SIGNAL PROCESSING ASPECTS OF SIGNAL DETECTION MASKING AND NOISE SUPPRESSION

ERIC DELORY & JOHN POTTER

Acoustic Research Laboratory
Tropical Marine Science Institute & EE Dept, National University of Singapore

ABSTRACT

Acoustic signal extraction and identification in the underwater environment is best achieved by adaptive methods as the signals encountered are generally non-stationary and corrupted by unpredictable noise sources, such as man-made noise, biological and seismic noises. While classical methods often fail in such an environment, the recent use of multiresolution methods like the adaptive wavelet transform and its dual, the cosine packet transform, provides a promising alternative. This paper introduces three applications where wavelet and cosine transforms are used for denoising and signal identification. Signal decomposition on these two sets of functions provide different representations, which are each specific to a certain noise type. The third application achieves background pink noise filtering along with a high signal compression rate, which can be used to optimize signal identification.

1 INTRODUCTION

In most underwater acoustic applications, the objective is to extract some information $s(t)$ out of a buffered signal $x(t)$, corrupted with noise $n(t)$ which is often assumed additive:

$$x(t) = s(t) + n(t)$$

Depending on the noise sources involved and the signal-to-noise ratio (SNR), the useful information the signal carries may or may not be entirely recovered. The signal extraction process is generally started with the input of some *a priori* knowledge on $s(t)$ and $n(t)$: e.g. in active sonar, $s(t)$ is an echo of a previously sent signal, which makes extraction easier, and achievable via the use of classical methods. Setting a hypothesis on background noise, such as assuming second-order stationarity, is generally fallacious as most underwater noise sources are non-stationary. A typical example in shallow waters is snapping shrimp noise, which can be viewed as a randomly time-distributed sequence of broadband transients.

However, in the general case, it is unlikely that no *a priori* information at all is known about signal and noise. According to geographical areas, certain types of noise can be expected: in high latitudes, low frequency ice breaking noise (localized source) is common and propagates on long distances, while in low latitude shallow waters, broadband time-localized snapping shrimp noise (multi-source) is ubiquitous. Another example is shipping noise in straits or near industrial harbors, which can mask any signal in the same frequency range. The difficulty here is then to use a known noise signature and to take into account its variability in time and frequency domains to be able to filter it out. However, in most cases, the computation complexity involved may be too high if all parameters need be implemented.

Besides, benchmarking an underwater acoustic signal extraction algorithm is not an easy task, considering the noise variety; white Gaussian noise is often used to compare algorithm efficiency, but it is likely that the algorithm performance in real applications will drop dramatically. In this paper, signal extraction is performed against three types of noise: pink noise, snapping shrimp noise, and shipping noise, which can respectively be classified in broadband colored noise, short time-localized transient noise, and long frequency-localized noise — where the adjectives short and long are obviously relative to the duration of $s(t)$. In this work, $s(t)$ can be any signal that does not correspond to the three predefined noise classes.

Prior noise signature detection is achieved using four different representations of $x(t)$, which is projected on four basis chracterized by a specific time-frequency resolution:
- basis W_0, which provides the highest time resolution σ_{tw0}, is trivial as no projection is performed: the sampled signal is stored on this basis during acquisition. To respect the terminology used in the following, $x(t)$ will be called x_{w0} when comparisons between different representations of the signal are needed.
- basis W_1 achieves a wavelet half-band coding (or level 1 wavelet packet decomposition), time-resolution is halved:
$\sigma_{tw1} = \sigma_{tw0}/2$
The projection produces two vectors: x_{w10} (low frequency band) and x_{w11} (high frequency band)
- basis C_0 performs a cosine transform of $x(t)$ (or level 0 cosine packet decomposition), and gives the highest frequency resolution $\sigma_{\omega c0}$
The projection produces one vector x_{c0}
- basis C_1 performs a cosine transform of the first and second halves of $x(t)$ (or level 1 cosine packet decomposition), frequency-resolution is halved:
$\sigma_{\omega c1} = \sigma_{\omega c0}/2$
The projection produces two vectors: x_{c10} (spectral representation of the first bin) and x_{c11} (for the second bin)

An information cost $I(x(t))$ is calculated on these four representations. Comparison of costs is then used for noise detection. If positive, a simple dot-product highlights high energy components generated by time-localized noise (here snapping shrimp) and/or frequency localized noise (shipping noise); these components in the respective bases are then set to zero and won't be selected in later compression.

Pink noise filtering is rarely dealt with in denoising literature although it is the most common background noise encountered in a natural environment. It is often approximated by an $1/f^n$ (with $n \geq 1$)decreasing power spectral density, which respects the fact that in a natural environment, high frequencies are more attenuated than low frequencies. Here, pink noise filtering is subsequently achieved by compression: $x(t)$ is decomposed into a set of wavelet-packet bases and a best basis is chosen according to the same information cost function used for the preliminary denoising part: the best basis is used to compact the signal energy in a low number of coefficients. This results in a few high coefficients containing only high energy time-frequency localized portions of $x(t)$, and a large number of small coefficients which represent the broadband pink noise. A simple threshold on the decomposition discards most of the noise components and achieves a high compression rate for $s(t)$, which facilitates the classification task.

2 SIGNAL DECOMPOSITION

Wavelets are increasingly used in signal processing as an alternative to the spectrogram, which results in a time-frequency plane representation of a signal, but with the drawback of a fixed predefined time-frequency resolution. The Heisenberg uncertainty principle (first stated in quantum mechanics, and proved in [1]) asserts the following:

$$\sigma_t \cdot \sigma_\omega \geq \frac{1}{4}$$

where σ_t and σ_ω are the variances of the energy distribution in time and frequency domains. A direct consequence of this low bound is the impossibility to get both high time and frequency resolutions in the same transform. The wavelet packet transform [2] and the cosine packet transform both provide a set of different bases with varying time-frequency resolutions.

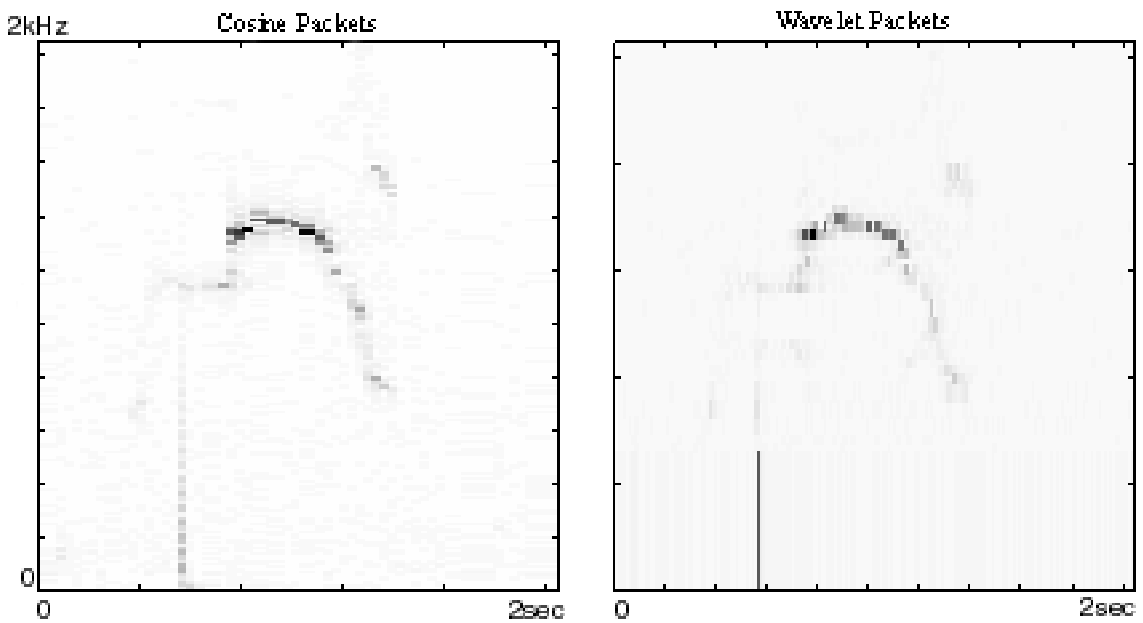

Figure 1: Cosine and wavelet packet transforms (CPT and WPT) of a humpback whale vocalization (2sec,0-2kHz), with selection of best basis. Dynamic range is 25 dB. For both transforms, a basis is chosen, which minimizes an information cost function. The resulting basis improves the compaction rate and consequently concentrates high energy time-frequency patterns in a few coefficients. These graphs show the time-frequency distribution of the resulting coefficients for each transform. As expected, CPT and WPT behave differently, especially in the face of transients and tonals. In this example, a low frequency transient is perfectly matched with the WPT best basis, resulting in only one coefficient, whereas CPT best-basis needs more than 20.

Figure (1) displays the cosine and wavelet packet transforms of a humpback whale vocalization in a noisy environment, duration is approximately 2 seconds and Nyquist frequency is 2 kHz. The time-frequency plane tiling reflects the behavior of both methods: The cosine-packet transform (CPT) decomposes $x(t)$ in time bins where the frequency spectrum shape is estimated, while the wavelet packet transform (WPT) decomposes $x(t)$ in frequency bins where the temporal shape is estimated. Perfect reconstruction can be

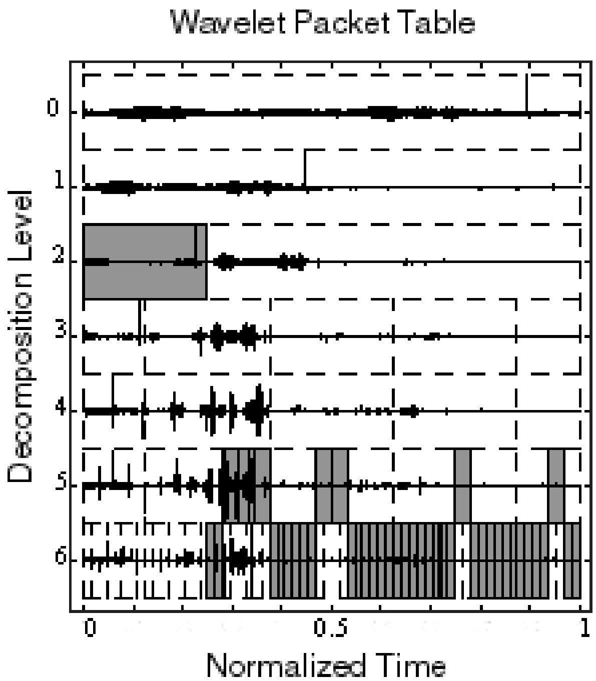

Figure 2: Best-basis selection (shaded blocks) selected by minimizing the pseudo-entropy function. The WPT decomposition is dyadic, and increasing the level by one involves subsampling by 2, so that each level has the same number of coefficients. Each delimited block "*can be seen*" as the time series that the decomposed signal carries in the corresponding frequency band. Frequency resolution is doubled and time resolution is halved each time the level is incremented.

processed with the respective inverse transforms as CPT and WPT are orthogonal projectors. In the WPT case the chosen wavelet must either be orthogonal or bi-orthogonal.

Although the forward orthogonal wavelet projection generates coefficient vectors of same Euclidian norm as the signal, it is not phase linear. On the contrary, bi-orthogonal wavelets provide phase linearity, but at the price of energy distortion: the Euclidian norm of a vector obtained from a bi-orthogonal wavelet decomposition does not equal the input signal energy, the reconstructed signal does. The wavelet pair used in this work is the 22-coefficients Daubechies Real Biorthogonal Most Selective (22-DRBMS). The reasons of this choice are beyond the scope of this paper, details can be found in [3, 4].

A Packet Transform (PT) decomposition until a given level L creates a set of 2^L basis, among which one can be selected according to a given criterion. A widely used criterion is the result of an information cost function applied to all basis. The information cost function used in this work is related to the Shannon entropy function [5] and is often called the pseudo-entropy $Ic(s)$:

$$Ic(s) = -\sum_j s_j^2 log(s_j^2)$$

where s is a signal of length N, and $1 \leq j \leq N$

As this function is additive, a translation invariant wavelet-packet algorithm is used, in order to get a consistent representation of the signal time-shifted versions[6].

The basis with minimum entropy (i.e., roughly, the one that allows the highest level of compaction) is the (best) one used for decomposition (figure(2)).

3 FIRST-STAGE DENOISING: EXAMPLES

3.1 Shipping noise

The more appropriate PT for shipping noise filtering is the frequency domain one, CPT. The number of decomposition levels M can be high if computation time is not a concern, or low e.g. in real-time applications, the computational cost being M-linear. If the signal size is N, a first level CPT decomposition will provide two basis: one with the poorest time resolution ($\sigma_t = N/2$) and the highest frequency resolution ($\sigma_\omega = 1$) (resolutions are scaled here for readability, formal definitions can be found in [7, 8]); in the second basis, time-resolution is doubled, whereas frequency-resolution is halved. A humpback whale vocalization corrupted with low frequency shipping noise is decomposed in figure (3), where the second basis is displayed in shaded blocks. The noise source comes from an underwater recording of shipping activity.

Here shipping noise is defined as a stationary tonal sound. Detection in the scope of CPT can be performed efficiently with an information cost comparison between level 0 ($\sigma_\omega = N/2$) , and level 1 where $\sigma_\omega = N/4$. If the pseudo-entropy is lower at level 0, that means frequency coding of the signal energy is alike in both time-bins at level 1. From a signal processing point of view, it means that a stationary tonal sound of relatively high energy is present. As such a signal is classified as noise, a more accurate detection phase is processed to locate it on the frequency scale and filter it out.: a dot-product $Corr_0$ (cross-correlation calculated at $\tau = 0$) is performed between the two level 1 coefficient vectors, to enhance matching frequency components and to reduce others. A threshold is then applied on $Corr_0$ to detect matching components between the two basis vectors at level 1. Detected components are set to zero and the denoised signal is built with the inverse transform applied on level 1. In short, the algorithm is:

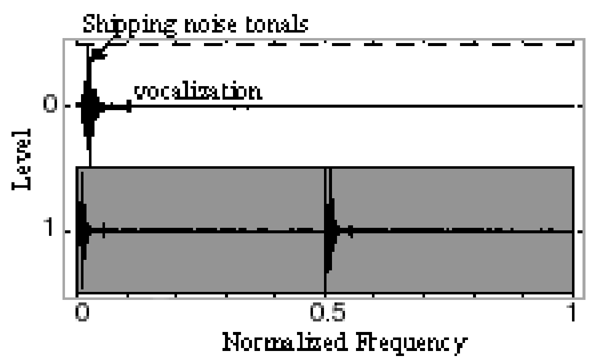

Figure 3: A -16dB SNR vocalization corrupted with boat noise, decomposed on a one level CPT. The vocalization is hidden by the high energy tones produced by shipping activity. Level one shows that these tonals appear in both time bins. Cross-correlation between these two bins is a way to localize the corrupted frequency bands and to extract them.

if $Ic(x_\zeta 0) < Ic(x_{\zeta 10}) + Ic(x_{\zeta 11})$
do $\{Corr_0 = |x_{\zeta 10}| \cdot |x_{\zeta 11}|$
for $(f = 0; f < N/2; inc(i))$
do{ if $Corr_0(f) > threshold$
$x_{\zeta 10}(f) = 0; x_{\zeta 11}(f) = 0;\}$
}

where ζ is c for cosine and w for wavelet transform, details on notations are provided in the introduction.

3.2 Snapping shrimp noise

Snapping shrimp noise is a broadband transient noise type. To detect and filter it a high temporal resolution is needed. WPT provides a high time resolution on the first decomposition level. A parallel can be observed with the former example, as snapping shrimp noise signature in the low-level wavelet space and shipping noise signature in the low-level cosine space are alike. The same paradigm as for shipping noise detection and thresholding is used.

3.3 Compression & noise filtering

The previous time or frequency localized noise filtering system is an intermediate step before the compression algorithm, which concentrates the signal energy in a few high coefficients and does not retain the low coefficients, e.g. the ones produced by pink noise. Therefore, the compression by-product used in this paradigm is that broadband stationary noise sources can be filtered out: even for low SNR, as noise energy is spread out on the whole time-frequency (or time-scale) plane, their components on the selected best-basis are low, and discarded when a threshold is applied.
In the former denoising methods, only two basis were used to detect noise sources, as time or frequency resolutions needed were obvious: high time resolution for transients, high frequency resolution for tones. In order to extract the time-frequency localized information via compression, the variety of time-frequency resolutions must be broadened, so that a basis that best matches $s(t)$ can be found. To broaden the searching range for best-basis selection, the number of decomposition levels is increased. Decomposition is stopped at a predefined level. The criteria to choose the limit are numerous, and depend on the application, the filter type (in the WPT case),and the method used (CPT or WPT). As the number of basis available increases exponentially with the level number, it is, in the general case, useless to go deeper than the 6^{th} or 7^{th} level.

Best-basis selection for WPT and CPT are presented and tested with different information cost functions and other interesting variations in [9]. Here, the method used is a shift-invariant WPT [6, 10], with a 6th level decomposition and the additive pseudo-entropy information cost function $I_c(s)$.

For the experiment, the signal is another humpback whale vocalization (frequency modulated type), corrupted with both shipping and snapping shrimp noise of equal energy. No pink noise was added as the shipping and snapping shrimp noise data were already highly corrupted with colored background noise. The global SNR is below -10dB.

The denoising experiment was processed in the following way:

1. CPT one level decomposition and frequency localized noise filtering
2. Signal reconstruction by inverse CPT
3. WPT one level decomposition and broadband transient noise filtering
4. WPT 2-to-6 level decomposition and best basis search
5. Compression

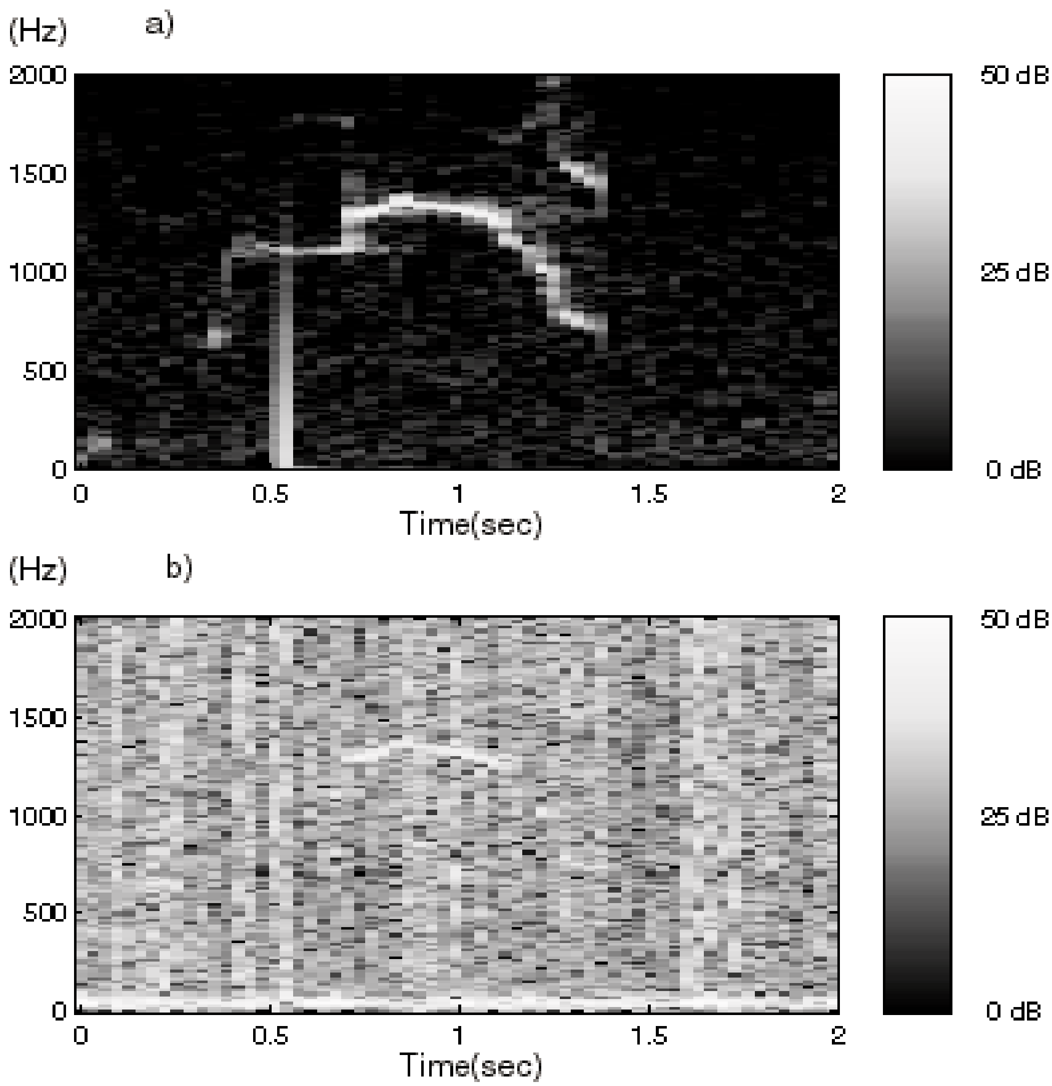

Figure 4: a) Humpback whale vocalization, b) the same vocalization corrupted with snapping shrimp, shipping noise and pink noise: SNR = -15dB

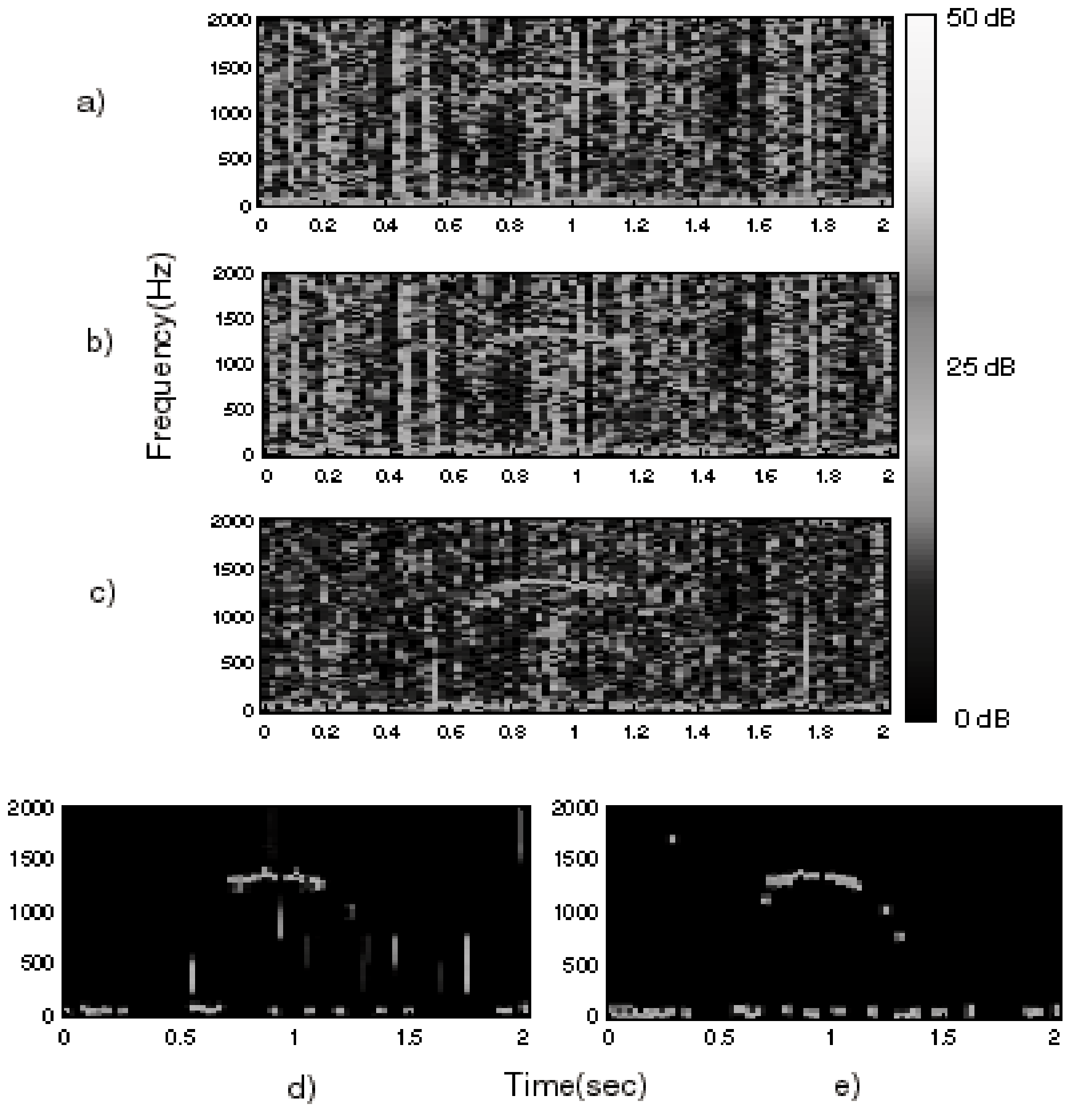

Figure 5: a) vocalization corrupted with snapping shrimp, shipping and pink noise. The three noises are of approximately equal energy, producing an SNR below -10dB. b) Shipping noise filtering is performed by extraction of all level1 cosine-packet coefficient producing $Corr_0 > 5$ c) Snapping shrimp noise filtering is performed by extraction of all level1 wavelet-packet coefficients producing $Corr_0 > 1$ d) Pink noise filtering by retaining 50 coefficients of the best-basis wavelet packet decomposition e) Pink noise filtering by retaining 50 coefficients of the best-basis cosine-packet decomposition.

Figure(4) shows the original signal and the corrupted version, with SNR = -15 dB, used for the numerical experiment. Figure(5) gives the intermediate results obtained after the two pre-filtering methods and the compression stage. Bottom images in Figure(5) show that wavelets and cosine transforms both provide efficient results for compression. Consequently, the choice of the method used for background noise filtering will depend on the feature extraction method for later signal classification. An example of the use of wavelets is given in [4].

4 CONCLUSION

New methods like WPT and CPT, which are part of a wider family called the block-transforms, prove efficient in environments where little is known on noise sources. They are adaptive and provide a large set of basis where signals can be optimally represented. Compacting the information in a few high coefficients facilitates the thresholding task and improves the signal extraction process. In this paper, no strict hypothesis on the signal was imposed before its extraction and the algorithm performance would certainly increase when the signal signature is known. Besides, it is important to notice that the algorithm was tested with real data, as the vocalization, snapping shrimp and shipping noise, as well as background noise, were not simulated.

An important feature that this denoising algorithm provides, is a high compression rate. Statistical or neuro-mimetic classifiers generally need a large training set of a size that must increase exponentially with the input size for good class discrimination. Being able to compress the information in a small set of coefficients is then relevant.

5 ACKNOWLEDGEMENTS

Noise data was recorded by Tze Wei Lim from PORL (National University of Singapore) with the underwater recording system PANDA. Humpback whale vocalizations were provided by the Cornell Bioacoustics Research Program.

References

[1] H. Weyl, *the theory of groups and quantum mechanics.* dutton, 1931.

[2] R. Coifman, Y. Meyer, and M. Wickerhauser, "Wavelet analysis and signal processing," in *Wavelets and their applications* (Jones and Bartlett, ed.), pp. 153–178, 1992.

[3] C. Taswell, "the systematized collection of wavelet filters computable by spectral factorization of the daubechies polynomial," tech. rep., Computational Toolsmiths, 1997.

[4] J. Potter and E. Delory, "a wavelet-packet transform as a preprocessor to an som-lvq neural network for humpback whale song identification." submitted to J. Acoust. Soc. Am., 1998.

[5] R. Hamming, *Coding and Information Theory.* Prentice Hall, 1980.

[6] J. Weiss, "translation invariance and the wavelet transform," tech. rep., Aware, Inc, 1994.

[7] S. Mallat, *a wavelet tour of signal processing.* Academic Press, 1998.

[8] G. Strang and N. Truong, *wavelets and filter bamks.* Wellesley-Cambridge Press, 1998 ed., 1997.

[9] C. Taswell, "satisficing search algorithms for selecting near-best basis in adaptive tree-strucured wavelet transforms," *IEEE Trans. on Sig. Proc.*, vol. 44, no. 10, pp. 2423–2438, 1996.

[10] J. Del Marco, S.and Weiss and K. Jagler, "Wavepacket-based transient signal detector using a translation invariant wavelet transform," in *Wavelet Applications*, no. 2242 in SPIE Proceedings.