Underwater Sound Detection based on Hilbert Transform Pairs of Wavelet Bases

Shun-Hsyung Chang, Fu-Tai Wang
Dept. of Electrical Engineering, National Taiwan Ocean University
2 Pei-Ning Rd., 20224 Keelung, Taiwan
E-mail: shchang@mail.ntou.edu.tw

Abstract- **We designed a new algorithm to detect multipath signals in underwater environments. The underwater sounds can be thought to comprise of acoustical signals of interest superimposed on a background of underwater noise. Though there are no obvious distribution assumptions which can be made to model underwater sound, the use of recursive density estimator of the initial underwater background noise, reconstructed by the discrete wavelet transform (DWT), can be considered to establish an empirical model for it. In a multipath environment, where signals are arbitrarily time delayed, the lack of translation invariance drawback of DWT is a pitfall for the model established. This paper proposes to use a Hilbert transform pairs of wavelet bases based on infinite-product formula and spectral factorization in dual-tree discrete wavelet transform structure (HISDT DWT) as a solution to it. The performance of the HISDT DWT exhibiting much better shift invariance than the DWT is illustrated. By utilizing the HISDT DWT to establish the recursive density estimator of the underwater background noise, the ability of detecting a multipath signal in an underwater environment were improved compared to that of by using the DWT's.**

I INTRODUCTION

This paper proposes the use of HISDT DWT, followed by using recursive density estimation as a solution to detecting multipath signals in underwater situations. We refer acoustic events of interest as "signals," and refer the background underwater sound environment as "noise." Some prior knowledge of frequency bands of signals allows us to focus on it and can be thought of as "transient features of potential interest." On the other hand, though there are no obvious distribution assumptions which can be made to model such noise, the authors of [1] used an empirical model of the noise and then allowed this model to adapt with time and let it be a baseline at any time for identifying the signals. They found that summary features of wavelet decompositions of underwater sounds appeared to exhibit distinctive multivariate behavior when signals occured. Under the sea, multipaths signals result from surface and bottom reflections and refractive phenomena due to the nonhomogenous ocean environment caused by variations in water temperature and various salinity levels. In such a multipath environment, where numbers of signal components arrive within arbitrary delays, that the DWT's lack of translation invariance is a major pitfall [2]. In order to reconstruct the initial background sound by the DWT reconstruction,arbitrary time delay is the problem that has to be solved.

By using a pair of wavelet transforms where the wavelets form a Hilbert transform pair, the authors of [3] demonstrated that the lack of time-invariance drawback of the DWT can be overcome. Based on Hilbert pairs of wavelets, the dual-tree complex wavelet transform (DT CWT) of [4] have many benefits including improved approximate shift invariance. The author of [5] designed the pairs of dyadic wavelet bases where the two wavelets form an approximate Hilbert transfrom pair based on spectral factorization. By the infinite-product formula, we can get two lowpass scaling filters that are offset from one another by a half sample, and their corresponding wavelets are a Hilbert transform pair. Using these two digital filters in the DT CWT structure [5], we have the HISDT DWT which is a nearly shift-invariant wavelet transform. The performance of the HISDT DWT exhibiting much better shift invariance than the DWT is illustrated later. By utilizing the HISDT DWT to establish the recursive density estimator of the underwater background noise, the following observations that are identified as an outlier by it are then flagged as potential signals. We show that the ability of detecting a multipath signal in an underwater environment by using the HISDT DWT is improved compared to that of by using the DWT.

II HILBERT TRANSFORM PAIRS OF WAVELET BASES AND SHIFT INVARANCE

Because wavelets are compactly supported in a time domain, the DWT is a good way for processing non-stationary signals such as transient. But the lacking of translation invariance for DWT analysis is a problem for processing signals that are arbitrarily delayed as in a multipath environment. We adopt the use of HISDT DWT as a solution to the translation invariance problem.

Let $h_0(n)$ and $h_1(n)$ are a class of exactly reconstructing filters called "conjugate quadrature filter" or CQF's [6]. That

0-933957-31-9

is

$$\sum_n h_0(n)h_0(n+2k) = \delta(k) = \begin{cases} 1, & k=0 \\ 0, & k \neq 0 \end{cases}$$

and $h_1(n) = (-z)^{(n)} h_0(M-n)$. The scaling and wavelet functions are given by the dilation and wavelet equations

$$\phi_h(t) = \sqrt{2}\sum_n h_0(n)\phi_h(2t-n)$$

$$\psi_h(t) = \sqrt{2}\sum_n h_1(n)\phi_h(2t-n)$$

Let $g_0(n)$ and $g_1(n)$ are a second CQF pair and assume filters $h_i(n)$ and $g_i(n)$, for $i = 1$ and 2, are real-valued. With $g_0(n)$ and $g_1(n)$, we can define the scaling function $\phi_g(t)$ and wavelet $\psi_g(t)$ similarly. If we define the CQF condition in terms of the autocorrelation functions

$$p_h(n) \doteq \sum_k h_0(k)h_0(k-n) = h_0(n) * h_0(-n)$$

$$p_g(n) \doteq \sum_k g_0(k)g_0(k-n) = g_0(n) * g_0(-n)$$

or equivalently as

$$P_h(z) \doteq H_0(z)H_0(1/z)$$

$$P_g(z) \doteq G_0(z)G_0(1/z)$$

then, $h_0(n)$ and $g_0(n)$ satisfy the CQF conditions if and only if $p_h(n)$ and $p_g(n)$ are halfband filters:

$$p_h(2n) = \delta(n) \quad and \quad p_g(2n) = \delta(n) \tag{2.1}$$

From the definition of the Hilbert transform, $\phi_g(t)$ is the Hilbert transform of $\phi_h(t)$ if

$$\Psi_g(\omega) = \begin{cases} -\Psi_h(\omega), & \omega > 0 \\ \Psi_h(\omega), & \omega < 0 \end{cases} \tag{2.2}$$

Let the two lowpass filters be related as

$$G_0(\omega) = H_0(\omega)e^{(-j\theta(\omega))}$$

where $\theta(\omega)$ is the 2π-periodic function. By the infinite-product formula [7], it was shown that if the phase $\theta(\omega)$ is defined as

$$\theta(\omega) = \tfrac{\omega}{2}, \quad |\omega| < \pi$$

then the two wavelets $\phi_h(t)$ and $\phi_g(t)$, generated by the digital filters h_0 and g_0 respectively, form a Hilbert transform pair. It means that the two lowpass filters are related as

$$G_0(\omega) = H_0(\omega)e^{-j(\frac{\omega}{2})} \tag{2.3}$$

and equivalently, $g_0(n)$ is a half-sample delayed version of $h_0(n)$

$$g_0(n) = h_0(n-1/2)$$

Thought a half-sample delay can not be designed by a finite impulse response (FIR) filter, it can be done by using an allpass filter with approximately constant fractional delay [5].

Let the two digital filters h_0 and g_0 have the form

$$h_0(n) = f(n) * d(n)$$

$$g_0(n) = f(n) * d(L-n)$$

We want to choose the filter $d(n)$ to achieve the (approximate) half-sample delay. In terms of the Z-transform, we have an allpass system

$$A(z) = \frac{G_0(z)}{H_0(z)} = \frac{z^{(-L)}D(1/z)}{D(z)}$$

According to Thiran's formula [8] for the maximally flat delay allpole filter of a delay τ sample

$$d(n) = (-1)^n \begin{pmatrix} L \\ n \end{pmatrix} \frac{(\tau - L)_n}{(\tau+1)_n}$$

where $(x)_n = (x)(x+1)\ldots(x+n-1)$. With this $d(n)$ and let a delay $\tau = 1/2$ samples, the allpass system $A(z)$ becomes an approximate half-sample delay

$$A(\omega) \approx e^{-j\frac{\omega}{2}}$$

and then the approximation (2.3) is achieved

$$G_0(\omega) \approx H_0(\omega)e^{-j(\frac{\omega}{2})} \quad \text{around} \quad \omega = 0$$

We now let

$$F(z) = Q(z)(1+z^{-1})^K$$

where K is the vanishing moments of wavelet bases. Then

$$H_0(z) = Q(z)(1+z^{-1})^K D(z)$$

$$G_0(z) = Q(z)(1+z^{-1})^K z^{(-L)} D(1/z)$$

and the autocorrelation functions of h_0 and g_0 are the same

$$\begin{aligned} P(z) &\doteq P_h(z) = P_g(z) \\ &= Q(z)Q(1/z)(z+2+z^{-1})^K D(z)D(1/z) \end{aligned}$$

Given $L = 2$ and $K = 4$ and using a spectral factorization approach, we can find $Q(z)$ such that h_0 and g_0 satisfy the CQF's condition (2.1) and they are illustrated in Fig. 1 (a),(b). The length of digital filters h_0 and g_0 is $2(L+K) = 12$. Their corresponding wavelets for which form a Hilbert transform pair (2.2) are shown in Fig. 2 (a),(b) and (c). Using these two digital filters in the dual-tree discrete wavelet transform structure [5], we make the HISDT DWT ploted in Fig. 3. Based on this structure, the lack of time-invariance drawback of the DWT can be overcome and the degree of shift invariance of it is illustrated in Fig. 4. The input is a unit step shifted to 17 adjacent sampling spaces in turn. Good shift invariance is shown when all the 17 output wavelet components from a given level are the same shape. This shows that the HISDT DWT is a solution to the DWT's lack of translation invariance.

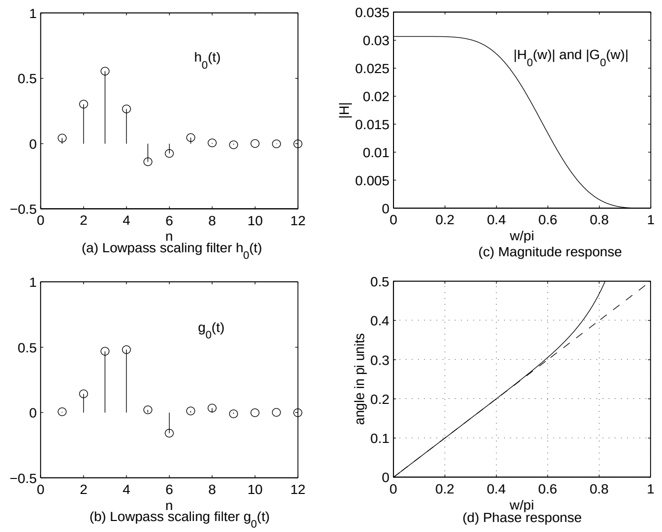

Figure 1: (a) The digital filter $h_0(n)$ with N=12, K=4, L=2 (b) A half-sample delayed version of $h_0(n)$ (c) The magnitude response of $h_0(n)$ and $g_0(n)$ (d) The phase of $G_0(\omega)/H_0(\omega)$.

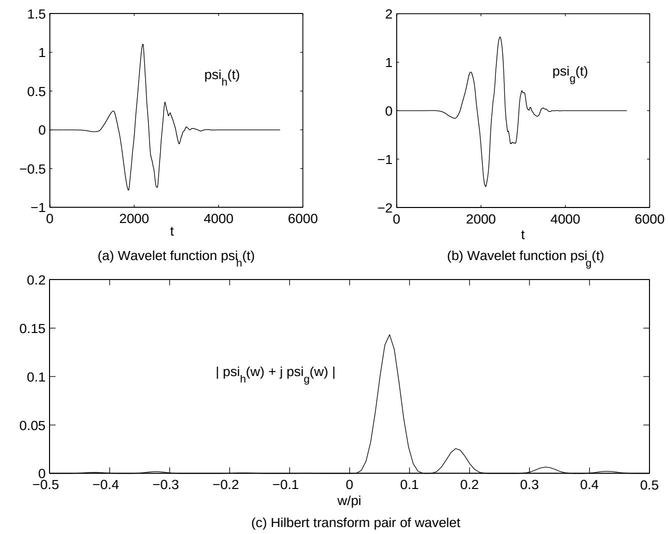

Figure 2: (a)(b) Approximate Hilbert transform pair of wavelet bases with N=12, K=4, L=2 and (c) The function $|\Psi_h(\omega) + j\Psi_g(\omega)|$.

III RECURSIVE DENSITY ESTIMATION AND SIGNAL DETECTION

Summary characteristics of the wavelet decomposition within temporal windows can be relevant to signal detection [1]. According to that article, we can divide given underwater sound recordings into appropriate periods. Using wavelet analyses, each of these segments had its wavelet decomposition. The wavelet coefficients in different levels are then summarized to several mean sums of squares, $(x_{t,1}\ ,\ \cdots\ ,\ x_{t,8})$ that covering the frequency range of 0.1-20KHz which would be of interest to us, so that

$$x_{t,j-6} = \frac{\sum_{k=1}^{2^{j-7}} d_{j,k}^2}{2^{j-7}}\ , \qquad j\ =\ 7\ ,\ \cdots\ ,\ 14$$

where $d_{j,k}^2$ are taken from time window t. Because the signals which we are concerned are in the higher frequency ranges, we focus on three mean sums of squares in time window t and form a vector of multivariate observation $\mathbf{x}_t = (x_{t,5}\ ,\ x_{t,6}\ ,\ x_{t,7})$, the behavior of which cover the range 2-10KHz.

The joint behavior of these mean sums of squares of noise is significantly different from that of signals. This result suggests to us to use density estimates to establish the initial density estimate of noise and then to detect the signal. We can obtain density estimates by using a multivariate kernel density estimator with a multivariate normal kernel [9]. So that, given data $\mathbf{x}_t, ..., \mathbf{x}_T$, the estimator used was

$$\hat{f}(\mathbf{x}) = \frac{(\det S)^{1/2}}{(2\pi)^{3/2} T h^3} \sum_{t=1}^{T} \exp(-\frac{(\mathbf{x}-\mathbf{x}_t)' S^{-1} (\mathbf{x}-\mathbf{x}_t)}{2h^2}) \tag{3.1}$$

where S is a robust estimate of the covariance matrix [10], [11], and h is a suitable global window width. Optimal choice of h can be determined by cross-validation. This means that we have to maximize the approximate log-likelihood for the data $\sum_{t=1}^{T} \log(\hat{f}_t(\mathbf{x}_t))$, where $\hat{f}_t$ refers to the kernel density estimate based on all given data $\mathbf{x}_1, ..., \mathbf{x}_T$, except $\mathbf{x}_t$ itself. Some sophisticated methods for window width selection, such as generalized cross-validation, could be adopted if required. The duration T of the initial sample size has to be large enough to ensure stable kernel estimate [9]. In our particular case, the duration $T = 256$ is selected. This kernel estimator $\hat{f}(\mathbf{x})$ is first applied to an initial sample, $\mathbf{x}_1, ..., \mathbf{x}_T$, taken from the sound recording that consist of pure background noise without signals of interest.

Because there are no obvious distribution assumptions which can be made to model such noise, we treated it as an empirical model of the noise, once this initial kernel estimate was established, and then considered the next observed vector of the selected mean sums of squares $\mathbf{x}$. Now we use a detection criterion to test if $\mathbf{x}$ is signal or not.

$$\frac{\#\{\hat{f}_t(\mathbf{x}_t) \le \hat{f}(\mathbf{x})\} + 1}{T+1} \tag{3.2}$$

where "$\#\{\cdot\}$" means "the number of" for $t = 1, ..., T$; $\hat{f}$ denotes the kernel density estimate based on $\mathbf{x}_1, ..., \mathbf{x}_T$; $\hat{f}_t$ denotes the kernel density estimate based on $\mathbf{x}_1, ..., \mathbf{x}_T$ but excluding $\mathbf{x}_t$. If this test produces a significantly small level, we identify $\mathbf{x}$ as an outlier from the current kernel density estimate and flag it as potential signal and then subsequently ignore it. Otherwise, this $\mathbf{x}$ will replaces $\mathbf{x}_T$, $\mathbf{x}_t$ isreplaced by $\mathbf{x}_{t+1}$, $t = 1, ..., T-1$, and the optimal window width and the kernel density are updated before the next $\mathbf{x}$. Then allow this model to adapt with time and let it be a baseline at any time for identifying the signals.

Under the sea, a multipath environment where numbers signal components arrive within arbitrary delays is the problem

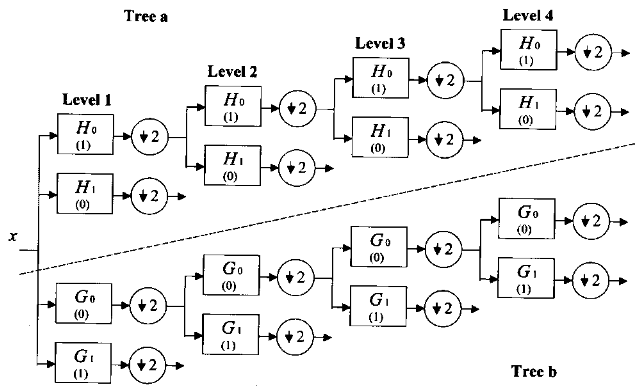

Figure 3: The structure of the HISDT DWT.

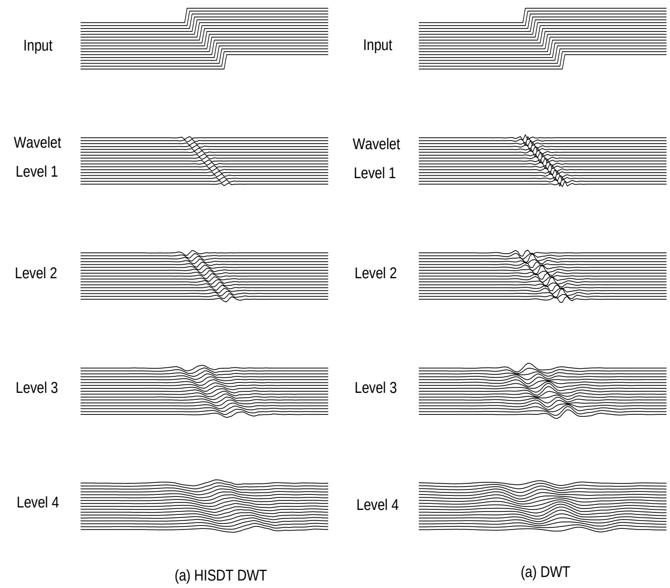

Figure 4: Wavelet components at level 1 to 4 of 17 shifted step responses of (a) HISDT DWT and (b) DWT.

which has to be solved. Assume that $g(t)$ denotes the signal of interest and the multipath signal is produced from a linear combination of delayed replicas,

$$g_M(t) = \sum_{k=1}^{K} \alpha_k g(t - \tau_k) \tag{3.3}$$

where the number of replicas K, and attenuation coefficient $\{\alpha_k\}$, and the delays $\{\tau_k\}$ are all unknown. We want to find out if we can improve the ability of detecting signals not only in a background underwater sound environment but also in a multipaths by taking the advantage of the ability of translation invariance of the HISDT DWT and the power of the recursive density estimation.

We examine the results when our signal detection method is applied to a multipath signal $g_M(t)$ in (3.3). In our simulation, the number of paths K is set to 4, all attenuation factors $\{\alpha_k\}$ are equal to 1, and the delay times $\{\tau_k, k = 1, ..., 4\}$ are set to different values to produce the multipath signal $g_M(t)$. The signal of interest $g(t)$ is the dolphin sounds. In order to have the result of being translation invariant, the vector of multivariate observation and three mean sums of squares in time window t $\mathbf{x}_t$= $(x_{t,5}\ ,\ x_{t,6}\ ,\ x_{t,7})$ are constructed from the HISDT DWT reconstruction instead of from the DWT decomposition. Let $\mathbf{x}_t, t = 1, ..., T$ be the set of initial points taken from the HISDT DWT reconstruction of the noise without signal $g_M(t)$ appeared.

From these initial points, a robust estimate of the covariance matrix S was computed. And by simple cross-validation, optimal choice of the window width $h = 0.38$ was determined, as illustrated in Fig. 5. Then the initial stable kernel estimate of the three-dimensional density of noise can be obtained by using a multivariate kernel density estimator with a multivariate normal kernel (3.1). Once the initial density estimate was established, it was treated as an empirical model of the noise. Subsequent vectors of multivariate observations were then successively considered. Finally, the detection criterion (3.2) was applied using a 1% significance level to detect the signals.

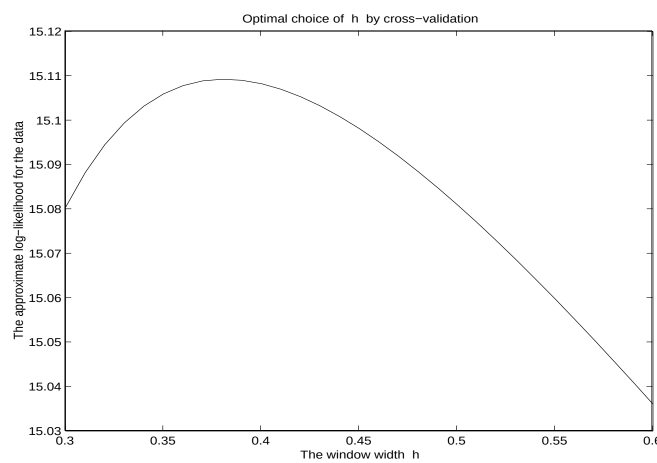

Figure 5: Optimal choice of window width h for given data $\mathbf{x}_{t,t=1,...,T}$..

Fig. 6 is a plot of the HISDT DWT reconstruction (Top), in frequency bands 2.5kHz, 5.1kHz, and 10.2kHz (with 2.5kHz at the top) and the HISDT DWT detection results (Bottom) of multipaths dolphin sound $g_M(t)$. In total, the HISDT DWT method identifies 75 signals in the 256 unclassified observations of the input $g_M(t)$. On the other hand, the signal detection results by using the DWT are depicted in Fig. 7. In that case, the DWT method identifies 71 signals in the 256 unclassified observations of the input $g_M(t)$. Although the two methods agree on the identification of the multipath signal, however, the DWT method has identified 4 fewer signals than the

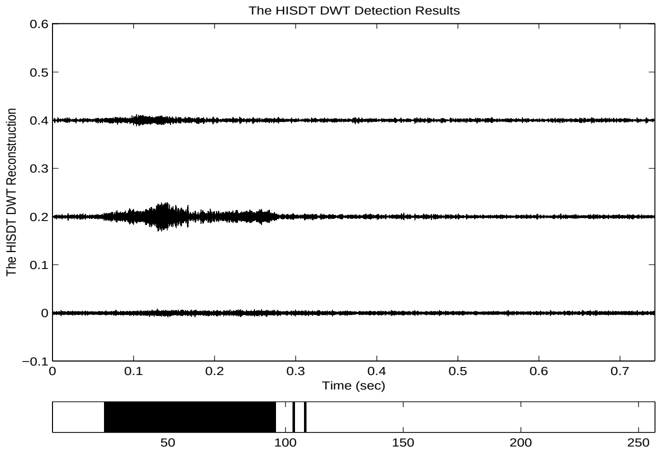

Figure 6: The HISDT DWT reconstruction (Top) and the HISDT DWT detection results (Bottom) of multipaths dolphin sound $g_M(t)$.

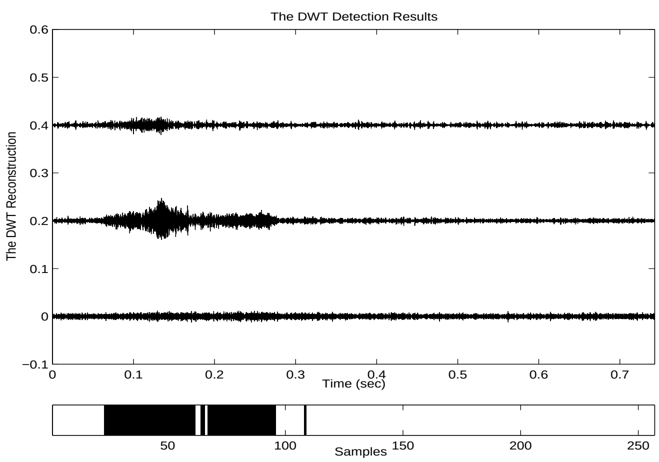

Figure 7: The DWT reconstruction (Top) and the DWT detection results (Bottom) of multipaths dolphin sound $g_M(t)$.

number identified by the HISDT DWT approach on the corresponding input of the multipath dolphin signal.

IV CONCLUSION

We propose the use of a Hilbert transform pairs of wavelet bases based on infinite-product formula and spectral factorization in dual-tree discrete wavelet transform structure (HISDT DWT), followed by recursive density estimation as a solution to detecting multipath signals in underwater situations. The recursive density estimator of the initial underwater background noise, reconstructed by the DWT can be considered to establish an empirical model for underwater sound. Instead of the DWT, the HISDT DWT is utilized to establish the recursive density estimator of the underwater background noise. The results of the previous section show that by taking advantage of the ability of translation invariance of the HISDT DWT and the empirical model built by the recursive density estimation, we can improve the ability of detecting signals not only in a background underwater sound environment but also in multipaths.

Acknowledgments

This work was supported by the National Science Council of the R.O.C. under contract NSC 91-2611-E-019-027.

REFERENCES

[1] T. C. Bailey, T. Sapatinas, K. J. Powell, and W. J. Krzanowski, “Signal detection in underwater sound using wavelets,” *Journal of the American Statistical Association*, vol. 93, no.441, pp. 73-83, March 1998.

[2] S. A. Benno and J. M. F. Moura, “On translation invariant subspaces and critically sampled wavelet transforms,” *Multidimensional Systems and Signal Process.*, vol. 8, pp. 89-110, Jan. 1997.

[3] P. Abry and P. Flandrin, “Multiresolution transient detection,” *Proc. IEEE-SP Int. Symp. Time-Freq. Time-Scale Anal.*, pp. 225-228, Oct. 1994.

[4] N. G. Kingsbury, “Complex wavelets for shift invariant analysis and filtering of signals,” *Appl. Comput. Harmon. Anal.*, vol. 10, no. 3, pp. 234-253, May 2001.

[5] I. W. Selesnick, “The design of approximate Hilbert transform pairs of wavelet bases,” *IEEE Trans. Signal Processing*, vol. 50, pp. 1144-1152, May 2002.

[6] M. J. T. Smith and T. P. Barnwell, III, “Exact reconstruction techniques for tree-structured subband coders,” *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 434-441, June 1986.

[7] I. W. Selesnick, “Hilbert transform pairs of wavelet bases,” *IEEE Signal Processing Lett.*, vol. 8, pp. 170-173, June 2001.

[8] J. P. Thiran, “Recursive digital filters with maximally flat group delay,” *IEEE Trans. Circuit Theory*, vol. CT-18, pp. 659-664, Nov. 1971.

[9] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*, London: Chapman and Hall, 1986.

[10] F. Mosteller and J. W. Tukey, *Data Analysis and Regression*,Addison-Wesley, 1977.

[11] P. A. Tukey and J. W. Tukey, 1981, *Graphical Display of Data sets in 3 or more Dimensions*,Chichester: Wesley, 1981.