# Acoustic Signal Characterisation using multi-resolution transforms.

Paul SEEKINGS, Frederic KERROUX and John POTTER
Acoustic Research Laboratory, Tropical Marine Science Institute, National University of Singapore,
12a Kent Ridge Road, Singapore 119223.
paul@arl.nus.edu.sg

## Abstract.

Most acoustic signals received by underwater systems in shallow waters are non-stationary and corrupted by unpredictable noise sources. In most cases, the noise has a dramatic influence on the performances of these systems. While classical methods often fail to characterise these noises in such an environment, recent multi-resolution methods like the adaptive wavelet transform and its dual, the cosine packet transform, provide a promising alternative. This paper treats the received signal as being made up of four components – tonals, transients, time/frequency transients, and spectrally smooth noise. We introduce an algorithm (ASC) that performs the automated detection and extraction of these four different types of signals. The ASC algorithm has already found applications in the processing of towed array data, humpback whale song and autonomous recorded acoustic datasets collected in Singapore waters.

## Introduction.

The main objective of this paper is to show that standard Wavelet packet and Cosine packet decompositions and de-noising algorithms can be used to separate a time signal into

- Tonals - long frequency-localized signal
- Transients - broadband time-localized signal
- Spectrally smooth noise
- Time-frequency transients

In the results section examples will be given with particular reference to acoustic signals recorded underwater in local Singapore waters.

Every stage has a similar approach that is
- Signal decomposition in an appropriate basis.
- Signal detection.
- Signal extraction and reconstruction.

The system is designed to be modular; hence any of the four types of signals can be extracted independently of others. The overall structure of the modular system is shown in Figure (1).

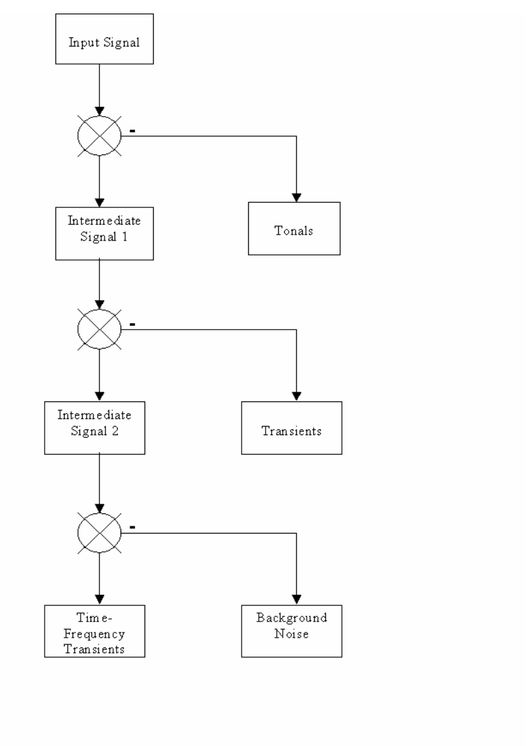



**Figure (1): ASC Schematic.**

## Tonal detection and Extraction.

By definition, a tonal signal is a long frequency-localized signal. To achieve the best representation in the frequency domain, the most appropriate transform is a Cosine Packet Transform (CPT).
Essentially the coefficients of the transform is given by

$$a_{j,k} = \left\langle s(n), y_{j,k} \right\rangle \tag{1}$$

Where s(n) is the time series of the signal. $y_{j,k}$ is the cosine packet scaled (j) and by translated (k) for $j = 0,1,......$ and $k = 0,..,2^j - 1$.

The Cosine packet is a cosine wave multiplied by a smooth envelope function, so is well localised in both time and space, initially it is scaled to be the same length as the signal. On the zero level, j=0, the transform is equivalent to a cosine transform of the whole signal. On the first level, j=1 and k=0, the cosine packet is scaled to half the length of the signal, the transform is equivalent to a cosine transform on the first half of the signal resulting in $a_{1,0}$. When j=1 and k=1, the cosine packet is scaled to half the length of the signal and translated such that the transform is equivalent to a cosine transform on the second half of the signal resulting in $a_{1,1}$. This can be process continued down to finite number of levels, each level of the decomposition contains the complete representation of the signal, but split over different time windows, hence it is a redundant basis representation of the original time signal. For this application we are only concerned with the first level coefficients.
Further details of Cosine packet decompositions and wavelets packet decompositions can be found in Wickerhauser(1) and Mallet(2).

Once the cosine packet transform has been taken of the signal down to first level, tonal detection can be achieved by the comparison of Shannon entropy function, Bouvet(3), of the coefficients.

$$P(a_{j,k}) = -\sum a_{j,k}^{2} \log(a_{j,k}^{2}) \tag{2}$$

If tonals are present, they will be present in both $a_{1,0}$ and $a_{1,1}$, since be definition they should be in the first and second half of the signal. Hence, the sum of the entropies of $a_{1,0}$ and $a_{1,1}$, should exceed the entropy of $a_{0,0}$ since the first level essentially contains 'more information' than the zero level. Hence,

If

$$P(a_{0,0}) \le P(a_{1,0}) + P(a_{1,1}) \tag{3}$$

then tonals are present.

In order to accurately locate tonal components a dot-product is performed between the two level 1-coefficient vectors, and the square root taken to preserve the amplitude of the common frequency components.
A threshold is then applied on the correlation result to detect matching components between the two basis vectors at level 1. The Donoho-Johnstone Estimator (DJE), Delory(3), Donoho(4), is used to estimate the threshold level. This requires the standard deviation of the assumed noise to be known, and estimate can be sought from the zero level of the cosine packet coefficients, Hirsh(5), Stahl(6), thus the threshold t is given by:

$$t = \sqrt{2 \ln N} \, \frac{\mathrm{M}\lfloor a_{0,0} \rfloor}{0.6745} \tag{4}$$

Where $\mathrm{M}[\cdot]$ denote the median, and N in the length of the signal.

The threshold allows splitting of the tonals into T, and the rest of the signal into R.

$$T_{1,i} = \left\langle c_{1,i}, \sqrt{\langle a_{1,0}, a_{1,1} \rangle} > t \right\rangle \Big|_{i=1,2} \tag{5}$$

$$R_{1,i} = \left\langle c_{1,i}, \sqrt{\langle a_{1,0}, a_{1,1} \rangle} \le t \right\rangle \Big|_{i=1,2} \tag{6}$$

T, R are then reconstructed to give the time series, T(n) and R(n).

## Transient detection, and extraction.

To detect transients, the same principle to detect tonals is applied. The only difference is that Wavelet packet decomposition is used. Daubechies Real Biorthogonal Most Selective (DRBMS) wavelets are used with length M=22, to obtain a good compromise between computation-time and efficiency, Delory(7). Wavelets are compactly supported over a time interval but are generally not compactly supported in frequency. They are the time/frequency dual of the Cosine packet decomposition. Translation of the wavelet function corresponds to a shift in frequency band (rather than time). Scaling of the wavelet function corresponds to a decrease in time resolution (rather than frequency).
The wavelet packet decomposition (to level 1) of the signal is calculated. A true time transient that is broadband in frequency should show up in both halves of the level 1 coefficients. The entropy test is used to check for the presence of transients, the threshold is set using the same scheme as in the tonal extraction; the coefficients are split and reconstructed to yield the transient time series and the remaining times series.

## Signal and Noise extraction.

The last stage of our algorithm is to split background noise and time-frequency transients. Essentially, the methodology is to use a 'standard' de-noising algorithm, with the slight change so that both the de-noised time series and the noise time series are available.
De-noising algorithms work by taking a 5 level wavelet decomposition:

$$a_{j,k} = \left\langle s(n), y_{j,k} \right\rangle \tag{7}$$

Where $k = 0,..,2^j - 1$. $j = 0,1,....,5$ and $y_{j,k}$ is mother wavelet function.

The entropy is calculated for each packet giving:

$$P_{j,k} = -\sum a_{j,k}^{\ 2} \log(a_{j,k}^{\ 2}) \tag{8}$$

The best basis search is performed using $P_{j,k}$, as the information cost function. The end result is basis set that has maximum compaction i.e. the a few large components and numerous smaller components. The general assumption is that the larger components represent the time/frequency transients, and the smaller components that are spread through out the best basis represents the spectrally smooth background noise. Thresholding can split the spectrally smooth noise and time/frequency transients, and reconstruction the two results time series for both components. Normally noise is assumed to be Gaussian, however in underwater applications noise does not have equal energy over all frequencies, thus a threshold is calculated for each packet in the bestbasis, effectively this sets the optimium threshold for each time/frequency band.

The threshold for each selected basis is given by:

$$t_{\hat{j},\hat{k}} = \sqrt{2 \ln N} \frac{\mathrm{M}\left[a_{\hat{j},\hat{k}}\right]}{0.6745} \tag{9}$$

where $\mathrm{M}[\cdot]$ denote the median, N in the length of the signal, $\hat{j},\hat{k}$ are the indices of the packets selected as the best basis.

Thresholding allows separation of the coefficients that represent the time/frequency transients $\partial_{\hat{j},\hat{k}}$ and the coefficients that represent the noise $h_{\hat{j},\hat{k}}$,

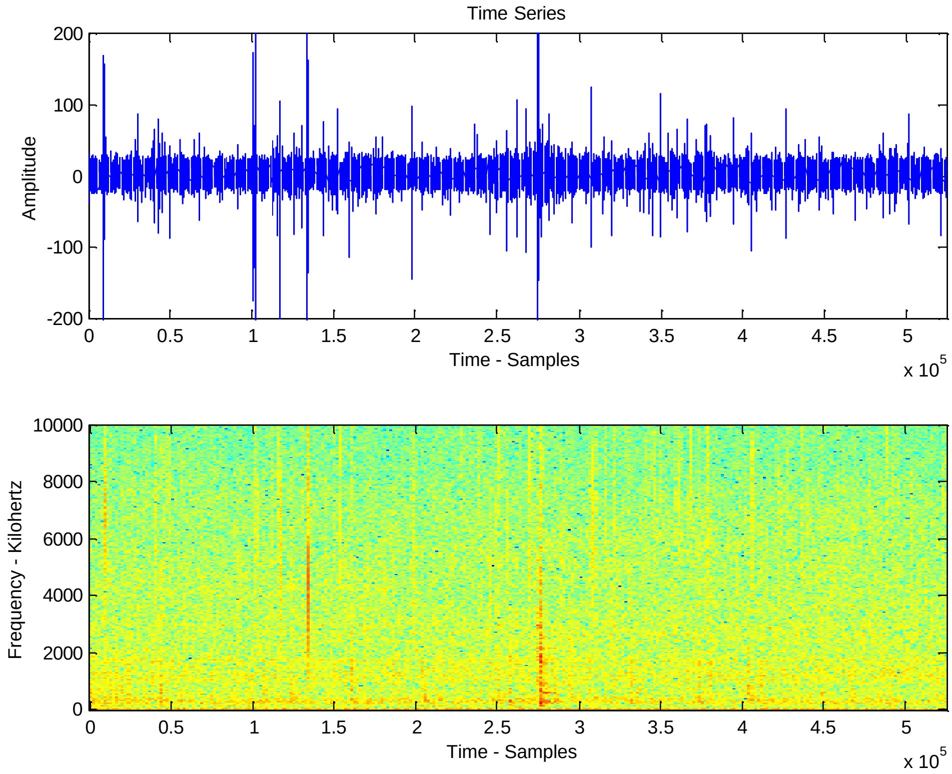
$$\partial_{\hat{j},\hat{k}} = \left\langle a_{\hat{j},\hat{k}}, \left| a_{\hat{j},\hat{k}} \right| > t_{\hat{j},\hat{k}} \right\rangle \tag{10}$$

$$h_{\hat{j},\hat{k}} = \left\langle a_{\hat{j},\hat{k}}, \left| a_{\hat{j},\hat{k}} \right| \le t_{\hat{j},\hat{k}} \right\rangle \tag{11}$$

$\partial_{\hat{j},\hat{k}}$ and $h_{\hat{j},\hat{k}}$ are then reconstructed to give the time series.

# Results

In order to demonstrate the algorithms we look use a section of a time series sampled at 20 kHz by an underwater data recorder in Singaporean shallow waters. The time series and spectrogram is shown in figure(2) below. Careful observation should reveal that the time series had tonals, and transients. A large time/frequency transient can be seen around the middle of the time series, and several low energy time/frequency transients throughout the time series.

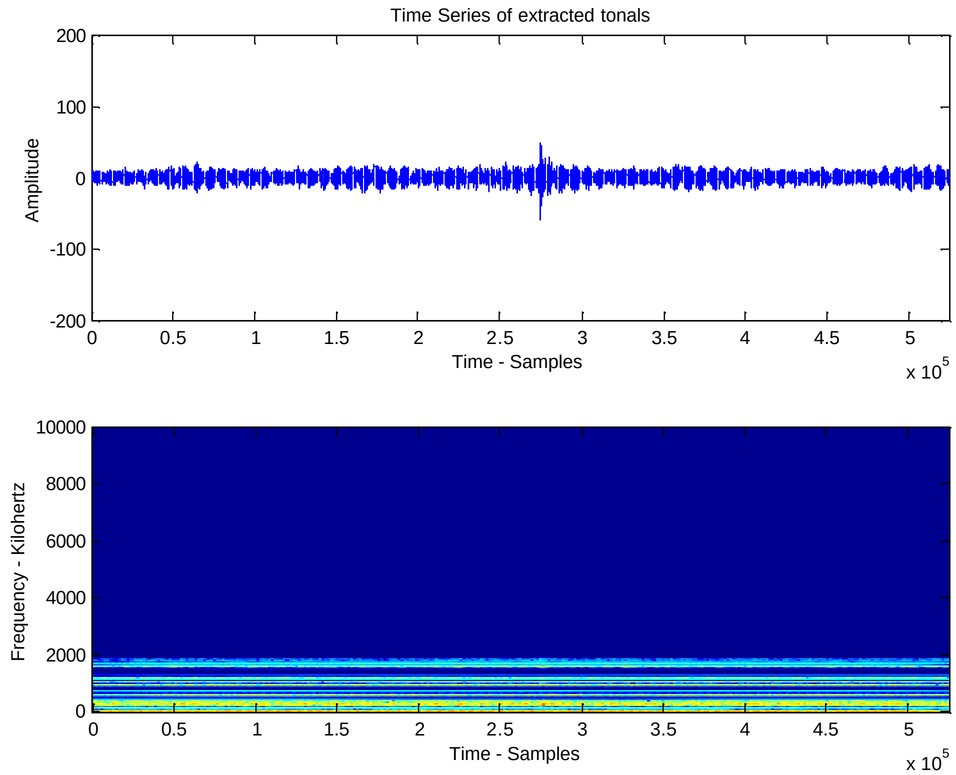



**Figure(2), The original time series.**

The example follows through the whole procedure where each signal type is extracted from the original signal.

## Results: Tonals

In figure(3) , the time series and spectrogram of extracted tonals can be seen.



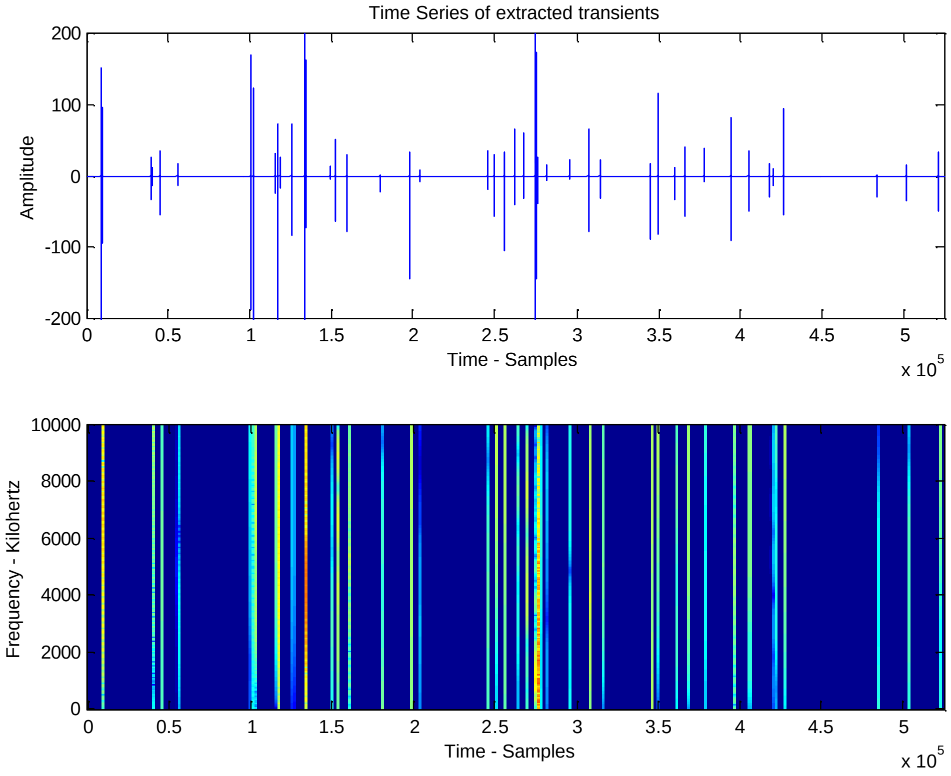

**Figure(3) The time series and spectrogram of extracted tonals.**

The algorithm has successfully extracted the tonals that can be seen in the spectrogram of the original time series. According to the area in which the recording was taken, we can assume that tonals are from different sources within the various frequency bands:

- 0-200 Hz => Shipping noise
- 500 Hz – 3 kHz => reclamation noise, Potter(8).
- Others => Electrical and/or mechanical noise.

## Results: Transients

In figure(4) , the time series and spectrogram of extracted transients can be seen

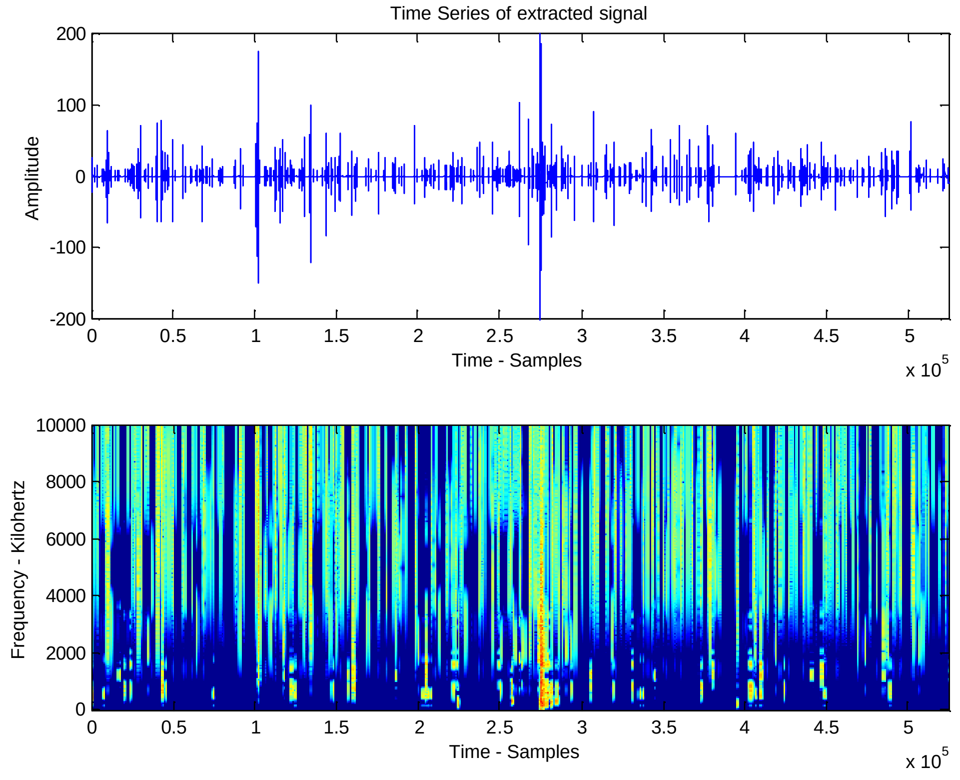



**Figure(4) The time series and spectrogram of extracted transients.**

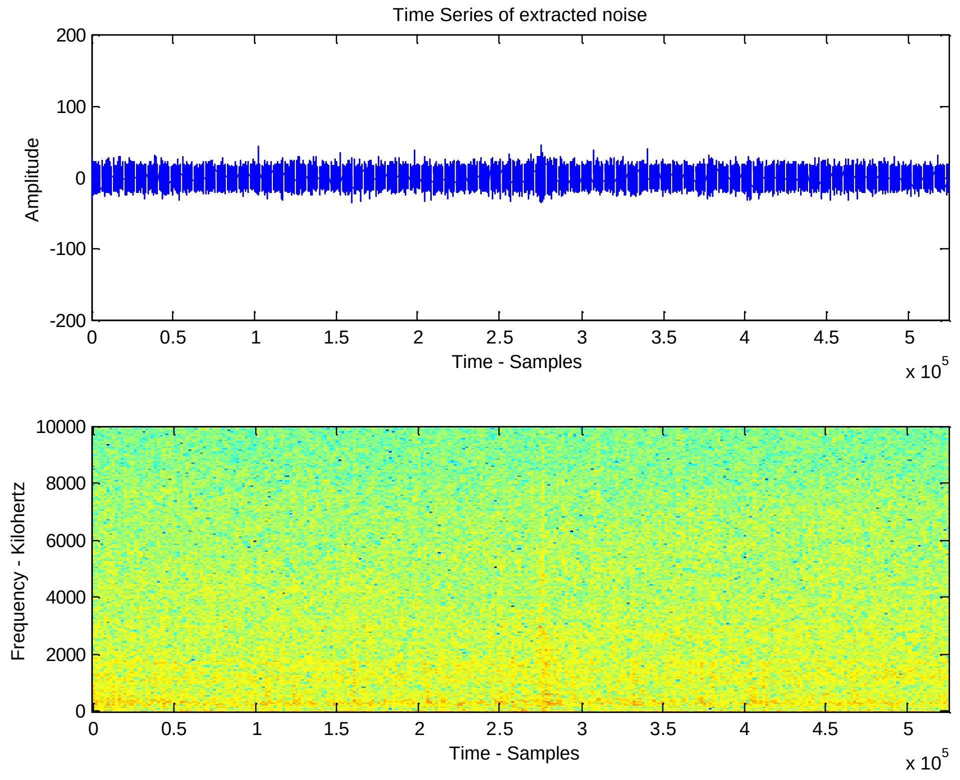
Figure(4), shows that most of highest components (i.e. “snaps” of snapping shrimps) have been detected. Furthermore, there is not any ambiguity between transients and time-frequency transients, on spectrogram, in spite of its low time resolution.
These results can be explained by the highly transients features of snapping shrimp noise, which is very short in time (83 µs) and broadband (1-200 kHz), Lim(9).
A possible method to improve the result should be to create a new wavelet using a snapping shrimp noise model,Versluis(10).

## Results: Time/Frequency transients and noise separation.

Figures (5 &6) show the result of the Time/Frequency transients and noise separation.




**Figure(5) The time series and spectrogram of the extracted Time/Frequency transients.**




**Figure(6) The time series and spectrogram of the extracted noise.**

The algorithm has successfully extracted the large energy time/frequency transient in the middle of the time series and has extracted the smaller energy time/frequency transients throughout the time series.

At this stage the level of the threshold will determine the relative split between the two output signals, and although the threshold is set automatically, manually adjusting the threshold away from that, is useful if a particular time/frequency transient is of interest. The time/frequency transient single to noise ratio limits the effectiveness of this procedure.

## Conclusion

The objective of ASC project was to develop an algorithm to split an input time-series into four signal components:

- Tonals - long frequency-localized signal
- Transients - broadband time-localized signal
- Spectrally smooth noise
- Time-frequency transients

Although this project was motivated by the need to analyse underwater acoustic signals recorded in the Singapore shallow waters, ASC has been developed with a non-application specific methodology thus it can be applied in a wide range of applications. A Matlab toolbox has been developed using ASC principle to analyse different signals as outlined in this paper. This toolbox contains functions to run each stage of ASC individually, and also to offset the automatic threshold to suit specialised applications.

**References**

[1] Wickerhauser,M.L, 'Comparison of picture compression methods: Wavelet, Wavelet packet and Local Cosine Transform coding', Dept. of Maths, Washington University.

[2] S.Mallat "A wavelet tour of signal processing ", Academic Press, 1998

[3] M. Bouvet, "Traitements des signaux pour les systemes sonar" Masson, 1992

[3] E. Delory & John R. Potter, "Signal Processing Aspect of Signal Detection, Masking and Noise Suppression", Acoustic Research Laboratory, National University of Singapore.

[4] David L. Donoho & Iain M. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage", Dpt of Statistics, Stanford University, 1994

[5] H. G. Hirsch & C. Ehrlicher, "Noise estimation techniques for robust speech recognition", Proc. ICASSP, volume1, pages 153-156, 1995

[6] V. Stahl, A. Fischer & R. Bippus, "Quantile based noise estimation for spectral substraction and wiener filtering ", Proc. ICASSP, volume 3, pages 1875-1878, 2000

[7] E. Delory & John R. Potter, "Underwater Background Noise Filtering", ARL website, 1999.

[8] John R. Potter, Lim Tze Wei & Mandar Chitre, "Acoustic Imaging & the Natural Soundscape in Singapore Waters", ARL paper.

[9] Lim Tze Wei, 'Measurement and analysis of marine ambient noise in local waters' MSc thesis, National University of Singapore.

[10] M. Versluis, B. Schmitz, A. Von der Heydt, D. Lohse, "How snapping shrimp snap: Through cavitating bubbles", Science Online (available at www.sciencemag.org), 2000.