REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

AD-A244 888

2. REPORT DATE
December 1991

3. REPORT TYPE AND DATES COVERED
professional paper

WAVELET PREPROCESSING OF ACOUSTIC SIGNALS

5. FUNDING NUMBERS
In-house

6. AUTHOR(S)
W. Y. Huang and M. R. Solorzano

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Ocean Systems Center
San Diego, CA 92152–5000

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Naval Ocean Systems Center
San Diego, CA 92152–5000

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

This paper describes results using the wavelet transform to preprocess acoustic broadband signals in a system that discriminates between different classes of acoustic bursts. This is motivated by the similarity between the proportional bandwidth filters provided by the wavelet transform and those found in biological hearing systems. The experiment involves comparing statistical pattern classifier effects of wavelet and FFT preprocessed acoustic signals. The data used was from the DARPA Phase I database, which consists of artificially generated signals with real ocean background. The results show that the wavelet transform did provide improved performance when classifying in a frame-by-frame basis. The DARPA Phase I database is well matched to proportional bandwidth filtering; i.e., signal classes that contain high frequencies do tend to have shorter duration in this database. It is also noted that the decreasing background levels at high frequencies compensate for the poor match of the wavelet transform for long duration (high frequency) signals.

Published in the *Proceedings of 25th Annual Asilomar Conference on Signals, Systems, and Computers*, November, 1991.

14. SUBJECT TERMS

15. NUMBER OF PAGES

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	SAME AS REPORT

UNCLASSIFIED

21a. NAME OF RESPONSIBLE INDIVIDUAL	21b. TELEPHONE (include Area Code)	21c. OFFICE SYMBOL
W. Y. Huang	(619) 553-9893	Code 535

Accession For	
NTIS GRA&I	☑
DTIC TAB	☐
Unannounced	☐
Justification	
By	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A-1	

presented at
25th Annual Asilomar Conference on
Signals, Systems and Computers
Pacific Grove, CA, Nov 6, 1991

Wavelet Preprocessing of Acoustic Signals

William Y. Huang and *Manuel R. Solorzano*
Code 535
Marine Sciences and Technology Dept.
Naval Ocean Systems Center
San Diego, CA 92152–5000

Abstract

This paper describes results using the wavelet transform to preprocess acoustic broadband signals in a system that discriminates between different classes of acoustic bursts. This is motivated by the similarity between the proportional bandwidth filters provided by the wavelet transform and those found in biological hearing systems. The experiment involves comparing statistical pattern classifier effects of wavelet and FFT preprocessed acoustic signals. The data used was from the "DARPA Phase I" database, which consists of artificially generated signals with real ocean background. The results show that the wavelet transform did provide improved performance when classifying in a frame–by–frame basis. The DARPA Phase I database is well matched to proportional bandwidth filtering; i.e., signal classes that contain high frequencies do tend to have shorter duration in this database. It is also noted that the decreasing background levels at high frequencies compensate for the poor match of the wavelet transform for long duration (high frequency) signals.

1 Introduction

An ocean acoustic event classification system includes a pre–processor, a frame–level classifier, and higher level decision logic. For known signal in background, there are a number of ways to optimize at each stage to maximize overall detection/classification probabilities. For short duration ocean acoustic events, however, we look for algorithms that are robust under different training conditions. The goal of this study is to compare wavelet and Fourier preprocessing for a system with the same classifier, in order to identify the characteristics in the pre–processor that lead to good overall performance.

There are a number of reasons for studying the wavelet transform. Humans are excellent acoustic event classifiers. The wavelet transform provides a proportional bandwidth (bandwidth increases in proportion with center frequency), similar to the filters in the human ear. There is considerable progress in the wavelet field. This means that systems that use the wavelet transform is supported by a rich understanding of the different aspects of the wavelet transform. Finally, the wavelet transform projects a signal into a multi–resolution space that is useful for image processing [1], speech coding [2], and sound pattern analysis [3]. For ocean acoustics, this means that wavelet transforms may be able to provide a small number of relevant parameters for the classifiers — a property that usually leads to good overall performance.

Nicolas has performed an extensive comparative study between wavelet, FFT, Wigner, and other processors for a database of short duration ocean acoustic events [4]. For the DARPA Phase I dataset, Desai [5] used wavelet a transform, with sophisticated feature extraction, to attain 0% error. Beck [6] performed comparative evaluation between wavelet and FFT preprocessing for neural net classification, and found that wavelets lead to better performance for this database. The contribution we make here is to offer some explanations as to why the wavelet transform seem to lead to better performance on the DARPA Phase I database. Section 2 will provide a brief introduction to the wavelet transform, which will leave out much of the mathematical properties but focus on the properties that we feel are important. Section 3 describes our approach to this problem. Section 4 describes the DARPA Phase I database and the experimental performed. Section 4.1 discusses the experimental results.

2 Background

Since the wavelet transform will be compared to the Fourier transform, perhaps the best way to introduce the wavelet transform is to compare and contrast it with the Short Time Fourier Transform (STFT) in the problem of time–frequency localization.

The STFT estimate of a signal, $f(t)$, in the time–frequency domain, for the mth frequency and nth time component, is defined as

$$c_{\mathrm{stft}}[m,n] = \int e^{-j(m\Delta\omega)t} g(t-n\Delta t) f(t)\,dt \qquad (1)$$

where

$$\begin{aligned} f(t) &= \text{the input signal,} \\ g(t) &= \text{a window,} \\ m\Delta\omega &= \text{frequency shift, and} \\ n\Delta t &= \text{time shift.} \end{aligned}$$

In other words, c_{sftf} is a discrete Fourier component of a windowed input. $g(t)$ is a window, and is required to have finite support in both the time and frequency domain, and centered around $t = 0$ and $\omega = 0$ in the

Table 1: Contrasts between STFT and wavelet transform.

STFT	*WAVELET*
$g(t)$ is centered at $t = 0$, $\omega = 0$ with compact support	$h(t)$ is centered at $t = 0$, with $H(\omega = 0) = 0$, and $H(\omega)$ is centered at k_0
$c_{\mathrm{stft}}[m,n]$ samples $f(t)$ at $\omega = m\Delta\omega$, $t = n\Delta t$	$c_W[m,n]$ samples $f(t)$ at $\omega = k_0/a^m$, $t = bn/a^m$
same bandwidth and time resolution for all m and n, the bandwidth and time resolution of $g(t)$	bandwidth scales as $a^m(BW)$, time window scales as $(Length)/a^m$

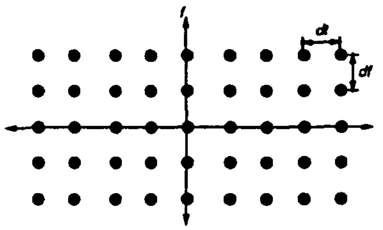

Figure 1: The sampling locations, bandwidth and time window length for STFT in the time–frequency domain.

time and frequency domain (in order to get resolution in time and frequency). The Wavelet transform, on the other hand, is defined as

$$c_W[m,n] = \int \sqrt{a^m} h(a^m t - bn) f(t) dt \qquad (2)$$

where

$$\begin{aligned} h(t) &= \text{wavelet function,} \\ a^m &= \text{a dilation factor, and} \\ nb/a^m &= \text{a time shift.} \end{aligned}$$

Here, $h(t)$, the wavelet function, is required to have finite support in both time and frequency, centered around $t = 0$ in the time domain, and centered around $\omega = k_0 \neq 0$ in the frequency domain. The reason for the requirement of it being not centered at 0 in the frequency domain is that frequency shifting is achieved by scaling to achieve a center frequency of $\omega = k_0/a^m$, and no frequency shifting can be achieved if $k_0 = 0$.

The differences between the wavelet transform and STFT are summarized in Table 1, and illustrated in Figs. 1 and 2. The main point is that wavelet transform provides proportional bandwidths, with wider bandwidth/higher time resolution at high frequencies, and finer bandwidth/wide time windows at lower frequencies.

To see why the wavelet is a biological model, consider the frequency response of the cochlea of a cat,

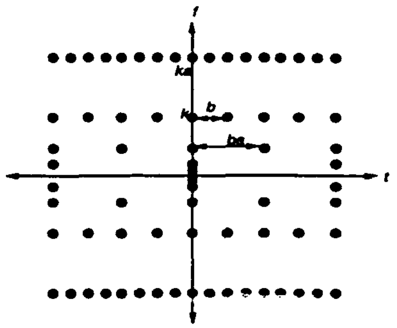

Figure 2: The sampling locations, bandwidth and time window length for wavelet transform in the time–frequency domain.

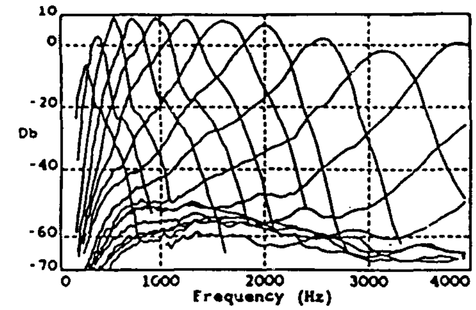

Figure 3: Frequency response of the hair cells at different parts of the cochlea of a cat. After [Bertrand, 1989]

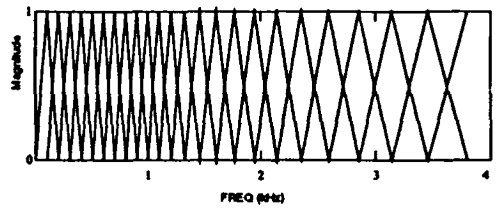

Figure 4: "Mel-scaled" filter bank: Linear spaced triangular filters between 0 and 1kHz, and logarithmic increments in both frequency and bandwidth afterwards.

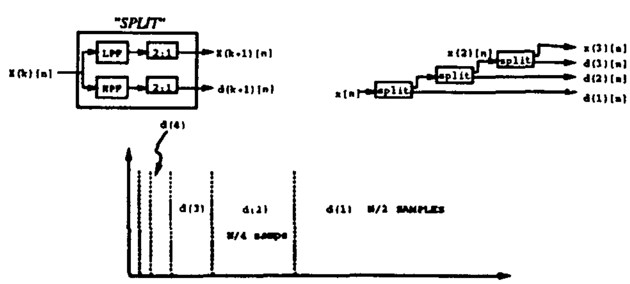

Figure 5: Octave band coder.

shown in Fig. 3. The biological ear, as in the case of wavelet, provide wider bandwidth in higher frequency than in lower frequency. This fact is taken advantage of in many speech processing front–ends. For instance, the "mel-scaled" cepstral coefficients (Fig. 4 leads to improved speech recognition performance [7]. The octave band speech coder, used for speech compression [2], has the same time–frequency resolution and sampling properties as the wavelet transform (Fig. 5).

2.0.1 Implementation

The octave band coder, shown in Fig. 5, represents the way the decimated wavelets are implemented in practice. The wavelet function is designed to pass frequencies from $\omega = \pi/2$ to π. At each stage or "octave", the wavelet function extracts the upper half of the bandwidth. The remaining signal (extracted with a similar low-pass filter) is decimated and then re–applied to the same set of filters. To improve the frequency resolution in the high frequency extraction step, the high frequency filter is divided into several high frequency, non–overlapping narrowband filter that collectively constitute the high frequency filter — and thereby creating several "voices". The "tree wavelet" [8], or "wavelet packets" [9] seeks to improve the resolution in the high frequency filter by splitting not only the low frequency component at the next octave, but also split the high frequency signal.

3 Approach

Performance comparison studies [10, 11] have demonstrated that a number of pattern classifiers will produces results near the Bayes' optimal value. Therefore, for the purpose of this study, evaluation will be restricted to the quadratic Gaussian classifier. Evaluation will be performed based on frame–level classification results. Recent studies have demonstrated that good frame–level recognition rates do not automatically extend to good event–level recognition. In [12], for instance, classifiers that provided best frame–level performance did not lead to good segmental–level performance. However, for the purpose of this study, it was felt that there is a strong correlation between the frame–level scores and segmental level scores so that we can make a valid interpretation from the frame–level scores. Here, the segmentation of data samples (assigning class labels to pre–processor frame outputs) is done by Hidden Markov Models (HMM's) and forced Viterbi decoding, where the HMM's were trained on the training set. Therefore, the segmentation boundaries are those that would lead to the highest recognition rates on both the training and testing set.

4 Experiments

The six signals of the DARPA Phase I dataset was processed by the conventional FFT (summarized in Table 2), and wavelets (Table 3). The frame–by–frame output of the pre–processors being studied are then fed into a quadratic Gaussian classifier.

Table 4 shows the percentage errors resulting from the Gaussian classifier for the various pre–processors.

Table 2: FFT Pre–processing schemes. *NAVG* is the number of FFT frames averaged together, *DATA WINDOW* is the amount of data that each frame of output examines. *FRAME SIZE* is the dimensionality of data supplied to the classifier.

id	*fft size*	*%overlap*	*navg*	*data window*	*frame size*
ft128	128	50	1	5 msec	65
ft128avg3	128	50	3	10msec	65
ft256	256	75	1	10msec	129
ft64avg4	64	50	4	6.25msec	33

Table 3: Wavelet Pre–processing schemes for the decimated wavelet and the tree wavelet. The Morlet wavelet is a modulated Gaussian pulse. Daubechies wavelet is a 4 point orthogonal wavelet function. ΔT is the time interval between wavelet outputs.

id	*# octaves*	*# voices*	*wavelet*
decwvlt	7	4	Morlet
tree	1	8	Daubechies

id	*window*	*ΔT*	*frame size*
decwvlt	168.4msec	.5msec	28
tree	2.56msec	.04msec	17

Table 4: Final classification percentage errors.

PRE-PROCESSOR	*PERCENTAGE ERROR*
decwvlt	0.14
ft128	7.93
ft128avg3	8.98
ft256	2.58
ft64avg4	6.37
tree	23.40

Table 5: Percentage confusion matrix for DECWVLT.

TRUE CLASS	*ESTIMATED CLASS* #	A	B	C	D	E
#	99.89	0	0	.01	.02	.08
A	0	100	0	0	0	0
B	0	0	100	0	0	0
C	0	0	0	100	0	0
D	0	0	0	0	100	0
E	.19	0	0	0	0	99.81

Table 6: Percentage confusion matrix for FT128.

TRUE CLASS	*ESTIMATED CLASS* #	A	B	C	D	E
#	91.65	0	0	0	0	8.35
A	0	100	0	0	0	0
B	31.25	0	56.15	0	0	12.5
C	0	0	0	100	0	0
D	0	0	0	0	97.62	2.38
E	6.96	0	0	0	0	93.04

Table 7: Percentage confusion matrix for FT128AVG3.

TRUE CLASS	*ESTIMATED CLASS* #	A	B	C	D	E
#	88.77	0	0	.15	.31	10.77
A	0	100	0	0	0	0
B	40	0	60	0	0	0
C	0	0	0	100	0	0
D	0	0	0	0	100	0
E	3.70	0	0	0	0	96.30

Table 8: Percentage confusion matrix for FT64AVG4.

TRUE CLASS	*ESTIMATED CLASS* #	A	B	C	D	E
#	94.87	0	.24	0	0	4.90
A	0	100	0	0	0	0
B	0	0	100	0	0	0
C	0	0	0	100	0	0
D	0	0	0	0	100	0
E	9.15	0	0	0	0	90.85

Table 9: Percentage confusion matrix for FT256.

TRUE CLASS	*ESTIMATED CLASS* #	A	B	C	D	E
#	98.3	0	.14	0	0	1.56
A	14.29	85.71	0	0	0	0
B	37.5	0	62.5	0	0	0
C	0	0	0	100	0	0
D	4.44	0	0	0	95.56	0
E	5.37	0	0	0	0	94.63

Table 10: Percentage confusion matrix for TREE.

TRUE CLASS	*ESTIMATED CLASS* #	A	B	C	D	E
#	75.94	0	3.14	0	0	20.92
A	0	96	0	0	0	4
B	1.04	0	98.96	0	0	0
C	0	0	0	100	0	0
D	0	0	0	0	100	0
E	20.97	0	1.33	0	.51	77.19

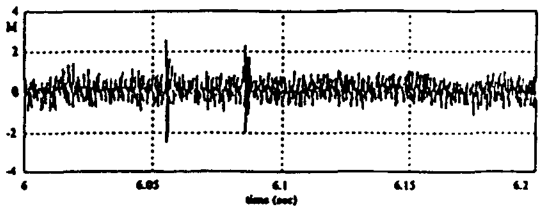

Figure 6: Signal "B".

The confusion matrices are shown in Tables 5 through 8.

4.1 Discussion of Results

The results, according to Table 4, shows that wavelet pre–processing is only marginally superior to FFT. This is consistent with the results of other studies into this problem. In fact, a statistical analysis may reveal that the difference is not significant. However, an examination of the confusion matrices reveal several distinct advantages of the wavelet transform.

First of all, the decimated wavelet performed better than the FFT based methods on signal E. Signal E is a long duration sinusoid. It occurs at a low frequency region that coincides to a region of high levels of ocean noise (high relative to other frequency bands). The time–frequency plots shows that the wavelet transform provided a sharp, distinct features for signal E. This is because at low frequency, the decimated wavelet have a large time window, corresponding to a significant SNR improvement via processing gain. The only other plots that appears to be quite distinct for signal E is that of the 256 point FFT and the 128 point FFT with 3 frame averaging. However, the decimated wavelet resulted in better performance on signal E than the 256 point FFT, because it can integrate over a time window of 168.4 msec's, whereas the 256 point FFT integrates only over 10 msec's.

In the case of signal B, Fig. 6, the wavelet transform was able to demonstrate its multi–resolution advantages. Signal B has important features in both its gross and fine details. Since signal B consists of two pulses that are approximately 30 to 40 msec apart, a pre–processor with a shorter width will likely mistaken the middle part of signal B for noise. Thus, large FFT's, such as the 256 point FFT, fared far better than the 128 point FFT with single frame averaging. The solution to this problem using FFT is to use larger FFT's; however, wavelets allow for classification of long events, such as signal B, without compromising temporal resolution.

Finally, note that the decimated wavelet provided the best overall performance. This may be an indication that it has provided a succinct, parsimonious representation of the essential features of this database. We note that the decimated wavelet output has only 28 points. The FFT can provide a small dimensionality (the 64 point FFT, for instance, provides 33 points), but not without degradation in performance.

The results provided by the tree wavelet are preliminary. It seems that 8 voices are not enough for this problem.

5 Conclusions

The use of the wavelet transform as a front end of an acoustic broadband discriminator is advantageous under at least two conditions. The first condition is for the recognition of a long duration low frequency burst in the ocean, where most of the background noise is also in the low frequency. In this low SNR case, longer wavelet filters in the low frequency provides higher SNR by a processing gain (a stronger signal component resulting from integrating the data over a longer period of time). The second condition is where multiresolution is required. Pulse trains emitted by dolphins, for instance, have a fine feature for the individual pulses, and a coarse feature characterized by the frequency and duration of the pulse train. If both features are important characteristics, the wavelet transform provides a good feature extraction.

The experiments here also suggest that wavelet transforms may provide a parsimonious representation of the signal, which can represent all the essential features of a signal with few parameters. However, the these experiments do not imply that wavelet transform is better for all other databases. Using the converse of the arguments made above, one would guess that wavelet might not perform as well as FFT for a long duration high frequency signal when the amount of noise is significant.

Acknowledgement

The authors are grateful to Dr. Mark Shensa for his many helpful discussions.

References

[1] S. G. Mallat, "Multifrequency channel decomposition of images and wavelet models," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, pp. 2091–2110, Dec 1989.

[2] M. J. T. Smith and T. P. Barnwell, "Exact reconstruction techniques for tree-structured subband coders," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 34, pp. 434–440, June 1986.

[3] R. Kronland-Martinet, J. Morlet, and A. Grossman, "Analysis of sound patterns through wavelet transforms," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 1, no. 2, pp. 274-302.

[4] J.-M. Nicolas, A. Lemer, and J.-C. Delvigne, "Automatic identification of transient biological noises in underwater acoustics using arborescent wavelets and neural networks," in *Colloque "Ondeletts et Premières Applications, Marseille*, June 1989.

[5] M. Desai and D. J. Shazeer, "Acoustic transient analysis using wavelet decomposition," in *Proceedings of the IEEE Conference on Neural Networks for Ocean Engineering*, pp. 29–40, IEEE, August 1991.

[6] S. Beck, L. Deuser, R. Still, and J. Whiteley, "A hybrid neural network classifier of short duration acoustic signals," in *Internation Joint Conference on Neural Networks*, pp. I-119–124, IEEE, July 8–12 1991.

[7] S. B. Davis and P. Mermelstein, "Comparison of parametric representation for monosyllabic word recognition in continuously spoken sentences," *IEEE Transactions on Acoustic, Speech and Signal Processing*, vol. ASSP-28, pp. 357–66, August 1980.

[8] M. J. Shensa, "The discrete wavelet transform," Technical Report 1426, Naval Ocean Systems Center, San Diego, June 1991.

[9] R. R. Coifman, Y. Meyer, S. Quake, and M. V. Wickerhauser, "Signal processing and compression with wavelet packets," tech. rep., Department of Mathematics, Yale University, April 1990.

[10] Y. Lee, "Classifiers: Adaptive modules in pattern recognition systems," Master's thesis, MIT, May 1989.

[11] W. Y. Huang and R. P. Lippmann, "Comparisons between conventional and neural net classifiers," in *1st International Conference on Neural Network*, pp. IV–485, IEEE, June 1987.

[12] W. Huang and R. Lippmann, "HMM speech recognition with neural net discrimination," in *Advances in Neural Information Processing Systems 2* (D. S. Touretzky, ed.), pp. 194–202, San Mateo, CA: Morgan Kaufmann, 1990.