Autocorrelation based denoising of manatee vocalizations using the undecimated discrete wavelet transform

Berke M. Gur[a] and Christopher Niezrecki
Department of Mechanical Engineering, University of Massachusetts-Lowell, Lowell, Massachusetts 01854

(Received 13 July 2006; revised 3 April 2007; accepted 4 April 2007)

Recent interest in the West Indian manatee (*Trichechus manatus latirostris*) vocalizations has been primarily induced by an effort to reduce manatee mortality rates due to watercraft collisions. A warning system based on passive acoustic detection of manatee vocalizations is desired. The success and feasibility of such a system depends on effective denoising of the vocalizations in the presence of high levels of background noise. In the last decade, simple and effective wavelet domain nonlinear denoising methods have emerged as an alternative to linear estimation methods. However, the denoising performances of these methods degrades considerably with decreasing signal-to-noise ratio (SNR) and therefore are not suited for denoising manatee vocalizations in which the typical SNR is below 0 dB. Manatee vocalizations possess a strong harmonic content and a slow decaying autocorrelation function. In this paper, an efficient denoising scheme that exploits both the autocorrelation function of manatee vocalizations and effectiveness of the nonlinear wavelet transform based denoising algorithms is introduced. The suggested wavelet-based denoising algorithm is shown to outperform linear filtering methods, extending the detection range of vocalizations. [DOI: 10.1121/1.2735111]

PACS number(s): 43.30.Sf, 43.60.Hj [EJS] Pages: 188–199

I. INTRODUCTION

The West Indian manatee (*Trichechus manatus latirostris*) was added to the endangered species list in 1967 and detailed records of manatee mortalities have been kept since 1974. In 1980, the United States Fish and Wildlife Service established a manatee protection plan. Within this plan, watercraft collisions were identified as the most significant cause of manatee deaths. Accordingly, idle-speed or no-wake zones were designated throughout the Florida waterways where manatee-watercraft collisions are most likely to occur. However, it has been reported that the rate of manatee-watercraft collision related mortalities continue to remain high, despite measures taken (U.S. Fish and Wildlife Service, 2001). The problem also has an economical aspect. The boating industry in Florida has been continuously expanding (24.8% annual growth rate over the last 25 years, 4% annual growth rate since 2001) and reached $18.4 billion in 2005 (Marine Industries Association of Florida Inc., 2007). The passive speed zones are increasing the travel time within the channels and have a negative impact on boating, the boating industry, real estate, and development.

Recent research has focused on developing more effective solutions that can reduce mortalities while forcing minimal restrictions on boaters. Boater warning systems based on active sonar (Jaffe *et al.*, 2007), above sea level infrared (Keith, 2002), and passive acoustic detection (Herbert *et al.*, 2002; Mann *et al.*, 2002; Niezrecki and Beusse, 2002) methods have been previously suggested for this purpose while Gerstein and Blue (2004) investigated an active acoustic manatee warning system. Passive acoustic detection methods exploit the harmonic nature of manatee vocalizations to discriminate them from background noise. However, the harmonic content of manatee vocalizations becomes less discernible as the signal-to-noise ratio (SNR) of the received acoustic signal decreases [see Figs. 1(a) and 1(b)]. Detection becomes impossible when the SNR drops below the critical detection threshold of the detector. Several factors affect the SNR of the received signal such as background noise levels and the distance of the manatee to the receiver hydrophone. Phillips *et al.* (2006) reported that in the absence of any boat noise, the detection range decreased from 250 to 2.5 m when the background noise sound pressure level (SPL) was increased from 70 to 100 dB, for a detection threshold of 6 dB. Taking into account the fact that typical overall SPL generated by boats are approximately 140 dB, detection of manatee calls at long distances in the presence of boat noise becomes a challenging task. Therefore, denoising of underwater acoustic signals, and particularly eliminating the contribution of boat noise is crucial for successful acoustic detection of manatee calls at reasonable ranges, as well as other marine mammal vocalizations within similar noisy habitats. This paper addresses the problem of denoising manatee vocalizations contaminated with boat noise.

A typical manatee vocalization lasts between 0.2 and 0.5 s and has a fundamental frequency followed by several harmonics (see Fig. 1). Both the fundamental and the harmonics generally are frequency modulated and the vocalizations are considered nonstationary. The fundamental frequency ranges between 2 and 5 kHz but can be as low as 600 Hz. In general, the first harmonic dominates the vocalization and contains the majority of the acoustic energy. Harmonics can be found at up to 18 kHz (Schevill and Watkins, 1965; Hartman, 1979; Steel, 1982; Bengston and Fitzgerald, 1985; Gerstein *et al.*, 1999; Nowacek *et al.*, 2003; Phillips *et*

[a]Electronic mail: berke_gur@student.uml.edu

 0001-4966/2007/122(1)/188/12/$23.00

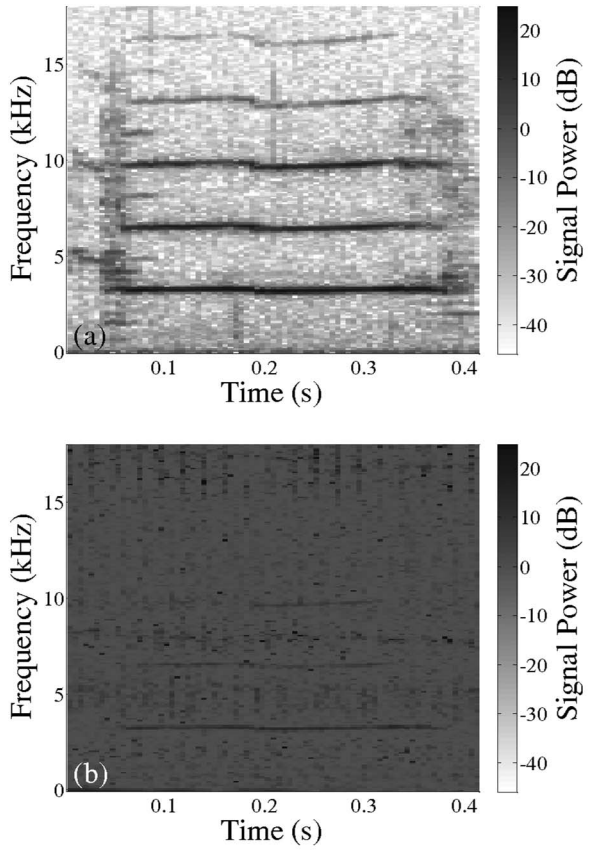

FIG. 1. (a) The spectrogram of a typical manatee call and (b) the spectrogram of the manatee call corrupted by boat noise (SNR=−5 dB).

al., 2004). The underwater acoustic environments of the shallow Florida channels are primarily corrupted with broadband boat noise. Snapping shrimp noise also contributes to the background noise and is more significant in brackish or saltwater environments. Typical snapping shrimp noise is broadband (up to or above 100 kHz), contains high energy spikes, with short durations (~1 ms).

The problem of denoising, detecting, and classifying underwater bioacoustic signals has been studied in the past with a majority of the effort directed toward marine mammals, in particular cetacean vocalizations. A variety of methods, including simple spectral energy ratios (Mellinger and Clark, 1994), matched filtering (Stafford *et al.*, 1998), time warping (Buck and Tyack, 1993), spectral correlation (Mellinger and Clark, 2000), and artificial neural networks (Potter *et al.*, 1994; Deecke and Janik, 2006) have been implemented for this purpose.

In an attempt to detect manatee vocalizations in noisy environments, bandpass and high-pass filters have been incorporated into a harmonic content detector (Niezrecki *et al.*, 2003). Yan *et al.*, (2005, 2006), benchmarked linear adaptive filtering algorithms against the denoising capability of a tenth-order Butterworth bandpass filter. Using test signals with a SNR between 0 and −15 dB, the denoising performance of the feedback adaptive line enhancer (FALE) was the finite-impulse-response filter structured adaptive line enhancer (FIR-ALE).

In the last two decades, the wavelet transform has emerged as an alternative to the short-time Fourier transform (STFT) for time-frequency analysis of nonstationary signals and has been used in analyzing transient signals in sonar applications (Chen *et al.*, 1998; Carevic, 2005). In addition, several researchers including Learned and Willsky (1995), Bailey *et al.* (1998), Huynh *et al.* (1998), and recently, Adam (2006) have implemented the wavelet transform in favor of the STFT in feature extraction, detection, and classification of marine mammal vocalizations.

However, since its introduction, the wavelet transform has found many other applications besides time-frequency analysis. In their innovative work, Donoho and Johnstone (1994) demonstrate that thresholding the wavelet coefficients obtained through an orthogonal wavelet transform results in a simple, nonlinear estimator that surpasses all previous linear estimators in estimating functions of a wide class of smoothness from noisy observations. The authors propose two nonlinear thresholding rules (hard thresholding and soft thresholding), along with an estimation scheme (VISUSHRINK) that performs thresholding on the wavelet coefficients of the noisy signal. The VISUSHRINK estimator is shown to be nearly asymptotically optimal in the minimax sense for the mean-square-error (MSE) risk. Following VISUSHRINK, Donoho and Johnstone (1995) derived the SURESHRINK estimator which utilizes level-dependent thresholding. Coifman and Donoho (1995) suggested that the wavelet decomposition be performed by the undecimated discrete wavelet transform (UDWT) and reported improvements in denoising performance compared to the discrete wavelet transform (DWT) algorithms. The UDWT was further employed in the work of Xu *et al.* (1994), Lang *et al.* (1996), and Pan *et al.* (1999).

The VISUSHRINK and SURESHRINK methods are simple and effective. Despite this, researchers implementing these algorithms emphasized that optimizing the estimator for minimax MSE performance was not suitable for some applications. This led to the development of a variety of wavelet domain denoising schemes. In medical image processing and image restoration applications, several researchers developed alternate thresholds and noise level estimation methods (Chang *et al.*, 2000; Bao and Zhang, 2003; Pizurica *et al.*, 2003). Weiss and Dixon (1997) applied the continuous wavelet transform (CWT) for denoising underwater high-frequency pulse returns.

A wavelet-based denoising strategy is fully described by the wavelet family, level of decomposition, thresholding rule, and the threshold. The selection of these parameters is dependent on the time, frequency, smoothness, and statistical characteristics of the target and noise signals, as well as the risk criteria (Taswell, 2000). The scope of this paper is limited to a single receiver hydrophone setup and possible improvements through more sophisticated multichannel techniques such as beamforming are not included in the discussion. Accordingly, the manatee vocalizations are referred to as the target signal and any signal other than the vocalizations are assumed to be noise. Denoising is defined as estimating the wave form of manatee vocalizations from received signals contaminated with typical underwater noise.

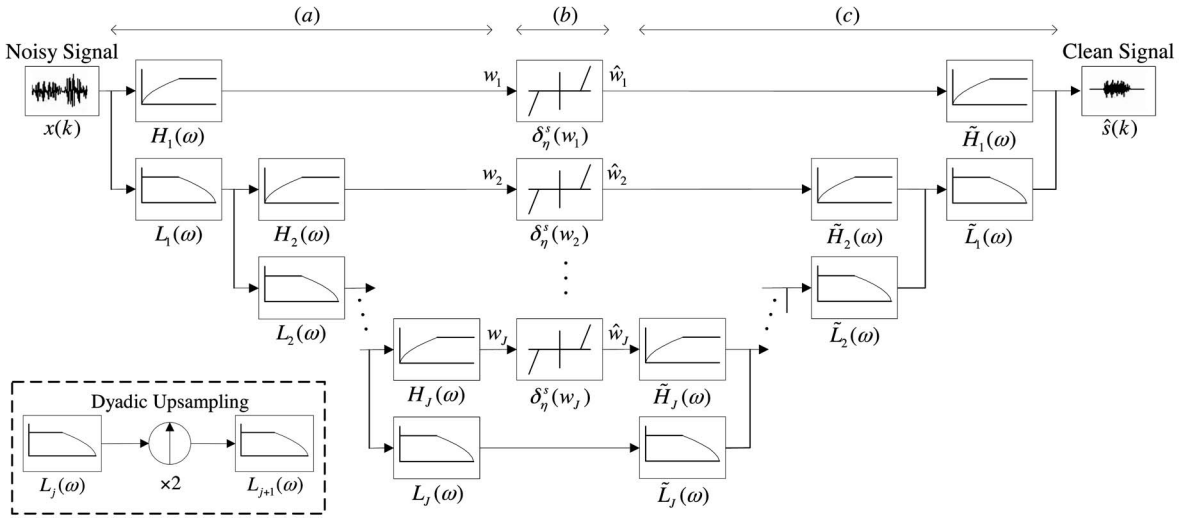

FIG. 2. The wavelet denoising scheme consists of (a) the forward UDWT; (b) thresholding; and (c) the inverse UDWT, where $L_j(\omega)$, $H_j(\omega)$, $\tilde{L}_j(\omega)$, $\tilde{H}_j(\omega)$, and $\delta_\eta^s(w_j)$ are the decomposition/reconstruction QMF pairs and the thresholding rule associated with level j. The low-/high-pass decomposition filter pairs at two consecutive levels are related through upsampling, as shown in the dashed box.

To the authors' knowledge, wavelet-based denoising of underwater acoustic signals for the detection of manatee vocalizations has not yet been investigated and, thus, represents the focus of this paper. In this study, a simple wavelet-domain thresholding scheme is proposed and tested. This paper is organized as follows. Section II provides an overview of the theoretical development of the wavelet transform, classical wavelet shrinkage denoising, and the newly developed autocorrelation-based denoising scheme. The performances of the denoising algorithms introduced in Sec. II are evaluated through a simulation setup defined in Sec. III, while the results of these simulations are presented in Sec. IV. Lastly, Sec. V presents conclusions and suggestions for future research.

II. THEORETICAL DEVELOPMENT

A. The undecimated discrete wavelet transform

The one-dimensional DWT of an analog signal $x(t) \in L^2(\mathbb{R})$ is the integral transform given in

$$\{\mathcal{W}_\psi x\}(j,k) = \int_{-\infty}^{\infty} x(t)\psi_{j,k}^*(t)dt, \tag{1}$$

where $\{\mathcal{W}_\psi x\}(j,k)$ are the wavelet coefficients of the signal at a scale $j \in \mathbb{Z}^+$ and translation $k \in \mathbb{Z}$, and $\psi_{j,k}(t) = 2^{-j/2}\psi(2^{-j}t-k)$ is a set of basis functions. The wavelet analysis of discrete signals is performed within the multirate signal processing and subband decomposition framework, in which the integral transform and the continuous basis functions $\psi_{j,k}(t)$ of Eq. (1) are replaced by the convolution operation and digital filters, respectively. In general, the decomposition is achieved through a pair of low/high pass FIR quadrature mirror filters (QMF), i.e., $L(\omega)$ and $H(\omega)$. Since the frequency content is reduced by half at the output of each filter, only half of the coefficients carry distinct frequency information. Therefore, following the decomposition, the output of each filter is downsampled (i.e., even samples of the signal are dropped) to reduce computational effort in the next step. This filtering/downsampling sequence is repeated on the output of the low-pass filter up to a desired level of decomposition, resulting in a tree-like structure. Each level of decomposition splits the frequency spectrum of the low frequency coefficients in half. Convolution of finite length filters and signals introduce boundary distortions which can be eliminated by zero padding or extending the signal sufficiently and taking only the central portion of the output of the convolution operation. The inverse transform or reconstruction of the decomposed signal is accomplished by upsampling the wavelet coefficients followed by convolving with the reconstruction QMF pairs. The pair of low/high pass filters used for decomposition cannot be used for reconstruction because this leads to aliasing and signal distortion. Therefore, reconstruction is accomplished by another set of low/high pass filters, i.e., $\tilde{L}(\omega)$ and $\tilde{H}(\omega)$, which are derived from of the decomposition QMF pair.

The choice of the coefficients to be downsampled is not trivial and conflicting decomposition results are obtained by choosing a different set of coefficients to eliminate in the downsampling process. This property is often referred to as the translation variance of the DWT, in the sense that a circular shift of the signal will result in a different set of coefficients compared to the unshifted signal. Initially, several methods were suggested to overcome this translation variance of the DWT. One such method, referred to as the UDWT, eliminates the downsampling (and therefore the translation variance) and accommodates for this by upsampling the filter coefficients following each decomposition level, as shown in Fig. 2. The lack of the downsampling step introduces redundancy to the transform.

B. Wavelet domain denoising

The core concept of wavelet-based denoising algorithms is that a target signal of a certain smoothness class can be represented in the wavelet domain with a few, relatively large amplitude coefficients, while noise is assumed to be

mapped to a large number of coefficients with small magnitudes. For example, it is well known that the mapping of white Gaussian noise (WGN) to the wavelet domain results in WGN with the same noise level. Therefore, it is anticipated that eliminating the small amplitude wavelet coefficients will result in eliminating the coefficients due to noise and the reconstruction of the remaining wavelet coefficients will result in a denoised signal. The outcome of the elimination scheme is determined by a thresholding rule and a threshold. The hard thresholding rule eliminates all coefficients below the threshold while the ones above the threshold remain unchanged. Although this method has a better MSE performance, it results in discontinuities and visually unappealing results. The soft thresholding rule, given in Eqs. (2)–(4), eliminates the coefficients below the threshold while the remaining coefficients are shrunk toward zero by an amount equal to the threshold. This shrinkage of the wavelet coefficients eliminates the discontinuities associated with hard thresholding,

$$\delta_\eta^s(\{\mathcal{W}_\psi x\}(j,k)) = \mathrm{sgn}(\{\mathcal{W}_\psi x\}(j,k))(|\{\mathcal{W}_\psi x\}(j,k)| - \eta)_+ \tag{2}$$

where $\delta_\eta^s(\cdot)$ denotes the soft thresholding operator using the threshold η and

$$\mathrm{sgn}(\{\mathcal{W}_\psi x\}(j,k)) = \begin{cases} +1 & \text{if } \{\mathcal{W}_\psi x\}(j,k) > 0 \\ 0 & \text{if } \{\mathcal{W}_\psi x\}(j,k) = 0 \\ -1 & \text{otherwise,} \end{cases} \tag{3}$$

$$(|\{\mathcal{W}_\psi x\}(j,k)| - \eta)_+ = \begin{cases} (|\{\mathcal{W}_\psi x\}(j,k)| - \eta) \\ \text{if } (|\{\mathcal{W}_\psi x\}(j,k)| - \eta) \geq 0 \\ 0 \quad \text{otherwise.} \end{cases} \tag{4}$$

The VISUSHRINK algorithm introduced by Donoho and Johnstone (1994) uses the soft thresholding rule with the universal threshold of $\eta_u = \sigma_v[2\log(N)]^{1/2}$ where the noise level σ_v is estimated through the median absolute deviation (MAD) estimate of the high frequency coefficients of the first level, i.e., $\sigma_v \approx \hat{\sigma}_{w_1} = \mathrm{median}(|w_1|)/0.6745$. The VISUSHRINK is near asymptotically optimal in the minimax sense. Following VISUSHRINK, Donoho and Johnstone (1995) introduced the SURESHRINK algorithm that minimizes Stein's unbiased estimate of risk, given as $\eta_{j,\mathrm{SURE}} = \mathrm{argmin}_\eta[N - 2\cdot\#\{i:|w_j(i)| \leq \eta\} + \Sigma_i \min(|w_j(i)|, \eta)^2]$, where $\#\{\cdot\}$ denotes the number of arguments satisfying the relation given inside the curly braces, for each level of decomposition. In both VISUSHRINK and SURESHRINK, the low frequency coefficients are not thresholded.

C. Optimum threshold denoising

Donoho and Johnstone (1994) argue that an ideal thresholding scheme can be achieved if the denoising algorithm is told which coefficients to eliminate. They show that the estimation error of such an ideal estimator is the lower limit of any estimator and that VISUSHRINK comes close to the performance of this ideal estimator for functions of certain smoothness. All the algorithms presented in this paper are tested using artificially generated test signals; hence, noise-free target signals are available. Therefore, a similar approach to that of the ideal thresholding scheme is adopted to determine the upper limit of denoising performance. The optimum threshold for a soft thresholding scheme is determined by minimizing the MSE between the target and noisy signals' wavelet coefficients, as shown in the following:

$$\eta_{j,\mathrm{opt}} = \arg\min_\eta \left[\sum_k (\delta_\eta^s(\{\mathcal{W}x\}(j,k)) - \{\mathcal{W}s\}(j,k))^2 \right] \tag{5}$$

for each level $j = 1, 2, \ldots, J$. The optimal threshold is not a realizable denoising method, as it calculates the denoised signal with prior knowledge of the noise free signal and represents the upper bound on the performance of any wavelet-based thresholding denoising scheme for a similar setup.

D. Undecimated discrete wavelet transform denoising using an autocorrelation-based threshold

The underwater acoustic noise signal is best modeled as additive, shown in the following:

$$x(n) = s(n) + v(n), \tag{6}$$

where $x(n)$ is the received noisy signal, $s(n)$ is the target, and $v(n)$ is the noise signal, respectively, with $n = 0, 1, 2, \ldots, N-1$. The target signals considered in this paper are manatee vocalizations that are known to have slow decaying and essentially harmonic autocorrelation functions. The noise signal, on the other hand, can be modeled as uncorrelated, normally distributed, nonstationary random sequence with a rapidly decreasing autocorrelation function (Yan *et al.*, 2005). However, the underwater acoustic signals that fall within the scope of this text can be assumed to be wide sense stationary (WSS) within a short time window of 3 ms or 128 samples at a sampling rate of 48 kHz. The distinct behaviors of the autocorrelation functions for both the manatee vocalizations and noise do not change when the signals are mapped to the wavelet domain via the UDWT, as shown in Fig. 3. The autocorrelation function of WSS signals can be estimated using

$$r_{xx}(\tau) = \frac{1}{N}\sum_{i=1}^{N-\tau} (x(i) - \bar{x})(x(i-\tau) - \bar{x}) \tag{7}$$

where τ is the lag and $\bar{x}$ is the mean of the signal. The biased autocorrelation estimator is preferred because it results in better noise reduction performance.

In this paper, the authors suggest an alternative thresholding scheme based on the autocorrelation function of the wavelet coefficients. The proposed autocorrelation-based UDWT algorithm (AUDWT) accomplishes denoising as follows. The received noisy signal is measured with a hydrophone and is processed as 128 sample buffers. The signal is then filtered through a Butterworth tenth-order bandpass filter with a passband ranging from 0.6 to 18 kHz to eliminate any signal contamination that does not overlap with the vocalization bandwidth. The performance of the proposed AUDWT method greatly depends on the effective mapping of the harmonic content of a manatee vocalization to the wavelet domain. This can be accomplished by selecting a wavelet

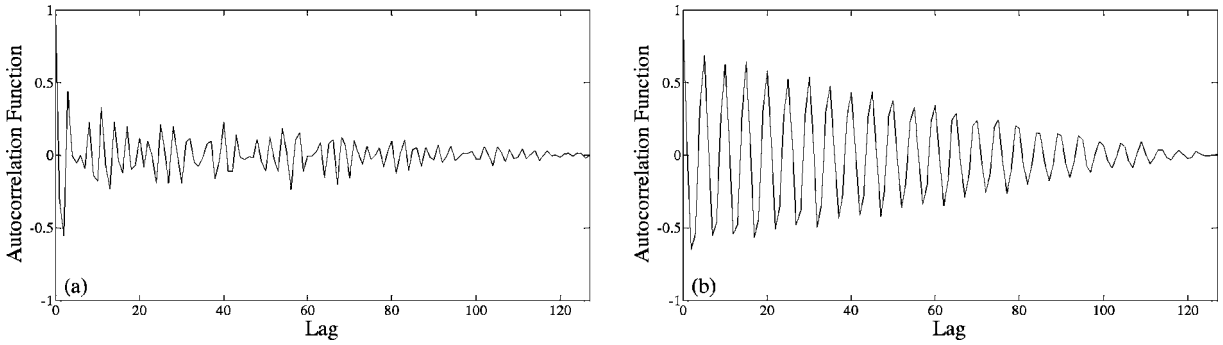

FIG. 3. The normalized autocorrelation function estimates of the detail wavelet coefficients of (a) background noise only and (b) of a manatee call plus background noise for a scale that contains vocalization energy.

with a sufficient number of vanishing moments (i.e., oscillations). Practical implementation of a subband filtering algorithm requires finite filter impulse responses. Furthermore, achieving high temporal resolution is also important since it is desired to suppress short duration signals such as snapping shrimp. The Daubechies family of wavelets achieves a maximum number of vanishing moments for a given temporal support (Daubechies, 1992). Therefore, the filtered signal, sampled at 48 kHz, is decomposed to five levels ($J=5$) using the UDWT with the Daubechies-8 (db8) wavelet. Considering the range of the fundamental frequency and the harmonic structure of manatee vocalizations, a five-level decomposition leads to the most appropriate frequency bands that result in approximate and detail coefficients $\{w_1, w_2, w_3, w_4, w_5, v_5\}$ representing the frequency bands [12–24], [6–12], [3–6], [1.5–3], [0.75–1.5], and [0–0.75] kHz, respectively. Following the UDWT decomposition, the autocorrelation function is estimated using Eq. (7). To distinguish a decaying autocorrelation function due to noise from that of a slow varying autocorrelation of a manatee call within a subband, the rms of the autocorrelation function beyond delay $\tau=20$ is calculated. The calculated rms value of the autocorrelation function is fed into a 16-point moving average filter to eliminate any contamination due to noise transients. The output of the filter is compared to a predetermined threshold constant for a given level of decomposition to determine whether thresholding will be applied to the signal or not. If the output of the filter exceeds the predetermined threshold constants, then the wavelet coefficients are shrunk using the soft thresholding rule. Several thresholds, including thresholds suggested for the UDWT (Berkner and Wells, 2001), were evaluated in terms of the MSE performance (see Sec. III). The universal threshold was determined to be the best performing threshold and is implemented in the AUDWT. If the output of the filter falls short of the threshold constants, then the wavelet coefficients are set to zero. This assures that the silent intervals within calls have no or very few unwanted transients that can increase the false alarm rate of a detector. The approximate coefficients are not included in the threshold rule and are set to zero because no or very little vocalization energy is expected within that subband. Following thresholding, the denoised signal is recovered through an inverse UDWT and the output can be applied to a signal detector or used as an input to a localization system. The block diagram implementation of the AUDWT is given in Fig. 4.

III. SIMULATION

In this study, the target signals used for simulation purposes are obtained from the manatee vocalization library developed by Yan *et al.* (2006). This library is composed of ten different categories of vocalizations based on the structure of the harmonics of the calls (the categories are labeled 0000, 1000, 1010, 1011, 1100, 1110, 1111, 1200, 1210, and 1211). Each category consists of ten calls bringing the total number of calls in the library to 100. All ten calls of each category are collected in a single file in which the vocalizations are spread with 2 s intervals, starting with the first call at 1 s.

Since the main concern of this paper is to evaluate the performance of denoising algorithms in the presence of boat noise, two noise recordings representing different boat speed configurations are used to contaminate the manatee vocaliza-

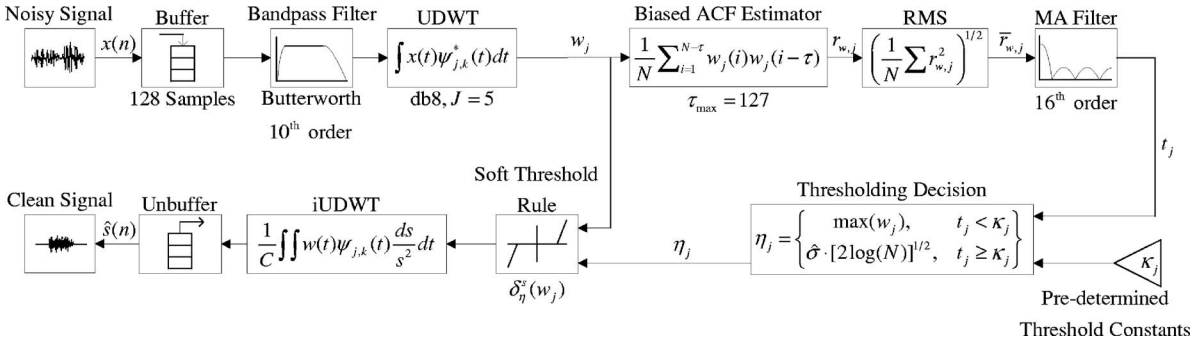

FIG. 4. The block diagram of the AUDWT method.

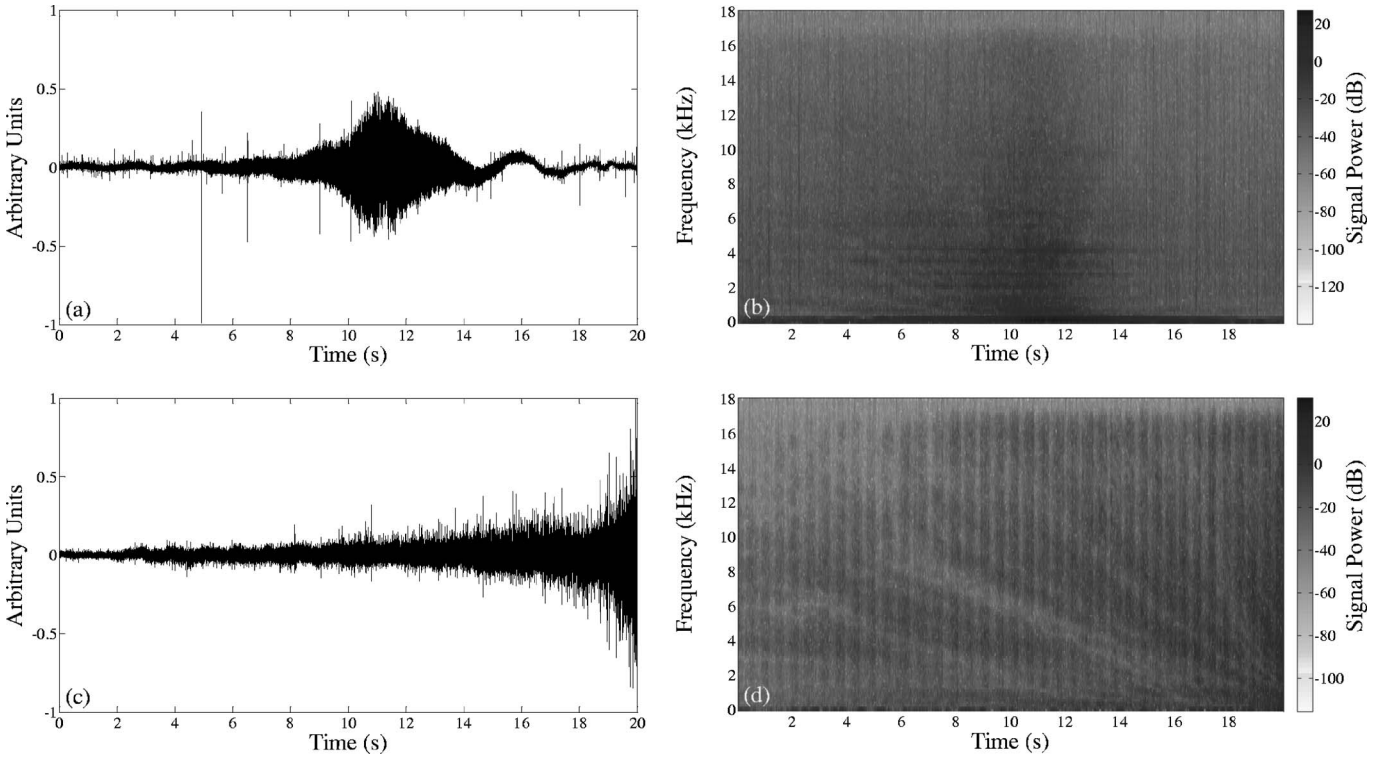

FIG. 5. Boat noise recordings used for simulation and the corresponding frequency spectra. (a), (b) Fast moving boat noise and (c), (d) slow moving boat noise.

tions. These noise recordings are selected from a collection of recordings that were made within Florida. The first recording consists of a boat making a high speed pass in a channel with snapping shrimp noise. The recording was made at Crystal River, FL in 2003. The second recording consists of sounds from a slow moving boat. This recording was made at Cedar Keys, FL in 2004. The time domain plots and spectra for both noise files are shown in Fig. 5. Both simulations consisted of files for each of the ten categories of calls contaminated with their respective noise files (slow or fast moving boat).

In order to evaluate the denoising performances of the methods, the signal power of manatee vocalizations were adjusted such that a constant SNR of −5 dB was achieved throughout the recordings. Yan *et al.* (2006), discuss the difficulty associated with accurately calculating the SNR of nonstationary signals after denoising and has suggested the following formula:

$$\overline{\mathrm{SNR}} = 10 \log_{10}\left[\sum_{i=1}^{10} P_i^2 \Big/ \sum_{i=1}^{10} Q_i^2\right], \tag{8}$$

where P_i is the rms value of the ith vocalization within a category and Q_i is the corresponding rms value of the preceding noisy section and is calculated over a length equal to that of P_i. The improvements in SNR of the vocalizations after denoising were calculated by taking the difference in SNR before and after denoising, both calculated using Eq. (8). Details of Eq. (8) can be found in the work of Yan *et al.* (2006) and the same formulation is adopted in this paper for all SNR measures. However, the SNR can be misleading when used for evaluating denoising performance. As was pointed out by Yan *et al.* (2006), the denoising algorithms in general fail to eliminate noise completely and some residue is left. This residue results in overestimation of the SNR performance. Nonetheless, SNR is an important measure in evaluating detection ranges and therefore is used as a means of measuring how well noise was suppressed. However, the SNR measure is complimented with another performance measure, the normalized mean square error (NMSE). Some applications require that the wave form of the acoustic signal be preserved. Such cases arise when the acoustic signal is used to localize the source and the performances of localizers degrade as the wave form is altered, or when the signals are used for classification and the changes in the wave form result in increased classification errors. Therefore, the NMSE, given in Eq. (9) is also used to evaluate the performance of denoising,

$$\mathrm{NMSE} = \sum_{i=1}^{10} |\hat{s}_i - s_i|^2 \Big/ \sum_{i=1}^{10} \tilde{N}_i, \tag{9}$$

where $\tilde{N}_i$ is the duration of each vocalization within the recording. The noisy vocalization has an initial NMSE, and it is assumed that the denoising reduces this value, while a NMSE of 0 corresponds to perfect preservation of the vocalization. The NMSE values reported in this paper are scaled such that the initial NSME of the noisy vocalization corresponds to 1. Therefore, the resulting NMSE performances of the denoising algorithms fall within the range of [0-1], with a lower value indicating better denoising performance. The NMSE of the optimal threshold method is the true upper performance of denoising and SNR values higher than those

TABLE I. The improvement in SNR for bandpass filtering, VISUSHRINK, FIR-ALE, and FALE, the AUDWT, and the optimal threshold methods for signals corrupted with a fast moving boat noise (original SNR=−5 dB).

Category	Bandpass filter	VISUSHRINK	FIR-ALE	FALE	AUDWT	Optimum
0000	5.95	6.30	14.89	15.65	113.62	68.69
1000	6.17	7.55	19.72	20.08	115.40	67.94
1010	5.85	7.06	17.98	18.42	114.10	69.11
1011	5.85	7.15	19.00	19.56	115.56	69.00
1100	4.43	3.20	16.71	17.16	114.10	67.18
1110	5.16	4.10	16.52	17.04	114.76	68.33
1111	6.09	5.85	18.09	19.04	16.22	69.49
1200	5.96	.48	18.80	19.43	116.32	68.54
1210	6.00	3.58	16.73	17.67	115.79	68.87
1211	5.80	4.14	18.28	18.96	116.60	68.62
Average	5.73	5.14	17.67	18.30	115.25	68.58

predicted by the optimum threshold indicate the presence of noise residue within the vocalization. All artificial recordings are generated such that they contain $N=960\,000$ samples, corresponding to a duration of 20 s for a sampling rate of $f_s=48$ kHz.

The performance of the AUDWT is compared to a bandpass filter, the VISUSHRINK of Donoho and Johnstone (1994), the FIR-ALE and FALE of a Yan *et al.* (2005, 2006), and the optimal threshold method. The bandpass filter used for comparison is the same tenth-order Butterworth bandpass filter with a passband of 0.6–18 kHz used within the AUDWT algorithm. It was observed that the SURESHRINK algorithm offered no distinct advantages compared to the VISUSHRINK method and therefore is not discussed any further. The wavelet domain algorithms (VISUSHRINK, AUDWT, and optimal threshold) process the data in buffers of 128 samples, corresponding to a window of 3 ms. The parameters of the linear adaptive filtering algorithms are set based on the suggestions given in the corresponding papers of Yan *et al.* (2005, 2006).

IV. RESULTS AND DISCUSSION

Previous papers test the performance of wavelet shrinkage algorithms with relatively clean signals having a SNR in the range of 7–20 dB (Donoho and Johnstone, 1994; Coifman and Donoho, 1995; Donoho and Johnstone, 1995; Lang *et al.*, 1996; Pan *et al.*, 1999). The results obtained using these signals are promising. However, the universal threshold of the VISUSHRINK algorithm is based on the asymptotic properties of normally distributed random numbers. Specifically, Donoho (1995) made use of the observation that for a normally distributed sequence of N random numbers $z=[z(0),z(1),\ldots,z(N-1)]$, the maximum absolute value within the sequence becomes bounded with a high probability, as shown in the following:

$$\Pr\{\max|z| \leqslant [2\log(N)]^{1/2}\} \rightarrow 1 \tag{10}$$

as $N\rightarrow\infty$. As the SNR of the signal decreases, the energy contribution of the target signal becomes weaker and the universal threshold estimation becomes too crude.

The resulting improvements in SNR (from an initial SNR of −5 dB) and NMSE of the denoised recordings for a fast traveling boat are given in Table I and Table II, respectively, for the bandpass filtering, VISUSHRINK, FIR-ALE, FALE, AUDWT, and optimal threshold methods. Similarly, the average SNR and NMSE results for denoising of vocalizations contaminated with a slow boat are given in Table III and Table IV, respectively. The bandpass filter performs much better for a fast moving boat than for a slower boat and can increase the SNR up to 6 dB. The fast moving boat noise file has more contribution from sources besides boat noise

TABLE II. The scaled NMSE for bandpass filtering, VISUSHRINK, FIR-ALE, and FALE, the AUDWT, and the optimal threshold methods for signals corrupted with a fast moving boat noise (original SNR=−5 dB).

Category	Bandpass filter	VISUSHRINK	FIR-ALE	FALE	AUDWT	Optimum
0000	0.14	0.13	0.33	0.34	0.14	0.09
1000	0.13	0.11	0.35	0.35	0.09	0.08
1010	0.09	0.07	0.26	0.27	0.06	0.05
1011	0.08	0.06	0.25	0.25	0.05	0.04
1100	0.20	0.20	0.21	0.21	0.17	0.16
1110	0.16	0.15	0.15	0.16	0.13	0.12
1111	0.12	0.12	0.22	0.23	0.09	0.07
1200	0.16	0.18	0.30	0.31	0.14	0.10
1210	0.14	0.15	0.20	0.19	0.13	0.08
1211	0.15	0.15	0.20	0.22	0.12	0.09
Average	0.14	0.13	0.25	0.25	0.11	0.09

TABLE III. The improvement in SNR for bandpass filtering, VISUSHRINK, FIR-ALE, and FALE, the AUDWT, and the optimal threshold methods for signals corrupted with a slow moving boat noise (original SNR= −5 dB).

Category	Bandpass filter	VISUSHRINK	FIR-ALE	FALE	AUDWT	Optimum
0000	0.18	1.39	4.40	3.30	23.39	50.68
1000	0.22	1.41	12.99	20.77	34.71	51.13
1010	0.23	0.53	11.61	18.45	103.63	53.38
1011	0.22	1.40	12.76	20.51	27.11	54.39
1100	0.26	1.22	11.30	19.07	31.66	53.54
1110	0.21	0.91	9.29	13.15	56.94	54.02
1111	0.25	1.11	10.11	17.27	32.69	54.21
1200	0.19	0.61	10.27	16.22	22.20	52.87
1210	0.20	0.90	7.57	14.231	24.09	52.92
1211	0.26	1.25	11.15	18.72	34.80	54.42
Average	0.22	1.07	10.14	16.17	39.12	53.45

such as snapping shrimp and wind. The better performance of the bandpass filter for this noise case can be attributed to the suppression of such noise sources that fall outside the bandwidth of manatee vocalizations. In both scenarios, in terms of the SNR, VISUSHRINK makes no significant improvement to the results achieved by bandpass filtering. The FALE works better than the FIR-ALE, especially in the case of a slow traveling boat, and results in a SNR improvement of 15–20 dB. It is evident from these results that the AUDWT performs best among the tested algorithms for both scenarios, both in the SNR and NMSE measures. However, the effects of noise residue on the SNR performance of the AUDWT are evident for the case of a fast moving boat. The improvements in SNR for the AUDWT are in the range of 115 dB while the predicted upper limit on denoising performance is about 70 dB. This difference is attributed to noise residue; hence the AUDWT has a poorer NMSE performance than the optimum threshold method. For the slow moving boat, the AUDWT achieves an average SNR improvement of 39.1 dB. It should be noted that the thresholds for the AUDWT method were determined empirically from tests conducted with a training data set such that the average NMSE performance was maximized. These thresholds were not modified during the tests with the two boat noise recordings. The noise recordings used to contaminate the manatee vocalizations in these two tests and in the training process come from different boats, set to different configurations and recorded at different locations. Thus, it can be concluded that the performance of the AUDWT algorithm appears to be robust to different sources of contamination. Plots demonstrating the typical resulting denoised signals for the tested algorithms are given in Figs. 6 and 7 for the fast and slow moving boats, respectively. A close inspection of Figs. 7(b) and 7(c) reveals that bandpass filtering alone is not effective in reducing the background noise, particularly for the slow moving boat.

The boat noise and manatee vocalization recordings were made independently, at different locations, using different equipment. To be able to convert the denoising performances to detection ranges, precise knowledge on the source levels of the boat noise and manatee vocalizations is required. Unfortunately, the environment in which these recordings are made makes it difficult to obtain such precise information. Even so, one can infer the detection ranges by assuming typical values for the unknown source levels. Several researchers have investigated background and boat noise levels of manatee habitats. Specifically, background noise levels have been reported to be in the range of 70–105 dB (ref 1 μPa) while the results for boat noise levels vary. However, an average source level of 140 dB (ref 1 μPa, at 1 m)

TABLE IV. The scaled NMSE for bandpass filtering, VISHUSHRINK, FIR-ALE, and FALE, the AUDWT, and the optimal threshold methods for signals corrupted with a slow moving boat noise (original SNR=−5 dB).

Category	Bandpass filter	VISHUSHRINK	FIR-ALE	FALE	AUDWT	Optimum
0000	1.20	0.53	0.28	0.27	0.26	0.21
1000	0.89	0.29	0.20	0.21	0.17	0.12
1010	0.98	0.31	0.19	0.18	0.27	0.17
1011	0.91	0.31	0.31	0.30	0.17	0.12
1100	0.90	0.32	0.25	0.24	0.22	0.15
1110	0.91	0.32	0.27	0.28	0.27	0.17
1111	0.89	0.31	0.26	0.26	0.21	0.14
1200	0.93	0.33	0.26	0.25	0.24	0.18
1210	0.92	0.33	0.26	0.25	0.26	0.18
1211	0.90	0.31	0.28	0.27	0.19	0.14
Average	0.94	0.34	0.26	0.25	0.23	0.16

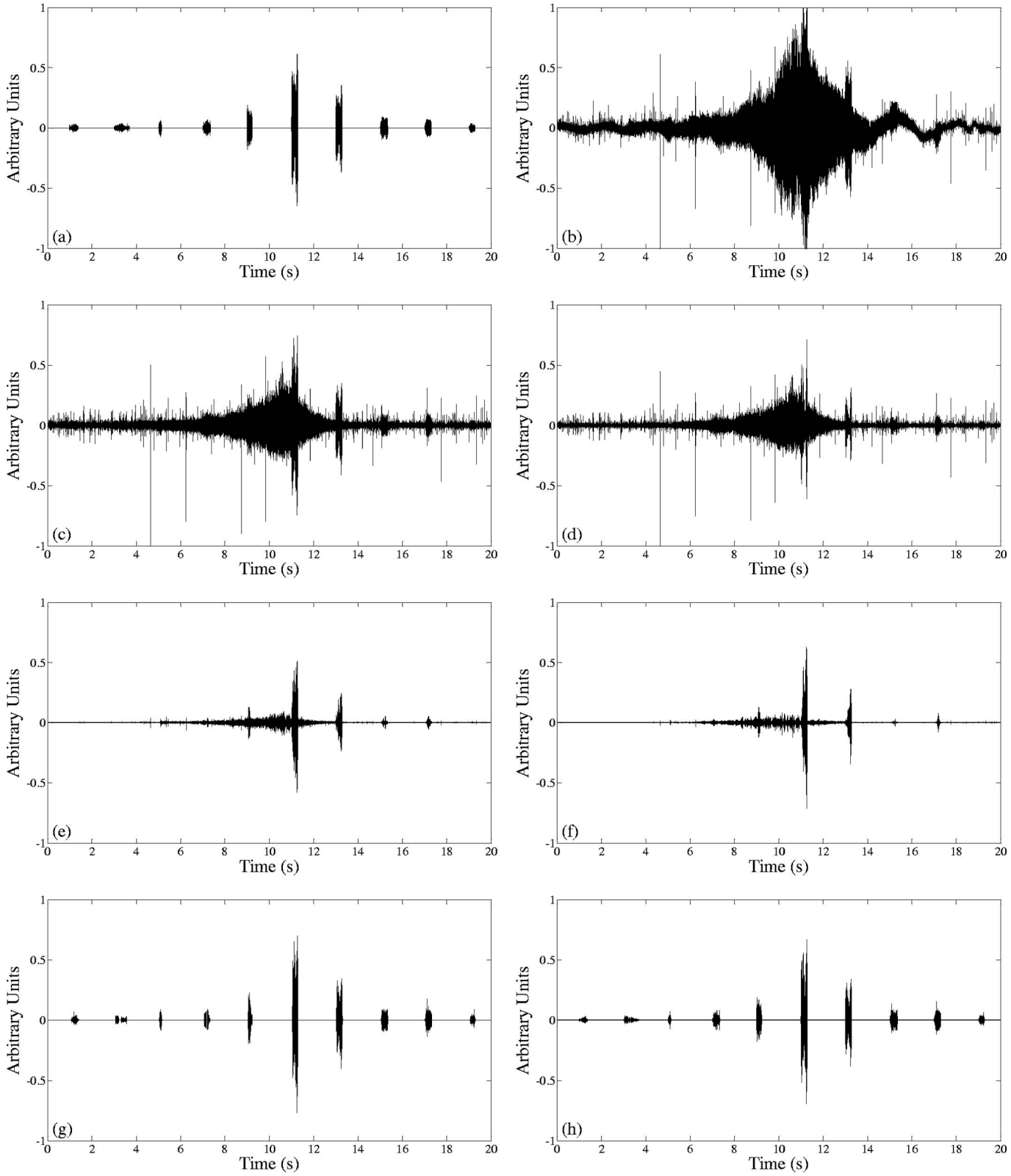

FIG. 6. (a) Noise free manatee vocalizations of type 1210, scaled for −5 dB; (b) test file obtained by the superposition of the scaled manatee vocalizations and fast moving boat noise; (c) denoising results of bandpass filtering; (d) denoising results of VISUSHRINK; (e) denoising results of FIR-ALE; (f) denoising results of FALE; (g) denoising results of AUDWT; and (h) denoising results of the optimal threshold method.

with a noise floor of 120 dB is typical for boat noise. Average manatee source levels (ref 1 μPa, at 1 m) have been reported to vary between 110 and 118 dB (Nowacek *et al.*, 2003; Phillips *et al.*, 2006). Once the boat and manatee source levels are known, the resulting SNR of a denoised recording can be converted to the distance of the manatee to the receiver hydrophone via the sonar equation given the following:

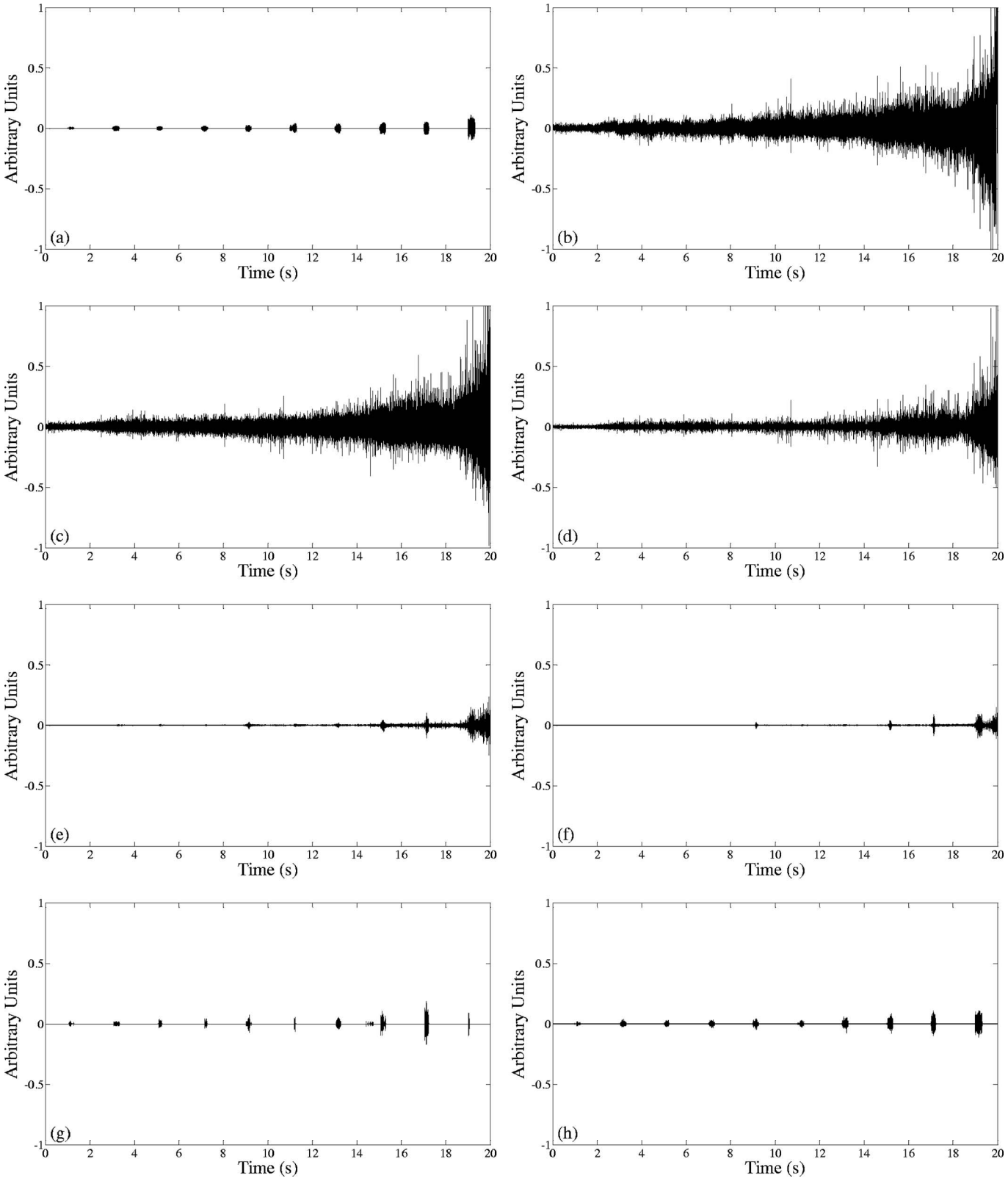

FIG. 7. (a) Noise free manatee vocalizations of type 1110, scaled for −5 dB; (b) test file obtained by the superposition of the scaled manatee vocalizations and slow moving boat noise; (c) denoising results of bandpass filtering; (d) denoising results of VISUSHRINK; (e) denoising results of FIR-ALE; (f) denoising results of FALE; (g) denoising results of AUDWT; and (h) denoising results of the optimal threshold method.

$$\mathrm{SNR}_f = \mathrm{SL}_m - 15\log_{10}(r_m) - \max[\mathrm{BL}, \mathrm{SL}_b - 15\log_{10}(r_b)] + \mathrm{DG}, \quad (11)$$

where SNR_f is the SNR after denoising, SL_m and r_m are the source level and distance of the manatee to the receiver hydrophone, respectively, BL is the background noise level, SL_b and r_b are the boat source level and range to hydrophone, respectively, and DG is the denoising gain. For detec-

tion at a range r_m, the resulting SNR_f for that range must be at least as high as the detection threshold (minimum required SNR) of a harmonic content detector, within the frequency bandwidth of interest. The noise floor of boat noise can be assumed to be 120 dB (ref 1 μPa, at 1 m) within the bandwidth of manatee vocalizations. Further, the manatee vocalizations are assumed to have a source level of 118 dB (ref 1 μPa, at 1 m). Using a mixed spreading model, the SPL of the boat noise received at the hydrophone can be estimated to peak at 105 dB when the boat is passing only 10 m away from the hydrophone. Assuming a detection threshold of 3 dB, the detection range can be calculated to be 4.6 m. Since the AUDWT can improve the SNR above the 3 dB detection threshold from −5 dB, this would correspond to an increase in the detection range up to 16 m. In an alternate scenario, keeping the source levels the same, a boat passing 50 m away from the receiver hydrophone results in a peak SPL of 94.5 dB at the hydrophone. Again, for a detection threshold of 3 dB, the detection range could be increased from 23.2 to 79.4 m.

V. CONCLUSIONS AND FUTURE RESEARCH

The problem of eliminating underwater background noise for detection of manatee vocalizations is challenging, mainly due to typical low SNR of the received signals. Bandpass filtering and the previously suggested wavelet domain VISUSHRINK method fail to achieve acceptable results at such low SNR. This paper demonstrates that the unique behavior of the autocorrelation function of manatee vocalizations can be used to perform nonlinear wavelet-based denoising when the vocalizations are corrupted with nonstationary broadband boat noise. In essence, the proposed AUDWT method outperforms linear filtering methods and has a NMSE performance approaching that of the optimal threshold method. This study demonstrates that nonlinear methods have a greater capability of improving the SNR of noisy manatee vocalizations over linear methods, thereby increasing the probability of detection and detection ranges of manatee vocalizations. Further improvement in noise reduction performance can be achieved by using narrower frequency bins to better isolate the harmonics. This can be accomplished in the wavelet domain using the wavelet packet transform. The increased computational workload associated with the wavelet packet transform can be eliminated to some degree by using an entropy-based cost function to select the most efficient decomposition tree. Noise reduction using the wavelet packet transform to denoise manatee vocalizations will be discussed in a future paper.

The AUDWT has been tested with signals of SNR lower than −5 dB and promising results were obtained. However, the denoising performance of the AUDWT degrades as the SNR approaches −10 dB for the current threshold settings. Therefore, future work can be directed toward identifying the maximum detection range (and thus the anticipated initial SNR of the noisy signals) that satisfies desired detection criteria in terms of correct and false detection rates. The proposed AUDWT makes a thresholding decision based on the presence of a coherent signal within a given frequency band. However, not all coherent signals within the manatee habitat are due to manatee vocalizations (e.g., dolphin vocalizations) and a dedicated manatee vocalization detector is necessary for achieving a satisfactory detection performance. The AUDWT noise reduction scheme can be implemented as a preprocessor to a dedicated manatee vocalization detector, such as the harmonic content detector developed by Niezrecki *et al.* (2003), resulting in a practical passive acoustic detection system to warn boaters of the presence of manatees.

ACKNOWLEDGMENTS

The authors would like to express their sincere appreciation to the Florida Sea Grant, Florida Fish and Wildlife Conservation Commission, and University of Florida Marine Mammal Program in supporting this research. The authors would also like to thank the two anonymous reviewers for their constructive comments.

Adam, O. (**2006**). "Advantages of the Hilbert Huang transform for marine mammals signals analysis," J. Acoust. Soc. Am. **120**, 2965–2973.

Bailey, T. C., Sapatinas, T., Powell, K. J., and Krazanowski, W. J. (**1998**). "Signal detection in underwater sound using wavelets," J. Am. Stat. Assoc. **93**, 73–83.

Bao, P., and Zhang, L. (**2003**). "Noise reduction for magnetic resonance images via adaptive multiscale products thresholding," IEEE Trans. Med. Imaging **22**, 1089–1099.

Bengston, J. L., and Fitzgerald, S. M. (**1985**). "Potential role of vocalizations in West Indian manatees," J. Mammal. **66**, 816–819.

Berkner, K., and Wells, R. O., Jr. (**2001**). "Denoising via nonorthogonal wavelet transforms," in *Wavelet Transforms and Time-Frequency Analysis*, edited by L. Debnath (Birkhäuser, Boston), pp. 68–80.

Buck, J. R., and Tyack, P. L. (**1993**). "A quantitative measure of similarity for *tursiops truncatus* signature whistles," J. Acoust. Soc. Am. **94**, 2497–2506.

Carevic, D. (**2005**). "Adaptive window-length detection of underwater transients using wavelets," J. Acoust. Soc. Am. **117**, 2904–2913.

Chang, S. G., Yu, B., and Vetterli, M. (**2000**). "Adaptive wavelet thresholding for image denoising and compression," IEEE Trans. Image Process. **9**, 1532–1546.

Chen, C. H., Lee, J. D., and Lin, M. C. (**1998**). "Classification of underwater signals using wavelet transforms and neural networks," Math. Comput. Modell. **27**, 47–60.

Coifman, R. R., and Donoho, D. L. (**1995**). "Translation-invariant denoising," in *Wavelets and Statistics*, edited by A. Antoniadis and G. Oppenheim (Springer, New York), pp. 125–150.

Daubechies, I. (**1992**). *Ten Lectures on Wavelets* (SIAM, Philadelphia), Chap. 6, pp. 167–213.

Deecke, V. B., and Janik, V. M. (**2006**). "Automated categorization of bioacoustic signals: Avoiding perceptual pitfalls," J. Acoust. Soc. Am. **119**, 645–653.

Donoho, D. L. (**1995**). "Denoising by soft-thresholding," IEEE Trans. Inf. Theory **41**, 613–627.

Donoho, D. L., and Johnstone, I. M. (**1994**). "Ideal spatial adaptation by wavelet shrinkage," Biometrika **81**, 425–455.

Donoho, D. L., and Johnstone, I. M. (**1995**). "Adapting to unknown smoothness via wavelet shrinkage," J. Am. Stat. Assoc. **90**, 1200–1224.

Gerstein, E. R., and Blue, J. E. (**2004**). "Investigation of the potential utility of manatee alerting devices, phase one: Acoustic warning system to alert manatees of approaching vessels," Report for the Fish and Wildlife Research Institute, http://www.floridamarine.org/features/view_article.asp?id=14362 (Last access date: 3 April 2007).

Gerstein, E. R., Gerstein, L., Forsythe, S. E., and Blue, J. E. (**1999**). "The underwater audiogram of the West Indian manatee (*Trichechus manatus*)," J. Acoust. Soc. Am. **105**, 3575–3583.

Hartman, D. S. (**1979**). "Ecology and behavior of the manatee (*Trichechus manatus*) in Florida," *Special Publication No. 5* (The American Society of Mammalogists, Lawrence, KS), pp. 98–100.

Herbert, T., Hitz, G., Mayo, C. C. S., Dobeck, G., Manning, B., Sandlin, M., Hansel, J., Bowden, T., and Artman, D. (**2002**). "Proof-of-concept for off

the shelf technology to identify acoustic signature to detect presence of manatee(s)," Report for the Fish and Wildlife Research Institute, http://www.floridamarine.org/features/view_article.asp?id=14362 (Last access date: 3 April 2007).
Huynh, Q. Q., Cooper, L. N., Intrator, N., and Shouval, H. (**1998**). "Classification of underwater mammals using feature extraction based on time-frequency analysis and BCM theory," IEEE Trans. Signal Process. **46**, 1202–1207.
Jaffe, J. S., Simonet, F., Roberts, P. L. D., and Bowles, A. E. (**2007**). "Measurement of the acoustic reflectivity of sirenia (Florida manatees) at 171 kHz," J. Acoust. Soc. Am. **121**, 158–165.
Keith, E. O. (**2002**). "Boater manatee awareness system," Report for the Fish and Wildlife Research Institute, http://www.floridamarine.org/features/view_article.asp?id=14362 (Last access date: 3 April 2007).
Lang, M., Guo, H., Odegard, J. E., Burrus, C. S., and Wells, R. O., Jr. (**1996**). "Noise reduction using an undecimated discrete wavelet transform," IEEE Signal Process. Lett. **3**, 10–12.
Learned, R. E., and Willsky, A. S. (**1995**). "A wavelet packet approach to transient signal classification," Appl. Comput. Harmon. Anal. **2**, 265–278.
Mann, D. A., Nowacek, D. P., and Reynolds, J. I. (**2002**). "Passive acoustic detection of manatee sounds to alert boaters," Report for the Fish and Wildlife Research Institute, http://www.floridamarine.org/features/view_article.asp?id=14362 (Last access date: 3 April 2007).
Marine Industries Association of Florida Inc. (**2007**). 7800 SW 57 Avenue, Suite 302, Miami, FL 33143, http://www.boatflorida.org/custom_pages/site_page_2708/index.html (Last access date: 3 April 2007).
Mellinger, D. K., and Clark, C. W. (**1994**). "A publicly accessible database for marine mammal call classification," J. Acoust. Soc. Am. **96**, 3298.
Mellinger, D. K., and Clark, C. W. (**2000**). "Recognizing transient low-frequency whale sounds by spectrogram correlation," J. Acoust. Soc. Am. **107**, 3518–3529.
Niezrecki, C., and Beusse, D. O. (**2002**). "A system for warning boaters in the presence of manatees," Report for the Fish and Wildlife Research Institute, http://www.floridamarine.org/features/view_article.asp?id=14362 (Last access date: 3 April 2007).
Niezrecki, C., Phillips, R., Meyer, M., and Beusse, D. O. (**2003**). "Acoustic detection of manatee vocalizations," J. Acoust. Soc. Am. **114**, 1640–1647.
Nowacek, D. P., Casper, B. M., Wells, R. S., Nowacek, S. M., and Mann, D. A. (**2003**). "Intraspecific and geographic variation of West Indian manatee (*Trichechus manatus spp*.) vocalizations (L)," J. Acoust. Soc. Am. **114**, 66–69.
Pan, Q., Zhang, L., Dai, G., and Zhang, H. (**1999**). "Two denoising methods by wavelet transform," IEEE Trans. Signal Process. **47**, 3401–3406.
Phillips, R., Niezrecki, C., and Beusse, D. O. (**2004**). "Determination of West Indian manatee vocalization levels and rate," J. Acoust. Soc. Am. **115**, 422–428.
Phillips, R., Niezrecki, C., and Beusse, D. O. (**2006**). "Theoretical detection ranges for acoustic based manatee avoidance technology," J. Acoust. Soc. Am. **120**, 153–163.
Pizurica, A., Phillips, W., Lemahieu, I., and Acheroy, M. (**2003**). "A versatile wavelet domain noise filteration technique for medical imaging," IEEE Trans. Med. Imaging **22**, 323–331.
Potter, J. R., Mellinger, D. K., and Clark, C. W. (**1994**). "Marine mammal call discrimination using artificial neural networks," J. Acoust. Soc. Am. **96**, 1255–1262.
Schevill, W. E., and Watkins, W. A. (**1965**). "Underwater calls of *Trichechus* (manatee)," Nature (London) **205**, 373–374.
Stafford, K., Fox, C. G., and Clark, D. S. (**1998**). "Long-range acoustic detection and localization of blue whale calls in the northeast Pacific Ocean," J. Acoust. Soc. Am. **104**, 3616–3625.
Steel, C. (**1982**). "Vocalization patterns and corresponding behavior of the West Indian manatee (Trichechus manatus)," Ph.D. dissertation, Biological Sciences, Florida Institute of Technology, Melbourne, FL.
Taswell, C. (**2000**). "The what, how, and why of wavelet shrinkage denoising," Comput. Sci. Eng. **2**, 12–19.
U.S. Fish and Wildlife Service (**2001**). "Florida manatee recovery plan," 5430 Gorsvenor Lane, Suite 110, Bethesda, MD 20814.
Weiss, L. G., and Dixon, T. L. (**1997**). "Wavelet-based denoising of underwater acoustic signals," J. Acoust. Soc. Am. **101**, 377–383.
Xu, Y., Weaver, J. B., Healy, D. M., Jr., and Lu, J. (**1994**). "Wavelet transform domain filters: A spatially selective noise filteration technique," IEEE Trans. Image Process. **3**, 747–758.
Yan, Z., Niezrecki, C., and Beusse, O. D. (**2005**). "Background noise cancellation for improved acoustic detection of manatee vocalizations," J. Acoust. Soc. Am. **117**, 3566–3573.
Yan, Z., Niezrecki, C., Cattafesta, L. N., III, and Beusse, O. D. (**2006**). "Background noise cancellation of manatee vocalizations using an adaptive line enhancer," J. Acoust. Soc. Am. **120**, 145–152.