LIDS-P-2120

WAVELET PACKET BASED TRANSIENT SIGNAL CLASSIFICATION *

Rachel E. Learned *William C. Karl* *Alan S. Willsky*

Massachusetts Institute of Technology
Room 35-439
Cambridge, Massachusetts, 02139
Email: learned@lids.mit.edu

ABSTRACT

Non-stationary signals are not well suited for detection and classification by traditional Fourier methods. An alternate means of analysis needs to be employed so that valuable time-frequency information is not lost. The wavelet packet transform [1] is one such time-frequency analysis tool. This paper summarizes efforts [2] which examine the feasibility of applying the wavelet packet transform to automatic transient signal classification through the development of a classification algorithm for biologically generated underwater acoustic signals in ocean noise. The formulation of a wavelet packet based feature set specific to the classification of snapping shrimp and whale clicks is given.

1 INTRODUCTION

Over the last decade much work has been done in applying time-frequency transforms to the problem of signal representation and classification. Mallat's work on the application of wavelets to image representation [3] and Daubechies's work on the development of smooth orthonormal wavelet basis functions with compact support [4] sparked a great deal of interest in wavelets in the engineering community. Most recently, the emergence of wavelet theory has motivated a considerable amount of research in transient and non-stationary signal analysis.

This paper discusses the use of the wavelet packet transform in the detection and classification of transient signals in background noise. Our approach focuses on the exploitation of class-specific differences obtained through careful examination of the feature separation attainable from the wavelet packet decomposition of the transients. The Charles Stark Draper Laboratory and the Naval Underwater Systems Center furnished an extensive collection of acoustic signals in background noise which allowed for an empirical study of some typical occurrences of snapping shrimp and whale clicks. A wavelet packet based feature set specific to the classification of snapping shrimp and whale clicks is formulated.

1.1 Motivation

The ability to classify underwater acoustic signals is of great importance to the Navy. Today, detection and classification, tailored for stationary signals, is done by the sonar officer who listens to incoming signals and determines their origins with the aid of a frequency display and look-up tables. Transient signals, lasting only a fraction of a second, are of particular concern because they will typically appear as broad band energy on the frequency display, thus, the sonar officer must be able to detect and classify these signals after only listening to them. These brief signals, such as the single acoustic transmission due to the closing of a door within a ship, may be missed by the sonar officer. Success or failure in the classification of transient signals using traditional methods relies solely on the officer's ability to detect and classify a signal after hearing it only once. An automatic method of classification for transient signals would greatly aid in the detection/classification process.

1.2 Shortcomings of Fourier Methods

Transient signals are not well matched to standard spectral analysis methods. In particular, Fourier-based methods are ideally suited to the extraction of narrow band signals whose duration exceeds or is at least on the order of the Fourier analysis window length. That is, for sources of this type Fourier analysis, particularly the short-term Fourier transform (STFT), does an excellent job of *focusing* the information, thus, providing features (spectral amplitudes)

*The work of the first author is supported by the Charles Stark Draper Laboratory under a research fellowship. The work of the last two authors is supported in part by the Draper Laboratory IR& D Program under agreement DL-H-418524, by the Air Force Office of Scientific Research under Grant AFSOR-92-J-0002, and by the Army Research Office under Grant DAAL03-92-G-0115.

perfectly suited to detection and discrimination. The STFT does allow for some temporal as well as frequency resolution, but it is not well suited for the analysis of many transient signals and, in particular, to the generation of features for detection and discrimination. The STFT may be viewed as a uniform division of the time-frequency space. It is calculated for consecutive segments of time using a predetermined window length. The accuracy of the STFT for extracting localized time/frequency information is limited by the length of this window relative to the duration of the signal. If the window is long in comparison with the signal duration there will be time averaging of the spectral information in that window. On the other hand, the window must be long enough so that there is not excessive frequency distortion of the signal spectrum. The STFT with its non-varying window is not readily adaptable for capturing signal-specific characteristics. Additionally, all time resolution is lost within each window. We look to the wavelet packet transform for a bit more freedom in dealing with this time-frequency trade off.

1.3 Current Work In This Area

Current work in the area of underwater acoustic transient classification using wavelet related concepts has been done by Nicolas [5] and, more recently, Desai and Shazeer [6]. They both employ a wavelet packet transform as a means of generating class dependent features from various classes of underwater acoustic transients for input to a neural network. In both studies, exploitation of class dependent frequency characteristics are suppressed by using a predetermined wavelet packet basis (or orthonormal division of the frequency space). The choice of the wavelet packet basis appears to be ad hoc in both cases. By limiting the input to one signal-independent feature set the adaptability of the neural network was left unexploited. These methods also ignore the redundancy between parent and children bins of the transform (discussed later in this paper). Additionally, by prohibiting signal specific division of the time-frequency space there can be no exploitation of any class dependent frequency variations. A natural expansion of their works is to address the issue of finding a wavelet packet based feature set that offers maximum feature separability due to class-specific characteristics.

2 THEORY

2.1 Wavelet Packet Decomposition (WPD)

The wavelet packet decomposition (WPD) of a signal can be viewed as a step by step transformation of the signal from the time domain to the frequency domain. The top level of the WPD tree is the time representation of the signal. As each level of the tree is traversed there is an increase in the trade off between time and frequency resolution. The bottom level of a fully decomposed tree is a frequency representation of the signal.

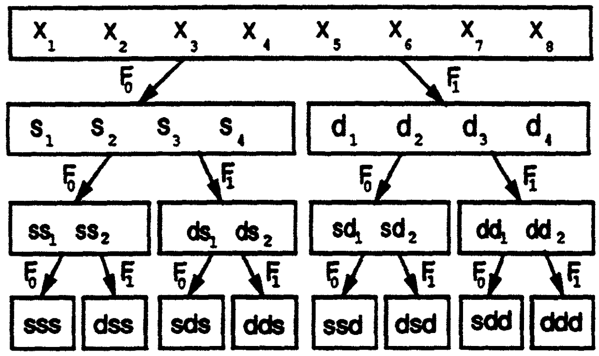
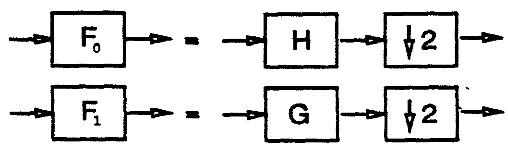


Figure 1: The fully decomposed wavelet packet tree for a signal of length eight.

This section presents the ideas developed by Wickerhauser [1] extending wavelet concepts to wavelet packets. Using Wickerhauser's notation, let $h(n)$ and $g(n)$ be the finite impulse response low-pass and high-pass filters derived from the wavelet chosen for the decomposition. Let $\mathbf{x}$ be the vector having elements $x_n = x(n)$, where $x(n)$ is the original discrete-time sequence that we wish to decompose via the wavelet packet method. Let F_0 and F_1 be the operators which perform the convolution with $h(n)$ and $g(n)$, respectively, followed by a decimation by two. The convolution and decimation steps in the WPD can be interpreted as a discrete time filtering and downsampling.



The full WPD can be displayed as a tree with a discrete sequence represented by a bin vector at the end of each branch. The the original discrete signal is on the first level. The bin vectors at each level are calculated by applying F_0 and F_1 to the bin vectors of the previous level. Figure 1 shows a WPD tree for a signal of length 8. Wickerhauser uses the notation $\mathbf{s}$ and $\mathbf{d}$ to represent the sequences resulting from the applications of F_0 and F_1 to $\mathbf{x}$.

$$F_0\,\mathbf{x} = \mathbf{s} \quad \& \quad F_1\,\mathbf{x} = \mathbf{d}$$

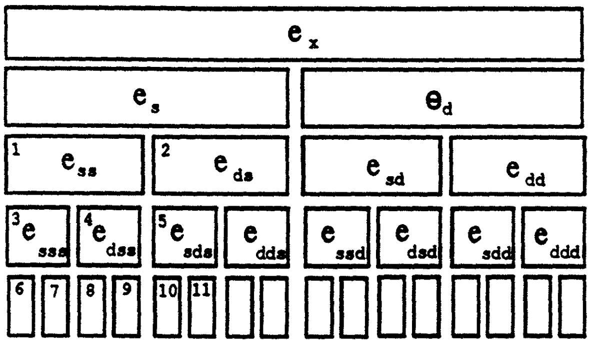


Figure 2: Energy mapping of the top five levels of a WPD tree. The eleven numbered bins comprise the energy vector used for analysis in Section 3.

In Figure 1, s_i and d_i are the i^{th} components of **s** and **d**.

The **s** and **d** are used as prefixes for the bin vector symbols throughout the tree because the low-pass filter-decimation operation can be compared to a sum and the high-pass filter-decimation can be compared to a difference. For example, at the third level of the tree **ss**, **ds**, **sd**, and **dd** are the vectors resulting from the filtering-decimation operations

$$F_0\,\mathbf{s} = \mathbf{ss} \quad \& \quad F_1\,\mathbf{s} = \mathbf{ds}$$

and

$$F_0\,\mathbf{d} = \mathbf{sd} \quad \& \quad F_1\,\mathbf{d} = \mathbf{dd}.$$

Due to the decimation, each bin vector contains half as many elements as its parent bin vector. The decomposition can be carried down to the final level where there is only one element in each bin vector.

2.2 Energy Mapping of the WPD

An intuitively pleasing representation of the WPD tree is one that highlights the energy distribution of the signal as it is decomposed down the tree. Such an energy map calculates an energy, e_y, from each bin vector, **y**. A simple energy calculation is the total energy in each bin

$$e_y = \frac{1}{2^N} \sum_{j=1}^{2^N} \mathbf{y_j}^2 \qquad (1)$$

where 2^N is the number of elements in **y**.

There is, however, freedom in the choice of the calculation of e_y. For example, due to the short duration of a shrimp snap relative to a whale click, it may be beneficial to calculate a windowed energy in each bin by calculating the energy in adjacent or overlapping segments of the bin vector, choosing a segment length small enough to encompass one snap at a time. Figure 2 shows the energy mapping the top five levels of a WPD tree.

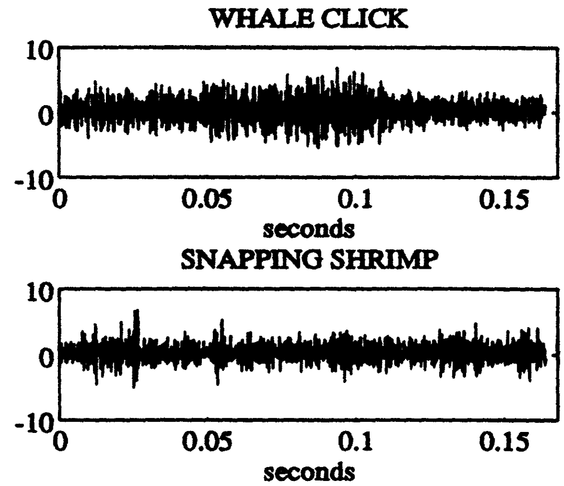


Figure 3: A typical whale click and snapping shrimp.

3 CHOOSING THE FEATURE SET

In the formulation of a decision rule, it is desirable to find a feature set which captures characteristics unique to each class of signals. Typically, the feature set uses a greatly reduced number of parameters in comparison with the number of samples in the signal.

Our feature set was found via an empirical study of the data using 54 excerpts from the NUSC data records. The duration of each excerpt is 4096 samples or 163.8 milliseconds. A typical whale click will have a duration of approximately 80 to 120 milliseconds and a single snap of a shrimp will have a duration on the order of 1 millisecond. The 4096 sample window will from one snap to an uncountably large number of snaps and can entirely encompass one whale click. The sample signal data base comprises 18 isolated whale clicks, 18 background noise excerpts, and 18 snapping shrimp excerpts. Figure 3 shows the time plots of a typical whale click and some snapping shrimp.

The transformation of the WPD trees into energy maps using (1) showed promising clarification of information. We began analysis of the 54 sample energy maps using the eleven bin energies corresponding to numbered bins shown in Figure 2. The choice of these eleven bins is discussed in greater detail in [2]. Let $\mathbf{e}_t^k$ be the energy vector containing these eleven bin energies, where $k = 1,\ldots,18$ for each t = shrimp, click, or noise.

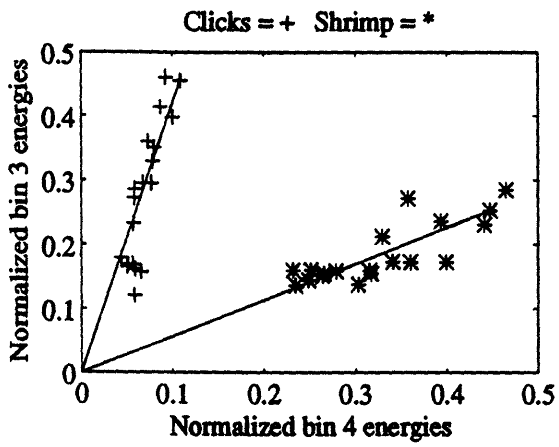


Figure 4: Normalized energies from bins 3 and 4 of the sample energy maps for snapping shrimp and whale clicks.

Each bin energy contains both signal and noise energies. Before continuing the search for a reduced parameter feature vector, the influence of noise on these eleven bin energies was compensated for by normalizing each of the eleven bin energies by the corresponding average noise energy as discussed in [2]. Each of the 54 energy vectors, $\mathbf{e}_t^k$, corresponds to the eleven bin energies from each of the 54 sample energy maps. From these we find 54 normalized energy vectors, $\hat{\mathbf{e}}_t^k$, containing the corresponding normalized bin energies.

A quantitative analysis was done by grouping the 54 normalized energy vectors into three classes and arranging them into three matrices, E_t, having columns $\hat{\mathbf{e}}_t^1$ through $\hat{\mathbf{e}}_t^{18}$ for each class, t = click, shrimp, and noise. Singular value decomposition of each E_t reveals one significant singular value, σ_t, and singular vector, $\mathbf{u}_t$ for each class, t. All other singular values were negligible. From examination of these three singular vectors we found that the bins numbered 3,4,7,8 and 9 in Figure 2 contain the dominant information.

Reduction of the feature vector is desirable for the simplification of the decision rule, and superfluous information should be avoided. A feature set which contains a parent bin energy and all of its descendant bin energies may be redundant because any parent bin vector of the WPD tree can be constructed from a linear combination of its children bin vectors. The feature set need not include all of the energies in bins of the energy map that are related in this way. This will also minimize the computational complexity of the energy vector because many bins of the WPD tree will not be used and will, therefore, not be calculated. Bins 3 and 4 are parents to 7, 8 and 9, thus, we begin with only the 3rd and 4th bin energies. Figure 4 plots the normalized energies from bins 3 and 4 of the 54 sample energy maps. There is excellent separation between the click and shrimp features.

4 CONCLUSION AND FUTURE WORK

This paper has presented results for the case of snapping shrimp and whale clicks; we are able to find a wavelet packet based feature set containing only two parameters which offers excellent separation of class specific characteristics. These features will greatly simplify the classification process for these two classes of signals.

Forthcoming, we are formulating a number of methods for detection and classification using neural networks and various pattern recognition techniques that lend themselves to the classification of signals using features of a limited number of sample signals as a training set. We are examining the robustness of the detection and classification algorithms derived from this reduced parameter feature set by running them on the entire NUSC data base which includes underwater sounds generated by popping ice, porpoise whistles, and whale cries in addition to many occurrences of snapping shrimp and whale clicks. A detailed discussion of the derivation and performance of different algorithms used with the wavelet packet based feature set is given in [2].

* References

[1] M. V. Wickerhauser. "Lectures on wavelet packet algorithms". Technical report, Washington University, Department of Mathematics, 1992.

[2] R. E. Learned. "Wavelet packet based transient signal classification". Master's thesis, Massachusetts Institute of Technology, 1992.

[3] S. Mallat. "A theory for multiresolution signal decomposition: The wavelet representation". *IEEE Trans. Pattern Anal. Machine Intell.*, 1989.

[4] I. Daubechies. "Orthonormal bases of compactly supported wavelets". *Commun. Pure Appl. Math.*, 1988.

[5] A. Lemer J. Nicolas and D. Legitimus. "Identification automatique de bruits impulsifs en acoustique sous-marine par reseaux multi-couches". In *International workshop on neural networks and their applications*, 1989.

[6] M. Desai and D. Shazeer. "Acoustic transient analysis using wavlet decomposition". In *Proceedings of the IEEE conference on neural networks for ocean engineering*, 1991.