A wavelet packet adaptive filtering algorithm for enhancing manatee vocalizations

M. Berke Gur[a)]
Department of Mechatronics Engineering, Bahcesehir University, Istanbul, Turkey

Christopher Niezrecki
Department of Mechanical Engineering, University of Massachusetts Lowell, Lowell, Massachusetts 01854

(Received 6 September 2010; revised 18 January 2011; accepted 19 January 2011)

Approximately a quarter of all West Indian manatee (*Trichechus manatus latirostris*) mortalities are attributed to collisions with watercraft. A boater warning system based on the passive acoustic detection of manatee vocalizations is one possible solution to reduce manatee–watercraft collisions. The success of such a warning system depends on effective enhancement of the vocalization signals in the presence of high levels of background noise, in particular, noise emitted from watercraft. Recent research has indicated that wavelet domain pre-processing of the noisy vocalizations is capable of significantly improving the detection ranges of passive acoustic vocalization detectors. In this paper, an adaptive denoising procedure, implemented on the wavelet packet transform coefficients obtained from the noisy vocalization signals, is investigated. The proposed denoising algorithm is shown to improve the manatee detection ranges by a factor ranging from two (minimum) to sixteen (maximum) compared to high-pass filtering alone, when evaluated using real manatee vocalization and background noise signals of varying signal-to-noise ratios (SNR). Furthermore, the proposed method is also shown to outperform a previously suggested feedback adaptive line enhancer (FALE) filter on average 3.4 dB in terms of noise suppression and 0.6 dB in terms of waveform preservation. © *2011 Acoustical Society of America*. [DOI: 10.1121/1.3557031]

PACS number(s): 43.60.Hj, 43.60.Mn, 43.30.Sf [EJS] Pages: 2059–2067

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. The manatee protection plan established by the United States Fish and Wildlife Service (FWS) in 1980 identified watercraft collisions as the most significant cause of manatee deaths. Accordingly, idle-speed and no-wake zones were designated throughout the Florida waterways where manatee–watercraft collisions are most likely to occur. Despite of measures taken, it is reported that the rate of manatee–watercraft collision related mortalities continue to remain high (U.S. Fish and Wildlife Service, 2001). Although the West Indian manatee was recently re-classified as a threatened species with a very high risk of extinction [Florida Fish and Wildlife Commission (FWC), 2007], data collected by the FWC indicate that watercraft related mortalities constitute approximately a quarter of all manatee mortalities (Florida Fish and Wildlife Conservation Commission, 2010). A factor that may contribute to the high rate of collisions is the lack of compliance of boaters to year-long and seasonal speed zones (Gorzelany, 2004), since these speed zones significantly increase travel times within the Florida channels.

The FWC announced two solicitations in 2001 and 2004 for developing systems, termed manatee avoidance technology, designed to prevent or reduce manatee–watercraft collisions. In response to these solicitations, several different approaches have been proposed for detecting manatees or warning manatees of a collision with watercraft. Among these, Gerstein *et al.* (2005) developed an active acoustic device which generates a narrowband, highly directional tone for warning manatees of an approaching watercraft. An active sonar based manatee detector was proposed by Jaffe *et al.* (2007). Another possible solution to reduce manatee–watercraft collisions by improving compliance rates to speed zones is an active boater warning system based on passive acoustic detection of manatee vocalizations and alerting nearby boaters of the presence of the animal. Niezrecki *et al.* (2003) demonstrated that a frequency domain passive acoustic energy detector is capable of satisfactorily detecting manatee vocalizations in moderate signal-to-noise ratio (SNR) measurements. However, early tests with the passive acoustic vocalization detector indicated that as noise levels increase relative to the vocalization source strength, a signal enhancement procedure becomes necessary prior to detection to achieve acceptable detection rates compared to the number of false alarms (Phillips *et al.*, 2006). In addition, once a detection is made, the vocalization signals may further be used for classification, localization, and related tasks. Such post-processing operations also require the enhancement of the vocalization signals prior to processing. In this paper, a single channel adaptive filtering algorithm implemented in the wavelet domain is presented and evaluated for enhancing manatee vocalizations.

A typical manatee vocalization lasts between 0.1 and 0.5 s, and may have several harmonics in the frequency band of 2–10 kHz with lower energy harmonics extending past 15 kHz. Like most marine mammal vocalizations, manatee

[a)]Author to whom correspondence should be addressed. Electronic mail: berke.gur@bahcesehir.edu.tr

vocalizations are non-stationary and may exhibit a complex time-frequency structure. While most manatee vocalizations possess a well defined harmonic structure, some vocalizations are transient acoustic energy bursts. The waveform and spectrogram of such a vocalization are depicted in Figs. 1 (a) and 1(b). A more detailed treatment of the signal properties of manatee vocalizations can be found in the papers of Phillips *et al*. (2004), Yan (2006), and the references therein. Several single channel and adaptive algorithms have been proposed in the literature for enhancing manatee vocalizations and other bioacoustic signals. Yan *et al*. (2005, 2006) implemented an adaptive line enhancer for enhancing manatee vocalizations. A wavelet domain *ad hoc* method for enhancing manatee vocalizations was developed by Gur and Niezrecki (2007). In this method, the amplitude of the wavelet coefficients obtained from the noisy vocalization signals was either zeroed or reduced after comparing to a threshold value (i.e., a non-linear soft thresholding scheme was used). Owing to the non-linear thresholding structure, the resulting algorithm was demonstrated to have very good noise suppression capabilities. However, the threshold value was empirically determined and fixed which prevented the algorithm from adapting to changes in the noise statistics. Furthermore, despite providing good noise suppression capabilities, the non-linear soft threshold rule resulted in clipping and local discontinuities in the vocalization waveform, reducing the audio quality of the vocalizations. A method for adaptively estimating the threshold for denoising the discrete wavelet transform (DWT) coefficients was proposed by Zhang (2001). Level-dependent thresholds were adapted to changing signal and noise statistics by minimizing the Stein's unbiased estimate (Stein, 1981) of the mean square error (MSE) cost function using a gradient descent optimization rule. The resulting algorithm was termed the thresholding neural network (TNN). More recently, Ren *et al*. (2008) proposed a wavelet domain non-linear adaptive filter for enhancing bioacoustic signals.

With this current paper, the TNN algorithm developed by Zhang (2001) is extended to the wavelet packet transform (WPT) domain and implemented for enhancing manatee vocalizations. Accordingly, the proposed method is referred to as the wavelet packet thresholding neural network (WP-TNN). The WP-TNN algorithm utilizes scale dependent and adaptive thresholds based on optimizing Stein's unbiased estimate of the MSE. The proposed algorithm is evaluated using real field recordings of vocalizations and background noise to estimate the improved detection range obtained with its use. Furthermore, the new algorithm overcomes some of the limitations of a previously proposed wavelet domain algorithm (Gur and Niezrecki, 2007) and is shown to outperform a linear adaptive line enhancer both in terms of noise suppression and waveform preservation.

This paper is organized as follows. In Sec. II, a theoretical treatment of wavelet domain denoising from an adaptive filtering perspective is provided and the proposed WP-TNN algorithm is introduced. Experimental evaluation of the proposed method using real measurements is presented in Sec. III. A simple but instructive analysis of the improvement in the detection range resulting from the proposed method is presented in Sec. IV. The paper is concluded with suggestions for future research in Sec. V.

II. THEORETICAL DEVELOPMENT

A. Wavelet domain denoising

The wavelet basis functions are very effective in compactly representing non-stationary signals that possess varying time-frequency energy distributions by using a few large amplitude coefficients. Noise, on the other hand, is mapped to the wavelet domain as numerous small amplitude

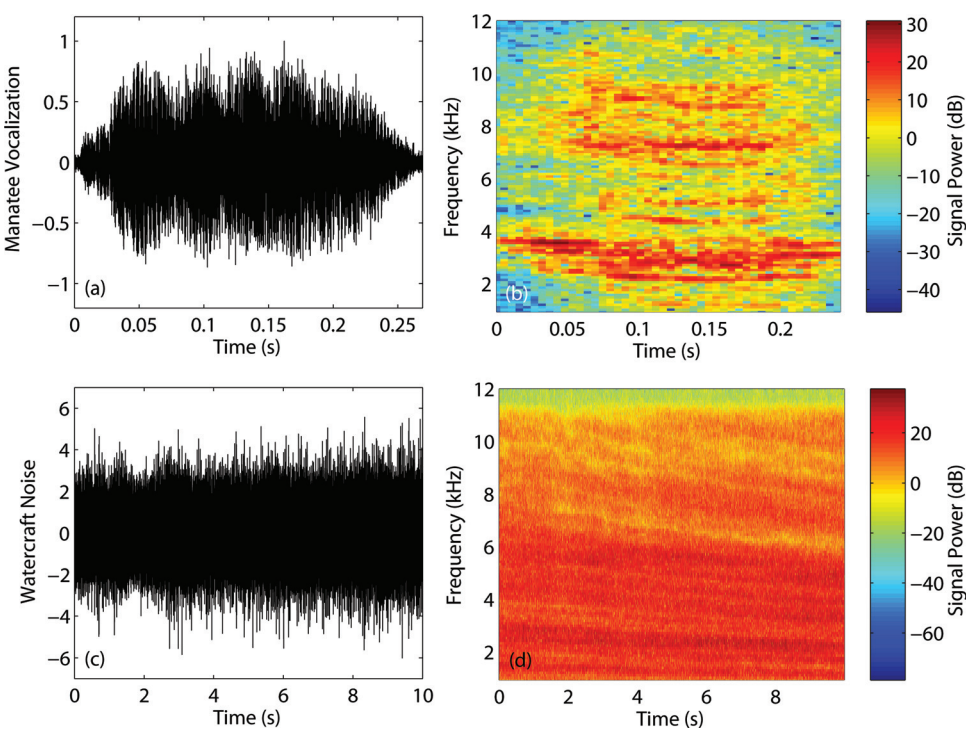

FIG. 1. (Color online) (a) and (c) Relative pressure-time domain plots and (b) and (d) the corresponding spectra of a manatee vocalization and watercraft emitted noise, respectively (due to high-pass filtering, lowest frequency plotted in spectra is 1 kHz).

coefficients. Thus, eliminating the small amplitude noise coefficients by thresholding forms the basis of wavelet domain denoising algorithms.

To explicitly formulate wavelet domain denoising, consider the problem of estimating the vocalization signal from a block of noisy measurement samples,

$$\mathbf{x} = \mathbf{s} + \mathbf{v}, \tag{1}$$

where $\mathbf{x} = [x(0) \quad x(1) \quad \cdots \quad x(N-1)]^T$ is the measurement vector, and the vectors $\mathbf{s}$ and $\mathbf{v}$ are similarly defined as the $N \times 1$ vocalization and noise vectors, respectively. The noise signal is assumed to be independent and identically distributed (IID) and zero mean Gaussian with unit variance. From a statistical perspective, the Gaussian noise assumption is justified in the presence of the dominant watercraft emitted noise (Gur, 2008).

Wavelet denoising can be visualized as a three-step, non-linear block filtering process which can conveniently be expressed using the matrix notation as a forward transform represented by the $N \times N$ orthogonal matrix $\mathcal{W}$, a diagonal threshold matrix $\mathbf{\Lambda}$, and the inverse transform represented by the orthogonal matrix $\mathcal{W}^{-1}$.

The forward transform will map the signals to the wavelet domain as

$$\mathbf{w}_x = \mathbf{w}_s + \mathbf{w}_v, \tag{2}$$

where $\mathbf{w}_x = \mathcal{W}\mathbf{x}$ is the $N \times 1$ vector of wavelet coefficients of the measurements. If the wavelet domain mapping of the vocalization signal is efficient, the wavelet coefficient vector will consist of a few large amplitude coefficients. The wavelet coefficients due to Gaussian noise, on the other hand, will be normally distributed with the same variance as the time domain noise signal. Essentially, denoising is performed by the diagonal filtering matrix $\mathbf{\Lambda}$ whose only non-zero terms are the non-linear threshold operators $\delta(n)$, $n = 0, 1, \ldots, N-1$. The entries $\delta(n)$ of the diagonal filtering matrix are generally defined as a function f (also termed the threshold rule) of a threshold parameter η and the corresponding wavelet coefficients $w_x(n)$ such that $\delta(n) = f(w_x(n), \eta(n))$. Hence, an estimate of the noise-free vocalization signal wavelet coefficients are obtained as

$$\hat{\mathbf{w}}_s = \mathbf{\Lambda} \cdot \mathcal{W} \cdot \mathbf{x}. \tag{3}$$

Once an estimate of the vocalization signal wavelet coefficients is obtained, the noise-free vocalization signal can be obtained from the inverse wavelet transform as $\hat{\mathbf{s}} = \mathcal{W}^{-1} \cdot \hat{\mathbf{w}}_s$.

The choice of the thresholding rule is detrimental in the performance of the denoising algorithm. An obvious cost function for measuring the denoising performance is the MSE defined as

$$\mathcal{J}_{\mathrm{MSE}} = E\left(\|\hat{\mathbf{w}}_s - \mathbf{w}_s\|_2^2\right). \tag{4}$$

Unfortunately, the MSE is an unpractical cost function because its computation requires the *a priori* knowledge of the clean vocalization signal wavelet coefficients $\mathbf{w}_s$. As an alternative, one can use Stein's unbiased estimate of the MSE (termed Stein's unbiased risk estimate, or SURE) defined as

$$\mathcal{J}_{\mathrm{SURE}} = N + E\left(\|\mathbf{g}(\mathbf{w}_x)\|_2^2 + 2\nabla_{w_x} \cdot \mathbf{g}(\mathbf{w}_x)\right), \tag{5}$$

where the estimator is assumed to be of the form $\hat{\mathbf{w}}_s = \mathbf{w}_x + \mathbf{g}(\mathbf{w}_x)$, the $N \times 1$ vector $\mathbf{g}(\mathbf{w}_x) = [g(w_x(0))g(w_x(1)) \cdots g(w_x(N-1))]^T$ is weakly differentiable and

$$\nabla_{w_x} \cdot \mathbf{g}(\mathbf{w}_x) = \sum_{n=0}^{N-1} \frac{\partial g(w)}{\partial w}\bigg|_{w=w_x(n)}. \tag{6}$$

The scale dependent (but non-adaptive) threshold value that minimizes the Stein's estimate of the MSE defined in Eq. (5) is given as

$$\eta_j = \arg\min_{\eta} \left[N - 2 \cdot \#\{n : |w_{x,j}(n)| \le \eta\} + \sum_{n=0}^{N-1} \min\left(|w_{x,j}(n)|, \eta\right)^2 \right], \tag{7}$$

where $\#\{n : \cdot\}$ denotes the number of arguments satisfying the relation given in the braces and j is the decomposition scale index.

B. The proposed method

Conventional SURE based denoising outlined in Sec. II A is a block processing approach in which a single scale dependent threshold [i.e., Eq. (7)] is used to denoise the wavelet coefficients in each decomposition level. In this block processing approach, the analysis window is selected long enough to ensure that the transient traits of the vocalization signals are captured. However, noise may also exhibit non-stationarity within this analysis window. Therefore, adapting the threshold to changing vocalization and noise statistics is necessary for achieving satisfactory MSE performance. For this purpose, the decomposition level specific threshold parameter η_j can be made to adapt to changing vocalization and noise statistics by updating it based on the current estimate of the MSE obtained from the SURE cost function defined in Eq. (5) (Zhang, 2001). Using a gradient descent optimization rule for the threshold update, the resulting scale dependent, adaptive threshold estimator has the form

$$\eta_j(n+1) = \eta_j(n) + \mu_j \frac{\partial}{\partial \eta_j} \mathcal{J}_{\mathrm{SURE}}(n), \; n = 0, 1, \ldots, N-1. \tag{8}$$

The gradient of the cost function $\partial \mathcal{J}_{\mathrm{SURE}}(n)/\partial \eta_j$ can be computed from Eq. (5) as

$$\frac{\partial}{\partial \eta_j} \mathcal{J}_{\mathrm{SURE}}(n) = 2g(w_{x,j}(n), \eta_j) \frac{\partial}{\partial \eta_j} g(w_{x,j}(n), \eta_j) + 2\frac{\partial^2}{\partial w_{x,j} \partial \eta_j} g(w_{x,j}(n), \eta_j), \tag{9}$$

where $g(w_{x,j}(n), \eta_j) = w_{x,j}(n) \cdot \left(\delta[w_{x,j}(n), \eta_j] - 1\right)$ is a non-linear function. Once the threshold parameter is computed

using Eqs. (8) and (9), the noisy wavelet coefficients are denoised using a threshold operator $\delta(w_{x,j}, \eta_j)$. For this purpose, the conventional soft threshold operator (Donoho and Johnstone, 1994) is modified as

$$\delta(w_{x,j}, \eta_j) = 1 + \frac{1}{2w_{x,j}}\Big([(w_{x,j} - \eta_j)^2 + \lambda]^{1/2} - [(w_{x,j} + \eta_j)^2 + \lambda]^{1/2}\Big), \tag{10}$$

where λ is the smoothness parameter to ensure that the gradients of Eq. (9) are non-zero (i.e., the adaptation does not stop) when $|w_{x,j}(n)| < \eta_j(n)$ (Zhang, 2001). When the modified soft threshold operator defined in Eq. (10) is used, the non-linear function becomes

$$g(w_{x,j}(n), \eta_j) = \frac{1}{2}\Big([(w_{x,j} - \eta_j)^2 + \lambda]^{1/2} - [(w_{x,j} + \eta_j)^2 + \lambda]^{1/2}\Big), \tag{11}$$

for which the gradients $\partial g(w_{x,j}, \eta_j)/\partial \eta_j$ and $\partial^2 g(w_{x,j}, \eta_j)/(\partial w_{x,j} \partial \eta_j)$ can be obtained in closed form. Thus, the SURE cost function $\mathcal{J}_{\mathrm{SURE}}$ has continuous derivatives with respect to both $w_{x,j}$ and η_j, enabling a smooth adaptation.

The temporal and spectral energy distribution of the noisy signals dictates the optimum partitioning of the time-frequency plane and the number of decomposition levels for wavelet domain denoising algorithms. The algorithm proposed in this paper is implemented using the orthogonal WPT. A full decomposition scheme is employed, resulting in a regular time-frequency plane partitioning. Although a regular partitioning is not optimal, no selective decomposition criteria was used to keep the algorithm simple. The use of scale dependent thresholds for denoising each subband is justified by the fact that the noise emitted from surface vessels exhibit some temporal correlation due to channel propagation. The SNR and SDR performance of the proposed algorithm 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 with a sufficient number of vanishing moments (i.e., oscillations). In addition, achieving high temporal resolution is also important since it is desired to suppress short duration transient signals such as snapping shrimp. The Daubechies family of wavelets possess the maximum number of vanishing moments for a given temporal support. Therefore, the WPT is implemented using the Daubechies-8 (db8) wavelet with seven levels of decomposition ($J = 7$). The block diagram of the resulting algorithm, termed the WP-TNN algorithm, is presented in Fig. 2.

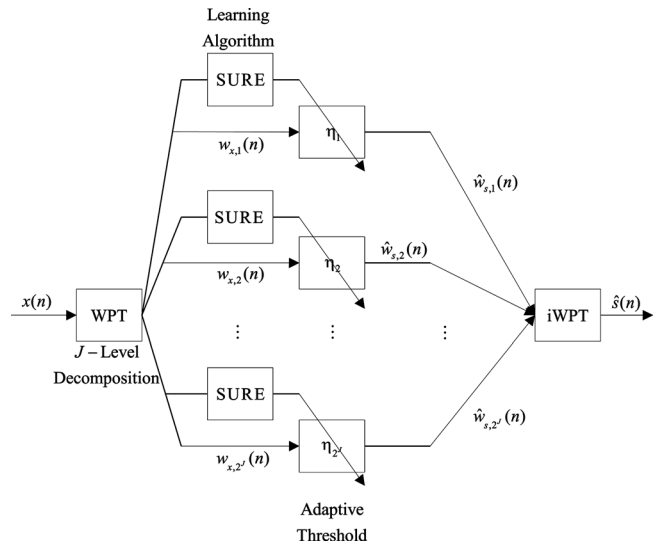

FIG. 2. The block diagram of the WP-TNN algorithm.

III. EXPERIMENTAL RESULTS

To evaluate the performance of the WP-TNN algorithm under realistic conditions, real vocalization samples and noise samples recorded at Crystal River, FL are used. The test location where the background noise data was collected is at close proximity to known manatee habitats and a busy waterway, and thus represents a pilot site where a manatee vocalization detector may potentially be implemented. Therefore, the results presented here are expected to be good indicators of how these algorithms are expected to perform in-field. A detailed map of the test site is shown in Fig. 3.

The feedback adaptive line enhancer (FALE) proposed by Yan *et al.* (2006) for enhancing manatee vocalizations is used as a benchmark to evaluate the performance of the WP-TNN algorithm. The WP-TNN and the FALE algorithms are evaluated using recordings constructed by superposing clean vocalization signals with the background noise signals measured at the test site. This approach is physically justified since small amplitude acoustic waves are additive. Both the noise and vocalization signals are first filtered with a tenth order Butterworth high-pass filter (with a 1 kHz cut-off frequency) before superposition to eliminate low frequency, high energy noise due to pressure waves, water current, and other natural noise sources that do not overlap with the vocalization harmonics. The input SNR is adjusted by changing the signal power of the vocalizations and is computed from these high-pass filtered vocalization and noise signals. Thus, the reported performance improvements in this section are in addition to what can be achieved by high-pass filtering alone. For input SNR below -10 dB after high-pass filtering, the noisy vocalization signal is severely distorted and none of the single channel algorithms are able to produce any useful output for detection. On the other hand, for input SNR above 5 dB, the local SNR at the harmonic frequencies are very high and most detectors will not require denoising to achieve a satisfactory detection performance. For this reason, the signal enhancement performances of the proposed algorithms are evaluated for input SNR in the range of -10 to 5 dB. The parameters used in implementing the WP-TNN and FALE algorithms, along with some observations regarding the performance of these algorithms are summarized in Table I.

The vocalizations employed for these tests were obtained from the vocalization library published by Yan (2006). This library is composed of ten different categories of vocalizations based on the structure of the harmonics of

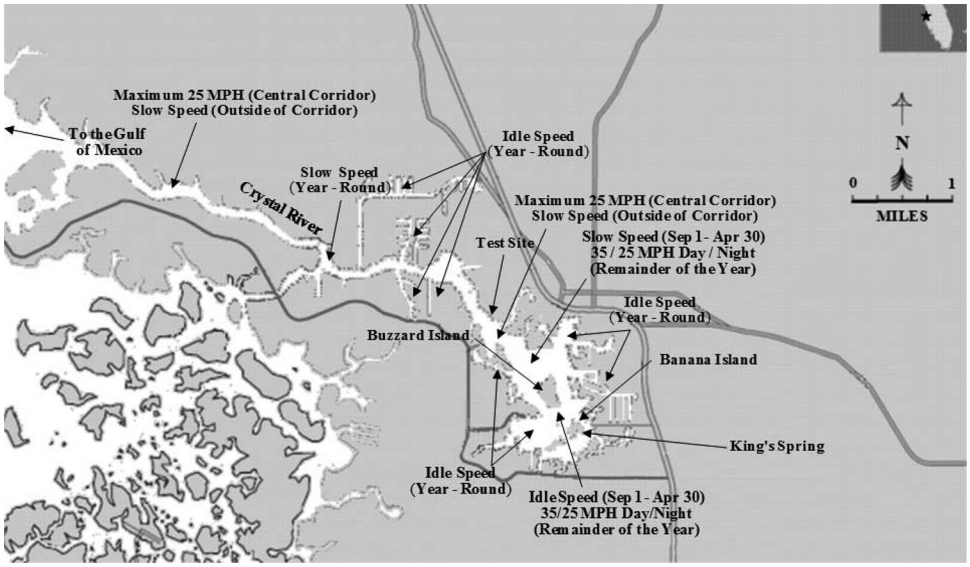

FIG. 3. The map of the Crystal River, FL noise recording test site indicating the surrounding manatee speed zones.

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. The vocalization and noise signals are sampled at 24 kHz. Typical background noise data that is dominated by watercraft emitted noise and recorded in Crystal River, FL is plotted in Figs. 1(c) and 1(d). Most of these noise signals are non-stationary over the duration of the vocalizations (due to the movement of an approaching surface vessel and other biological noise such as snapping shrimp) and exhibit temporal correlation (due to the convolutive and frequency selective transmission channel) with some narrowband non-vocalization harmonics (resulting from mechanical propeller shaft rotation) that could not be suppressed by high-pass filtering. Such traits make denoising vocalizations corrupted with these noise signals particularly challenging since the noisy signal has some temporal characteristics that are similar to the vocalizations to be detected.

Two performance measures are used to evaluate the two algorithms. The SNR performance measure is the ratio of the signal over the noise energy throughout the duration of the vocalization and is used to quantify the pre- and post-denoising quality of a noisy vocalization signal. The signal-to-distortion ratio (SDR) is defined as the ratio of the vocalization signal over the distortion power and is an indicator of how well the vocalization waveform is preserved after the denoising operation. Detailed derivations of these two performance measures can be found in Gur (2008) and Gur and Niezrecki (2009).

TABLE I. The parameters used in implementing the two signal enhancement algorithms.

Algorithm	Algorithm specific parameters	Remarks
WP-TNN	$J = 7$	May exhibit poor noise suppression
	$\mu = 0.005$, $\lambda = 1.0$	immediately after a vocalization
FALE	$\Delta = 50$, $L = 20$	Disproportional enhancement of harmonics
	$\mu = 0.005$, $\beta = 0.5$	distorts vocalization waveform

The output SNR and SDR results averaged over the ten manatee vocalizations (within each call type) and four different noise recordings for varying levels of noise are presented in Figs. 4 and 5, respectively. The output SNR and SDR results averaged over 100 manatee vocalizations and four different noise recordings are presented in Fig. 6. These results indicate that the WP-TNN algorithm achieves a slightly more consistent performance for the different

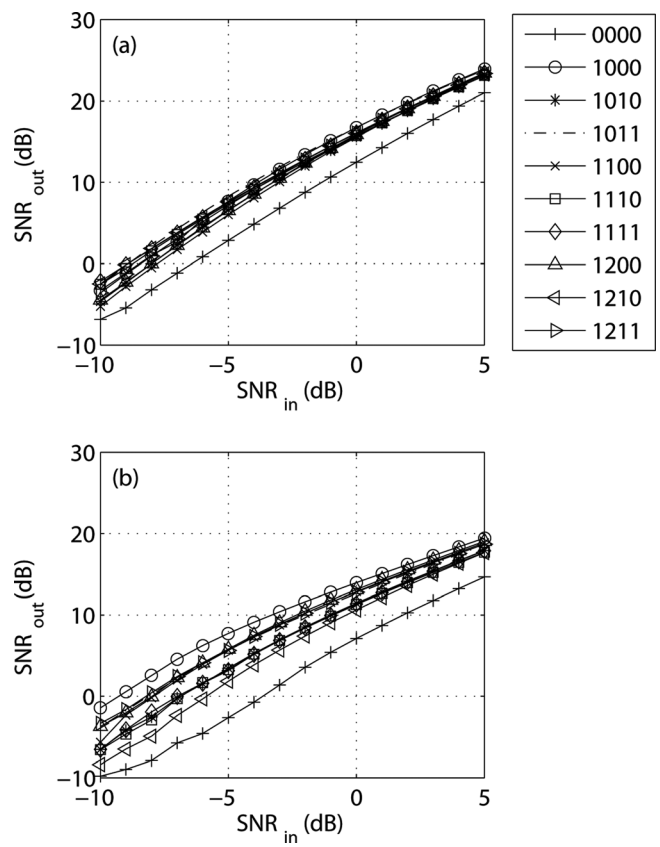

FIG. 4. The output SNR performance measures as a function of the input SNR for (a) the WP-TNN and (b) the FALE algorithms averaged over the four different noise recordings and ten vocalizations within each of the ten categories.

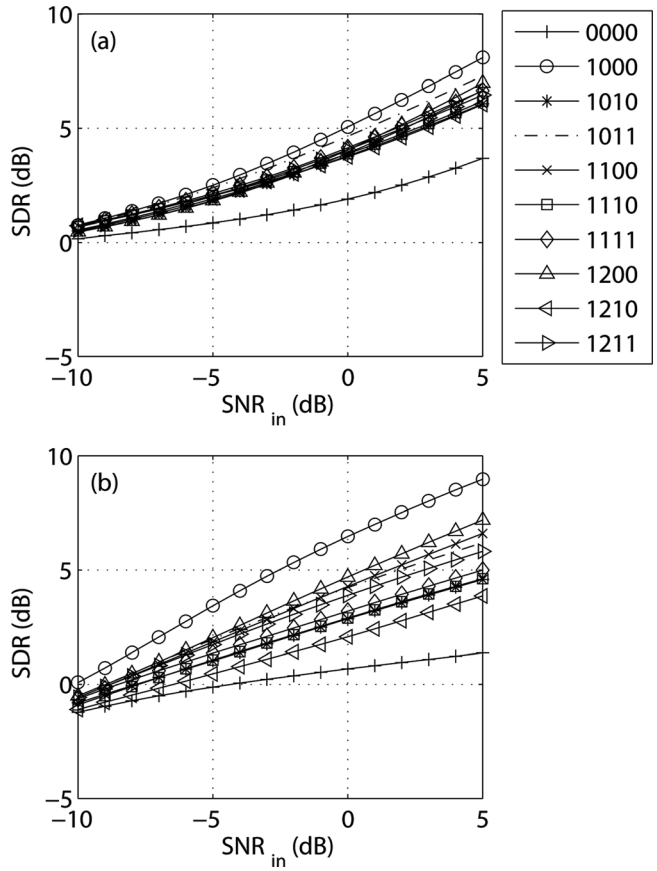

FIG. 5. The SDR performance measures as a function of the input SNR for (a) the WP-TNN and (b) the FALE algorithms averaged over the four different noise recordings and ten vocalizations within each of the ten categories.

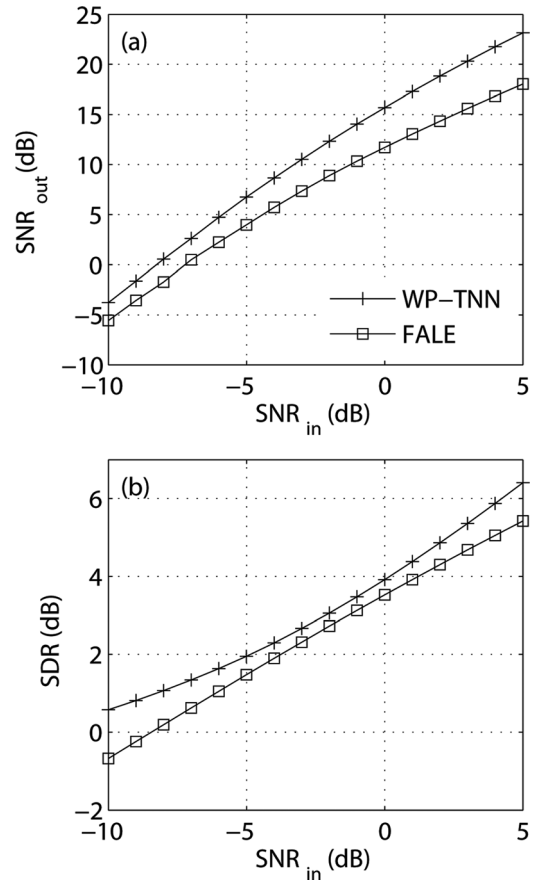

FIG. 6. The output SNR and SDR performance measures as a function of the input SNR for the two evaluated adaptive algorithms averaged over 100 manatee vocalizations and four different noise recordings.

vocalization categories and is more robust to different vocalization time-frequency structures compared to the FALE. The FALE achieves the best output SNR performance for the 1000 type vocalizations. The 1000 type vocalizations are characterized by a clear harmonic content and a single, consistent dominant harmonic, resulting in a higher local SNR at this dominant frequency and making them easier to detect. The two tested algorithms achieved the worst output SNR performance for the 0000 type vocalizations. The 0000 type vocalizations do not possess a clear harmonic structure and the vocalization energy is distributed over a wide frequency band (Yan, 2006). The wavelet transform is more appropriate for compactly representing such signals and the SNR performance of the WP-TNN algorithm is better for the 0000 type vocalization signals compared to the FALE algorithm for all four test cases. On average, the WP-TNN algorithm achieves a 3.4 dB improved noise suppression performance compared to the FALE algorithm. As it is in the case of output SNR, the WP-TNN algorithm is more robust to variations in the vocalization signal's time-frequency statistics, in terms of the SDR performance. Again, both algorithms achieved their best and worst SDR performance for the 1000 and 0000 type vocalizations, respectively. As is the case for the output SNR performance measure, the WP-TNN algorithm consistently resulted in higher SDR performance (on average, 0.6 dB) over the entire range of input SNR values compared to the FALE algorithm. WP-TNN and FALE algorithms result in denoised vocalization signals with similar audio quality. Typical time domain outputs of the evaluated denoising algorithms are presented in Fig. 7 for 0 dB input SNR.

IV. DETECTION RANGE IMPROVEMENTS

A simple detection range analysis based on the result outlined in Sec. III can provide an insight into the actual in-field performance of the WP-TNN algorithm. Following the approach presented in Gur and Niezrecki (2009), the signal enhancement performance of the WP-TNN algorithm is related to the improvements in the detection range through the passive sonar equation written in terms of the figure of merit (FOM)

$$\mathrm{FOM} = \mathrm{SL_m} - \max(\mathrm{NL_a}, \mathrm{SL_w} - \mathrm{TL_w}) + \mathrm{AG} - \mathrm{DT}, \qquad (12)$$

where $\mathrm{SL_m}$ is the source level of the vocalizations, $\mathrm{NL_a}$ is the ambient noise levels, $\mathrm{SL_w}$ is the boat source level, $\mathrm{TL_w}$ is the transmission loss associated with the watercraft, AG is the array gain, and DT is the detection threshold of the vocalization detector (all source levels are defined with reference to 1 μPa at a distance of 1 m from the source). The FOM represents the maximum allowable transmission loss; and hence, the maximum range at which the animal can be detected by the passive acoustic system. Assuming that the

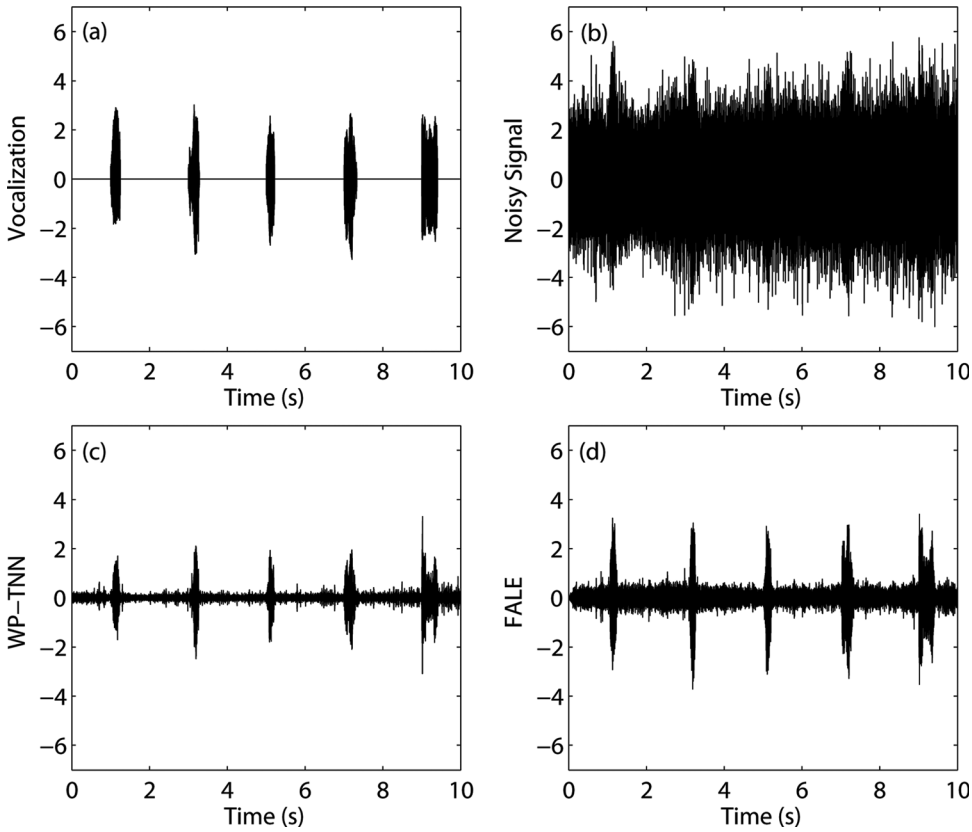

FIG. 7. (a) Five clean manatee vocalization in the time domain, (b) the noisy measurement with input SNR equal to 0 dB, and the time domain outputs of (c) the WP-TNN algorithm and (d) the FALE algorithm.

transmission loss obeys a mixed spreading model (Phillips *et al.*, 2006), the FOM can also be expressed as a function of the range of the manatee to the hydrophones

$$\mathrm{FOM} = 15\log_{10}(r_m). \tag{13}$$

Noting that the AG term in Eq. (12) is by definition zero for the single channel WP-TNN algorithm and solving for the range of the animal from Eqs (12) and (13) one obtains

$$r_{\mathrm{m}} = 10^{(\mathrm{SL_m} - \max(\mathrm{NL_a}, \mathrm{SL_w} - 15\log_{10}(r_{\mathrm{w}})) - \mathrm{DT})/15}. \tag{14}$$

The improvement in SNR resulting from processing the noisy vocalizations is incorporated into Eq. (14) as a decrease in the DT. The ratio of the detection range resulting from the WP-TNN algorithm to high-pass filtering detection range can be computed using the relation,

$$\frac{(r_{\mathrm{m}})_{\mathrm{WP-TNN}}}{(r_{\mathrm{m}})_{\mathrm{HPF}}} = 10^{(-\mathrm{DT_{WP-TNN}} + \mathrm{DT_{HPF}})/15}, \tag{15}$$

where $(r_{\mathrm{m}})_{\mathrm{WP-TNN}}$, $(r_{\mathrm{m}})_{\mathrm{HPF}}$, and $\mathrm{DT_{WP-TNN}}$ and $\mathrm{DT_{HPF}}$ are the detection ranges after WP-TNN and high-pass filtering, and the detection thresholds (DT) associated with WP-TNN and high-pass filtering, respectively. The ratio of the detection ranges resulting from the WP-TNN and the FALE algorithms over high-pass filtering for the four Crystal River test cases are presented in Fig. 8. In the upper end of the input SNR range, the WP-TNN algorithm can improve the detection range by a factor of 10–16. As the input SNR is reduced to −10 dB or below, none of the proposed algorithms can improve the detection range significantly over the high-pass filter.

Next, the detection ranges resulting from processing the vocalizations with the WP-TNN algorithm are presented using an example. The watercraft noise and manatee vocalization source levels are assumed to be $\mathrm{SL_w} = 140$ dB and $\mathrm{SL_m} = 118$ dB, respectively. If ambient noise is assumed to be $\mathrm{NL_a} = 70$ dB, based on the noise term in Eq. (12), watercraft emitted noise will dominate the background noise environment within the feasible detection range of the vocalization detector. If the ambient noise levels are increased to $\mathrm{NL_a} = 100$ dB, the watercraft emitted noise will still

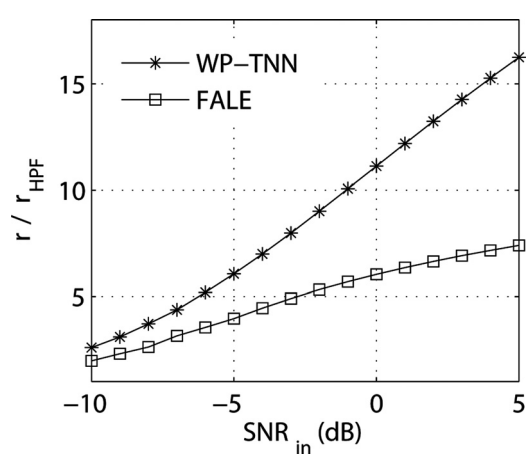

FIG. 8. The average ratio of the detection range resulting from the WP-TNN and FALE algorithms to high-pass filtering for the four Crystal River tests. A unity ratio indicates no improvement in detection range.

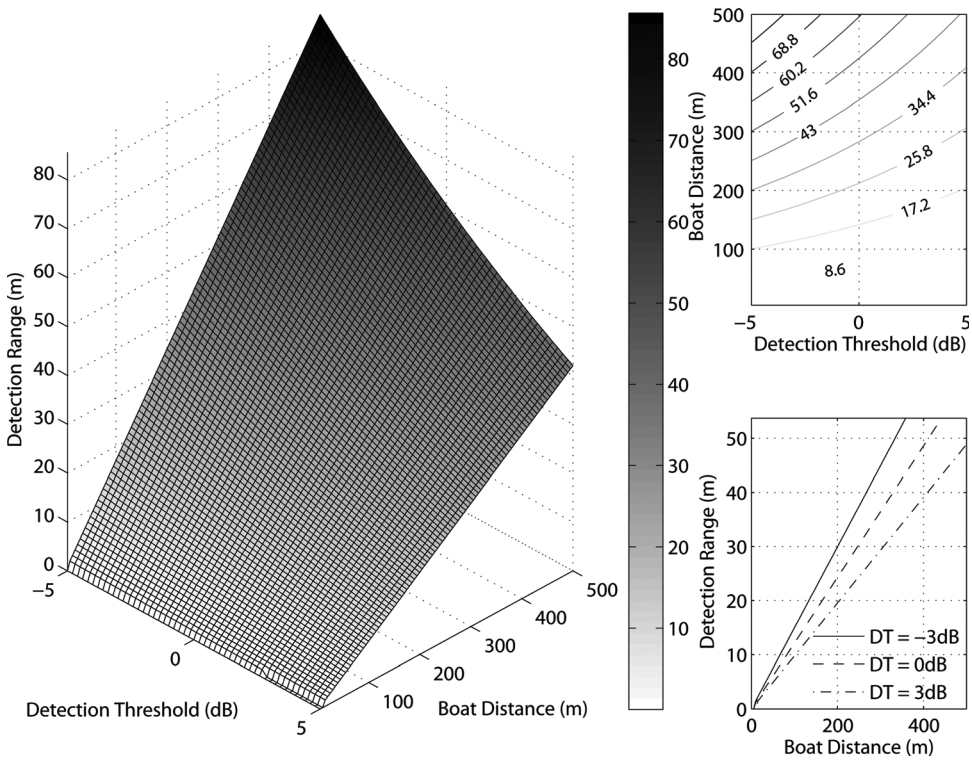

FIG. 9. Manatee vocalization detection ranges after processing with the WP-TNN algorithm in the presence of watercraft emitted noise.

dominate the noise levels provided that the watercraft is within 464.2 m of the receivers. The experimentally obtained performance measures presented in Sec. III are computed from recordings made in an environment where the ambient background noise levels are significantly lower (−12 dB or lower) compared to the sound levels in the presence of watercraft. For that reason, the ambient noise level and DT are assumed to be 70 dB and 3 dB, respectively.

In the above described scenario, if the watercraft is $r_w = 500$ m away, using Eq. (14) the detection range is computed to be limited at $r_m = 10.8$ m. If the watercraft moves to a range of 250 m, the maximum detection range of manatee vocalizations drops to 5.4 m. It becomes impossible to detect manatee vocalizations (irrespective of the distance of the manatee to the receivers) if the watercraft is closer than 46.4 m.

For improved detection range estimations (obtained after processing the noisy vocalizations with the WP-TNN algorithm), the DT in the case of high-pass filtering corresponds to the output SNR of the WP-TNN algorithm, and the input SNR corresponds to the improved DT obtained with the WP-TNN algorithm. To evaluate the improved detection ranges, it is necessary to compute the input SNR corresponding to a given output SNR. However, the output SNR for the WP-TNN algorithm which is estimated through experiments in Sec. III is available only at non-equally spaced discrete values. To obtain an analytic relation between the input SNR and the output SNR for the WP-TNN algorithm, the output SNR averaged over the four Crystal River test cases is curve fit with a second order polynomial,

$$\mathrm{SNR}_{\mathrm{out}} = -0.0314 \cdot (\mathrm{SNR}_{\mathrm{in}})^2 + 1.64 \cdot (\mathrm{SNR}_{\mathrm{in}}) + 15.7038, \tag{16}$$

in the least square sense (Yan, 2006). From Eq. (16), the minimum input SNR necessary for a DT of 3 dB for the WP-TNN algorithm is computed as −6.8 dB. Thus, with the WP-TNN and a detector with a DT of 3 dB combined, the vocalizations can be detected at an input SNR of −6.8 dB or higher. From the detection range ratios provided in Fig. 7, the average improvement ratio in the detection range for the WP-TNN algorithm can be inferred to be in the range of 3.2–6.5 for an input SNR of −6.8 dB. Using Eq. (12), a more precise value for the ratio $r_{\mathrm{WP-TNN}}/r_{\mathrm{HPF}}$ is computed as 4.5. If the watercraft is at a distance of 500 m, the improved detection range will become 48.5 m (compare to 10.8 m with high-pass filtering) for a DT of 3 dB. If the watercraft comes within 250 m of the receivers, the improved detection range will be 24.4 m (compare to 5.4 m with high-pass filtering). The maximum range of the watercraft to the receivers for which manatee vocalizations cannot be detected, irrespective of the range of the manatee, improves to 10.3 m. (compare to 46.4 m with high-pass filtering). The improved detection ranges are depicted in Fig. 9 for the scenario described above.

V. CONCLUSIONS

Within this paper, a wavelet domain adaptive filtering algorithm is described for enhancing noisy manatee vocalizations. First, a theoretical treatment of wavelet domain denoising is introduced from an adaptive filtering perspective. Next, the proposed WP-TNN algorithm is presented. The WP-TNN algorithm is implemented and tested using real manatee vocalizations and background noise recordings dominated with watercraft noise in which the input SNR is varied from −10 to 5 dB. The performance of the algorithm is evaluated in terms of the SNR and SDR performance

measures. These experiments suggest that the proposed method at least doubles the detection range for low input SNR (approximately -10 dB) and can extend it up to 16 times for high input SNR (approximately 5 dB) compared to high-pass filtering alone. Hence, the detection range of manatee vocalizations can significantly be improved when the vocalization signals are pre-processed with the WP-TNN algorithm, which will be pivotal in the practical implementation of a passive acoustic manatee vocalization detector.

The proposed WP-TNN algorithm incorporates an adaptive thresholding scheme enabling it to adapt to different vocalization signals and changing ambient conditions. In addition, the proposed method is linear and the outputs obtained are free from adverse effects of non-linear processing such as clipping. Hence, the WP-TNN algorithm overcomes some of the limitations of a previously proposed wavelet domain algorithm. Furthermore, the WP-TNN method is shown to outperform a linear adaptive line enhancer, in terms of both noise suppression and waveform preservation when evaluated using real vocalization and watercraft emitted noise measurements.

Several improvements to the proposed WP-TNN method are worth further investigation. The wavelet coefficients obtained from a full decomposition at level $J = 7$ are used for the WP-TNN algorithm, which results in a regular decomposition of the time-frequency plane. A more appropriate decomposition structure (one that perhaps results in an irregular time-frequency grid) can be obtained based on the signal properties or some other criteria (e.g., see Ren *et al.*, 2008), reducing the computational load of the algorithm. The cost function used to adaptively compute the level dependent threshold values was Stein's unbiased estimate of the MSE. Alternative cost functions; in particular, those that are more suitable for non-Gaussian noise may extend the applicability of the proposed method. Wavelet domain denoising methods are not confined to single channel approaches (Rao and Jones, 2000) and extensions of the proposed method to multi-channels can result in viable alternatives to conventional adaptive beamforming. Finally, wavelet coefficients are known to be robust features for transient signal detection (e.g., see Carevic, 2005). A natural extension of this work would be to investigate vocalization detectors that incorporate a detection test statistic defined in terms of the wavelet coefficients.

ACKNOWLEDGMENTS

The authors would like to express their sincere appreciation to the Florida Sea Grant, Florida Fish and Wildlife Conservation Commission, University of Florida Marine Mammal Program, and the University of Massachusetts Lowell Provost's Office for supporting this research.

Carevic, D. (**2005**). "Adaptive window-length detection of underwater transients using wavelets," J. Acous. Soc. Am. **117**(5), 2904–2913.

Donoho, D. L., and Johnstone, I. M. (**1994**). "Ideal spatial adaptation by wavelet shrinkage," Biometika **81**(3), 425–455.

Florida Fish and Wildlife Commission (**2007**). *Florida Manatee Management Plan* (Florida Fish and Wildlife Commission Tallahassee, FL), 280 pp.

Florida Fish and Wildlife Conservation Commission (**2010**). *Manatee Mortality Statistics* (1974–2009), http://research.myfwc.com/features/category_sub.asp?id=2241 (Last viewed August 2, 2010).

Gerstein, E. R., Blue, J. E., and Forysthe, S. E. (**2005**). "The acoustics of vessel collisions with marine mammals," in Proc. IEEE/MTS Oceans 2005 Conf. **2**, 1190–1197.

Gorzelany, J. F. (**2004**). "Evaluation of boater compliance with manatee speed zones along the Gulf Coast of Florida," Coastal Manage. **32**, 215–226.

Gur, B. M. (**2008**). "Adaptive enhancement of marine mammal vocalizations," Ph.D. dissertation, Department of Mechanical Engineering, University of Massachusetts, Lowell, MA.

Gur, B. M., and Niezrecki, C. (**2007**). "Autocorrelation based denoising of manatee vocalizations using the undecimated discrete wavelet transform," J. Acous. Soc. Am. **122**(1), 188–199.

Gur, B. M., and Niezrecki, C. (**2009**). "A source separation approach to enhancing marine mammal vocalizations," J. Acoust. Soc. Am. **126**(6), 3062–3070.

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**(1), 158–165.

Niezrecki, C., Phillips, R., Meyer, M., and Beusse, D. O. (**2003**). "Acoustic detection of manatee vocalizations," J. Acoust. Soc. Am. **114**(3), 1640–1647.

Phillips, R., Niezrecki, C., and Beusse, D. O. (**2004**). "Determination of West Indian manatee vocalization levels and rate," J. Acous. Soc. Am. **115**(1), 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**(**1**), 153–163.

Rao, A. M., and Jones, D. L. (**2000**). "A denoising approach to multisensory signal enhancement," IEEE Trans. Signal Process. **48**(5), 1225–1234.

Ren, Y., Johnson, M. T., and Tao, J. (**2008**). "Perceptually motivated wavelet packet transform for bioacoustic signal enhancement," J. Acous. Soc. Am. **124**(1), 316–327.

Stein, C. M. (**1981**). "Estimation of the mean of a multivariate normal distribution," Ann. Statis. **9**(6), 1135–1151.

U.S. Fish and Wildlife Service (**2001**). *Florida Manatee Recovery Plan (Trichechus manatus latirostris)* (U.S. Fish and Wildlife Service, Bethesda, MD), 3rd revision, 193 pp.

Yan, Z. (**2006**). "Background noise cancellation for acoustic detection of manatee vocalizations," Ph.D. dissertation, Department of Mechanical Engineering, University of Florida, Gainesville, FL.

Yan, Z., Niezrecki, C., Cattafesta III, L. N., and Beusse, D. O. (**2005**). "Background noise cancellation for improved acoustic detection of manatee vocalizations," J. Acous. Soc. Am. **117**(6), 3566–3573.

Yan, Z., Niezrecki, C., Cattafesta III, L. N., and Beusse, D. O. (**2006**). "Background noise cancellation of manatee vocalizations using an adaptive line enhancer," J. Acous. Soc. Am. **120**(1), 145–152.

Zhang, X.-P. (**2001**). "Thresholding neural network for adaptive noise reduction," IEEE Trans. Neural Networks **12**(3), 567–584.