Proceedings of the Institute of Acoustics

THE DETECTION OF UNDERWATER ACOUSTIC TRANSIENTS USING ADAPTIVE FILTERS AND TIME-FREQUENCY METHODS

P.R. White

University of Southampton, Highfield, Southampton, Hants., U.K.

1 INTRODUCTION

A recent area of much research activity in the field of passive SONAR surveillance systems, has dedicated to the development of signal processing tools for detecting and analysing short term, broad-band signals (transients). This paper derives two transient detection algorithms which utilise existing processing techniques. The transient signals we study are of many varieties, some of which are of a random nature. By random we mean that the exact nature of a transient signal generated by an event is unpredictable *a priori*. In this paper we shall only discuss algorithms which aim to detect transients and avoid discussion of the analysis problem. The algorithms considered must be capable of detecting a wide variety of signals. One common feature of these transients is that they tend to be short in duration and are relatively rare events. These transients signals are immersed in oceanic background noise and may be at relatively low Signal to Noise Ratios (SNRs).

The premise on which this work is based is that one is interested in detecting any events which do not appear to be consistent with the background noise. The advantage of this approach is that one makes no assumptions about the nature of the transients one is endeavouring to detect, so one does not unfairly discriminate against any signal types, resulting in techniques which are widely applicable. The disadvantage is that for any given class of signal one may be able to construct a detector which has better performance than those proposed herein.

Most of the transient detection algorithms previously proposed in the literature make assumptions about the nature of the transient signals which they aim to detect. These assumptions may be quite specific, as in [1] where it is assumed that the transients are decaying exponentials or relatively general, as in [2] where it is assumed that the transient signal produces a significant 'foot print' in the outer regions of the bispectrum. These, and other techniques [3,4,5], by restricting the signal type tend to exchange good performance on a small sub-set of transients for a loss of performance on transients outside this sub-set.

2 FUNDAMENTALS

Since this paper deliberately avoids modelling the transient signals to be detected, the only realistic route remaining is to model the background noise. Here we shall assume that the background noise can be modelled as a wide sense stationary Gaussian random process with zero mean. There are several parts to this assumption the implications of which can be profound. The assumption of wide sense stationarity is more specifically stated by saying that one requires the statistics of the signal (up to second order) to be approximately time independent, over any analysis window considered. The assumption of Gaussianity enables us to restrict our attention to the first two moments of the data. However, not all oceanic

background noise signals are Gaussian in nature. One example of where Gaussianity is a poor assumption is in the polar regions where the noise of 'ice cracking' [6] can produce a skewed non-Gaussian random process. Such problems may be mitigated since SONAR systems often beamform data from many hydrophones prior to any detection algorithm. The process of beamforming essentially involves summing the signals from many hydrophones. The tendency will be to make the beamformer output 'more like' a Gaussian process than the input signal (as a consequence of the Central Limit Theorem). Another example of a scenario where the background noise is non-Gaussian occurs when a tonal component is present in the background noise. These tonal signals cannot be incorporated into the Gaussian assumption and simple beamforming does not necessarily make such components 'more Gaussian'; indeed it is often the case that the beamformer aims to enhance these tonal signals.

If one considers a window of L data samples arranged in a vector, $\underline{x}_n$, defined as

$$\underline{x}_n = [x(n)\ x(n-1)\ x(n-2)\ \ldots\ x(n-L+1)]^t$$

where $x(n)$ is the input data sampled at time interval n, and ^{t} denotes matrix transposition. Under the assumption of Gaussianity one can simply write down the value probability density function (p.d.f.) evaluated for a particular data window. This is

$$p(\underline{x}_n) = \frac{1}{\sqrt{2\pi|R|}} \exp(-\underline{x}_n R^{-1} \underline{x}_n / 2) \qquad (1)$$

where R is the so called auto-correlation matrix defined as $E[\underline{x}_n \underline{x}_n^t]$. Thus if one knows the auto-correlation matrix one can evaluate *(1)* the result of which is a measure of the likelihood that the data in $\underline{x}_n$ was generated from a Gaussian process with the ascribed auto-correlation function. Note since *(1)* describes a probability density function its evaluation can not be interpreted as giving a probability. Yet it is still appropriate to say that if $\underline{x}_n$ relates to a small value of $p(\underline{x}_n)$ then the data window is unlikely to have arisen from a Gaussian process with a correlation matrix R.

Indeed one need not evaluate the entire expression in *(1)* since the only data dependent term is

$$\sigma(\underline{x}_n) = \underline{x}_n^t R^{-1} \underline{x}_n \qquad (2)$$

This quantity is positive (due the semi-positive nature of the auto-correlation matrix). Clearly large values of the above quadratic term correspond to small values of the probability density function, *i.e.* data windows which do not conform well with the background noise model.

The aim of the methods discussed here is to approximate the quadratic term in *(2)* via various techniques. These approaches can be viewed via different factorisations of the inverse correlation matrix.

Firstly we shall consider the Cholesky factorisation [7], so that

$$R^{-1} = C^t C$$

where C is a lower triangular matrix. Substituting this into *(2)* yields

$$\sigma(x_n) = x_n C^t C x_n = q_n^t q_n \qquad (3)$$

where $q_n = C x_n = [\, q_1(n)\ q_2(n)\ \ldots\ q_L(n)\,]^t$ the lower triangular structure of the matrix C ensures that $q_k(n)$ is a linear combination of only the data samples : $x(n), x(n-1), \ldots, x(n-k+1)$.

The multiplication of the data vector by the matrix C can be viewed as a transformation of the data vector x_n into the new vector q_n. The transformed data vector's correlation matrix can also be evaluated as :

$$E[\, q_n q_n^t \,] = C\, E[\, x_n x_n^t \,]\, C^t = C R C^t = I$$

where I is the $L \times L$ identity matrix. This illustrates that the data transformation serves to whiten the data vector. This whitening is achieved by forming linear combinations of the current data element and only the preceding samples in the data vector. As such this is simply a Gram-Schmidt orthogonalisation [8].

The Cholesky factorisation of the inverse auto-correlation does not produce a unique whitening transformation matrix. Another factorisation of the inverse auto-correlation matrix is

$$R^{-1} = V \Lambda^{-1/2} \Lambda^{-1/2} V'$$

where $'$ denotes the conjugate transpose, V is the orthonormal matrix whose columns are the eigenvectors of R and $\Lambda^{-1/2}$ is the diagonal matrix whose elements are the eigenvalues of R, raised to the power $-1/2$. Thus by defining the matrix Γ as

$$\Gamma = V \Lambda^{-1/2} \qquad \text{so that} \qquad R^{-1} = \Gamma \Gamma'$$

The transformed data vector $p_n = \Gamma x_n$ is also a whitened vector. Thus the test statistic is

$$\sigma(x_n) = x_n^t \Gamma \Gamma' x_n = p_n^t p_n \qquad (4)$$

It should be noted that the quantities defined by *(2)*, *(3)* and *(4)* are all the same. If the data vector, x_n, is Gaussian then because the transformations are linear, then the vectors, p_n and q_n, are also Gaussian. We can also see that the elements of p_n also have unit variance, thus one can conclude that the test statistic $\sigma(x_n)$ is a Chi-squared random variable with L degrees of freedom.

3 ALGORITHMS

The preceding section describes the quantity we wish to employ to test the hypothesis that a data window under examination is due to a Gaussian process with a correlation matrix R. We now examine the practicalities of calculating estimates of this quantity. One requires that any such algorithm is not overly computationally demanding. It is assumed that the correlation matrix is unknown *a priori*, and an on line estimate of it must be implicitly, or explicitly, calculated. The two classes of methods discussed herein use adaptive filters and time-frequencies methods which in turn exploit the factorisations used in *(3)* and *(4)* respectively.

The broad class of adaptive filters which interests us here are the Finite Impulse Response (FIR) lattice filters. The lattice structures which they incorporate perform the Gram-Schmidt orthogonalisation required to use *(3)*. The class of FIR lattices can further be subdivided into the gradient based algorithms [9] and the exact least squares algorithms. The recent development of numerically stable exact least squares algorithms, referred to as fast QR Lattices [10], which only require $O(L)$ operations per update cycle, make these a feasible candidate. The gradient methods require less computation than the fast QR formulation but their performance tends to be marginally inferior. The above distinctions only apply to the coefficient updates within the lattice and as such do not significantly affect the following discussion.

The general form of an FIR lattice is depicted in Figure 1. As the input sequence is passed down the structure its dependence on previous stages is removed. This is achieved by use of the reflection coefficients, which are updated via the chosen adaptive scheme. The output of each stage of the lattice is called the backward prediction error, and is orthogonal to all the preceding backward prediction errors, *i.e.* $E[\, b_i(n)\, b_j(n)\,] = 0$ for $i \neq j$.

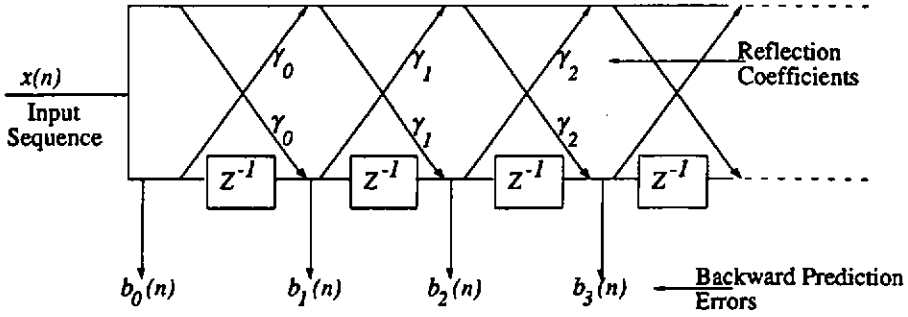

Figure 1 : Finite Impulse Response Lattice

The variance of the k^{th} backward prediction error is denoted as $E_b(k)$ and is given by

$$E_b(k) = E[\, b_k(n)^2\,] \approx (1-\lambda) \sum_{m=0}^{n} \lambda^{n-m}\, b_k(m)^2 = B_k(n)$$

Note the estimates $B_k(n)$ are obtained by exponentially averaging the $b_k(m)^2$. The constant λ should lie in the range $(0, 1)$ and controls the time constant associated with the averaging process, this time constant is $1/(1-\lambda)$. The $B_k(n)$ can be recursively calculated by :

$$B_k(n) = \lambda B_k(n-1) + (1-\lambda)\, b_k(n)^2$$

Using these quantities (3) can be evaluated using :

$$\sigma(\underline{x}_n) \approx \sum_{k=1}^{L} \frac{b_k(n)^2}{B_k(n)} \qquad (5)$$

This approach is most suitable for use with the gradient based lattice filter algorithms. The fast QR lattices calculate this quantity in a far more direct fashion, their update equations use a variable called the likelihood variable, which can be expanded as

$$\sigma(\underline{x}_n) \approx \underline{x}_n^t \left(\sum_{m=0}^{n} \lambda^{n-m}\, \underline{x}_m\, \underline{x}_m^t \right)^{-1} \underline{x}_n \qquad (6)$$

where the term in the brackets is an approximation to the input auto-correlation matrix, so the above is a scaled approximation to $\sigma(\underline{x}_n)$.

The second general approach we consider is based on the factorisation in (4). We can re-express (4) as :

$$\sigma(\underline{x}_n) = \underline{x}_n^t\, \Gamma\, \Gamma^t\, \underline{x}_n = \sum_{k=1}^{L} \frac{|\, \underline{v}_k^t\, \underline{x}_n\, |^2}{\lambda_k}$$

where the $\underline{v}_k$'s are the eigenvectors of R and the λ_k are the corresponding eigenvalues. The problem normally is that this eigen decomposition is difficult to evaluate. However a result presented by Gray [11] allows us to approximate this formulation. This result can be paraphrased as saying that the eigenvectors of an auto-correlation matrix tend to the vectors

$$\underline{v}_k = [\, 1 \;\; e^{2\pi ik/L} \;\; e^{4\pi ik/L} \;\; \ldots \;\; e^{2\pi ik(L-1)/L} \,]^t$$

as $L \to \infty$. The inner products of the above vector with the data vector gives the k^{th} Discrete Fourier coefficient, $X(k,n)$. of the n^{th} data window. The corresponding approximation to the eigenvalues is

$$\lambda_k \approx E[\, |X(k,n)|^2 \,]$$

Thus another approximation to our test statistic is

$$\sigma(\Sigma_n) \approx \sum_{k=1}^{L/2} \frac{|X(k,n)|^2}{E[\,|X(k,n)|^2\,]} \qquad (7)$$

The quantity $|X(k,n)|^2$ represents the spectrogram of the data, evaluated using no windowing. The divisors, $E[\,|X(k,n)|^2\,]$, in the above can be estimated by forming averages of the spectrogram. From this stand point the above expression can be thought of as representing sums of the normalised spectrograms along lines of constant frequency. In many applications normalising the spectrogram is a standard procedure, justified on heuristic grounds, yet here we present a theoretical reason for employing such a procedure.

The exact choice of averaging procedure employed to form estimates of $E[\,|X(k,n)|^2\,]$ is not a matter for discussion within this paper. We shall use the following averaging scheme, mainly because of its computational simplicity when employed on a real time system.

$$E[\,|X(k,n)|^2\,] \approx (1-\lambda) \sum_{m=0}^{n} \lambda^{n-m} |X(k,n)|^2$$

4 DETECTION THRESHOLDS

As we have seen the test statistic is a chi-squared random variable with L degrees of freedom when the input sequence is a Gaussian process. Using this we can set the detection thresholds for any of the above methods. Such thresholds can be obtained directly from tabulated values of the area under the chi-squared curve. An alternative approach is to use a third order approximation to the chi-squared statistic, based on the normal distribution [12]. Specifically if Z_α is the 'Z value' for the normal distribution which leaves an area of α units under the normal p.d.f. Then the corresponding point on the chi-squared curve with L degrees of freedom is given by

$$L\left\{1 - 2/(9L) + Z_\alpha\sqrt{2/(9L)}\right\}^3$$

5 THE DATA

In order to test the performance of these methods a selection of measurements of underwater acoustic transient signals were acquired. These measured transients were recorded at a relatively high SNR. Besides the measured transients, a time history of oceanic background noise was also obtained. The transients were then scaled and added to the background noise. A scale factor was determined for each transient which rendered the transient 'barely audible' in the background noise to generate signals at various SNRs. These scales factors are clearly subjective and do not represent absolute values, they do however permit comparison of the methods across the set of transients in an intuitively appealing fashion. We denote this subjective SNR as $0\ dB_{sub}$. Two examples of time histories for the transients studied are shown in Figure 2.

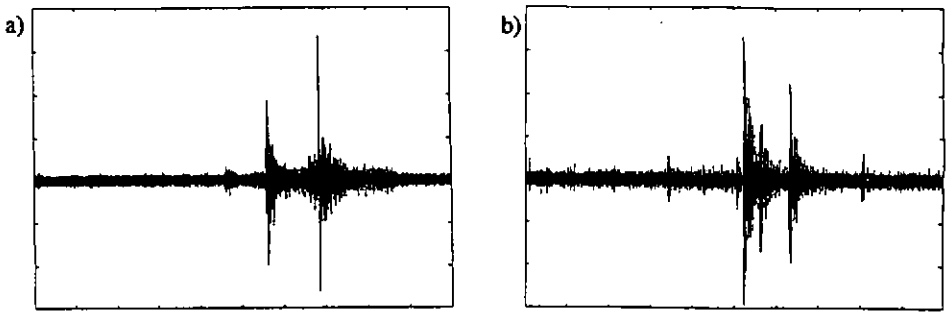

Figure 2 : Time Histories of Two Example Transients

6 RESULTS

We firstly demonstrate that the methods summarised by equations *(5)*, *(6)* and *(7)* produce broadly the same results. Figure 3 shows the test statistic, calculated for the transient depicted in Figure 2a) immersed in the background noise at *6* dB_{sub}, calculated using each of the three techniques described herein. The threshold level shown is set using a value of Z_α equal to *6* and the normal approximation to the chi-squared has been employed. In this example each of the three methods clearly detects the transient. From these curves one is unable to easily discern if any one of the methods performs significantly better than either of the other two. We require a comparison technique which is more discriminating.

To obtain a more sensitive measure of which method performs best we proceed by selecting a value of Z_α at which to set the threshold. Then the transients are added to the background noise at various SNRs. The lowest SNR where a transient is detected is then found for each of the methods. Here to reduce the simulation burden we restrict our attention to only the spectral method, as summarised by *(7)*, and the fast QR lattice, as summarised by *(6)*. These threshold levels, measured in dB_{sub} are shown in Table 1 for a selection of transients, numbered from *1* to *6*. For both of the algorithms $\lambda=0.9999$ and $L = 64$.

	Transient Number					
Detection Thresholds for :	1	2	3	4	5	6
Spectral Detector (dB_{sub})	-2.75	-5	-1.25	3	0.75	-8.75
Adaptive Detector (dB_{sub})	-2.5	-4	-1.5	1.25	1.25	-6.75

Table 1 : Detection Thresholds for the Spectral Detector and the QR Adaptive Detector

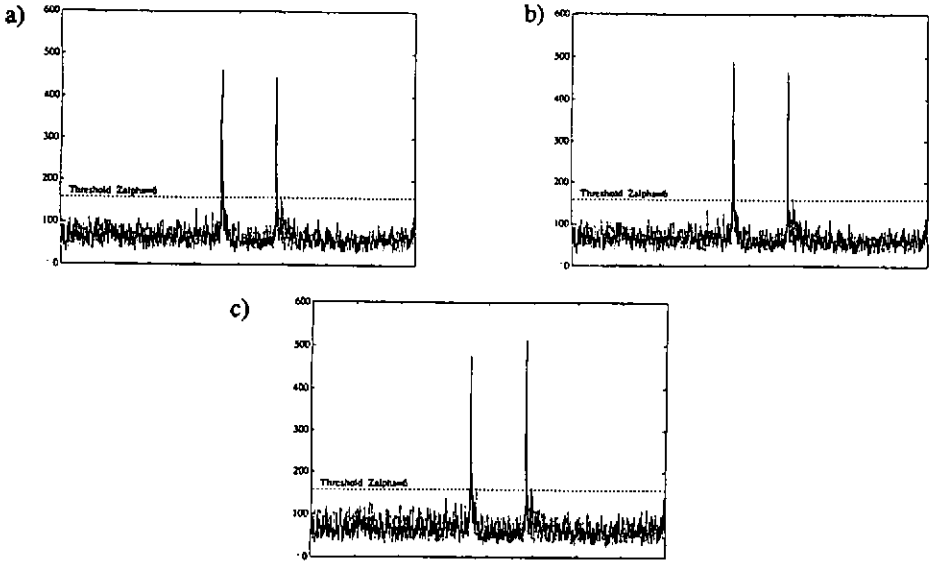

Figure 3 : Test Statistic for the Three Detection Algorithms
a) Spectral Detector *(7)*, (b) Adaptive QR Detector *(6)*, (c) Gradient Adaptive Detector *(5)*

In the above table we see that in general the two methods are successful at detecting transients down to levels where the untrained ear is struggling to detect the signals. The only two examples where the ear performs better that the automatic detector, *i.e.* for transients *4* and *5*, are the two examples where the transient signals are significantly longer in duration than the analysis window. Clearly if the analysis window is not of a suitable length then a degradation in the performance of a given detector is expected.

7 CONCLUSIONS

We have illustrated that transient detection algorithms can be constructed from existing signal processing tools. Three algorithms were presented whose performance often produces accurate detections down to the same level as the untrained human ear and in some examples significantly below that level. These algorithms are relatively simple and can be, and have been, programmed on real-time signal processing chip.

8 ACKNOWLEDGEMENTS

I would like to thank DRA Farnborough (Maritime) for their financial and technical support throughout this work.

9 REFERENCES

[1] *Friedlander, B and Porat, B.*, " Detection of Transients by the Gabor Representation", IEEE Trans. on Acoustics, Speech and Signal Processing, Vol. 37, No. 2, pp 169-180, 1989.

[2] *Hinich, M.J.*, "Detection of Transient Signals by Bispectral Analysis", IEEE Trans. on Acoustics, Speech and Signal Processing, Vol. 38, No. 7, pp 1277-1283, 1990.

[3] *Chen, C.H.*, "On Segmentation Algorithms for Seismic Signal Analysis", Geoexploration, Vol. 23, pp 35-40, 1984.

[4] *Porat, B. and Friedlander, B.*, "Adaptive Detection of Transient Signals", IEEE Trans. on Acoustics, Speech and Signal Processing, Vol. 34, No. 6, pp 485-490, 1986.

[5] *Boashash, B. and O'Shea, P.*, "A Methodology for Detection and Classification of Some Underwater Acoustic Signals using Time-Frequency Analysis Techniques", IEEE Trans. on Acoustics, Speech and Signal Processing, Vol. 38 No. 11, pp 1829-1841, 1990.

[6] *Parsons, A.*, "Bispectral Content of Passive Sonar Signals Part 1", Report No. 92157, DRA Portland, 1992

[7] *Golub, G.H and Van Loan, C.H.*, **Matrix Computations**, Pub. John Hopkins Press, 1989.

[8] *Scharf, L.*, **Statistical Signal Processing : Detection, Estimation and Time Series Analysis**, Pub Addison-Wesley, 1991.

[9] *Honig, M., and Messerschmidt, D.*, **Adaptive Filters: Structures, Algorithms and Applications**, Pub. Kluwer Academic Press, 1985.

[10] *Proudler, I.K. and McWhirter, J.G.*, "QR Decomposition Based Algorithms and Architectures for Least Squares Adaptive Filtering", RSRE Memorandum No. 4562, 1992.

[11] *Gray, R.M.* "On the Asymptotic Eigenvalue Distribution of Toeplitz Matrices", IEEE Trans. on Information Theory, Vol. 18, No. 6, pp 725-730, 1972.

[12] *Abramowitz, M. and Stegun, I.A.*, **Handbook of Mathematical Functions**, Dover Publications, 1970.