmingham University.

Creasey, D. J., Smith, B. V. and Gazey, B. K. (1989b) *Underwater Acoustics and Sonar Systems—Signal Processing and Systems*, Vol. 2. Lecture Notes, Birmingham University.

Daubechies, I. (1988) Orthonormal bases of compactly supported wavelets. *Communications in Pure and Applied Mathematics*, **41,** 909–96.

Daubechies, I. (1992) *Ten Lectures on Wavelets.* SIAM, Philadelphia, PA.

Donoho, D. L. (1993) Non-linear wavelet methods for recovery of signals, densities and spectra from indirect and noisy data. In *Proceedings of Symposia in Applied Mathematics: Different Perspectives on Wavelets*, **47,** Daubechies I. (ed), pp. 173–205. American Mathematical Society, Providence, RI.

Donoho, D. L. and Johnstone, I. M. (1994) Ideal spatial adaption by wavelet shrinkage. *Biometrika*, **81,** 425–455.

Donoho, D. L. and Johnstone, I. M. (1995) Adapting to unknown smoothness via wavelet shinkage. *Journal of the American Statistical Association*, (to appear).

Donoho, D. L., Johnstone, I. M., Kerkyacharian, G. and Picard, D. (1995) Wavelet shrinkage: asymptopia? (with discussion). *Journal of the Royal Statistical Society*, Series B, **57,** 301–369.

Mallat, S. G. (1989) A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **11,** 674–93.

Nason, G. P. (1993) The WaveThresh package; wavelet transform and thresholding for S. Available from the Statlib archive.

Nason, G. P. (1994) Wavelet regression by cross-validation. Technical Report **447,** Department of Statistics, Stanford University, USA.

Nason, G. P. and Silverman, B. W. (1994) The discrete wavelet transform in S. *Journal of Computational and Graphical Statistics*, **3**, 163–91.

Siegel, S. (1956) *Nonparametric Statistics for the Behavioral Sciences*. McGraw-Hill Kogakusha, Tokyo.

Smith, D. J., Bailey, T. C. and Munford, A. G. (1993) Robust classification of high dimensional data using artifical neural networks. *Statistics and Computing*, **3**, 71–81.

Stein, C. (1981) Estimation of the mean of a multivariate normal distribution. *Annals of Statistics*, **9**, 1135–51.