A source separation approach to enhancing marine mammal vocalizations

M. Berke Gur[a)]
Department of Mechatronics Engineering, Bahcesehir University, Besiktas, Istanbul 34353, Turkey

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

(Received 10 September 2009; accepted 2 October 2009)

A common problem in passive acoustic based marine mammal monitoring is the contamination of vocalizations by a noise source, such as a surface vessel. The conventional approach in improving the vocalization signal to noise ratio (SNR) is to suppress the unwanted noise sources by beamforming the measurements made using an array. In this paper, an alternative approach to multi-channel underwater signal enhancement is proposed. Specifically, a blind source separation algorithm that extracts the vocalization signal from two-channel noisy measurements is derived and implemented. The proposed algorithm uses a robust decorrelation criterion to separate the vocalization from background noise, and hence is suitable for low SNR measurements. To overcome the convergence limitations resulting from temporally correlated recordings, the supervised affine projection filter update rule is adapted to the unsupervised source separation framework. The proposed method is evaluated using real West Indian manatee (*Trichechus manatus latirostris*) vocalizations and watercraft emitted noise measurements made within a typical manatee habitat in Florida. The results suggest that the proposed algorithm can improve the detection range of a passive acoustic detector five times on average (for input SNR between −10 and 5 dB) using only two receivers. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3257549]

PACS number(s): 43.60.Fg, 43.30.Sf, 43.60.Mn [EJS] Pages: 3062–3070

I. INTRODUCTION

A frequently encountered problem in passive acoustic based marine mammal monitoring is the enhancement of vocalization signals in the presence of other interfering sources and ambient noise. The presence of dominant noise sources reduces the signal-to-noise ratio (SNR) of the measurements, can degrade detection and classification performance, or reduce the effective range of a passive acoustic monitoring system. A feasible implementation of a passive acoustic based monitoring system operating in noisy environments generally requires the enhancement of the vocalization signals. One such application related to marine mammals is the enhancement of manatee vocalizations for more effective passive acoustic based detection in the presence of recreational watercraft. In this paper, a two-channel second-order-statistics (SOS) based blind source separation (BSS) approach is developed and evaluated for enhancing manatee vocalizations.

The West Indian manatee (*Trichechus manatus latirostris*) was added to the endangered species list in 1967. In 1980, the U.S. Fish and Wildlife Service established a manatee protection plan. Within this plan, collisions with recreational boats and other watercraft were identified as the most significant cause of unnatural manatee mortalities. Accordingly, idle-speed and no-wake zones have been designated throughout the shallow Florida waterways where manatee-watercraft collisions are most likely to occur. However, a 2001 recovery review revealed that the rate of manatee-watercraft collision related mortalities continued to remain high despite measures taken.[1] Although the West Indian manatee was recently re-classified as a threatened species with a very high risk of extinction,[2] data collected by the Florida Fish and Wildlife Conservation Commission indicate that watercraft related mortalities remain steady at 25% among all manatee mortalities.[3] Several reasons for the ineffectiveness of the speed zones have been put forth in the recent literature. 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 since these speed zones significantly increase travel times within the Florida channels. In a survey conducted at 15 sites in Florida, overall compliance rates to speed zones were reported as 58% and 63%.[4]

One possible solution to improve 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. 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 [see Figs. 1(a) and 1(b)]. Detailed information on manatee vocalizations can be found in the works of Steel,[5] Nowacek *et al.*,[6] Phillips *et al.*,[7] and the references therein. It was shown by Niezrecki *et al.*[8] that a frequency domain energy detector is capable of satisfactorily detecting manatee vocalizations in moderate SNR measurements. However, as the noise levels increase relative to the vocalization source strength, a signal enhancement procedure becomes necessary prior to detection.[9]

In general, fluctuations in the underwater ambient noise statistics warrant adaptive signal enhancement algorithms. Several single channel, adaptive algorithms have been pro-

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

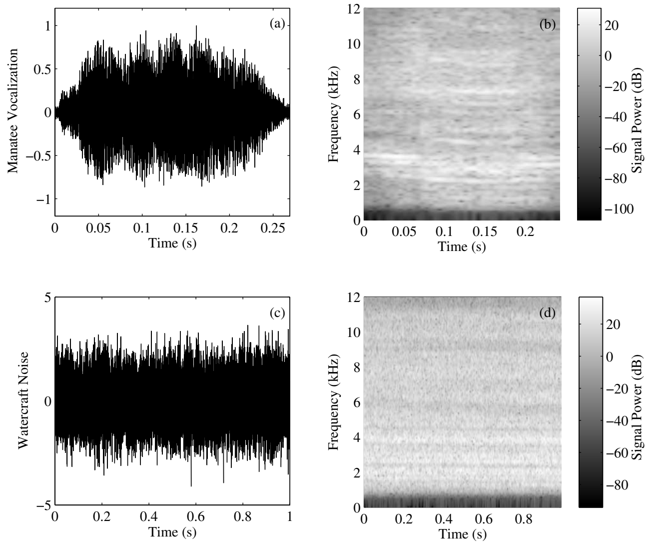


FIG. 1. [(a) and (c)] Time domain plots and [(b) and (d)] spectra of a manatee vocalization and watercraft emitted noise, respectively (after high-pass filtering).

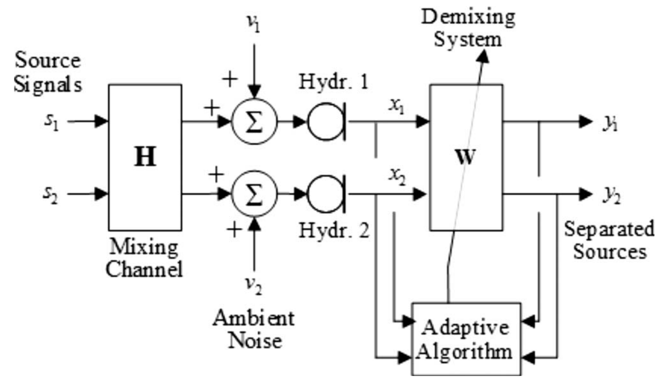


FIG. 2. The typical setup of a two-channel BSS problem.

posed for enhancing manatee vocalizations and other bioacoustic signals. Yan *et al.*[10,11] proposed an adaptive line enhancer for enhancing manatee vocalizations. An adaptive wavelet domain *ad-hoc* method for enhancing manatee vocalizations was developed by Gur and Niezrecki.[12] More recently, Ren *et al.*[13] proposed a wavelet domain non-linear adaptive filter for enhancing bioacoustic signals. The enhancement performances of conventional single channel adaptive filtering methods drastically decline as the SNR of the measurements decreases. However, if a significant portion of background noise is emitted from a point noise source such as a surface vessel, signal enhancement performance can substantially be improved through multi-channel adaptive algorithms.

The conventional multi-channel approach to signal enhancement is beamforming through which measurements from a uniform linear array (ULA) are constructively combined to enhance the signal incident from the target source location.[14] Adaptive beamforming algorithms can also be designed to suppress a dominant point noise source by placing a null in the corresponding incidence direction. Several passive acoustic based detection, localization, and classification systems that incorporate beamforming are described in the literature.[15,16] However, beamforming has some important drawbacks. It requires the precise knowledge of the target and noise source locations. In general, the location of the sources is not known a priori, and the direction of arrival (DOA) estimation for the vocalization signals must be performed. The estimation of DOA and the related problem of time difference of arrival (TDOA) estimation from noisy manatee vocalization recordings were investigated by Muanke and Niezrecki[17] in the context of source localization. The cited authors concluded that an input SNR of 8 dB or higher was necessary on all the input channels for accurately estimating the TDOA. Alternatively, the DOA can be estimated using blind sub-space algorithms such as the multiple signal classification algorithm or even by scanning each radial angle for high power incoming acoustic signals. However, the signal power of weak vocalization signals will generally be too low to determine the corresponding DOA using such methods. The ability of a beamformer to resolve the location of the target is inversely proportional to the length of the discrete aperture formed by the ULA. Combined with sensor spacing restrictions necessary to avoid spatial aliasing, enhancing marine mammal vocalizations requires an array that consists of many (on the order of 10 or more) hydrophones. Hence, beamforming is not a cost-effective solution to enhance manatee vocalizations in the numerous manatee idle-speed/no-wake zones within the Florida waterways.

As an alternative, BSS is a class of adaptive signal processing algorithms that serve for retrieving the original signals emitted from multiple point sources from multi-channel mixtures. These algorithms are referred to as “blind” because both the source signals and the mixing channels are assumed to be unknown. The signals emitted from multiple acoustic point sources are assumed to be statistically independent, which can be physically justified. Due to transmission through a multi-input, multi-output mixing channel, the acoustic signals measured at the receivers become statistically dependent. The original sources can be extracted from the measurements by solving for a separating solution that makes the multi-channel measurements statistically independent. The separated sources are retrieved from the measurements using unsupervised adaptive filtering (see Fig. 2 for a two-channel setup). BSS algorithms can be separated into several different groups based on the implementation of the statistical independence criterion. Most of these algorithms utilize the known or estimated probability density function (PDF) and/or higher-order-statistics (HOS) of the acoustic source signals to achieve source separation. Unfortunately, accurate estimates of the source PDF and HOS of weak vocalizations from noisy measurements are very difficult to obtain. However, a separate group of SOS based algorithms achieves source separation by making the outputs uncorrelated, which result in less complex and more robust algorithms for two-channel setups.

Source separation, particularly its applications in speech enhancement and communications, is a topic that has received high interest among researchers over the past decade (for a concise review, see Ref. 18). In the context of signal enhancement, separation of the vocalization and noise signals suggests an improvement in the SNR of the vocalization signal. Therefore, BSS algorithms are versatile methods for enhancing marine mammal (e.g., manatee) vocalizations corrupted by an interfering point source such as watercraft navigating through the channel. Despite the extensive literature on speech and communications source separation, only a few studies implement source separation in the context of underwater acoustics. Gaeta *et al.*[19] suggested a HOS based blind separation of artificially mixed underwater acoustic signals. They numerically computed the channel impulse responses using ray propagation theory. Bonnifay *et al.*[20] incorporated prior knowledge of the channel impulse responses in a HOS based source separation algorithm and experimented with artificially mixed underwater communication signals. More recently, Mansour *et al.*[21] investigated blind separation of underwater acoustic signals (including artificial mixtures of ship noise and whale vocalizations) for passive acoustic tomography and reported that SOS based frequency domain iterative algorithms exploiting the non-stationarity of the source signals resulted in better separation performance.

In this study, a new SOS based BSS algorithm for enhancing marine mammal vocalizations is proposed and evaluated using real vocalization and watercraft emitted noise measurements. This paper is organized as follows. A novel SOS-BSS algorithm based on the affine projection (AP) filter update rule is proposed in Sec. II. The signal enhancement performance of the proposed SOS-BSS algorithm is evaluated using real measurements in Sec. III. The improvements in the detection range resulting from preprocessing the measurements with these algorithms are presented in Sec. IV. Lastly, the concluding remarks are provided in Sec. V.

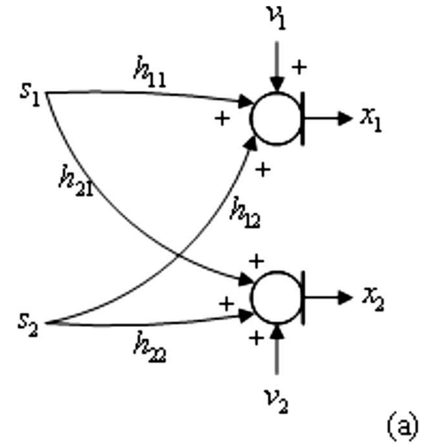
II. THEORETICAL DEVELOPMENT

A. The underwater acoustic channel model

The mixing channel model is depicted in Fig. 3(a) where the two measurements $\mathbf{X}=[x_1\ x_2]^T$ are a result of a linear mixture of the two sources $\mathbf{S}=[s_1\ s_2]^T$ (assuming that the s_1 is the vocalization signal) and ambient noise $\mathbf{V}=[v_1\ v_2]^T$,

$$\mathbf{X} = \mathbf{H} * \mathbf{S} + \mathbf{V}, \qquad (1)$$

where $*$ represents time domain convolution, and $\mathbf{H}$ is a matrix of finite impulse response (FIR) filters modeling multipath channel transmission. In the context of source separation, the source signals are assumed to be more dominant compared to ambient noise. The channel mixing process can be separated into two stages, as shown in Fig. 3(b); the propagation of the source signals through the convolutive channel is represented by the impulse responses h_{11} and h_{22}, and the cross-channel mixing of the sources is represented by the pseudochannel impulse responses $\tilde{h}_{12}=h_{12}/h_{11}$ and $\tilde{h}_{21}=h_{21}/h_{22}$. Simple decorrelation can be used to extract the source signals $\tilde{s}_1$ and $\tilde{s}_2$, but is not sufficient to retrieve the original sources s_1 and s_2. Additional known properties of the source signals such as non-stationarity[22] or temporal correlation[23] need to be exploited to perform simultaneous separation and deconvolution. Alternatively, deconvolution can be performed postseparation.[24] Although $\tilde{s}_1$ is a filtered version of the original vocalization signal, it is sufficient to extract this filtered signal for detection purposes.



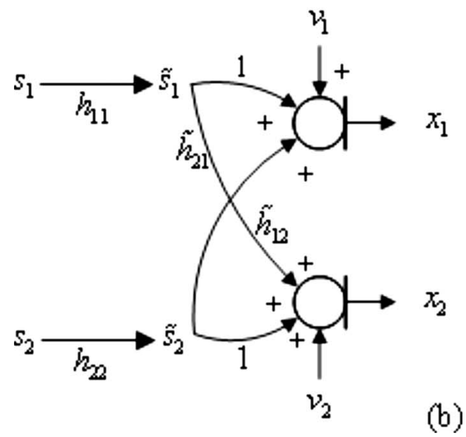


FIG. 3. (a) The two-channel transmission model and (b) the channel model with channel propagation and mixing stages decoupled.

B. Second-order-statistics based blind source separation

Decorrelation based algorithms employ a SOS based cost function of the de-mixing system outputs for achieving source separation. Adaptively decorrelating the outputs

$$E[y_p(n)y_q(n-k)] = r_{y_p y_q}(n,k) = 0, \quad p \neq q = 1,2, \qquad (2)$$

where $E[\cdot]$ is the expectation operator, y_p is the output at the pth channel and $r_{y_p y_q}(n,k)$ is the cross correlation coefficient between the outputs at time n and lag k, is anticipated to result in source separation. Assuming a two-channel mixing system, a cost function that uses the instantaneous estimate of the correlation between the two outputs at L lags combined with a stochastic gradient update rule results in the SOS based symmetric adaptive decorrelator (SAD) algorithm[25] with a filter update structure similar to the least mean squares (LMS) algorithm. The SAD algorithm is a simple and robust BSS algorithm suitable for source separation in the two-input, two-output configuration.

However, the SAD algorithm also inherits the limitations of the LMS algorithm. Due to the large eigenvalue spread of the input autocorrelation matrix, the convergence

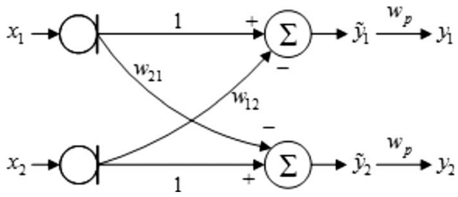


FIG. 4. The block diagram of the two-channel feedforward structured de-mixing system.

of the filter coefficients is rather slow in the presence of temporally correlated inputs (such as the harmonic manatee vocalizations that require small step sizes to prevent the separating filters from destabilizing).[26] The small step sizes required for stability reduce the ability of the de-mixing filters to adapt to changes in the noise signal statistics or to track the changes in the channel impulse responses due to moving watercraft. One possible solution to improve the convergence rate of the SAD algorithm is to update the adaptive filters using the recursive least squares (RLS) algorithm.[27] The resulting double-RLS (DRLS) algorithm eliminates the large eigenvalue spread of the filter inputs by pre-multiplying the inputs with a recursive estimate of the inverse of the cross-correlation matrix between the inputs and filtered outputs. However, the DRLS algorithm is prone to the same divergence phenomenon observed in supervised RLS algorithms caused by numerical instabilities.[28] To further aggravate the problem, the use of the input and output cross-covariance matrices in the DRLS algorithm (in contrast to the Toeplitz input autocorrelation matrix, in case of the supervised RLS algorithm) prevents fast and numerically more stable QR-decomposition based implementations.

C. The proposed method

The AP algorithm[29] was proposed in the context of supervised adaptive filtering as an intermediate algorithm between the LMS and RLS algorithms. Instead of the inverse of the full order $L\times L$ input autocorrelation matrix (where L is the filter order), the AP algorithm uses the inverse of a $K\times K$ lower order partial autocorrelation matrix (where $K<L$) to pre-whiten the input data. In this paper, an unsupervised BSS algorithm based on the supervised AP algorithm is derived and implemented for the two-channel configuration. The extension of the supervised AP algorithm to the unsupervised adaptive filtering framework was first proposed by Gabrea.[30] The cited author proposed a feedback structured block-update double affine projection (DAP) algorithm in the context of two-channel speech enhancement. However, the feedback structure may cause instability. In addition, the $K\times K$ partial autocorrelation matrix will always be incomplete due to the block-update restriction, leading to degraded separation performance. Hence, to circumvent these problems, a feedforward (FF) implementation of the DAP algorithm with sequential update (SU) of the filter coefficients is proposed. The block diagram of the feedforward setup is depicted in Fig. 4. The FF/SU DAP algorithm solves for the separating filters by computing the minimum squared norm filter update such that the updated filter decorrelates the past K lag cross-correlations between the intermediate outputs $\tilde{y}_1(n)$ and $\tilde{y}_2(n)$, where $K<L$ and L is the maximum order of the cross-channel filters. This criterion can be expressed as the cost function

$$\mathcal{J}_p(n) = (\|\mathbf{w}_{pq}(n+1) - \mathbf{w}_{pq}(n)\|_2)^2 + E[\tilde{\mathbf{y}}_{p,K}^T(n)\tilde{y}_q(n)]\boldsymbol{\lambda}, \quad p \neq q = 1,2, \tag{3}$$

where $\boldsymbol{\lambda}$ is the $K\times 1$ vector of Lagrange multipliers and the $K\times 1$ intermediate output vector $\tilde{\mathbf{y}}_{p,K}(n)$ is defined as

$$\tilde{\mathbf{y}}_{p,K}(n) = \mathbf{x}_{p,K}(n) - \mathbf{X}_q^H(n)\mathbf{w}_{pq}(n+1), \tag{4}$$

and where $\mathbf{X}_q(n)=[\mathbf{x}_{q,L}(n)\ \mathbf{x}_{q,L}(n-1)\ \cdots\ \mathbf{x}_{q,L}(n-K+1)]$ is an $L\times K$ matrix of the past filter outputs. Replacing the cross-correlations with their instantaneous estimates (i.e., $E[y_p y_q] = y_p y_q$), using a stochastic gradient descent optimization rule, and after some algebraic manipulation, the resulting filter update equations take the form

$$\mathbf{w}_{pq}(n+1) = \mathbf{w}_{pq}(n) + \mu\mathbf{X}_q(n)(\mathbf{X}_q^T(n)\mathbf{X}_q(n) + \delta\mathbf{I})^{-1}\tilde{\mathbf{y}}_{p,K}(n). \tag{5}$$

Here, μ is the step size and δ is a regularization term added diagonally to prevent numerical difficulties in inverting the possibly rank deficient matrix $\mathbf{X}_q^T\mathbf{X}_q$. More detailed derivations of the FF/SU DAP algorithm as well as some other related BSS algorithms can be found in Ref. 31.

D. Performance measures

Well defined, quantitative performance measures are necessary for objectively evaluating the signal enhancement performances of the proposed BSS algorithm. In this paper, two performance measures are utilized to evaluate the signal enhancement performance of the FF/SU DAP algorithm. The pre-denoising quality of a noisy vocalization signal is quantified in terms of the input SNR which is defined for each input channel as the ratio of the squared root-mean-square (rms) values of the vocalization and noise signals

$$\mathrm{SNR}_{\mathrm{in},p} = 10\log_{10}((s_{\mathrm{rms},p})^2/(v_{\mathrm{rms},p})^2), \quad p=1,2, \tag{6}$$

where p is the channel index, and $s_p(n)$ and $v_p(n)$ are the vocalization and noise signals at channel p, respectively. The rms value of the length N signal $x(n)$ is defined as

$$x_{\mathrm{rms}} = \left[\frac{1}{N}\sum_{n=0}^{N-1}(x(n))^2\right]^{1/2}. \tag{7}$$

The noise signal in the denominator of Eq. (6) represents all signals (including watercraft emitted noise) other than the vocalization signal. The rms value of the noise signal is computed over the duration of the vocalization signal. In general, the SNR at the input channels may vary. In contrast, a single output SNR is defined as

$$\mathrm{SNR}_{\mathrm{out}} = 10\log_{10}[((y_{\mathrm{rms},s})^2 - (y_{\mathrm{rms},v})^2)/(y_{\mathrm{rms},v})^2], \tag{8}$$

where $y_{\mathrm{rms},s}$ is the rms value of the enhanced estimate of the vocalization signal and $y_{\mathrm{rms},v}$ is the rms value of the output if no vocalization is present. Thus, $y_{\mathrm{rms},v}$ represents the noise residue that is not suppressed by the algorithm over the duration of the vocalization.

The signal-to-distortion ratio (SDR) is defined as the ratio of the vocalization signal and distortion power

$$\mathrm{SDR} = 10 \log_{10}((s_{\mathrm{rms}})^2/(e_{\mathrm{rms}})^2), \tag{9}$$

where $s(n)$ is the vocalization signal, and the distortion is defined as the mean-squared-error

$$e_{\mathrm{rms}} = \left[\frac{1}{N}\sum_{n=0}^{N-1}(e(n))^2\right]^{1/2} = \left[\frac{1}{N}\sum_{n=0}^{N-1}(s(n)-y_s(n))^2\right]^{1/2}, \tag{10}$$

and $y_s(n)$ is the enhanced estimate of the vocalization signal. The SDR is an indicator of how well the vocalization waveform is preserved, which strongly affects the detection rate of a matched filter detector or other similar correlation based detectors, and the accuracy of source localization algorithms based on TDOA estimates.

III. EXPERIMENTAL RESULTS

The underwater acoustic environment is very challenging in terms of propagation. The acoustic propagating channel has certain dynamics that cannot be modeled accurately with a FIR filter. Unfortunately, these undesired channel effects are more pronounced in shallow water channels, primarily due to the complex interaction of the acoustic waves with the waveguide boundaries, volumetric inhomogeneities in the water, and other uncertainties inherent to the underwater acoustic channel. The underwater acoustic channel is frequency selective and can significantly attenuate high frequency signal energy. Another factor that results in frequency selective attenuation of underwater acoustic signals is the Lloyd mirror effect. All of these factors reduce the coherence of the signals measured at different receivers, which may affect the convergence and signal enhancement performance of the BSS algorithms. Thus, it is essential to evaluate the proposed FF/SU DAP algorithm under real environments to be able to fully comprehend their signal enhancement performances.

To evaluate the performance of the FF/SU DAP algorithm under realistic conditions, real vocalization and noise data were recorded at Crystal River, FL. The test location where the background noise data were 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.

In speech processing, speech signals are generally recorded in an anechoic chamber in order to obtain reverberant free source signals, and then convolved with an experimentally measured or numerically simulated reverberant impulse response. Thus, the source signals and the channel impulse responses are known a priori and can be used to evaluate the SNR and SDR performances of the algorithms as well as the convergence of the de-mixing filters to the optimum separating solution. Alternatively, to obtain more realistic mixtures, the speech signals are recorded separately (but in the same environment) and are numerically superposed to create the

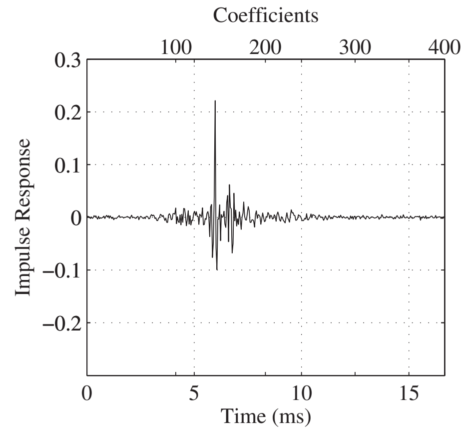


FIG. 5. A typical channel impulse response estimated from broadband broadcast tests conducted in Crystal River, FL.

noisy measurements.[32] Both experimental setups used in speech processing are not feasible for marine mammal monitoring applications. Nevertheless, in an effort to generate as realistic noisy vocalization recordings as possible, the vocalizations in the manatee vocalization library[33] are convolved with actual measured underwater acoustic channel impulse responses and added to real background noise recordings. The channel impulse responses used to convolve manatee vocalizations are estimated from a series of broadband broadcast tests conducted at Crystal River, FL on the same day and location as the background noise recordings. A typical impulse response obtained from the broadcast tests is presented in Fig. 5 where the underwater speaker is placed 10.3 m away from the reference hydrophone.

The performance of the FF/SU DAP algorithm is evaluated in a two-channel setup which consists of a manatee and a watercraft as the only two active acoustic point sources. In general, the input SNR is a function of the vocalization and watercraft emitted noise source levels (SLs) as well as the channel attenuation. Channel attenuation is determined by the channel impulse responses, whereas both the vocalization and watercraft emitted noise SL may change.[7,9] To compensate for the variance in the SL of both the manatee vocalizations and watercraft emitted noise, the input SNR is controlled by scaling the power of the vocalization signal such that the input SNR at the reference channel of the vocalizations is equal to pre-specified values. Another important factor that reduces the SNR and the separation performance of the proposed BSS algorithm is the presence of extraneous noise. The ambient background noise levels measured at the Crystal River test site when no point source was active (i.e., no watercraft in the vicinity) were determined to be lower than −12 dB compared to typical watercraft emitted noise throughout the measurements. The diffuse ambient noise levels were used to estimate the lowest input SNR that the FF/SU DAP algorithm was expected to achieve acceptable signal enhancement performance. Since manatee vocalizations are generally 0.5 s or shorter in duration, several watercraft noise recordings, each of 1 s duration, were selected from the measurements [see Figs. 1(c) and 1(d)]. The approach direction and the relative speed of each watercraft were noted during the recordings. Both the vocalization and

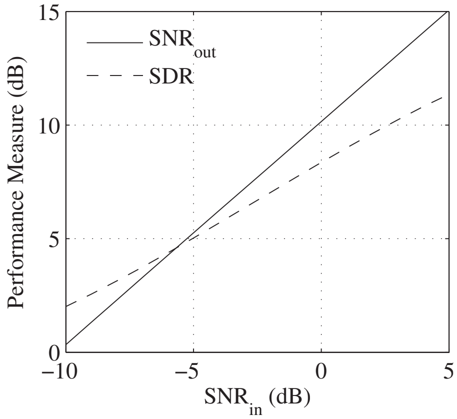


FIG. 6. The output SNR and SDR performance measures as a function of the input SNR for the FF/SU DAP algorithm averaged over ten manatee vocalizations and four different noise recordings.

noise signals were high-pass filtered using tenth order Butterworth filter with a cutoff frequency of 1 kHz. The filter order, step size, and cross correlation delay for the FF/SU DAP algorithm were set to $L=200$, $\mu=0.01$, and $K=10$, respectively.

The output SNR and SDR results averaged over ten manatee vocalizations and four different noise recordings are presented in Fig. 6. For these tests, it was assumed that the manatee was 1 m away from the reference hydrophone. The other hydrophone is located 9.3 m away. The input SNR at the vocalization reference channel is varied from −10 to 5 dB. The input SNR at the other channel is not manipulated, but rather is determined by the channel impulse responses, and is usually 5–10 dB lower than the input SNR. A typical time domain output of the FF/SU DAP algorithm is presented in Fig. 7 for −5 dB input SNR on channel 1.

As is discussed in Sec. II, the cross channel transfer function is more relevant for BSS algorithms. The results presented above are obtained for manatee vocalizations convolved with a channel impulse response between two hydrophones (separated by 9.3 m) where the manatee is assumed to be 1 m away from the reference hydrophone.

In the following tests, the effect of changing the distance between the manatee and the corresponding reference hydrophone on the performance of the FF/SU DAP algorithm is investigated. The separation between the hydrophones is fixed at 9.3 m, and the manatee is assumed to be at distances of 1, 4.7, 10.3, and 13.9 m away from the reference hydrophone. Although the distance between the hydrophones remains the same, the pseudochannel impulse response changes as the range of the manatee is increased. The output SNR and SDR results obtained with the manatee assumed to be at the four different locations are presented in Fig. 8. These results suggest that the signal enhancement performance of the FF/SU DAP algorithm is not significantly affected by changes in the channel transfer function between the manatee and the corresponding reference receiver, particularly at low input SNR values. Moreover, these results prove that the detection range resulting from processing the noisy vocalizations with the FF/SU DAP algorithm is only a function of the input SNR. This conclusion simplifies the detection range computations presented in Sec. IV.

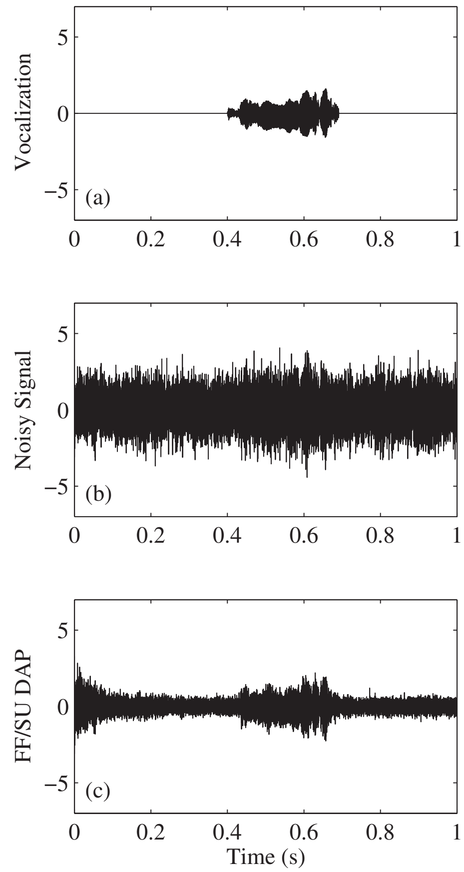


FIG. 7. (a) A clean manatee vocalization in the time domain, (b) the noisy measurement with input SNR equal to −5 dB on channel 1, and (c) the time domain output of the FF/SU DAP algorithm (plots are for the channel 1 measurements; the input SNR for channel 2 is −13.4 dB).

IV. DETECTION RANGE IMPROVEMENTS

The signal enhancement performance of the FF/SU DAP algorithm can be related to the improvements in the detection range through the passive sonar equation written in terms of the figure of merit (FOM)

$$\mathrm{FOM} = \mathrm{SL}_m - \max(\mathrm{NL}_a, \mathrm{SL}_w - \mathrm{TL}_w) + \mathrm{AG} - \mathrm{DT}, \quad (11)$$

where SL_m is the source level of the vocalizations, NL_a is the ambient noise levels, SL_w is the boat source level, 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.[9] The source levels are the sound pressure levels (referenced to 1 μPa) located 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 transmission loss obeys a mixed spreading model,[9] the FOM can also be expressed as a function of the range of the manatee to the hydrophones

$$\mathrm{FOM} = 15 \log_{10}(r_m). \quad (12)$$

Equating Eqs. (11) and (12) and solving for the range of the manatee, one obtains

$$r_m = 10^{(\mathrm{SL}_m - \max(\mathrm{NL}_a, \mathrm{SL}_w - 15 \log_{10}(r_w)) + \mathrm{AG} - \mathrm{DT})/15}. \quad (13)$$

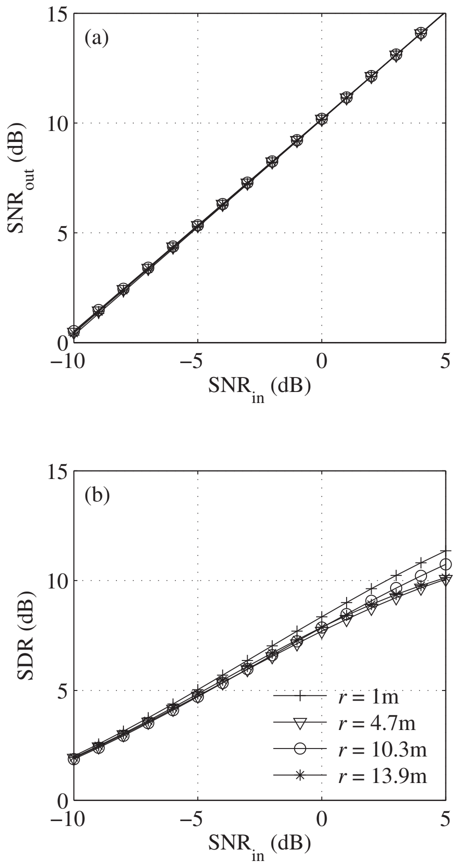


FIG. 8. (a) The output SNR and (b) the SDR performance measures for different manatee ranges as a function of the input SNR for the FF/SU DAP algorithm.

The ratio of the detection range resulting from the FF/SU DAP algorithm to high-pass filtering detection range can be computed using the relation

$$\frac{(r_m)_{\mathrm{DAP}}}{(r_m)_{\mathrm{HPF}}} = 10^{\mathrm{AG}/15}. \tag{14}$$

The AG term is defined as the improvement in the SNR with respect to the single receiver input SNR. Thus, for the FF/SU DAP algorithm, AG is defined as the improvement in the SNR,

$$\mathrm{AG} = \mathrm{SNR}_{\mathrm{out}} - \mathrm{SNR}_{\mathrm{in}}. \tag{15}$$

The average ratio of the detection ranges resulting from the FF/SU DAP algorithm over high-pass filtering for the Crystal River test cases is presented in Fig. 9. The FF/SU DAP algorithm achieves a relatively uniform output SNR performance over the input SNR values and can improve the detection range by a factor of 4.7 or higher for input SNR varying from −10 to 5 dB.

Next, the detection ranges resulting from processing the vocalizations with the FF/SU DAP algorithm are presented using an example. The watercraft noise and manatee vocalization SL are assumed to be 140 and 118 dB, respectively. Ambient noise is assumed to be 70 dB. For a given output SNR, it is necessary to compute the input SNR (and thus, the AG) for the FF/SU DAP algorithm. To obtain an analytic

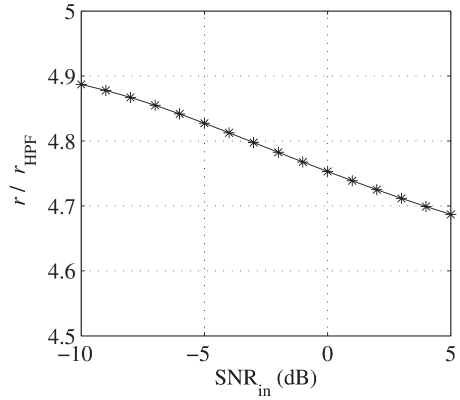


FIG. 9. The average ratio of the detection range resulting from the FF/SU DAP algorithm to high-pass filtering for the Crystal River tests.

relation between the input and the output SNRs, the output SNR for the Crystal River test cases is curve fitted with a first order polynomial

$$\mathrm{SNR}_{\mathrm{out}} = 0.9734 \cdot (\mathrm{SNR}_{\mathrm{in}}) + 10.2057 \tag{16}$$

in the least squared sense.[34] Equation (16) suggests that an average of 10.2 dB AG can be achieved with the FF/SU DAP algorithm. For the FF/SU DAP algorithm implemented together with a 3 dB DT passive acoustic detector, the AG is 10.4 dB and the minimum input SNR necessary for detection is computed as −7.4 dB. The ratio of the estimated detection range resulting from the FF/SU DAP algorithm and high-pass filtering is computed as 4.9 for a DT of 3 dB. The maximum detection range as a function of the detection threshold and range of the watercraft are depicted in Fig. 10. Although the detection range drops below 10 m when the watercraft is 100 m away from the receivers with the FF/SU DAP algorithm, the proposed method significantly increases the detection range (approximately five times) and the effective coverage area (approximately 25 times) of a passive acoustic based detector compared to a high-pass filter alone.

For the setup in this study, the manatee and watercraft are assumed to be closer to hydrophones 1 and 2, respec-

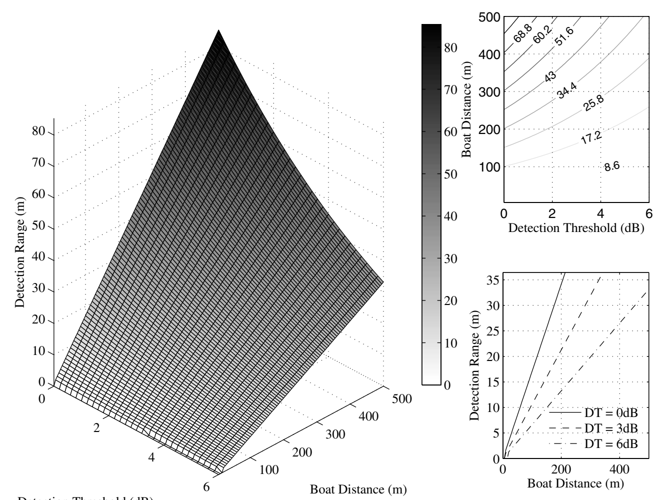


FIG. 10. Manatee vocalization detection ranges after processing with the FF/SU DAP algorithm in the presence of watercraft emitted noise.

tively. Accordingly, it can be shown that the separated vocalization and watercraft noise signals are retrieved from channels 1 and 2, respectively.[27] Since the location of the animal is not known in advance, both channels have to be monitored by a detector.

The separation performance of the FF/SU DAP algorithm depends on the coherence between the measurements. The separation and signal enhancement performance reduces as the cross-channel coherences drop. There are several factors that may affect inter-channel coherence. A greater separation between the hydrophones increases the coverage area of the passive detection system, but also reduces the coherence. Hence, there is a trade-off between coverage and performance. In addition, the acoustic signals recorded from a small recreational watercraft departing away from the hydrophones are relatively incoherent due to the scatter from cavitation of the propeller. Thus, the proposed system will enhance vocalization signals particularly in the presence of an approaching vessel. Increase in the ambient diffuse noise field (e.g., when precipitating) also results in reduced coherence. For cases in which the coherence drops below a certain level, the FF/SU DAP algorithm can be reconfigured to function as a single channel supervised adaptive noise canceller implemented with the affine projection update rule (by forcing one of the de-mixing filters to function as an all-pass unity filter).

V. CONCLUSIONS

The problem of enhancing marine mammal vocalizations in the presence of an interfering acoustic source is addressed in this paper. The conventional approach to underwater signal enhancement is to use expensive directional receivers such as vector sensors or to beamform the measurements from an array of many omni-directional hydrophones. As a low-cost alternative to these methods, a two-channel, unsupervised, adaptive source separation approach is proposed for enhancing noisy manatee vocalizations. The presented FF/SU DAP algorithm uses the affine projection update rule to be able to maintain a satisfactory convergence speed in the presence of dynamic underwater mixing channels (due to the motion of watercraft) and temporally correlated vocalizations. The FF/SU DAP algorithm is evaluated using real watercraft emitted noise and manatee vocalizations 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 extends the detection range on average five times compared to high-pass filtering alone. In contrast to conventional beamforming, the presented source separation approach does not require an array of many hydrophones and is not sensitive to uncertainties in the sensor locations. One limitation of the proposed FF/SU DAP algorithm is that the signal enhancement performance declines as the inter-channel coherence between the two channels is reduced (e.g., due to increases in diffuse ambient noise levels). For low coherence measurements, the implementation structure of the FF/SU DAP algorithm allows it to be reconfigured as the conventional supervised adaptive noise canceller which is a robust algorithm for enhancing signals in the presence of multiple uncorrelated noise measurements.

Although the proposed FF/SU DAP algorithm was implemented and evaluated using manatee vocalizations, the approach makes very generic assumptions about properties of the vocalization signals and the shallow underwater acoustic channel. Hence, the proposed method is suitable for extending to other marine mammal monitoring applications such as the detection and the classification of whale calls in deep oceans with minimal modifications to the algorithm parameters.

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 providing financial support for this research.

[1]U.S. Fish and Wildlife Service, "Florida manatee recovery plan (*Trichechus manatus latirostris*)," 3rd revision, US Fish and Wildlife Service, 5430 Grosvenor Lane, Suite 110, Bethesda, MD 20814 (2001).
[2]Florida Fish and Wildlife Commission, "Florida manatee management plan. Florida Fish and Wildlife Commission," 620 South Meridian Str., Tallahassee, FL 32399-1600 (2007).
[3]Florida Fish and Wildlife Conservation Commission, "Manatee mortality statistics," http://research.myfwc.com/features/category_sub.asp?id=2241 (Last viewed 8/14/2009).
[4]J. F. Gorzelany, "Evaluation of boater compliance with manatee speed zones along the Gulf Coast of Florida," Coastal Manage. **32**, 215–226 (2004).
[5]C. Steel, "Vocalization patterns and corresponding behavior of the West Indian manatee (*Trichechus manatus*)," Ph.D. thesis, Florida Institute of Technology, Melbourne, FL (1982).
[6]D. Nowacek, B. M. Casper, R. S. Wells, S. M. Mowacek, and D. A. Mann, "Intraspecific and geographic variation of West Indian manatee (*Trichechus manatus spp.*) vocalizations," J. Acoust. Soc. Am. **114**, 66–69 (2003).
[7]R. Phillips, C. Niezrecki, and D. O. Beusse, "Determination of West Indian manatee vocalization levels and rate," J. Acoust. Soc. Am. **115**, 422–428 (2004).
[8]C. Niezrecki, R. Phillips, M. Meyer, and D. O. Beusse, "Acoustic detection of manatee vocalizations," J. Acoust. Soc. Am. **114**, 1640–1647 (2003).
[9]R. Phillips, C. Niezrecki, and D. O. Beusse, "Theoretical detection ranges for acoustic based manatee avoidance technology," J. Acoust. Soc. Am. **120**, 153–163 (2006).
[10]Z. Yan, C. Niezrecki, L. N. Cattafesta III, and D. O. Beusse, "Background noise cancellation for improved acoustic detection of manatee vocalizations," J. Acoust. Soc. Am. **117**, 3566–3573 (2005).
[11]Z. Yan, C. Niezrecki, L. N. Cattafesta III, and D. O. Beusse, "Background noise cancellation of manatee vocalizations using an adaptive line enhancer," J. Acoust. Soc. Am. **120**, 145–152 (2006).
[12]B. M. Gur and C. Niezrecki, "Autocorrelation based denoising of manatee vocalizations using the undecimated discrete wavelet transform," J. Acoust. Soc. Am. **122**, 188–199 (2007).
[13]Y. Ren, M. T. Johnson, and J. Tao, "Perceptually motivated wavelet packet transform for bioacoustic signal enhancement," J. Acoust. Soc. Am. **124**, 316–327 (2008).
[14]B. Van Veen and K. M. Buckley, "Beamforming techniques for spatial filtering," in *The Digital Signal Processing Handbook*, edited by V. K. Madisetti and D. Williams (CRC, Boca Raton, FL, 1998), pp. 61/1–61/22.
[15]P. Miller and P. L. Tyack, "A small towed beamforming array to identify vocalization *resident* killer whales (*Orcinus orca*) concurrent with behavioral observation," Deep-Sea Res., Part II **45**, 1389–1405 (1998).
[16]F. Desharnais, G. R. Ebbeson, M.-N. R. Matthews, G. J. Heard, D. J. Thomson, and G. H. Brooke, "A generalized beamformer for localization

of marine mammals," Appl. Acoust. **67**, 1213–1225 (2006).
[17]P. B. Muanke and C. Niezrecki, "Manatee position estimation by passive acoustic localization," J. Acoust. Soc. Am. **121**, 2049–2059 (2007).
[18]M. S. Pedersen, J. Larsen, U. Kjems, and L. C. Parra, "Convolutive blind source separation methods," in *Springer Handbook of Speech Processing*, edited by J. Benesty, S. M. Sondhi, and Y. Huang (Springer, Berlin, 2008), pp. 1065–1095.
[19]M. Gaeta, F. Briolle, and P. Esparcieux, "Blind separation of sources applied to convolutive mixtures in shallow water," in Proceedings of the IEEE Signal Processing Workshop on Higher Order Statistics (1997), pp. 340–343.
[20]S. Bonnifay, K. Yao, and C. Jutten, "Underwater acoustic signal separation based on prior estimation of the channel impulse response," in Proceedings of the ICASSP (2000), Vol. 5, pp. 3156–3159.
[21]A. Mansour, N. Benchekroun, and C. Gervaise, "Blind separation of underwater acoustic signals," in *Proceedings of the ICA, Lecture Series on Computer Science*, edited by J. Rosca, D. Erdogmus, J. C. Principe, and S. Haykin (Springer, Berlin, 2006), Vol. 3889, pp. 181–188.
[22]L. Parra and C. Spence, "Convolutive blind separation of non-stationary sources," IEEE Trans. Speech Audio Process. **8**, 320–327 (2000).
[23]H. Buchner, R. Aichner, and W. Kellerman, "A generalization of blind source separation algorithms for convolutive mixtures based on second-order statistics," IEEE Trans. Speech Audio Process. **13**, 120–134 (2005).
[24]Z. H. Michalopoulou, "Underwater transient signal processing: Marine mammal identification, localization, and source signal deconvolution," in Proceedings of the ICASSP (1997), Vol. **1**, pp. 503–506.
[25]S. Van Gerven and D. Van Compernolle, "Signal separation by adaptive symmetric decorrelation: Stability, convergence, and uniqueness," IEEE Trans. Signal Process. **43**, 1602–1612 (1995).
[26]S. Haykin, *Adaptive Filter Theory* (Prentice-Hall, Upper Saddle River, NJ, 2002), Chap. 5, pp. 231–320.
[27]E. Weinstein, M. Feder, and A. V. Oppenheim, "Multi-channel signal separation by decorrelation," IEEE Trans. Speech Audio Process. **1**, 405–413 (1993).
[28]G. E. Bottomley and S. T. Alexander, "A novel approach for stabilizing recursive least squares filters," IEEE Trans. Signal Process. **39**, 1770–1779 (1991).
[29]K. Ozeki and T. Umeda, "An adaptive filtering algorithm using orthogonal projection to an affine subspace and its properties," Electron. Commun. Jpn. **67**, 19–27 (1984).
[30]M. Gabrea, "Double affine projection algorithm-based speech enhancement algorithm," in Proceedings of the ICASSP (2003), Vol. **1**, pp. 904–907.
[31]B. M. Gur, "Adaptive enhancement of marine mammal vocalizations," Ph.D. thesis, University of Massachusetts, Lowell, MA (2008).
[32]D. Schobben, K. Torkkala, and P. Smaragdis, "Evaluation of blind signal separation methods," in Proceedings of the ICA (1999), pp. 261–266.
[33]T. O'Shea, "Manatee vocalization-catalog of sounds," produced by Coastal Systems Station, Naval Surface Warfare Center, Dahlgren Division, Panama City, FL (1981–1984).
[34]Z. Yan, "Background noise cancellation for acoustic detection of manatee vocalizations," Ph.D. thesis, University of Florida, Gainesville, FL (2006).