ADAPTIVE ENHANCEMENT OF MARINE MAMMAL VOCALIZATIONS

BY

BERKE M. GÜR
B.S. MIDDLE EAST TECHNICAL UNIVERSITY (1999)
M.S. UNIVERSITY OF SOUTHERN CALIFORNIA (2003)
M.B.A. BOĞAZİÇİ UNIVERSITY (2006)

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY
IN MECHANICAL ENGINEERING
UNIVERSITY OF MASSACHUSETTS LOWELL

Signature of Author: M. Berke Gür **Date:** 4 DEC. 2008

Signature of Dissertation Supervisor: Christopher Niezrecki

Signature of Other Dissertation Committee Members: Matthew Chamberlain

Peter Avitabile

ADAPTIVE ENHANCEMENT OF MARINE MAMMAL VOCALIZATIONS

BY

BERKE M. GÜR

ABSTRACT OF A DISSERTATION SUBMITTED TO THE FACULTY OF THE
DEPARTMENT OF MECHANICAL ENGINEERING
IN PARTIAL FULLFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY IN MECHANICAL ENGINEERING
UNIVERSITY OF MASSACHUSETTS LOWELL
2008

Dissertation Supervisor: Christopher Niezrecki, Ph.D.
Associate Professor, Department of Mechanical Engineering

ABSTRACT

In this dissertation, adaptive signal enhancement and noise reduction algorithms for marine mammal vocalizations (specifically, manatee vocalizations) are investigated motivated by the recent advances in speech processing. There has been a recent growing interest in the scientific community for acoustically monitoring marine mammals. Noise-free vocalization measurements are crucial for effective detection and classification. However, the vocalizations are generally distorted due to multi-path channel transmission, and further corrupted with background noise. In this research, the problem is approached in the unsupervised adaptive filtering framework. Several signal models are proposed for modeling the vocalizations and background noise. First, wavelet domain denoising (WDD) is proposed as a single channel noise reduction approach. Numerical simulations and tests conducted using real vocalization and noise data show that the WDD algorithm outperforms linear filtering. Second, blind source separation (BSS) is employed for enhancing vocalization signals buried in acoustic noise such as propeller cavitation noise emitted from surface vessels. A new time domain BSS algorithm based on the affine projection (AP) algorithm is proposed. The signal enhancement performance of the new algorithm is compared to several well-known decorrelation based BSS algorithms. Numerical simulations and experiments conducted with real recordings indicate that the AP based BSS algorithm can enhance weak vocalization signals in the presence of other point noise sources. Finally, weak signal detectors that complement the signal enhancement algorithms are proposed. The contribution of this dissertation is the introduction of contemporary adaptive signal enhancement schemes which constitute viable alternatives to conventional noise cancelling and beamforming techniques for enhancing marine mammal vocalizations.

To my wife, my mom, my dad, and my sister

ACKNOWLEDGEMENTS

I would like to express my gratitude to my dissertation advisor Dr. Niezrecki for providing me with the opportunity to pursue my Ph.D. studies at his laboratory and for his encouragement throughout the past four years. I would like to acknowledge and thank my committee members, Dr. Avitabile, for the invaluable engineering intuition he has provided, Dr. Chandra for patiently enduring all my questions, and Dr. Mahmoud for showing a keen interest in me and my research. I would also like to thank my colleagues in our research group for their company and for making the never-ending workdays an enjoyable experience.

I am much obliged to the Florida Sea Grant, and the Florida Fish and Wildlife Conservation Commission for providing partial support that enabled this research, the University of Massachusetts Lowell Provost's Office for funding the final year of my studies, and the University of Massachusetts Lowell Graduate Student Association for providing the funding for field-trips and data collection.

I am very grateful to my family, to my mother, to my father, and to my sister for relentlessly supporting me throughout my graduate studies and having confidence in me. I would like to extend my sincere appreciation to my extended family, my father-in-law for all his career guidance, and my mother-in-law for her many helpful comments on an earlier version of this dissertation.

Last but most importantly, I would like to thank my beloved wife Nükhet, who patiently supported me throughout the past years and gave me hope when times were tough. Although we had to be far apart for some time, her devotion kept me motivated to move forward and made me the better person and scholar I am today.

TABLE OF CONTENTS

Page

Page

Page

Page

LIST OF TABLES

Page

LIST OF ILLUSTRATIONS

Page

Page

Page

Page

Page

Page

ABBREVIATIONS

AG	Array Gain
ALE	Adaptive Line Enhancer
ANC	Adaptive Noise Canceller
AP	Affine Projection
APD	Amplitude Probability Distribution
AR	Autoregressive
ARMA	Autoregressive Moving Average
AUTEC	Atlantic Undersea Test and Evaluation Center
AWPT	Adaptive Wavelet Packet Transform
BSS	Blind Source Separation
BU	Block-Update
CPSD	Cross-Power Spectral Density
CSDM	Cross-Spectral Density Matrix
DAP	Double Affine Projection
DCT	Discrete Cosine Transform
DFT	Discrete Fourier Transform
DOA	Direction of Arrival
DRLS	Double Recursive Least Squares
DT	Detection Threshold
DWT	Discrete Wavelet Transform
EM	Estimate-Maximize

FALE	Feedback Adaptive Line Enhancer
FB	Feedback
FF	Feedforward
FIR	Finite Impulse Response
FIR-ALE	Finite Impulse Response-Adaptive Line Enhancer
FOM	Figure of Merit
FWC	(Florida) Fish and Wildlife Conservation (Commission)
FWS	Florida Wildlife Service
HOS	Higher Order Statistics
IID	Independent and Identically Distributed
IIR	Infinite Impulse Response
KLT	Karhunen-Loeve Transform
LMS	Least Mean Squares
LO	Locally Optimum
MIMO	Multi-Input Multi-Output
ML	Maximum Likelihood
NPES	Normalized Partial Energy Sum
MSE	Mean Square Error
MVDR	Minimum Variance Distortionless Response
NL	Noise Level
PDE	Partial Differential Equation
PDF	Probability Density Function
PSD	Power Spectral Density

QMF	Quadrature Mirror Filter
RLS	Recursive Least Squares
RMS	Root Mean Square
SαS	Symmetric Alpha-Stable
SAD	Symmetric Adaptive Decorrelator
SDR	Signal to Distortion Ratio
SE	Signal Excess
SL	Source Level
SNR	Signal to Noise Ratio
SOS	Second Order Statistics
SPL	Sound Pressure Level
STFT	Short-Time Fourier Transform
SU	Sequential-Update
SVD	Singular Value Decomposition
TDOA	Time Difference of Arrival
TITO	Two-Input Two-Output
TL	Transmission Loss
UDWT	Undecimated Discrete Wavelet Transform
USCG	Unites Stated Coast Guard
UWPT	Undecimated Wavelet Packet Transform
WDD	Wavelet Domain Denoising
WGN	White Gaussian Noise
WPT	Wavelet Packet Transform

WSS	Wide Sense Stationary

CHAPTER 1: INTRODUCTION

This first chapter begins with a description of the scope of the dissertation. The motivation for this research, which provides a practical application of passive acoustic monitoring, is presented in detail, followed by the scholarly contributions of the dissertation. The chapter is concluded with an overview of the remaining chapters.

1.1 Scope of the Dissertation

This dissertation focuses on the enhancement and the detection of weak marine mammal vocalizations corrupted with background noise. The research described herein is motivated by the practical problem of detecting manatees from their vocalizations for warning watercraft in the vicinity of their presence. Hence, the research herein provides a practical solution to a real problem. This study combines signal processing and underwater acoustics with marine mammal biology; and thus, represents a multi-disciplinary research effort. In particular, contemporary speech and other related signal processing techniques are investigated for passive sonar applications in the context of enhancing manatee vocalizations. While the primary objectives of the research presented are to enhance and to detect the vocalizations of the Florida manatee, the methods

proposed are developed within the constraints of very generic acoustic signal models and can be readily extended to other marine mammal species.

Through this dissertation, two fundamentally different signal enhancement techniques are investigated; wavelet domain denoising (WDD) and blind source separation (BSS). Marine mammal vocalizations, including manatee vocalizations, are non-stationary and have a variety of time-frequency and statistical distributions. The motivation for investigating wavelet domain algorithms stems from the ability of the wavelet transform to efficiently represent a variety of non-stationary signals with diverse time-frequency and statistical properties. WDD is a non-linear noise reduction method, which offers improved noise reduction and signal waveform preservation properties compared to linear methods.

The enhancement performance of conventional single channel adaptive filtering methods drastically degrades as the signal to noise ratio (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 approach to multi-channel signal enhancement problems is adaptive array processing and beamforming. However, adaptive beamforming has several drawbacks, which are discussed in the next chapter. As an alternative, BSS is a multi-channel approach in which a mixture of signals is statistically separated. Thus, the multi-channel BSS algorithms provide a viable signal processing framework for enhancing marine mammal vocalizations in the presence of interfering point sources. Both the WDD and BSS

approaches are very versatile in the sense that they can accommodate signals with a variety of temporal, spectral and statistical properties. In this respect, the proposed algorithms constitute possible alternatives to conventional adaptive signal processing techniques such as adaptive noise cancellation and adaptive beamforming, and are suitable for marine mammal monitoring applications. This dissertation focuses primarily on the more challenging problem of processing weak vocalization signals. The vocalization signals are assumed to have an acoustic signal power on the order of or lower than the contaminating noise power and thus, require enhancement for detection. A novel wavelet domain detector is also proposed for detecting weak manatee vocalizations.

1.2 Manatee-Watercraft Collisions

The West Indian manatee (*Trichechus manatus latirostris*), depicted in Figure 1-1, was added to the endangered species list in 1967 by the Florida Wildlife Service (FWS), and detailed records of manatee mortalities have been kept consistently since 1976. In 1980, the FWS 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 Florida waterways where manatee-watercraft collisions were most likely to occur (see Figure 1-2). These speed zones were established to reduce the threat of watercraft to manatees by increasing the reaction time of a boat operator and the manatee, and reducing the severity of the injuries in the event that a

collision occurs [1]. In a series of surveys, Nowacek et al. [2] determined that manatees responded to approaching watercraft by swimming towards deeper waters and that slower moving watercraft provided manatees with a greater time to swim to the relative safety of deeper waters.

However, a 2001 recovery review revealed that the rate of manatee-watercraft collision related mortalities continue to remain high, despite measures taken [3]. The number of watercraft related mortalities from 1976 to 2007 and the corresponding ratio to all mortalities are presented in Figure 1-3 (a) and (b), respectively. On the average, a quarter of all manatee mortalities are attributed to watercraft collisions (see Figure 1-4).

Figure 1-1. A manatee cow and her calf (FWC, 2008).

Figure 1-2. A slow speed manatee zone sign in New Smyrna Beach, FL.

Several reasons explaining the ineffectiveness of the speed zones have been put forth in the recent literature. Gerstein et al. [4] determined that high acoustic levels were necessary for manatees to hear and react to approaching watercraft, and suggested that slowing the watercraft made it more difficult for manatees to sense a watercraft at a

sufficient distance. Calleson and Frohlich [1] identified the increase in the draft of planning hull boats at slower speeds as one possible cause for the ineffectiveness of the speed zones. Another 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 were reported as 58% and 63% [5].

The problem also has both economic and social consequences. The State of Florida, with approximately one million registered watercraft, was the highest ranked state in terms of the total number of registered watercraft in 2006 [6]. The boating industry in Florida has been continuously expanding and has reached $ 18.4 billion in 2005 [7]. The increased travel times within the Florida channels negatively affect recreational boating, the boating industry, real estate, and development (due to speed regulations and restrictions to marine and dock construction in manatee sanctuaries). Recently, the Florida manatee was re-classified as a threatened species with a very high risk of extinction. A very detailed review of the current status of the Florida manatee and the state of the manatee protection plan enforced in Florida is presented in a report by the Florida Fish and Wildlife Conservation Commission (FWC) [8].

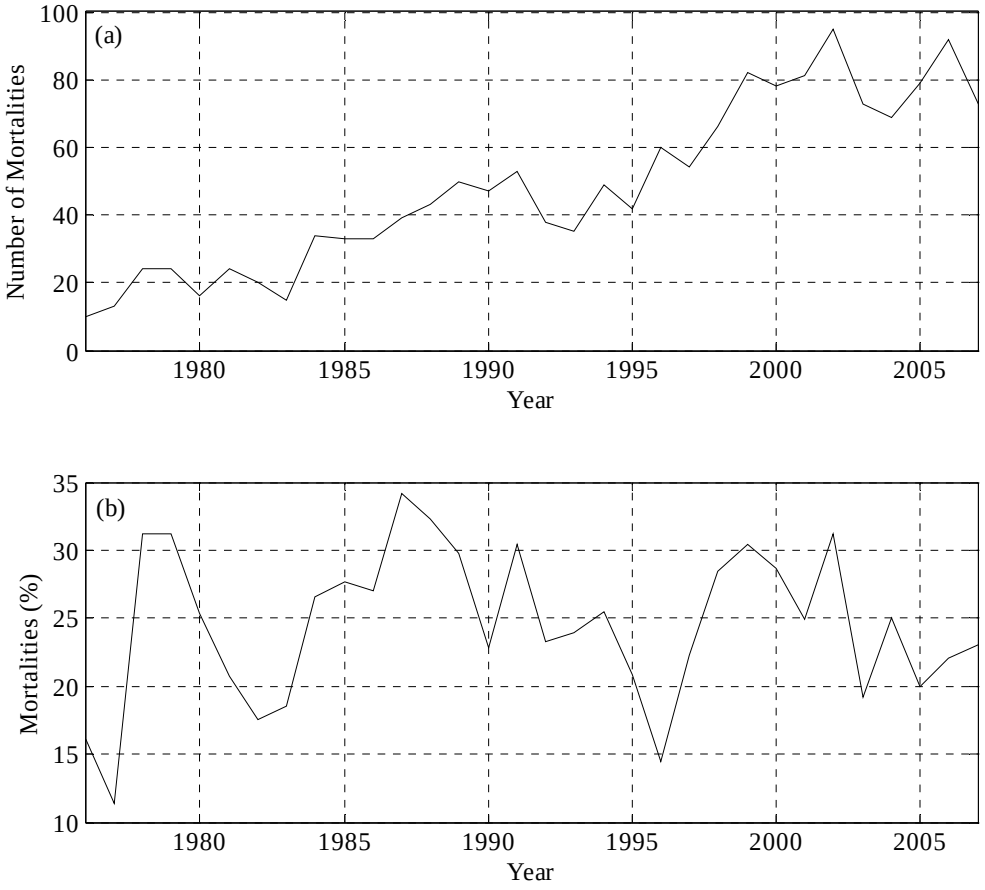


Figure 1-3. Manatee mortality statistics. (a) The number of watercraft related manatee mortalities, and (b) the ratio of watercraft related mortalities to overall mortalities from 1976 to 2007[a].

[a] Mortality statistics are compiled from FWC data which can be accessed through the URL http://research.myfwc.com/features/category_sub.asp?id=2241 (last accessed: 18 Oct. 2008).

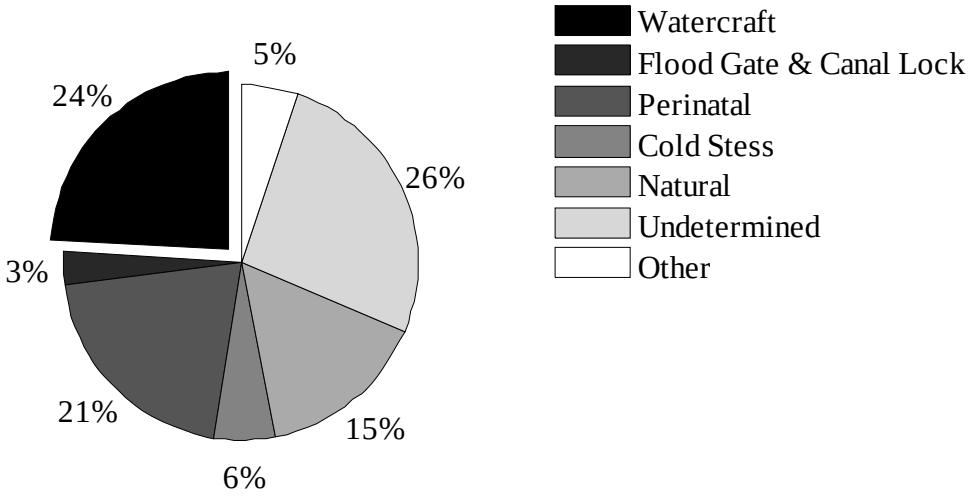


Figure 1-4. The causes of total manatee mortalities from 1976 to 2007.

1.3 Contributions

The primary objectives of this dissertation are to develop single and multi-channel signal processing algorithms for passive acoustic marine mammal monitoring systems, and in particular, to develop such systems for enhancing and detecting the vocalizations of the Florida manatee. The dissertation represents the first attempt to investigate, develop, and evaluate wavelet denoising and unsupervised source separation algorithms for enhancing manatee vocalizations. However, the proposed methods do not make any exclusive assumptions on the vocalization signals, and thus can be extended to other marine mammals. The contributions of the research are as follows.

1. *This dissertation provides a comprehensive understanding on the acoustic environment of the Florida manatee with respect to the manatee vocalizations, background noise, and the shallow water acoustic channel transmission in typical*

manatee habitats. The autoregressive moving average/autoregressive (ARMA/AR) and the symmetric alpha-stable (SαS) signal models are proposed as unifying mathematical models for marine mammal vocalizations and background noise signals, respectively. Underwater acoustic propagation models for modeling acoustic transmission in very shallow water environments are evaluated and verified through experiments.

2. *Adaptive WDD algorithms are proposed and evaluated for enhancing noisy manatee vocalizations.* This research represents the first study in the literature to implement coefficient thresholding based enhancement of marine mammal vocalizations in the wavelet domain.
3. *Multi-channel, unsupervised source separation algorithms are investigated for enhancing weak manatee vocalizations. A novel second order statistics (SOS) based BSS method is proposed for improving the convergence speed and signal enhancement performance of sequential decorrelation based BSS algorithms.* This dissertation is the first to propose and evaluate BSS for the purpose of enhancing marine mammal vocalizations. In particular, an unsupervised BSS algorithm is developed based on the affine projection algorithm.
4. *The proposed WDD and BSS algorithms are evaluated with real vocalization and background noise signals.* Theoretical improvements in the detection range of a vocalization detector are estimated and analyzed. The results obtained from these tests conducted with real vocalization and noise signals provide an accurate indicator of the in-field performance to be expected in a typical manatee habitat.

5. *Optimum detectors for detecting weak vocalization signals in Gaussian and non-Gaussian noise environments are investigated.*

1.4 Organization of the Dissertation

Chapter 2 provides a comprehensive review of literature and theory relevant to the scope of this dissertation, beginning with the introduction of the recent developments and current research trends in manatee avoidance technology. The need to enhance manatee vocalizations to achieve satisfactory detection performance is explained by exploring several examples. The conventional single and multi-channel adaptive filtering algorithms for enhancing signals are reviewed, followed by an introduction of the two primary approaches proposed (i.e., WDD and BSS). The rationale for selecting the two approaches for further investigation is also provided.

Chapter 3 presents the underlying mathematics for the problem addressed in this research and serves as a foundation for the remainder of the dissertation. The first sections of the chapter introduce the underwater acoustic source and noise signals. Pertinent properties of the acoustic signals are presented, followed by the development of signal models. These signal models serve to generate reproducible test cases for evaluating the performances of the algorithms proposed in the following two chapters. The properties of the underwater acoustic channel also influence the criterion used for designing adaptive enhancement and detection algorithms. Analytic and numerical underwater acoustic channel models are presented next. Finally, several measures used for performance evaluation are defined.

In Chapter 4, WDD algorithms for denoising noisy manatee vocalizations are investigated. Different aspects of the wavelet denoising strategy for processing manatee vocalizations are investigated first. Next, the conventional WDD approach is presented in a more instructive optimal filtering framework. An adaptive WDD algorithm based on the wavelet packet transform (WPT) is introduced in the next section, followed by the description of a similar Fourier domain algorithm. In the final section of the chapter, the performance of the proposed algorithms is evaluated by Monte-Carlo trials using the signal and channel models developed in Chapter 3.

Chapter 5 is primarily concerned with multi-channel unsupervised adaptive source separation algorithms. First, a novel SOS based BSS algorithm for two-input two-output (TITO) systems is developed as an extension of the supervised affine projection algorithm. Several alternative implementation structures are proposed in the next section, followed by extensions to multi-input multi-output (MIMO) configurations. The chapter is concluded with a performance evaluation obtained from Monte-Carlo trials.

In Chapter 6, the algorithms developed in the preceding two chapters are thoroughly tested and evaluated with real manatee vocalizations and background noise. The experimental setup is briefly described first. The WDD algorithms are evaluated for different input SNR in the next section and compared to linear adaptive filtering. In the following section, a similar performance analysis is conducted for the BSS algorithms proposed in Chapter 5.

In Chapter 7, the improvements in the detection range resulting from processing noisy manatee vocalizations with the proposed WDD and BSS algorithms is evaluated.

The passive sonar equations and the methodology used to estimate the detection range is provided first. In the following section, the detection range improvements resulting from the proposed adaptive WDD algorithm are compared to highpass filtering, linear adaptive filtering and Fourier domain methods. A similar analysis for the proposed multi-channel BSS algorithm is presented next. Estimates of the improved detection ranges resulting from both the WDD and BSS algorithms are provided using several examples.

Several important conclusions regarding the WDD and BSS algorithms are presented in Chapter 8. The extension of the proposed algorithms to other marine mammals and noise environments are presented. Some key aspects of the proposed algorithms that require further research is outlined.

A concise review of detection theory is presented in Appendix A. An undecimated wavelet packet transform (UWPT) based locally optimum (LO) detector for detecting weak manatee vocalization in snapping shrimp dominated noise environments is proposed and evaluated using real vocalization and background noise recordings. Several considerations for extending the proposed method to watercraft emitted noise dominated environments are presented. Important derivations omitted in the main text and descriptions of the test setups are presented in Appendix B to Appendix E.

CHAPTER 2: RESEARCH BACKGROUND

The purpose of this chapter is to provide a comprehensive review of the existing literature including the algorithms and theories on the proposed adaptive signal enhancement methods, and to provide justification for these methodologies. The chapter begins with an introduction to manatee avoidance technology. A concise review of single channel adaptive noise cancellation and multi-channel adaptive array processing is presented next. This is followed by an introduction to the two methods investigated in this dissertation; wavelet domain denoising (WDD) and blind source separation (BSS). A review of the relevant literature on both these methods is also presented.

2.1 Manatee Avoidance Technology

The Florida Fish and Wildlife Conservation Commission (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 or warning manatees of a collision with watercraft. Gerstein and Blue [9] developed an active acoustic device which generates a narrowband, highly directional tone for warning manatees of an approaching watercraft. The watercraft borne device generates

narrowband tones in the 10 to 20 kHz range by combining a 200 kHz signal with another ultrasonic signal in the frequency band of 180 to 190 kHz. The center frequency of the tone is optimized to the measured hearing ability of the manatees, obtained from audiogram experiments [10]. Although the manatee audiogram suggests that a tone in the 10 to 20 kHz range may be detected by a manatee at a far greater range than the propeller generated noise of a slow moving recreational watercraft, the reaction of a manatee to such a tone has not been measured yet. It may take several collisions with a watercraft before a manatee can learn the threat associated with the tone. Furthermore, speed zones may still be necessary to allow sufficient time for the manatees to react to the approaching surface vessel as it is unclear that the possible advanced warning from a fast approaching watercraft equipped with the device will give the manatee adequate time to respond to the threat. Finally, despite using ultrasound frequencies that are well above the vocalization and hearing thresholds of manatees and other marine animals that inhabit the Florida waterways, active acoustic systems introduce artificial noise contamination into the habitat and have faced regulatory problems in the past.

An active sonar based manatee detector was proposed by Bowles et al. [11]. Unlike most other proposed methods, sonar systems do not require the manatee to perform a specific action (i.e., vocalize, surface, or exhale) for detection. The speed of sound, the acoustic attenuation coefficients and the density of various manatee tissues were measured from a post-mortem subject to evaluate the effectiveness of active sonar detection of manatees. Relative target strengths of live manatees in a pool were also measured using sonar with an operating frequency of 171 kHz. These experiments

revealed that the effective detection range using 171 kHz sonar was limited to 5 m, partially due to lower than expected reflectivity of the manatee and reverberant noise caused by the pool. The study concluded that sonar operating at 171 kHz was ineffective for detecting manatees to warn boaters of their presence [12]. Keith [13-14] proposed surface and underwater infrared cameras for detecting manatees through their exhalations and silhouette. The lack of heat signature in manatee exhalations made surface infrared detection very difficult and underwater infrared cameras were reported to have a maximum usable range of 15 m, under ideal conditions in clear water. A more realistic upper range in which accurate detection can be made in Florida waterways was estimated as 5 m, which is not sufficient for a feasible implementation of an underwater infrared based manatee detector.

Finally, several researchers have investigated passive acoustic based vocalization detection for warning watercraft of the presence of manatees. Warner [15] proposed a stand-alone Manatee Proximity Indicator which employed a neural network based vocalization recognizer. The measured acoustic signal is fed through a filter bank and the output signal energy for each filter is used as the input to the neural network. However, the detector was not adequately evaluated to make any conclusions on its performance. Herbert et al. [16] developed a PC-based manatee vocalization detector. The detector estimates the Welch periodogram from overlapping recording segments and performs noise reduction on the data, by computing the normalized spectrum and eliminating spurious peaks. A score is assigned to the recording based on the number, center frequency, and energy of resonant peaks and a detection decision is made by comparing

the score to a predetermined threshold. Although some promising detection results were reported using artificially mixed noisy vocalization recordings, these tests were not able to assess the effective detection range of the system. A patent was issued to this passive acoustic based manatee vocalization detector [17]. Mann et al. [18] developed a single channel passive acoustic vocalization detector that makes a detection decision based on the variation of the peak frequency for a given time frame within the band of 4 to 7 kHz. In-field vocalization broadcast tests conducted in a noisy habitat revealed that the vocalizations were easily discernable from the dominant background noise sources such as snapping shrimp and boat propeller generated noise. These tests suggest the effective detection range of the detector to be 10 m. The study concluded that the two dominant background noise sources (i.e., snapping shrimp and watercraft) limited the detection range. A multi-receiver configuration was proposed to perform beamforming for improving the signal to noise ratio (SNR) and to localize the manatee.

Niezrecki and Beusse [19] developed and tested three threshold based frequency domain detection algorithms. The most promising algorithm with the highest detection rates resulted in a detection decision based on the spectral energy content of four frequency bands: 0 to 5 kHz, 5 to 10 kHz, 10 to 15 kHz, and 15 to 20 kHz. To reduce the effects of noise, the frequency spectrum is computed from the unbiased estimate of the autocorrelation function. Two bandpass filters are also incorporated into the algorithm at various stages to reduce low and high frequency noise contamination. The persistency of the signal energy within each band is tracked to ensure that high energy shrimp snaps not be misclassified as vocalizations. The proposed detector was shown to achieve a 96%

detection rate for 16% false alarm rate for moderate SNR (2 to 4 dB) vocalizations predominantly corrupted with snapping shrimp noise [20-22].

Frisch and Haubold [23] reported that the most promising results for manatee avoidance technology have been obtained using passive acoustic systems based voice recognition and other advanced acoustic signal processing techniques developed for speech processing. The persistent harmonic structure of manatee vocalizations suggests that the time-frequency features provide a robust vocalization detection framework due to the high SNR at the harmonic frequencies. However, as explained in detail in the next section, the high acoustic power emitted from watercraft and other typical background noise sources diminish this high local SNR of the vocalizations and make the enhancement of the vocalization necessary prior to detection.

2.2 Enhancement of Manatee Vocalizations

Preliminary research conducted on passive acoustic detection based manatee avoidance technology revealed that a high SNR at the detector input to be a key factor in achieving satisfactory detector performance and detection range for threshold based detectors using spectral features. Phillips et al. [24] estimated the theoretical detection ranges of manatee vocalizations in the presence of diffuse ambient background noise and point noise sources (i.e., a watercraft navigating in the channel) with the assumption of simple transmission loss models. For example, assuming that channel transmission obeys the mixed spreading model, the theoretical detection range of an average manatee vocalization with 118 dB^a source level (SL) is estimated to drop from 1000 m to 10 m as

the ambient background noise sound pressure level (SPL) increase from 70 dB[b] to 100 dB for a 3 dB detection threshold (DT). The detection ranges as a function of background noise SPL, and DT are presented in Figure 2-1. If the pre-processing of the measured signals can improve the SNR by 10 or 20 dB in the presence of 100 dB ambient noise level and 3 dB DT, the detection ranges increase to 46.4 m and 215.4 m, respectively.

Theoretically, for a diffuse background noise SPL of 70 dB, watercraft emitted noise with a SL of 140 dB will be the dominant noise source if the watercraft is within the typical practical operating range of the vocalization detector. If the diffuse background noise SPL is increased to 100 dB, a 140 dB SL watercraft noise will dominate the noise field up to a range of 464.2 m from the receiver. With the ambient background noise level and the DT at 70 dB and 3 dB respectively, and the watercraft at a 500 m distance to the receivers, the maximum detection range is estimated to be 10.8 m. If the watercraft moves closer to a range of 250 m from the receivers, the maximum detection range drops to 5.4 m. It becomes impossible to detect the vocalizations (i.e., irrespective of the range of the manatee) if the boat is closer than 46.4 m to the receiver. This is demonstrated in Figure 2-2.

[a] All acoustic SL values are defined with reference to 1 µPa measured at a distance of 1 m, unless indicated otherwise.

[b] All acoustic SPL values are defined with reference to 1 µPa, unless indicated otherwise.

If pre-processing the measured signals can improve the input SNR by an average of 5 dB, manatee vocalizations can be detected from a range of 23.2 m if the watercraft is 500 m away, and at a range of 11.6 m if the watercraft is 250 m away. Likewise, if pre-processing can improve the input SNR at an average of 10 dB, the vocalizations can be detected at a range of 50 m and 25 m if the watercraft is at a range of 500 m and 250 m, respectively. A more detailed analysis of the improvements in the detection ranges is presented in Chapter 7.

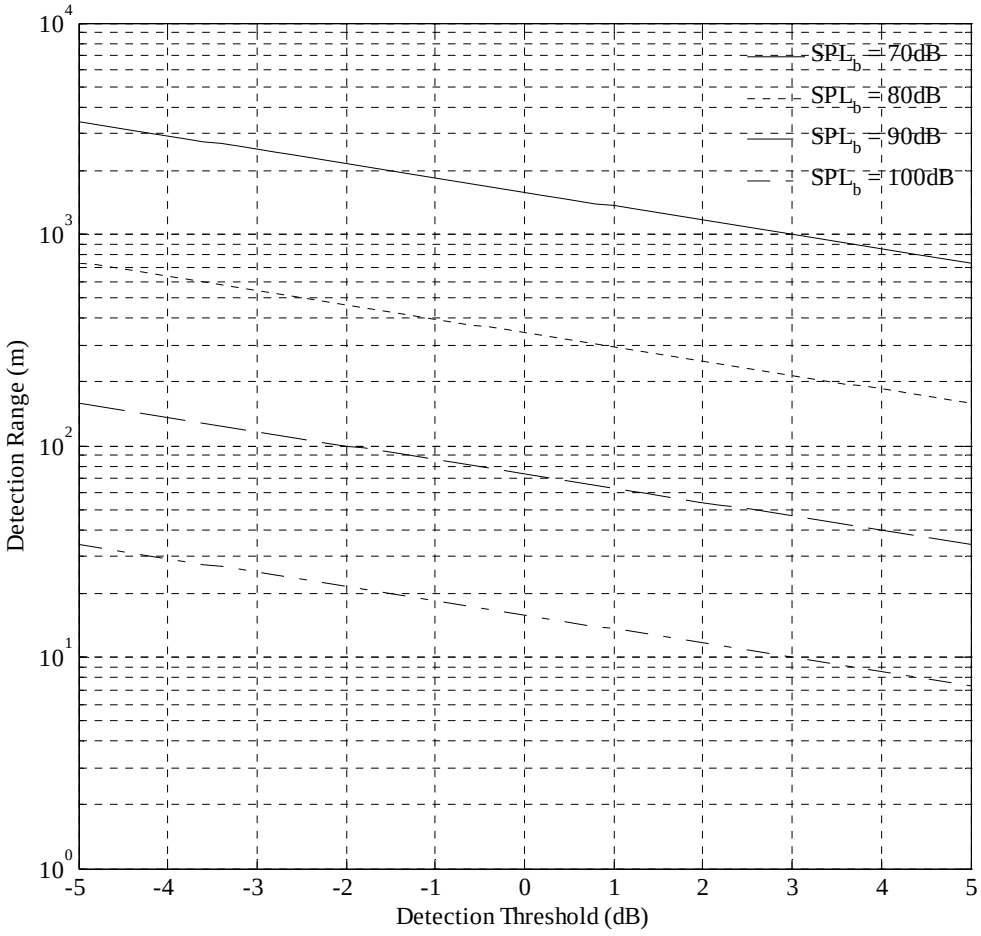


Figure 2-1. Manatee detection ranges for various diffuse background noise levels.

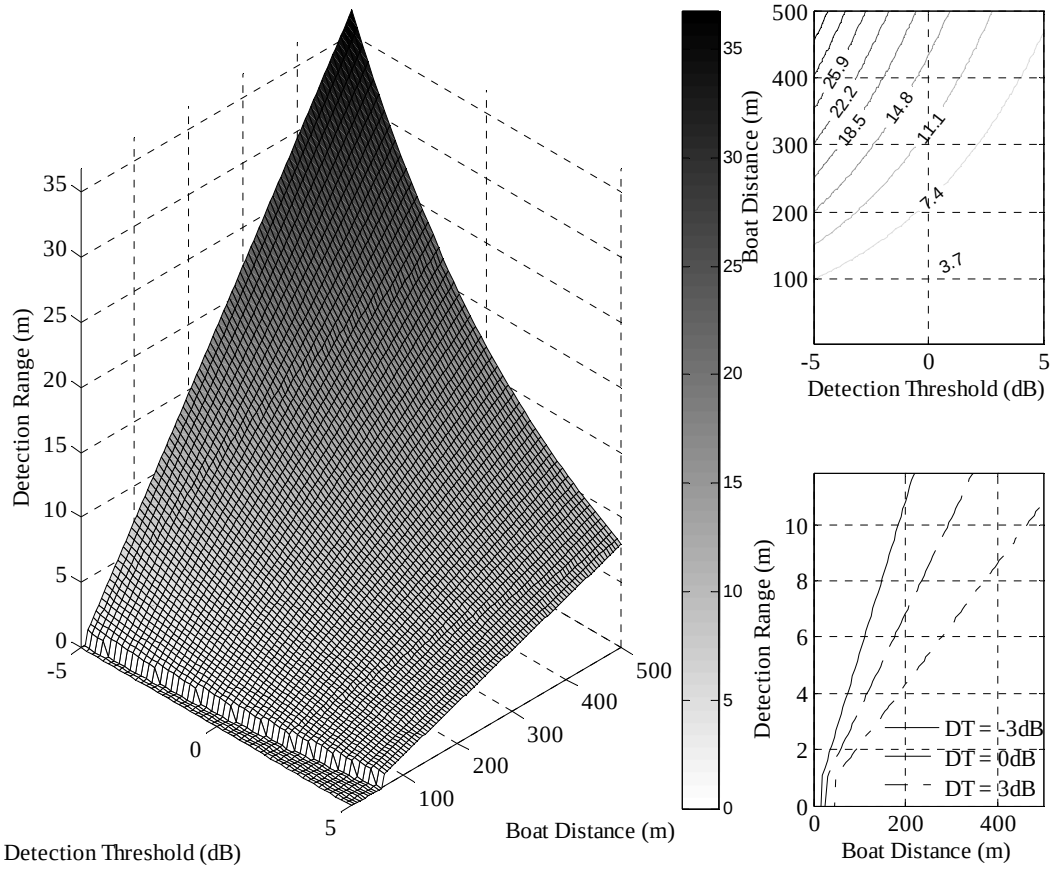


Figure 2-2. Manatee detection distances in the presence of watercraft emitted noise.

Having provided compelling arguments for enhancing manatee vocalization in the presence of watercraft, the remainder of this section is devoted to introducing the conventional single channel adaptive filtering approaches that have been previously proposed in the literature for enhancing manatee vocalizations.

The dynamic characteristics of both marine mammal vocalizations and the background noise environment make adaptive signal processing approaches very viable,

if not mandatory, for enhancing marine mammal vocalizations. The fundamental adaptive signal enhancement technique is the adaptive noise canceller (ANC) proposed by Widrow et al. [25]. The concept of the ANC is depicted in Figure 2-3. Essentially, the ANC is a two channel signal enhancer in which the adaptive filter input is constrained to a noise-only reference signal, and the desired signal represents the noisy vocalizations. The noise in the primary channel is related to the reference input signal through a filter $h(z)$ which represents the relative transfer function between the two receivers with respect to the point noise source $v(n)$. The adaptive filter converges to the pseudo channel impulse response to minimize the error signal. At convergence, the adaptive filter output is the noise signal at the primary channel, and the error is the estimated vocalization signal.

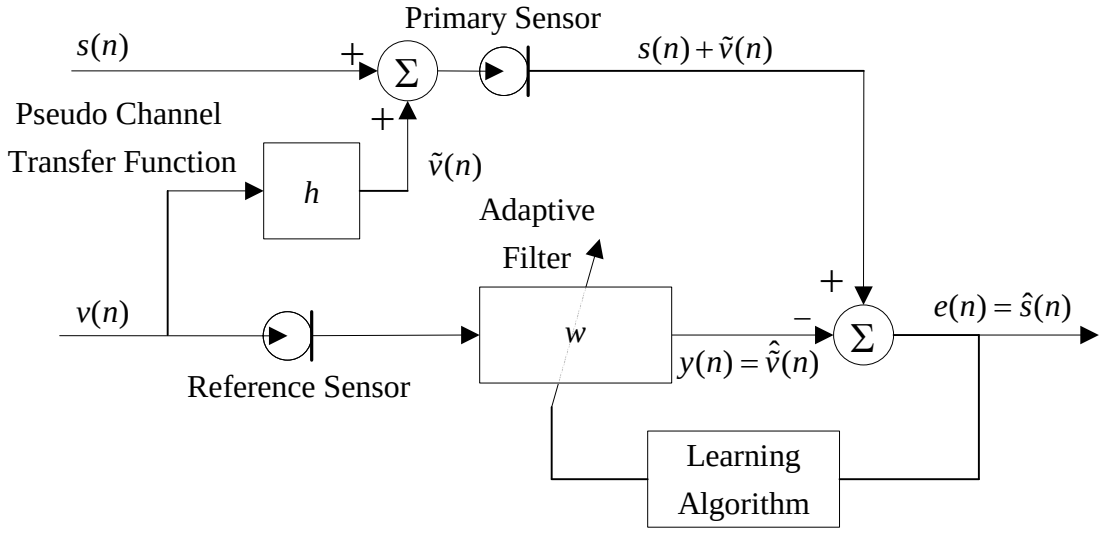


Figure 2-3. The block diagram of the ANC.

However, the ANC has a fundamental limitation. The vocalization free, noise only reference signal is very difficult to measure under most realistic scenarios and the

vocalization signals leak into the reference channel. This leakage results in the cancellation of the vocalization signal along with the noise signal, which limits the maximum achievable output SNR for the ANC. Adaptive enhancement of manatee vocalizations buried in background noise was investigated by Yan et al. [26-28]. To overcome the limitations of the ANC, the cited authors proposed the adaptive line enhancer (ALE) [25], [29] for enhancing manatee vocalizations. Manatee vocalizations are assumed to be narrowband compared to the background noise, and a delayed version of the primary sensor signal is used as the reference channel measurement in the ALE setup (see Figure 2-4).

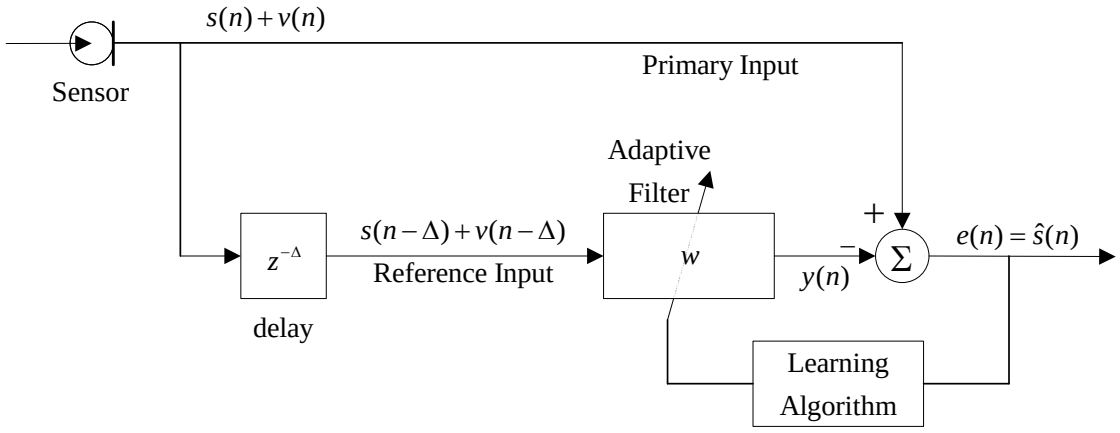


Figure 2-4. The ALE setup.

The delay is selected sufficiently large such that the broadband background noise becomes uncorrelated while the vocalization signals remain correlated in the primary and reference channels. The adaptive filter converges to a solution that minimizes the error signal between the filter output and primary channel measurements by suppressing the background noise signal and enhancing the narrowband vocalization. The resulting error

signal is an estimate of the noise-free vocalization signal. A detailed theoretical review of the ALE is presented by Yan [28]. The finite impulse response (FIR) structured ALE (FIR-ALE) [26] and a FIR structured ALE with feedback (FALE) [27] are implemented and are shown to enhance manatee vocalizations buried under typical underwater background acoustic noise. As an alternative to single channel linear adaptive filtering of enhancing manatee vocalization, non-linear WDD is investigated in this dissertation. A review of WDD is presented in Section 2.4 and an adaptive WDD algorithm is developed in Chapter 4.

The signal enhancement performance of single channel algorithms can be improved by extensions to multi-channel setups. The conventional multi-channel adaptive signal enhancement approach for passive sonar applications is the adaptive beamformer, which is reviewed in the next section. Some limitations of the adaptive beamformer that make it unsuitable for enhancing weak manatee vocalizations are also presented.

2.3 Adaptive Beamforming

The conventional approach to multi-channel noise reduction is adaptive beamforming. The beamforming utilizes the known geometric information about the location of the sensors relative to each other and the location of the acoustic point sources to adaptively filter the measured multi-channel signals. Adaptive processing is necessary to adjust the beamformer response to changing source locations and source signal

statistics. The schematic diagram of a conventional Q-channel adaptive beamformer is depicted in Figure 2-5. The beamformer output $y(n)$ is expressed as

$$\begin{aligned} y(n) &= \sum_{q=1}^{Q}\sum_{l=0}^{L-1} w_{q,l}(n)x_q(n-l) \\ &= \mathbf{w}^H(n)\mathbf{x}(n), \end{aligned} \tag{2-1}$$

where $\mathbf{w}(n) = \begin{bmatrix} \mathbf{w}_1^T(n) & \mathbf{w}_2^T(n) & \cdots & \mathbf{w}_Q^T(n) \end{bmatrix}^T$ is the $QL \times 1$ vector of the adaptive filter weights (each filter is of order L) for each channel and $\mathbf{x}(n)$ is the similarly defined $QL \times 1$ input vector. Assuming that the incoming signal emitted from a point source is a unit amplitude time-harmonic plane wave (see Appendix B) at frequency $\omega = 2\pi f$ and zero phase at the first sensor (i.e., $x_1(n) = \exp(j\omega n)$), the beamformer output can be expressed as

$$\begin{aligned} y(n) &= x_1(n)\sum_{q=1}^{Q}\sum_{l=0}^{L-1} w_{q,l}(n)\exp(-j\omega l)\exp\left(-j\omega k_q(\theta)\right) \\ &= x_1(n)\mathbf{w}^H(n)\mathbf{d}(\omega,\theta), \end{aligned} \tag{2-2}$$

where $\mathbf{d}(\omega,\theta) = \begin{bmatrix} \mathbf{d}_1^T(\omega,\theta) & \mathbf{d}_2^T(\omega,\theta) & \cdots & \mathbf{d}_Q^T(\omega,\theta) \end{bmatrix}^T$ is the steering vector that consists of the $L \times 1$ vectors

$$\mathbf{d}_q(\omega,\theta) = \begin{bmatrix} \exp\left(-j\omega\left(k_q(\theta)\right)\right) & \exp\left(-j\omega\left(1+k_q(\theta)\right)\right) & \cdots & \exp\left(-j\omega\left(L-1+k_q(\theta)\right)\right) \end{bmatrix}^T, \tag{2-3}$$

and $k_q(\theta)$ is the propagation delay between first the qth receiver that is a function of the direction of arrival (DOA) θ.

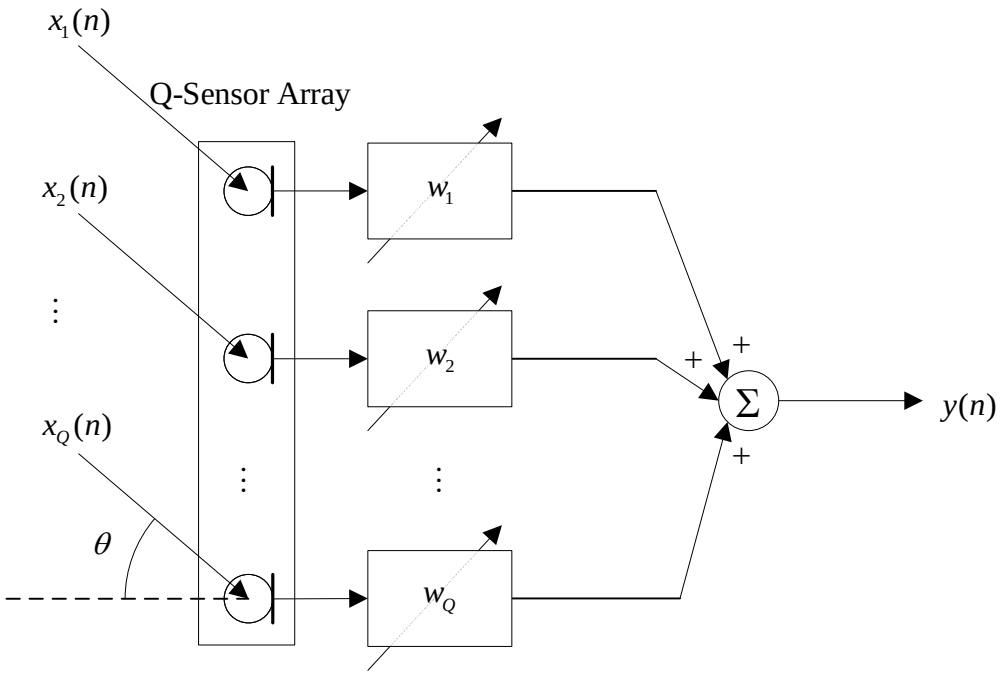


Figure 2-5. The schematic of an adaptive beamformer.

Depending on the statistical optimization criteria, several different types of beamformers have been proposed in the literature (for a review, see [30]). In particular, the minimum variance distortionless response (MVDR) beamformer (also known as the Capon beamformer) adaptively adjusts the filters to minimize the output signal power

$$\min_{\mathbf{w}} \left\{ E[y^2(n)] \right\} = \mathbf{w}^H(n)\mathbf{R}_{xx}(n)\mathbf{w}(n), \tag{2-4}$$

where $E[\cdot]$ is the expectation operator, and $\mathbf{R}_{xx}$ is the input signal covariance matrix, subject to the constraint that the signal incoming from the location of the point source of interest (i.e., the vocalizing mammal) remain unaltered [31]

$$\mathbf{w}^H(n)\mathbf{d}(\theta,\omega)=1. \tag{2-5}$$

The resulting MVDR beamformer takes the form

$$\mathbf{w}_{\text{MVDR}}(n)=\frac{\mathbf{R}_{xx}^{-1}(n)\mathbf{d}(\theta,\omega)}{\mathbf{d}^H(\theta,\omega)\mathbf{R}_{xx}^{-1}(n)\mathbf{d}(\theta,\omega)}. \tag{2-6}$$

The above presented formulation of the MVDR can be expanded to incorporate multiple constraints and is very suitable for adaptive estimation of the filter weights using the least mean squared (LMS) or a similar algorithm [32].

In general, the location of the sources is not known a priori and DOA estimation (see [33] for a review) 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 [34] 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. In addition, the experiments conducted in the cited research were made with the manatees as the only acoustic point source. For the passive acoustic based manatee avoidance system proposed in this research, the DOA estimation for the harmonic manatee vocalization signals need to be performed in the presence of temporally and spatially correlated Gaussian noise emitted from surface vessels, which may require the use of higher order statistics (HOS) based algorithms (e.g., [35]). Alternatively, the DOA can be estimated by making the beamformer scan each radial angle for high power incoming acoustic signals. However, the signal power of weak vocalization signals will generally be low and cannot be used to determine the corresponding DOA. Furthermore, in the presence

of other dominant point sources such as surface vessels, the adaptive MVDR beamformer may suffer from the signal cancellation phenomenon [36], which has similar effects as the leakage phenomenon of the ANC. The signal cancellation phenomenon can adversely affect the mean squared error (MSE) performance of an adaptive beamformer by cancelling the vocalization signal as well, in particular, if the harmonics of manatee vocalizations and the harmonics created by the propeller shaft nearly overlap. In contrast to adaptive beamforming, BSS does not utilize geometric information and does not require DOA information, which eliminates the need to precisely know the sensor locations and to have high SNR measurements. Furthermore, BSS algorithms separate the mixed source signals and are not prone to the adverse effects of the signal cancellation phenomenon. Therefore, in this dissertation, BSS is investigated as an alternative to adaptive beamforming. A review the generic BSS problem is presented in Section 2.5. A novel BSS algorithm for enhancing manatee vocalizations is developed in Chapter 5.

2.4 Wavelet Domain Signal Enhancement

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 non-stationary signals. A major drawback of the STFT is that it divides the time-frequency domain into a regular grid. The wavelet transform, on the other hand, provides an irregular grid with varying time and frequency resolutions. In particular, the wavelet transform sacrifices time resolution in favor of frequency resolution at low frequencies, and vice versa. This

decomposition is more intuitive for analyzing non-stationary signals since high frequency transients are generally shorter in duration and require more precise time resolution. In contrast, low-frequency signals are of longer duration, and coarser time resolution is sufficient for their representation. In addition, the wavelet transform enables the use of different basis functions which may be more suitable for irregular transient signals.

The J-level discrete wavelet transform (DWT) coefficients of a signal $x(n)$ are defined as the output of a J-level octave band perfect reconstruction quadrature mirror filter (QMF) bank depicted in Figure 2-6 (a). If the individual cascaded lowpass and highpass filters of the octave band decomposition are combined into a single filter as shown in Figure 2-6 (b), the wavelet coefficients can be expressed as

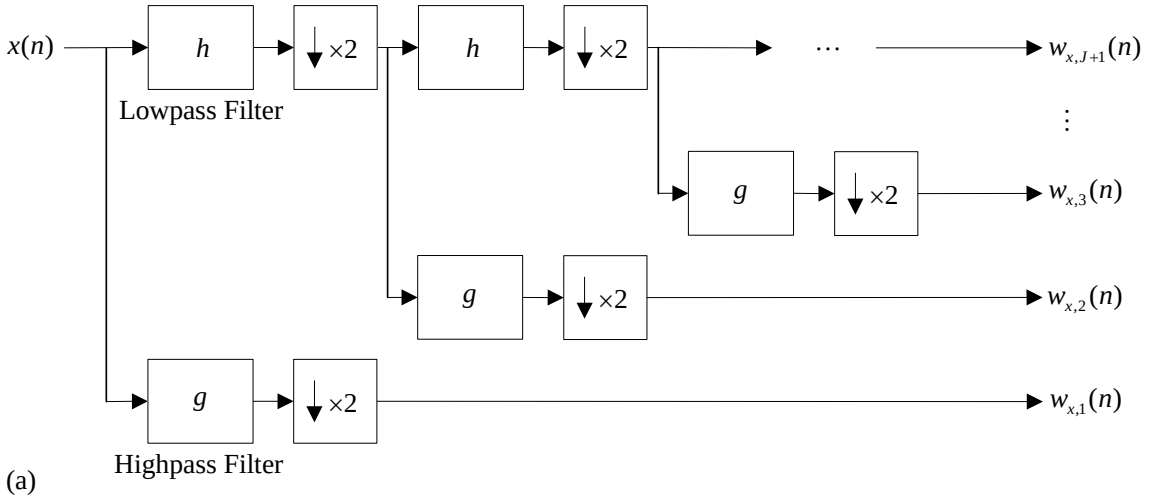
$$\begin{aligned} w_{x,J+1}(n) &= \sum_{m=-\infty}^{\infty} \tilde{h}_J\left(2^J n - m\right) x(m) \\ w_{x,j}(n) &= \sum_{m=-\infty}^{\infty} \tilde{g}_j\left(2^j n - m\right) x(m), \quad 1 \le j < J, \end{aligned} \tag{2-7}$$

where the combined highpass and lowpass filters are defined as

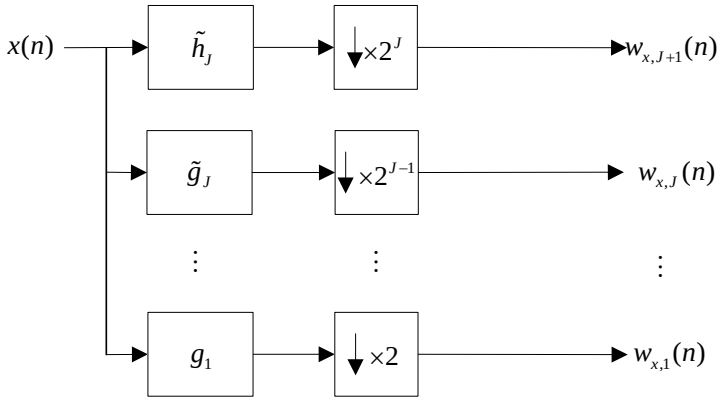
$$\begin{aligned} \tilde{h}_J(z) &= \prod_{k=1}^{J} h\left(z^{2^k}\right) \\ \tilde{g}_j(z) &= g\left(z^{2^j}\right) \prod_{k=1}^{J-1} h\left(z^{2^k}\right). \end{aligned} \tag{2-8}$$

The impulse responses $\tilde{g}_j\left(2^j n\right)$ and $\tilde{h}_J\left(2^J n\right)$ are the wavelet basis functions at level j and the scaling function at level J, respectively [37-38]. The DWT can be reversed by replacing downsampling with upsampling and the decomposition filters $\tilde{h}_J(z)$ and $\tilde{g}_j(z)$ by their corresponding perfect reconstruction QMF pairs.

The downsampling that follows the lowpass and highpass filtering at each level of decomposition makes the DWT a non-redundant transform. If the downsampling is eliminated and the decomposition filters are upsampled to account for the lack of downsampling the coefficients, the resulting transform is the redundant undecimated discrete wavelet transform (UDWT). Only the lowpass filter outputs are further decomposed in the DWT, which results in the dyadic decomposition time-frequency plane. Alternatively, if both the lowpass and highpass filter outputs are decomposed, the resulting non-redundant transform is termed the wavelet packet transform (WPT). Similar to the UDWT, eliminating downsampling from the WPT results in the redundant undecimated wavelet packet transform (UWPT).



(a)



(b)

Figure 2-6. The filter bank implementation of the DWT. (a) The J -level DWT filter bank implementation and (b) an equivalent representation with the cascaded filters combined. The blocks with a downward pointing arrows represent downsampling.

In the context of signal enhancement, the wavelet transform and the related concept of subband filtering has found two major applications. These are wavelet domain or subband implementation of adaptive filters for improving the convergence rate

of the filters and wavelet coefficient shrinkage based denoising of noisy signals. In what follows, brief descriptions of both these applications are provided.

2.4.1 Wavelet Domain and Subband Adaptive Filtering

The convergence speed of the LMS algorithm is influenced by the eigenvalue spread of the input autocorrelation matrix [39]. The eigenvalue spread of a Toeplitz autocorrelation matrix can be shown to asymptotically approach the ratio of the highest and lowest values of the power spectral density (PSD) [40]; and the eigenvalue spread for temporally correlated signals such as speech or marine mammal vocalizations are large. Thus, the convergence rate of LMS based signal enhancement algorithms suffers in the presence of temporally correlated inputs. The recursive least squares (RLS) algorithm has improved convergence properties compared to the LMS algorithm, mainly because the input signals are pre-whitened (and thus, the eigenvalue spread of their autocorrelation matrix is reduced) by incorporating the inverse of an estimate of the input signal autocorrelation matrix into the filter update equation. Alternatively, the discrete Fourier transform (DFT) and the discrete cosine transform (DCT) are known to approximately diagonalize Toeplitz matrices. Thus, DFT and DCT domain implementations of adaptive filters enjoy improved convergence rates for inputs that have autocorrelation matrices with large eigenvalue spreads [41].

Wavelet basis functions have also been shown to approximate the eigenvectors of the autocorrelation matrix of a wide class of signals. Thus, the wavelet transform behaves similar to the data dependent Karhunen-Loeve transform (KLT) and

approximately diagonalizes the input autocorrelation matrix [42]. Based on this property of the DWT, wavelet domain adaptive filters have been proposed as alternatives to conventional transform domain approaches such as the DCT and DFT [43-45]. Implementing separate adaptive filters for each subband have also lead to improvements in the convergence rate [46], the tracking ability, and the computational complexity [47] of the filters compared to a regular full-band filters in adaptive echo cancellation applications. For temporally correlated signals such as manatee vocalizations, the eigenvalue spread is considerably large, forcing a small step size to ensure stability. Therefore, wavelet domain and subband adaptive filters may improve the convergence rate of the ALE implementations proposed by Yan et al. [26-27]. The block diagram of a possible implementation of wavelet domain ALE is depicted in Figure 2-7.

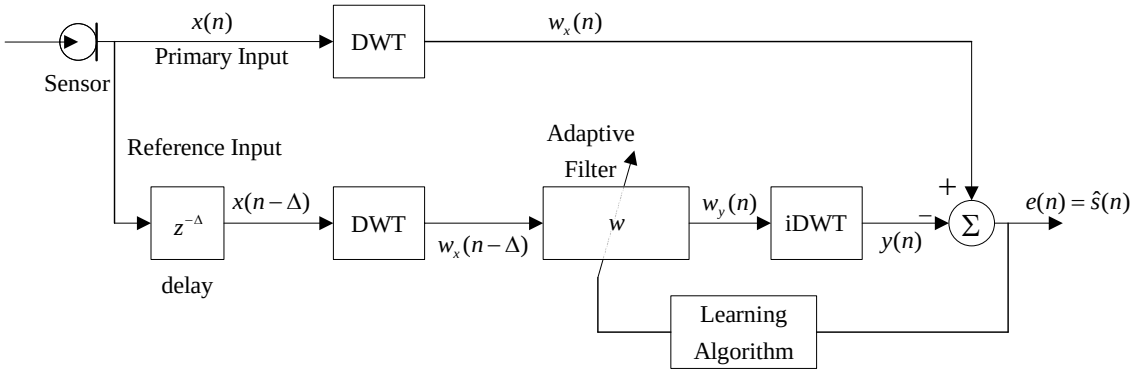


Figure 2-7. The wavelet domain implementation of the FIR-ALE.

However, wavelet domain and subband adaptive filtering requires the use of perfect reconstruction QMF filter banks (with no amplitude and phase distortion) which introduce a delay to the error signal. This delay limits the improvement that can be

achieved in the convergence rate of the subband adaptive filter [48]. For this reason, wavelet domain and subband adaptive filtering is not investigated any further in this dissertation.

2.4.2 Wavelet Domain Denoising

WDD was first introduced in the seminal papers of Donoho and Johnstone [49] and Donoho [50] as a transform domain method for estimating a deterministic function from a set of noisy measurements. The appeal of WDD stems from the ability of wavelet basis functions to compactly represent signals with different time-frequency properties, as was described in the previous subsection. More specifically, most of the signal energy is efficiently compressed and represented with a few large amplitude coefficients in the wavelet domain. Noise signals, on the other hand, are not compressed as effectively and are mapped as many small amplitude coefficients. Thus, the wavelet transform maps noisy signals to a domain in which signal and noise can be discerned more easily, simplifying signal estimation and noise reduction. WDD algorithms exploit this property of the wavelet coefficients and achieve noise reduction by eliminating small amplitude coefficients below a threshold. The estimate of the denoised signal is retrieved through the inverse wavelet transform of the remaining coefficients. The above cited authors show that wavelet domain non-linear signal estimators outperform linear estimators and asymptotically achieve near optimal performance for the minimax MSE criterion (i.e., minimizes the maximum possible MSE for all functions that belong to a certain smoothness class, [51]).

In the last two decades, the wavelet transform has been extensively used for improving feature extraction in passive sonar (e.g., [52-53]) and in marine mammal monitoring applications (e.g., [54-56]). Weiss and Dixon [57] applied the wavelet transform for eliminating noise contamination due to dominant scatters in multi-channel active sonar applications. The preliminary research conducted by the dissertation author was the first published work [58] on WDD of noisy marine mammal vocalizations. More recently, Ren et al. [59] proposed a wavelet domain non-linear adaptive filter for enhancing bioacoustic signals. In this dissertation, an alternative WDD approach, based on an adaptive threshold rule is proposed for enhancing manatee vocalizations in the presence of watercraft emitted noise.

2.5 Blind Source Separation

BSS is a class of adaptive signal processing algorithms that serve for retrieving the original source signals from multi-channel mixtures. These algorithms are referred to as ‘blind’ because both the source signals and the mixing channels are unknown. Therefore, BSS algorithms are versatile methods for enhancing noisy vocalizations of a variety of marine mammal species recorded in environments with very limited and generic information regarding underwater acoustic channel. Separation of sources from the measured signals is achieved by unsupervised adaptive filtering, which is in contrast to conventional supervised adaptive filtering methods for noise cancellation that require an often difficult to measure reference signal. BSS is related to, but significantly different than, blind deconvolution. The latter refers to a group of unsupervised

algorithms that recover an original independent and identically distributed (IID) source signal from only the measurements of the source after being transmitted through a (possible non-minimum phase) channel. Blind deconvolution algorithms involve a single source and a single channel whereas BSS algorithms process multi-channel mixtures of several sources [60]. Figure 2-8 depicts the generic model for blind separation of P source signals from Q mixture measurements.

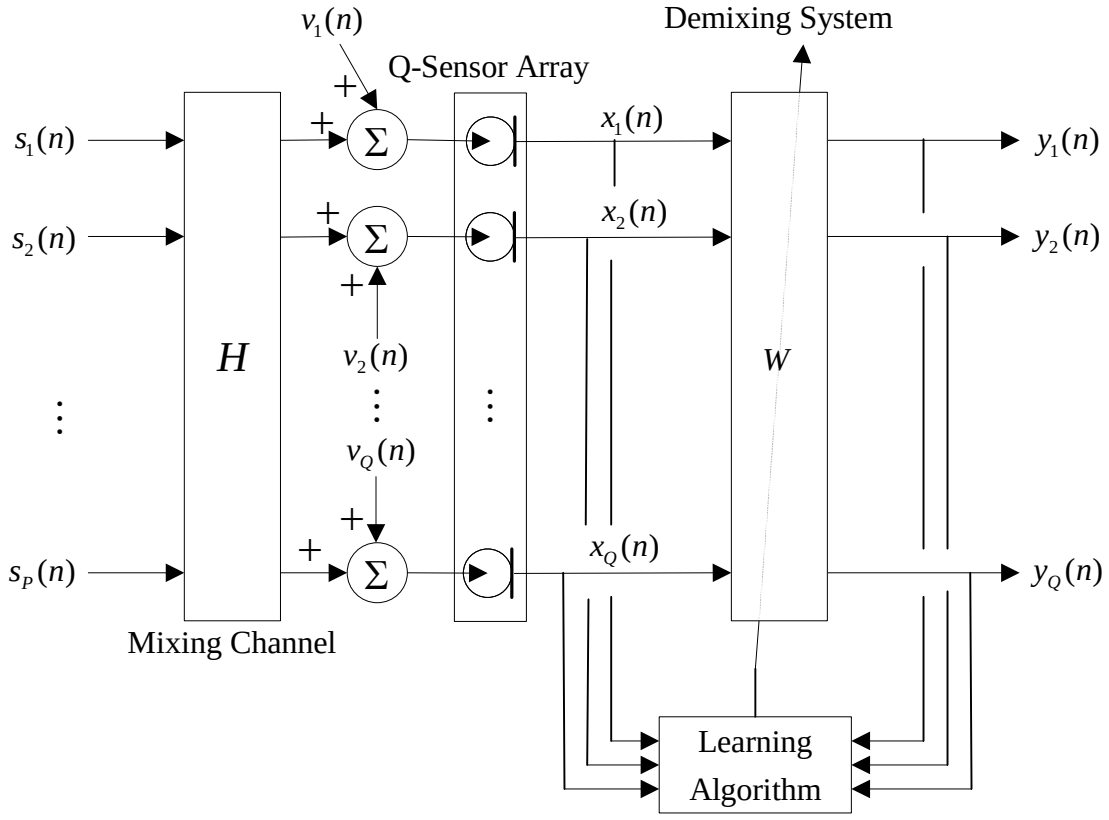


Figure 2-8. The generic model for BSS.

The Q measurements are a superposition of a linear mixture of the P sources and IID noise

$$\mathbf{X} = \mathbf{H} \cdot \mathbf{S} + \mathbf{V}, \tag{2-9}$$

where $\mathbf{X}$, $\mathbf{S}$, and $\mathbf{V}$ are the multi-channel measurements, sources, and noise, respectively. The mixing channel is represented by the matrix $\mathbf{H}$ with entries that model the individual transfer functions between the sources and the receivers. The two structures of the channel transmission matrix $\mathbf{H}$ that are relevant to underwater acoustics are explained below.

1. If $\mathbf{H}$ is a matrix of time varying scalars, it represents a time varying instantaneous mixing channel. Multi-path acoustic channels (e.g., the underwater acoustic channel) cannot be modeled as instantaneous mixing environments. Nevertheless, instantaneous mixing is important since convolutive acoustic channels (defined next) can be approximated as instantaneous mixing channels in the frequency domain.
2. If $\mathbf{H}$ is a matrix of time varying polynomials, it represents a convolutive mixing channel. The underwater acoustic channel is a multi-path, convolutive channel and is modeled exclusively as a matrix of FIR filters in the time domain.

The estimates of the sources are obtained at the output of the de-mixing system

$$\mathbf{Y} = \mathbf{W} \cdot \mathbf{X}. \tag{2-10}$$

For complete source recovery, the de-mixing system is required to converge to the unique solution which satisfies $\mathbf{W} \cdot \mathbf{H} = \mathbf{I}$. Thus, an intuitive condition for source separation is that the mixing is reversible (i.e., $\mathbf{H}$ is invertible). Even so, the blind separation problem is ill-conditioned without further assumptions. The common presumption in all BSS

algorithms is the statistical independence of source signals, which can be physically justified. Essentially, all BSS algorithms solve for a de-mixing solution that will effectively make the outputs statistically independent, but differ in the independence criterion. In general, the de-mixing system can only converge within a permutation matrix $\mathbf{P}$ and diagonal shaping matrix $\mathbf{D}$ of the true solution

$$\mathbf{W} \cdot \mathbf{H} = \mathbf{P} \cdot \mathbf{D}, \tag{2-11}$$

which will result in source separation rather than complete source retrieval. The diagonal entries of the matrix $\mathbf{D}$ are scalars in the frequency domain and polynomials in the time domain for convolutive mixtures. The permutation and scaling/filtering ambiguities are inevitable for generic setups as the one depicted in Figure 2-8, if only statistical independence of the outputs is used as a criterion. This is because two statistically independent signals will remain statistically independent after an arbitrary scaling/filtering. In addition, the order at which the separated signals appear at the different channels of the de-mixing system does not affect and cannot be manipulated by the statistical independence criteria. Permutation and filtering ambiguities are relatively trivial in time domain BSS algorithms since retrieving filtered versions of the source signals in an arbitrary order is sufficient for most applications. In contrast, correcting for the permutation ambiguity is essential in frequency domain approaches since such ambiguities result in the random distribution of the spectral components of the separated source signals over the output channels.

BSS algorithms can be grouped into four categories based on their statistical independence criterion; entropy, HOS, Bussgang, and second order statistics (SOS)

algorithms. Entropy based algorithms exploit the fact that the joint probability density function (PDF) of the separated outputs (which should be statistically independent if separation is achieved) can be expressed as a product of the marginal PDF of each output or equivalently, of each source PDF. Thus, entropy based approaches require precise knowledge of the source PDF, which is in general not available a priori. Alternatively, source PDF can be estimated from the measurements. Unfortunately, accurately estimating the source PDF from noisy measurements of weak vocalizations is very difficult.

Statistical independence of outputs implies vanishing cross-cumulants or cross-polyspectra which can be cast into HOS based separation criterion. HOS based approaches require the source signals to be non-Gaussian since the third and fourth order cumulants of Gaussian processes are zero [61]. In Section 3.3, watercraft will be modeled as a Gaussian noise source. Hence, HOS based BSS approaches are not suitable for extracting manatee vocalizations from noisy measurements dominated by watercraft emitted noise.

The autocorrelation sequence of an IID Bussgang process can be expressed in terms of a non-linear function of the process. This non-linearity is generally derived from the PDF of the Bussgang random variable. Thus, the Bussgang criterion implicitly assumes that the source PDF is known a priori. Adaptive algorithms similar to those of supervised adaptive filtering can be devised by minimizing the cross-correlations of the estimated outputs expressed in terms of this non-linearity. Bussgang approaches are not suitable for enhancing manatee vocalizations which are not IID and in contrast, exhibit

strong temporal correlation. Among these four approaches, SOS based methods that impose decorrelation of the measurements rather than independence is investigated in this dissertation for the reasons outlined in the next subsection.

2.5.1 Second Order Statistics Based Blind Source Separation

Decorrelation based source separation algorithms employ SOS of the de-mixing system outputs. Adaptively decorrelating the outputs

$$E[y_p(n)y_q(n-k)] = r_{y_p y_q}(n,k) = 0, \quad \forall n,k, \quad p \neq q = 1,2,\ldots,P, \tag{2-12}$$

where $y_q(n)$ is the output at the qth 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. In contrast to the other BSS approaches, decorrelation based methods can accommodate multiple Gaussian sources. In addition, SOS based approaches do not require the a priori knowledge or the estimation of the source probability density function (PDF). Furthermore, most vocalization signals exhibit strong temporal correlation and, thus violate the IID source signals assumption of some BSS approaches. Finally, for a predefined level of consistency, SOS estimates require considerably fewer samples compared to HOS, making decorrelation based approaches more attractive for on-line implementation in time varying, dynamic environments. Experimental results provided in subsequent chapters suggest that decorrelation based SOS-BSS algorithms perform well for signals with temporal time structures such as marine mammal vocalizations in the presence of Gaussian noise sources. The above presented attributes of SOS based

criterion provide compelling arguments to pursue multi-channel enhancement of marine mammal vocalizations through decorrelation based BSS approaches.

2.5.2 Unsupervised Source Separation for Adaptive Noise Cancelling

An intuitive understanding of the two-input two-output (TITO) source separation algorithms can be achieved by considering the ANC setup depicted in Figure 2-3. The ANC requires a vocalization-free noise signal in the input of the adaptive filter that is correlated with the noise component of the measurements made in the primary receiver. Alternatively, if noise was of primary interest, the ANC would require a noise-free vocalization signal at the input of the adaptive filter. Merging these two configurations, the output of each adaptive filter can be used as the input of the other adaptive filter resulting in the double feedback structure shown in Figure 2-9 (a). An equivalent feedforward setup is given in Figure 2-9 (b). In general, the separating filters are restricted to causal FIR filters. The resulting algorithm depicted in Figure 2-9 (b) is termed the symmetric adaptive decorrelation (SAD) algorithm [62].

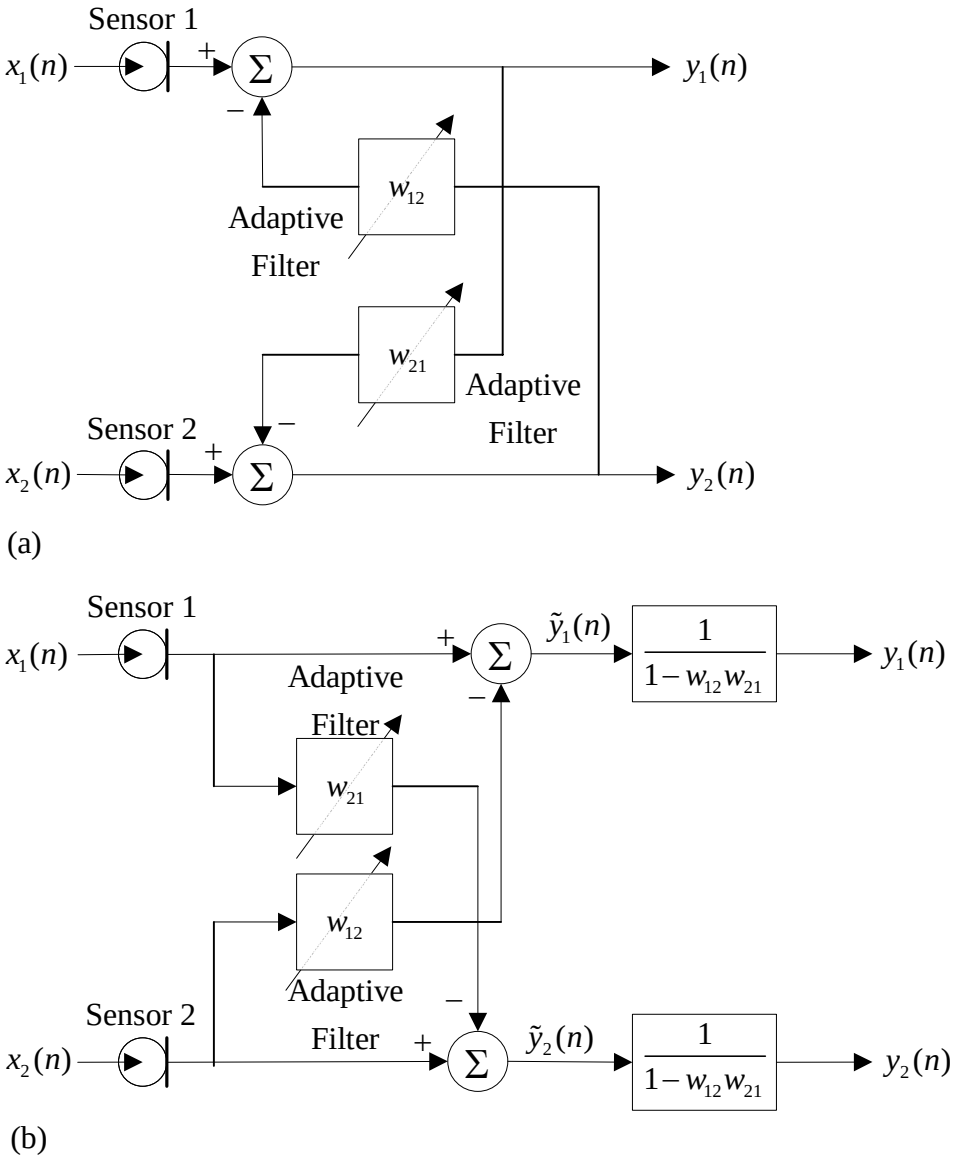


Figure 2-9. The block diagram of the double ANC setup. (a) In the feedback configuration and (b) the equivalent feedforward setup.

In general, the feedforward setup is preferred over the feedback setup to avoid the potentially unstable feedback and the infinite impulse response (IIR) post-processing filter $1/[1-w_{12}(z)w_{21}(z)]$ is stable provided that $\left|w_{12}(z)w_{21}(z)\right|<1, \quad \forall z=\exp(j\omega)$. In

adaptive filtering theory, the principle of orthogonality requires that the reference input to each adaptive filter be uncorrelated to the error signal at convergence. The filter input and error signals for the two adaptive filters in Figure 2-9 (a) are the outputs of the de-mixing system $y_q(n), \quad q=1,2$. Therefore, the SAD cost function can be expressed as the decorrelation of the two de-mixing system outputs at L lags

$$\mathcal{J}(n) = E[y_1(n)y_2(n-k)], \quad 0 \le k \le L-1, \tag{2-13}$$

where L is the separating filter order. Source separation is anticipated when the cost function is minimized. The SAD algorithm uses a stochastic gradient update rule which results in the following de-mixing filter update equations

$$\begin{aligned} \mathbf{w}_{12}(n+1) &= \mathbf{w}_{12}(n) + \mu \tilde{y}_2(n)\tilde{\mathbf{y}}_1(n) \\ \mathbf{w}_{21}(n+1) &= \mathbf{w}_{21}(n) + \mu \tilde{y}_1(n)\tilde{\mathbf{y}}_2(n), \end{aligned} \tag{2-14}$$

where μ is the step size and $\tilde{y}_q(n)$ are the intermediate outputs (see Figure 2-9 (b)) for the qth channel. The SAD algorithm developed above is used as a benchmark to compare the proposed BSS algorithms developed in Chapter 6.

2.5.3 Alternative Second Order Statistics Based Blind Source Separation Algorithms

Aside from the SAD algorithm, numerous decorrelation based SOS-BSS algorithms have been proposed for both the TITO and the multi-input multi-output (MIMO) configurations in the context of multi-channel speech processing in the last decade. The cost function given in Eq. (2-13) can be reformulated and combined with different optimization methods to derive block-based iterative, recursive, or sequential

source separation algorithms implemented in the time domain (e.g., [63-65]). However, while statistical independence implies decorrelation, the opposite is not true. More specifically, it is possible for the de-mixing filters to converge to a solution that effectively decorrelates the outputs, but fails to achieve source separation. For TITO system, most such incorrect solutions require IIR separating filters. Weinstein et al. [63] suggested constraining the structure of the mixing channel and the de-mixing system to FIR filters in order to reduce the possibility of the separating filters converging to incorrect solutions. For the TITO configuration, Lindgren and Broman [65] proved that separation can be achieved by decorrelation alone if the mixing system is FIR structured and is minimum phase, and thus possesses a stable and causal inverse. A more detailed analysis of the TITO system is presented in Chapter 6.

Another constraint that can be imposed for achieving separation by decorrelation is to exploit the non-stationarity of speech signals. In this approach, the separating filters are obtained as the solution that simultaneously diagonalizes several input cross-correlation matrices computed at different times. Parra and Spence [66] expanded on this formulation and developed an iterative, FIR structured frequency domain source separation algorithm that minimizes the cost function

$$\mathcal{J} = \sum_{\omega} \sum_{k=1}^{K} \left(\mathbf{W}(\omega)[\mathbf{P}_{xx}(\omega,k) - \mathbf{P}_{vv}(\omega,k)]\mathbf{W}^{H}(\omega) \right)^{2}, \tag{2-15}$$

where $\mathbf{P}_{xx}(\omega)$ and $\mathbf{P}_{vv}(\omega)$ are the cross-spectral density matrices (CSDM) for the input signals and exogenous noise, respectively, K is the total number of CSDM that are jointly diagonalized, and $\mathbf{W}(\omega)$ is the matrix of the separating filters in the frequency

domain. Each of the K CSDM of the input signals can be estimated from N consecutive or overlapping blocks using the M -point DFT of the measurements

$$\mathbf{P}_{xx}(\omega) = \frac{1}{N}\sum_{n=1}^{N}\mathbf{x}(\omega,n)\mathbf{x}^{H}(\omega,n), \tag{2-16}$$

where $\mathbf{x}(\omega)$ is the DFT of the time domain noisy measurements. In this algorithm, the FIR filtering of the noisy measurements (i.e., linear convolution) is approximated with the DFT which implements circular convolution. The approximation error associated with modeling the time domain linear convolution with the frequency domain circular convolution is minimized by selecting a larger DFT order with respect to the separating filter orders. In that case, most of the separating filter coefficients will be zero. The permutation of the filter coefficients is also avoided by restricting the latter part of the time domain filter coefficients to be zero. This restriction ensures that the separating filters have continuous and smooth frequency responses. In Chapter 6, the Parra algorithm outlined above is converted to a block update form [67] by replacing Eq. (2-16) with the recursive estimator

$$\mathbf{P}_{xx}(\omega,n) = \gamma\mathbf{P}_{xx}\left(\omega,n-1\right) + (1-\gamma)\mathbf{x}(\omega,n)\mathbf{x}^{H}(\omega,n), \tag{2-17}$$

where γ is the forgetting factor, and used as a benchmark in evaluating the performance of the proposed BSS algorithms.

In addition to the non-stationarity, the temporal correlation of the source signals can be used to achieve or improve source separation. Weinstein et al. [68] proposed exploiting the temporal correlation of noisy speech signals for signal enhancement by

developing a parametric TITO state-space mixture model. The cited authors modeled the cross coupling channels as FIR filters. Speech and background noise are modeled as an autoregressive (AR) process and IID Gaussian noise, respectively. The maximum likelihood (ML) estimates of the FIR filter and AR speech signal parameters are computed using the estimate-maximize (EM) algorithm. An estimate of the noise-free speech signal is obtained as a by-product of the EM algorithm. The iterative EM algorithm is converted to a sequential, time domain algorithm using Kalman filtering. A similar algorithm was also proposed for MIMO BSS of noisy speech signals [69]. Alternatively, if the source signals are cyclo-stationary (i.e., their autocorrelation sequence is periodic), this property can be exploited to incorporate the temporal correlation into the source separation criteria [70]. More recently, Buchner et al. [71] proposed a generic SOS based BSS algorithm that incorporates both the non-stationarity and temporal correlation of the source signals for separation. In the next chapter, manatee vocalizations and watercraft emitted noise will be modeled as AR and IID Gaussian noise signal, respectively. Furthermore, due to their harmonic structure, manatee vocalizations have an almost periodic autocorrelation sequence and can be considered as cyclo-stationary signals. As a part of this research, algorithms based on the AR signal model or the cyclo-stationarity of manatee vocalizations were evaluated and shown to perform exceptionally well with artificial signals in numerical simulations. However, simulations with real data revealed that these algorithms are very sensitive to accuracy of the assumed signal and transmission channel models, and non-modeled signal and channel dynamics frequently caused these algorithms to diverge or to converge

to incorrect solutions. Thus, the multi-channel BSS approaches that rely on specific signal models for modeling manatee vocalizations are not investigated in this dissertation. An excellent recent review of the blind speech separation literature was published by Pedersen et al. [72].

2.5.4 Blind Source Separation in Underwater Acoustics

Despite the extensive literature on speech source separation, only a few papers discuss source separation in the context of underwater acoustics. Gaeta et al. [73] suggested blind separation of artificially mixed underwater acoustic signals using HOS based independence criteria. The cited authors numerically computed the channel impulse responses using ray propagation theory. Bonnifay et al. [74] 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. [75] investigated blind signal 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 acceptable separation. A novel SOS-BSS algorithm, based on the affine projection filter update rule is proposed in this dissertation for enhancing manatee vocalizations in the presence of watercraft emitted noise.

2.6 Summary

With this chapter, potential solutions for manatee avoidance technology are reviewed, and passive acoustic detection of the vocalizations is established as the most promising approach. Manatee vocalizations are generally weak compared to the underwater noise environment. Therefore, the feasibility and success of a passive acoustic vocalization detector is contingent on satisfactory reduction in background noise levels. A concise review of wavelet domain processing and multi-channel source separation methods, and the rationale for investigating these two approaches as an alternative to conventional single and multi-channel adaptive noise reduction method have been presented in this chapter. In Chapter 4 and Chapter 5, a WDD and a BSS algorithm are proposed for enhancing manatee vocalizations in the presence of watercraft emitted noise. Before presenting these proposed approaches, the signal and channel models, which are pivotal in the development and evaluation of these methods, are presented in the next chapter.

CHAPTER 3: ACOUSTIC AND SIGNAL MODELS

The purpose of this chapter is to interpret the physical problem of enhancing manatee vocalizations in the acoustics and signal processing framework, and thereby linking the preceding and subsequent chapters. For this purpose, mathematical models of the underwater acoustic signals and the underwater acoustic channel are developed in the following sections. Artificial noisy vocalization signals generated based on these mathematical models provides a means for conducting realistic and reproducible in-lab experiments for preliminary performance evaluation. An overview of the wet-end is presented in the first section. Mathematical signal and channel models that represent the idealized underwater acoustic signals and the underwater acoustic channel are introduced in the subsequent sections. The last section of this chapter is devoted to introducing the performance measures used to test and evaluate wavelet domain denoising (WDD) and blind source separation (BSS) algorithms that are proposed in the following chapters.

3.1 The Wet-End Model

The physical framework of the underwater acoustic signal processing algorithms presented in this research can be described in two parts; the wet-end and the dry-end. The wet-end represents the domain in which acoustic waves are generated by some point

sources, transmitted through the channel and received at the sensor(s). The dry-end consists of the signal processing hardware and software. In general, the sources and the channel transmission are not known in advance, and the wet-end constitutes the unobservable part of the problem. The generic multi-input multi-output (MIMO) wet-end model considered for this research is depicted in Figure 3-1 where P point sources (s_p, $p = 1, 2, \ldots, P$) are transmitted through a channel and recorded at the Q receivers. Each sensor measurement is corrupted by extraneous noise (v_q, $q = 1, 2, \ldots, Q$) that results in the Q noisy measurements (x_q). Extraneous noise collectively represents local noise (e.g., pressure fluctuations), measurement errors, random errors resulting from modeling the transmission channel, and diffuse ambient noise (e.g., biological activity such as snapping shrimp). Low frequency local noise due to pressure fluctuations can be eliminated with highpass filtering. The remaining extraneous noise is modeled as spatially and temporally independent and identically distributed (IID) signals with some parametric probability density function (PDF). For $Q = 1$, the MIMO wet-end model presented in Figure 3-1 reduces to a single channel setup.

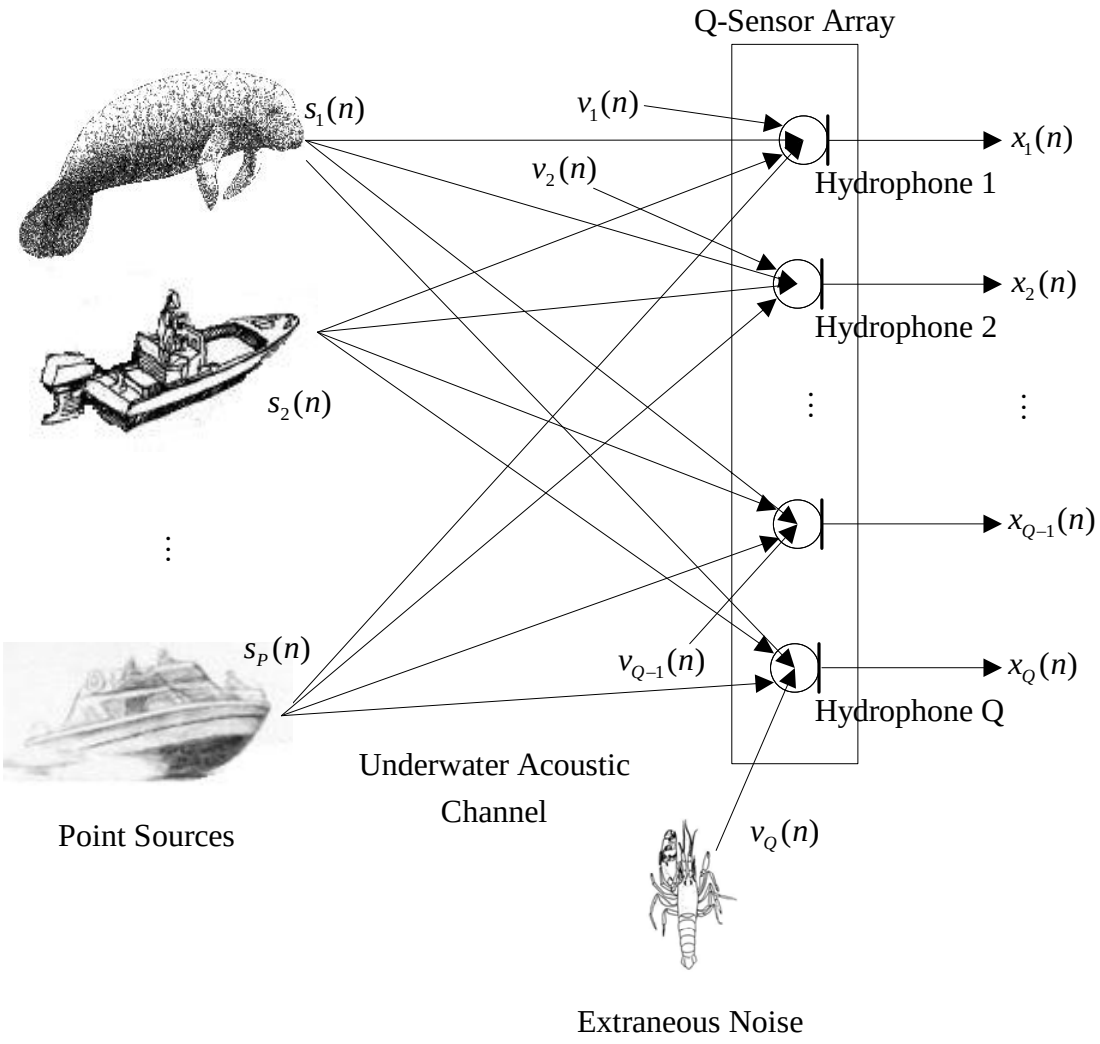


Figure 3-1. The physical model of a MIMO channel representing the wet-end.

The distinction between point sources and diffuse ambient acoustic noise is crucial for the choice of the signal enhancement algorithm. More specifically, BSS algorithms are more effective in the presence of point noise sources such as surface vessels. On the other hand, if the only point source present in the channel is the manatee, a single channel algorithm is more appropriate. The extent of the measurements being

due to point sources can be inferred using the coherence defined as a function of frequency as

$$\gamma^2_{x_p x_q}(\omega) = \frac{\left|P_{x_p x_q}(\omega)\right|^2}{P_{x_p x_p}(\omega) P_{x_q x_q}(\omega)}, \quad p \neq q, \tag{3-1}$$

where $P(\omega)$ is the power spectral density (PSD), and p and q are the measurement channel indices. The coherence satisfies $0 \leq \gamma^2(\omega) \leq 1$, where 0 and 1 indicate no coherence and perfect coherence between the two channel measurements, respectively. Theoretically, perfect coherence is an indication that the measured acoustic signals are due to linear transmission of point sources. In reality, however, the coherence will always be less than 1 due to non-linear and absorptive channel propagation or due to the presence of extraneous noise. Aichner [76] formulated the coherence of point sources and the diffuse noise field, and investigated problems arising in the estimation of the coherence in the context of room acoustics and speech processing. The instantaneous estimate of the coherence can be utilized to determine the optimum algorithm for signal enhancement. For example, Yen and Zhao [77] proposed switching between the symmetric adaptive decorrelator (SAD) algorithm and the adaptive noise canceller (ANC) depending on the coherence estimate.

The underwater acoustic noise environment of marine mammals varies greatly depending on the dominant noise sources present in the channel. These noise sources include a combination of natural and man-made sources. Three geographically distinct underwater noise environments exist; the deep ocean, the shallow water, and the arctic

noise environments. The focus of this research is the tropical, shallow underwater noise environments, which are considered to be more challenging than deep ocean environments. The typical habitat of the Florida manatee is contaminated with numerous noise sources including natural, biological, and anthropogenic noise. Natural noise results from precipitation, surface waves, and water current. Biological noise sources are due to the marine life such as fish and snapping shrimps. The latter is the dominant biological noise source in brackish or saltwater environments, in the absence of anthropogenic noise. The dominant anthropogenic noise source is generally emitted by recreational watercraft navigating through the channels and along the shoreline.

3.2 Marine Mammal Vocalizations

Marine mammals vocalize narrowband tones (e.g., dolphin whistles, right whale contact calls, Beluga whale vocalizations) with time varying instantaneous frequencies and amplitudes; although more complex structures have been observed using high-resolution time-frequency transforms (e.g., [78]). Typical spectrograms representing marine mammal vocalizations are depicted in Figure 3-2. Although directivity in the vocalizations of marine mammals has been reported in several studies [79-80], vocalizing marine mammals are modeled as omni-directional point sources to keep the acoustic models at a manageable complexity. The scope of the research presented in this dissertation is limited to multi-tone harmonic type vocalizations, and in particular, emphasis is given to manatee vocalizations. However, no absolute assumption is identified regarding the vocalization signals such as the number of harmonics, the center

frequency of the fundamental tone, or the contour of the instantaneous frequency of the harmonics, making the results of the dissertation generalizable to a wide range of marine mammal species.

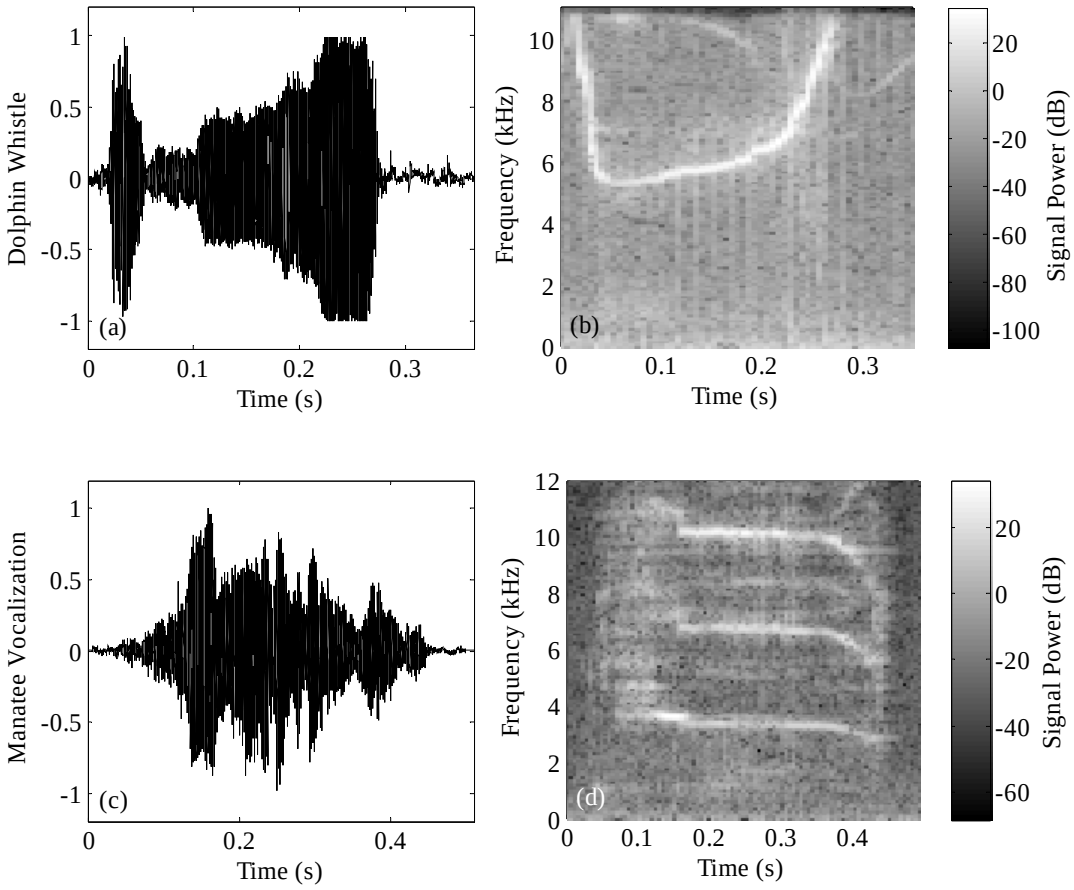


Figure 3-2. Narrowband marine mammal vocalizations. The time domain and spectrogram plots of a (a-b) dolphin whistle and (c-d) a manatee vocalization.

3.2.1 The Florida Manatee

The Florida manatee is a subspecies of the West Indian manatee (*Trichechus manatus*) which belongs to the Trichechidae family of the order Sirenia. A scientific classification of the order Sirenia is presented in Figure 3-3.

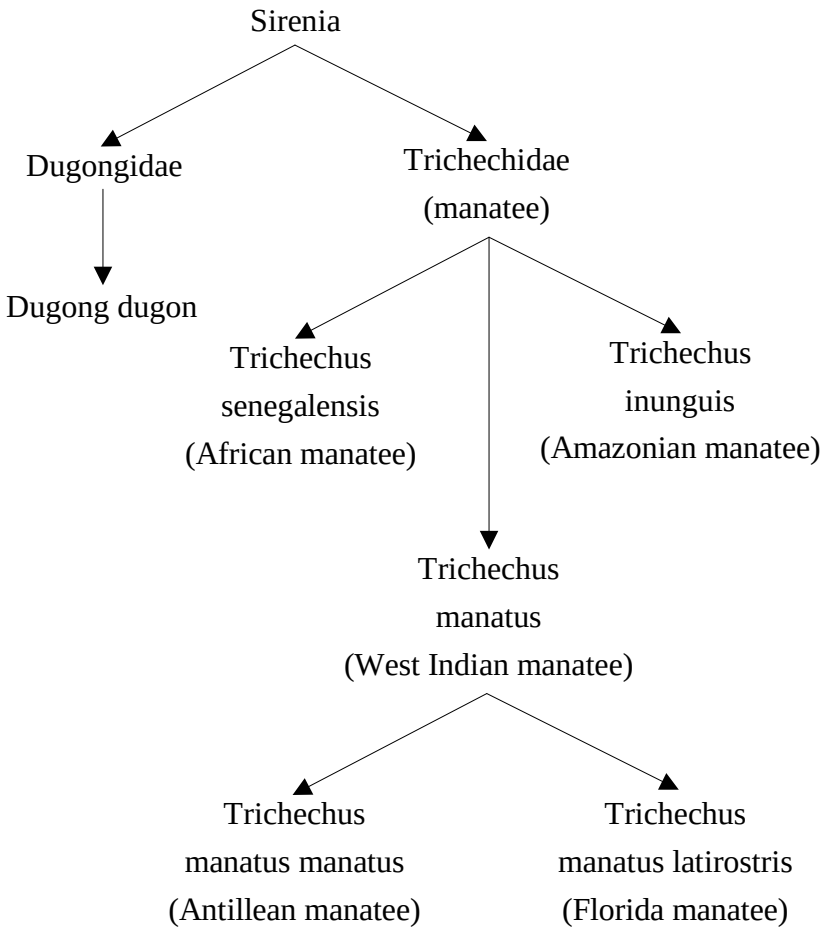


Figure 3-3. The scientific classification of the order Sirenia.

The first spectrogram of manatee vocalizations appeared in a publication by Schevill and Watkins [81]. Evans and Herald [82] reported that both the Amazon and Florida manatee vocalizations have a dominant second harmonic with a higher acoustic

energy content compared to the fundamental harmonic, and that complex time-frequency energy distributions were observed at the beginning and end of the vocalizations. Sonoda and Takemura [83] classified Antillean manatee vocalizations as “very weak, with a narrow main frequency range”. Similar characteristics were observed in dugong vocalizations [84]. Steel [85] perceptually grouped West Indian manatee vocalizations and reported the presence of frequency modulation. Richard-Clark [86] and more recently, Yan et al. [27] classified Florida manatee vocalizations based on the time-frequency properties. A detailed review of the Florida manatee by Hartman [87] related types of vocalizations to different physiological states. Bengston and Fitzgerald [88] reported the average rate of vocalizations during various activities and concluded that the West Indian manatee vocalizes more often when engaged in social activities such as cavorting (4.5 vocalizations per manatee per 5 min) or milling (3 vocalizations per manatee per 5 min). The lowest vocalization rates occurred during feeding (0.3 vocalizations per manatee per 5 min). Similar vocalization rates were also observed by Nowacek et al. [89] for a group of manatees (1.3 vocalizations per manatee per minute) and Phillips et al. [90] for isolated manatees (1.88 and 1.09 vocalizations per manatee per 5 min). More recent research showed that specific vocalization traits of the Amazonian [91] and the Florida manatee [92] are effective features for discerning individual animals. Nowacek et al. [89] measured the mean sound pressure level (SPL) of the dominant harmonic of manatee vocalizations as 103.6 dB from a distance of 20 m. Phillips et al. [90] estimated the mean source level (SL) of manatee vocalizations as 113 dB. The SL was calculated by adjusting the measured vocalization SPL for transmission loss. A

cylindrical spreading model was assumed and the range of the manatees to the four hydrophone array was estimated using differences in time-of-arrivals. Gerstein et al. [93] performed audiogram tests with two captive-born Florida manatees and concluded a U-shaped audiogram with a peak hearing sensitivity of 50 dB ref. 1 μPa between 16 to 18 kHz. The hearing sensitivity was observed to remain high between 6 to 18 kHz. Mann et al. [94] reported the presence of sub-harmonics and other non-linear phenomena in manatee vocalizations. A typical Florida manatee vocalization is a harmonic signal that lasts between 0.1 to 0.5 s, and has a fundamental frequency along with several harmonics (see Figure 3-2 (c-d)). The fundamental harmonic ranges between 2 to 5 kHz, but in general, the second or the third harmonic has the highest acoustic energy and dominates the vocalization. Although decreasing in energy content with increasing frequency, harmonics can be observed up to 18 kHz within the audible frequency range. A sum of multiple sinusoids can form the basis of an idealized, simple manatee vocalization model. The sum of p real sinusoids

$$s(n) = \sum_{k=1}^{p} A_k \sin(\omega_k n + \phi_k), \tag{3-2}$$

can be modeled as the impulse response of a causal autoregressive moving average (ARMA) filter with $2p$ complex conjugate poles, all located on the unit circle

$$s(n) = -\sum_{k=1}^{2p} a_k s(n-k) + \sum_{l=0}^{q} b_l \delta(n-l), \tag{3-3}$$

or equivalently in the z-domain as

$$S(z) = \frac{\sum_{l=0}^{q} b_l z^{-l}}{1 - \sum_{k=1}^{2p} a_k z^{-k}}. \quad (3\text{-}4)$$

The pole phases determine the frequencies of the sinusoids, while the zeros determine the sinusoidal amplitudes and phases. The magnitude of the poles control the bandwidth of the harmonics and slightly shifting the poles towards the center of the unit circle on the z-plane results in narrowband harmonic signals which are more accurate representations of marine mammal vocalizations compared to deterministic sinusoids. Frequency, amplitude, and phase modulation can be introduced by assuming time-varying ARMA coefficients. Speech signals, which have signal energy distributed over a broader frequency band compared to harmonic vocalizations, are also commonly modeled as an ARMA (e.g., [95-96]) or an autoregressive (AR; e.g., [97-98]) process. Similar to speech modeling, the stochastic structure of the marine mammal vocalizations can be captured through exciting the ARMA filter with a periodic pulse or a random signal. A wide selection of methods has been developed for estimating sinusoidal parameters from ARMA coefficients ([99-100], for an exhaustive review of the subject, see [101]). In addition to harmonic signals, manatees have been observed to vocalize non-harmonic signals [85-86], [28], which can also be effectively represented with a broader band ARMA model.

The precise amplitude and phase of the harmonics are less relevant for enhancement and detection tasks addressed with this dissertation. Therefore, to keep the computations manageable, the output of an AR filter excited with a zero-mean random

signal is employed in modeling manatee vocalizations. This AR model of the manatee vocalizations is used to generate artificial, vocalization-like signals with realistic time-frequency characteristics that are used in the following chapters for evaluating the proposed algorithms through Monte-Carlo simulations. The pole magnitudes and order and of the AR filter are selected based on the distinctiveness of the harmonics and the number of harmonics modeled, respectively. All harmonics of the artificial vocalizations are assumed to have equal acoustic energy. The amplitude envelope, the waveform, and the duration of the vocalizations are controlled by passing the output of the AR filter through a time domain window function. Audible similarity to actual vocalizations is not intended with the AR model. An artificially generated vocalization and the corresponding spectrogram are depicted in Figure 3-4 (cf. to Figure 3-2). A similar approach to the AR/ARMA model presented in this section is taken by Hernandez-Perez et al. [102] for maximum likelihood (ML) identification of sperm whale clicks, which is an indication of the flexibility provided by ARMA models in modeling marine mammal vocalizations.

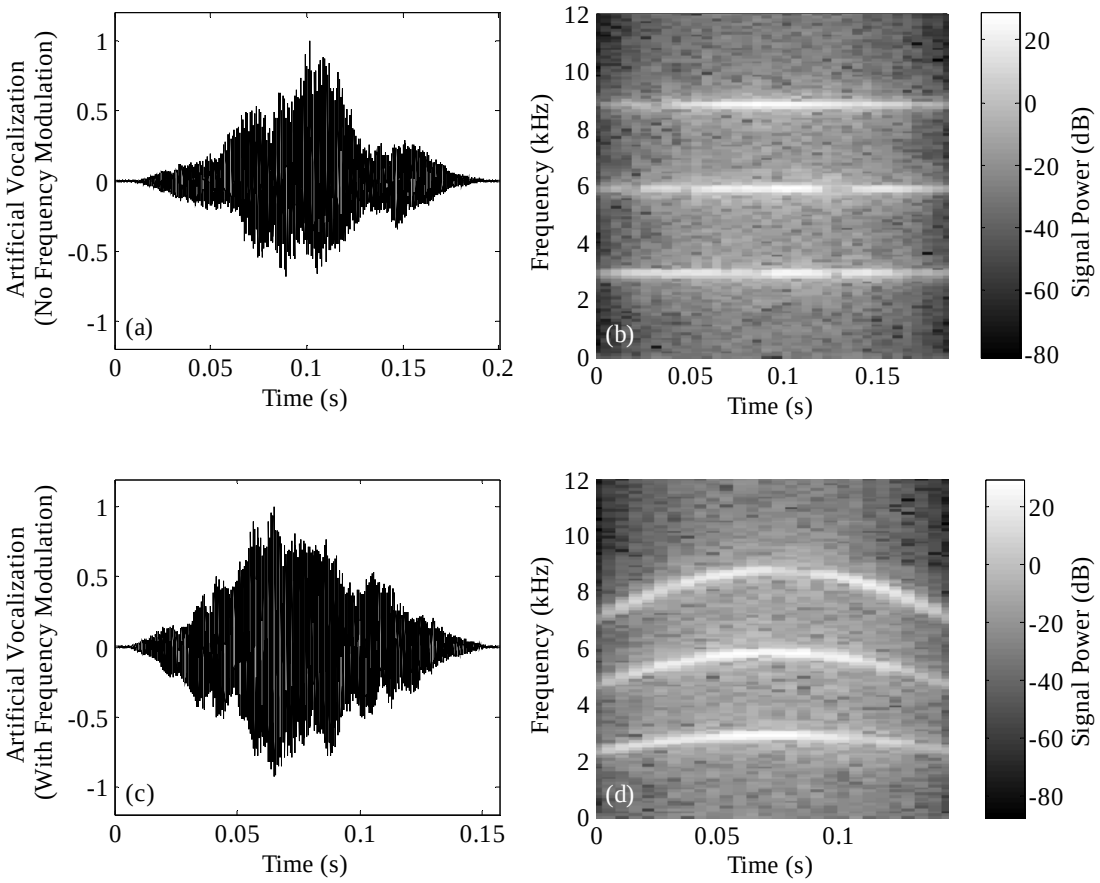


Figure 3-4. Artificial manatee vocalizations. The time domain and spectrogram plots of artificially generated manatee vocalization signals with (a-b) no frequency modulation, and (c-d) concave type frequency modulation.

3.3 Surface Vessels

Noise emitted from surface vessels has been investigated mainly in the context of classification, primarily due to its significance in naval applications. Hinich et al. [103] evaluated the Gaussianity and linearity of ship radiated noise through the bispectrum.

The linearity of a stochastic process implies that it is the output of causal finite impulse response (FIR) filter (possibly of infinite size) driven by IID noise [104]. The Hinich tests performed on a very narrow frequency band centered at 1100 to 1200 Hz revealed the non-Gaussian and non-linear characteristics of ship radiated noise within this narrow frequency band. However, cavitation noise, which is the most dominant acoustic emission from a surface vessel, was omitted in this analysis by selecting a very narrow analysis band. Pentek et al. [105] and Lennartsson et al. [106] suggested the second order delay differential equation to extract features from radiated small boat noise signals for detection and classification purposes. Although a low order nonlinear correlation is included in the model as a discriminating characteristic of boat noise, this model is not intended to accurately model the time domain received acoustic signal.

A complete time series model for noise emitted by surface vessels was presented by Lourens and Du Preez [107]. The cited authors modeled the acoustic background noise signal measured at the receivers in the presence of a transiting ship as a product of two signals; a white Gaussian noise (WGN) component due to randomly collapsing bubbles generated through cavitation on the surfaces of the propeller, and a non-zero, deterministic, and periodic component modulating the amplitude of the IID Gaussian signal. The waveform and frequency of the amplitude modulating signal is determined by the flow pattern behind the vessel and the shaft rotation rate, respectively. Other background noise sources such as precipitation, wind and wave induced noise are modeled as WGN, independent of the cavitation noise. The resulting noise model is

$$v(n) = v_c(n) \cdot u(n) + v_a(n), \tag{3-5}$$

where $v_c(n)$, $u(n)$, and $v_a(n)$ are the cavitation noise, amplitude modulating component, and ambient noise, respectively. The emitted acoustic noise signal $v(t)$ is a superposition of a non-stationary (due to amplitude modulation) WGN signal generated by the propeller and a stationary WGN signal due to other background noise sources.

In 1993, a study that evaluated the acoustic spectrum of a small recreational watercraft was published by the US Navy [108]. The measured spectra were characterized by multiple, well-defined tones below 500 Hz and broadband spectrum above 1.5 to 2 kHz. The signal power was observed to decrease with increasing frequency, which was attributed to absorption of the higher frequencies at the seabed boundary. In some of the recordings, the Lloyd mirror effect was observed to be very dominant, indicated by the rolling peaks and dips in the spectra. More recently, noise emitted by smaller recreational watercraft navigating through the Florida channels was studied by Phillips et al. [24]. The cited authors reported flat spectrum and a mean SPL of 140 dB when frequencies beyond 2 kHz are considered only. The mean SPL increased to 146 dB and could reach up to 170 dB when the entire audible frequency range was considered. The study also revealed that the noise floor dropped below 120 dB beyond 2 kHz.

A number of recordings of watercraft noise made in various locations at Florida between 2003 and 2007 were investigated for modeling recreational boat noise. A typical PSD estimate of watercraft emitted noise recorded in Crystal River, FL is shown in Figure 3-5. Most of these recordings lacked the amplitude modulation suggested by Lourens and Du Preez [107]. The spectrum and autocorrelation estimates presented in

Yan [28] suggest that measurements of noise emitted from recreational watercraft are temporally correlated to some extent. However, in the absence of amplitude modulation, the temporal correlation is introduced solely by transmission through a frequency selective and multi-path channel. Therefore, recreational watercraft are modeled as point sources that emit broadband IID Gaussian acoustic signals. This signal model is used in numerical simulations in the subsequent chapters for evaluating the performance of the proposed methods.

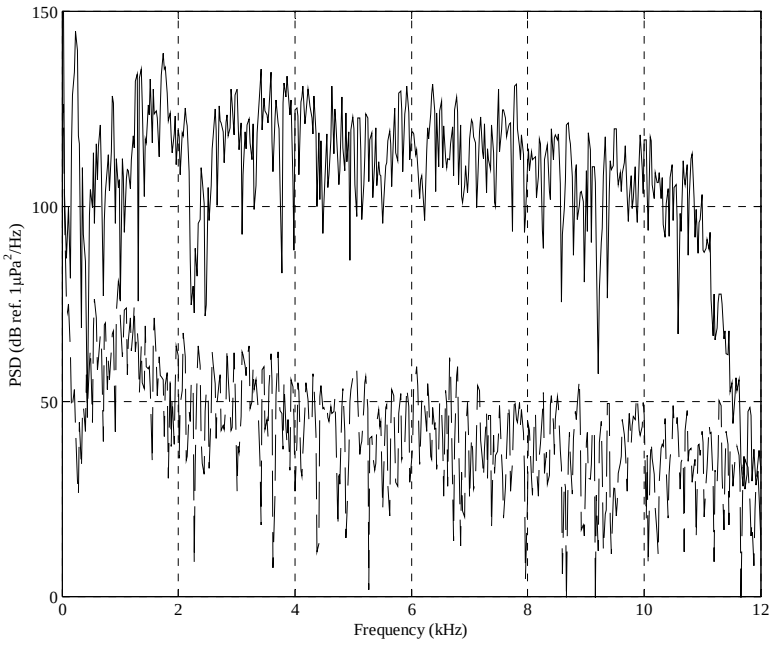


Figure 3-5. Background noise spectra. The noise levels recorded in the presence of a watercraft (solid) and the ambient noise levels without the watercraft (dashed) measured at Crystal River, FL.

3.4 Background Noise

In this dissertation, natural and biological noise sources are considered as ambient background noise sources. Natural noises such as precipitation, surface waves and wind (through breaking surface waves) generate noise through the common mechanism of collapsing air bubbles formed and trapped underwater. These air bubbles create an acoustic ring at a resonant frequency which is dependent on the bubble diameter. Rainfall has been measured to have a relatively flat spectrum up to 10 kHz while wind induced noise decreases logarithmically. The SPL associated with wind induced noise and drizzle (rainfall below 10 mm/hr) was determined to stay below 65 dB. The SPL was measured to remain below 75 dB for heavy downpours (100-200 mm/hr, [109]). Ambient background noise levels at some typical manatee habitats in Florida were reported by Phillips et al. [24]. The SPL varied between 70 to 104 dB, depending on the geographical location. In particular, the SPL in Crystal River, FL dropped from 104 dB at the river mouth to the Gulf of Mexico to 69 dB further inland. The significant difference in SPL is primarily caused by the decline of the snapping shrimp populations inland towards fresh water.

Having distinctly different statistical and time-frequency characteristics compared to the typical deep ocean noise sources has resulted in significant research devoted to snapping shrimp noise for optimizing sonar systems to shallow tropical water environments [110-111]. Snapping shrimp noise is associated with a distinct click sound caused by the closing of the shrimp's claw. Isolated shrimp snaps are short duration (less than 1 ms) and broadband (up to 200 kHz) with a low frequency peak typically less than

20 dB higher relative to the broadband noise floor. The SL are very high reaching up to 190 dB [112]. Although an individual shrimp is a point source, snapping shrimp noise is a characterized by the superposition of numerous spatially spread shrimp snaps at a given instant. Thus, snapping shrimp noise can be idealized as non-local, diffuse noise emitted by infinitely many point sources uniformly distributed throughout the underwater acoustic channel. The acoustic signals measured in environments dominated by snapping shrimp noise are impulsive, as shown in Figure 3-6 (a). The PDF of such recordings are characterized by a narrower and taller mid section, and heavier tails compared to the Gaussian PDF, which is depicted in Figure 3-6 (c) and 3-6 (d), respectively. The non-parametric kernel density estimate shown in Figure 3-6 (c) and (d) was computed using Gaussian kernels of width 0.2 from a set of samples 1000 samples [113].

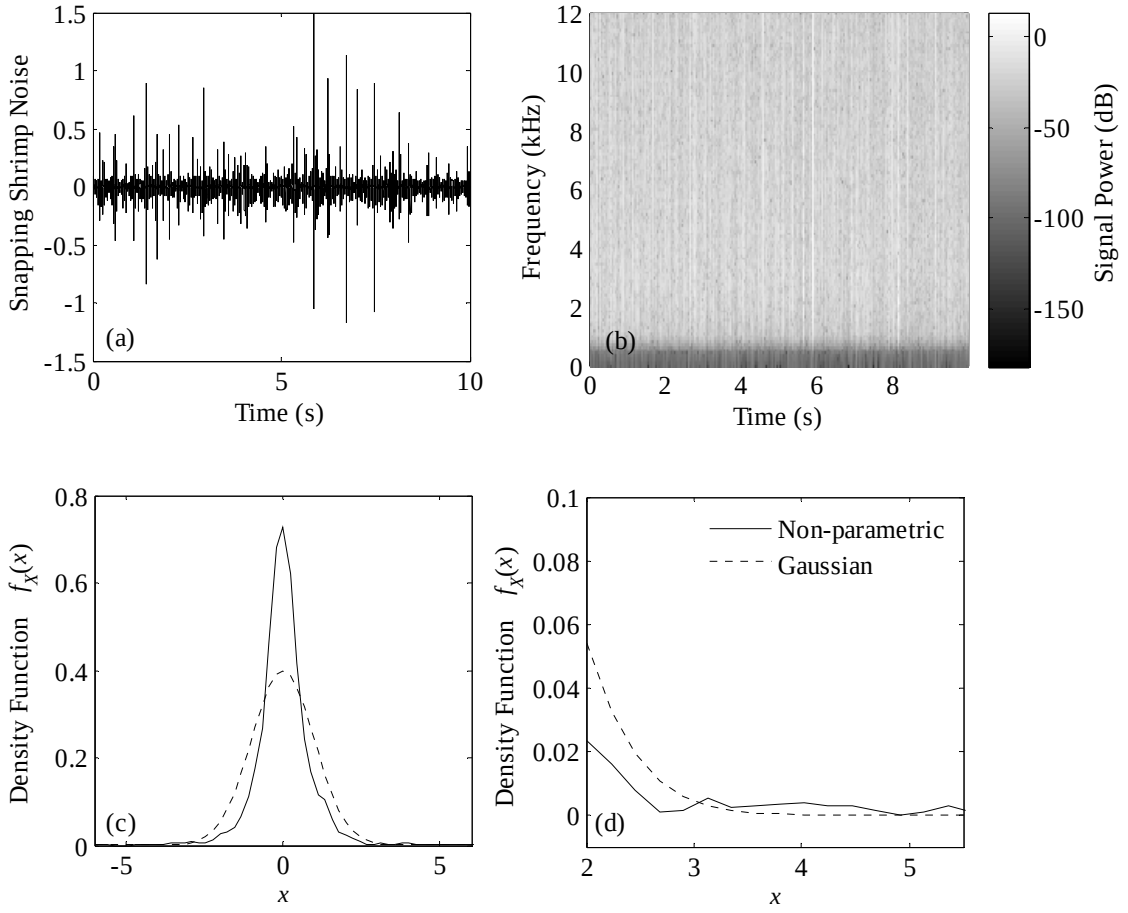


Figure 3-6. Snapping shrimp noise. (a) Time domain plot, (b) the spectrogram, (c) the comparison of a non-parametric PDF estimate of the snapping shrimp noise data to the ML Gaussian PDF estimate, and (d) the tails of the non-parametric and Gaussian PDF estimates.

Powell and Wilson [114] were the first researchers to model the underwater acoustic environment dominated by snapping shrimp noise with the Middleton Class A [115] density functions. The Middleton Class PDF consists of an infinite summation of weighted Gaussian PDF with different variances, and this infinite summation is a major drawback limiting its practical application in signal processing. Bouvet and Schwartz

[116] investigated modeling snapping shrimp dominated underwater acoustic environments using a two term truncated Middleton A Class PDF (i.e., the Gaussian-Gaussian mixture). The cited authors concluded that the full Middleton A Class model was more accurate for modeling snapping shrimp noise compared to the Gaussian-Gaussian mixture. Recently, the symmetric alpha stable (SαS) distribution model has been proposed for modeling impulsive noise [117]. The SαS distributions are exponential distributions defined using two parameters; α and γ are the characteristic exponent and scale parameters, respectively. The characteristic exponent α is a measure of the heaviness of the tails (i.e., impulsiveness of the underlying random variable) and the scale parameter γ is similar to the variance. The SαS distribution is very convenient for modeling both the watercraft emitted noise and snapping shrimp noise dominated background noise environments as it reduces to the Gaussian distribution ($\alpha = 2$) and the heavy tailed Cauchy distribution ($\alpha = 1$) as special cases. The SαS PDF for various characteristic exponent values (scale parameter is kept constant at $\gamma = 0.5$) is depicted in Figure 3-7. Unfortunately, the SαS distribution does not possess a closed form PDF except for these two special cases. Chitre et al. [118] modeled snapping shrimp noise dominated tropical underwater acoustic noise environments using symmetric SαS distributions with α in the range of 1.6 to 1.9, and concluded that the SαS PDF is a more accurate model of the data compared to the Gaussian PDF. The snapping shrimp noise dominated underwater ambient acoustic noise signals are modeled as an SαS random variable in the following chapters.

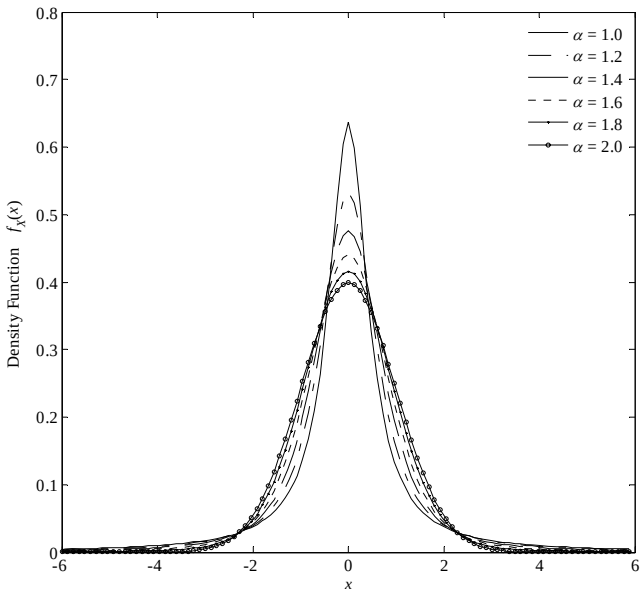


Figure 3-7. The SαS PDF. Shown for different characteristic exponent values from 1.0 to 2.0 (scale parameter is kept constant at $\gamma = 0.5$). Note that for $\alpha = 2$ and $\gamma = 0.5$, the SαS distribution is equal to the zero mean Gaussian PDF with unit variance. For $\alpha = 1$, SαS distribution reduces to the Cauchy distribution.

3.5 The Underwater Acoustic Channel

The transmission properties of the underwater acoustic channel dictate the statistical properties of the measured acoustic signals and define the wet-end mixing process. More specifically, the underwater acoustic channel affects the noise signal statistics, and influences the design of the WDD algorithms proposed in Chapter 4. On the other hand, the channel also defines the de-mixing structure of the multi-channel BSS algorithms proposed in Chapter 5. Thus, an accurate model for underwater acoustic propagation and a thorough understanding of the properties of this model are vital for the

design of effective denoising and signal enhancement algorithms. This section presents a comprehensive analysis of underwater acoustic propagation models which enables the formulation of the underwater channel as a finite impulse response (FIR) filter. Due to the natural habitat of the Florida manatee, emphasis is given to modeling very shallow water acoustic transmission.

The underwater sounds considered in this dissertation are within the scope of linear acoustics. The resulting acoustic pressures are computed from the linear wave equation. Due to the physical waveguides of the underwater acoustic channel, the cylindrical coordinate system is more relevant in shallow water acoustics. A concise formulation of the linear acoustic wave equation in the cylindrical coordinates is presented in Appendix B. The underwater acoustic channel impulse response can be computed analytically or numerically, or can be estimated from experimental data. Numerical models are more accurate in describing the underwater acoustic transmission. However, numerically computed channel FIR filter models are generally of very large order and the resulting channel transmission matrices are difficult to verify in terms of the causality, stability, and invertibility, which are vital for evaluating the performance of the multi-channel BSS algorithms. On the other hand, analytic models are simpler and less accurate, but are very instrumental in developing simple channel transmission models which possess certain desirable mixing properties. Hence, both simple analytic and more complex numerical transmission models are used in this dissertation, and are described next. The discussion of experimentally obtained channel models is provided in Chapter 6.

3.5.1 Simple Spreading Models

A derivation of the cylindrical, spherical, and mixed spreading models from the linearized wave equation are presented in Appendix B. If channel transmission is assumed to obey a simple spreading model, the transmission loss will be independent of the frequency, and will be a function of the range only. For example, the transmission loss associated with the mixed spreading model is given as

$$\mathrm{TL}_{\mathrm{mxd}} = 15\log_{10}(r), \quad r \geq 1, \tag{3-6}$$

in Appendix B. Hence, the simple spreading models are useful in determining a crude estimate of the detection range. However, the resulting underwater acoustic channel model is a single tap FIR filter, and the measured acoustic wave at a receiver is a scaled and delayed version of original source signal. In contrast, the shallow underwater acoustic channel is known to exhibit multi-path propagation and other frequency dependent phenomena such as the Lloyd mirror effect, and the channel is more accurately modeled as a multi-tap FIR filter. The time-harmonic wave assumption (see Appendix B) can be used to analytically obtain simple, multi-path shallow water channel impulse responses. First, the total number of acoustic paths is determined, and the simple spreading model is applied to each path to determine the contribution of the path to the aggregate pressure field at a given point. As an example, if only three paths are to be included in the pressure field computations (as shown in Figure 3-8) and spherical spreading is assumed, the amplitude of the pressure field in the frequency domain becomes

$$P(\mathbf{r},\omega) = \frac{\exp(jkr_1)}{r_1} - \frac{\exp(jkr_2)}{r_2} + \frac{R_{12}\exp(jkr_3)}{r_3}, \tag{3-7}$$

where $k = \omega / c$ is the wave number and c is the speed of sound, and

$$R_{12} = \frac{Z_2\cos(\theta_i) - Z_1\cos(\theta_t)}{Z_2\cos(\theta_i) + Z_1\cos(\theta_t)}, \tag{3-8}$$

is the reflection coefficient for the sea-seabed boundary. The acoustic impedances are computed as $Z = \rho c$ where ρ is the density of the medium. The negative sign of the second term on the right-hand-side of Eq. (3-7) is due to the phase shift caused by the sea-air boundary which is modeled as an ideal pressure release surface.

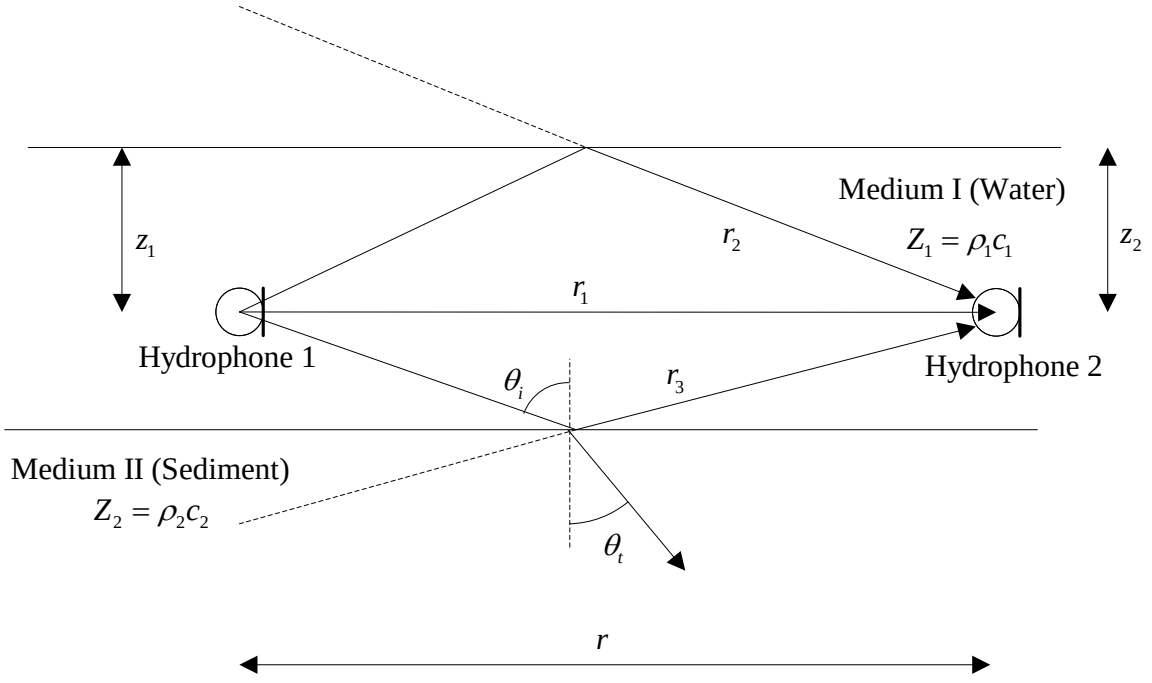


Figure 3-8. Transmission paths associated with a three-path simple spread model.

Once the pressure field is computed, the impulse response can be obtained from the inverse Fourier transform. The order of the mixing filters can be controlled by changing

the number of arrival paths. The transmission loss estimated using the above procedure and the corresponding channel impulse response at a range of 20 m computed assuming three arrivals is presented in Figure 3-9 (a) and (c), respectively. The analytic method for computing the multi-path FIR filter channel model outlined above is used exclusively in Chapter 5 to model the pseudo impulse responses of the cross-coupling channels. Furthermore, based on the transmission loss experiments conducted in typical manatee habitats and presented by Phillips et al. [24], the mixed transmission loss model is used to estimate the improved detection ranges in Chapter 7.

3.5.2 Numerical Models

The wave equation can also be solved numerically using three principle approaches; the normal mode approach, the parabolic approximation to the wave equation, and the ray propagation theory. All three approaches numerically solve the Helmholtz equation for the complex valued acoustic particle velocity potential $\varphi(\mathbf{r},\omega)$, as described in Appendix B. Once the velocity potential (and the acoustic pressure field) is obtained for each frequency in the desired Nyquist band, the channel impulse response can be computed from the inverse Fourier transform. The shallow water medium is considered to be bounded by the two parallel seabed-water and water-air boundaries. Acoustic propagation between these two boundaries can be viewed as the radial propagation of standing waves in the vertical direction between the two boundaries. The normal mode approach solves the Helmholtz equation for the individual vertical standing waves (or modes) at a given location $\mathbf{r} = (r, \psi, z)$. The superposition of the modes at a

certain range yields the acoustic field at that range. The total number of modes that exist is a function of the speed of sound, the depth of the water column, and the frequency of the acoustic source. The normal mode approach is suitable estimating the propagation of low frequencies (< 1 kHz) in shallow water. However, typical manatee habitats are only several meters in depth, and the normal mode approach fails to predict any modes (and propagation) at those depths.

The Helmholtz equation is an elliptic partial differential equation (PDE), which is difficult to solve numerically. Based on certain assumptions and simplifications, the elliptic Helmholtz equation can be transformed to a parabolic PDE, which is numerically easier to solve using Fourier transform methods. This approximation is more accurate for high frequencies, and the parabolic approximation methods generally produce erroneous results at low frequencies. However, to estimate the impulse response from the inverse Fourier transform of the acoustic particle velocity field, the velocity potential for all frequencies in the Nyquist band is necessary. Thus, accurate channel impulse responses for the shallow water manatee habitat cannot be estimated using parabolic approximation techniques.

The ray propagation theory is used in this dissertation to generate accurate models of the underwater acoustic transmission channel. The acoustic particle velocity potential $\varphi(\mathbf{r},\omega)$ is complex valued and can be expressed as a product of a magnitude and phase. Substituting the velocity potential magnitude-phase pair into the Helmholtz equation results in two separate equations. Based on some approximations, the solution to these two equations provides the paths of propagation and the corresponding acoustic energy

propagating. The superposition of the various acoustic paths at a given coordinate results in the acoustic particle velocity field at that point. The ray propagation theory is generally more accurate for high frequency propagation, but also yields acceptable results for lower frequencies and shallow water propagation. A complete theoretical development of the different solutions to the wave equation can be found in [119-120].

The BELLHOP software[a] (for a theoretical treatment see [121]) is used in this dissertation to numerically model the underwater acoustic channel using the ray propagation theory. The numerically computed channel impulse responses are used for the Monte-Carlo simulations in Chapter 4 and Chapter 5. The magnitude of the pressure field estimated in a typical manatee habitat is presented in Figure 3-9 (b). The corresponding impulse response is plotted in Figure 3-9 (d).

[a] The BELLHOP software is a part of the Acoustics Toolbox software package written by M. Porter and can be accessed through the URL http://oalib.hlsresearch.com/ (last accessed: 12 Nov. 2008).

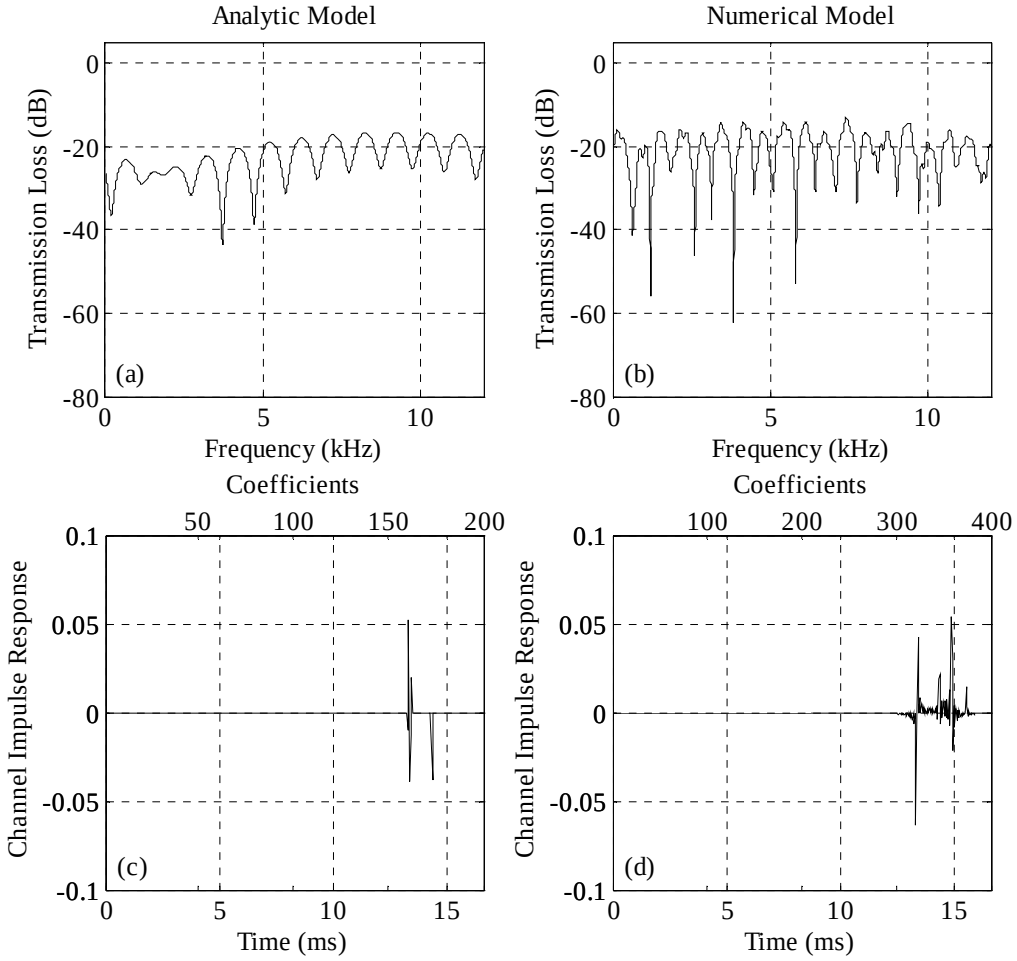


Figure 3-9. Analytic and numerical channel transmission models. (a)-(b) The magnitude of the analytically computed and BELLHOP generated (both at a range of 20 m) complex pressure fields, respectively, and (c)-(d) the corresponding truncated impulse responses.

3.6 Performance Measures

Well defined, quantitative performance measures are necessary for objectively evaluating the denoising and signal enhancement performances of the algorithms

proposed in following chapters. In this section, two performance measures are introduced. Each one of these performance measures is an indicator of a specific attribute of the enhanced signal, which is a prerequisite for effectively performing a certain marine mammal monitoring task. For tests conducted with real vocalizations, these quantitative performance measures are supported with qualitative performance indicators such as the audible quality and clarity of the vocalizations.

3.6.1 Signal to Noise Ratio

The pre-denoising quality of a noisy vocalization signal is quantified in terms of the input signal to noise ratio (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

$$\mathrm{SNR}_{\mathrm{in},q} = 10 \cdot \log_{10}\left(\left(s_{\mathrm{rms},q}\right)^2 \Big/ \left(v_{\mathrm{rms},q}\right)^2 \right), \quad q = 1, 2, \ldots, Q, \tag{3-9}$$

where q is the channel index, $s_q(n)$ and $v_q(n)$ are the vocalization and noise signals at channel q, respectively. The RMS value of the length N signal $x(n)$ is defined as

$$x_{\mathrm{rms}} = \left[\frac{1}{N} \sum_{n=0}^{N-1} \left(x(n) \right)^2 \right]^{1/2}. \tag{3-10}$$

The noise signal in the denominator of Eq. (3-9) 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 case of BSS algorithms, watercraft emitted noise power

must be considered separately from extraneous noise power. For the Monte-Carlo simulations conducted in Section 5.5, the extraneous noise power is defined relative to watercraft emitted noise power whereas the overall input SNR is defined as a ratio of the vocalization signal power and total (both watercraft emitted and extraneous) noise power.

In contrast, a single output SNR is defined as

$$\mathrm{SNR}_{\mathrm{out}} = 10 \cdot \log_{10}\left[\left(\left(y_{\mathrm{rms},s}\right)^2 - \left(y_{\mathrm{rms},v}\right)^2\right) \Big/ \left(y_{\mathrm{rms},v}\right)^2\right], \tag{3-11}$$

where $y_{\mathrm{rms},s}$ is the RMS value of the denoised or enhanced estimate of the vocalization signal and $y_{\mathrm{rms},v}$ is the RMS value of the output if no vocalization was present. Thus, $y_{\mathrm{rms},v}$ represents the noise residue that is not suppressed by the tested algorithm over the duration of the vocalization. The RMS values used for the computation of the SNR are illustrated in Figure 3-10.

The computation methodology of the SNR performance measure implicitly assumes that the denoising or signal enhancement algorithm is linear. More specifically, it is assumed that the denoised output, when only the vocalization signal is present (not shown in Figure 3-10), can be combined with the denoised output when only the noise signal is present (see Figure 3-10 (e)) to yield the denoised vocalization signal (see Figure 3-10 (d)). While this assumption may be justified for linear filtering methods such as the feedback adaptive line enhancer (FALE) algorithm, it is not valid for WDD which involves a non-linear threshold operator. The bias introduced into the SNR performance measure is alleviated to some extent by the use of the soft threshold rule in the WDD methods and does not appear to significantly affect the performance measures. Thus, the

output SNR as defined in Eq. (3-11) is determined to be satisfactory in evaluating the performance of the non-linear WDD algorithms. The output SNR is an indicator of how well noise is suppressed in the presence of a vocalization signal and is relevant for vocalization detectors based on signal energy.

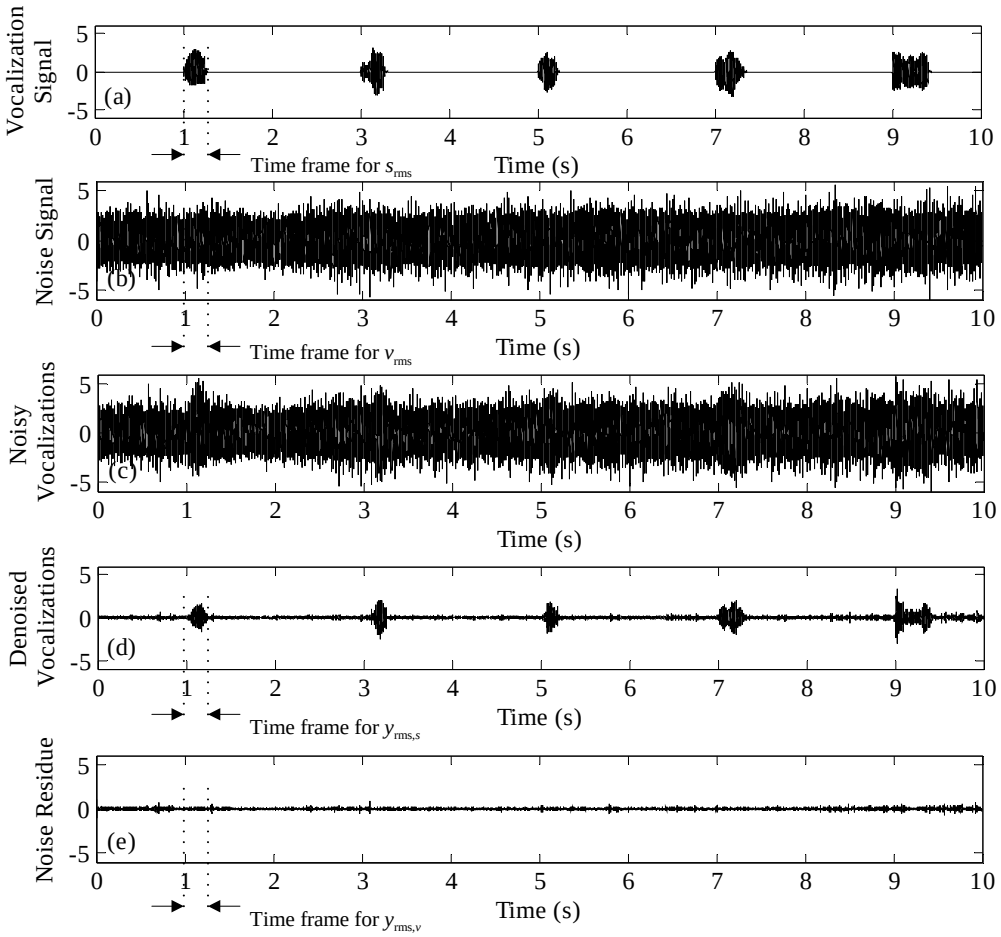


Figure 3-10. Different RMS values used to calculate the performance measures. (a) The RMS value of the vocalization, (b) the RMS value of noise, (c) the noisy vocalization signal, (d) the RMS value of the denoised vocalization, and (e) the RMS value of the noise residue.

3.6.2 Signal to Distortion Ratio

The signal to distortion ratio (SDR) is defined for time domain signals as the ratio of the vocalization signal and distortion power

$$\mathrm{SDR} = 10 \cdot \log_{10}\left(\left(s_{\mathrm{rms}}\right)^2 \Big/ \left(e_{\mathrm{rms}}\right)^2 \right), \tag{3-12}$$

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

$$e_{\mathrm{rms}} = \left[\frac{1}{N} \sum_{n=0}^{N-1} \left(e(n)\right)^2 \right]^{1/2} = \left[\frac{1}{N} \sum_{n=0}^{N-1} \left(s(n) - y_s(n)\right)^2 \right]^{1/2}, \tag{3-13}$$

and $y_s(n)$ is the denoised or enhanced estimate of the vocalization signal. By definition, the RMS value of the distortion is computed over the duration of the vocalization signal. The permutation ambiguity inherent to BSS algorithms (which is discussed in Section 2.5) may prevent precisely knowing which channel the enhanced vocalization signal will appear. This permutation ambiguity is manually accounted for by selecting the channel at which the estimate of the vocalization appears, to ensure accurate SDR computations. The scaling ambiguity associated with the BSS algorithms result in an overestimated MSE. This overestimation is corrected for by normalizing the vocalization and output signals to unit energy. The SDR is an indicator of how well the vocalization waveform is preserved, which strongly affects the detection rate of the matched filter detector or other similar correlation based detectors, and the accuracy of source localization algorithms based on time difference of arrival (TDOA). The SDR defined in Eq. (3-12) is the

inverse of the distortion performance measure frequently used in speech processing to measure the quality of the enhanced speech [122].

3.7 Summary

With this chapter, watercraft emitted noise and snapping shrimp noise are established as dominant noise sources in the manatee habitat. The pertinent characteristics of manatee vocalizations and background noise, as well as the underwater acoustic propagation are introduced. Based on these characteristics, mathematical models for the vocalizations, noise signals, and the acoustic transmission channel are developed. More specifically, the AR signal model is shown to produce accurate artificial signals that possess statistical and time-frequency properties similar to manatee vocalizations. Watercraft emitted noise is modeled as an IID Gaussian noise sequence. The heavy tailed SαS distribution is proposed to model snapping shrimp dominated noise environments. The SαS signal model is shown to possess Gaussian noise as a special case and is established as the unifying statistical framework for modeling both watercraft emitted noise and snapping shrimp noise dominated acoustic environments. The underwater acoustic channel is modeled as a FIR filter with multiple non-zero coefficients representing the multi-path arrivals. Both analytic methods based on simple spreading models and more complex numerical methods are presented for estimating the FIR filter coefficients. The numerical models are more accurate, while simple analytic models are easier to evaluate and verify. The SNR and SDR performance measures are

suggested for measuring the noise suppression and waveform preservation properties of the proposed algorithms.

These signal and channel models presented in this chapter are used to develop a thorough understanding of the problem of enhancing and detecting manatee vocalizations, and serve to generate Monte-Carlo tests for evaluating the performances of the proposed algorithms in next two chapters. In particular, numerical channel models are used to simulate the path channel impulse responses for the channel WDD algorithms proposed in Chapter 4. However, the numerical models are too complex to evaluate and verify the mixing properties of the simulated channels. Therefore, the simpler analytic models are used to model the pseudo channel impulse responses for the BSS algorithms proposed in Chapter 5. The mixed spreading model is also used to evaluate the improved detection ranges in Chapter 7.

CHAPTER 4: WAVELET DOMAIN DENOISING

Wavelet domain denoising (WDD) is presented in this chapter as a non-linear noise reduction algorithm. The conventional WDD approach is reviewed in the first section of the chapter. In the following section, the optimal denoising strategy for manatee vocalizations is analyzed. Based on the conclusions of this analysis, an unsupervised, adaptive WDD algorithm based on the wavelet packet transform (WPT) is presented next, followed by the introduction of a similar non-adaptive Fourier domain denoising algorithm. In the final section of the chapter, the proposed methods are evaluated by Monte-Carlo simulations.

4.1 Wavelet Domain Denoising

The wavelet basis functions have been shown to be very effective in compactly representing non-stationary signals that possess varying time-frequency energy distributions with a few large amplitude coefficients. Noise, on the other hand, is mapped to the wavelet domain as numerous small amplitude coefficients. Thus, eliminating the small amplitude noise coefficients by thresholding forms the basis of WDD algorithms. Like most marine mammal vocalizations, manatee vocalizations are non-stationary and may exhibit a varying and non-linear time-frequency structure. In

Section 2.4, such properties of the vocalization signals are presented as the rationale for investigating WDD for denoising manatee vocalizations. In this section, a theoretical development of WDD is presented to provide a foundation for the adaptive WDD method proposed later in the chapter.

To explicitly formulate WDD, consider the problem of estimating the vocalization signal from a block of N noisy measurement samples

$$\mathbf{x} = \mathbf{s} + \mathbf{v}, \tag{4-1}$$

where $\mathbf{x} = \begin{bmatrix} x(0) & x(1) & \cdots & x(N-1) \end{bmatrix}^T$ 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. Assume that the noise signal is independent and identically distributed (IID) and Gaussian with a distribution $v \sim \mathcal{N}(0, \sigma)$. Wavelet denoising can be visualized as a three-step, non-linear block filtering process (see Figure 4-1) which can conveniently be expressed using the matrix notation as a forward discrete wavelet transform (DWT) transform represented by the $N \times N$ orthogonal DWT matrix $\mathcal{W}$, a diagonal threshold matrix $\mathbf{\Lambda}$, and the inverse transform represented by the orthogonal matrix $\mathcal{W}^{-1} = \mathcal{W}^H$

$$\hat{\mathbf{s}}_{\text{DWT}} = \left(\mathcal{W}^H \mathbf{\Lambda} \mathcal{W} \right) \mathbf{x}. \tag{4-2}$$

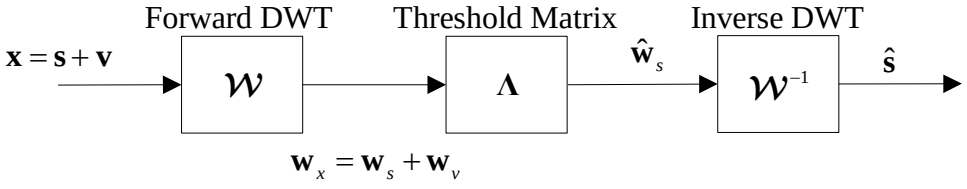


Figure 4-1. The matrix representation of wavelet domain denoising filters.

The DWT will map the signals to the wavelet domain as

$$\mathbf{w}_x = \mathbf{w}_s + \mathbf{w}_v, \tag{4-3}$$

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 $\mathbf{w}_s$ 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 $\mathbf{\Lambda} = \text{diag}\left[\delta(0) \quad \delta(1) \quad \cdots \quad \delta(N-1)\right]$ whose entries are non-linear threshold operators. A generic unsupervised criterion for the non-linear filtering process can be expressed in terms of the penalized least squares [123] defined as

$$\mathcal{J}_{\text{PLS}} = \frac{1}{2}\left(\left\|\mathbf{w}_x - \hat{\mathbf{w}}_s\right\|_2\right)^2 + p(\boldsymbol{\eta}, \hat{\mathbf{w}}_s), \tag{4-4}$$

where $p(\cdot)$ is the penalty function which is necessary to prevent the trivial solution of $\hat{\mathbf{w}}_s = \mathbf{w}_x$ (i.e., no denoising), $\boldsymbol{\eta}$ is the regularization or threshold parameter vector, and $\|.\|_2$ represents the ℓ_2-norm. Conventional wavelet denoising is a block processing approach and a single threshold parameter η is used for the entire set of measurements. The orthogonal DWT results in approximately IID wavelet coefficients, which justifies the individual processing of each coefficient. Thus, the cost function defined in Eq. (4-4) is reduced to

$$\mathcal{J} = \frac{1}{2}\sum_{n=0}^{N-1}\left(w_x(n) - \hat{w}_s(n)\right)^2 + \eta\sum_{n=0}^{N-1} p\left(\hat{w}_s(n)\right). \tag{4-5}$$

In general, the penalty function is only imposed on some of the wavelet coefficients and in particular, the coefficients associated with the lowest frequency subband are not penalized. Different penalty functions were proposed in the literature which led to the development of some of the well-known threshold rules. For example, penalizing the ℓ_1-norm of the coefficients $p\left(\hat{w}_s(n)\right) = \left|\hat{w}_s(n)\right|$ results in the soft threshold rule [124] which eliminates the coefficients below a certain threshold η and shrinks other coefficients according to

$$\delta^s(w,\eta) = \begin{cases} 1 - \eta/|w| & |w| > \eta \\ 0 & \text{otherwise,} \end{cases} \tag{4-6}$$

where sgn(x) is the sign function. The penalty function [123]

$$p\left(\hat{w}_s(n)\right) = \eta^2 - \left(\left|\hat{w}_s(n)\right| - \eta\right)^2 \cdot \mathcal{I}_{\{|w|<\eta\}}\left(\hat{w}_s(n)\right), \tag{4-7}$$

where $\mathcal{I}$ is the indicator function such that

$$\mathcal{I}_{\{|w|<\eta\}}\left(w(n)\right) = \begin{cases} 1 & |w(n)| < \eta \\ 0 & \text{otherwise,} \end{cases} \tag{4-8}$$

results in the hard threshold rule which eliminates all coefficients with amplitudes less than η

$$\delta^h(w,\eta) = \begin{cases} 1 & |w| > \eta \\ 0 & \text{otherwise.} \end{cases} \tag{4-9}$$

Donoho and Johnstone [124] proposed combining the soft or hard threshold rules with universal threshold defined as

$$\eta_u = \sigma_v \left(2\log(N)\right)^{1/2}, \tag{4-10}$$

where the noise level σ_v is estimated from the median absolute deviation estimate of the high frequency wavelet coefficients $\mathbf{w}_{x,1}$ of the first level of decomposition

$$\sigma_v \approx \hat{\sigma}_{\mathbf{w}_{x,1}} = \frac{\text{med}\left(\left|\mathbf{w}_{x,1}\right|\right)}{0.6745}. \tag{4-11}$$

The choice of the universal threshold was motivated by the asymptotic properties of normally distributed random variables. In particular, for N realizations of a normally distributed zero mean random variable $x \sim \mathcal{N}(0,1)$, the maximum absolute value within the sequence becomes bounded with a high probability

$$\lim_{N\to\infty}\left(\Pr\left\{\max(\mathbf{x}) \le \left(2\log(N)\right)^{1/2}\right\}\right) = 1, \tag{4-12}$$

where Pr{·} denotes probability. For temporally correlated noise, Johnstone and Silverman [125] showed that the variance of the noise coefficients will depend of the level of decomposition and the authors proposed level dependent universal thresholds of the form

$$\eta_{u,j} = \hat{\sigma}_{w_{x,j}} \left(2\log(N)\right)^{1/2}. \tag{4-13}$$

Alternatively, Donoho and Johnstone [126] proposed minimizing Stein's unbiased risk estimate (SURE)

$$\mathscr{J}_{\text{SURE}} = E\left(\left\|\hat{\mathbf{w}}_s - \mathbf{w}_s\right\|_2^2\right) = N + E\left(\left\|\mathbf{g}(\mathbf{w}_x)\right\|_2^2 + 2\nabla_{w_x} \cdot \mathbf{g}(\mathbf{w}_x)\right), \tag{4-14}$$

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

$$\nabla_{w_x} \cdot \mathbf{g}(\mathbf{w}_x) = \sum_{n=0}^{N-1} \left.\frac{\partial g(w)}{\partial w}\right|_{w=w_x(n)}. \tag{4-15}$$

Both the soft and hard threshold rules can be recast into the estimator structure of $\hat{\mathbf{w}}_s = \mathbf{w}_x + \mathbf{g}(\mathbf{w}_x)$. The scale dependent threshold value that minimizes the mean squared error (MSE) defined in Eq. (4-14) is given as

$$\eta_{j,\text{SURE}} = \arg\min_{\eta}\left[N - 2 \cdot \#\left\{n : \left|w_{x,j}(n)\right| \le \eta\right\} + \sum_{n=0}^{N-1} \min\left(\left|w_{x,j}(n)\right|, \eta\right)^2\right], \tag{4-16}$$

where $\#\{n : \cdot\}$ denotes the number of arguments satisfying the relation given in the braces.

In this section, WDD is presented as a three stage noise reduction method. A concise theoretical development of WDD is presented. In the following section, several aspects of the WDD strategy are reviewed in the context of denoising manatee

vocalizations. Based on these observations, an adaptive WDD algorithm that uses the SURE cost function given in Eqs. (4-14) and (4-15) is presented in Section 4.3.

4.2 Denoising Manatee Vocalizations

A wavelet based denoising strategy is fully described by the type of transform (i.e., non-redundant, redundant, dyadic, packet), the basis function, the level of decomposition, the thresholding rule, and the threshold. The optimum (e.g., in the minimum MSE sense) choice for these elements depends on the time, frequency, smoothness and statistical characteristics of the vocalization and noise signals. The downsampling of the outputs at each decomposition level (see Section 2.4) results in a non-redundant transform. Algorithms based on a redundant wavelet transform are reported to result in a better denoising performance compared to otherwise equivalent non-redundant algorithms [127-128]. This improvement is a consequence of the redundant denoising algorithms using more samples compared to the non-redundant algorithm to estimate the noise-free signal. However, certain desirable properties of the non-redundant algorithms are not valid for redundant algorithms. In particular, the wavelet coefficients resulting from a redundant transform of IID Gaussian noise are no longer independent, but are rather correlated [129]. Thus, thresholds and other criteria based on the IID Gaussian noise assumption are no longer applicable.

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. The dyadic (or octave band) partitioning, as in the case of the DWT, is more

suitable than a regular grid partition resulting from the short-time Fourier transform (STFT) for non-stationary signals that exhibit both low and high frequency transients within the analysis window. In particular, the high frequency transients are relatively short duration compared to low frequency transients and the dyadic partitioning algorithms sacrifice some frequency resolution for improved time localization of these high frequency transients and vice versa. An irregular grid can also be obtained by packet transformations (such as the WPT decomposition) through a best basis decomposition based on optimizing energy or entropy cost functions [130].

It can be shown that the wavelet coefficients of a signal at a given scale and translation are the correlation coefficients between the signal and the corresponding wavelet basis functions. Thus, wavelet basis functions that match the noise-free signal or some of its components result in large amplitude wavelet coefficients. Therefore, the selection of the wavelet basis functions should be influenced by signal waveform and statistical properties. Motivated by these conclusions, several researchers proposed designing wavelets that match the signal waveform [131] or the signal statistics [132]. On the other hand, the optimal threshold rule and the threshold should also be selected based on the prior knowledge regarding the time-frequency and statistical properties of the signals and the type of transform implemented. Conventional thresholds implemented in WDD (e.g., see Eqs. (4-10) and (4-16)) are constant and do not adapt to changing signal and noise statistics.

Manatee vocalizations are harmonic signals with multiple narrowband components such that much of the signal energy is localized about the harmonic

frequencies. This structure suggests that a selective or irregular decomposition of the time-frequency plane with high frequency resolution (and lower time resolution) around the harmonic frequencies would provide a more compact representation of the vocalization signals. However, the octave band partitioning of the frequency axis resulting from the dyadic DWT generally does not yield such a desirable decomposition and a compact representation. Rather, an irregular decomposition of the time-frequency plane using a packet transform (such as the WPT) supported with a best basis selection criterion is more suitable for analyzing manatee vocalizations. Unfortunately, the number of harmonics and the harmonic frequencies may change from one vocalization to the next. Thus, it is not possible to define a single packet decomposition that is optimal for all types of vocalizations. Furthermore, since vocalizations are assumed to be corrupted with significant noise, a best basis decomposition estimated from noisy measurements may not result in the optimum partitioning in the MSE sense. Such a suboptimal decomposition derived from noisy vocalizations may reduce the maximum achievable denoising performance [133]. Therefore, a complete decomposition at a predefined scale that results in a regular time-frequency grid is proposed for analyzing manatee vocalizations.

The efficiency of the WPT decomposition over the DWT can be quantitatively measured through the compression efficiency expressed in terms of the normalized partial energy sequence (NPES) [134].

$$\gamma_c = \frac{\sum_{n=0}^{n_1-1} \theta_n^2}{\sum_{n=0}^{N-1} \theta_n^2}, \tag{4-17}$$

where θ_n are the sorted (according to energy) N transform domain coefficients and n_1 is some pre-determined cut-off value. In Figure 4-2, the NPES averaged over all calls within a specific category of the vocalization library [27] is depicted for the DFT, the DWT, and the WPT implemented using the Daubechies-8 wavelet basis functions. The depth of the decomposition for both the DWT and WPT are arbitrarily set to 7 levels. The full binary expansion is constructed for the WPT (i.e., no best basis selection is performed). The sampling rate is 24 kHz and the analysis window is set to 2048 samples. The DFT is the most effective in representing the vocalization signals. For most vocalization categories, the WPT achieves a similar compression performance to the DFT. The difference in performance between the DFT and the WPT is most significant for type 1200 vocalization presented in Figure 4-2 (h). This result is expected since the 1200 type vocalizations are characterized by very narrowband harmonics with relatively no frequency modulation, which are optimally represented by the DFT. The lower compression rates obtained by the DWT compared to the WPT for all categories are in agreement with the above discussion regarding the sub-optimality of octave band partitioning. Rather than a fixing the number of coefficients used to compute the compression efficiency as done in Eq. (4-17), the percentage of coefficients required to compress 90% of the total signal energy is computed and presented in Table 4-1.

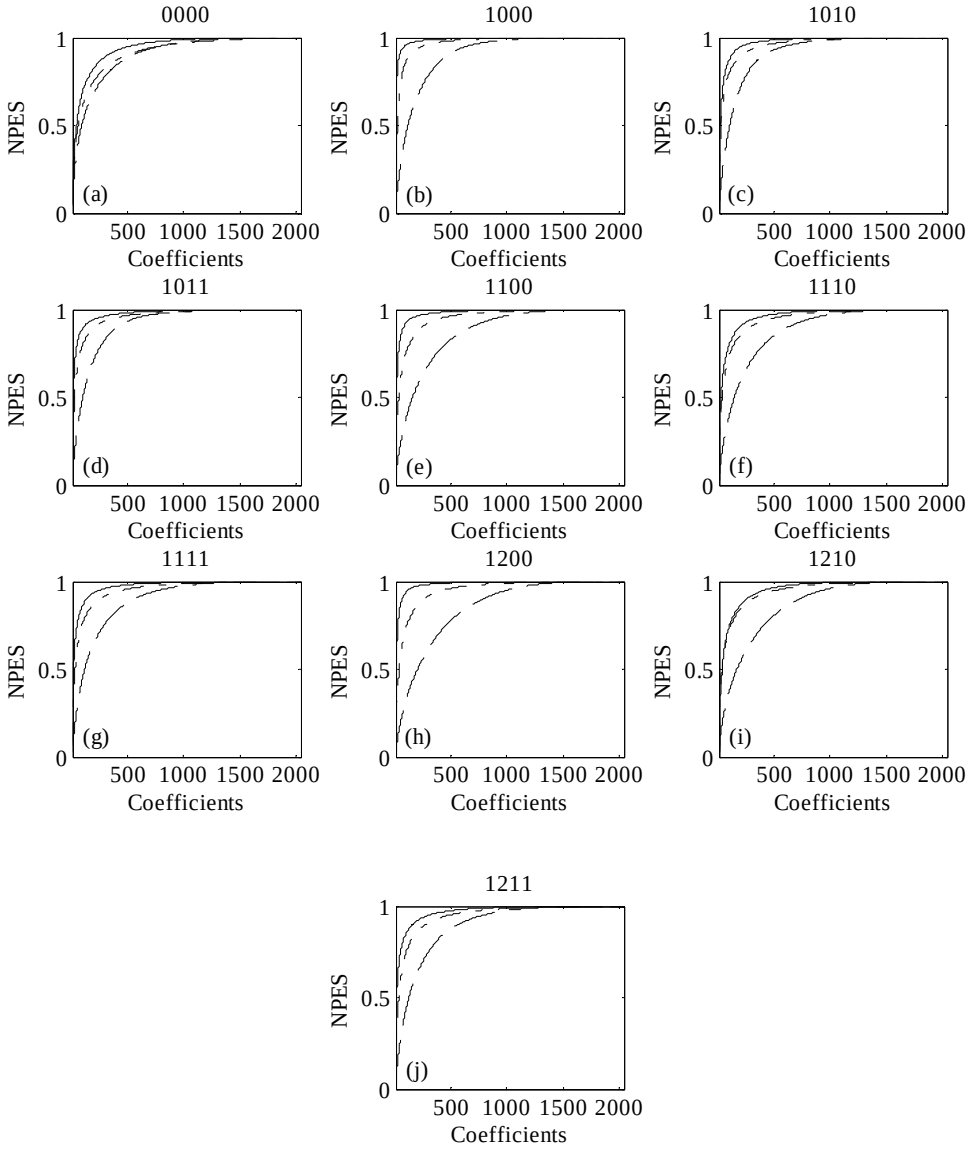


Figure 4-2. The NPES curves for manatee vocalizations. The curves are computed using the DFT (solid), the DWT (dashed), and the WPT (dash-dot) averaged over the ten vocalizations within each of the ten vocalization categories of the vocalization library. A steeper slope in the first few coefficients is an indication of more efficient compression.

Table 4-1. The compression efficiencies of manatee vocalizations. For each vocalization category of the vocalization library, the average percentage of coefficients required for compressing 90% of the vocalization signal energy using the DFT, DWT, and WPT is presented. A smaller number is an indication of more efficient compression.

Category	90% Compression Efficiency		
	DFT	DWT	WPT
0000	15.4	25.7	23.5
1000	2.2	21.7	6.6
1010	5.3	19.4	10.0
1011	5.7	19.6	10.1
1100	4.1	31.9	11.7
1110	9.2	29.1	13.5
1111	6.2	25.7	12.3
1200	3.0	37.0	12.5
1210	11.4	33.9	14.5
1211	7.2	26.2	13.3
Average	7.0	27.0	12.8

Recall that wavelet domain coefficient shrinkage denoising relies on the mapping of the vocalization signals as a few large amplitude coefficients and noise as many small amplitude coefficients. Thus, while it is necessary for a time-frequency transform to efficiently represent vocalization signals with only a few coefficients, this is not sufficient for achieving a high denoising performance. In addition, the noise signal energy must be spread over all the coefficients, which means that the transform must be very inefficient in compactly representing noise signals. The discernability of the

vocalizations signals may be lost if the noise is mapped as efficiently as the vocalizations. Thus, the NPES of the DFT, DWT, and WPT are also computed and presented in Figure 4-3 for four underwater background noise recordings. The percentage of coefficients required to compress 90% of the total noise signal energies are presented in Table 4-2.

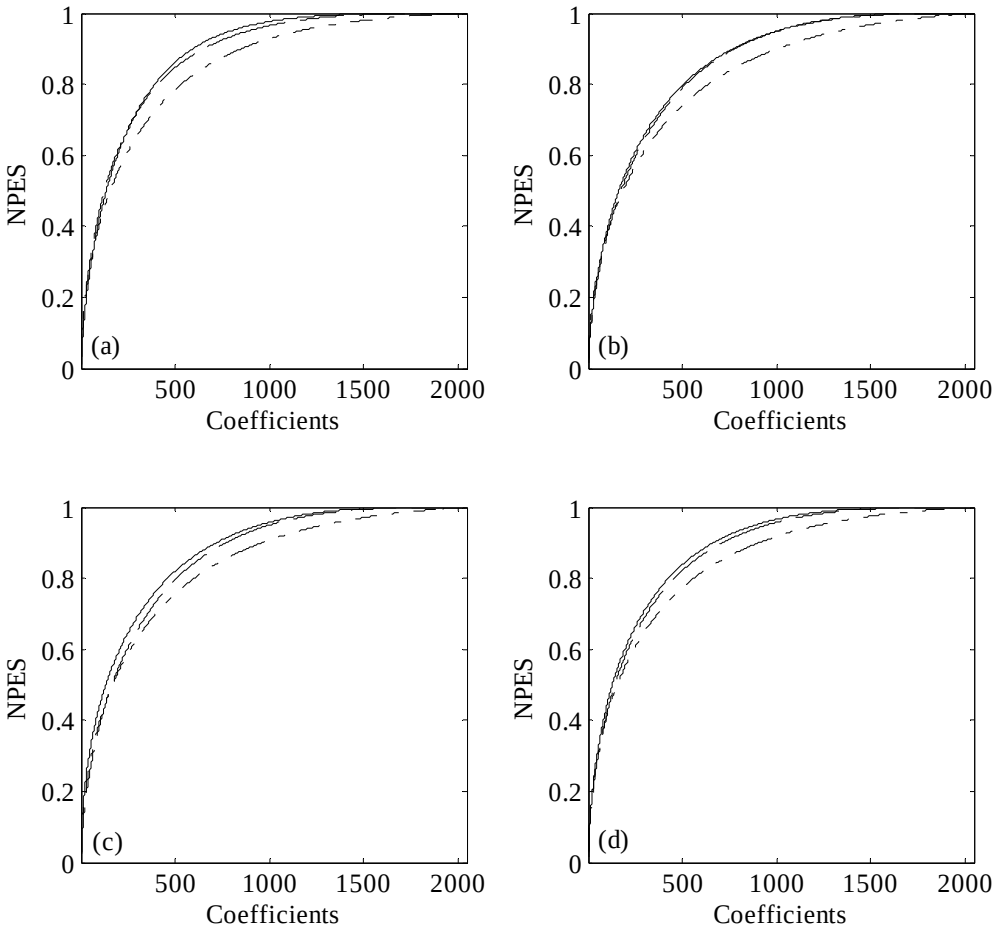


Figure 4-3. The NPES curves for watercraft emitted noise. The curves are computed using the DFT (solid), the DWT (dashed), and the WPT (dash-dot) for four typical noise recordings made in Crystal River, FL. A steeper slope in the first few coefficients is an indication of more efficient compression.

Table 4-2. The compression efficiencies of watercraft emitted noise. The average percentage of coefficients required for compressing 90% of the watercraft emitted noise signal using the DFT, DWT, and WPT is presented. A smaller number is an indication of more efficient compression.

Noise Recording	90% Compression Efficiency		
	DFT	DWT	WPT
1	29.2	32.0	41.8
2	37.6	38.2	49.2
3	35.2	37.7	47.0
4	32.6	34.9	44.3
Average	33.6	35.7	45.6

On average, the DFT and the WPT can compress 90% of the vocalization energy into the largest 7% and 12.8% of the Fourier and wavelet coefficients, respectively. However, the WPT spreads 90% of the background noise energy over 45.6% of the coefficients compared to the 33.6% of the DFT. Thus, while the WPT is slightly less effective in efficiently representing the manatee vocalizations, thresholding the WPT coefficients can result in a better denoising performance compared to thresholding the Fourier coefficients.

A wavelet basis that matches the waveform of the vocalizations or that possesses an autocorrelation sequence similar to the autocorrelation of the vocalizations can efficiently represent the vocalization signals with a few coefficients and thus, simplify the

estimation of the noise-free vocalizations. However, the variability in the time-frequency structure and statistical properties of manatee vocalizations does not permit the design of such an optimal wavelet basis function. Instead, the Daubechies wavelet basis functions with an 8th order vanishing moment (i.e., number of oscillations about the abscissa) is selected for decomposing manatee vocalizations. The Daubechies wavelet family is optimal in the sense that it possesses the maximum order of vanishing moments (which enables efficiently mapping more complex waveforms) for a given time duration or temporal resolution [135]. A plot of the lowpass and highpass perfect reconstruction quadrature mirror filters (QMF) of the Daubechies-8 wavelet family is presented in Figure 4-4. Although not an optimal decomposition for manatee vocalizations, the WPT decomposition using the Daubechies-8 wavelet basis functions has significant advantages over the STFT in terms of discerning vocalizations signals from watercraft emitted noise. This conclusion is supported by the experimental results presented in Chapter 6.

The two remaining elements of the WDD strategy that are not yet addressed are the threshold rule and the threshold. In particular, the selection of the threshold is crucial for effective denoising and the threshold should adapt to changing vocalization and noise signal statistics. More specifically, the threshold value should increase to a sufficiently high value to eliminate all the wavelet coefficients during the silent regions when no vocalization is present. In contrast, the threshold should decrease to an appropriate level such that the coefficients due to the vocalization signals are not eliminated. An insightful perspective to the optimal threshold and threshold rule is presented in the context of optimal Wiener filtering in the next section.

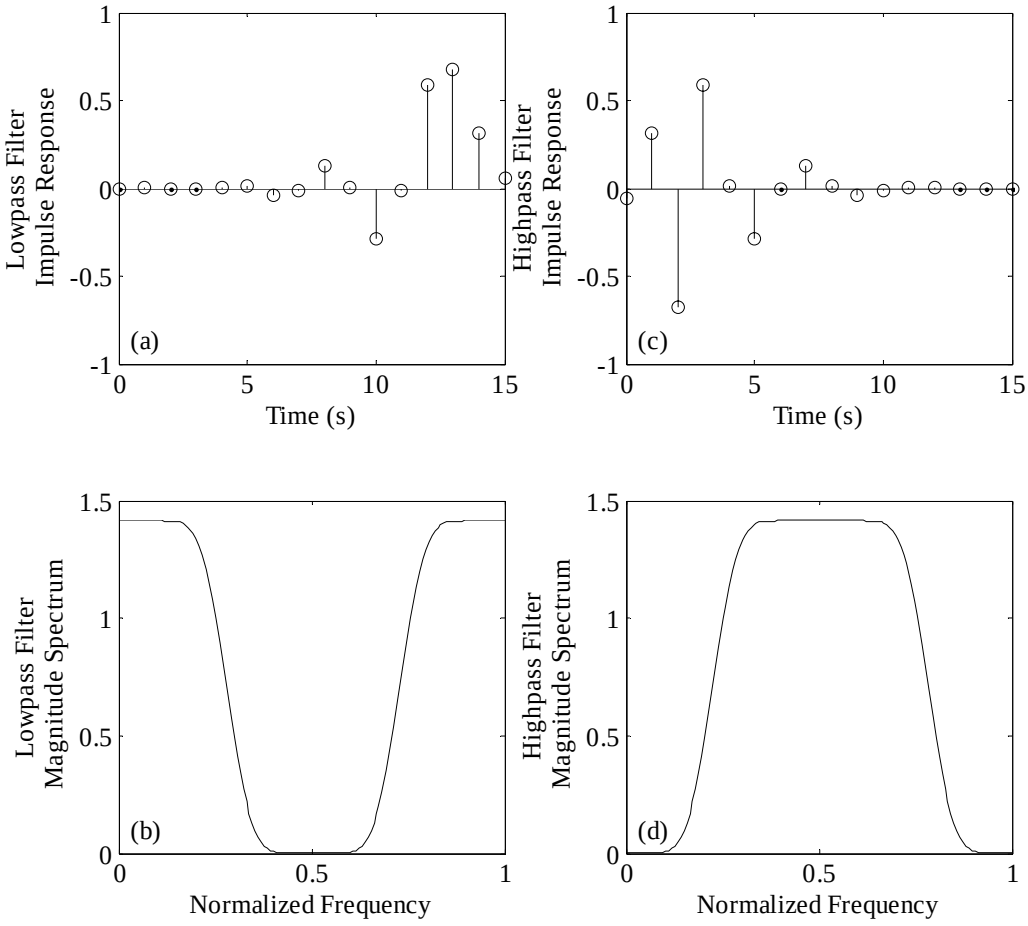


Figure 4-4. The 16 tap Daubechies-8 decomposition filters. (a)-(b) Lowpass and highpass filter impulse responses, respectively, and (c)-(d) the corresponding magnitude spectra of the lowpass and highpass decomposition filters, respectively.

4.3 Adaptive Wavelet Denoising

An interesting analogy between the wavelet denoising filters and the Wiener filter has been revealed in the work of Ghael et al. [136]. The Wiener filter, which is defined as

$$\begin{aligned}\mathbf{W}_{\text{opt}} &= \mathbf{R}_{xx}^{-1}\mathbf{r}_{xs} \\ &= \left\{E[(\mathbf{s}+\mathbf{v})(\mathbf{s}+\mathbf{v})^H]\right\}^{-1} E[(\mathbf{s}+\mathbf{v})\mathbf{s}^H] \\ &= [\mathbf{R}_{ss}+\mathbf{R}_{vv}]^{-1}\mathbf{R}_{ss},\end{aligned} \tag{4-18}$$

where $\mathbf{R}_{xx}$, $\mathbf{R}_{ss}$, $\mathbf{R}_{vv}$ are the input signal, vocalization, and noise autocorrelation matrices, and $\mathbf{r}_{xs}$ is the cross-correlation vector between the input and vocalization signals, uses prior information available on the filter input and noise-free vocalization signal statistics to obtain the minimum MSE estimate of the vocalization signal. The singular value decomposition (SVD) of the Toeplitz vocalization signal autocorrelation matrix

$$\mathbf{R}_{ss} = \mathbf{U}_s^H \mathbf{\Lambda}_s \mathbf{U}_s, \tag{4-19}$$

can be combined with the noise autocorrelation matrix $\mathbf{R}_{vv} = \sigma_v^2 \mathbf{I}$ to obtain an alternate representation of the Wiener filter

$$\mathbf{W}_{\text{opt}} = \mathbf{U}_s^H \left[\frac{\mathbf{\Lambda}_{ss}}{\mathbf{\Lambda}_{ss}+\sigma_v^2\mathbf{I}}\right]\mathbf{U}_s. \tag{4-20}$$

The minimum MSE estimate of the target signal can be computed from

$$\hat{\mathbf{s}}_{\text{MMSE}} = \left\{ \mathbf{U}_s^H \left[\frac{\mathbf{\Lambda}_{ss}}{\mathbf{\Lambda}_{ss} + \sigma_v^2 \mathbf{I}} \right] \mathbf{U}_s \right\} \cdot \mathbf{x}. \tag{4-21}$$

The transform matrix $\mathbf{U}_s$ that consists of eigenvectors of the signal covariance matrix is computed using the Karhunen-Loeve transform (KLT). Therefore, the optimum unitary transform in the MSE sense for threshold based denoising in the presence of Gaussian noise is the KLT. Comparing the optimal Wiener estimate of Eq. (4-21) to the wavelet denoising estimate given by Eq. (4-2) reveals two key relations between the wavelet domain filters and the Wiener filter. First, the target signal eigenvector matrix $\mathbf{U}_s$ is replaced by the orthogonal DWT matrix $\mathcal{W}$ in the wavelet denoising approach. The columns of the eigenvector matrix $\mathbf{U}_s$ are the N mutually uncorrelated signal subspaces. Secondly, the threshold rules δ_n (i.e., the diagonal entries of the threshold matrix) are replaced by the ratio

$$\frac{\lambda_{s,n}^2}{\lambda_{s,n}^2 + \sigma_v^2}, \quad n = 0,1,\ldots,N-1. \tag{4-22}$$

The Wiener filter enhances the dominant signal modes with large eigenvalues and the enhancement constant approaches to 1 when $\lambda_{s,n}^2 \gg \sigma_v^2$. The enhancement constant diminishes for modes with $\lambda_{s,n}^2 \ll \sigma_v^2$ (i.e., dominated by noise). In contrast to the continuous set of values in $[0,1]$ that the enhancement constants can take in Wiener filtering, the constants are restricted to zero and one for the hard threshold rule and to the range $[0,1-\eta/|w_x|]$ for the soft threshold rule. Two important conclusions can be drawn

from the above discussion for developing wavelet denoising strategies for minimum MSE performance. First, for effective denoising of vocalizations, the eigenvectors of the vocalization autocorrelation matrix of the vocalization signals should coincide with the wavelet basis functions. As is discussed in the previous section, without prior knowledge of the vocalization signals, the eigenvectors of the vocalization signal autocorrelation matrix (or the modes of the vocalization signals) will not completely be known in advance. These modes will also differ from one vocalization to another. Therefore, selecting an optimum wavelet that fits all vocalizations may be difficult or impossible. Still, for stereotypical vocalizations, the time-frequency properties of the vocalizations should influence the choice of wavelet basis functions. The second conclusion is that the minimum MSE threshold operator is a smooth function that linearly operates on the input coefficients and adapts according to changing noise and signal statistics.

Conventional orthogonal wavelet denoising is a block processing approach in which a single threshold is used to denoise a set of measurements. 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. Adaptively estimating the optimum threshold for DWT based wavelet denoising was proposed by Zhang [137] where it was proven that the optimal threshold (if it exists) is unique for the soft thresholding rule, but that this result does not generalize to the hard threshold rule. The cited author proposed an adaptive wavelet denoising algorithm based on the gradient

descent optimization rule to estimate the threshold by minimizing the instantaneous estimate of Stein's unbiased risk (see Eq. (4-14)). The resulting scale dependent, adaptive threshold estimator has the form

$$\eta_j(n+1) = \eta_j(n) + \mu_j \frac{\partial}{\partial \eta_j} \mathcal{J}_{\text{SURE}}(n), \quad \begin{matrix} n = 0,1,\ldots,N-1 \\ j = 1,2,\ldots J, \end{matrix} \tag{4-23}$$

where

$$\frac{\partial}{\partial \eta_j} \mathcal{J}_{\text{SURE}}(n) = 2g(w_{x,j}(n),\eta_j) \frac{\partial}{\partial \eta_j} g(w_{x,j}(n),\eta_j) + 2 \frac{\partial^2}{\partial w_{x,j} \partial \eta_j} g(w_{x,j}(n),\eta_j), \tag{4-24}$$

and $g(w_{x,j}(n),\eta_j) = \delta(w_{x,j}(n),\eta_j) - w_{x,j}(n)$ is a nonlinear function. The soft and hard threshold rules were modified to ensure that the gradients of Eq. (4-24) are non-zero (i.e., the adaptation does not stop) when $\left|w_{x,j}(n)\right| < \eta_j(n)$.

As shown in Section 4.2, the octave band decomposition structure of the dyadic DWT is too coarse and less efficient for processing manatee vocalizations compared to the WPT. Therefore, the orthogonal WPT is proposed for the SURE based adaptive threshold denoising algorithm that is presented in this section. The resulting algorithm, termed the adaptive wavelet packet algorithm (AWPT), is implemented using the Daubechies-8 wavelet. The block diagram of the AWPT algorithm is presented in Figure 4-5.

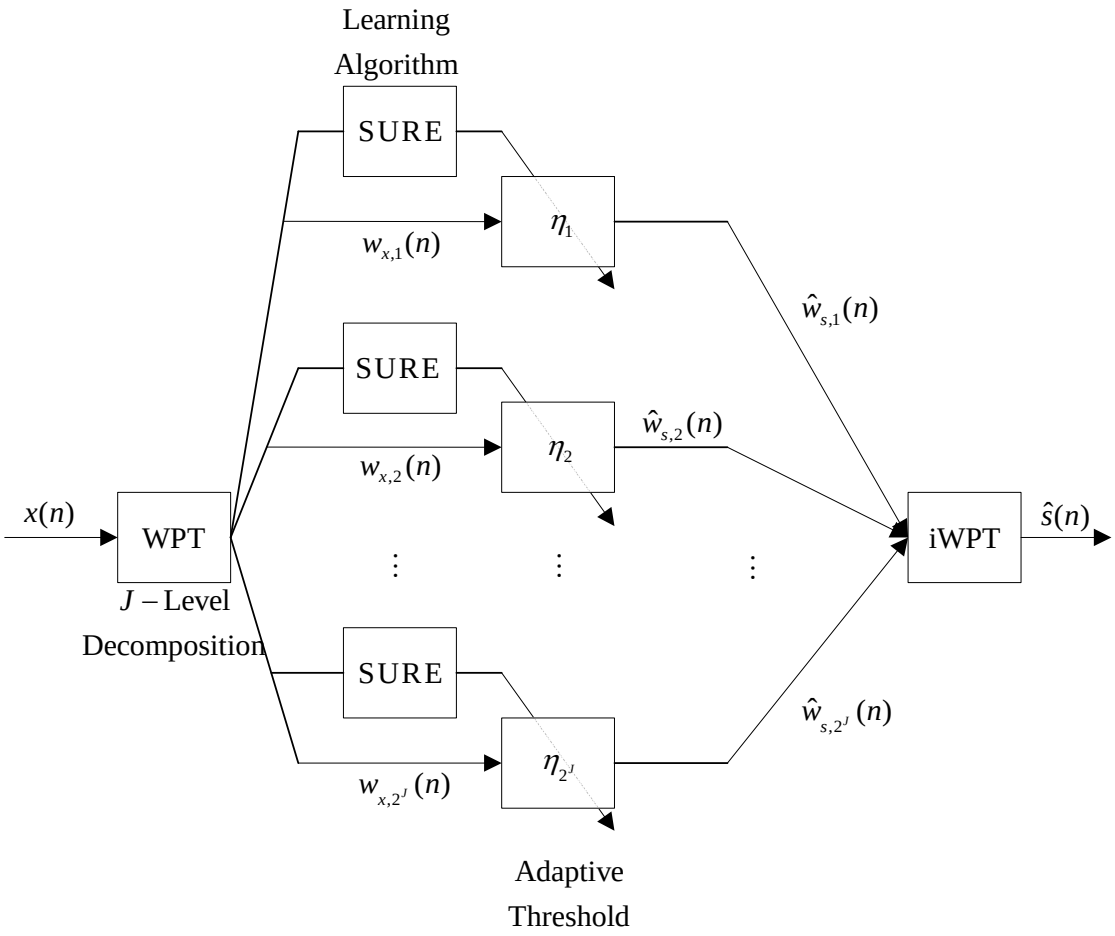


Figure 4-5. The block diagram of the AWPT algorithm.

Several hard and soft threshold rules with continuous derivatives were evaluated using artificial manatee vocalizations and watercraft emitted noise. These simulations revealed that the soft threshold rule results in better MSE performance for weak vocalization signals. This result is in agreement with the conclusions drawn by Zhang [137]. Therefore, the modified soft threshold rule proposed by Zhang

$$\delta(w,\eta) = w + \frac{1}{2}\left([(w-\eta)^2 + \lambda]^{1/2} - [(w+\eta)^2 + \lambda]^{1/2}\right), \quad (4\text{-}25)$$

where λ is the smoothness parameter, is used for the proposed AWPT algorithm. As mentioned in the theoretical development presented in Section 4.1, level dependent thresholds are more appropriate for denoising signals contaminated with temporally correlated noise. Since the noise emitted from surface vessels exhibit some temporal correlation due to channel propagation, level dependent thresholds are used for denoising each subband.

4.4 Fourier Domain Shrinkage

As described in Chapter 3, manatee vocalizations are harmonic signals that can be very efficiently be represented in the Fourier domain. Furthermore, the principle eigenvectors of the covariance matrix of a harmonic signal are the individual sinusoids at the different harmonic frequencies. Therefore, as an alternative to wavelet domain denoising, a linear Fourier coefficient shrinkage based denoising is proposed in this section. The proposed method is a non-adaptive block processing algorithm based on the James-Stein estimator and was discussed in Donoho and Johnstone [126]. James and Stein [138] proposed a linear estimator of a signal corrupted with Gaussian noise $\mathbf{v} \sim \mathcal{N}(0, \mathbf{R}_{vv})$ as

$$\hat{s}(n) = \left\{1 - \frac{N-2}{\mathbf{x}^T \hat{\mathbf{R}}_{vv} \mathbf{x}}\right\} x(n), \tag{4-26}$$

where $\hat{\mathbf{R}}_{vv}$ is an independent estimate of the noise covariance matrix. The discrete Fourier transform (DFT) of the noisy measurements are computed and the resulting Fourier coefficients are grouped into equal size bins indexed by $k = 1, 2, \ldots, K$. Because the DFT approximately decorrelates the time domain measurements, the signal coefficients are estimated from

$$\hat{S}_k(\omega) = \max\left\{1 - \frac{(N_k - 2)}{\mathbf{X}_k^T \mathbf{X}_k}, 0\right\} \cdot X_k(\omega), \quad k = 1, 2, \ldots, K, \tag{4-27}$$

where N_k is the number of coefficients in the k^{th} frequency bin. A separate shrinkage factor is computed for the real and imaginary parts of the coefficients. The larger of these shrinkage factors is compared to zero and used to shrink both the imaginary and real parts. The estimate of the signal is obtained from the inverse DFT of the estimates. This method is referred to as FourierShrink in the remainder of this dissertation.

4.5 Numerical Simulations

Within this section, the denoising performances of the AWPT and FourierShrink algorithms are evaluated using artificially generated autoregressive (AR) signals corrupted with Gaussian noise. As described in Section 3.2, the AR signals and Gaussian noise are idealized signal models representing the harmonic manatee vocalizations and watercraft emitted noise. A Monte-Carlo test that consists of 200 independent trials for

each of the 16 input signal to noise ratio (SNR) values in the range from -10 dB to 5 dB is conducted to obtain upper bounds on the expected denoising performance of the tested algorithms. For each Monte-Carlo trial, a new realization of the AR signal (representing the manatee vocalization) and the IID Gaussian signal (representing watercraft emitted noise) is generated. The sampling rate is assumed to be 24 kHz. An artificial manatee vocalization with three frequency modulated harmonics is generated in each trial. The time domain and spectrogram plots of a typical artificial vocalization signal used in these Monte-Carlo trials is presented in Figure 3-4 (c)-(d). The duration of each vocalization is set to 6400 samples (0.27 s). The manatee is assumed to be 20 m away from the hydrophone in a channel of 5 m water depth and the vocalization signal is convolved with the corresponding channel impulse response. The channel impulse response is obtained from the inverse Fourier transform of the complex pressure field computed using the BELLHOP underwater acoustic propagation software as outlined in Section 3.5. Noise is modeled as a zero-mean, temporally correlated Gaussian random process obtained by filtering IID Gaussian noise with the channel impulse response between the receiver and a broadband point source located at an arbitrarily selected distance of 100 m. A more detailed explanation of the simulated underwater channel and the test setup is provided in Appendix C. After filtering with the channel impulse responses, the vocalization and noise signal powers are scaled to match the predefined input SNR and to have a unit signal power, respectively.

The AWPT algorithm is implemented at $J = 7$ scales, which results in a frequency resolution of approximately 100 Hz. The step size and the threshold parameter

(λ, see Eq. (4-25)) are set to 0.005 and 1, respectively. The resulting threshold rule is very smooth and is empirically determined to result in the best signal to distortion ratio (SDR) performance in the range of -10 to 5 dB SNR. The FourierShrink algorithm is implemented using $2^J = 128$ frequency bins (for positive frequencies only), which is equal to the number of subbands resulting from $J = 7$ scale decomposition of the AWPT algorithm. The output SNR (averaged over the 200 Monte-Carlo trials for each input SNR) and the SDR performance resulting from the AWPT and FourierShrink algorithms as a function of input SNR is depicted in Figure 4-6.

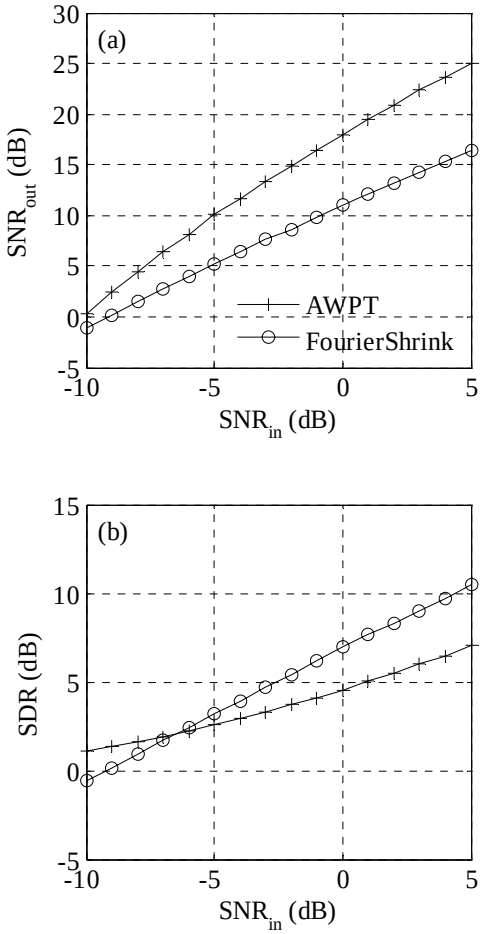


Figure 4-6. Plots of the performance measures for the denoising algorithms obtained from Monte-Carlo trials. (a) The output SNR and (b) the SDR as a function of input SNR for the AWPT and FourierShrink algorithms averaged over the 200 Monte-Carlo trials for each input SNR.

The AWPT algorithm achieves a significantly better output SNR performance compared to the FourierShrink method over the entire input SNR range. On average, the

output SNR resulting from the AWPT algorithm is 7 dB higher than the output SNR of the FourierShrink algorithm. The FourierShrink algorithm, in general, results in a lower signal distortion. The SDR of the FourierShrink algorithm improves at a rate slightly higher than the SDR of the AWPT algorithm as the input SNR is increased. The time domain plots of the denoised signal for a single realization of the Monte-Carlo simulations is shown in Figure 4-7 for an input SNR of 0 dB.

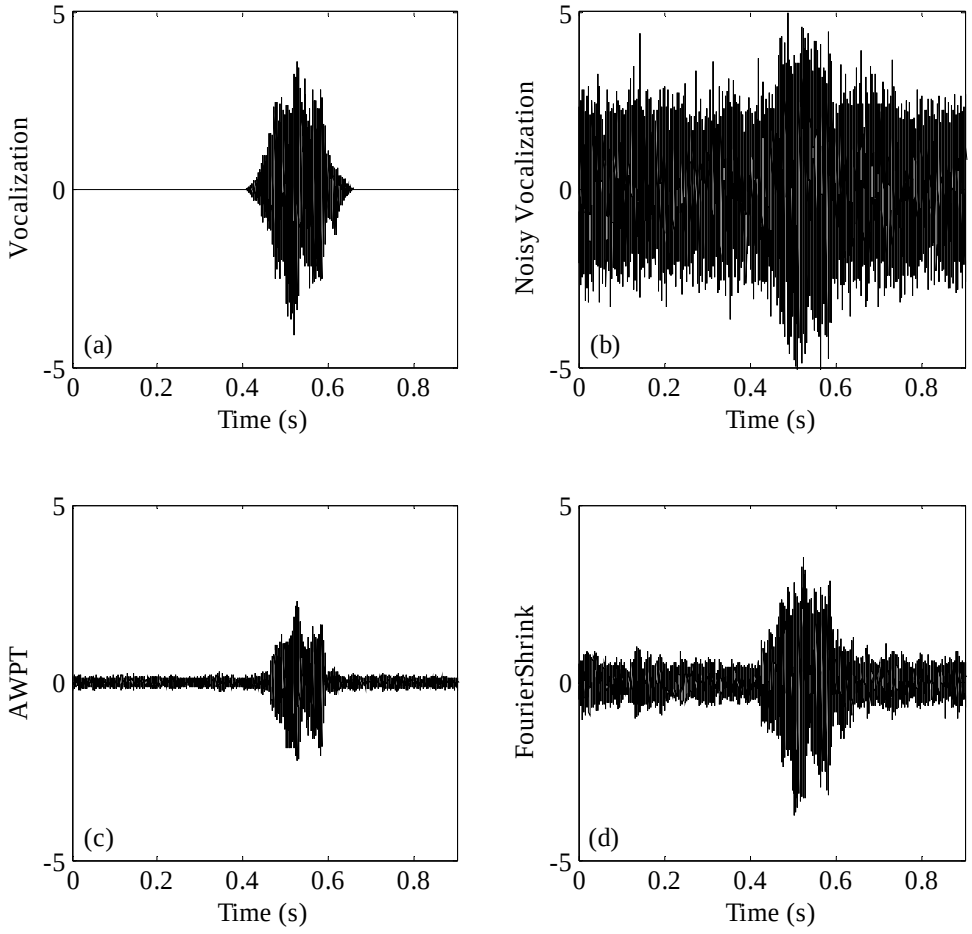


Figure 4-7. Time domain outputs of the denoising algorithms for a single Monte-Carlo trial. The input SNR is set to 0 dB.

4.6 Summary

Wavelet domain enhancement and denoising of manatee vocalizations is presented in this chapter. WDD is introduced from an optimal filtering perspective and similarities between the wavelet domain coefficient thresholding and the Wiener filter are reviewed. This comparison between WDD and the Wiener filter reveals that the wavelet basis functions must match the principle components of the vocalization signal and that the threshold value must adapt to changing signal and noise statistics for the WDD methods to achieve high denoising performance. The optimal denoising strategy for manatee vocalizations is investigated and several conclusions are drawn. First, although the DFT is very effective for efficiently representing manatee vocalizations, the WPT coefficients obtained using the Daubechies-8 wavelet are better for discerning vocalizations from background noise. Second, due to the variations in the time-frequency structure of manatee vocalizations, it is not possible to define a single wavelet basis function or a WPT decomposition structure optimal for all vocalization types. Based on these conclusions, a WPT denoising algorithm that utilizes an adaptive threshold is proposed. As a benchmark, a non-adaptive Fourier coefficient shrinkage algorithm (FourierShrink) is also implemented. The performances of the AWPT and the FourierShrink algorithms are evaluated in terms of the output SNR and SDR performance measures for varying input SNR values through Monte-Carlo trials. The AWPT algorithm achieves the highest output SNR performance while the FourierShrink algorithm outperforms the AWPT in terms of the SDR performance measure.

CHAPTER 5: BLIND SOURCE SEPARATION

In this chapter, blind source separation (BSS) is presented as an alternative to conventional adaptive beamforming for enhancing manatee vocalizations in the presence point acoustic noise sources such as watercraft. The two-input two-output (TITO) configuration for BSS is reviewed first. In the following section, a feedback BSS algorithm based on the affine projection filter update rule is presented.

5.1 The Two-Input Two-Output System

In Section 2.5, the TITO system depicted in Figure 5-1 (in the feedback configuration) was used to show that the symmetric adaptive decorrelator (SAD) algorithm is an extension of the adaptive noise canceller (ANC). In Chapter 6, the affine projection based BSS algorithm proposed in the following sections is implemented exclusively in the TITO configuration for processing real vocalization and noise data. Therefore, the TITO configuration is analyzed in greater detail and some properties of the TITO configuration are presented in this section.

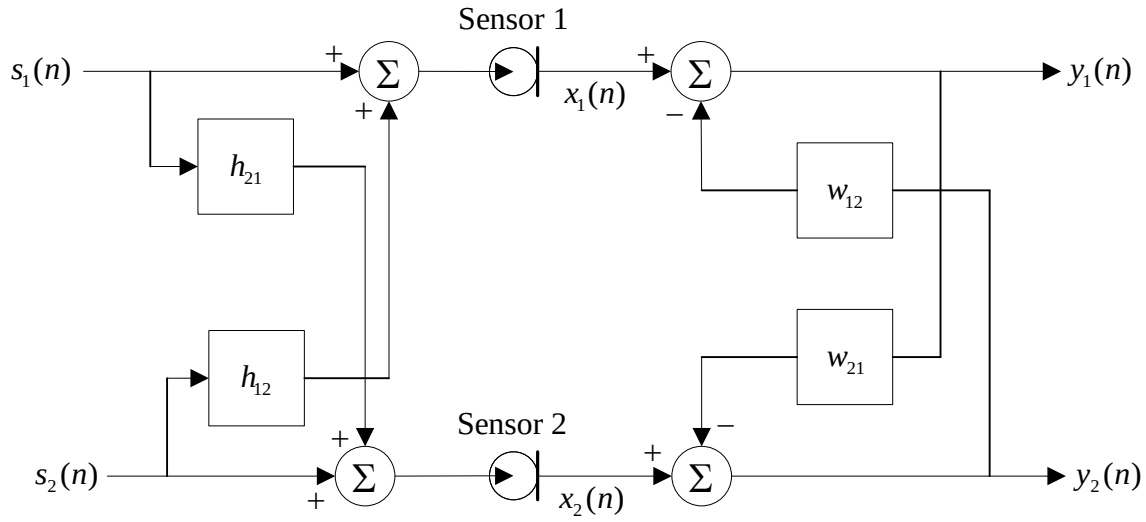


Figure 5-1. The TITO system configuration with a feedback structured separator stage.

For the TITO configuration, the decorrelation criterion can be expressed in terms of the cost function

$$\mathcal{J}(n) = E[y_1(n)y_2(n-k)], \quad 0 \le k \le L-1, \tag{5-1}$$

where $E[\cdot]$ is the expectation operator, $y_q, \quad q = 1, 2,$ are the separated outputs, and L is the separating filter order. Following the underwater acoustic channel models presented in Section 3.5, the pseudo channel models h_{12} and h_{21} are assumed to be finite impulse response (FIR) filters. The two FIR filters h_{12} and h_{21} represent cross-channel coupling paths of the source signals $s_1(n)$ and $s_2(n)$, respectively. The coupling effects of these cross-channel filters are eliminated by the two adaptive filters w_{12} and w_{21}, which are implemented in a feedback structure in Figure 5-1. In the z-domain, the decorrelation

criteria requires the cross-power spectral density (CPSD) $P_{y_1y_2}(z)=0,\ \forall z=\exp(j\omega)$ at convergence, which can be expanded as [63]

$$\begin{aligned} P_{y_1y_2}(z) &= \left[\left(1-w_{12}(z^{-1})h_{21}(z^{-1})\right)\cdot\left(h_{21}(z^{-1})-w_{21}(z^{-1})\right)\right]P_{s_1s_1}(z) \\ &\quad +\left[\left(1-w_{21}(z)h_{12}(z)\right)\cdot\left(h_{12}(z)-w_{12}(z)\right)\right]P_{s_2s_2}(z) \\ &= 0. \end{aligned} \tag{5-2}$$

For broadband source signals such that the power spectral density (PSD) $P_{s_qs_q}(z)\neq 0$, $q=1,2$, a straightforward FIR filter solution that satisfies Eq. (5-2) is given by the pair $w_{12}(z)=h_{12}(z)$ and $w_{21}(z)=h_{21}(z)$. It can be shown that this FIR filter solution also achieves source separation without introducing any permutation ambiguities [63]. However, as can be inferred from Eq. (5-2), the FIR filter solution is not unique. For example, the infinite impulse response (IIR) filter solution $w_{12}(z)=1/h_{21}(z)$ and $w_{21}(z)=1/h_{12}(z)$ also achieves decorrelation and source separation solution (but the outputs are permuted, i.e., $s_p(n)=y_q(n),\quad p\neq q=1,2$). Other IIR filter solutions that satisfy Eq. (5-2) but fail to achieve source separation also exist [62]. Lindgren and Broman [65] proved that the only possible solution to the TITO configuration is the FIR filter solution, provided that the separating filters are constrained to be FIR filters and that $1-h_{12}(z)h_{21}(z)$ is minimum phase.

5.2 The Feedback Double Affine Projection Algorithm

The SAD algorithm (which is discussed at the end of Section 2.5) is a simple and robust BSS algorithm suitable for source separation in the TITO configuration. However, the SAD algorithm is based on the least mean square (LMS) filter update rule and inherits the limitations of the LMS algorithm. More specifically, the convergence of the filter weights is rather slow in the presence of temporally correlated inputs such as the harmonic manatee vocalizations (due to the large eigenvalue spread of the input autocorrelation matrix) which require small step sizes to prevent the separating filters from destabilizing ([139], also see Section 2.4). The small step sizes required for stability degrades the ability of the de-mixing filters to track the changes in the channel impulse responses due to moving watercraft. Simulations conducted with the SAD algorithm using real manatee vocalizations and background noise revealed that the convergence rate and tracking ability of the SAD algorithm is insufficient for achieving a satisfactory enhancement performance. 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 rather than the LMS algorithm [63]. 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, simulations with real vocalization and noise data show that the DRLS algorithm is prone to the same divergence phenomenon observed in supervised RLS algorithms caused by numerical instabilities [140]. To further aggravate the problem, the use of the input and output

cross-covariance matrix 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 [141].

The affine projection (AP) algorithm [142] 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 section, an unsupervised BSS algorithm based on the supervised AP algorithm is derived and implemented for the TITO channel configuration depicted in Figure 5-1. The extension of the supervised AP algorithm to the unsupervised adaptive filtering framework was first proposed by Gabrea [143]. More specifically, the cited author proposed a feedback structured block-update (FB/BU) double affine projection (DAP) algorithm in the context of speech enhancement. The FB/BU DAP algorithm was evaluated using artificial mixtures of real speech and car noise recordings, and was reported to result in 1 to 3 dB performance improvement compared to the SAD algorithm. However, no derivation of the algorithm was provided. In what follows, the derivation of the FB/BU DAP algorithm is provided.

The FB/BU DAP algorithm solves for the separating filters by computing the minimum squared norm filter update

$$\min\left(\left\|\Delta\mathbf{w}_{pq}(n+1)\right\|_2\right)^2 = \min\sum_{l=0}^{L-1}\left(w_{pq,l}(n+1)-w_{pq,l}(n)\right)^2, \quad p\neq q=1,2, \tag{5-3}$$

such that the updated filter decorrelates the past K lag cross-correlations between the filter outputs, where $K<L$ and L is the maximum order of the cross-channel filters. This criterion presents a constrained optimization problem which can be expressed in terms of Lagrange multipliers as

$$\mathcal{J}_1(n)=\left(\left\|\mathbf{w}_{12}(n+1)-\mathbf{w}_{12}(n)\right\|_2\right)^2+\sum_{k=0}^{K-1}\lambda_k r_{y_1y_2}(k), \tag{5-4}$$

where $r_{y_1y_2}(k)$ are the cross-correlations and λ_k, $k=0,1,\ldots,K-1$ are the corresponding Lagrange multipliers. The cross-correlations are unknown and are estimated from their instantaneous realizations as

$$\hat{r}_{y_1y_2}(k)=y_1(n-k)y_2(n). \tag{5-5}$$

To derive the update rule for the first filter $\mathbf{w}_{12}(n)$, Eq. (5-4) is expressed as

$$\mathcal{J}_1(n)=\left(\left\|\mathbf{w}_{12}(n+1)-\mathbf{w}_{12}(n)\right\|_2\right)^2+\left(\mathbf{y}_{1,K}^T(n)y_2(n)\right)\boldsymbol{\lambda}, \tag{5-6}$$

where $\boldsymbol{\lambda}=\begin{bmatrix}\lambda_0 & \lambda_1 & \cdots & \lambda_{K-1}\end{bmatrix}^T$ is the $K\times 1$ vector of Lagrange multipliers and the vector $\mathbf{y}_{1,L}=\begin{bmatrix}y_1(n) & y_1(n-1) & \cdots & y_1(n-L+1)\end{bmatrix}^T$ consists of the past K outputs computed with the updated filters $\mathbf{w}_{12}(n+1)$ such that

$$\mathbf{y}_{1,K}(n) = \mathbf{x}_{1,K}(n) - \mathbf{Y}_2^H(n)\mathbf{w}_{12}(n+1), \tag{5-7}$$

where $\mathbf{Y}_2(n) = \begin{bmatrix} \mathbf{y}_{2,L}(n) & \mathbf{y}_{2,L}(n-1) & \cdots & \mathbf{y}_{2,L}(n-K+1) \end{bmatrix}$ is a $L \times K$ matrix of the past filter outputs and $\mathbf{x}_{1,K}(n) = \begin{bmatrix} x_1(n) & x_1(n-1) & \cdots & x_1(n-K+1) \end{bmatrix}^T$ is the $K \times 1$ vector of inputs in channel 1. The cost function can be expanded by substituting $\mathbf{y}_{1,K}(n)$ as defined Eq. (5-7) into Eq. (5-6) as

$$\begin{aligned} \mathcal{J}_1(n) = &\left(\left\| \mathbf{w}_{12}(n+1) - \mathbf{w}_{12}(n) \right\|_2 \right)^2 \\ &+ \left(\mathbf{x}_1(n) - \mathbf{Y}_2^T(n)\mathbf{w}_{12}(n+1) \right)^T y_2(n)\boldsymbol{\lambda}. \end{aligned} \tag{5-8}$$

Defining the $K \times 1$ cross-correlation vector

$$\hat{\mathbf{r}}_{y_2 x_1} = y_2(n)\mathbf{x}_{1,K}(n) = \begin{bmatrix} y_2(n)x_1(n) & y_2(n)x_1(n-1) & \cdots & y_2(n)x_1(n-K+1) \end{bmatrix}^T, \tag{5-9}$$

and the $L \times K$ cross-correlation matrix

$$\begin{aligned} \hat{\mathbf{R}}_{y_2 y_2} &= y_2(n)\mathbf{Y}_2(n) \\ &= \begin{bmatrix} y_2(n)y_2(n) & y_2(n)y_2(n-1)) & \cdots & y_2(n)y_2(n-K+1) \\ y_2(n)y_2(n-1) & y_2(n)y_2(n-2) & \cdots & y_2(n)y_2(n-K) \\ \vdots & \vdots & \ddots & \vdots \\ y_2(n)y_2(n-L+1) & y_2(n)y_2(n-L) & \cdots & y_2(n)y_2(n-K-L+2) \end{bmatrix}, \end{aligned} \tag{5-10}$$

the cost function can be rewritten in the matrix notation as

$$\mathcal{J}_1(n) = \left(\left\| \mathbf{w}_{12}(n+1) - \mathbf{w}_{12}(n) \right\|_2 \right)^2 + [\hat{\mathbf{r}}_{y_2 x_1} - \hat{\mathbf{R}}_{y_2 y_2}^T \mathbf{w}_{12}(n+1)]^T \boldsymbol{\lambda}. \tag{5-11}$$

A gradient descent optimization method requires differentiating $\mathcal{J}_1(n)$ with respect to $\mathbf{w}_{12}(n+1)$ which results in

$$\frac{\partial}{\partial \mathbf{w}_{12}(n+1)} \mathcal{J}_1(n) = 2[\mathbf{w}_{12}(n+1) - \mathbf{w}_{12}(n)] - \mathbf{R}_{y_2 y_2} \boldsymbol{\lambda}. \tag{5-12}$$

Setting Eq. (5-12) to zero results in

$$\Delta \mathbf{w}_{12}(n+1) = \frac{1}{2} \mathbf{R}_{y_2 y_2} \boldsymbol{\lambda}. \tag{5-13}$$

Pre-multiplying both sides in Eq. (5-13) with $\mathbf{R}^T_{y_2 y_2}$ gives

$$\hat{\mathbf{R}}^T_{y_2 y_2} \Delta \mathbf{w}_{12}(n+1) = \frac{1}{2} \hat{\mathbf{R}}^T_{y_2 y_2} \hat{\mathbf{R}}_{y_2 y_2} \boldsymbol{\lambda}. \tag{5-14}$$

The updated filter $\mathbf{w}_{12}(n+1)$ is expected to decorrelate the filter outputs at the past K lags, which can be expressed in the matrix notation as

$$\begin{aligned} \hat{\mathbf{r}}_{y_2 x_1} &= \hat{\mathbf{R}}^T_{y_2 y_2} \mathbf{w}_{12}(n+1) \\ &= \hat{\mathbf{R}}^T_{y_2 y_2} [\Delta \mathbf{w}_{12}(n+1) + \mathbf{w}_{12}(n)], \end{aligned} \tag{5-15}$$

or, combining with Eq. (5-14)

$$\frac{1}{2} \hat{\mathbf{R}}^T_{y_2 y_2} \hat{\mathbf{R}}_{y_2 y_2} \boldsymbol{\lambda} = \hat{\mathbf{r}}_{y_2 x_1} - \hat{\mathbf{R}}^T_{y_2 y_2} \mathbf{w}_{12}(n). \tag{5-16}$$

Solving for the Lagrange multiplier vector from Eq. (5-16), one gets

$$\boldsymbol{\lambda} = 2\left(\hat{\mathbf{R}}^T_{y_2 y_2} \hat{\mathbf{R}}_{y_2 y_2}\right)^{-1} \cdot \left(\hat{\mathbf{r}}_{y_2 x_1} - \hat{\mathbf{R}}^T_{y_2 y_2} \mathbf{w}_{12}(n)\right), \tag{5-17}$$

And upon substituting Eq. (5-17) into Eq. (5-13), the filter update is derived to be

$$\mathbf{w}_{12}(n+1) = \mathbf{w}_{12}(n) + \hat{\mathbf{R}}_{y_2y_2}\left(\hat{\mathbf{R}}_{y_2y_2}^T\hat{\mathbf{R}}_{y_2y_2}\right)^{-1}\cdot\left(\hat{\mathbf{r}}_{y_2x_1} - \hat{\mathbf{R}}_{y_2y_2}^T\mathbf{w}_{12}(n)\right). \tag{5-18}$$

However, the update equation can further be simplified by rewriting Eq. (5-18) as

$$\begin{aligned}\mathbf{w}_{12}(n+1) &= \mathbf{w}_{12}(n) + y_2(n)\mathbf{Y}_2(n)\left(y_2(n)\mathbf{Y}_2^T(n)\mathbf{Y}_2(n)y_2(n)\right)^{-1}\\ &\quad\cdot y_2(n)\left(\mathbf{x}_{1,K}(n) - \mathbf{Y}_2^T(n)\mathbf{w}_{12}(n)\right)\\ &= \mathbf{w}_{12}(n) + \mathbf{Y}_2(n)\left(\mathbf{Y}_2^T(n)\mathbf{Y}_2(n)\right)^{-1}\cdot\left(\mathbf{x}_{1,K}(n) - \mathbf{Y}_2^T(n)\mathbf{w}_{12}(n)\right)\\ &= \mathbf{w}_{12}(n) + \mathbf{Y}_2(n)\left(\mathbf{Y}_2^T(n)\mathbf{Y}_2(n)\right)^{-1}\cdot\mathbf{y}_{1,K}(n).\end{aligned} \tag{5-19}$$

The stability and rate of convergence of the algorithm can be controlled by introducing a step size constant μ into the filter update. Numerical problems that may be encountered during the computation of $\left(\mathbf{Y}_2^T(n)\mathbf{Y}_2(n)\right)^{-1}$ can be eliminated by introducing a regularization factor $\delta\mathbf{I}$. The resulting algorithm is the FB/BU DAP implemented in the time domain for TITO systems

$$\mathbf{w}_{12}(n+1) = \mathbf{w}_{12}(n) + \mu\mathbf{Y}_2(n)\left(\mathbf{Y}_2^T(n)\mathbf{Y}_2(n) + \delta\mathbf{I}\right)^{-1}\mathbf{y}_{1,K}(n). \tag{5-20}$$

The update equation for the other filter w_{21} can be obtained similarly from the cost function

$$\mathcal{J}_2(n) = \left(\left\|\mathbf{w}_{21}(n+1) - \mathbf{w}_{21}(n)\right\|_2\right)^2 + \left(y_1(n)\mathbf{y}_{2,K}^T(n)\right)\boldsymbol{\lambda}, \tag{5-21}$$

where the cross-correlation is computed from the relation

$$\mathbf{y}_{2,K}(n) = \mathbf{x}_{2,K}(n) - \mathbf{Y}_1^H(n)\mathbf{w}_{21}(n+1). \tag{5-22}$$

The resulting update rule is

$$\mathbf{w}_{21}(n+1) = \mathbf{w}_{21}(n) + \mu_2 \mathbf{Y}_1(n)\left(\mathbf{Y}_1^T(n)\mathbf{Y}_1(n) + \delta\mathbf{I}\right)^{-1} \mathbf{y}_{2,K}(n). \tag{5-23}$$

In Eqs. (5-20) and (5-23), each of the filters with L unknown coefficients are obtained by solving the K decorrelation equations $K < L$ and picking the minimum norm solution from a set of possible filters. The computation involves the inversion of a $K \times K$ matrix.

The FB/BU DAP algorithm uses the $L \times K$ matrices

$$\begin{aligned}\mathbf{Y}_p(n) &= \begin{bmatrix} \mathbf{y}_{p,L}(n) & \mathbf{y}_{p,L}(n-1) & \cdots & \mathbf{y}_{p,L}(n-K+1) \end{bmatrix} \\ &= \begin{bmatrix} y_p(n) & y_p(n-1) & \cdots & y_p(n-K+1) \\ y_p(n-1) & y_p(n-2) & \cdots & y_p(n-K) \\ \vdots & \vdots & \ddots & \vdots \\ y_p(n-K+1) & y_p(n-K) & \cdots & y_p(n-K) \\ \vdots & \vdots & \ddots & \vdots \\ y_p(n-L+1) & y_p(n-L) & \cdots & y_p(n-L-K+2) \end{bmatrix},\end{aligned} \tag{5-24}$$

where $p = 1, 2$, to compute the filter outputs. However, at the time of this computation, the last K outputs $\mathbf{y}_{p,K}(n) = \begin{bmatrix} y_p(n) & y_p(n-1) & \cdots & y_p(n-K+1) \end{bmatrix}^T$ will not be available to complete the upper left $K \times K$ triangular portion of $\mathbf{Y}_p(n)$. Depending on the number of cross-correlations to be cancelled (i.e., K), the missing upper left portion of $\mathbf{Y}_p(n)$ can reduce the convergence rate of the filters and the separation performance of the algorithm. A simple fix to correct this problem is to convert the FB/BU DAP algorithm to a sequential update algorithm (FB/SU DAP) by updating the separation filters with each new measurement. In the sequential algorithm, the only missing entry of $\mathbf{Y}_p(n)$ is the top left corner entry $y_p(n)$.

5.3 The Feedforward Double Affine Projection Algorithm

To circumvent the stability and causality problems associated with the FB/BU DAP algorithm, a feedforward implementation of the DAP algorithm with sequential update of the filter coefficients (FF/SU DAP) is derived in this section. The block diagram of the feedforward setup is depicted in Figure 5-2.

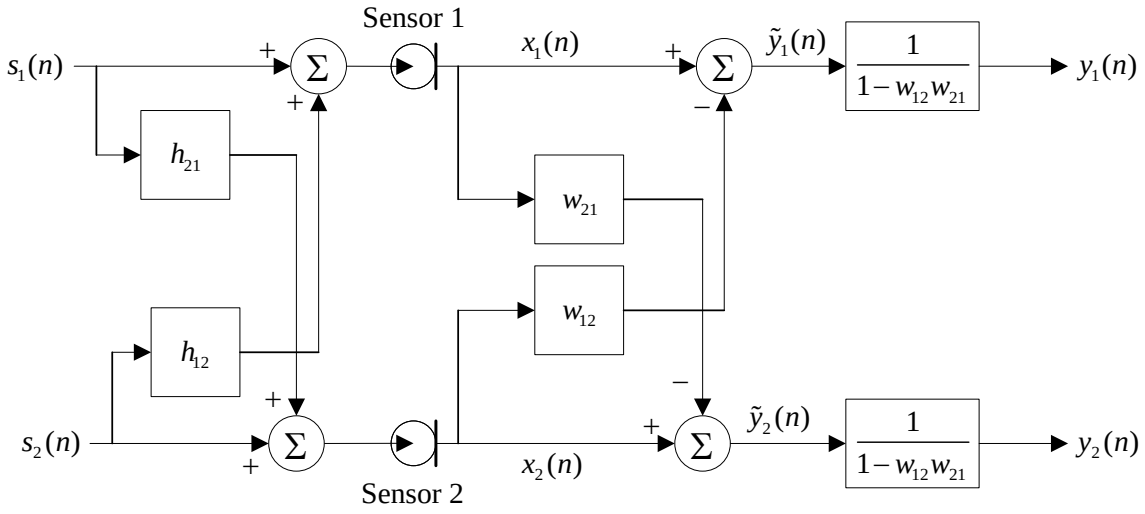


Figure 5-2. The TITO system with a feedforward structured separator stage.

As in the case for the FB/BU DAP algorithm, 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

$$\mathcal{J}_1(n) = \left(\left\| \mathbf{w}_{12}(n+1) - \mathbf{w}_{12}(n) \right\|_2 \right)^2 + \left(\tilde{\mathbf{y}}_{1,K}^T(n) \tilde{y}_2(n) \right) \boldsymbol{\lambda}, \tag{5-25}$$

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

$$\tilde{\mathbf{y}}_{1,K}(n) = \mathbf{x}_{1,K}(n) - \mathbf{X}_2^H(n) \mathbf{w}_{12}(n+1), \tag{5-26}$$

where $\mathbf{X}_2(n) = \left[\mathbf{x}_{2,L}(n) \quad \mathbf{x}_{2,L}(n-1) \quad \cdots \quad \mathbf{x}_{2,L}(n-K+1) \right]$ is a $L \times K$ matrix of the past filter outputs. Following the derivation of the FB/BU DAP algorithm as in the previous section, the resulting filter update equations take the form

$$\mathbf{w}_{12}(n+1) = \mathbf{w}_{12}(n) + \mu \mathbf{X}_2(n) \left(\mathbf{X}_2^T(n) \mathbf{X}_2(n) + \delta \mathbf{I} \right)^{-1} \tilde{\mathbf{y}}_{1,K}(n), \tag{5-27}$$

and

$$\mathbf{w}_{21}(n+1) = \mathbf{w}_{21}(n) + \mu \mathbf{X}_1(n) \left(\mathbf{X}_1^T(n) \mathbf{X}_1(n) + \delta \mathbf{I} \right)^{-1} \tilde{\mathbf{y}}_{2,K}(n). \tag{5-28}$$

A detailed derivation of the FF/SU DAP algorithm is presented in Appendix D.

5.4 Enhancement of Manatee Vocalizations

Manatee vocalizations are, in general, narrowband signals that have signal energy at a few frequencies. Examining the decorrelation equation presented in Eq. (5-2), where $s_1(n)$ is assumed to represent manatee vocalizations, reveals that the convergence of only a few filter coefficients of w_{21} (i.e., those that correspond to the harmonic frequencies $z = \exp(j\omega)$ for non-zero values of $P_{s_1 s_1}(z)$) is sufficient to satisfy the decorrelation

condition. For coefficients of the de-mixing filter w_{21} that correspond to frequencies without any vocalization energy, the decorrelation criteria will not be able to force them to converge to the corresponding coefficients of the mixing filter h_{21}. However, since the de-mixing filter coefficients that do not converge correspond to zero vocalization signal energy, separation will still be achieved as long as the non-converging filter coefficients do not randomly drift away form zero values. Thus, it is expected that changing the manatee path pseudo impulse response will have a lesser effect on the performance of the BSS algorithms compared to the pseudo impulse response corresponding to the watercraft noise transmission. Simulations conducted with real noise data presented in Chapter 6 will verify this conclusion.

5.5 Numerical Simulations

The noise reduction performance of the affine projection based BSS algorithms derived in the previous sections are evaluated by Monte-Carlo simulations. As was done in Section 4.5, 6th order Gaussian AR signals are used to represent vocalizations with frequency modulated harmonics, while watercraft emitted noise is modeled an independent and identically distributed (IID) Gaussian signal. The IID Gaussian noise is filtered with a BELLHOP computed channel impulse response to introduce temporal correlation. However, the synthetic mixing of the artificial vocalization and watercraft emitted noise signals is performed using analytically computed multi-tap FIR filters (as outlined in Section 2.5). For each mixture, Monte-Carlo simulations that consist of 100 independent trials are conducted in the range of -10 to 5 dB input signal to noise ratio

(SNR) measured at the reference receiver of the vocalizations. The input SNR of the other channel is determined through the corresponding pseudo channel impulse response and is always lower than the reference channel input SNR. A single step size parameter, which results in an optimal signal to distortion ratio (SDR) performance, is selected for each tested source separation algorithm over the range of input SNR values. In particular, the step sizes used in the FB/BU DAP, FB/SU DAP, and FF/SU DAP algorithms are 0.1, 0.01, and 0.01, respectively.

For the first set of tests, the order of the three DAP algorithms are fixed to $K = 10$, channel impulses responses are computed assuming three arrivals, and extraneous noise is ignored. Hence, the coherence of the measurements between the two channels is close to unity (or 0 dB) for all frequencies. The output SNR and the SDR performance measures resulting from these three affine projection based algorithms as a function of input SNR is depicted in Figure 5-3 for the 100 Monte-Carlo trials.

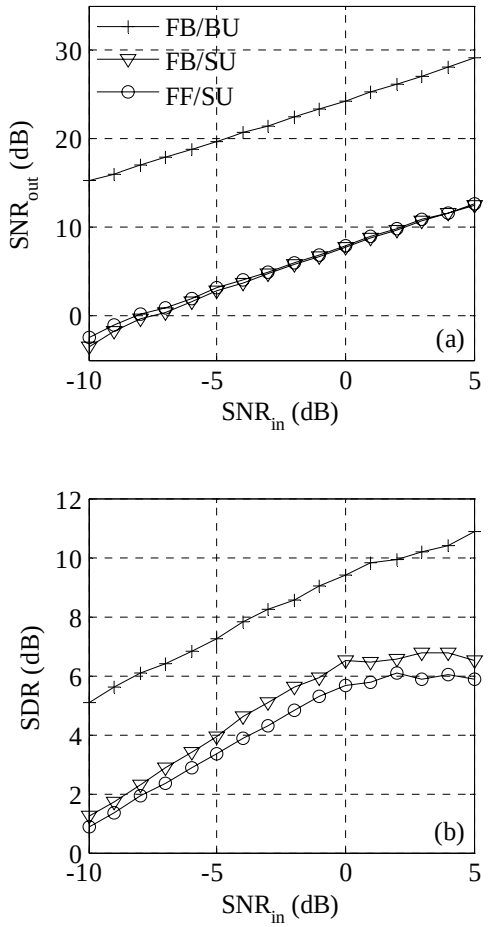


Figure 5-3. Plots of the performance measures for the DAP algorithms obtained from Monte-Carlo trials. (a) The output SNR and (b) the SDR as a function of input SNR for the FB/BU DAP, FB/SU DAP, and FF/SU DAP algorithms averaged over the 100 Monte-Carlo trials for each input SNR.

The FF/SU implementation of the DAP algorithm clearly outperforms both feedback implementations in terms of the output SNR and SDR performance measures. The improved performance of the FF/SU DAP algorithm for both performance measures

is partially attributed to the completeness of the cross-correlation matrices used in updating the separation filters. For this reason, no further simulation results are reported for the feedback DAP algorithms in the remainder of this chapter. The simulations suggest that the output SNR resulting from the feedforward implementation is in the range of 15 dB to 30 dB.

Synthetic mixing and the lack of extraneous noise results in a coherence that is typically higher than what is observed with real noise and vocalization recordings. Hence, the results presented in Figure 5-3 overestimate the expected in-field performance of the FF/SU DAP algorithm. Next, the performance of the FF/SU DAP algorithm in the presence of extraneous noise is evaluated. The coherence functions in the presence of IID Gaussian noise with extraneous signal power resulting in -15 dB, -10 dB, and -5 dB SNR with respect to the temporally correlated Gaussian noise (representing watercraft emitted noise) and the corresponding output SNR and SDR performance measures for the FF/SU DAP algorithm is presented in Figure 5-4. The first two tested extraneous noise levels are typical of what is observed in the real watercraft noise recordings used to test the FF/SU DAP algorithm in the next chapter. The output SNR performance of the FF/SU DAP algorithm is observed to vary linearly with input SNR between 4 to 18 dB for -15 dB extraneous noise. When the extraneous noise levels increase to -5 dB, the output SNR drops to -5 dB to 10 dB.

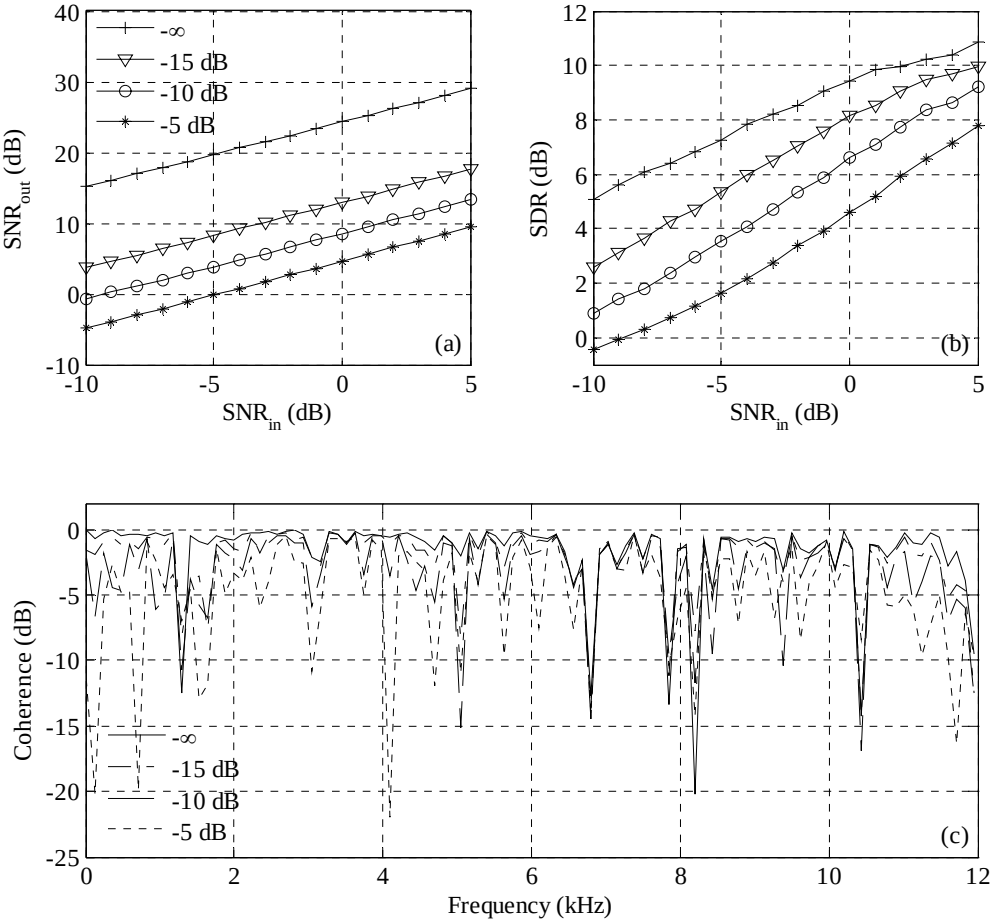


Figure 5-4. Plots of the performance measures for the FF/SU DAP algorithm for various extraneous noise powers obtained from Monte-Carlo trials. (a) The output SNR, (b) the SDR as a function of input SNR averaged over the 100 Monte-Carlo trials for each input SNR, and (c) the coherence function (plotted at 117 Hz intervals for clarity) of the measurements over the duration of the AR signals. Note that negative infinity corresponds to no extraneous noise.

For the above tests, the order (K) of the FF/SU DAP algorithm is fixed to 10. In the next set of simulations, the effect of the order of the algorithm on the signal

enhancement performance is investigated. The IID Gaussian extraneous noise power is set to -15 dB with respect to the artificial watercraft emitted noise power at each channel. The output SNR and SDR performance measures are plotted as a function of the input SNR in Figure 5-5.

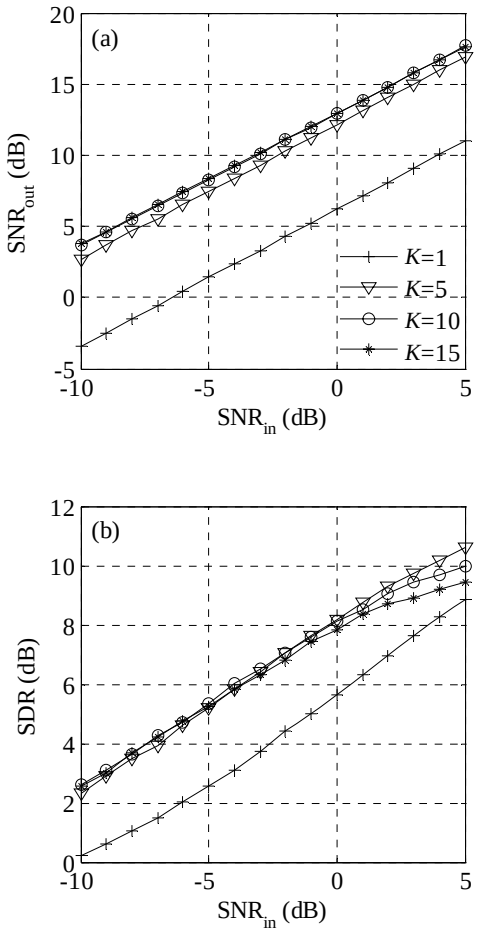


Figure 5-5. Plots of the performance measures for the FF/SU DAP algorithm for various orders K obtained from Monte-Carlo trials. (a) The output SNR and (b) the SDR performance measures as a function of the input SNR averaged over the 100 Monte-Carlo trials for each input SNR.

The filter update rule with $K=1$ corresponds to the unsupervised equivalent of the normalized LMS algorithm. There is a significant improvement as K is increased from 1 to 5, but the same incremental improvement is not observed as K is increased further. A slight improvement in the output SNR is observed as K is increased from 5 to 10, but no consistent improvement is achieved for $K=15$. Hence, the order of the FF/SU DAP algorithm is fixed to $K=10$ for the remainder of this dissertation.

Next, the effect of the channel impulse responses on the output SNR and SDR performance of the FF/SU DAP algorithm is evaluated. For this set of simulations, the number of arrivals (M) assumed in modeling the pseudo channel impulse responses are changed in each Monte-Carlo simulation, effectively changing the number of non-zero weights of the mixing FIR filters h_{12} and h_{21}. The output SNR and SDR results are presented in Figure 5-6 as a function of the number of arrivals assumed. The channel impulse responses for $M=1$ corresponds to signal tap FIR filter model and represents the case when only the direct path transmission is significant. For $M>1$, the higher number of arrivals represents indirect paths for which the wave reflects from the boundaries at least once. As M is increased, the acoustic wave that traverses these paths interacts more frequently with the waveguide boundaries and consequently, attenuates more. Arrivals that traverse the paths for $M>7$ are significantly attenuated and can be ignored. Hence, the maximum number of arrivals in these simulations is limited to $M=7$. The results presented in Figure 5-6 indicate that the output SNR performance is not significantly affected by the increase in the cross-coupling filter orders.

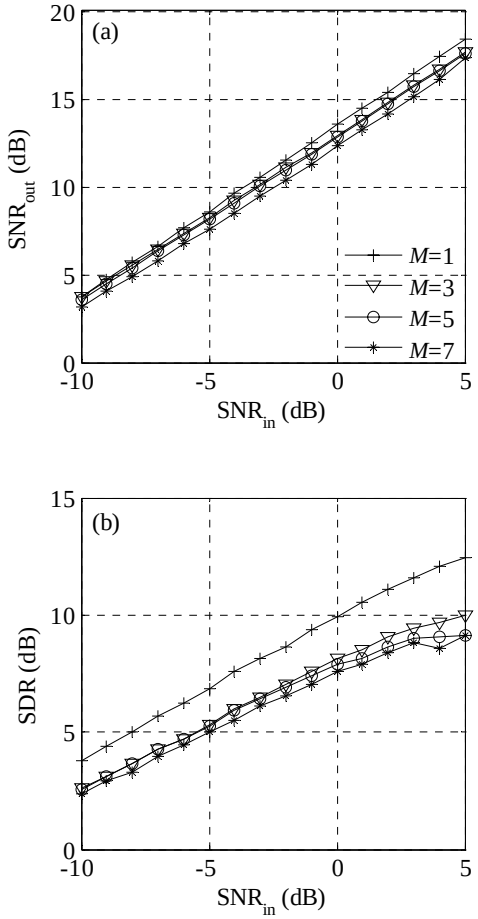


Figure 5-6. Plots of the performance measures for the FF/SU DAP algorithm various channel model orders (M) obtained from Monte-Carlo trials. (a) The output SNR and (b) the SDR performance measures as a function of the input SNR averaged over the 100 Monte-Carlo trials for each input SNR.

The time domain plots of the enhanced vocalization signals obtained from the three DAP algorithms for a single realization of the Monte-Carlo simulations is shown in Figure 5-7 for an input SNR of 0 dB with extraneous noise power at -15 dB. Only three paths are assumed for modeling the pseudo channel impulse responses in generating these plots.

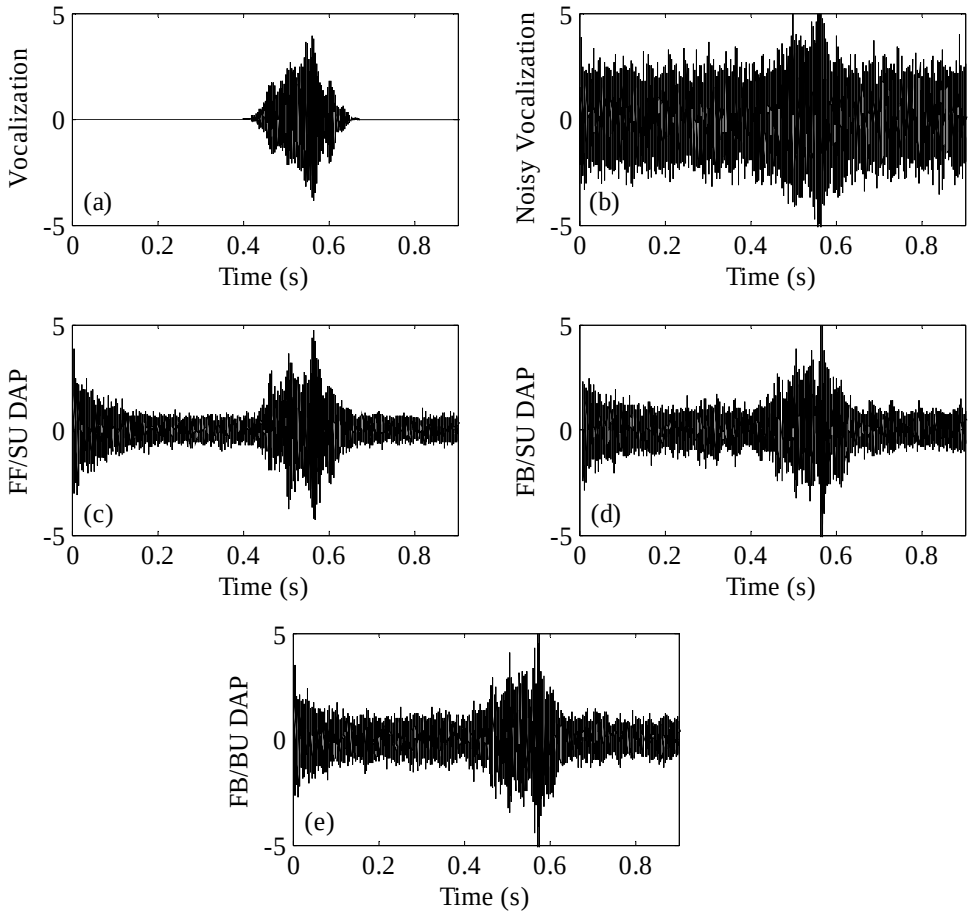


Figure 5-7. Time domain outputs of the DAP algorithms for a single Monte-Carlo trial. The input SNR and the extraneous noise level are set to 0 dB and -15 dB, respectively.

5.6 Summary

Three different implementations of the affine projection based BSS algorithm are developed in this chapter. These algorithms are implemented in a TITO configuration in the subsequent chapters. Hence, a review of the important mixing properties of the TITO

channel is presented. Numerical Monte-Carlo simulations are performed to evaluate the signal enhancement performances of the three proposed algorithms. These simulations reveal that the FF/SU DAP algorithm outperforms the feedback structured implementations.

Next, the performance of the FF/SU DAP algorithm is evaluated in the presence of extraneous noise with varying signal powers. BSS algorithms in general cannot suppress extraneous or diffuse noise, and as expected, the performance of the FF/SU DAP algorithm is shown to decrease with increasing extraneous noise power. The effect of the order of the FF/SU DAP algorithm on the signal enhancement performance is also evaluated. The order K represents a trade-off between computational complexity and performance; as the order is increased, the signal enhancement performance of the algorithm increases, as well as the computational complexity. An order of $K = 10$ is found to provide a suitable compromise between performance and computational complexity, and is adopted in the next chapter for simulations with real data.

Finally, the signal enhancement performance is investigated as the number of arrivals assumed in analytically computed pseudo channel impulse responses is varied. Increasing the number of arrivals increases the order of the FIR filter model of the channel and the cross-coupling between the source signals. Although the performance of the FF/SU DAP algorithm drops slightly with increasing model complexity, the algorithm is shown to provide a relatively robust performance up to 7 dominant arrival paths (which is typical of what is observed in real channels, see experimentally estimated channel impulse response plots in Chapter 6).

CHAPTER 6: EXPERIMENTAL RESULTS

In this chapter, the proposed wavelet domain denoising (WDD) and blind source separation (BSS) algorithms are tested with real vocalization and underwater noise recordings. The denoising and signal enhancement performances of the tested algorithms are evaluated in terms of the signal to noise ratio (SNR) and signal to distortion ratio (SDR) performance measures. The chapter is concluded with some remarks and observations regarding the experiments presented in this chapter.

6.1 Experimental Setup

The denoising and signal enhancement performances of the proposed WDD and BSS algorithms are estimated by Monte-Carlo simulations in Chapter 4 and Chapter 5. However, the real underwater acoustic environment is significantly more challenging. First, the test signals used in generating the Monte-Carlo simulations are idealized test signals with well defined statistical properties. In reality, both manatee vocalization signals and other acoustic noise signals have some statistical, temporal and spectral components that will deviate from the models, which may degrade the signal enhancement performance of the proposed algorithms. In addition to non-ideal signals, the acoustic channel will also have dynamics that cannot be modeled accurately with a

finite impulse response (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. Finally, 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. Both these factors may result in a complex noise spectrum and deviation in the noise signal statistics. Most importantly, the coherence of the signals measured at different receivers will be reduced, which may affect the convergence and signal enhancement performance of the proposed BSS algorithms. Thus, it is essential to evaluate the proposed algorithms under real environments to be able to fully comprehend their denoising and signal enhancement performances. In the following sections, the performances of the WDD and BSS algorithms are evaluated with real vocalization and noise data recorded at Crystal River, FL. This test location is at close proximity to known manatee habitats and a busy waterway, thus represent a pilot site where a manatee vocalization detector may potentially be implemented. Therefore, the results presented here are indicators of how these algorithms are expected to perform in-field. A detailed description of the test site and the experimental setup is provided in Appendix C.

6.2 Wavelet Domain Denoising

In this section, the denoising performance of the adaptive wavelet packet transform (AWPT) algorithm is evaluated using real vocalization and noise signals. In general, it is very difficult, if not impossible, to conduct a controlled experiment in the presence of both a manatee and a watercraft. Instead, the proposed AWPT algorithm and other approaches discussed in Chapter 4 are evaluated using recordings constructed by superposing actual vocalizations recorded previously on background noise signals measured at the test site. This approach is physically justified since small amplitude acoustic waves are additive. The raw underwater acoustic time data is first filtered with a 10th order Butterworth highpass filter (with a 1 kHz cutoff frequency) before superposition to eliminate low frequency, high energy noise due to pressure waves, water current and other natural noise sources. Highpass filtering also eliminates some low frequency and high energy harmonics due to mechanical shaft rotation which do not spectrally overlap with vocalization harmonics. The input SNR is adjusted by changing the signal power of the vocalizations and is computed from these highpass filtered signals. Thus, the reported performance improvements in this section are in addition to what can be achieved by highpass filtering alone. For input SNR below -10 dB after highpass 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 signal enhancement 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 of the AWPT and the FourierShrink algorithms implemented in the Monte-Carlo trials presented in Chapter 4 are not changed for the following experiments. In addition to the methods proposed in Chapter 4, the feedback adaptive line enhancer (FALE) is implemented as a benchmark. The parameters used to implemented the three algorithms, along with some observations regarding the performance of the algorithms are summarized in Table 6-1.

Table 6-1. The parameters used in the implementation of the WDD algorithms.

Algorithm	Algorithm Specific Parameters	Remarks
AWPT	$J = 7$ $\mu = 0.005, \lambda = 1.0$	May exhibit poor noise suppression immediately after a vocalization
FourierShrink	128 frequency bins	Sensitive to non-vocalization harmonics
FALE	$\Delta = 50, L = 20$ $\mu = 0.001, \beta = 0.5$	Disproportional enhancement of harmonics distorts vocalization waveform

The vocalizations employed for these tests were obtained from the vocalization library published by Yan [28]. The vocalization and noise signals are sampled at 24 kHz. The background noise recordings for four test cases recorded in Crystal River, FL which represents different watercraft approach speeds are plotted in Figures 6-1 and 6-2. Most of these noise signals are non-stationary over the duration of the vocalizations (due to an approaching surface vessel and biological noise) and exhibit temporal correlation (due to

the convolutive and frequency selective transmission channel) with some narrowband non-vocalization harmonics resulting from mechanical shaft rotation which could not be suppressed with highpass filtering. Such traits make denoising vocalizations corrupted with these noise signals particularly challenging.

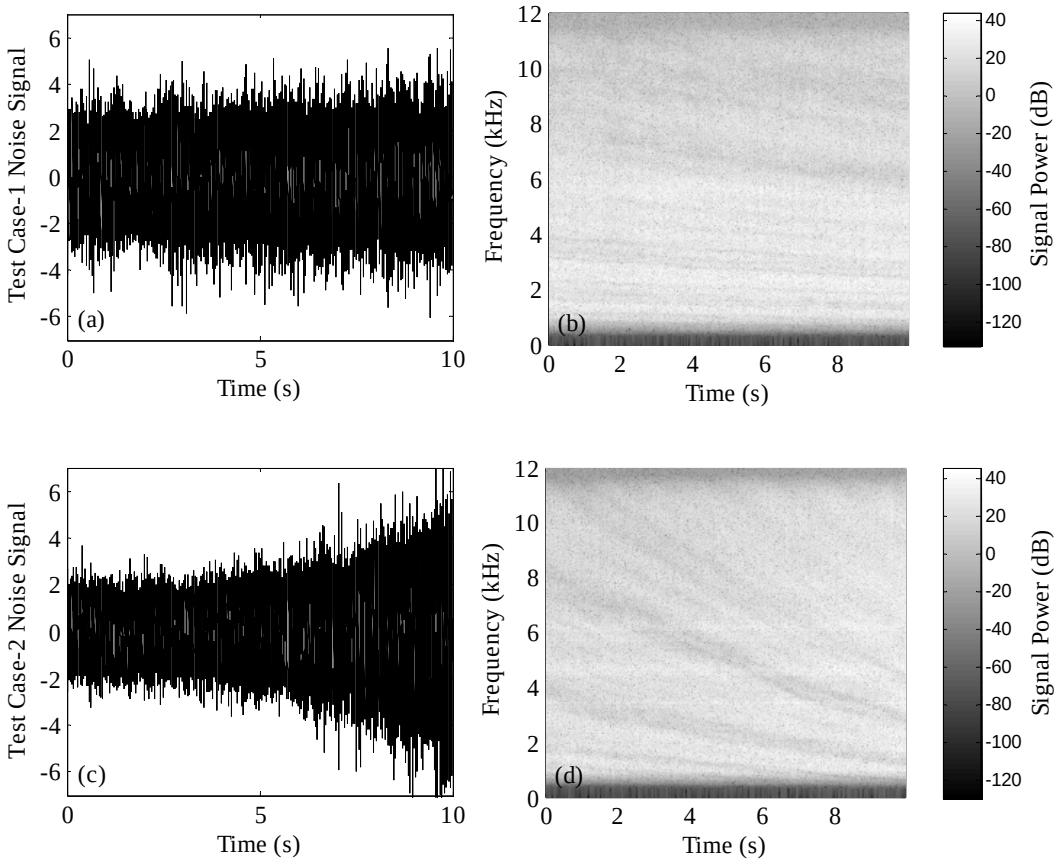


Figure 6-1. The noise measurements for the first and second Crystal River test cases used to evaluate the WDD algorithms. Time domain plots (a), (c) and the spectra (b), (d) of watercraft emitted noise after highpass filtering.

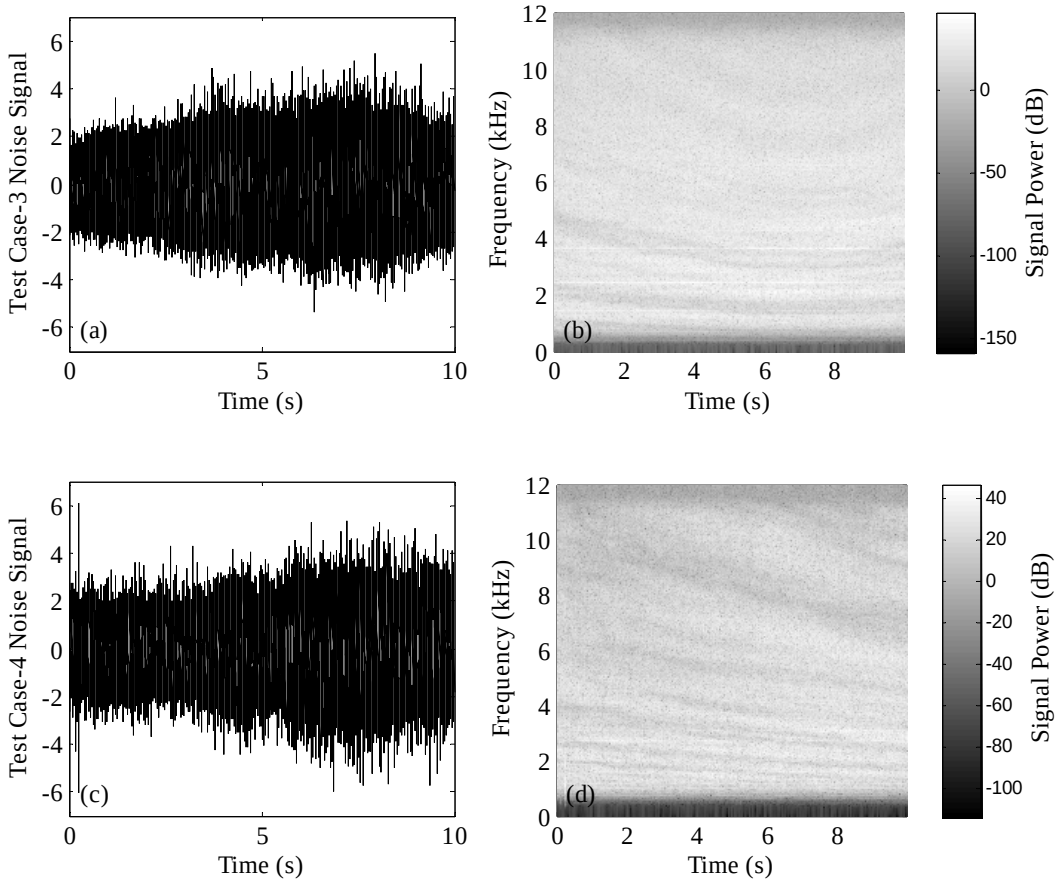


Figure 6-2. The noise measurements for the third and fourth Crystal River test cases used to evaluate the WDD algorithms. Time domain plots (a), (c) and the spectra (b), (d) of watercraft emitted noise after highpass filtering.

The output SNR averaged over the 10 vocalizations within each category of the vocalization library is depicted in Figures 6-3 to 6-6.

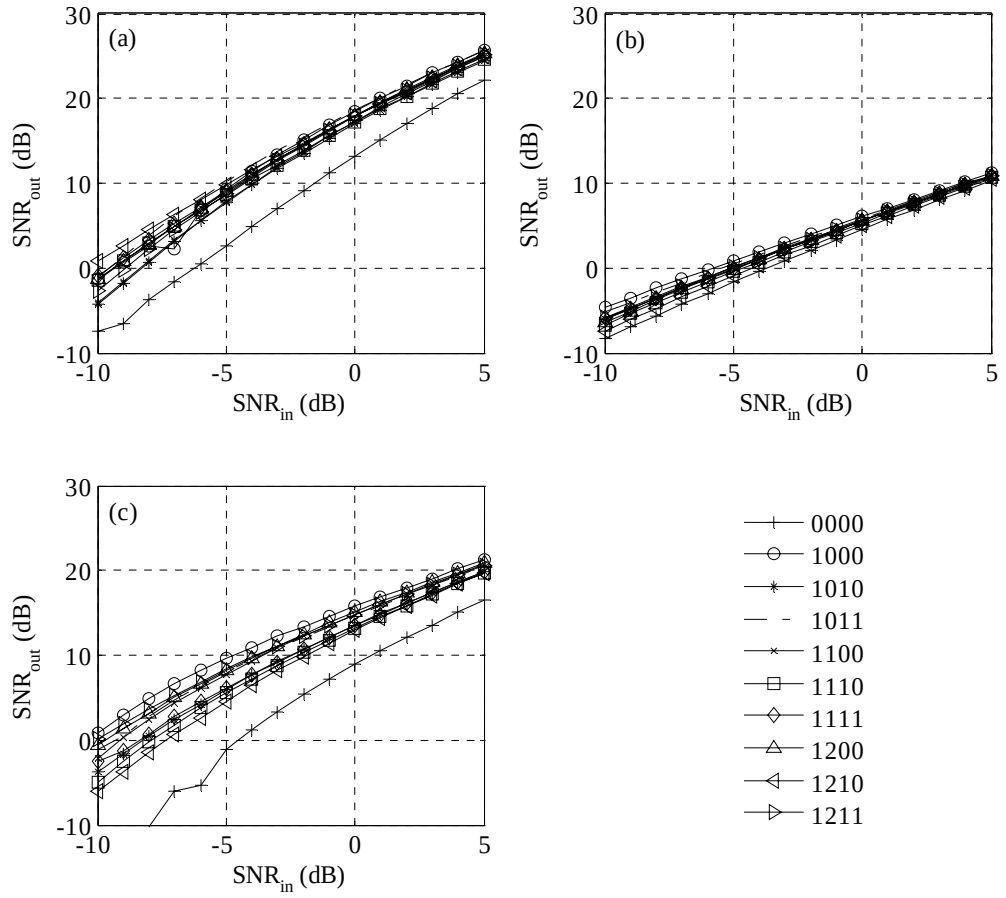


Figure 6-3. The output SNR obtained from the WDD methods for the first Crystal River test case. For (a) AWPT, (b) FourierShrink, and (c) FALE algorithms as a function of the input SNR and vocalization categories.

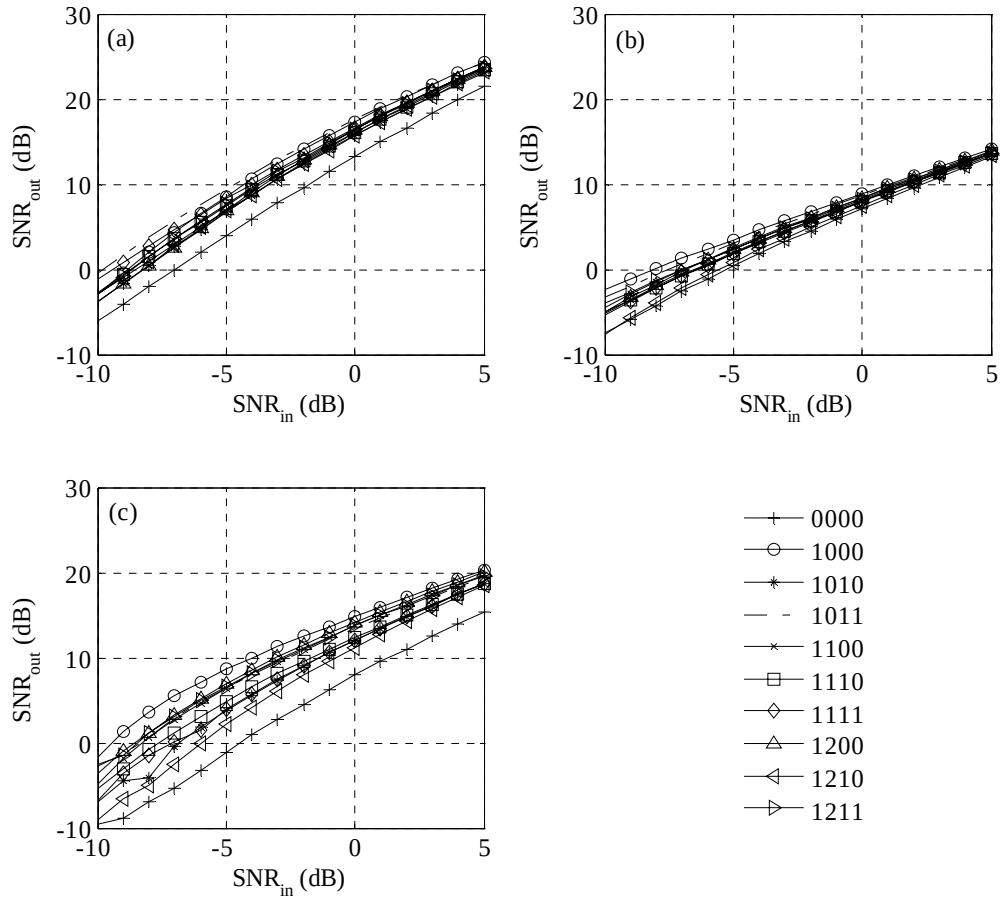


Figure 6-4. The output SNR obtained from the WDD methods for the second Crystal River test case. For (a) AWPT, (b) FourierShrink, and (c) FALE algorithms as a function of the input SNR and vocalization categories.

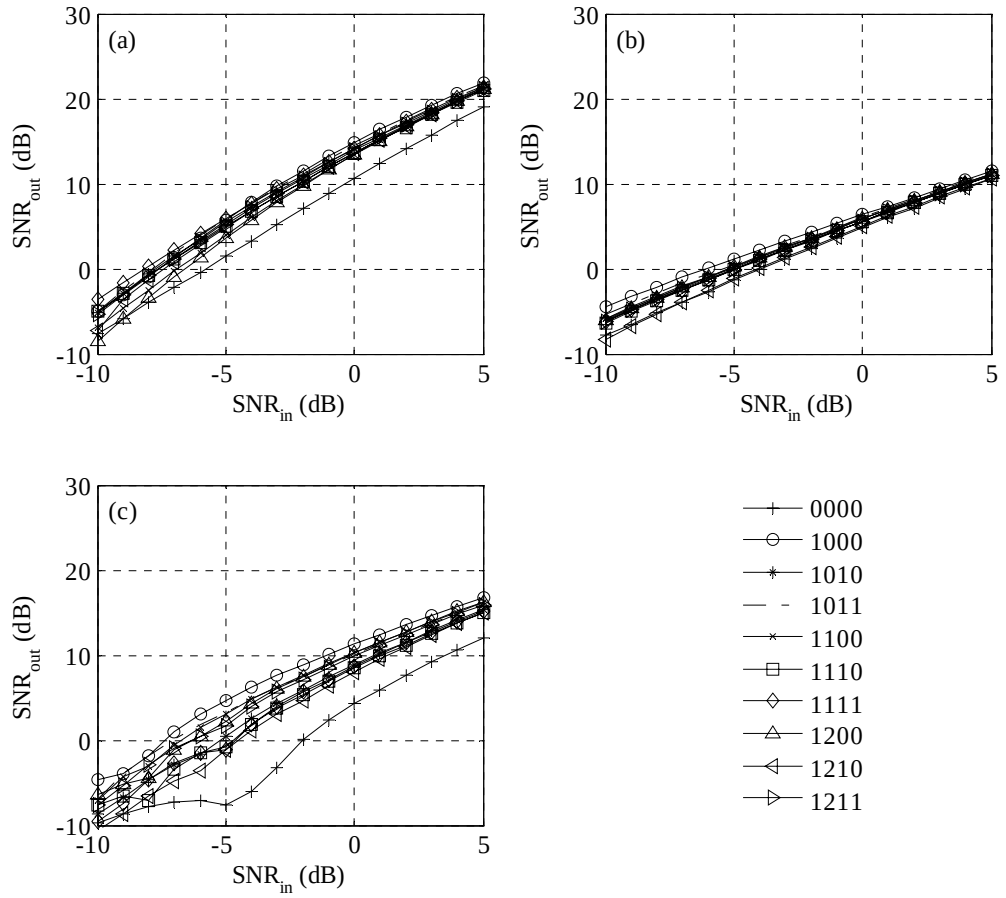


Figure 6-5. The output SNR obtained from the WDD methods for the third Crystal River test case. For (a) AWPT, (b) FourierShrink, and (c) FALE algorithms as a function of the input SNR and vocalization categories.

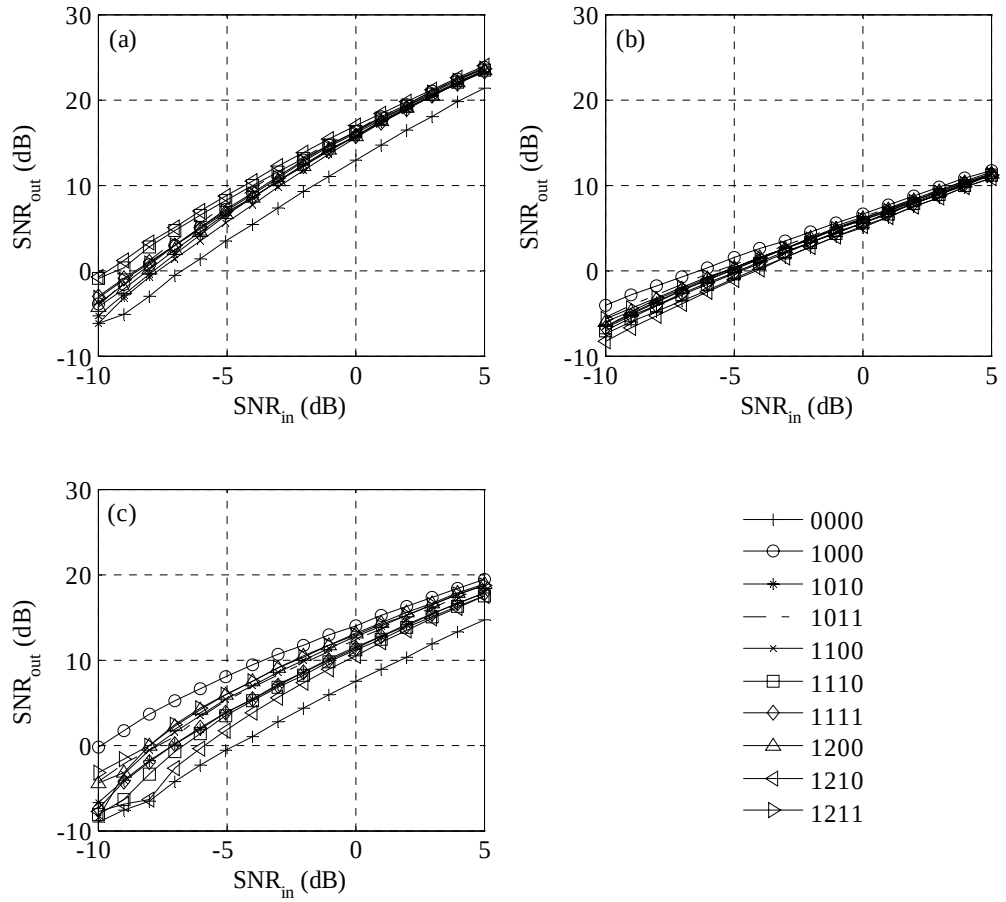


Figure 6-6. The output SNR obtained from the WDD methods for the fourth Crystal River test case. For (a) AWPT, (b) FourierShrink, and (c) FALE algorithms as a function of the input SNR and vocalization categories.

These results indicate that both the AWPT and FourierShrink algorithms achieve a more consistent performance for the vocalization categories and are more robust to different vocalization time-frequency structures compared to the FALE. The FourierShrink and FALE achieve 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. All three 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 [28]. The wavelet transform is more appropriate for compactly representing such signals and the SNR performance of the AWPT algorithm is better for the 0000 type vocalization signal compared to the FourierShrink and FALE algorithms for all four test cases. On average, the AWPT algorithm achieves an improved noise suppression performance compared to both the FourierShrink and FALE algorithms. This is illustrated in Figure 6-7 where the output SNR performance of the three algorithms averaged over all the vocalizations in the 10 categories are plotted for each test case.

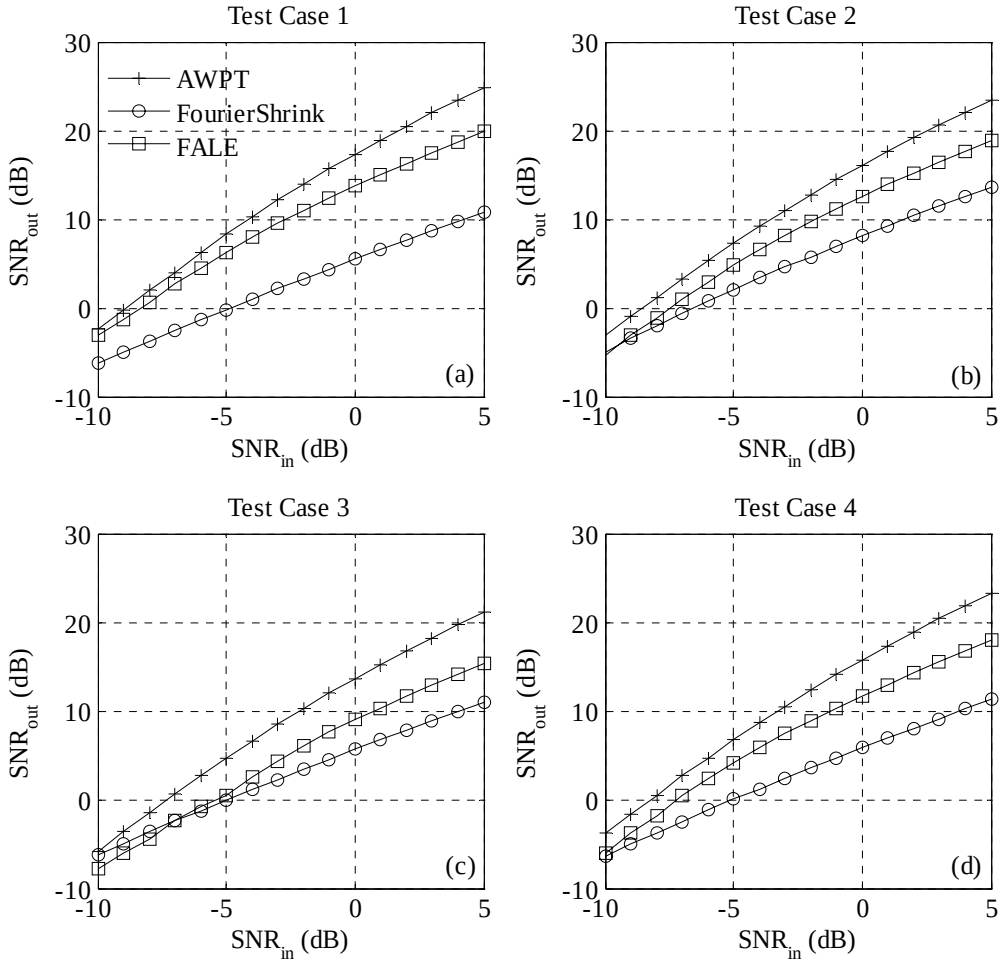


Figure 6-7. The output SNR obtained from the WDD methods averaged over all vocalization categories for the four Crystal River test cases. For (a) test case 1, (b) test case 2, (c) test case 3, and (d) test case 4.

The SDR averaged over the 10 vocalizations within each category of the vocalization library is depicted in Figures 6-8 to 6-11.

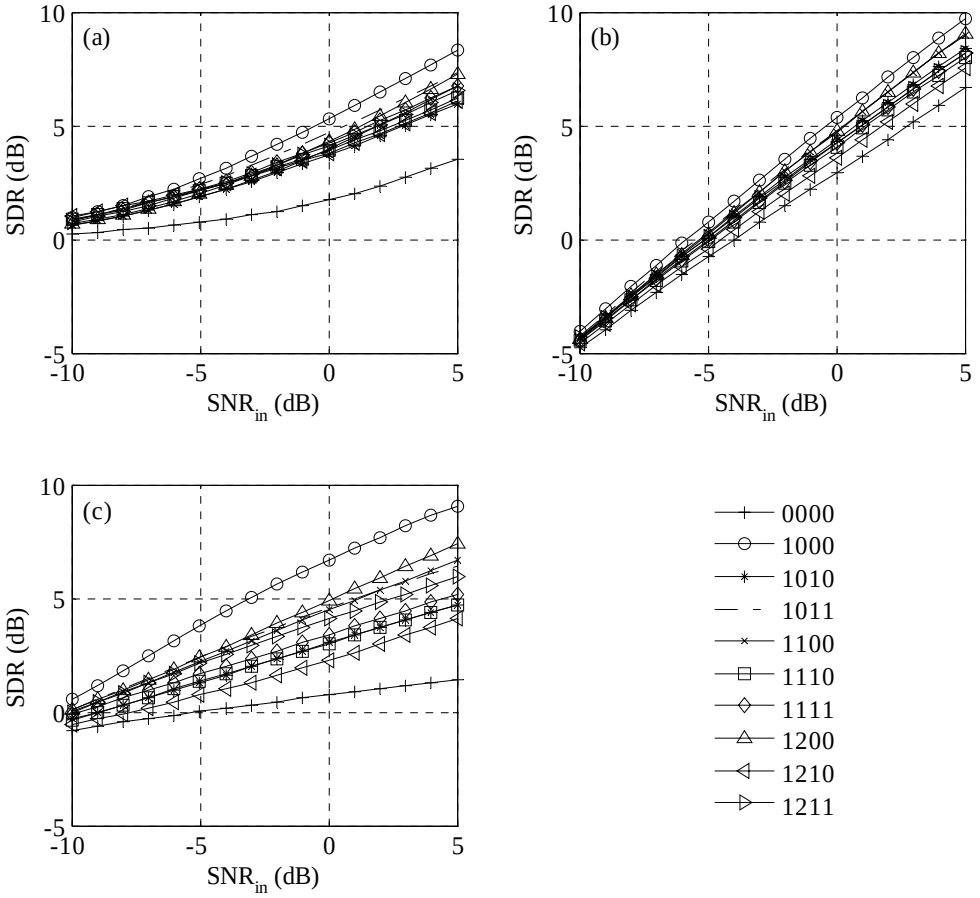


Figure 6-8. The SDR obtained from the WDD methods for the first Crystal River test case. For (a) AWPT, (b) FourierShrink, and (c) FALE algorithms as a function of the input SNR and vocalization categories.

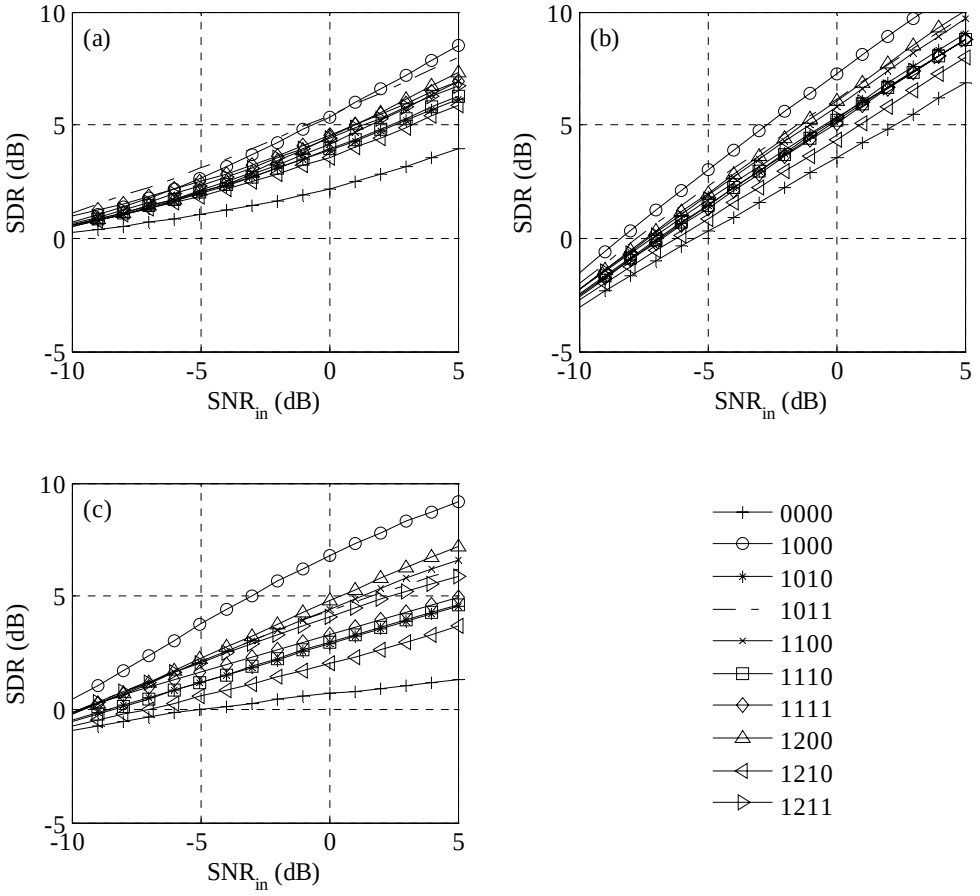


Figure 6-9. The SDR obtained from the WDD methods for the second Crystal River test case. For (a) AWPT, (b) FourierShrink, and (c) FALE algorithms as a function of the input SNR and vocalization categories.

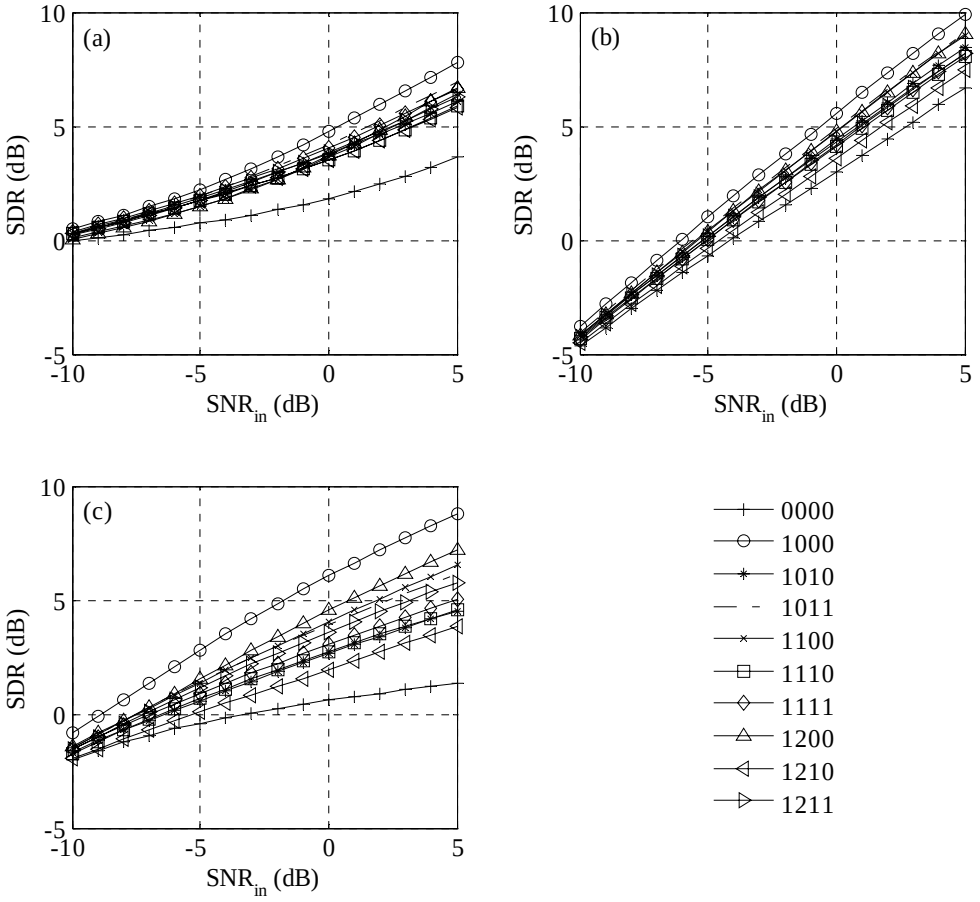


Figure 6-10. The SDR obtained from the WDD methods for the third Crystal River test case. For (a) AWPT, (b) FourierShrink, and (c) FALE algorithms as a function of the input SNR and vocalization categories.

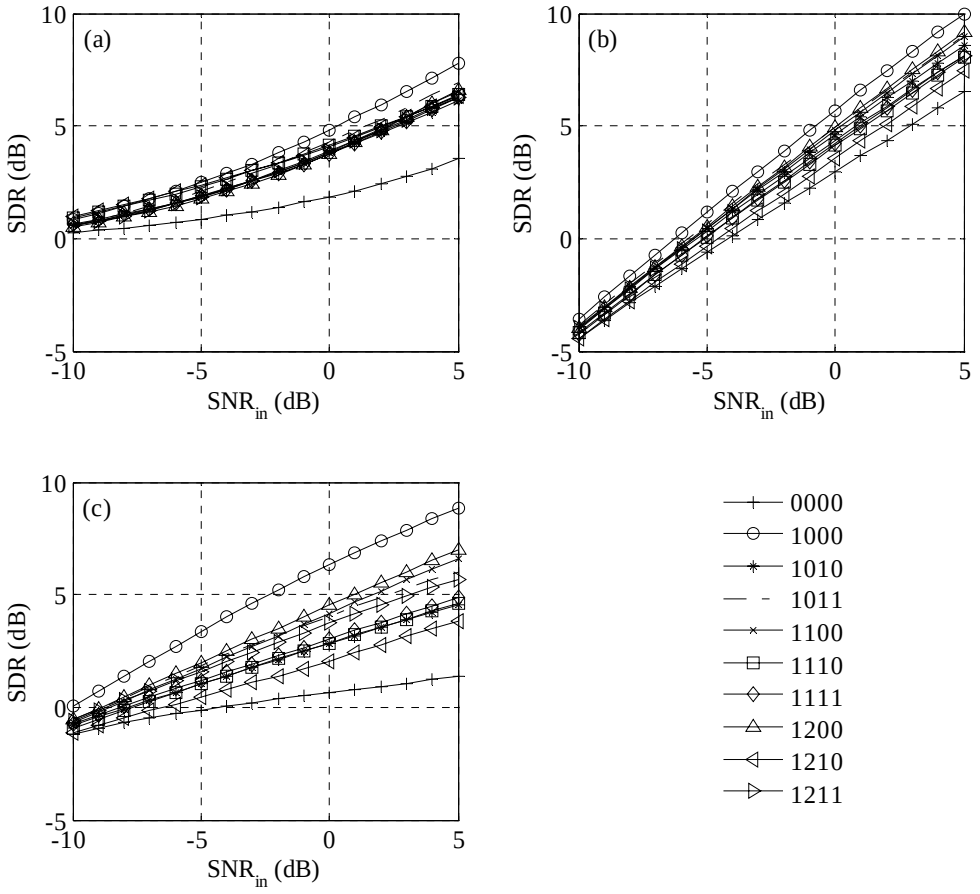


Figure 6-11. The SDR obtained from the WDD methods for the fourth Crystal River test case. For (a) AWPT, (b) FourierShrink, and (c) FALE algorithms as a function of the input SNR and vocalization categories.

As is in the case of output SNR, the AWPT and FourierShrink algorithms are more robust to variations in the vocalization signals time-frequency statistics. All the three tested algorithms achieved their best and worst performance for the 1000 and 0000 type vocalizations, respectively. Among the three tested methods, FourierShrink resulted in the worst SDR performance at low input SNR, in particular, below -5 dB. However,

the FourierShrink algorithm outperformed both the AWPT and FALE algorithms at higher input SNR values. As is the case for the output SNR performance measure, the AWPT algorithm consistently resulted in higher SDR performance over the entire range of input SNR values compared to the FALE algorithm. The SDR values averaged over the 100 vocalizations of the vocalization library are plotted in Figure 6-12.

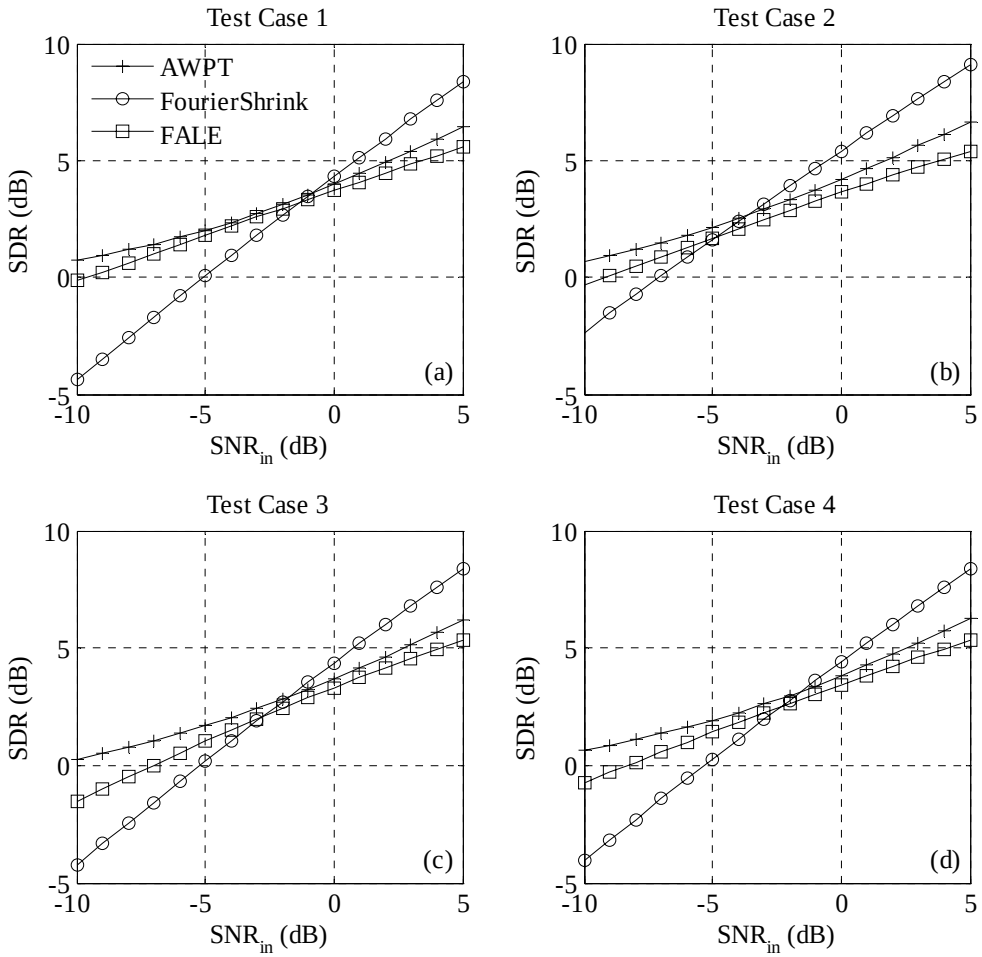


Figure 6-12. The SDR obtained from the WDD methods averaged over all vocalization categories for the four Crystal River test cases. For (a) test case 1, (b) test case 2, (c) test case 3, and (d) test case 4.

The plots of the time domain outputs of the AWPT, FourierShrink and the FALE algorithms are presented for the first five vocalizations of category 1200 at 0 dB input SNR for the four Crystal River test cases in Figures 6-13 to 6-16.

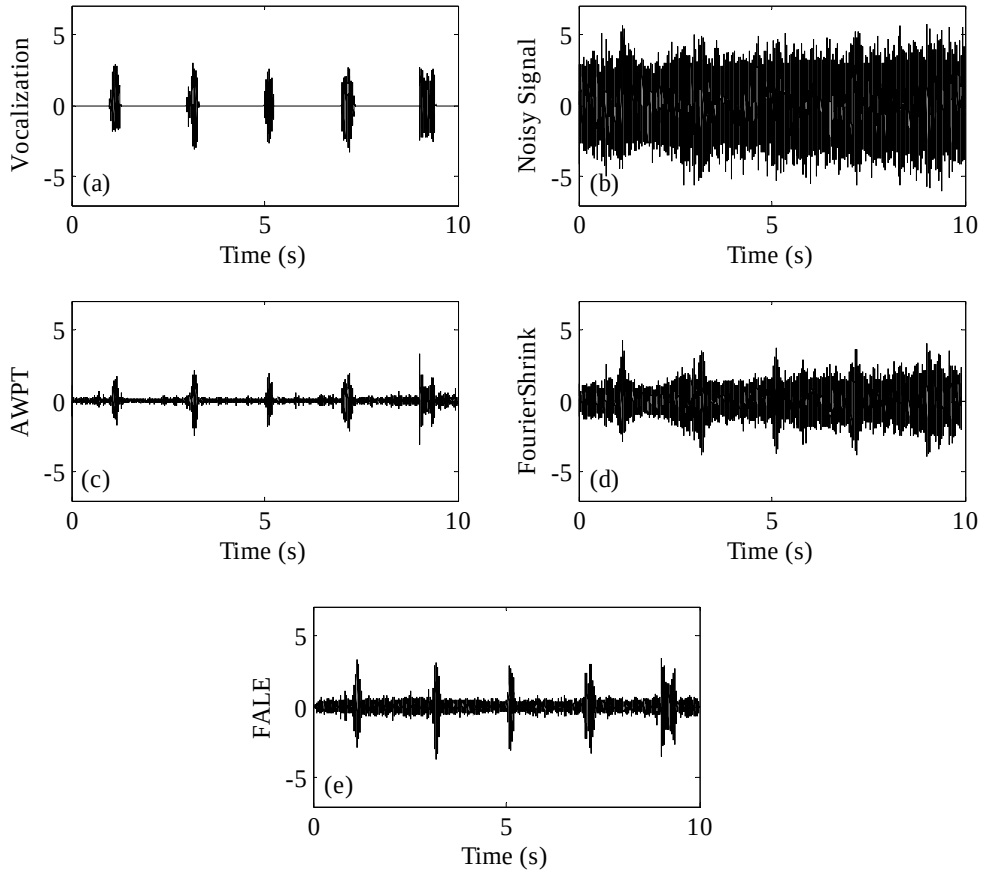


Figure 6-13. Time domain outputs of the denoising algorithms for the first Crystal River test case. Input SNR is set to 0 dB.

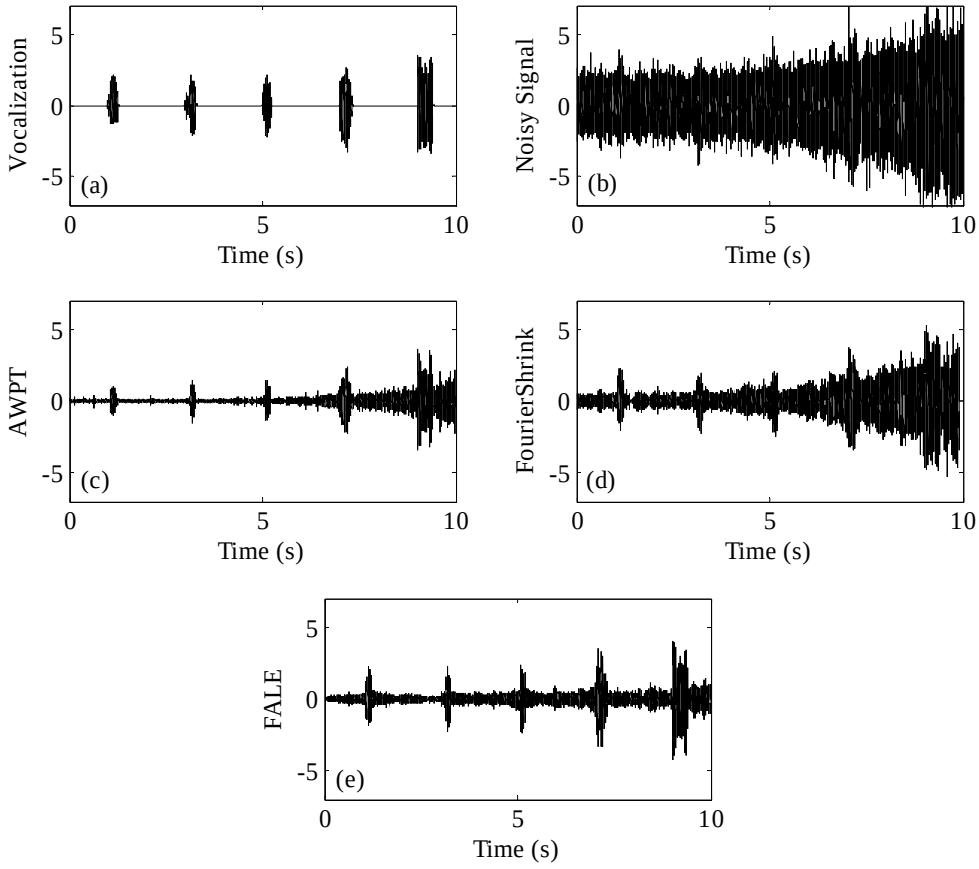


Figure 6-14. Time domain outputs of the denoising algorithms for the second Crystal River test case. Input SNR is set to 0 dB.

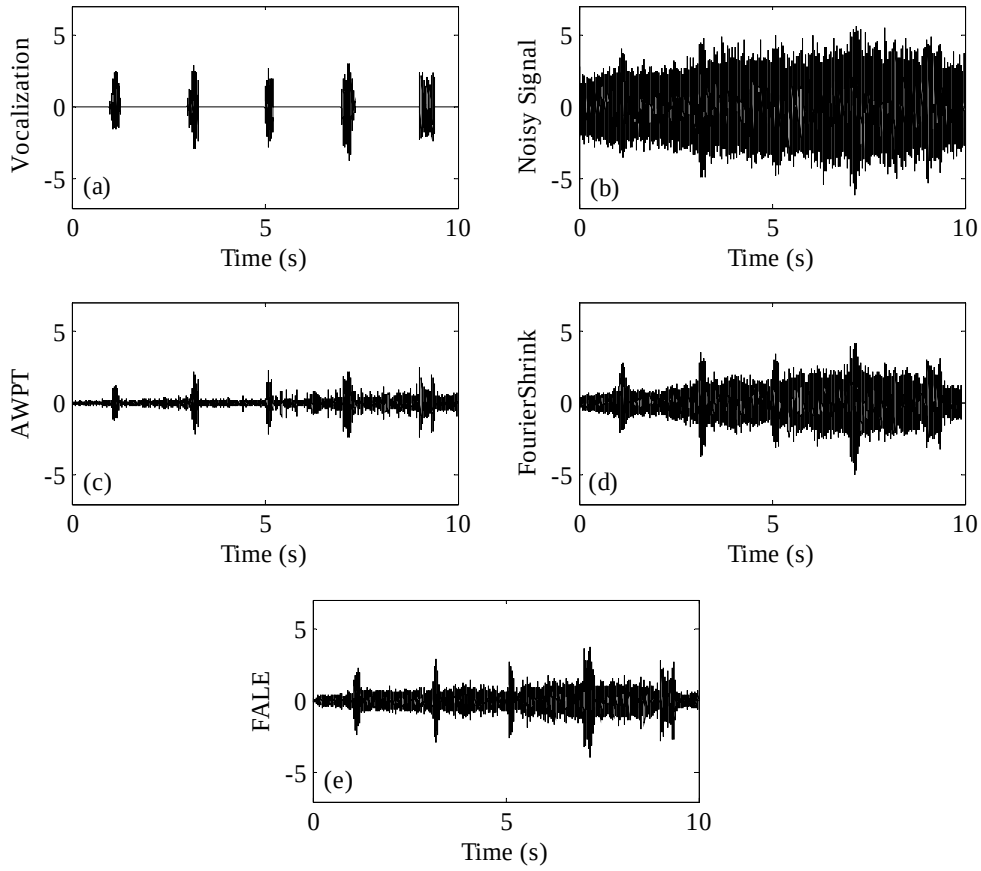


Figure 6-15. Time domain outputs of the denoising algorithms for the third Crystal River test case. Input SNR is set to 0 dB.

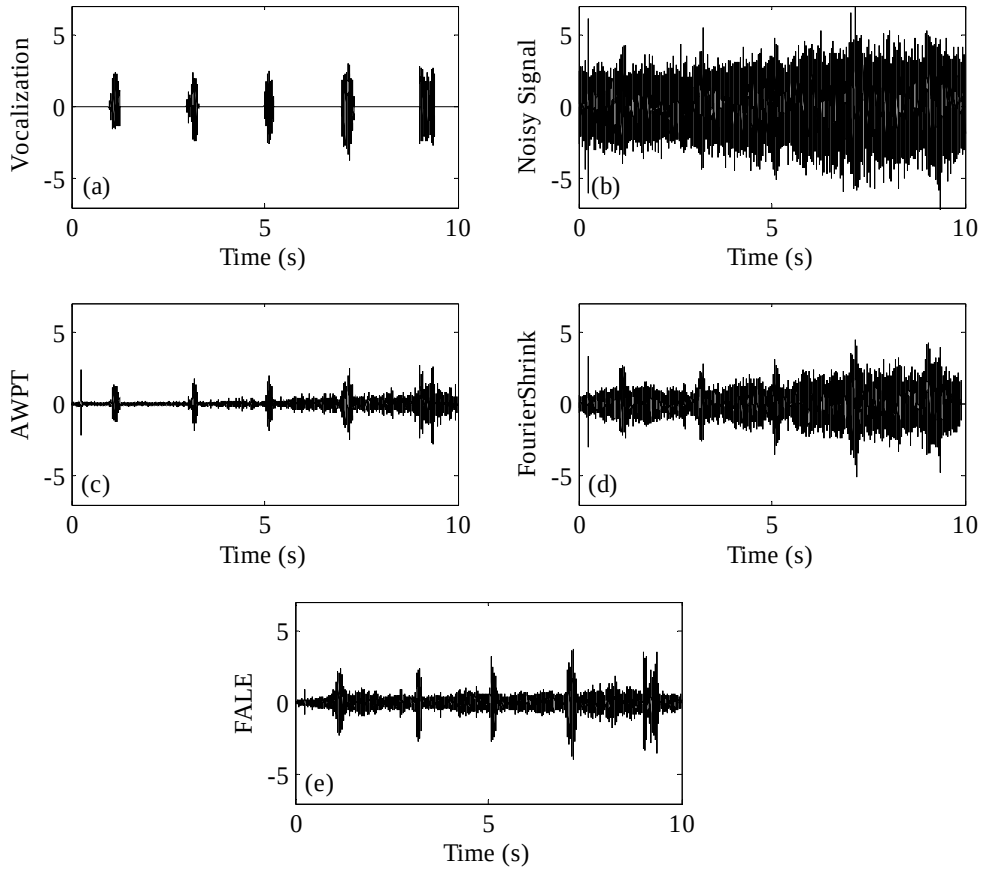


Figure 6-16. Time domain outputs of the denoising algorithms for the fourth Crystal River test case. Input SNR is set to 0 dB.

Qualitatively, AWPT and FALE provide better results compared to FourierShrink. In most of the noisy vocalization signals processed with the FourierShrink algorithm, the watercraft emitted noise is more audible in the background and can sometimes mask the vocalizations even at relatively high input SNR. In particular, the FourierShrink algorithm is not as effective in suppressing spurious noise peaks that result from the Lloyd mirror effect and other frequency selective phenomenon,

which results in a lower output SNR performance compared to the AWPT and FALE algorithms as is depicted in Figure 6-7. While the Fourier transform is in general more effective in compressing and sparsely representing manatee vocalizations, the discrete Fourier transform (DFT) based FourierShrink method is too sensitive to non-vocalization peaks and harmonics and thus results in an inferior performance. As the watercraft approaches closer to the receivers, the mechanical shaft rotation induced harmonics, which are not modeled as a part of watercraft emitted noise, become more significant and start to overlap with the vocalization harmonics. This increase in the signal power of the mechanical harmonics is the main cause for deterioration in the noise suppression performance in all three algorithms towards the end of the noise recordings.

6.3 Blind Source Separation

In this section, the signal enhancement performance of the feedforward structured, sequential update double affine projection algorithm is evaluated and compared to other selected algorithms using real vocalization and noise signals. Generating multi-channel test recordings for evaluating the performance of BSS algorithms is slightly more involved compared to the single channel approaches of the previous section. 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

demixing filters to the optimum separating solution. This method (based on the synthetic mixing of the source signals) is similar to the Monte-Carlo simulations used in Chapter 5 to evaluate the performance of the BSS algorithms with artificial vocalization and noise recordings. Alternatively, to obtain more realistic mixtures, the speech signals are recorded separately (but in the same environment) and are numerically superposed to create the noisy measurements [122]. The latter experimental setup is not possible 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 are convolved with actual underwater acoustic channel impulse responses and added to real background noise recordings. As was described in Section 3.1, BSS algorithms require the measurements at the multiple receivers to be coherent. Hence, the background noise recordings used to evaluate the BSS algorithms were selected based on the average inter-channel coherence. 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 more detailed description of the experimental setup and the procedure used in estimating the channel impulse responses are provided in Appendix C. The convolved vocalizations are added to the noise recordings such that the vocalization signal reference channel is different than the watercraft noise reference channel. This ensures that the channel transmission matrix is causal.

As described in Section 3.2, the fundamental harmonic of manatee vocalizations fall generally within the 2 to 5 kHz frequency band and harmonics can reach up to 18

kHz or higher. Thus, to effectively capture at least the first few harmonics, a sampling rate 12 kHz or higher is necessary. This relatively high sampling rate will generally result in very high order FIR filters for modeling the acoustic transmission between the point sources and the receivers. However, the multi-channel BSS algorithms (introduced in Chapter 5) achieve source separation by converging to the pseudo channel impulse responses which are computed relative to a reference channel. Hence, the pseudo transfer function between an acoustic source and the receivers (i.e., relative transfer function with respect to a reference receiver) is more relevant for multi-channel adaptive signal enhancement algorithms compared to the absolute transfer functions between each of the sources and receivers. The FIR filters that model the pseudo channel impulse responses are of significantly lower order compared to those that model the absolute transfer functions between the sources and the receivers, which enable low computational complexity algorithms that can be implemented in real-time. A typical impulse response obtained from the broadcast tests is presented in Figure 6-17 where the underwater speaker is placed 1 m away from the reference hydrophone.

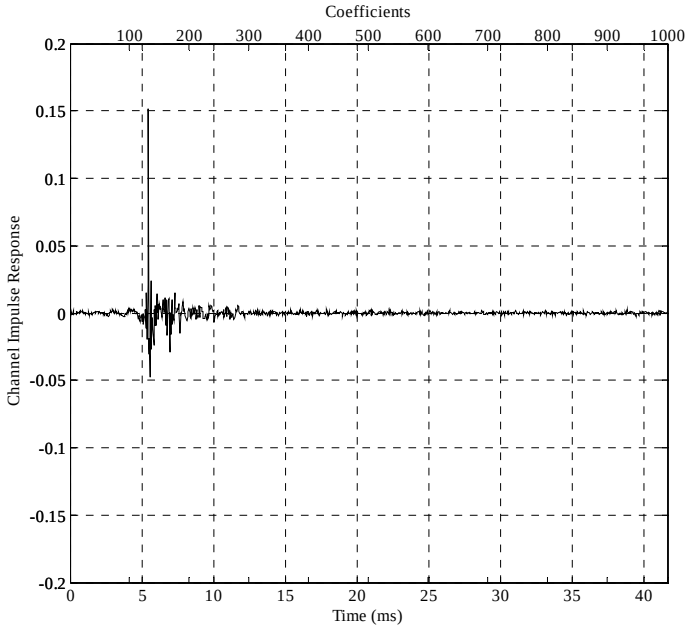


Figure 6-17. A typical channel impulse response estimated from broadband broadcast tests conducted in Crystal River, FL.

In general, the input SNR is a function of the vocalization and watercraft emitted noise source levels (SL) 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 [90], [24]. For consistency with the results provided in the previous section for WDD and 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 the pre-specified values.

The performances of the BSS algorithms are evaluated in a two-input two-output (TITO) setup which consists of a manatee and a watercraft as the only two active acoustic

point sources. Another important factor affecting the performance of the proposed BSS algorithms is the presence of extraneous noise (see Section 5.5). The ambient background noise levels were measured at the Crystal River test site when no point source was active (i.e., no watercraft in the vicinity). The ambient noise levels were determined to be significantly lower (-12 dB or lower) than typical watercraft emitted noise.

Since manatee vocalizations are generally 0.5 s or shorter in duration, four watercraft noise recordings, each of 1 s duration, were selected from the measurements (see Figures 6-18 to 6-21). In all four noise recordings, the watercraft was approaching the channel from the broadside (i.e., parallel to the main axis) of the linear hydrophone array. A diagram of the experimental setup is provided in Appendix C. The approach direction and relative speed of each watercraft were annotated during the recordings.

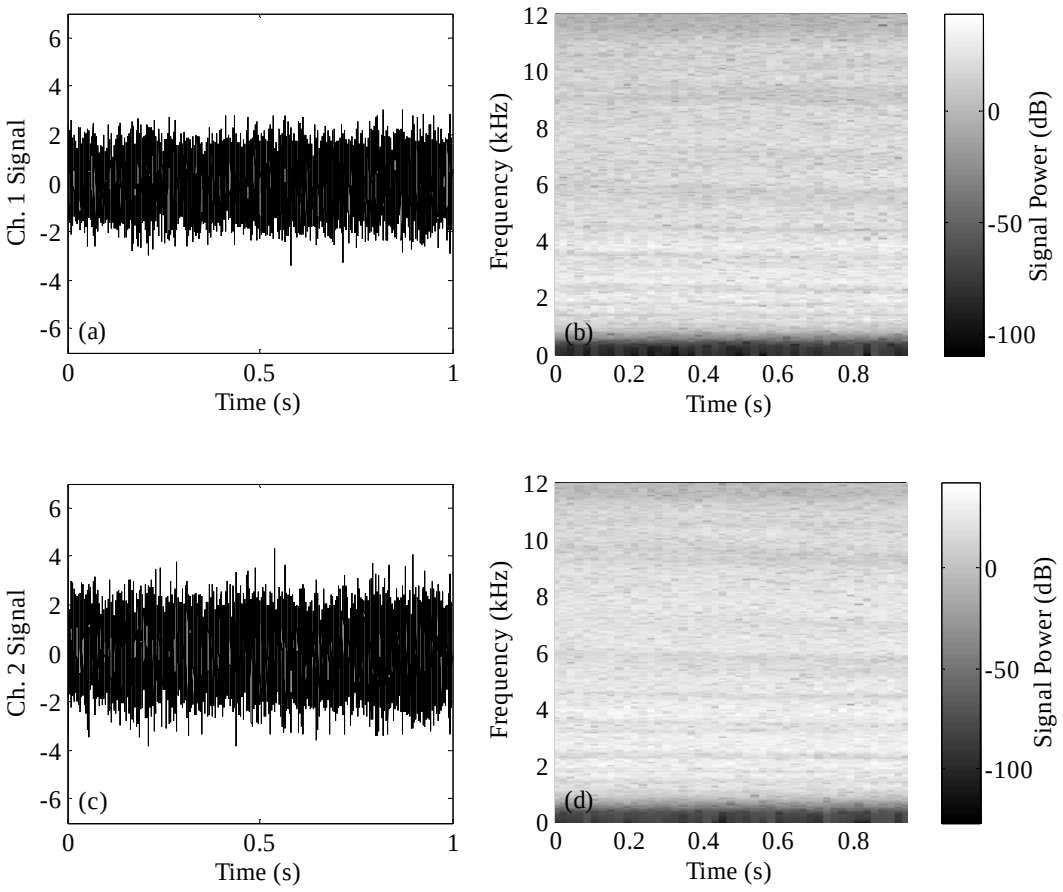


Figure 6-18. The noise measurements for the first Crystal River test case used to evaluate the BSS algorithms. Time domain plots (a), (c) and the spectra (b), (d) of the two channel watercraft emitted noise recordings after highpass filtering.

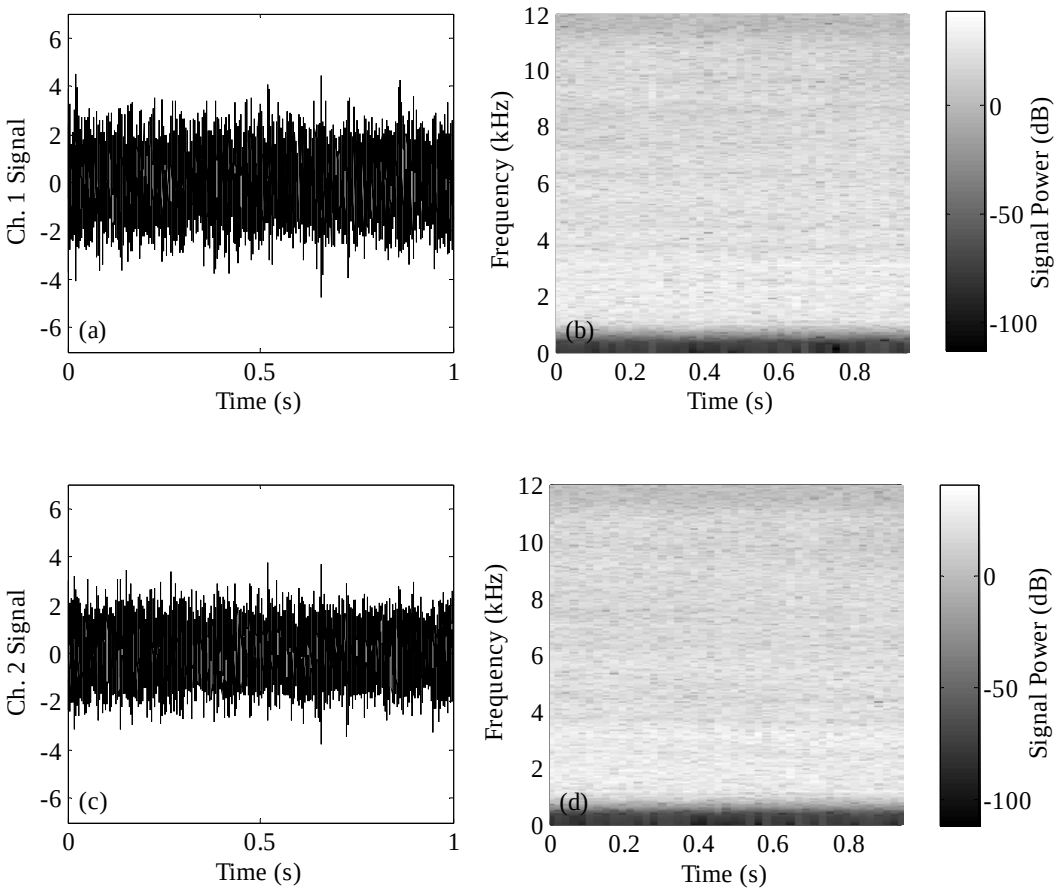


Figure 6-19. The noise measurements for the second Crystal River test case used to evaluate the BSS algorithms. Time domain plots (a), (c) and the spectra (b), (d) of the two channel watercraft emitted noise recordings after highpass filtering.

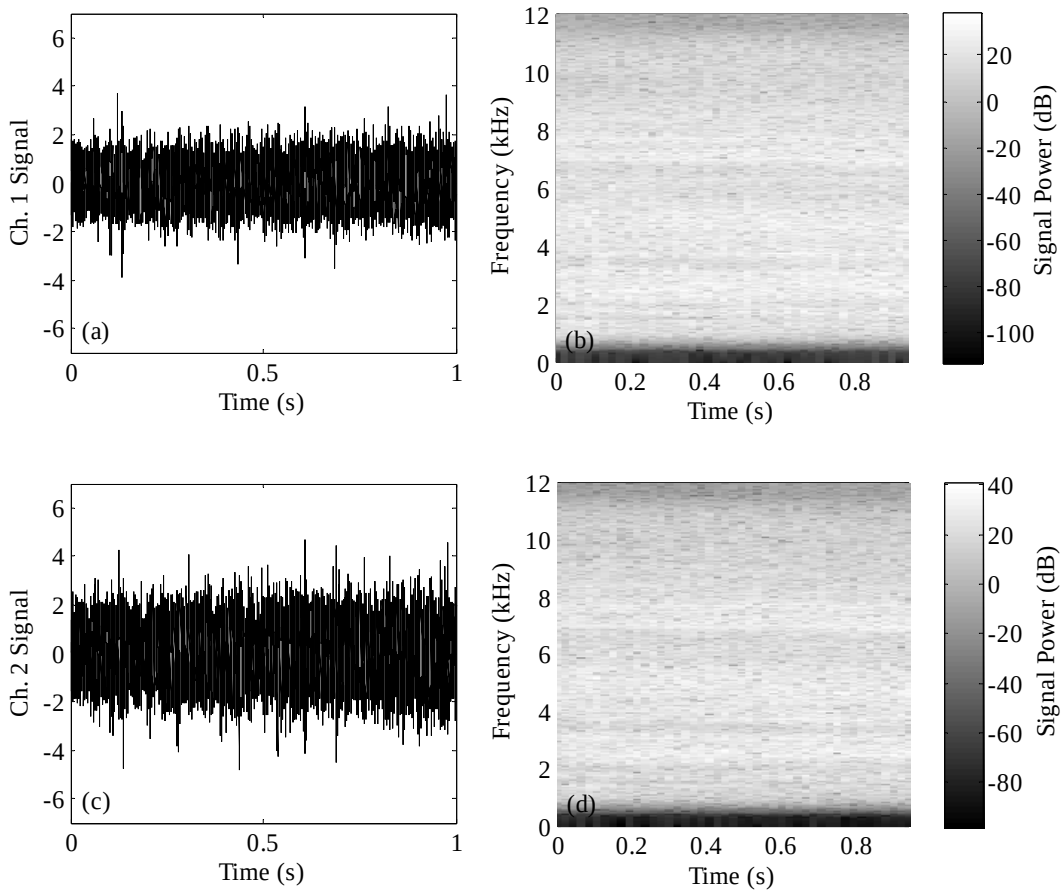


Figure 6-20. The noise measurements for the third Crystal River test case used to evaluate the BSS algorithms. Time domain plots (a), (c) and the spectra (b), (d) of the two channel watercraft emitted noise recordings after highpass filtering.

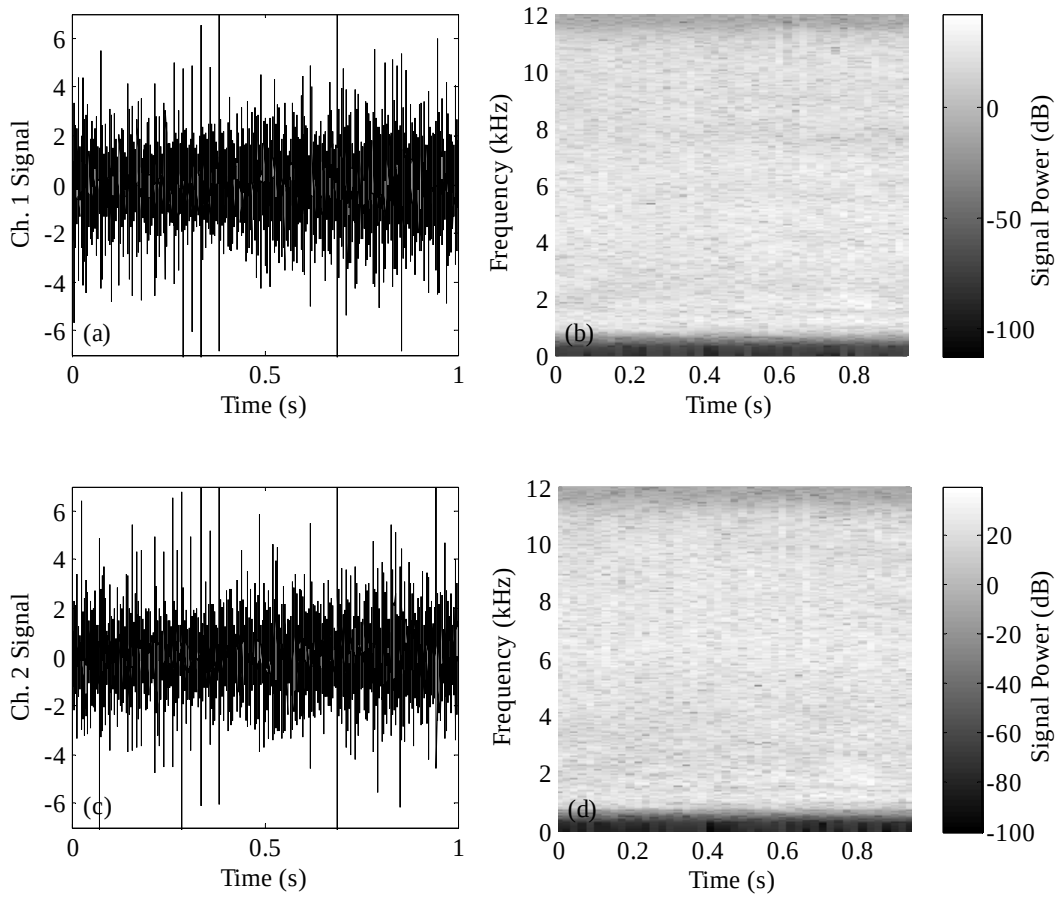


Figure 6-21. The noise measurements for the fourth Crystal River test case used to evaluate the BSS algorithms. Time domain plots (a), (c) and the spectra (b), (d) of the two channel watercraft emitted noise recordings after highpass filtering.

The parameters for the evaluated BSS algorithms are selected to ensure highest SDR performance while staying stable over the entire range of the input SNR. These parameters are provided in Table 6-2 for each algorithm along with some observations and remarks regarding their performance.

Table 6-2. The parameters used in the implementation of the BSS algorithms.

Algorithm	Filter Order (L)	Step Size (μ)	Algorithm Specific Parameters	Remarks
FF/SU DAP	200	0.01	$K = 10$	High computational complexity
FB/SU DAP	200	0.01	$K = 10$	High computational complexity
FB/BU DAP	200	0.1	$K = 10$	Erratic convergence filter weights
SAD	200	0.0001	N/A	Small step size necessary for stability, slow convergence
Parra	200	1.0	$T = 1200$ $N = 2400$ 50% overlap $\gamma = 0.5$	Slow convergence due to large block sizes (N) necessary to the avoid permuation ambiguity

As can be seen from Figure 6-17, a separation filter order of 200 is sufficient to capture the direct path arrival and the subsequent multi-path arrivals. A very small step size was necessary to keep the symmetric adaptive decorrelator (SAD) algorithm stable at high input SNR. The step size of the feedback, block-update double affine projection (FB/BU DAP) algorithm is selected a factor larger than the feedforward, sequential-update (FF/SU DAP) and the feedback, sequential-update (FB/SU DAP) implementations to ensure comparable convergence while keeping the algorithm stable. The cross-spectral density matrix (CSDM) of the measurements for Parra's algorithm is estimated from 2400 samples of data which corresponds to 0.1 s at a sampling rate of 24 kHz. The

vocalization signals and background noise are generally considered to be non-stationary over that time frame.

For the first set of tests, the performance of the various implementations of the proposed DAP algorithms are evaluated with the assumption that the manatee is 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 always 10 to 15 dB lower than the input SNR. It is worth re-iterating that the input SNR is measured post-highpass filtering. The performances of the FB/BU, the FF/BU, and the FF/SU implementations of the DAP algorithm are compared in terms of the output SNR and the SDR averaged over 10 randomly selected vocalizations (one vocalization for each category) from the vocalization library. The results are plotted in Figures 6-22 and 6-23, respectively, for each of the Crystal River test cases. The FF/SU DAP outperforms the FB/SU and FB/BU implementations for both the output SNR and SDR performance measures. The improved performance of the FF/SU DAP algorithm for both performance measures is consistent with the results obtained by Monte-Carlo trials in Chapter 5 and can be attributed to the completeness of the cross-correlation matrices used in updating the separation filters. Since the FF/SU DAP algorithm consistently outperforms the other affine projection based source separation algorithms, no further results are reported for the FB/SU DAP and FB/BU DAP algorithms.

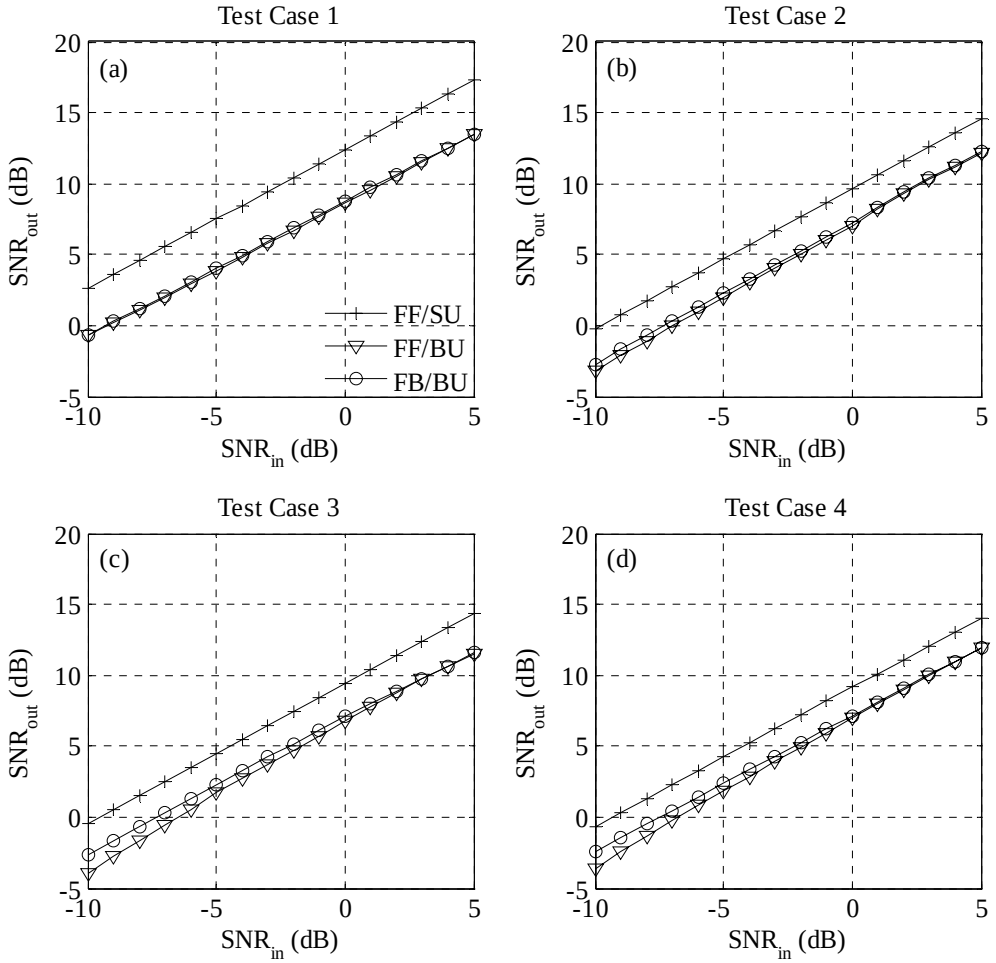


Figure 6-22. The output SNR obtained from the DAP algorithms for the four Crystal River test cases. The output SNR is averaged over the 10 vocalizations and plotted as a function of the input SNR for the FF/SU DAP, FB/SU DAP, and FB/BU DAP algorithms.

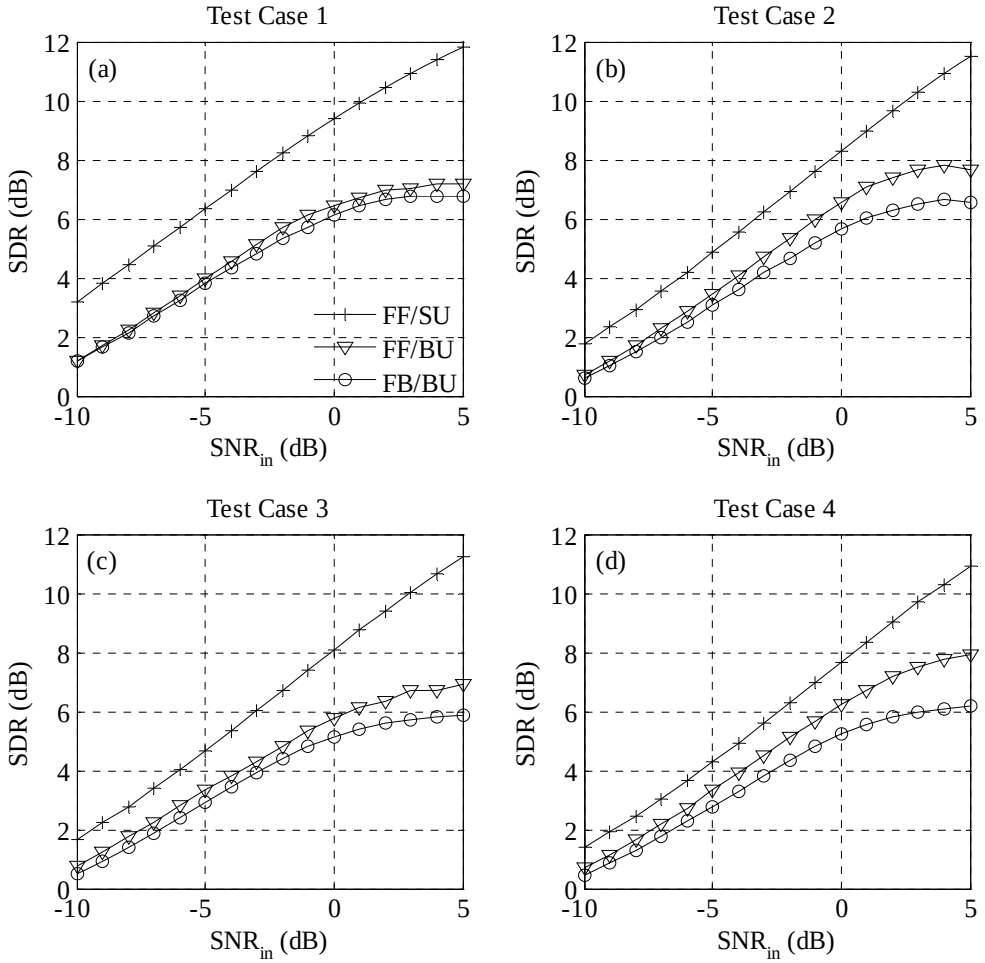


Figure 6-23. The SDR obtained from the DAP algorithms for the four Crystal River test cases. The SDR is averaged over the 10 vocalizations and plotted as a function of the input SNR for the FF/SU DAP, FB/SU DAP, and FB/BU DAP algorithms.

Next, the performance of the FF/SU DAP algorithm is compared to the second order statistics (SOS) based SAD and Parra algorithms. The output SNR and SDR results averaged over the same randomly selected 10 vocalizations of the vocalization library for the four Crystal River test cases are presented in Figures 6-24 and 6-25. Overall, the

FF/SU DAP algorithm results in the best output SNR and SDR performance measure for all four test cases, followed by the Parra algorithm. The output SNR and SDR performance of the SAD algorithm is lower than both the FF/SU DAP and Parra algorithms, most likely due to the very small step sizes required to keep the algorithm stable.

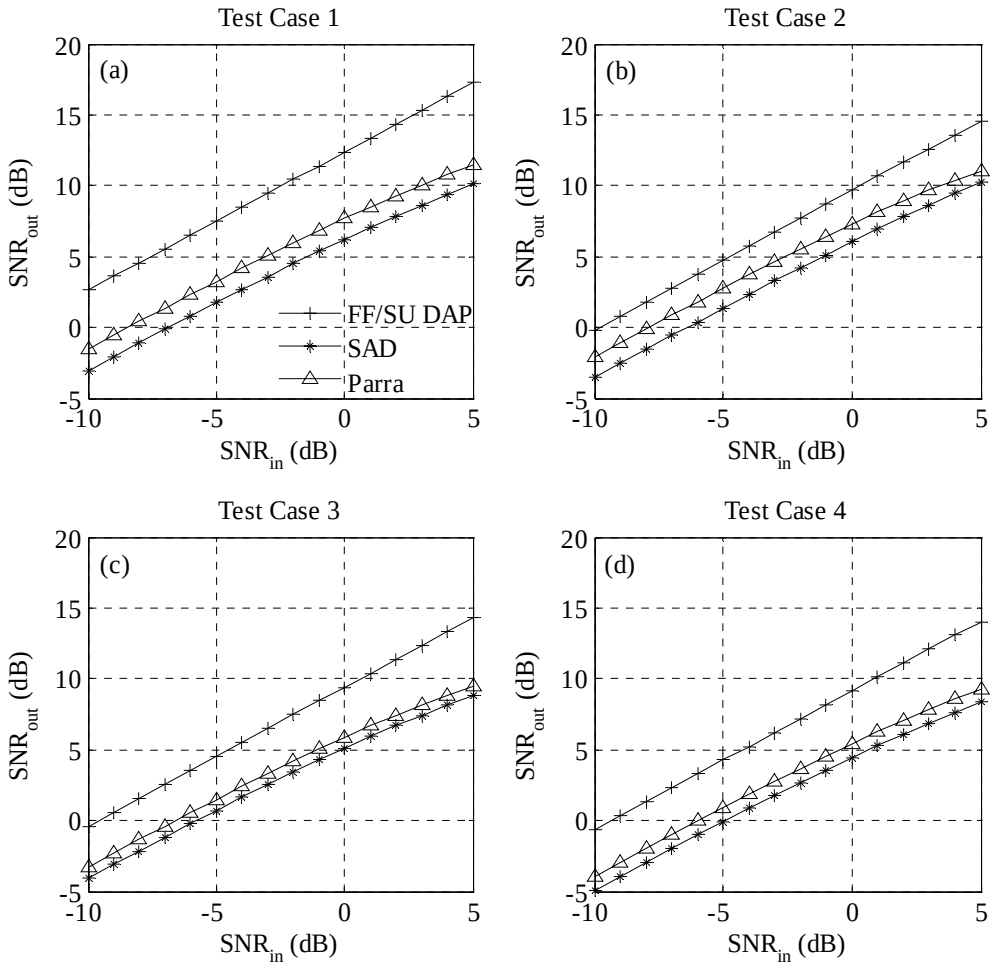


Figure 6-24. The output SNR obtained from the SOS-BSS algorithms for the four Crystal River test cases. The output SNR is averaged over the 10 vocalizations and plotted as a function of the input SNR for the FF/SU DAP, SAD, and Parra algorithms.

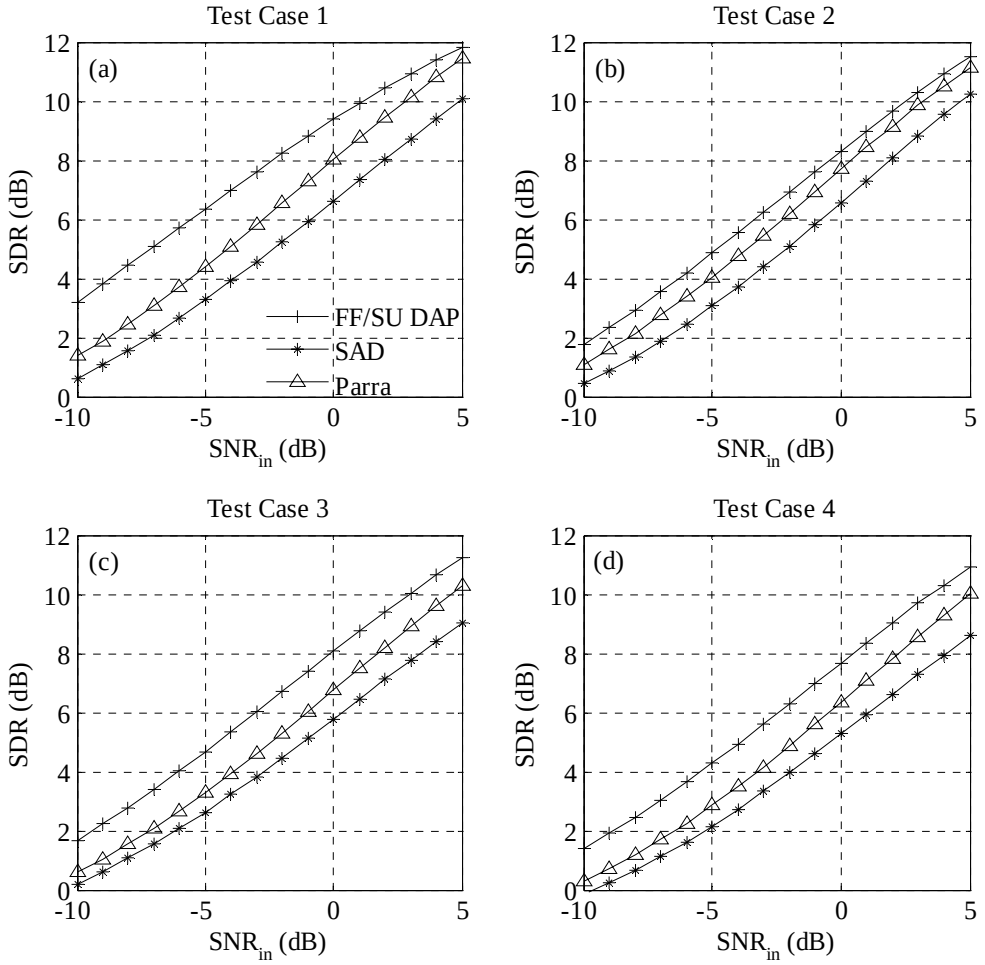


Figure 6-25. The SDR obtained from the SOS-BSS algorithms for the four Crystal River test cases. The SDR is averaged over the 10 vocalizations and plotted as a function of the input SNR for the FF/SU DAP, SAD, and Parra algorithms.

The plots of the time domain outputs of the FF/SU DAP, the SAD, and the Parra algorithms are presented for the randomly selected vocalization of the category 1200 at 0 dB input SNR for the four Crystal River test cases in Figures 6-26 to 6-29.

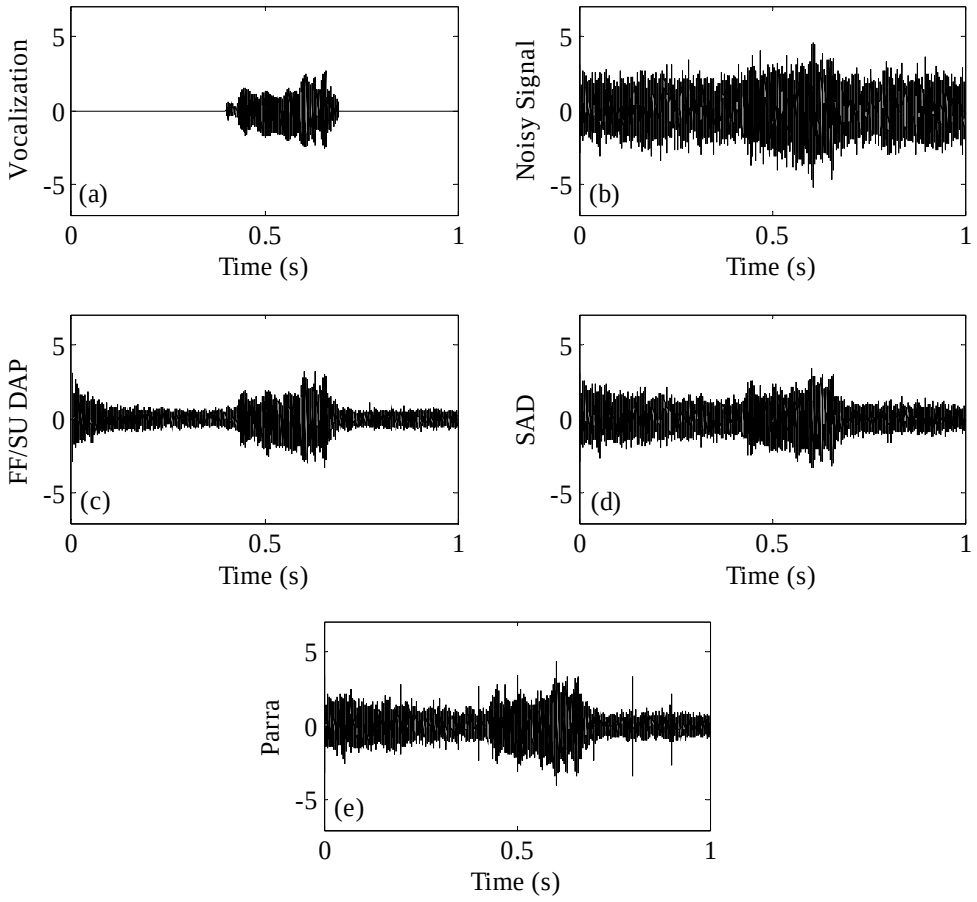


Figure 6-26. Time domain outputs of the SOS-BSS algorithms for the first Crystal River test case. Input SNR is set to 0 dB. Plots are for the channel 1 measurements.

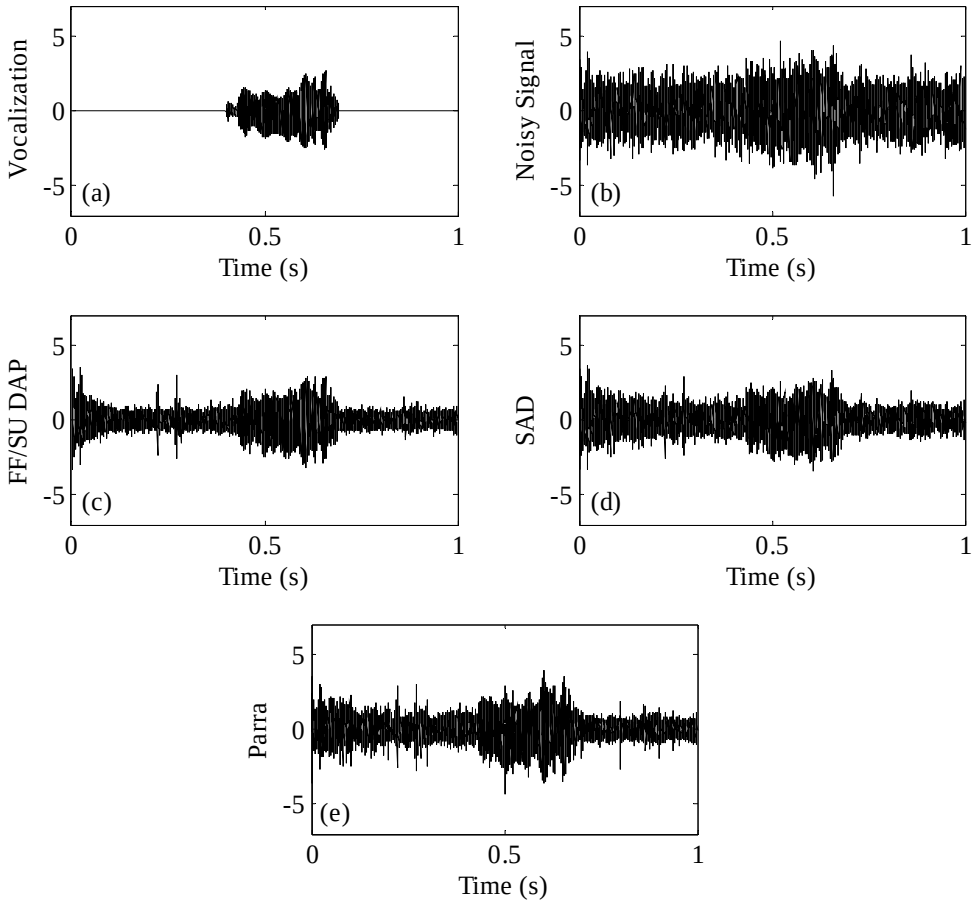


Figure 6-27. Time domain outputs of the SOS-BSS algorithms for the second Crystal River test case. Input SNR is set to 0 dB. Plots are for the channel 1 measurements.

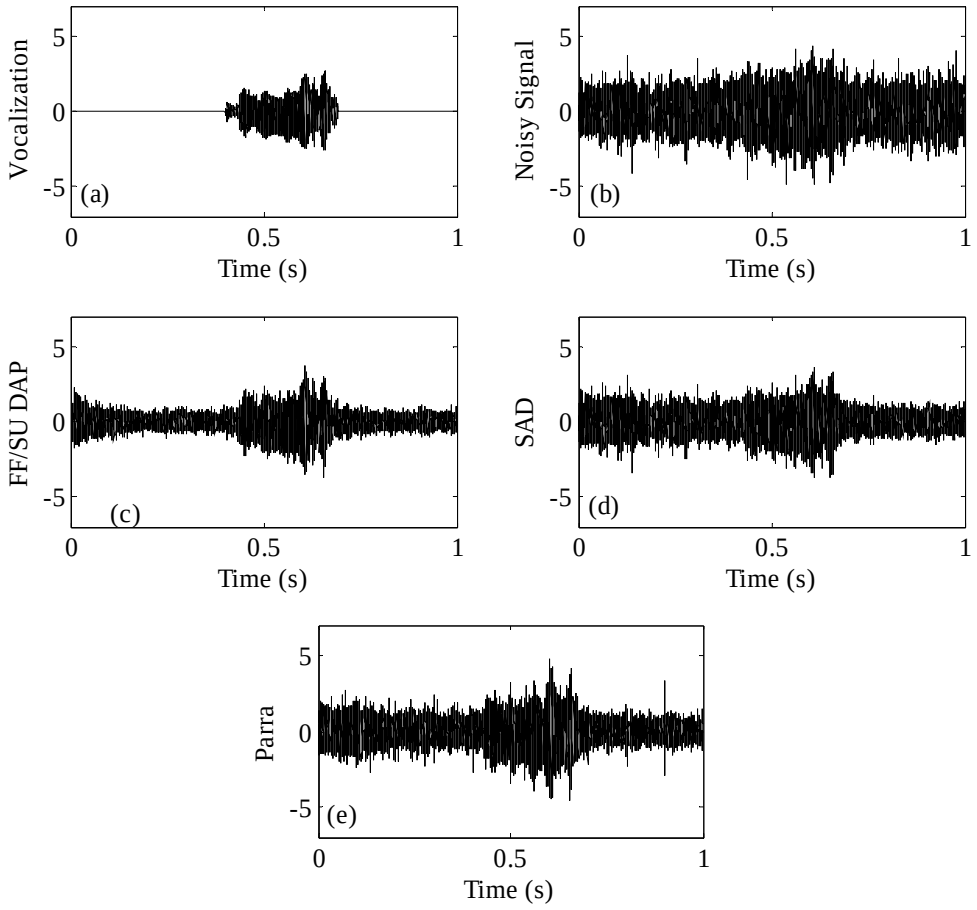


Figure 6-28. Time domain outputs of the SOS-BSS algorithms for the third Crystal River test case. Input SNR is set to 0 dB. Plots are for the channel 1 measurements.

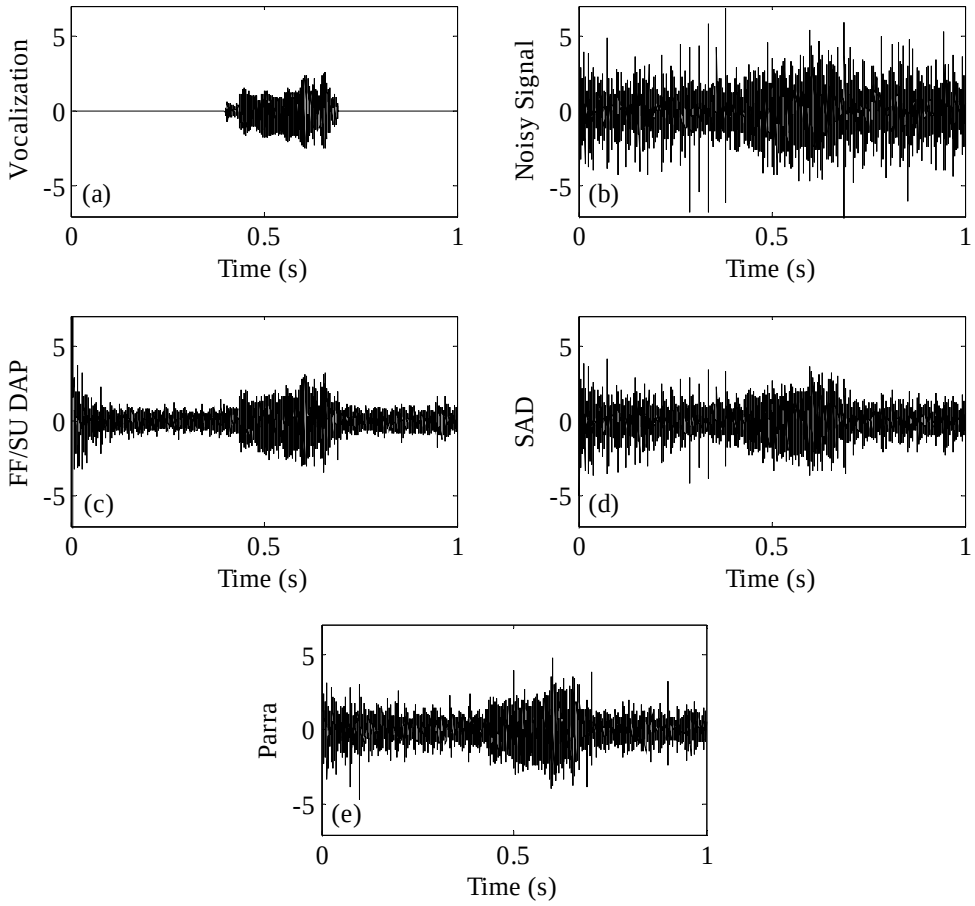


Figure 6-29. Time domain outputs of the SOS-BSS algorithms for the fourth Crystal River test case. Input SNR is set to 0 dB. Plots are for the channel 1 measurements.

As is discussed in Section 3.5, the pseudo channel transfer function (which is defined as the relative transfer function between a reference channel and the other channels) is more relevant for BSS algorithms. The results presented above are obtained for manatee vocalizations convolved with a pseudo 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 source (i.e., manatee) and the corresponding reference receiver on the performance of the FF/SU DAP algorithm is investigated. The separation between the hydrophones is fixed at 9.3 m and the speaker is placed at 1 m, 4.7 m, 10.3 m, and 13.9 m away from the reference hydrophone. Although the distance between the hydrophones stays the same, the pseudo channel impulse response will change as the range of the speaker is increased (as a consequence of the multi-path and frequency selective acoustic channel). Typical pseudo channel impulse responses estimated for the above four speaker-hydrophone distances are plotted in Figure 6-30. The output SNR and SDR results obtained with the speaker placed at the four different locations is presented in Figures 6-31 and 6-32.

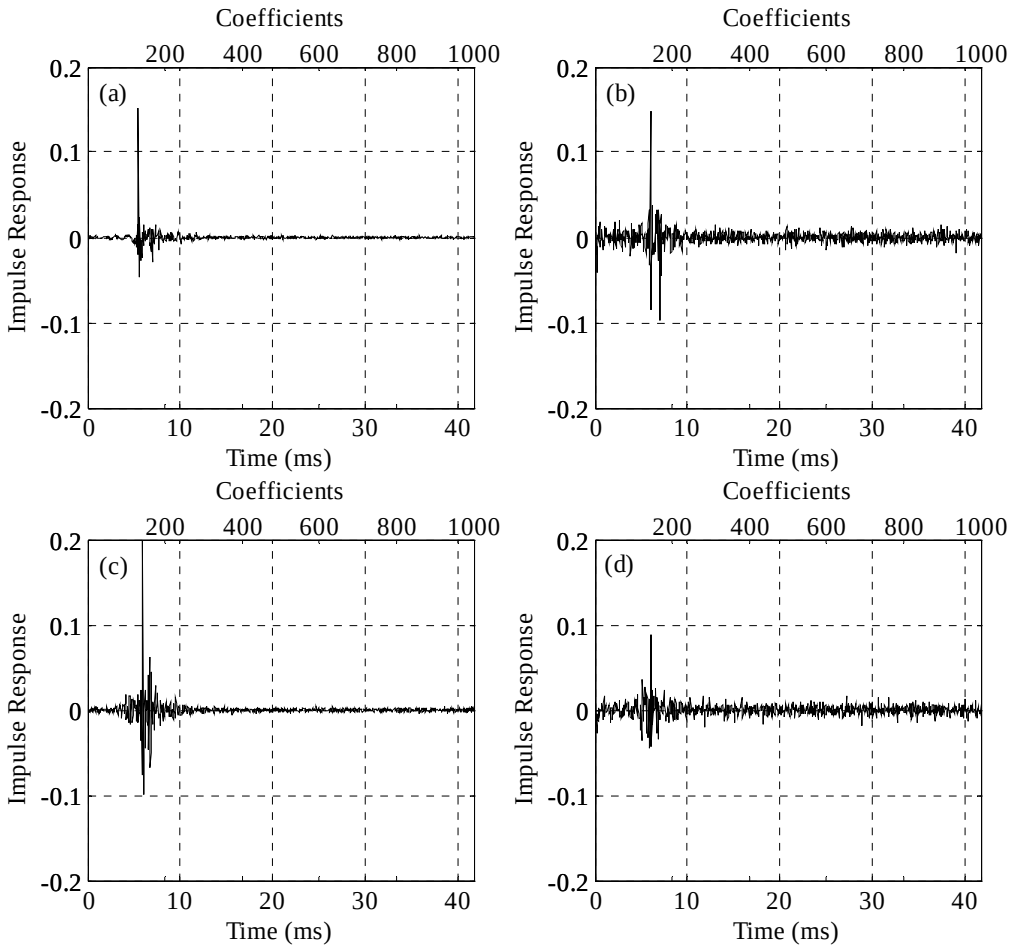


Figure 6-30. The pseudo channel impulse responses due to a source at various distances to the receivers. The results are shown for (a) 1 m, (b) 4.7 m, (c) 10.3 m, and (d) 13.9 m distances obtained from broadband Gaussian noise broadcast rest in Crystal River, FL.

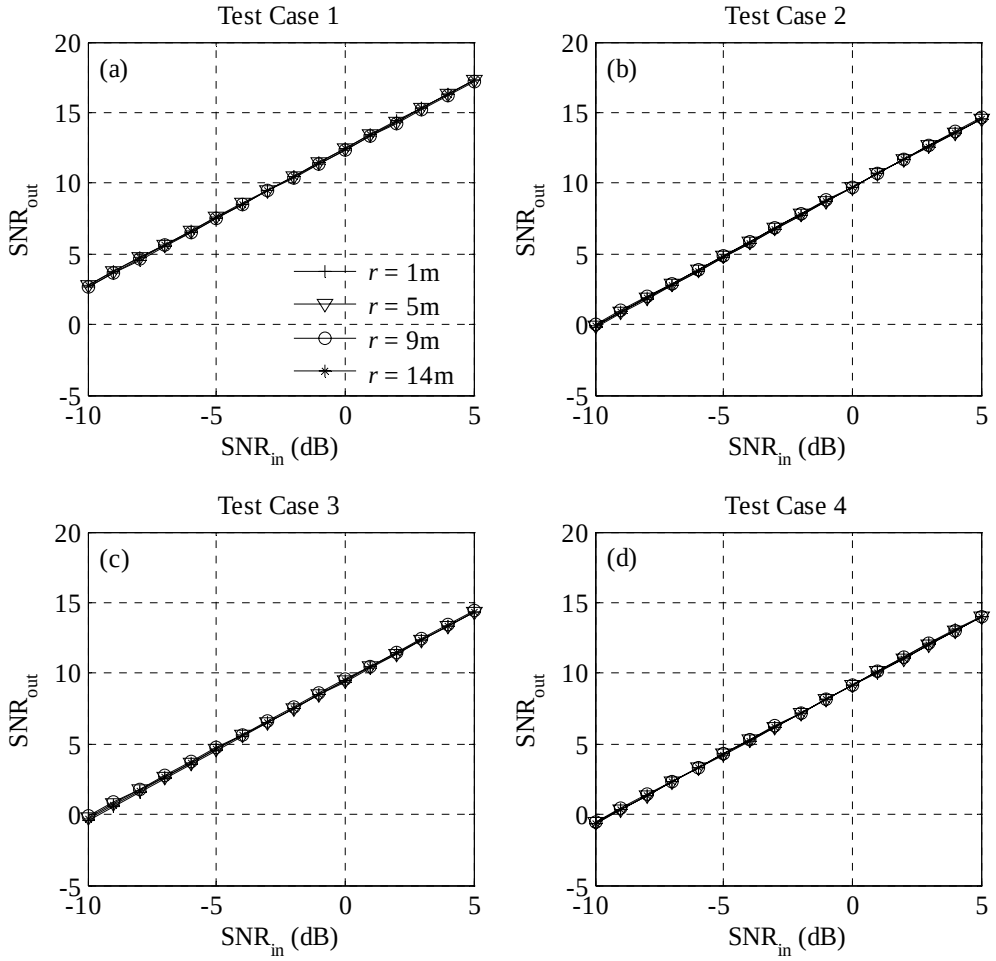


Figure 6-31. The output SNR obtained from the FF/SU DAP algorithm with various source distances for the four Crystal River test cases. The output SNR is averaged over the 10 vocalizations and the separation between hydrophones is fixed at 9.3 m.

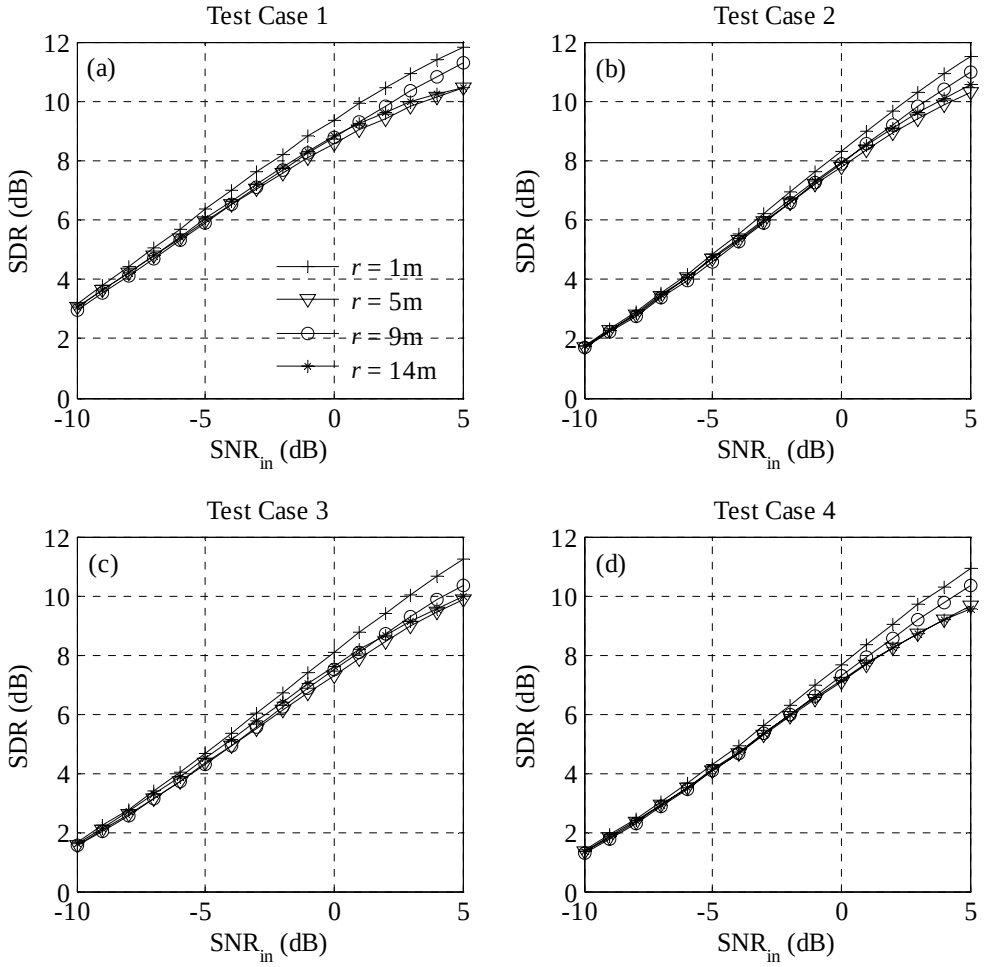


Figure 6-32. The SDR obtained from the FF/SU DAP algorithm with various source distances for the four Crystal River test cases. The SDR is averaged over the 10 vocalizations and the separation between hydrophones is fixed at 9.3 m.

The results presented in Figures 6-31 and 6-32 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, in particular, at low input SNR values. A key conclusion that is drawn

from these results is 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 Chapter 7.

6.4 Summary

In this chapter, the performances of the proposed adaptive WDD and affine projection BSS algorithms are evaluated using real vocalizations and background noise recordings. First, the output SNR and SDR performances of the single channel AWPT algorithm is evaluated and compared to the FourierShrink and FALE algorithms. The performances of the AWPT and FourierShrink algorithms are determined to be more robust to variations in the time-frequency structure of the vocalizations compared to the FALE algorithm. The AWPT algorithm outperforms both the FourierShrink and FALE algorithms in terms of the output SNR performance measure and achieves the best SDR performance for low input SNR values. On average, the AWPT algorithm improves the SNR by 5 dB to 20 dB. The improvement is more pronounced at higher input SNR values.

Next, the signal enhancement performances of the three implementations of the DAP algorithm are compared. As predicted by the Monte-Carlo simulations presented in Chapter 5, the FF/SU DAP algorithm outperforms the feedback implementations. The FF/SU DAP algorithm also achieves a better output SNR and SDR compared to the SAD and Parra algorithms. The FF/SU DAP algorithm achieves a constant 10 dB SNR improvement for all tested input SNR values. In addition, the FF/SU DAP algorithm

generally achieves a significantly higher SDR performance compared to the single channel AWPT algorithm. In the next chapter, the performance measures presented in this chapter are used to estimate the improved detection ranges associated with the proposed algorithms.

CHAPTER 7: DETECTION RANGE EVALUATION

The improvements in detection range resulting from denoising and enhancing manatee vocalizations with the proposed wavelet domain denoising (WDD) and blind source separation (BSS) algorithms are presented in this chapter. First, the passive sonar equations are reviewed and the relationship between the output signal to noise ratio (SNR) performance metric and detection range is established. Next, the improved detection ranges obtained after processing the noisy vocalizations with the adaptive wavelet packet transform (AWPT) algorithm is presented and compared to other single channel approaches. In the final section of the chapter, the improvements in the detection range resulting from the feedforward structure, sequential-update double affine projection (FF/SU DAP) algorithm proposed in Chapter 5 is presented.

7.1 Estimation of the Detection Range

The denoising and enhancement performances of the proposed algorithms can be related to the improvements in detection through the passive sonar equations. In this section, a detailed theoretical development of the passive sonar equations and their relation to the detection range is presented [144], [120].

The sound pressure level (SPL) of a distant point source measured at a receiver will be lower compared to the source level (SL) due to channel transmission loss (TL)

$$\mathrm{SPL} = \mathrm{SL} - \mathrm{TL}, \tag{7-1}$$

where SL is measured at 1 m away from the source. However, the signal to noise ratio (SNR) of the input signal at the receiver will also be affected by ambient noise levels (NL)

$$\mathrm{SNR}_{\mathrm{in}} = \mathrm{SL} - \mathrm{TL} - \mathrm{NL}. \tag{7-2}$$

The conventional passive sonar consists of several receivers which can be beamformed (see Section 2.3) and made to 'listen' to the direction of the vocalization source (e.g., the manatee) while ignoring the noise signals emitted from other point noise sources. Theoretically, the beamformer will preserve the vocalization signal and only the noise signals that are emitted from the direction of the vocalization source will be picked up. Thus, beamforming will increase the input SNR, and the increase will be a function of the spatial resolution of the beamformer. Array gain (AG) is defined as the improvement in the SNR that can be achieved by beamforming compared to a single omni-direction receiver. With AG, the input SNR can be computed using

$$\mathrm{SNR}_{\mathrm{in}} = \mathrm{SL} - \mathrm{TL} - \mathrm{NL} + \mathrm{AG}. \tag{7-3}$$

The detection threshold (DT) is defined as the minimum input SNR necessary to achieve a satisfactory detection performance (e.g., 90% probability of detection). The DT will in generally be a function of the detector, the vocalization and noise spectra, as well as some

other factors. The passive sonar equation is expressed in terms of the signal excess (SE) as

$$SE = SL - TL - NL + AG - DT, \tag{7-4}$$

where SE is defined as the input signal power in excess of what is required for an average detector performance. As the signal excess increases, the detection performance of the detector will improve. The figure of merit (FOM) is defined as the transmission loss that results in a zero SE, and thus the desired detector performance

$$FOM = SL - NL + AG - DT. \tag{7-5}$$

If the TL can be formulated as a function of the range, Eq. (7-5) can be used to estimate the detection range in terms of the detector performance, vocalization source levels, background noise levels, and array processing gain.

For the detection of manatee vocalizations, two different noise sources are considered in this dissertation. The ambient noise level is already included (as the term NL) in the passive sonar equation presented in Eq. (7-4). Watercraft, on the other hand, is modeled as point acoustic noise sources and is also associated with a TL. Since watercraft emitted noise is statistically independent from ambient noise, the noise level emitted from watercraft must be added to ambient noise levels to determine the total noise level at the receiver. However, it is difficult to isolate watercraft emitted noise from ambient noise and measure the SL of watercraft emitted noise alone. Thus, the noise level at the receiver is estimated as the greater of the ambient noise levels in the

absence of watercraft and the noise level expected in the presence of watercraft. Therefore, the FOM can be expressed as [24]

$$\mathrm{FOM} = \mathrm{SL_m} - \max(\mathrm{NL_a}, \mathrm{SL_b} - \mathrm{TL_b}) + \mathrm{AG} - \mathrm{DT}, \tag{7-6}$$

where $\mathrm{NL_a}$ is the ambient noise levels, $\mathrm{SL_b}$ is the source level of the watercraft, and $\mathrm{TL_b}$ is the transmission loss for the watercraft. Based on the experimental results presented by Phillips et al. [24], the mixed spreading model

$$\mathrm{TL_{mxd}} = 15\log_{10}(r), \quad r \geq 1, \tag{7-7}$$

where r is the range to the receiver is assumed. For mixed spreading, Eq. (7-6) can be recast in terms of the distance between the manatee and the receiver (r_m) as

$$r_\mathrm{m} = 10^{\left(\mathrm{SL_m} - \max\left(\mathrm{NL_a}, \mathrm{SL_b} - 15\log_{10}(r_\mathrm{b})\right) + \mathrm{AG} - \mathrm{DT}\right)/15}, \tag{7-8}$$

where r_b is the distance between the watercraft and receiver. The DT is generally a positive value for conventional passive sonar applications. However, while the average watercraft emitted noise SL was determined to be 140 dB (above 2 kHz), the noise floor in this frequency range is reported to be 120 dB or less [24]. As a consequence, the local input SNR at the harmonic frequencies of the vocalizations will be higher than the overall input SNR. Thus, for a detector that exploits the harmonic structure of the vocalizations, the DT will be lower than typical DT values for passive sonar systems.

In the following sections, the improvements in the detection range that can be achieved by processing the manatee vocalizations with the proposed WDD and BSS algorithm are evaluated. The output SNR performance measure is more appropriate for

relating the denoising and signal enhancement performance of the algorithms to detection range through the passive sonar equation. The following detection range computations are based on the output SNR performance measure with the assumption that the vocalization waveform is not distorted during denoising or enhancement, and does not affect the DT. In this dissertation, the waveform distortion is quantified by the signal to distortion ratio (SDR) performance indicator. As presented in Sections 6-2 and 6-3, both the WDD and BSS algorithms distort the vocalization signal to some extent which may increase the DT and decrease the range improvements that are achieved. Thus, the detection range estimates are biased for algorithms that sacrifice preserving the waveform in favor of high SNR. To minimize the effects of waveform distortion on detection range estimates, the parameters of the evaluated algorithms are optimized for best SDR performance rather than the SNR performance.

7.2 Detection Range Estimation for Wavelet Domain Denoising

For the single channel algorithms proposed in Chapter 4, the AG is zero and Eq. (7-8) reduces to

$$r_{\mathrm{m}} = 10^{\left(\mathrm{SL_m} - \max\left(\mathrm{NL_a}, \mathrm{SL_b} - 15\log_{10}(r_{\mathrm{b}})\right) - \mathrm{DT}\right)/15}. \tag{7-9}$$

The improvement in SNR resulting from processing the noisy vocalizations is incorporated into Eq. (7-9) as a decrease in the DT (i.e., the minimum input SNR required for average detector performance). This approach was used by Yan [28] to evaluate the detection performance of the adaptive line enhancer (ALE) and is also

adopted in this study for single channel algorithms. As an example, the output SNR of the adaptive wavelet packet transform (AWPT) algorithm is 8.3 dB when the input SNR after highpass filtering is -5 dB for the first Crystal River test case. Thus, if the DT is 8.3 dB, the detector would be able to detect a vocalization signal with an input SNR -5 dB if the AWPT is implemented prior to detection. This improvement in the DT adjustment is illustrated in Figure 7-1. The average output SNR and average SDR resulting from the three algorithms evaluated in Section 6.2 are presented in Figures 6-7 and 6-12.

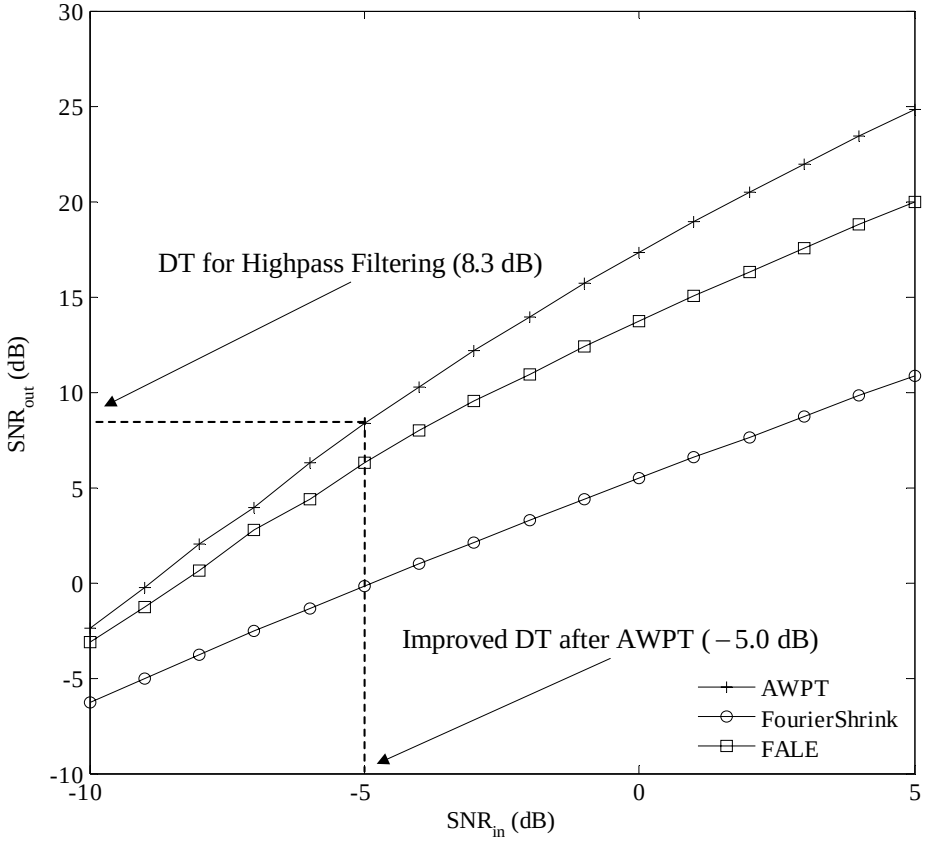


Figure 7-1. An example illustrating the improvement in the DT resulting from denoising the vocalizations.

Using the detection range formula presented in Eq. (7-9), the ratio of the detection range after the AWPT and after highpass filtering can be determined as

$$\frac{(r_{\mathrm{m}})_{\mathrm{AWPT}}}{(r_{\mathrm{m}})_{\mathrm{HPF}}} = 10^{(-\mathrm{DT}_{\mathrm{AWPT}}+\mathrm{DT}_{\mathrm{HPF}})/15}, \tag{7-10}$$

where $(r_{\mathrm{m}})_{\mathrm{AWPT}}$, $(r_{\mathrm{m}})_{\mathrm{HPF}}$ and $\mathrm{DT}_{\mathrm{AWPT}}$, $\mathrm{DT}_{\mathrm{HPF}}$ are the detection ranges after AWPT and highpass filtering, and the detection thresholds associated with AWPT and highpass filtering, respectively. For the particular case depicted in Figure 7-1, the detection range after the AWPT algorithm is determined to be 7.7 times greater than the range associated with highpass filtering alone. The ratio of the detection ranges resulting from the AWPT, FourierShrink and feedback adaptive line enhancer (FALE) algorithms evaluated in Chapter 4 over highpass filtering for the four Crystal River test cases are presented in Figure 7-2. In the upper end of the input SNR range, the AWPT algorithm can improve the detection range by a factor of 10 to 20. As the input SNR is reduced to -10 dB, none of the proposed algorithms can improve the detection range over the highpass filter. The improvement in detection range estimated for the FALE algorithm is consistent with the results presented by Yan [28]. The improvement in detection range resulting from FourierShrink algorithm is significantly inferior those resulting from the AWPT and FALE algorithms. Therefore, the improvement in detection ranges that result from the FourierShrink algorithm are not investigated further. A more detailed analysis of the improvements in the detection range resulting from the FALE algorithm is beyond the scope of this dissertation and can be found in the work of Yan [28].

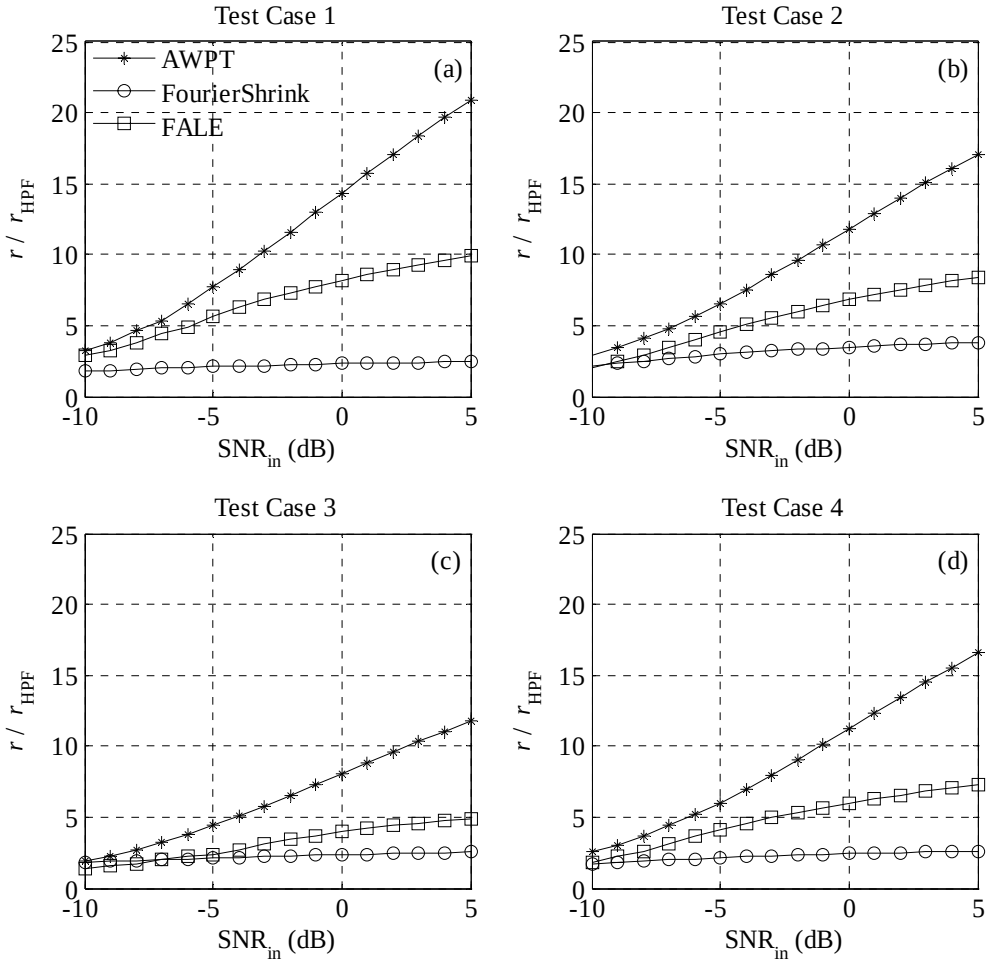


Figure 7-2. The ratio of the detection range resulting from the WDD algorithms to highpass filtering for the four Crystal River test cases.

In the following, the example that was provided in Section 2.2 for estimating the detection ranges without signal processing is revisited. In that example, the watercraft emitted noise SL after highpass filtering is assumed to be 140 dB, whereas the manatee vocalization SL is taken as 118 dB [24]. For a given ambient noise level of 70 dB, 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 100 dB, the watercraft emitted noise will still dominate the noise levels provided that the watercraft is within 464.2 m of the receivers. The experimentally obtained performance measures presented in Section 6-2 are computed from recordings made in an environment where the ambient background noise levels are significantly lower (-12 dB or lower) compared to SPL 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 this scenario, if the watercraft is 500 m away, the detection range is limited to 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. The detection ranges for highpass filtering is depicted in Figure 7-3 where the detection range is plotted as a function of the DT and the distance between the watercraft and receivers.

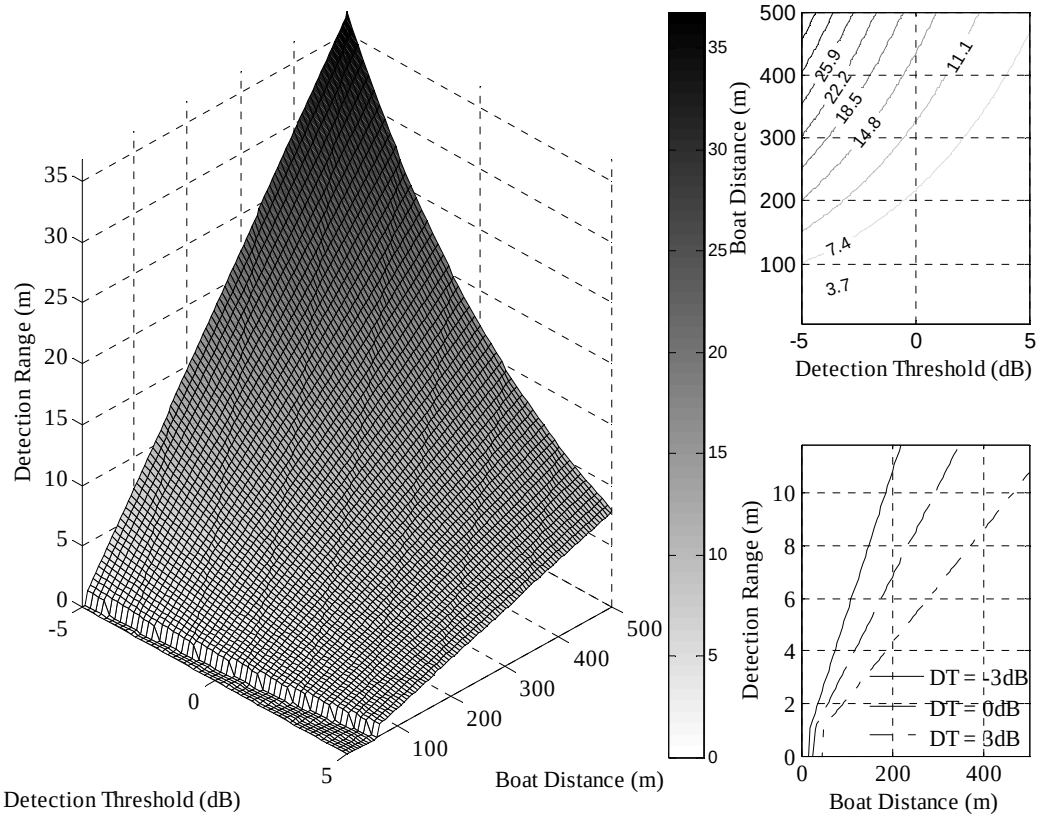


Figure 7-3. Manatee vocalization detection distances with highpass filtering only in the presence of watercraft emitted noise.

For improved detection range estimations (obtained after processing the noisy vocalizations with the AWPT algorithm), the DT in the case of highpass filtering corresponds to the output SNR of the AWPT algorithm, and the input SNR corresponds to the improved DT obtained with the AWPT algorithm (see Figure 7-1). 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 AWPT algorithm which is estimated through experiments in Chapter 6 is available only at non-equally spaced

discrete values. To obtain an analytic relation between the input SNR and the output SNR for the AWPT algorithm, the output SNR averaged over the four Crystal River test cases is curve fit with a second order polynomial

$$y(x) = -0.0314\,x^2 + 1.64x + 15.7038, \tag{7-11}$$

in the least square sense [28]. The curve fit is depicted in Figure 7-4. For a DT of 3 dB, the input SNR for the AWPT algorithm is computed as -6.8 dB. Thus, with the AWPT 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 Figure 7-2, the average improvement ratio in the detection range for the AWPT algorithm can be inferred to be in the range of 3.2 to 6.5 for an input SNR of -6.8 dB. Using Eq. (7-11), a more precise value for the ratio $r_{\mathrm{AWPT}}/r_{\mathrm{HPF}}$ is computed as 4.5. If the watercraft is at a distance of 500 m, the improved detection range will become 48.5 m (cf. to 10.8 m with highpass 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 (cf. to 5.4 m with highpass filtering). The minimum 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. (cf. to 46.4 m with highpass filtering). The improved detection ranges are depicted in Figure 7-5 for the scenario described above. As mentioned in the previous section, the detection range estimates provided in this section ignore the possible increases in the DT due to signal distortion. The AWPT algorithm, as well as the other algorithms evaluated in Chapter 4 and Chapter 5 distort the vocalization signal waveform

to some extent while suppressing noise. The relationship between SDR and the DT depends on the detector characteristics and is in general difficult to quantify analytically. However, the experiments conducted in Chapter 6 (see Figure 6-12) reveals that the AWPT algorithm achieves a slightly improved SDR performance for all the tested input SNR over the FALE algorithm. Thus, the detection threshold is expected to increase relatively less compared to the FALE algorithm.

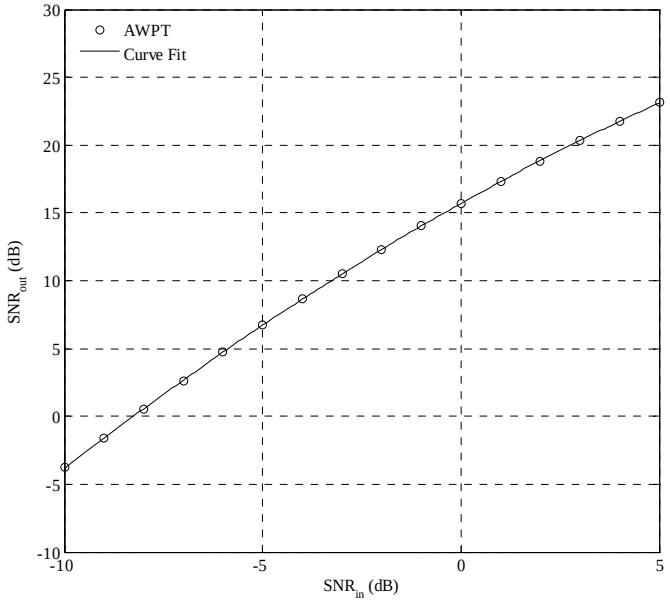


Figure 7-4. The output SNR for the AWPT algorithm and the second order polynomial fit. The output SNR result is averaged over the four Crystal River test cases and is plotted as a function of the input SNR.

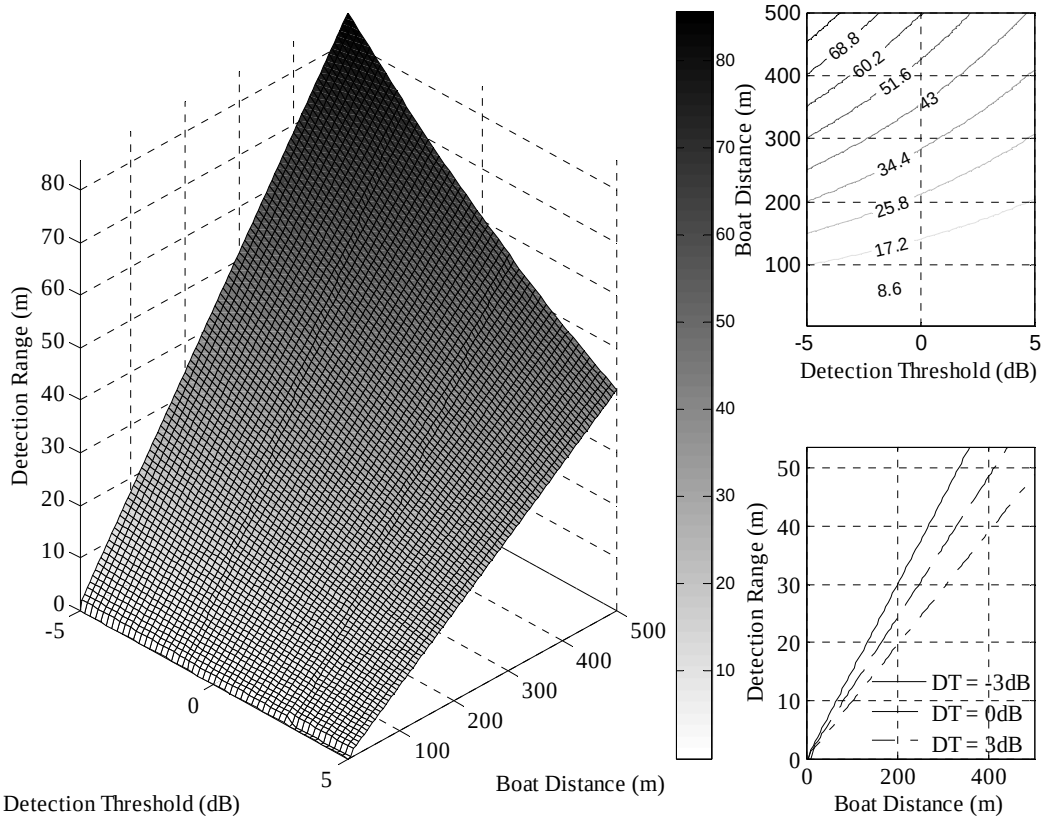


Figure 7-5. Manatee vocalization detection distances after processing with the AWPT algorithm in the presence of watercraft emitted noise.

7.3 Detection Range Estimation for Blind Source Separation

In Section 7.1, the AG is defined for multi-channel setups as the improvement in the SNR with respect to the single receiver input SNR. Thus, for the multi-channel BSS algorithms, AG is defined as the improvement in the SNR

$$\mathrm{AG} = \mathrm{SNR}_{\mathrm{out}} - \mathrm{SNR}_{\mathrm{in}}. \tag{7-12}$$

For the two channel recordings used to test BSS algorithms in Section 6.3, each input channel has a different input SNR value, and the highest channel SNR (measured at the reference channel for the vocalizations) is defined as the input SNR. This assumption is also adopted in the following analysis. The maximum range at which a manatee can be detected using BSS algorithms is expressed by modifying Eq. (7-8) to

$$r_{\mathrm{m}} = 10^{\left(\mathrm{SL_m} - \max\left(\mathrm{NL_a}, \mathrm{SL_b} - 15\log_{10}(r_{\mathrm{b}})\right) + \mathrm{AG} - \mathrm{DT}\right)/15}. \tag{7-13}$$

The ratio of the detection range resulting from the DAP algorithm to highpass filtering detection range can be computed using the relation

$$\frac{\left(r_{\mathrm{m}}\right)_{\mathrm{DAP}}}{\left(r_{\mathrm{m}}\right)_{\mathrm{HPF}}} = 10^{\mathrm{AG}/15}. \tag{7-14}$$

The average output SNR and SDR resulting from the feedforward structured, sequential-update double affine projection (FF/SU DAP) algorithm are evaluated in Section 6.3 and are presented in Figures 6-22 and 6-23. The distance between the manatee and the receivers will affect both the input SNR and the pseudo channel impulse response. However, the output SNR and SDR performances of the BSS algorithms are shown to be rather insensitive to the channel impulse responses (see Figures 6-31 and 6-32). Thus, the detection range for manatee vocalizations is assumed to be a function of the SNR only. For example, the output SNR of the feedforward structured, sequential-update double affine projection (FF/SU DAP) algorithm is 7.5 dB when the input SNR after highpass filtering is -5 dB for the first Crystal River test case, which corresponds to a

12.5 dB AG. Using Eq. (7-14), the resulting improved detection threshold is estimated to be 6.8 times what can be achieved by a single channel highpass filter. The ratio of the detection ranges resulting from the FF/SU DAP algorithm over highpass filtering for the four Crystal River test cases are presented in Figure 7-6.

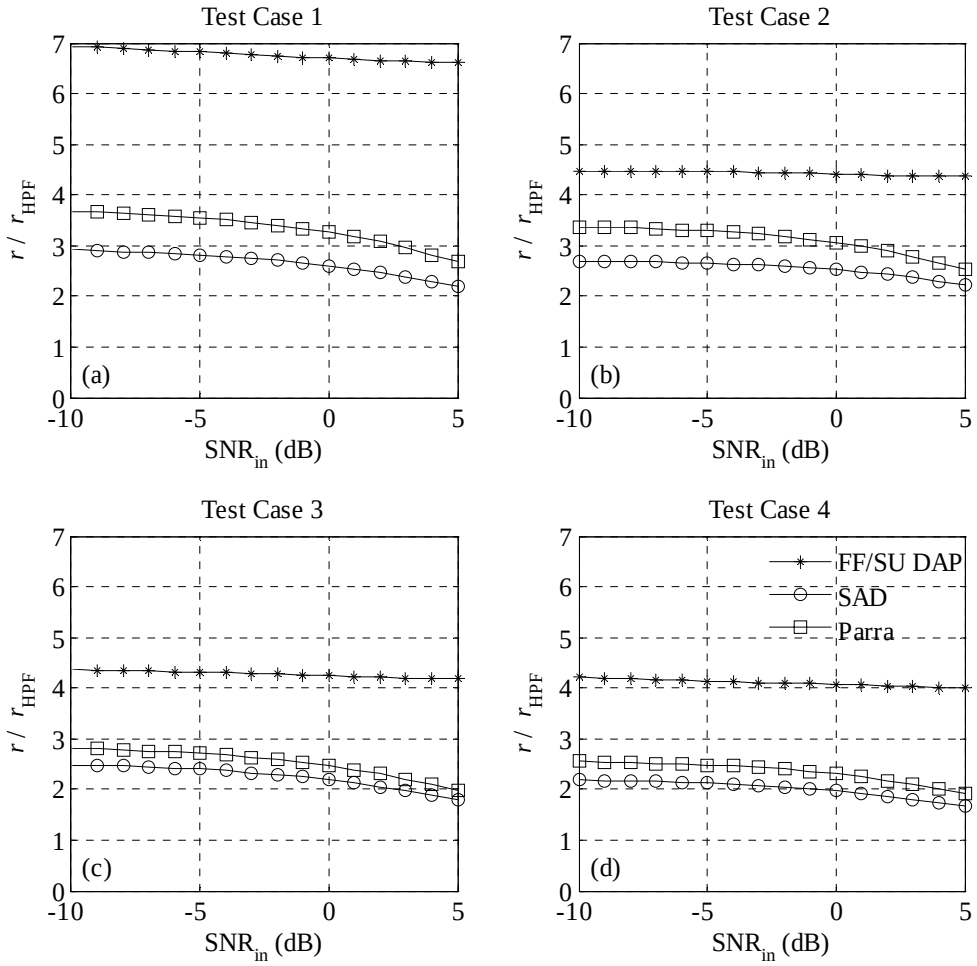


Figure 7-6. The ratio of the detection range resulting from the FF/SU DAP algorithm to highpass filtering for the four Crystal River test cases.

Unlike the single channel WDD approaches presented in the previous section, BSS algorithms achieve a relatively uniform output SNR performance over the input SNR values. The FF/SU DAP algorithm can improve the detection range by a factor of 4 to 7 for input SNR varying from -10 to 5 dB. As is discussed in Section 6.3, the output SNR performance of the BSS algorithms is limited by the presence of extraneous noise, lower than ideal coherence of watercraft emitted noise, and channel dynamics that cannot be modeled accurately with FIR filters.

Next, the detection ranges resulting from processing the vocalizations with the FF/SU DAP algorithm is presented through an example. The watercraft noise and manatee vocalization SL are assumed to be 140 dB 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. As in the case of the WDD methods, the output SNR for the FF/SU DAP algorithm estimated through experiments in Chapter 6 is available only at non-equally spaced discrete values. To obtain an analytic relation between the input SNR and the output SNR, the output SNR averaged over the four Crystal River test cases is curve fit with a first order polynomial

$$y(x) = 0.9734\, x + 10.2057, \tag{7-15}$$

in the least squared sense. The curve fit is depicted in Figure 7-7. Note that Eq. (7-15) 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 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

highpass filtering is computed as 4.9 for a DT of 3 dB (cf. to 4.5 for the AWPT algorithm). For DT lower than 3 dB (i.e., lower input SNR), the improvement in the detection range that can be achieved with the FF/SU DAP algorithm over highpass filtering or the single channel AWPT algorithm will further increase. This can be observed from Figure 7-8 where the ratio of the improvement in the detection range resulting from the AWPT algorithm to highpass filtering, the FF/SU DAP algorithm to highpass filtering, and the FF/SU DAP algorithm to the AWPT algorithm are plotted. The FF/SU DAP algorithm achieves a higher detection range compared to the AWPT algorithm for input SNR below -6 dB. The SDR performance measure of the FF/SU DAP algorithm is significantly better than the AWPT (cf. Figures 6-12 and 6-25), hence, the actual input SNR value which the AWPT algorithm starts outperforming the FF/SU DAP will most likely be higher than -6 dB. If the watercraft is at a distance of 500 m, the improved detection range will become 53.2 m (cf. to 10.8 m for highpass filtering and 48.5 m for AWPT algorithm). If the watercraft comes within 250 m of the receivers, the improved detection range will be 26.6 m (cf. to 5.4 m for highpass filtering and 24.4 m for the AWPT algorithm). The minimum range of the watercraft to the receivers for which manatee vocalizations cannot be detected, irrespective of the range of the manatee to the receivers, improves to 9.4 m. (cf. to 46.4 m for highpass filtering and 10.3 m for the AWPT algorithm). The improved detection ranges are depicted in Figure 7-9 for the scenario described above.

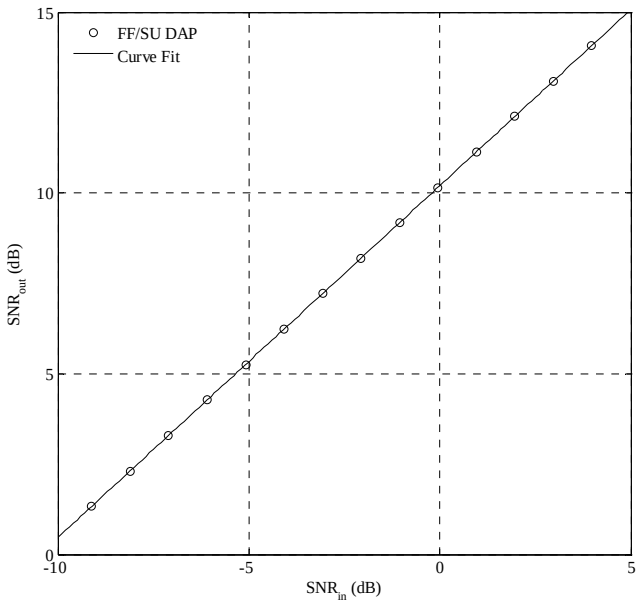


Figure 7-7. The output SNR for the FF/SU DAP algorithm and the first order polynomial fit. The output SNR result is averaged over the four Crystal River test cases and is plotted as a function of the input SNR.

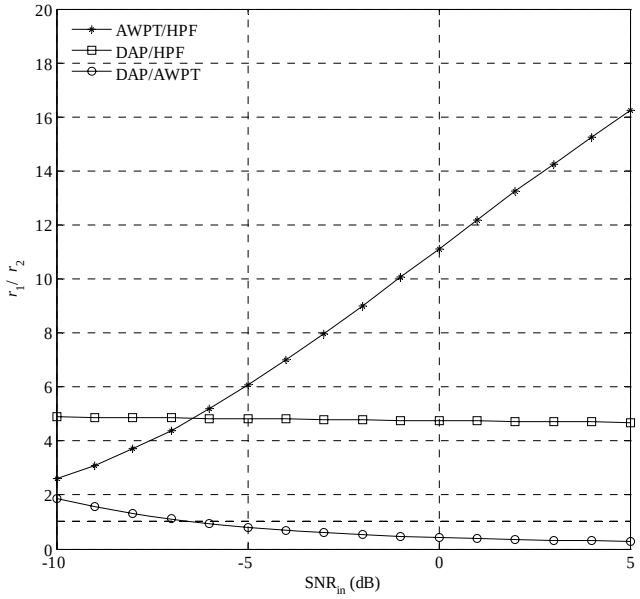


Figure 7-8. The comparison of the detection range ratios of AWPT and FF/SU DAP algorithms.

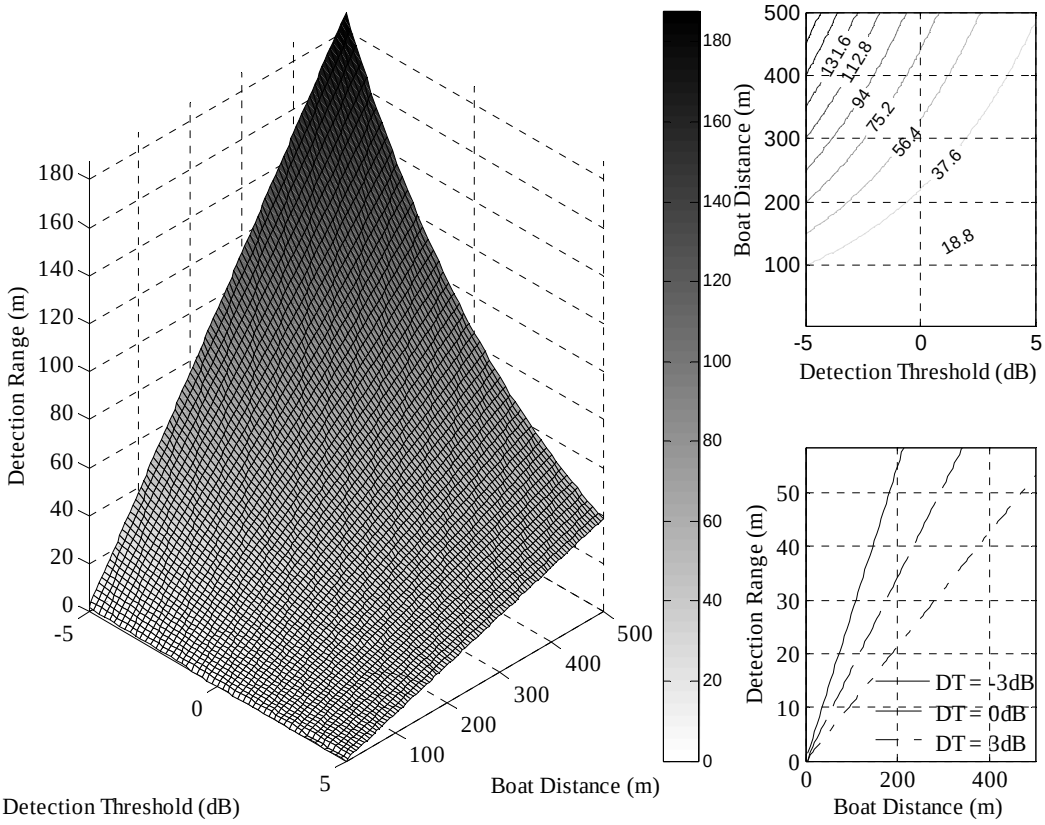


Figure 7-9. Manatee vocalization detection distances after processing with the FF/SU DAP algorithm in the presence of watercraft emitted noise.

7.4 Summary

In this chapter, the improved detection ranges resulting from processing the noisy vocalizations with the AWPT and FF/SU DAP algorithms are presented as a ratio to the maximum detection range that can be achieved with highpass filtering alone. The improved detection ranges are estimated as a function of the output SNR using the passive sonar equation and assuming the mixed spreading transmission loss model. The AWPT algorithm outperforms both the FourierShrink and FALE algorithms, and results

in 10 to 20 times better detection ranges compared to highpass filtering at high input SNR. As the input SNR decreases, the improvements in detection that can be achieved using the AWPT algorithm diminishes, and no improvement can be achieved (over highpass filtering) at -10 dB input SNR.

Unlike the single channel AWPT algorithm, the FF/SU DAP algorithm sustains a uniform output SNR performance (between 4 to 7 times better than highpass filtering alone) over the different input SNR values. In general, the FF/SU DAP algorithm outperforms the AWPT algorithm at lower input SNR, typically below -6 dB. In general, it is very difficult to quantify the effect of signal distortion on the detection range. For this analysis, the distortion of the vocalization signals resulting from adaptive denoising or enhancement is assumed to be negligible. In reality, this distortion may cause the DT to increase, which may reduce the gains achieved in the detection range. The FF/SU DAP algorithm significantly outperforms the AWPT algorithm in terms of the SDR. Thus, the FF/SU DAP algorithm may actually outperform the AWPT algorithm at input SNR above -6 dB.

Finally, the estimated detection ranges for highpass filtering alone, for the AWPT algorithm, and for the FF/SU DAP algorithm are provided. If the watercraft is 500 m away, manatee vocalizations can be detected at a maximum distance of approximately 50 m for the AWPT and FF/SU DAP algorithms. When the watercraft approaches to 250 m, the maximum detection range drops to approximately 25 m.

CHAPTER 8: CONCLUSIONS AND FUTURE RESEARCH

In this final chapter, the concluding remarks regarding passive acoustic based detection of manatee vocalizations are presented. The results obtained for manatee vocalizations are interpreted for enhancing other marine mammal vocalizations. Finally, several key aspects of the proposed algorithms that require further research are outlined.

8.1 Conclusions

A promising approach for detecting manatees is based on passive acoustic detection of their vocalizations. However, manatee vocalizations are relatively weak compared to typical background noise levels and in particular, compared to watercraft emitted noise levels. Ironically, detection of manatees is most crucial in the presence of watercraft since a quarter of all unnatural manatee mortalities are attributed to watercraft collisions. For a feasible implementation of a manatee vocalization detector with a satisfactory detection range, the noisy vocalization signals must be processed prior to detection. With this dissertation, two distinctly different methods are proposed for enhancing weak manatee vocalizations in the presence of watercraft emitted noise. First, single channel wavelet domain denoising (WDD) is proposed for denoising manatee vocalizations. This research represents the first effort to enhance marine mammal

vocalizations in the wavelet domain. More specifically, the adaptive wavelet packet transform (AWPT) algorithm is developed and evaluated. Next, several multi-channel blind source separation (BSS) methods based on the affine projection filter update rule are developed and the feedforward structured, sequential-update double affine projection (FF/SU DAP) is proposed for enhancing manatee vocalizations in the presence of watercraft emitted noise. Both the AWPT and the FF/SU DAP algorithms are evaluated using real manatee and watercraft recordings, and are shown to significantly improve the effective detection ranges. As the input signal to noise ratio (SNR) decreases, it becomes more difficult to suppress background noise with minimum distortion of the vocalization signal waveform and time-frequency properties. The multi-channel BSS algorithms generally provide better performance compared to WDD approaches at low input SNR values. On the other hand, WDD methods outperform BSS algorithms at higher input SNR. The symmetric alpha-stable (SαS) distribution model is established as a unifying statistical framework for modeling watercraft emitted noise and snapping shrimp dominated noise environments. Locally optimum (LO) detectors optimized for detecting weak manatee vocalizations that complement the signal enhancement algorithms are developed and evaluated.

Although several researchers have proposed extensions of WDD to heavy tailed noise distributions (e.g., [145]), the AWPT algorithm proposed in Chapter 4 employs the Stein's unbiased risk estimate (SURE) which assumes Gaussian noise contamination. Hence, the AWPT method is not applicable for snapping shrimp noise environments. In addition, the AWPT is based on the non-redundant wavelet packet transform (WPT)

which is not translation invariant (i.e., different circular shifts of the input signal results in different outputs). This reduces the consistency in the denoising results and may be eliminated by replacing the WPT with the translation invariant undecimated wavelet packet transform (UWPT) at the cost of increased computational complexity.

The signal enhancement performance of the proposed BSS algorithms is dependent on coherence of the vocalization and watercraft emitted noise measurements at the multiple receivers. The shallow Florida channels that are inhabited by the manatees are generally very absorptive [24], [146], which can cause the inter-channel coherence to decrease below acceptable levels for satisfactory BSS performance. The inter-channel coherence may be improved by reducing the hydrophone spacing at the cost of reducing the effective coverage area of the passive acoustic detection system. Hence, the BSS algorithms require a trade-off between performance and feasibility.

The methods and algorithms proposed in this dissertation are very generic and can easily be extended to enhance or detect vocalizations of other marine mammals. In fact, the wavelet transform and the associated WDD methods may be more effective for vocalizations with harmonics that exhibit more pronounced frequency modulation such as whale vocalizations. Several researchers (e.g., [54], [56]) have already suggested using wavelet coefficients as effective discriminating features for whale vocalizations. In addition, for denoising broadband vocalizations or echolocation clicks that do not exhibit a well defined time-frequency structure, wavelet domain methods may be the most viable approach.

For BSS methods on the other hand, the primary factor that affects the signal enhancement performance is related to the underwater acoustic channel, rather than the properties of the vocalization and noise signals. As mentioned previously, the very shallow water habitat of the Florida manatee is in particular a very challenging environment due to the inhomogenieties and other uncertainties inherent to the channel, as well as numerous and complex interactions of the acoustic waves with the waveguide boundaries. These interactions generally attenuate the acoustic waves more drastically in shallow waters compared to deeper ocean environments, and significantly reduce the coherence at multiple receivers. The acoustic channel in deeper waters is generally more predictable and sound can travel greater distances. Hence, the BSS methods proposed in this dissertation may perform significantly better in deep ocean environments and can provide a significant improvement in the detection ranges. However, it should be pointed out that as the detection range increases, the effects of the direct path transfer function may become significant and should also be taken into account. The most dominant point acoustic noise source in the deep ocean and arctic environments will most likely be a large surface vessel such as a container ship or an icebreaker. Noise emitted from such surface vessels can also be modeled as a Gaussian random variable. The ambient noise in the deep ocean generally has a Gaussian distribution. However, the Arctic environments may also be contaminated with impulsive noise due to cracking ice. Ice cracking can conveniently be modeled as an impulsive random variable and the LO detectors developed for snapping shrimp noise dominated environments may be extended to Arctic environments.

The application of the proposed AWPT and FF/SU DAP algorithms are not limited to marine mammal monitoring applications and can be extended to other signal enhancement problems. For example, wavelet basis functions are very effective in compactly representing singularities in a signal compared to other bases [147-148]. Hence, the AWPT algorithm can be utilized in structural health monitoring and fault diagnostics applications for denoising sensor data and improving failure detection. Alternatively, the FF/SU DAP algorithm can be extended to multimedia and teleconferencing applications which require the separation of multiple simultaneous speakers.

8.2 Future Research

The WPT implemented using Daubechies-8 wavelet basis functions is shown in Section 4.2 to better discriminate vocalizations from background noise compared to the Fourier transform. The denoising performance of the AWPT proposed in this dissertation can be improved by implementing the WPT with efficient wavelet basis functions or by optimizing the time-frequency decomposition of the WPT based on some criterion. The Hilbert transform can be incorporated into the denoising algorithm to improve the denoising performance for harmonic vocalization signals as proposed by Olhede and Walden [149]. Alternatively, the performance of the WDD methods can be improved by extending the AWPT to multi-channel setups. One such possible multi-channel threshold based WDD algorithm was proposed by Rao and Jones [150].

The proposed affine projection based BSS algorithms ignore the distortion caused by the direct path channel impulse response on the vocalization signals. While this distortion is not significant enough to affect the detection of manatee vocalizations, it may become significant if the marine mammal is more distant [151]. Hence, an alternative decorrelation criterion that incorporates the direct paths should be investigated. The proposed FF/SU DAP algorithm does not exploit the non-stationarity and temporal correlation of the source signals. However, including the direct path channel filters into the separation criterion will increase the number of unknown filter coefficients that need to be solved for, and the resulting algorithm will need to incorporate the non-stationarity and/or temporal correlation. As discussed in Section 2.5, several source separation and signal enhancement algorithms that exploit the temporal correlation and signal model of manatee vocalizations were tested. These algorithms failed to achieve a satisfactory enhancement performance because the algorithms were very sensitive to divergence in the noise statistics from the models. These methods may favor better in noise environments with more stable statistics such as the deep ocean environment.

The two-input two-output (TITO) setup is sufficient for enhancing vocalizations of a manatee in the presence of a single watercraft. In other circumstances, the total number of marine mammals and surface vessels can be more than two. For such cases, the FF/SU DAP algorithm must be extended to multi-input multi-output (MIMO) setups. The multi-channel extension of the supervised affine projection algorithm proposed by Benesty et al. [152] can be adopted for this purpose. The computational complexity of

the FF/SU DAP can be reduced by performing the matrix inversion using the Levinson algorithm [153]. Further computational improvements can be achieved by extending the fast implementations of the supervised affine projection algorithm [154-155] to BSS approaches.

APPENDIX A: DETECTION OF WEAK VOCALIZATIONS

Detectors optimized for detecting weak manatee vocalizations are investigated in this appendix. First, a concise review of detection theory is provided, with emphasis placed on non-parametric locally optimum (LO) detectors optimized for detecting weak signals. Next, a wavelet domain LO manatee vocalization detector is developed and evaluated using real vocalization and noise recordings.

A.1 Detection Theory

Signal detection in the presence of noise is a commonly encountered problem in underwater acoustic signal processing and arises in many applications involving active and passive sonar. The detection problem can be conveniently formulated in the framework of binary hypothesis testing. In what follows, a concise review of detection theory is provided following the development presented by Kassam [156]. The derivations presented here are used in Section A.2 to develop a LO detector for manatee vocalizations. Consider a block of N measurements is expressed as

$$\mathbf{x} = \theta \cdot \mathbf{s} + \mathbf{v}, \tag{A-1}$$

where $\mathbf{x} = \begin{bmatrix} x(0) & x(1) & \dots & x(N-1) \end{bmatrix}^T$ is the $N \times 1$ vector of measurements, and $\mathbf{s}$ and $\mathbf{v}$ are similarly defined vocalization and noise signals, respectively. The scalar parameter θ represents the square root of the vocalization signal power and without losing generality, the vocalization signal $\mathbf{s}$ is normalized to unit power (i.e., $E[s^2(n)] = 1$, where $E[\cdot]$ is the expectation operator). The null hypothesis (H) represents the assumption that no signal is present and the measurements are due to noise only; the alternate hypothesis (K) represents the assumption that the signal is present and the measurements are due to both the vocalization and noise signals. For these two hypotheses, the signal power parameters are defined as

$$\begin{aligned} H &: \theta_H = 0 \\ K &: \theta_K > 0. \end{aligned} \tag{A-2}$$

Thus, in the binary hypothesis testing framework, the complicated problem of detecting vocalizations buried in noise reduces to the simpler problem of estimating the value of the parameter θ. In general, the input signal to noise ratio (SNR) of the measurements is not available a priori. Therefore, the null hypothesis is a simple hypothesis (i.e., $\theta_H = 0$), while the alternate hypothesis is composite (i.e., $\theta_K \in \Theta_K$ where Θ_K is a infinite set of possible positive values that θ_K can take).

If noise is modeled as stochastic random variable with a parametric probability density function (PDF), the distribution model parameters (and the PDF) can be estimated from the measurements. Detectors that exploit information available on the

distribution of the noise field are termed parametric detectors. On the other hand, non-parametric detectors rely on very limited and generic information available from the measurements. Parametric detectors generally outperform non-parametric detectors if the assumed parametric noise model is accurate. However, non-parametric detectors are more robust to the uncertainties in the noise statistics. As is discussed in Chapter 3, the two dominant noise sources in the manatee habitat can both be modeled accurately using the symmetric alpha-stable (SαS) distribution. Therefore, only parametric detectors are considered in this dissertation. For parametric detectors, the null and alternate hypotheses can be expressed in terms of the noise PDF as

$$\begin{aligned} H &: f_X\left(\mathbf{x} \mid \theta = 0\right) = f_V(\mathbf{x}), \\ K &: f_X\left(\mathbf{x} \mid \theta = \theta_K\right) = E_s[\, f_V(\mathbf{x} - \theta \mathbf{s})], \end{aligned} \tag{A-3}$$

where $f_V(x \mid \theta)$ denotes the conditional PDF of v given θ, evaluated at x, and $E_s(x) = \int x \cdot f_S(s) ds$ is the expectation with respect to the stochastic vocalization signal $s(n)$.

Given a set of measurements, the probability of a detector to accept the alternate hypothesis is defined as the test function of the detector

$$\delta(\mathbf{x}) = \Pr\{\theta = \theta_K \mid \mathbf{x}\}, \tag{A-4}$$

where $\Pr\{\cdot\}$ is the probability. The power function of a detector is defined as the probability of detection, given the power parameter θ

$$p[\delta(\mathbf{x}\,|\,\theta)] = \int_{-\infty}^{\infty} \delta(\mathbf{x})\, f_V(\mathbf{x}\,|\,\theta)\, d\mathbf{x}. \tag{A-5}$$

Four possible outcomes exist in the binary hypothesis testing problem. The detector may choose the alternate hypothesis when there is a vocalization signal (i.e., correct detection) or may select the null hypothesis when there is no vocalization (i.e., correct rejection). However, the detector can also choose the alternate hypothesis when there is no vocalization signal. This error is termed the false detection or type I error. Finally, the detector may select the null hypothesis when there actually is a vocalization. This error is a missed detection or type II error. It is desirable to have a detector with a high power function when the alternate hypothesis is true and a zero power function for the null hypothesis. Unfortunately, these are conflicting requirements and are generally impossible to achieve simultaneously. Therefore, most detectors are designed to maximize the power function for a given acceptable false detection rate. A detector is said to be the uniformly most powerful (UMP) detector if it has a larger power function than all other possible detectors that achieve a given false alarm rate for the corresponding non-zero power parameter θ_K. The generalized Neyman-Pearson lemma can be used to formulate a UMP detector that optimizes the probability of detection for a given false detection rate

$$\delta(\mathbf{x}) = \begin{cases} 1 & L(\mathbf{x}) > \lambda \\ r(\mathbf{x}) & L(\mathbf{x}) = \lambda \\ 0 & L(\mathbf{x}) < \lambda, \end{cases} \tag{A-6}$$

where $L(\mathbf{x})$ is the likelihood ratio and is defined as

$$L(\mathbf{x}) = \frac{f_X\left(\mathbf{x} \mid \theta \in \Theta_K\right)}{f_X\left(\mathbf{x} \mid \theta = 0\right)} = \frac{f_V(\mathbf{x} - \theta\mathbf{s})}{f_V(\mathbf{x})}, \tag{A-7}$$

and $0 \le r(\mathbf{x}) \le 1$. The likelihood ratio is a function of the stochastic measurements $\mathbf{x}$ and thus, is also a random variable fully defined by its PDF [157]. The detection threshold (λ) is related to the false alarm rate (α) and can be determined from the integral equation [158]

$$\alpha = \int_{\lambda}^{\infty} f[L(\mathbf{x}) \mid \theta = 0] dL, \tag{A-8}$$

or can be estimated from simulations.

Unfortunately, UMP detectors exist for only for a limited number of special cases. For example, if the noise is independent and identically distributed (IID) with a Gaussian PDF, the linear correlator (LC), whose likelihood ratio can be simplified to the test statistic

$$T_{\mathrm{LC}}(\mathbf{x}) = \sum_{n=0}^{N-1} x(n)s(n), \tag{A-9}$$

is the UMP detector [159-160]. In addition, if the parameter θ is known a priori (i.e., the input SNR can be estimated from the measurements), it can be shown that a UMP detector exists and can be derived using the Neyman-Pearson lemma [161].

For problems that do not permit UMP detectors, further restrictions can be placed on the signal models and detectors optimized in some sense under these constraints can be derived. If the measured signals are assumed to be weak compared to noise measurements, the unknown parameter θ for the alternate hypothesis will be in the

vicinity of $\theta_H = 0$ and the detection problem becomes difficult. In that case, the weak signal detection performance can be optimized by maximizing the slope of the power function in the vicinity of the null hypothesis (i.e., $\theta_K \approx 0$). Since the power function is the probability of detection, maximizing the slope of the power function within the vicinity of the null hypothesis for a given false detection rate results in LO detectors. The LO detector can be derived by reinterpreting the power function defined in Eq, (A-5) as a function of θ

$$p(\theta) = \int_{-\infty}^{\infty} \delta(\mathbf{x}) f_V(\mathbf{x} \mid \theta) d\mathbf{x}, \tag{A-10}$$

and taking its derivative with respect to θ. Replacing the alternate hypothesis PDF in the numerator of the likelihood ratio defined in Eq. (A-7) results in

$$T_{\mathrm{LO},1}(\mathbf{x}) = \frac{\dfrac{d}{d\theta} f_V\left(\mathbf{x} \mid \theta_K\right)}{f_V\left(\mathbf{x} \mid \theta_H\right)}, \quad \theta_K \approx 0. \tag{A-11}$$

After some manipulation, the test statistic of a weak stochastic signal in IID noise is derived as

$$T_{\mathrm{LO},1}(\mathbf{x}) = -\sum_{n=0}^{N-1} g_{\mathrm{LO}}[x(n)] E[s(n)], \tag{A-12}$$

where the locally optimum non-linearity is defined as

$$g_{\mathrm{LO}}(x) = -\frac{1}{f_V(x)}\frac{d}{dv} f_V(x) = -\frac{d}{dv}\ln f_V(x). \tag{A-13}$$

Like most bioacoustic signals, manatee vocalizations are zero-mean (i.e., $E[s(n)] = 0$) stochastic signals. Hence, the slope of the power function at $\theta = 0$ is zero and the test statistic given in Eq. (A-12) does not yield a useful detector. As an alternative, the curvature (i.e., the second derivative) of the power function can be maximized for very small θ. The second order LO test statistic is defined as

$$T_{\mathrm{LO},2}(\mathbf{x}) = \frac{\frac{d^2}{d\theta^2} f_V\left(\mathbf{x} \mid \theta_K\right)}{f_V\left(\mathbf{x} \mid \theta_H\right)}, \quad \theta_K \approx 0. \tag{A-14}$$

Expanding the expression in Eq. (A-14) for a IID noise signal, and assuming that the vocalization signal is wide sense stationary (WSS) within the measurement block results in

$$T_{\mathrm{LO},2}(\mathbf{x}) = \sum_{n=0}^{N-1}\sum_{k=0}^{N-1} g_{\mathrm{LO}}[x(n)]g_{\mathrm{LO}}[x(k)]r_s(n-k) + \sum_{n=0}^{N-1} h_{\mathrm{LO}}[x(n)], \tag{A-15}$$

where $r_s(n-k) = E[s(n)s(k)]$ is the autocorrelation of the vocalization signal and the second order LO non-linearity is defined as

$$h_{\mathrm{LO}}(x) = -\frac{d}{dv} g_{\mathrm{LO}}(x) = \frac{d^2}{dv^2}\ln f_V(x). \tag{A-16}$$

The first term in the summation of the right hand side of Eq. (A-15) can be expressed in terms of the square of the signal filtered by the cascaded first order non-linearity $g_{\mathrm{LO}}(x)$ and a matched filter or generalized correlator. The LO detector defined

in Eq. (A-15) is proposed in the next section for detecting weak manatee vocalization signals in snapping shrimp dominated environments. The block diagram of the LO detector is depicted in Figure A-1.

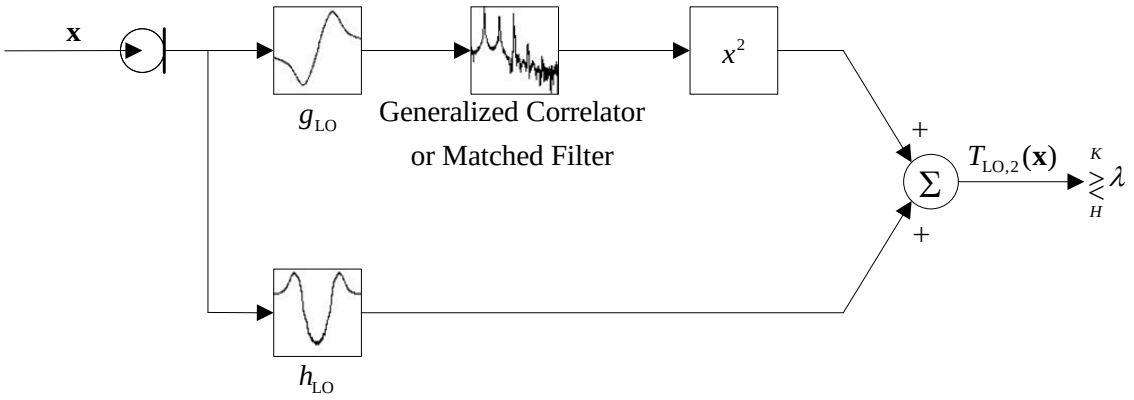


Figure A-1. The block diagram of a LO detector for zero-mean stochastic signals.

A.2 Locally Optimum Detectors for Manatee Vocalizations

In Chapter 3, the two dominant noise sources in the manatee habitat were established as watercraft emitted noise and snapping shrimp noise. Watercraft emitted noise (as measured at the receiver) is modeled as a temporally and spatially correlated Gaussian signal. The IID SαS noise model was proposed for modeling snapping shrimp noise dominated measurements. Furthermore, the SαS noise model was shown to possess the Gaussian noise distribution as a special case, and was proposed as a unifying noise model for watercraft emitted noise and snapping shrimp dominated ambient noise measurements. In this section, a wavelet domain LO detector is proposed for detecting weak manatee vocalizations in snapping shrimp noise dominated environments. Some

comments regarding the extension of the proposed LO to detecting weak vocalization signals in the presence of watercraft emitted noise is also provided.

The SαS distribution model which is described by the exponential PDF function [161]

$$f_V(x;\alpha,\gamma) = \frac{1}{\pi}\int_0^{\infty} \exp(-\gamma t^{\alpha})\cos(xt)dt, \quad 0 < \alpha \le 2, \tag{A-17}$$

where α is the characteristic exponent and γ is the scale parameter, is proposed for modeling snapping shrimp dominated ambient noise recordings. The characteristic exponent is a measure of the impulsiveness of the underlying random variable and the scale parameter is related to the signal power of the random variable. If $\alpha = 2$, the SαS distribution reduces to the Gaussian distribution and the scale parameter is half of the variance. As the other extreme case, for $\alpha = 1$, the SαS distribution reduces to the Cauchy distribution. For other intermediate values of the characteristic exponent, the integral of Eq. (A-17) cannot be calculated analytically and the SαS distribution does not have a closed form PDF. However, the SαS distribution and density functions can be computed numerically [162]. The accuracy of the SαS distribution in modeling snapping shrimp noise is illustrated in Figure A-2 where the asymptotic probability distributions (APD) computed empirically from the measurements is compared to SαS, Gaussian, and Cauchy noise model based computations. A good match between the empirical APD obtained from the measurements and the APD computed for the noise model is an indicator that the distribution model accurately represents the measurement statistics.

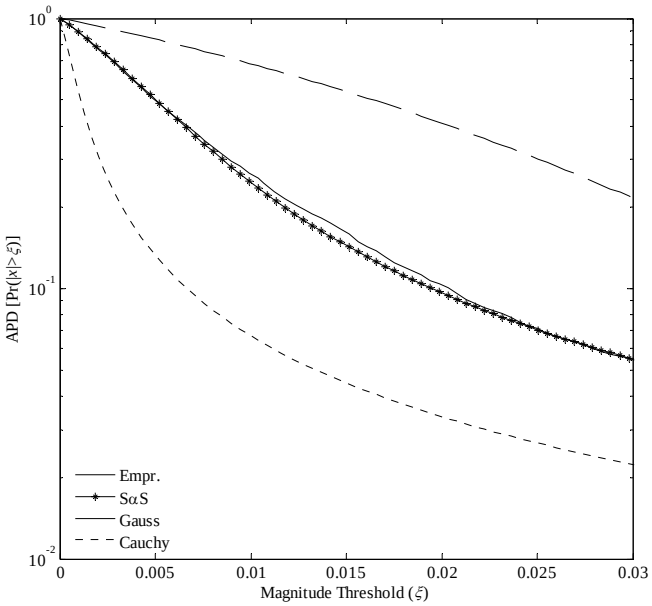


Figure A-2. The APD plot for a typical snapping shrimp noise measurement. The APD is plotted as a function of the magnitude threshold computed empirically from the measurement data and from the different distribution models (the parameters of the distributions are estimated from the measurements).

Once the PDF of the noise distribution is computed, the first and second order LO non-linearities can be computed by numerical differentiation. In Figure A-3, numerically computed LO nonlinearities of the SαS distribution for various characteristic exponents are presented. Examining the plots of Figure A-3 reveal that the first and second order non-linearities associated with Gaussian noise ($\alpha = 2$) are linear (i.e., the output of $g_{\mathrm{LO}}(x)$ is the input samples) and constant, respectively. This observation indicates that the LC is the LO detector for stochastic signals with a non-zero mean in IID Gaussian, which is an expected result since the LC was defined as the UMP detector for Gaussian

noise. Furthermore, the matched filter or the generalized correlator is the LO detector for zero-mean stochastic signals buried in IID Gaussian noise.

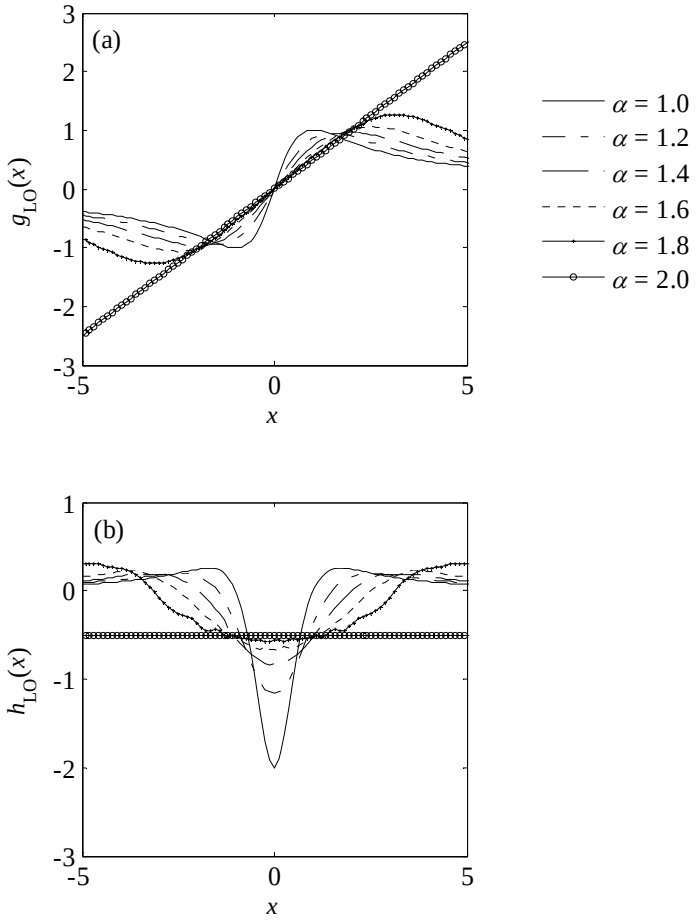


Figure A-3. The SαS LO non-linearities for different characteristic exponents. (a) The first order non-linearity and (b) the second order non-linearity.

For the LO detector optimized in the Neyman-Pearson sense (i.e., the detector that achieves the highest detection probability for a given false detection rate), the matched filter or the generalized correlator is applied to the outputs of the first order LO non-linearity. Designing a matched filter or a generalized correlator requires a prior

knowledge on the vocalization's autocorrelation sequence or spectrum. For example, since the vocalization signals are modeled as autoregressive (AR) signals, the generalized correlator is the AR filter with coefficients equal to the AR coefficients used to model the vocalizations. If the noise-free vocalizations are available, the AR coefficients can be estimated by solving the Yule-Walker equations (for a detailed review of AR spectral estimation, see [101]). Since the noise-free vocalization signals are not available, this approach is not feasible. The AR coefficients estimated for the vocalizations within each vocalization category exhibit some similarity. Hence, a separate correlator can be designed for each vocalization category, and implemented as a bank of filters. The detection decision can be based on the comparison between the maximum of the test statistics obtained from the different filters and a pre-defined threshold. A similar series of matched filters (up to 80 filters) is implemented in the generalized likelihood ratio detector of Urazghildiiev and Clark [163] for detecting North Atlantic right whale contact calls. A spectral correlator is proposed in Mellinger and Clark [164] for detecting bowhead whale end notes where the well-defined time-frequency energy distribution of the end notes are exploited in designing a kernel function that matches the vocalizations. More recently, Urazghildiiev and Clark [165] proposed wavelet basis functions as kernels for automatically detecting North Atlantic right whale contact calls.

Alternatively, the matched filter of the LO detector can be replaced by a bank of lowpass and highpass decomposition filters used to implement the wavelet packet transform (WPT). Although the WPT filters are not optimal matched filters, the WPT has been shown to discriminate vocalization signals from background noise more

effectively compared to the discrete Fourier transform (DFT) in Section 4.2. However, the WPT is a non-redundant and a multi-rate sampling transform (due to the downsampling of the outputs following each filtering operation). Hence, the WPT cannot be implemented with the second order non-linearity $h_{\mathrm{LO}}(x)$ which does not change the sampling rate. The undecimated wavelet packet transform (UWPT) is proposed in this appendix as an unsupervised and robust generalized correlator for detecting weak manatee vocalizations. The UWPT is implemented by eliminating the downsampling that follows each filtering operation in the WPT and upsampling the filter coefficients at each level to account for the lack of downsampling. The redundant UWPT is translation invariant which also improves the consistency of the detector. The detection decision is based on the comparison between the maximum of the test statistics resulting from the UWPT filter bank and a pre-defined threshold. The UWPT based LO detector is depicted in Figure A-4.

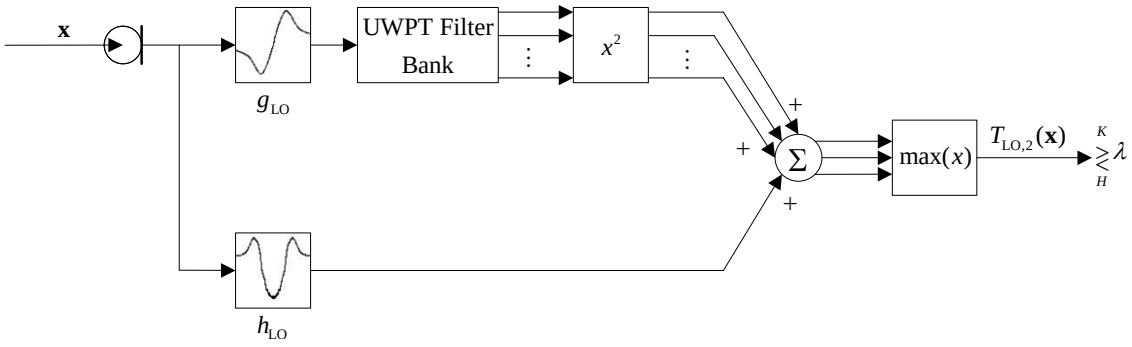


Figure A-4. The schematic of the proposed UWPT based LO detector.

The performance of the proposed detector is evaluated using the vocalizations collected in the vocalization library [28] and real ambient noise recordings made in

brackish waters at the mouth of Crystal River, FL to the Gulf of Mexico. As the salinity of the water increases, snapping shrimp noise dominates the ambient noise environment in the absence of watercraft. A plot of a typical ambient noise recording dominated with snapping shrimp is presented in Figure 3-6 (a). Several tests are conducted with different ambient noise recordings and using all 100 vocalizations of the vocalization library. The input SNR of the vocalizations is also varied between -10 dB to 5 dB. The detector performance is averaged over each test case and over all 100 vocalizations for each input SNR. The resulting receiver operator characteristics (ROC) curve at -4 dB input SNR and the detector performance curve for 1% probability of false detection are depicted in Figure A-5. If the input SNR is assumed to be -4 dB, the parametric LO detector is capable of achieving a %72 detection rate for a false detection rate of 0.1%. If a detection rate of 90% is desired, the false detection rate increases to 11%. The detection rate is estimated to be relatively constant about 94% for an input SNR in the range of -10 dB to -6 dB. The LO detector is optimized assuming that the vocalization signals are very weak. As the input SNR increases, the signal power parameter becomes much larger than the assumed value of $\theta_K \approx 0$, and the detector performance degrades. Therefore, another detector optimized for higher vocalization signal power (e.g., the harmonic detector proposed by Niezrecki et al. [20]) should be preferred for high input SNR values. In addition, based on the lower performance of the LO detector at high input SNR, it can be concluded that a UMP detector (optimum for both small and large θ_K) does not exist for snapping shrimp noise dominated environments that exhibit similar noise statistics to those tested in this Appendix.

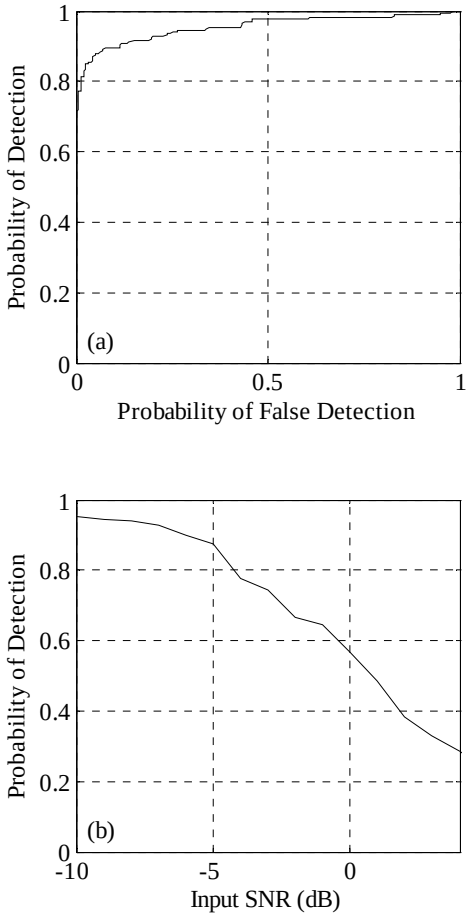


Figure A-5. Plots of the performance measures for the UWPT based LO detector. (a) The ROC for -4 dB input SNR of the UWPT based LO detector estimated from tests conducted with real vocalization and snapping shrimp noise recordings and (b) the corresponding performance curve of the detector as a function of the input SNR for a fixed false detection rate of 1 %.

The UWPT based LO detector developed in this section can also be implemented for detecting weak vocalization signals in the presence of watercraft emitted noise. Earlier in this Appendix, the matched filter or the generalized correlator was shown to be

the LO detector for zero-mean stochastic signals in IID Gaussian noise. However, since watercraft emitted noise exhibits temporal correlation, the covariance matrix of the noise signals must be estimated and used to pre-whiten the measurements prior to detection. The statistics of noise emitted from an approaching watercraft will change in time, and the inverse of the noise covariance matrix can be estimated recursively from the measurements [166].

A.3 Summary

In this appendix, a wavelet domain LO detector for detecting weak manatee vocalizations in snapping shrimp dominated noise environments are developed and evaluated using real recordings. The LO detector for manatee vocalizations are derived using the second derivative of the power function. This detector achieves the highest possible detection rate for a given false detection rate and for weak vocalization signals. The matched filter that requires a prior knowledge of the vocalization signals are replaced by a bank of filters used to implement the UWPT. The detector is shown to achieve a detection rate of approximately 94% for 1% false detection rate when the input SNR is in the range of -10 dB to -6 dB. The performance of the detector is also shown to degrade as the input SNR approaches 0 dB. A discussion regarding the extension of the proposed LO detector to detect weak vocalization signals in watercraft emitted noise environments is also presented.

APPENDIX B: THE LINEAR WAVE EQUATION IN CYLINDRICAL COORDINATES

For typical properties of seawater $(c = 1500 \text{ m/s})$, the linearized wave equation can be approximated as

$$\nabla^2 p(\mathbf{r},t) - \frac{1}{c^2(\mathbf{r})}\frac{\partial^2}{\partial t^2} p(\mathbf{r},t) = 0, \tag{B-1}$$

where $\mathbf{r} = (r,\psi,z)$ is the cylindrical coordinate vector, c is the speed of sound, and $p(\mathbf{r},t)$ is the acoustic pressure wave. The Laplacian operator for the cylindrical coordinate system is defined as

$$\nabla^2 = \frac{\partial^2}{\partial r^2} + \frac{1}{r}\frac{\partial}{\partial r} + \frac{1}{r^2}\frac{\partial^2}{\partial \psi^2} + \frac{\partial^2}{\partial z^2}. \tag{B-2}$$

In Eq. (B-1), the presence of the exciting point source that induces the acoustic pressure wave is ignored, which results in a homogenous partial differential equation (PDE). The velocity potential can be computed from the acoustic pressure from the relation

$$p(\mathbf{r},t) \approx -\rho\frac{\partial}{\partial t}\varphi(\mathbf{r},t), \tag{B-3}$$

where ρ is the density of seawater and $\varphi(\mathbf{r},t)$ is the acoustic particle velocity potential. Alternatively, the wave equation is commonly expressed in terms of the particle velocity potential. The frequency domain equivalent of the time domain wave equation presented in Eq. (B-1) is termed the Helmholtz equation

$$\nabla^2 p(\mathbf{r},\omega) + k^2(\mathbf{r})\, p(\mathbf{r},\omega) = 0, \tag{B-4}$$

where $k(\mathbf{r})$ is the complex wave number defined as

$$k(\mathbf{r}) = \frac{\omega}{c(\mathbf{r})}, \tag{B-5}$$

and $p(\mathbf{r},\omega)$ represents the frequency domain equivalent of the acoustic pressure wave [119-120].

The homogenous wave equation presented in Eq. (B-1) can be considerably simplified if the acoustic pressure is assumed to be a function of the radial coordinate and time only, and the speed of sound in water is assumed to be constant. For far-field propagation at a significant distance from the source, the linear wave equation in cylindrical coordinates reduces to

$$\frac{\partial^2}{\partial r^2} p(r,t) + \frac{1}{r}\frac{\partial}{\partial r} p(r,t) - \frac{1}{c^2}\frac{\partial^2}{\partial t^2} p(r,t) = 0, \tag{B-6}$$

where r is the radial coordinate, and c is the speed of sound. Defining $\eta = r^{1/2} p(r,t)$, Eq. (B-6) can be written in terms of $\eta(r,t)$ as

$$\frac{\partial^2}{\partial r^2}\eta(r,t) - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\eta(r,t) + \frac{\eta(r,t)}{4r^2} = 0, \tag{B-7}$$

where the last term in the left-hand-side can be neglected at the far-field. The solution to the resulting one-dimensional, planar wave equation is of the form

$$p(r,t) = r^{-1/2} f(r - ct), \tag{B-8}$$

where $f(r)$ is the waveform function of the outgoing wave [167]. The simple, frequency independent solution in Eq. (B-8) indicates that at the far-field, acoustic pressure decreases in proportion to the square root of the radial distance or range. Transmission loss associated with this simple spreading model is given as

$$\mathrm{TL}_{\mathrm{cyl}} = 10\log_{10}(r), \tag{B-9}$$

for $r \geq 1$. Alternatively, solving the one-dimensional wave equation in the spherical coordinates results in a solution of the form

$$p(r,t) = r^{-1} f(r - ct), \tag{B-10}$$

with a transmission loss given through

$$\mathrm{TL}_{\mathrm{sph}} = 20\log_{10}(r), \quad r \geq 1. \tag{B-11}$$

Mixed spreading is another simple spreading model which results in an intermediate transmission loss between cylindrical and spherical spreading models

$$\mathrm{TL}_{\mathrm{mxd}} = 15\log_{10}(r), \quad r \geq 1. \tag{B-12}$$

Alternatively, the acoustic pressure can be assumed to be a time-harmonic wave of the form

$$p(r,t) = P(r)\exp(j\omega t), \tag{B-13}$$

where $P(r)$ is the complex valued amplitude of the pressure wave. Substituting Eq. (B-13) into Eq. (B-6) results in the ordinary differential equation

$$\frac{d^2}{dr^2}P(r) + \frac{1}{r}\frac{d}{dr}P(r) + k^2 P(r) = 0, \tag{B-14}$$

Performing the necessary simplifications as outlined above and assuming that no incoming waves exists, the only solution to Eq. (B-14) is the outgoing wave of the form

$$P(r) = \frac{A}{r^{1/2}}\exp(-jkr), \tag{B-15}$$

where A is a complex valued constant which depends on the boundary conditions. Thus, the outgoing pressure wave can be expressed as

$$p(r,t) = \frac{A}{r^{1/2}}\exp[\,j(\omega t - kr)]. \tag{B-16}$$

Using Eq. (B-16), it is possible to obtain a crude estimate of the complex pressure field at a given frequency ω and range r.

APPENDIX C: DESCRIPTION OF THE TEST CASES

In this appendix, several test cases used throughout the dissertation for simulations and performance evaluation are described. First, the numerical channel impulse responses used to evaluate the performance of the wavelet domain denoising (WDD) methods in Chapter 4 is presented, followed by the test case used in Chapter 5 for blind source separation (BSS) methods. In the last section of this Appendix, the experimental setup of the Crystal River tests used in Chapter 6 is described.

C.1 The Wavelet Domain Denoising Test Case

The performance of the WDD algorithms proposed in Chapter 4 is evaluated through Monte-Carlo simulations. As described previously, watercraft emitted noise is modeled as an independent and identically distributed (IID) Gaussian sequence. However, the convolutive acoustic transmission channel introduces temporal correlation and watercraft emitted noise measured at the receivers. Hence, IID Gaussian noise is convolved with a numerical computed channel impulse response to generate a temporally correlated Gaussian signal which models watercraft emitted noise as measured at the receivers. For the sake of completeness, the Gaussian autoregressive (AR) signal which represent manatee vocalization is also convolved with a different numerical channel

impulse response model. The channel impulse response models are computed using the BELLHOP software which is based on the ray propagation theory and outputs an estimate of the complex pressure field at a given point. The channel impulse responses are obtained from the inverse Fourier transform of the complex valued acoustic pressure field. The geophysical parameters of the acoustic environment reported for dredged manatee habitats [146] are used in computing the channel impulse responses, and are presented in Table C-1. The manatee and watercraft are assumed to be 20 m and 102 m away from the receiver, respectively. The simulation setup is presented in Figure C-1.

Table C-1. The geophysical parameters used in the computation of the numerical transmission channel models.

BELLHOP Input Parameters	Value
Water Depth (m)	5.0
Speed of Sound (in water, m/s)	1500.0
Density (of water, g/cm^3)	1.02
Speed of Sound (in sediment, m/s)	1700.0
Density (of sediment, g/cm^3)	1.80
Attenuation (of sediment dB/(km Hz))	0.06
Depth of Source (m)	0.75
Depth of Receiver (m)	4.00

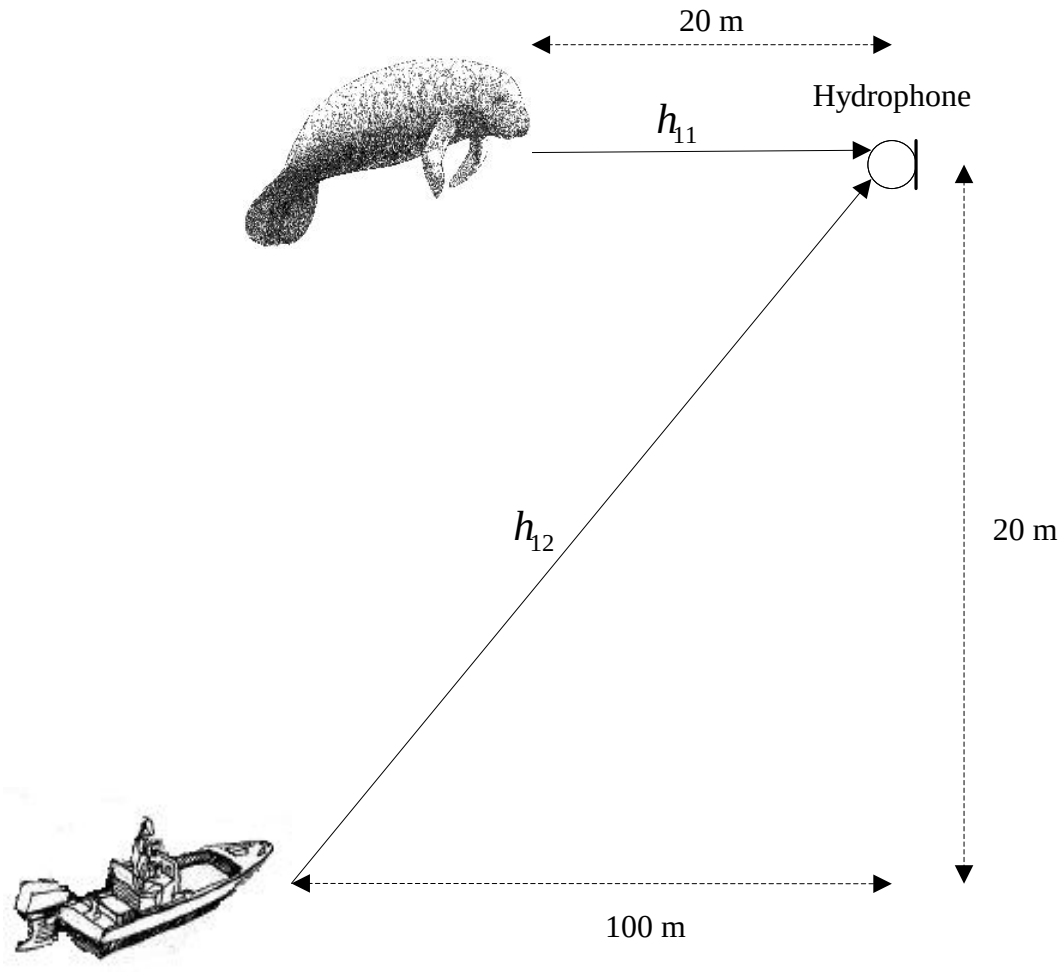


Figure C-1. The physical setup assumed for the Monte-Carlo simulations conducted to evaluate the proposed WDD algorithms.

The impulse responses that are convolved with the AR and IID Gaussian noise signal are presented in Figure C-2.

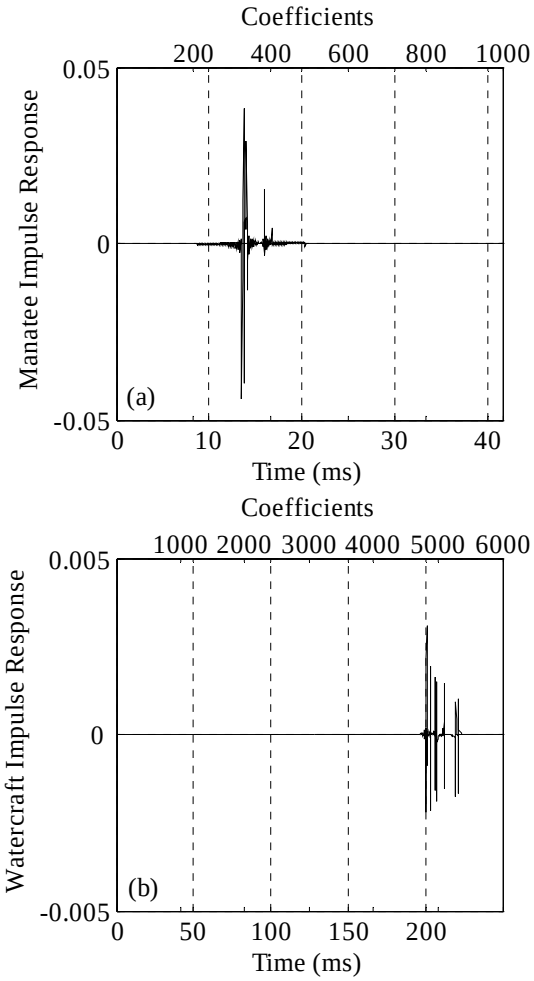


Figure C-2. The channel impulse responses used for the Monte-Carlo tests conducted to evaluate the proposed WDD algorithms. The channel impulse response used to convolve (a) the manatee vocalizations and, (b) the watercraft emitted noise recordings.

C.2 The Blind Source Separation Test Case

The performances of the proposed affine projection based BSS algorithms are evaluated by Monte-Carlo simulations using artificial vocalization and watercraft emitted noise recordings in Chapter 5. The two-input two-output (TITO) channel configuration used for these tests is presented in Figure C-3. As is done in the previous section for the WDD methods, the artificial vocalization and watercraft emitted noise signals are first convolved with the numerically computed impulse responses shown in Figure C-2. However, the BELLHOP generated impulse responses are generally very complex and the channel mixing matrix consists of very high order FIR filters. In general, it is very difficult to verify the stability, causality, and invertibility properties of this channel mixing matrix. For this reason, the pseudo channel impulse responses that are used for mixing the artificial signals is analytically computed as outlined in Section 2.5. The physical parameters used in the numerical impulse responses computations (presented in Table C.1) are also used for the analytic computations. The ranges at which the analytic impulse responses are computed are based on the difference between the ranges of the manatee and the watercraft to the receivers. This difference is computed as 5 m for the manatee paths and 4 m for the watercraft paths.

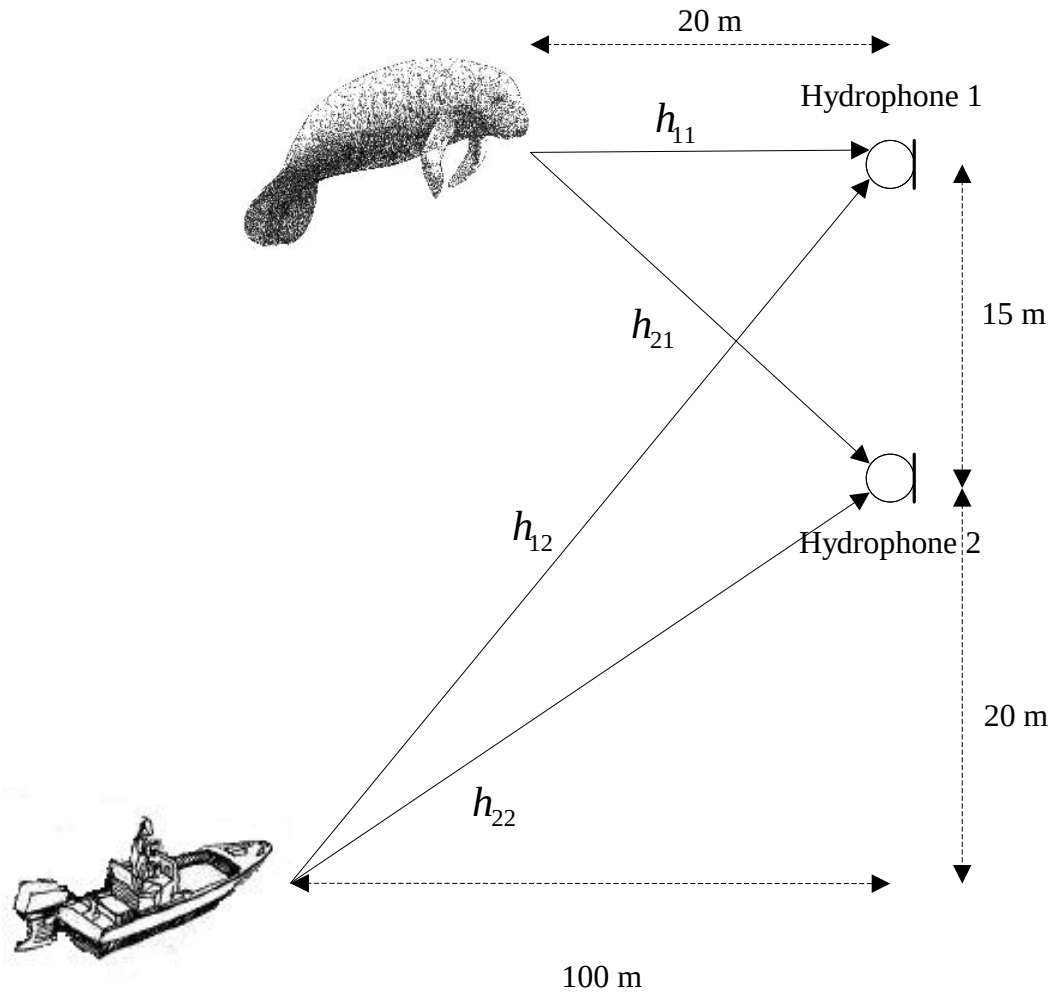


Figure C-3. The physical setup assumed for the Monte-Carlo simulations conducted to evaluate the proposed BSS algorithms.

C.3 Crystal River Experiments

The real watercraft emitted and ambient background noise recordings used in Chapter 6 were recorded on 14 and 15 August, 2007 using a 4-hydrophone linear array deployed in an active waterway along the river bank at Crystal River, FL. Crystal River is a populated area with numerous docks and heavy recreational watercraft traffic along

the main channel that provides access to the Gulf of Mexico. King Springs, just south of Banana Island, is a natural warm water refuge for manatees during the winter months when the surrounding water temperatures drop. Speed restrictions are enforced in Crystal River, and in particular, in the waterways surrounding King Springs. These aspects make Crystal River, FL an ideal experiment site. A map of Crystal River and the test site is presented in Figure C-4. The sky was clear and the waters were calm during the tests, resulting in low ambient natural background noise levels. Manatees generally migrate to cooler waters to the North during summer months, and only a single manatee was visually observed in the channel during these tests. However, no vocalizations were recorded. The linear array consisted of HTI 90-U (2 Hz to 20 kHz flat frequency response) and HTI-96 MIN (2 Hz to 30 kHz flat frequency response) hydrophones. For all watercraft emitted noise recordings, the distance between the hydrophones was constant and equal to 9.3 m. The hydrophones were placed approximately 1m above the river bed. The experimental setup is depicted in Figure C-5.

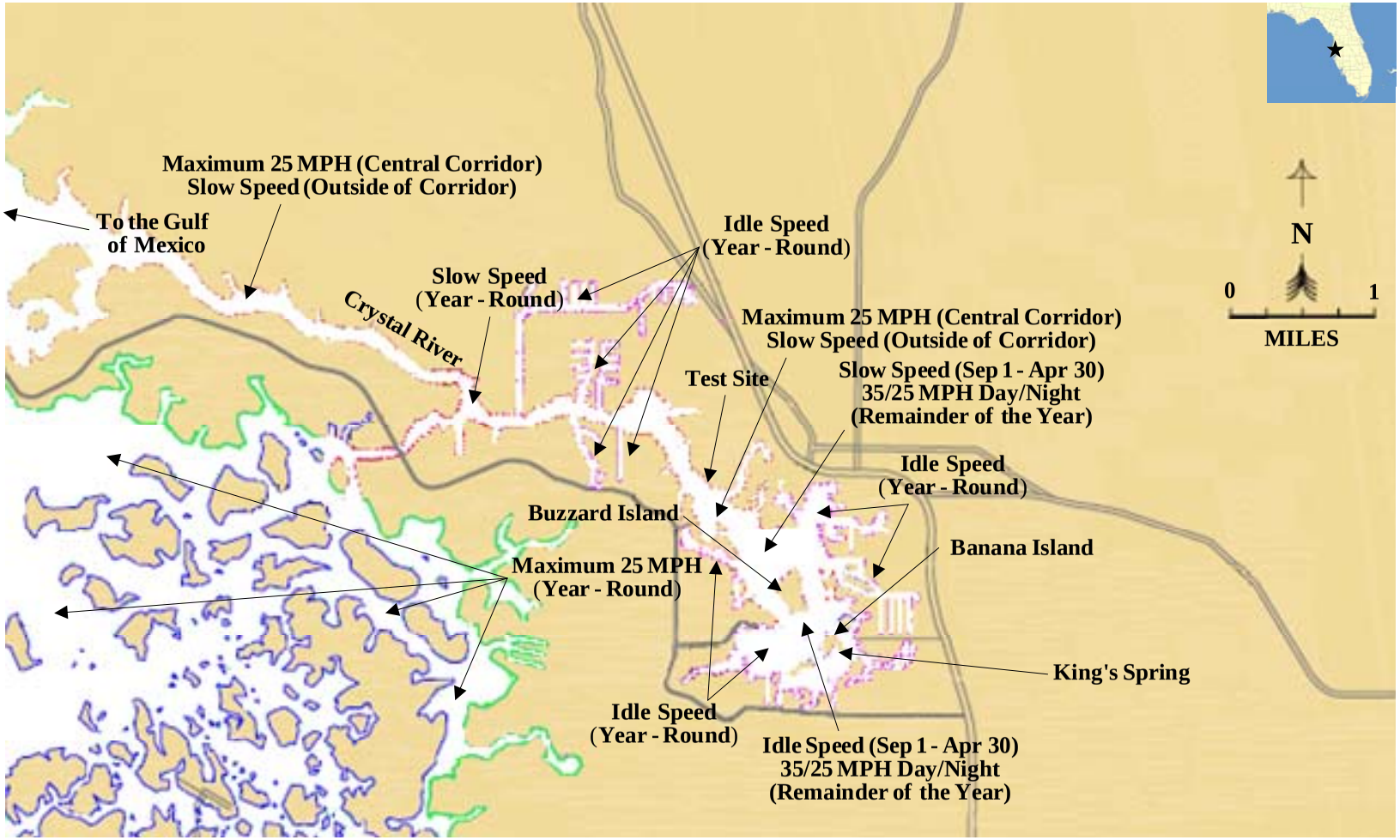


Figure C-4. The map of the Crystal River, FL test site[a].

[a] The map of Crystal River, FL is obtained from the Florida Fish and Wildlife Conservation Commission's (FWC) Marine Resources Geographic Information System which can be accessed through the URL http://ocean. floridamarine.org/mrgis/viewer.htm (last accessed: 9 Mar. 2008).

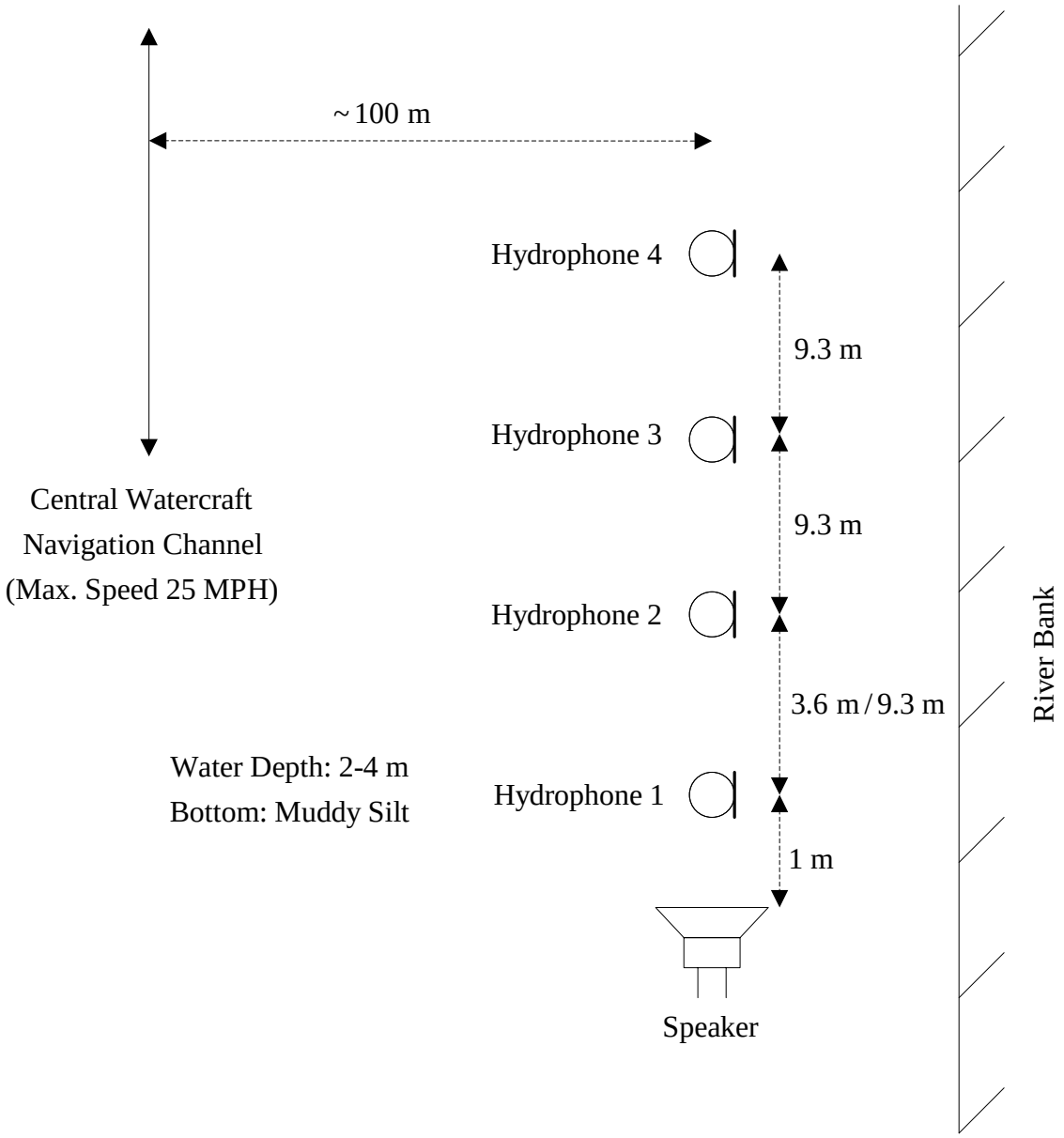


Figure C-5. The experimental setup for the Crystal River, FL test site.

Acoustic signals measured from watercraft navigating through a channel are separated into two regions [24]. As the watercraft departs away from the receivers, the transmission path between the propeller and the receivers is corrupted with cavitation bubbles that increase attenuation and significantly reduce the coherence between the

measurements at each channel. Therefore, watercraft emitted noise measurements recorded only for approaching watercraft is used in Chapter 6.

Aside from the background noise recordings, pseudo channel impulse responses used to convolve manatee vocalizations for the tests conducted in Chapter 6 are also estimated from broadcast tests conducted at the above location. For these tests, an IID Gaussian noise signal was broadcast using a Clark Synthesis AQ339 underwater speaker (frequency range 0.02 to 17 kHz). A single hydrophone was fixed at a distance of 1 m away from the underwater speaker and the remaining 3 hydrophones were placed at different distances from the speaker, including 4.6 m, 10.3 m, 13.9 m. Except for one set of tests which the separation between hydrophones 1 and 2 was 3.6 m, the 9.3 m separation between the hydrophones was maintained. However, for consistency with watercraft emitted noise recordings, the data recorded between hydrophones 1 and 2 when the separation was 3.6 m was not used to convolve manatee vocalizations. Although measurements were made at distances further away from the speaker than 13.9 m, the data collected at those distances were too noisy and incoherent to accurately estimate an impulse response.

The estimation of the pseudo channel impulse responses from the measurements is treated in the context of system identification. As was discussed in Chapter 2 and Chapter 5, the cross-coupling channel between the measurements is modeled as a finite impulse response (FIR) filter

$$x_2(n) = \sum_{k=0}^{L-1} h_k x_1(n-k) + v(n), \tag{C-1}$$

where $x_2(n)$ and $x_1(n)$ are the measurements at the two receivers, $v(n)$ is the component of $x_2(n)$ which cannot be explained with the FIR model (i.e., extraneous noise) and h_k, $k = 0,1,\ldots,L-1$ are the unknown time-invariant FIR filter coefficients. The channel impulse response is estimated adaptively using the time domain least mean squares (LMS) algorithm [168-169]. The use of the LMS algorithm was justified in this case because the channel transmission properties did not vary significantly throughout the duration of the broadcasts and the measurements did not exhibit significant temporal correlation (which would slow down the convergence of the LMS algorithm). The sampling rate was set to 24 kHz and broadcasts of 0.5 s durations were used to estimate a single realization of the channel impulse response. The impulse responses estimated in the time domain were compared to and verified with non-parametric, frequency domain estimates. The frequency domain estimates of the channel impulse response are computed from the inverse Fourier transform of the channel transfer function, which is obtained from the ratio of the auto- and cross-power spectral density (PSD) estimates of the measured signals at the hydrophones. The Welch periodogram method is used for computing consistent estimates of the PSD.

APPENDIX D: THE DERIVATION OF THE FEEDFORWARD STRUCTURED SEQUENTIAL-UPDATE DOUBLE AFFINE PROJECTION ALGORITHM

The derivation of the feedback structured, block-update double affine projection (FB/BU DAP) algorithm was presented in Section 5.2. In this Appendix, a similar derivation for the feedforward structured, sequential-update double affine projection (FF/SU DAP) algorithm is presented. The FF/SU DAP algorithm the solves for the separating filters by computing the minimum squared norm filter update

$$\min\left(\left\|\Delta\mathbf{w}_{pq}(n+1)\right\|_2\right)^2 = \min\sum_{l=0}^{L-1}\left(w_{pq,l}(n+1)-w_{pq,l}(n)\right)^2, \quad p\neq q=1,2, \qquad \text{(D-1)}$$

such that the updated filter decorrelates the past K lag cross-correlations between the intermediate filter outputs, where $K<L$ and L is the maximum order of the cross-channel filters. This criterion presents a constrained optimization problem that can be expressed in terms of Lagrange multipliers as

$$\mathcal{J}_1(n)=\left(\left\|\mathbf{w}_{12}(n+1)-\mathbf{w}_{12}(n)\right\|_2\right)^2+\sum_{k=0}^{K-1}\lambda_k r_{\tilde{y}_1\tilde{y}_2}(k), \qquad \text{(D-2)}$$

where $r_{\tilde{y}_1\tilde{y}_2}(k)$ are the cross-correlations and λ_k, $k=0,1,\ldots,K-1$ are the corresponding Lagrange multipliers. The cross-correlations are unknown and are estimated from their instantaneous realizations as

$$\hat{r}_{\tilde{y}_1\tilde{y}_2}(k) = \tilde{y}_1(n-k)\tilde{y}_2(n). \tag{D-3}$$

To derive the update rule for the first filter $\mathbf{w}_{12}(n)$, Eq. (D-2) is expressed as

$$\mathscr{J}_1(n) = \left(\left\|\mathbf{w}_{12}(n+1) - \mathbf{w}_{12}(n)\right\|_2\right)^2 + \left(\tilde{\mathbf{y}}_{1,K}^T(n)\tilde{y}_2(n)\right)\boldsymbol{\lambda}, \tag{D-4}$$

where $\boldsymbol{\lambda} = \begin{bmatrix}\lambda_0 & \lambda_1 & \cdots & \lambda_{K-1}\end{bmatrix}^T$ is the $K\times 1$ vector of Lagrange multipliers and the vector $\tilde{\mathbf{y}}_{1,L} = \begin{bmatrix}\tilde{y}_1(n) & \tilde{y}_1(n-1) & \cdots & \tilde{y}_1(n-L+1)\end{bmatrix}^T$ consists of the past K outputs computed with the updated filters $\mathbf{w}_{12}(n+1)$ such that

$$\tilde{\mathbf{y}}_{1,K}(n) = \mathbf{x}_{1,K}(n) - \mathbf{X}_2^H(n)\mathbf{w}_{12}(n+1), \tag{D-5}$$

where $\mathbf{X}_2(n) = \begin{bmatrix}\mathbf{x}_{2,L}(n) & \mathbf{x}_{2,L}(n-1) & \cdots & \mathbf{x}_{2,L}(n-K+1)\end{bmatrix}$ is a $L\times K$ matrix of the past filter inputs in channel 2 and $\mathbf{x}_{1,K}(n) = \begin{bmatrix}x_1(n) & x_1(n-1) & \cdots & x_1(n-K+1)\end{bmatrix}^T$ is the $K\times 1$ vector of inputs in channel 1. The cost function can be expanded by substituting $\tilde{\mathbf{y}}_{1,K}(n)$ as defined Eq. (D-5) into Eq. (D-4) as

$$\begin{aligned}\mathscr{J}_1(n) = &\left(\left\|\mathbf{w}_{12}(n+1) - \mathbf{w}_{12}(n)\right\|_2\right)^2 \\ &+\left(\mathbf{x}_1(n) - \mathbf{X}_2^T(n)\mathbf{w}_{12}(n+1)\right)^T \tilde{y}_2(n)\boldsymbol{\lambda}.\end{aligned} \tag{D-6}$$

Defining the $K\times 1$ cross-correlation vector

$$\hat{\mathbf{r}}_{\tilde{y}_2x_1} = \tilde{y}_2(n)\mathbf{x}_{1,K}(n) = \begin{bmatrix}\tilde{y}_2(n)x_1(n) & \tilde{y}_2(n)x_1(n-1) & \cdots & \tilde{y}_2(n)x_1(n-K+1)\end{bmatrix}^T, \tag{D-7}$$

and the $L\times K$ cross-correlation matrix

$$\begin{aligned}\hat{\mathbf{R}}_{\tilde{y}_2x_2} &= \tilde{y}_2(n)\mathbf{X}_2(n) \\ &= \begin{bmatrix} \tilde{y}_2(n)x_2(n) & \tilde{y}_2(n)x_2(n-1)) & \cdots & \tilde{y}_2(n)x_2(n-K+1) \\ \tilde{y}_2(n)x_2(n-1) & \tilde{y}_2(n)x_2(n-2) & \cdots & \tilde{y}_2(n)x_2(n-K) \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{y}_2(n)x_2(n-L+1) & \tilde{y}_2(n)x_2(n-L) & \cdots & \tilde{y}_2(n)x_2(n-K-L+2) \end{bmatrix},\end{aligned} \tag{D-8}$$

the cost function can be rewritten in the matrix notation as

$$\mathcal{J}_1(n) = \left(\left\| \mathbf{w}_{12}(n+1) - \mathbf{w}_{12}(n) \right\|_2 \right)^2 + [\hat{\mathbf{r}}_{\tilde{y}_2 x_1} - \hat{\mathbf{R}}^T_{\tilde{y}_2 x_2} \mathbf{w}_{12}(n+1)]^T \boldsymbol{\lambda}. \tag{D-9}$$

A gradient descent optimization method requires differentiating $\mathcal{J}_1(n)$ with respect to $\mathbf{w}_{12}(n+1)$ which results in

$$\frac{\partial}{\partial \mathbf{w}_{12}(n+1)} \mathcal{J}_1(n) = 2[\mathbf{w}_{12}(n+1) - \mathbf{w}_{12}(n)] - \mathbf{R}_{\tilde{y}_2 x_2} \boldsymbol{\lambda}. \tag{D-10}$$

Setting Eq. (D-10) to zero results in

$$\Delta \mathbf{w}_{12}(n+1) = \frac{1}{2} \mathbf{R}_{\tilde{y}_2 x_2} \boldsymbol{\lambda}. \tag{D-11}$$

Pre-multiplying both sides in Eq. (D-11) with $\mathbf{R}^T_{\tilde{y}_2 x_2}$ gives

$$\hat{\mathbf{R}}^T_{\tilde{y}_2 x_2} \Delta \mathbf{w}_{12}(n+1) = \frac{1}{2} \hat{\mathbf{R}}^T_{\tilde{y}_2 x_2} \hat{\mathbf{R}}_{\tilde{y}_2 x_2} \boldsymbol{\lambda}. \tag{D-12}$$

The updated filter $\mathbf{w}_{12}(n+1)$ is expected to decorrelate the filter outputs at the past K lags, which can be expressed in the matrix notation as

$$\begin{aligned} \hat{\mathbf{r}}_{\tilde{y}_2 x_1} &= \hat{\mathbf{R}}^T_{\tilde{y}_2 x_2} \mathbf{w}_{12}(n+1) \\ &= \hat{\mathbf{R}}^T_{\tilde{y}_2 x_2} [\Delta \mathbf{w}_{12}(n+1) + \mathbf{w}_{12}(n)], \end{aligned} \tag{D-13}$$

or, combining with Eq. (D-12)

$$\frac{1}{2} \hat{\mathbf{R}}^T_{\tilde{y}_2 x_2} \hat{\mathbf{R}}_{\tilde{y}_2 x_2} \boldsymbol{\lambda} = \hat{\mathbf{r}}_{\tilde{y}_2 x_1} - \hat{\mathbf{R}}^T_{\tilde{y}_2 x_2} \mathbf{w}_{12}(n). \tag{D-14}$$

Solving for the Lagrange multiplier vector from Eq. (D-14), one gets

$$\boldsymbol{\lambda} = 2 \left(\hat{\mathbf{R}}^T_{\tilde{y}_2 x_2} \hat{\mathbf{R}}_{\tilde{y}_2 x_2} \right)^{-1} \cdot \left(\hat{\mathbf{r}}_{\tilde{y}_2 x_1} - \hat{\mathbf{R}}^T_{\tilde{y}_2 x_2} \mathbf{w}_{12}(n) \right), \tag{D-15}$$

And upon substituting Eq. (D-15) into Eq. (D-11), the filter update is derived to be

$$\mathbf{w}_{12}(n+1) = \mathbf{w}_{12}(n) + \hat{\mathbf{R}}_{\tilde{y}_2 x_2} \left(\hat{\mathbf{R}}^T_{\tilde{y}_2 x_2} \hat{\mathbf{R}}_{\tilde{y}_2 x_2} \right)^{-1} \cdot \left(\hat{\mathbf{r}}_{\tilde{y}_2 x_1} - \hat{\mathbf{R}}^T_{\tilde{y}_2 x_2} \mathbf{w}_{12}(n) \right). \tag{D-16}$$

However, the update equation can further be simplified by rewriting Eq. (D-16) as

$$\begin{aligned}\mathbf{w}_{12}(n+1) &= \mathbf{w}_{12}(n)+ \tilde{y}_2(n)\mathbf{X}_2(n)\left(\tilde{y}_2(n)\mathbf{X}_2^T(n)\mathbf{X}_2(n)\tilde{y}_2(n)\right)^{-1} \\ &\quad \cdot\tilde{y}_2(n)\left(\mathbf{x}_{1,K}(n)-\mathbf{X}_2^T(n)\mathbf{w}_{12}(n)\right) \\ &= \mathbf{w}_{12}(n)+ \mathbf{X}_2(n)\left(\mathbf{X}_2^T(n)\mathbf{X}_2(n)\right)^{-1}\cdot\left(\mathbf{x}_{1,K}(n)-\mathbf{X}_2^T(n)\mathbf{w}_{12}(n)\right) \\ &= \mathbf{w}_{12}(n)+ \mathbf{X}_2(n)\left(\mathbf{X}_2^T(n)\mathbf{X}_2(n)\right)^{-1}\cdot\tilde{\mathbf{y}}_{1,K}(n).\end{aligned} \tag{D-17}$$

The update equation for the other filter w_{21} can be obtained similarly from the cost function

$$\mathcal{J}_2(n) = \left(\left\|\mathbf{w}_{21}(n+1)-\mathbf{w}_{21}(n)\right\|_2\right)^2 + \left(\tilde{y}_1(n)\tilde{\mathbf{y}}_{2,K}^T(n)\right)\boldsymbol{\lambda}, \tag{D-18}$$

where the cross-correlation is computed from the relation

$$\tilde{\mathbf{y}}_{2,K}(n) = \mathbf{x}_{2,K}(n)-\mathbf{X}_1^H(n)\mathbf{w}_{21}(n+1). \tag{D-19}$$

The resulting update rule is

$$\mathbf{w}_{21}(n+1) = \mathbf{w}_{21}(n)+\mathbf{X}_1(n)\left(\mathbf{X}_1^T(n)\mathbf{X}_1(n)\right)^{-1}\tilde{\mathbf{y}}_{2,K}(n). \tag{D-20}$$

As is done with the FB/BU DAP algorithm, the rate of convergence of the algorithm is controlled by introducing a step size constant μ into the update equation. Also, numerical problems that may be encountered during the computation of $\left(\mathbf{X}^T(n)\mathbf{X}(n)\right)^{-1}$ can be eliminated by introducing a regularization factor $\delta\mathbf{I}$. The resulting algorithm is the FF/SU DAP implemented in the time domain for TITO systems

$$\mathbf{w}_{pq}(n+1) = \mathbf{w}_{pq}(n)+ \mu\mathbf{X}_q(n)\left(\mathbf{X}_q^T(n)\mathbf{X}_q(n)+\delta\mathbf{I}\right)^{-1}\tilde{\mathbf{y}}_{p,K}(n), \quad p\neq q=1,2. \tag{D-21}$$

APPENDIX E: THE CLASSIFICATION OF MANATEE VOCALIZATIONS

A collection of 100 manatee vocalizations were grouped into 10 categories according to their time-frequency properties by Yan [28]. In Chapter 6, these vocalization signals are used to evaluate the performances of the proposed algorithms. The classification criteria of the vocalizations are provided in Figure E-1. A more detailed explanation and plots of these vocalization signals can be found in the dissertation of Yan [28].

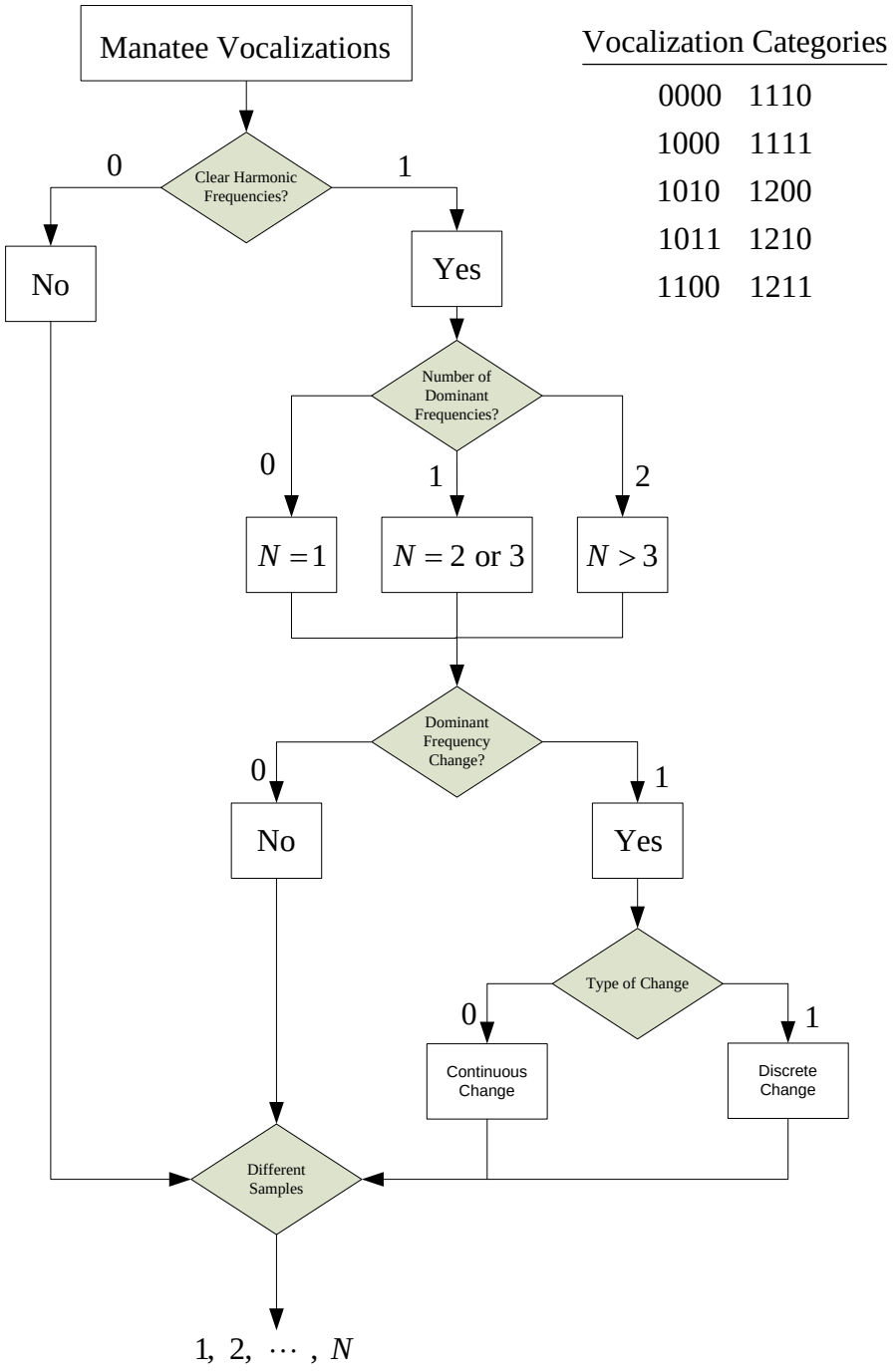


Figure E-1. The time-frequency structure based categorization of manatee vocalizations (adopted from Yan [28]).

LITERATURE CITED

[1] Calleson, C. S. and Frohlich, R. K. 2007. Slower boat speeds reduce risk to manatees. Endang. Species Res. 3: 295-304.

[2] Nowacek, S. M., Wells, R. S., Owen, E. C. G., Speakman, T. R., Flamm, R. O., and Nowacek, D. P. 2004. Florida manatees, *Trichechus manatus latirostris*, respond to approaching vessels. Biological Conservation 119: 517-523.

[3] FWS. 2001. Florida manatee recovery plan (*Trichechus manatus latirostris)*. Third revision. US Fish and Wildlife Service. 5430 Gorsvenor Lane, Suite 110 Bethesda, MD 20814. 193 pp.

[4] Gerstein, E. R., Gerstein, L. A., Forsythe, S. E., and Blue, J. E. 2001. It's all about SOUND science: Manatees, masking and boats. Proc. 142nd Meeting Acoust. Soc. Am. 110: 2722 (abstract only).

[5] Gorzelany, J. F. 2004. Evaluation of boater compliance with manatee speed zones along the Gulf Coast of Florida. Coastal Manage. 32: 215-226.

[6] USCG. 2006. 2006 Boating statistics. COMDTPUB P16754.20. US Coast Guard, 2100 Second Str. SW Washington, DC, 20593-0001. 47 pp.

[7] Murray, T. J. 2005. Florida's recreational marine industry-Economic impact and growth 1980-2005. Marine Industries Association of Florida, 6526 S. Kanner Highway #338 Stuart, FL 34997.

[8] FWC. 2007. Florida manatee management plan. Florida Fish and Wildlife Commission, 620 South Meridian Str. Tallahassee, FL 32399-1600. 280 pp.

[9] 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. FWC final report for PO No S-7701-619163. Available online at http://www.floridamarine.org/features/view_article.asp?id= 14362. Last access date: 3 Apr. 2007.

[10] Gerstein, E. R., Gerstein, L., E., F. S., and Blue, J. E. 1999. The underwater audiogram of the West Indian manatee (*Trichechus manatus*). J. Acoust. Soc. Am. 105(6): 3575-3583.

[11] Bowles, A. E. and Jaffe, J. S. 2004. Design for a manatee finder: Sonar techniques to prevent manatee-vessel collisions. FWC final report for PO No S-7701-617591. Available online at http://www.floridamarine.org/features/view_article.asp?id=14362. Last access date: 3 Apr. 2007.

[12] 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.

[13] Keith, E. O. 2002. Boater manatee awareness system. FWC 100 8th Ave., SE St. Petersburg, FL 33701. Available online at http://www.floridamarine.org/features/view_article.asp?id=14362. Last access date: 3 Apr. 2007.

[14] Keith, E. O. 2005. Development of an underwater infrared camera to detect manatees. FWC final report for contract FWC 03/04-28. Available online at http://www.floridamarine.org/features/view_article.asp?id=14362. Last access date: 14 Apr. 2008.

[15] Warner, W. R. (2003). Manatee proximity detector. FWC final report for contract FWC 01/02-14. Available online at http://www.floridamarine.org/features/view_article.asp?id=14362. Last access date: 3 Apr. 2007.

[16] 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). FWC final report for PO No S-7701-617592. Available online at http://www.floridamarine.org/features/view_article.asp?id=14362. Last access date: 3 Apr. 2007.

[17] Sermarini, C. A. 2005. Manatee vocalization detection method and system. US Patent No. 6879544. J. Acoust. Soc. Am. 118(3)-Pt.1: 1261.

[18] Mann, D. A., Nowacek, D. P., and Reynolds, J. I. 2002. Passive acoustic detection of manatee sounds to alert boaters. FWC final report for PO No S-7701-617553. Available online at http://www.floridamarine.org/features/view_article.asp?id=14362. Last access date: 3 Apr. 2007.

[19] Niezrecki, C. and Beusse, D. O. 2002. A system for warning boaters in the presence of manatees. FWC final report for PO No S-7701-617590. Available online at http://www.floridamarine.org/features/view_article.asp?id=14362. Last access date: 3 Apr. 2007.

[20] Niezrecki, C., Phillips, R., Meyer, M., and Beusse, D. O. 2003. Acoustic detection of manatee vocalizations. J. Acoust. Soc. Am. 114(3): 1640-1647.

[21] Meyer, S. M. 2003. Acoustic detection of manatee vocalizations. MS thesis, Dept. of Mechanical Eng., Univ. of Florida, Gainesville, FL. 85 pp.

[22] Niezrecki, C., Phillips, R. L., Meyer, M., and Beusse, D. 2005. Manatee warning system. US Patent No. 6845062. J. Acous. Soc. Am. 118(1):17.

[23] Frisch, K. and Haubold, E. 2003. Florida manatee avoidance technology: A pilot program by the Florida Fish and Wildlife Conservation Commission. J. Acoust. Soc. Am. 114(4)-Pt.2: 2415 (abstract only).

[24] 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.

[25] Widrow, B., Glover, J. R. Jr., McCool, J. M., Kaunitz, J., Zeidler, J. R., Dong, E. Jr., and Goodlin, R. C. 1975. Adaptive noise cancelling: Principles and applications. Proc. IEEE 63(12): 1692-1716.

[26] 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.

[27] 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.

[28] Yan, Z. 2006. Background noise cancellation for acoustic detection of manatee vocalizations. PhD dissertation, Dept. of Mechanical Eng., Univ. of Florida, Gainesville, FL. 205 pp.

[29] Zeidler, J., Satorius, E. H., Chabries, D. M., and Wexler, H. T. 1978. Adaptive enhancement of multiple sinusoids in uncorrelated noise. IEEE Trans. Acous. Speech and Sign. Processing ASSP-26(3): 240-254.

[30] Van Veen, B. and Buckley, K. M. 1998. Beamforming techniques for spatial filtering. In: Madisetti, V. K. and Williams, D. eds. The digital signal processing handbook. CRC Press, Boca Raton, FL, pp. 61/1-61/22.

[31] Owsley, N. L. 1984. Sonar array processing. In: Haykin, S. ed. Array signal processing. Prentice Hall, Englewood Cliffs, NJ, pp. 115-193.

[32] Frost, O. L. III. 1972. An algorithm for linearly constrained adaptive array processing. Proc. IEEE 60(8): 926-935.

[33] Krim, H. and Viberg, M. 1996. Two decades of array signal processing research: A parametric approach. IEEE Sign. Processing Mag. 13(4): 67-94.

[34] Muanke, P. B. and Niezrecki, C. 2007. Manatee position estimation by passive acoustic localization. J. Acous. Soc. Am. 121(40): 2049-2059.

[35] Swami, A. and Mendel, J. 1991. Cumulant-based approach to the harmonic retrieval and related problems. IEEE Trans. Sign. Processing 39(5): 1099-1109.

[36] Widrow, B., Duvall, K. M., Gooch, R. P., and Newman, W. C. 1982. Signal cancellation phenomena in adaptive antennas: Causes and cures. IEEE Trans. Antennas Propag. AP-30(3): 469-478.

[37] Vaidyanathan, P. P. 1990. Multirate digital filters, filter banks, polyphase networks, and applications: A tutorial. Proc. IEEE 78(1): 56-93.

[38] Mitra, S. K. 2006. Digital signal processing: A computer based approach. McGraw-Hill, New York. Ch. 14, pp. 835-843.

[39] Widrow, B., McCool, J. M., Larimore, M. G., and Johnson, C. R. Jr. 1976. Stationary and nonstationary learning characteristics of the LMS adaptive filter. Proc. IEEE 64(8): 1151-1162.

[40] Gray, R. M. 1972. On the asymptotic eigenvalue distribution of Toeplitz matrices. IEEE Trans. Inf. Theory IT-18(6): 725-730.

[41] Naranyan, S. S., Peterson, A. M., and Narasimha, M. J. 1983. Transform domain LMS algorithm. IEEE Trans. Acous. Speech, Sign. Processing ASSP-31(3): 609-615.

[42] Tewfik, A. H. and Kim, M. 1994. Fast positive definite linear system solvers. IEEE Trans. Sign.Processing 42(3): 572-585.

[43] Erdol, N. and Basbug, F. 1996. Wavelet transform based adaptive filters: Analysis and new results. IEEE Trans. Sign. Processing 44(9): 2163-2171.

[44] Doroslovacki, M. I. and Fan, H. 1996. Wavelet-based linear system modeling and adaptive filtering. IEEE Trans. Sign. Processing 44(5): 1156-1167.

[45] Hosur, S. and Tewfik, A. H. 1997. Wavelet transform domain adaptive FIR filtering. IEEE Trans. Sign. Processing 45(3): 617-630.

[46] Pradhan, S. S. and Reddy, V. U. 1999. A new approach to subband adaptive filtering. IEEE Trans. Sign. Processing 47(3): 655-664.

[47] Gilloire, A. and Vetterli, M. 1992. Adaptive filtering in subbands with critical sampling: Analysis, experiments, and application to acoustic echo cancellation. IEEE Trans. Sign. Processing 40(8): 1862-1875.

[48] Benesty, J., Gansler, T., Morgan, D. R., Sondhi, M. M., and Gay, S. L. 2001. Advances in network and acoustic echo cancellation. Springer-Verlag, Berlin. Ch. 1, pp. 1-31.

[49] Donoho, D. L. and Johnstone, I . M. 1994. Ideal spatial adaptation by wavelet shrinkage. Biometrika 81(3): 425-455.

[50] Donoho, D. L. 1995. De-noising by soft-thresholding. IEEE Trans. Inf. Theory 41(3): 613-627.

[51] Donoho, D. L. and Johnstone, I. M. 1998. Minimax estimation via wavelet shrinkage. The Annals of Statist. (26)3: 879-921.

[52] Chen, C. H., Lee, J. D., and Lin, M. C. 1998. Classification of underwater signals using wavelet transforms and neural networks. Math. Comput. Modeling 27(2): 47-60.

[53] Carevic, D. 2005. Adaptive window-length detection of underwater transients using wavelets. J. Acous. Soc. Am. 117(5): 2904-2913.

[54] Learned, R. E. and Willsky, A. S. 1995. A wavelet packet approach to transient signal classification. Appl. Comp. Harm. Anal. 2: 265-278.

[55] Farques, M. P. and Bennett, R. 1995. Comparing wavelet transforms and AR modeling as feature extraction tools for underwater signal classification. ASILOMAR Conf. Sign. Sys. 2: 915-919.

[56] 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. Sig. Processing 46(5): 1202-1207.

[57] Weiss, L. G. and Dixon, T. L. 1997. Wavelet-based denoising of underwater acoustic signals. J Acoust. Soc. Am. 101(1): 377-383.

[58] 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.

[59] 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.

[60] Douglas, S. 2002. Blind signal separation and blind deconvolution. In: Hu, Y. H. and Hwang, J. N. eds. Handbook of neural network processing. CRC Press, Boca Raton, FL, pp. 7/1-7/31.

[61] Mendel, J. 1995. Lessons in estimation theory for signal processing, communications, and control. Prentice Hall, Englewood Cliffs, NJ. App. B, pp. 450-472.

[62] Van Gerven, S. and Van Compernolle, D. 1995. Signal separation by adaptive symmetric decorrelation: Stability, convergence, and uniqueness. IEEE Trans. Sign. Processing 43(7): 1602-1612.

[63] Weinstein, E., Feder, M., and Oppenheim, A. V. (1993). Multi-channel signal separation by decorrelation. IEEE Trans. Speech Audio Processing 1(4): 405-413.

[64] Chan, D. B. 1997. Blind signal separation. PhD dissertation, Dept. of Engineering, Univ. of Cambridge, Cambridge, UK. 161 pp.

[65] Lindgren U. A. and Broman, H. 1998. Source separation using a criterion based on second-order statistics. IEEE Trans. Sign. Processing 46(7): 1837-1850.

[66] Parra, L. and Spence, C. 2000. Convolutive blind separation of non-stationary sources. IEEE Trans. Speech Audio Processing 8(3): 320-327.

[67] Parra, L. and Fancourt, C. 2002. An adaptive beamforming perspective on convolutive blind source separation. In: Davis, G. M. ed. Noise reduction in speech applications. CRC Press, Boca Raton, FL, pp. 15/1-15/17.

[68] Weinstein, E., Oppenheim, A. V., Feder, M., and Buck, J. R. 1994. Iterative and sequential algorithms for multisensory signal enhancement. IEEE Trans. Sign. Processing 42(4): 846-859.

[69] Olsson, R. K. and Hansen, L. K. 2004. A harmonic excitation state-space approach to blind separation of speech. Proc. Advances Neural Inf. Processing Sys. Conf. 17: 993-1000.

[70] Bradaric, I., Petropulu, A. P., and Diamantaras, K. I. 2003. On identifiability of FIR-MIMO systems with cyclostationary inputs using second order statistics. IEEE Trans. Sign. Processing 51(2): 434-441.

[71] Buchner, H., Aichner, R., and Kellerman, W. 2005. A generalization of blind source separation algorithms for convolutive mixtures based on second-order statistics. IEEE Trans. Speech Audio Processing 13(1): 120-134.

[72] Pedersen, M. S., Larsen, J., Kjems, U., and Parra, L. C. 2008. Convolutive blind source separation methods. In: Benesty, J., Sondhi, S. M., Huang, Y., eds. Springer handbook of speech processing. Springer, Berlin, pp. 1065-1095.

[73] Gaeta, M., Briolle, F., and Esparcieux, P. 1997. Blind separation of sources applied to convolutive mixtures in shallow water. Proc. IEEE Sign. Processing Workshop on Higher Order Statist., pp. 340-343.

[74] Bonnifay, S., Yao, K., and Jutten, C. 2000. Underwater acoustic signal separation based on prior estimation of the channel impulse response. Proc. ICASSP 5:3156-3159.

[75] Mansour, A., Benchekroun, N., and Gervaise, C. 2006. Blind separation of underwater acoustic signals. In: Rosca, J., Erdogmus, D., Principe, J. C., and Haykin, S. eds. Proc. ICA, Lecture Series on Computer Science, Springer, Berlin Vol. 3889, pp. 181-188.

[76] Aichner, R. 2007. Acoustic blind source separation in reverberant and noisy environments. PhD dissertation, University of Erlangen-Nuremberg, Germany. 275 pp.

[77] Yen, K. C. and Zhao, Y. 1999. Adaptive co-channel speech separation and recognition. IEEE Trans. Speech Audio Processing 7(2): 138-151.

[78] Tyack, P. L., Williams, W. J., and Cunningham, G. 1992. Time-frequency fine structure of dolphin whistles. Proc. IEEE-SP Int. Symp. pp. 17-20.

[79] Lammers, M. O. and Au, W. W. L. 2003. Directionality in the whistles of Hawaiian spinner dolphins (*Stenella longirostris*): A signal design feature to cue direction of movement? Marine Mamm. Sci. 19: 249-264.

[80] Zimmer, W. M., Tyack, P. L., Johnson, M. P., and Madsen, P. T. 2005. Three-dimensional beam pattern of regular sperm whale clicks confirms bent-horn hypothesis. J. Acous. Soc. Am. 117(3)-Pt. 1: 1473-1485.

[81] Schevill, W. E. and Watkins, W. A. 1965. Underwater calls of *Trichechus* (manatee). Nature 205(4969): 373-374.

[82] Evans, W. E., and Herald, E. S. 1970. Underwater calls of a captive Amazon manatee, *Trichechus inunguis*. J. Mamm. 51(4): 820-823.

[83] Sonoda, S. and Takemura, A. 1973. Underwater sounds of the manatees, *Trichechus manatus manatus* and *T. inunguis* (Trichechidae). Institute of Breeding Research, Tokyo Univ. of Agriculture, Report No: 1, pp. 19-24.

[84] Anderson, P. K., and Barclay, R. M. 1995. Acoustic signals of solitary dugongs: Physical characteristics and behavioral correlates. J. Mamm. 76(4): 1226-1237.

[85] Steel, C. 1982. Vocalization patterns and corresponding behavior of the West Indian manatee (*Trichechus manatus*). PhD dissertation, Dept. of Biological Sciences, Florida Inst. of Technology, Melbourne, FL. 189 pp.

[86] Richard-Clark, E. M. 1991. A vocalization comparison of the West Indian manatee in Florida. MS thesis, Dept. of Biological Sciences, Florida Inst. of Technology, Melbourne, FL. 61 pp.

[87] Hartman, D. S. 1979. Ecology and behavior of the manatee (*Trichechus manatus*) in Florida. The American Society of Mammalogists Special Publication, No: 5. 153 pp.

[88] Bengston, J. L. and Fitzgerald, S. M. 1985. Potential role of vocalizations in West Indian manatees. J. Mamm. 66(4):816-819.

[89] Nowacek, D., Casper, B. M., Wells, R. S., Mowacek, S. M., and Mann, D. A. 2003. Intraspecific and geographic variation of West Indian manatee (*Trichechus manatus spp*.) vocalizations. J. Acous. Soc. Am. 114(1): 66-69.

[90] 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.

[91] Sousa-Lima, R. S., Paglia, A. P., and Da Fonseca, G. A. B. 2002. Signature information and individual recognition in the isolation calls of Amazonian manatees, *Trichechus inunguis* (Mammalia: Sirenia). Animal Behavior 63: 301-310.

[92] O'Shea, T. J. and Poche, L. B. Jr. 2006. Aspects of underwater sound communication in Florida manatees (*Trichechus manatus latirostris*). J. Mamm. 87(6): 1061-1071.

[93] Gerstein, E., Gerstein, L., Forsythe, S. E., and Blue, J. E. 1999. The underwater audiogram of the West Indian manatee (*Trichechus manatus*). J. Acous. Soc. Am. 105(6): 3575-3583.

[94] Mann, D. and O'Shea, T. J. (2006). Nonlinear dymanics in manatee vocalizations. Marine Mamm. Sci. 22(3): 548-555.

[95] Atal, B., S. and Schroeder, M. R. 1978. Linear prediction analysis of speech based on a pole-zero representation. J. Acoust. Soc. Am. 64(5): 1310-1318.

[96] Morikawa, H. and Fujisaki, H. 1982. Adaptive analysis of speech based on a pole-zero representation. IEEE Trans. Acoust. Speech Sign. Processing ASSP-30(1): 77-88.

[97] Makhoul, J. 1975. Linear prediction: A tutorial review. Proc. IEEE 63(4): 561-580.

[98] Lim, J. S. and Oppenheim, A. V. 1978. All-pole modeling of degraded speech. IEEE Trans. Acoust. Speech Sign. Processing ASSP-26(3): 197-210.

[99] Tufts, D. W. and Kumaresan, R. 1982. Estimation of frequencies of multiple sinusoids: Making linear prediction perform like maximum likelihood. Proc. IEEE 70(9): 975-989.

[100] Matausek, M. R., Stankovic, S. S., and Radovic, D. V. 1983. Iterative inverse filtering approach to the estimation of frequencies of noisy sinusoids. IEEE Trans. Acoust. Speech Sign. Processing ASSP-31(6): 1456-1463.

[101] Kay, S. M. 1988. Modern spectral estimation: Theory and applications. Prentice Hall, Upper Saddle River, NJ. 576 pp.

[102] Hernandez-Perez, E., Nvarro-Mesa, J. L., and Millan_Munoz, M. J. 2004. An approach to maximum likelihood identification of autoregressive marine mammal sources by passive sonar. Proc. IEEE Int. Geoscience and Remote Sensing Symp. 2: 1435-1438.

[103] Hinich, M., Maradino, D., and Sullivan, E. J. 1989. Bispectrum of ship-radiated noise. J. Acoust. Soc. Am. 85(4): 1512-1517.

[104] Box, G. E. P. and Jenkis, G. M. 1976. Time series analysis, forecasting and control. Holden-Day, San Francisco. 592 pp.

[105] Pentek, A., Lennartsson, R. K., and Kadtke, J. B. 2000. Dynamical classification of hydroacoustic data from a Baltic Sea experiment. Proc. MTS/IEEE Oceans Conf. 3: pp. 1761-1767.

[106] Lennartsson, R. K., Kadtke, J. B., and Pentek, A. 2001. Passive signature characterization and classification by means of nonlinear dynamics. Proc. IEEE Sign. Processing Workshop, pp. 186-189.

[107] Lourens, J. G. and Du Preez, J. A. 1998. Passive sonar ML estimator for ship propeller noise. IEEE J. Oceanic Eng. 23(4): 448-453.

[108] AUTEC. 1993. Radiated noise report: Florida Marine Research Institute's 17 ft. MAKO 171. Naval Undesea Warfare Center, AUTEC Detachment West Palm Beach, FL. 68 pp.

[109] Ma, B. B., Nystuen, A., and Lien, R. C. 2005. Prediction of underwater sound levels from rain and wind. J. Acous. Soc. Am. 117(6): 3555-3565.

[110] Everest, F. A. and Young, R. W. 1948. Acoustical characteristics of noise produced by snapping shrimp. J. Acous. Soc. Am. 20(2): 137-142.

[111] Widener, W. 1967. Ambient-noise levels in selected shallow water off Miami, Florida. J. Acous. Soc. Am. 42(4): 904-905.

[112] Au, W. W. L. and Banks, K. 1998. The acoustics of the snapping shrimp *Synalpheus parneomeris* in Kaneohe Bay. J. Acous. Soc. Am. 103(1): 41-47.

[113] Silverman, B. W. 1986. Density estimation for statistics and data analysis. Chapman-Hall/CRC Press, Boca Raton, FL. Ch.3, pp. 34-74.

[114] Powell, D. R. and Wilson, G. R. 1984. Class A modeling of ocean acoustic noise process. In: Topics in non-Gaussian signal processing. Wegman, E. J., Schwartz, S. C., and Thomas, J. B. eds. Springer, New York. pp. 17-28.

[115] Middleton, D. 1977. Statistical-physical models of electromagnetic interference. IEEE Trans. Electromagnetic Compatibility EMC-19(3): 106-127.

[116] Bouvet, M. and Schwartz, S. C. 1988. Underwater noises: Statistical modeling, detection, and normalization. J. Acous. Soc. Am. 83(3): 1023-1033.

[117] Shao, M. and Nikias, C. 1993. Signal processing with fractional lower order moments: Stable processes and their applications. Proc. IEEE 81(7): 986-1010.

[118] Chitre, M. A., Potter, J. R., and Ong, S. H. 2006. Optimal and near-optimal signal detection in snapping shrimp dominated ambient noise. IEEE J. Oceanic Eng. 31(2): 497-503.

[119] Ziomek, L. J. 1995. Fundamentals of acoustic field theory and space-time signal processing. CRC Press, Boca Raton, FL. 720 pp.

[120] Jensen, F. B., Kuperman, W. A., Porter, M. B., Schmidt, H. 2000. Computational ocean acoustics. Springer, New York. 616 pp.

[121] Porter, B. and Bucker, H. P. 1987. Gaussian beam tracing for computing ocean acoustic fields. J. Acous. Soc. Am. 82(4): 1349-1359.

[122] Schobben, D., Torkkala, K., and Smaragdis, P. 1999. Evaluation of blind signal separation methods. Proc. ICA pp. 261-266.

[123] Antoniadis, A. and Fan, J. 2001. Regularization of wavelet approximations. J. Am. Statist. Assoc. 96(455): 939-967.

[124] Donoho, D. L. and Johnstone, I. M. 1994. Ideal spatial adaptation by wavelet shrinkage. Biometika 81(3): 425-455.

[125] Johnstone, I. M. and Silverman, B. W. 1997. Wavelet threshold estimators for data with correlated noise. J. Roy. Statist. Soc. B 59(2): 319-351.

[126] Donoho, D. L. and Johnstoen, I. M. 1995. Adapting to unknown smoothness via wavelet shrinkage. J. Am. Statist. Assoc. 90(432): 1200-1224.

[127] Coifman, R. R. and Donoho, D. L. 1995. Translation-invariant de-noising. In: Antoniadis, A. and Oppenheim, G., eds. Lecture notes in statistics Vol. 103. Wavelets and statistics. Springer, New York. pp. 125-150.

[128] 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 Sign. Proc. Lett. 3(1): 10-12.

[129] Berkner, K. and Wells, R. O. Jr. 2001. Denoising via nonorthogonal wavelet transforms. In: Debnath, L. ed. Wavelet transforms and time-frequency analysis. Birkhauser, Boston, pp. 68-80.

[130] Wickerhauser, M. V. 1994. Adapted wavelet analysis from theory to software. IEEE Press, New York. Ch. 8, pp. 273-299.

[131] Chapa, J. O. and Rao, R. M. 2000. Algorithms for designing wavelets to match a specified signal. IEEE Trans. Sign. Processing 48(12): 3395-3406.

[132] Caglar, H., Liu, Y., and Akansu, A. N. 1993. Optimal PR-QMF design for subband image coding. J. Visual Comm. Image Representation 4(3): 242-253.

[133] Krim, H., Tucker, D., Mallat, S., and Donoho, D. 1999. On denoising and best signal representation. IEEE Trans. on Inf. Theory 45(7): 2225-2238.

[134] Percival, D. B. and Walden, A. T. 2000. Wavelet methods for time series analysis. Cambridge University Press, New York. Ch. 10, pp. 393-457.

[135] Daubechies, I. 1992. Ten lectures on wavelets. SIAM, Philadelphia. Ch. 6, pp. 167-214.

[136] Ghael, S. P., Sayeed, A. M., and Baraniuk, R. G. 1997. Improved wavelet denoising via empirical Wiener filtering. Proc. SPIE 3169: 389-399.

[137] Zhang, X.-P. 2001. Thresholding neural network for adaptive noise reduction. IEEE Trans. Nueral Networks 12(3): 567-584.

[138] James, W. and Stein, C. 1961. Estimation with quadratic loss. Proc. 4th Berkeley Symp. on Math. Statist. Prob. 1: 361-379.

[139] Haykin, S. 2002. Adaptive filter theory. Prentice Hall, Upper Saddle River, NJ. Ch. 5, pp. 231-320.

[140] Bottomley, G. E. and Alexander, S. T. 1991. A novel approach for stabilizing recursive least squares filters. IEEE Trans. Sign. Processing 39(8): 1770-1779.

[141] Sayed, A. H. and Kailath, T. 1994. A state-space approach to adaptive RLS filtering. IEEE Sign. Processing Mag. 11(4): 18-60.

[142] Ozeki, K. and Umeda, T. 1984. An adaptive filtering algorithm using orthogonal projection to an affine subspace and its properties. Electron. Comm. Jpn. 67(5): 19-27.

[143] Gabrea, M. 2003. Double affine projection algorithm-based speech enhancement algorithm. Proc. ICASSP 1: 904-907.

[144] Caruthers, J. W. 1977. Fundamentals of marine acoustics. Elsevier Scientific Publishing, Amsterdam. Ch. 5, pp. 63-66.

[145] Antoniadis, A., Leporini, D., and Pesquet, J. C. 2002. Wavelet thresholding for some classes of non-Gaussian noise. Statistica Neerlandica 56(4): 434-453.

[146] Miksis-Olds, J. L. and Miller, J. H. 2006. Tranmission loss in manatee habitats. J Acoust. Soc. Am. 120(4): 2320-2327.

[147] Mallat, S. and Hwang, W. L. 1992. Singularity detection and processing with wavelets. IEEE Trans. Inf. Theory 38(2): 617-643.

[148] Donoho, D. L. 1993. Unconditional bases are optimal bases for data compression and statistical estimation. Appl. Comp. Harm. Anal. 1: 100-115.

[149] Olhede, S. C. and Walden, A. (2004). ‘Analytic’ wavelet thresholding. Biometrika 91(4): 955-973.

[150] Rao, A. M. and Jones, D. L. 2000. A denoising approach to multisensory signal enhancement. IEEE Trans. Sign. Processing 48(5): 1225-1234.

[151] Michalopoulou, Z. H. 1997. Underwater transient signal processing: Marine mammal identification, localization, and source signal deconvolution. Proc. ICASSP 1: 503-506.

[152] Benesty, J. Duhamel, P., and Grenier, Y. 1996. A multichannel affine projection algorithm with applications to multichannel acoustic echo cancellation. IEEE Sign. Processing Lett. 3(2): 35-37.

[153] Montazeri, M. and Duhamel, P. 1995. A set of algorithms linking NLMS and block RLS algorithms. IEEE Tran. Sign. Processing 43(2): 444-453.

[154] Gay, S. L. and Tavathia, S. 1995. The fast affine projection algorithm. Proc. ICASSP 5: 3023-3026.

[155] Tanaka, M., Makino, S., and Kojima, J. 1999. A block exact fast affine projection algorithm. IEEE Trans. Speech Audio Processing 7(1): 79-86.

[156] Kassam, S. A. 1988. Signal detection in non-Gaussian noise. Springer-Verlag, New York. 234 pp.

[157] Tsihrintzis, G. A. and Nikias, C. L. 1995. Performance of optimum and suboptimum receivers in the presence of impulsive noise modeled as an alpha-stable process. IEEE Trans. Comm. 43(2/3/4): 904-914.

[158] Gibson, J. D. and Melsa, J. L. 1995. Introduction to nonparametric detection with applications. IEEE Piscataway, NJ. Ch. 2, pp. 9-37.

[159] Lehmann, E. L. 1959. Testing statistical hypothesis. John Wiley, New York. Ch. 3, pp. 60-124.

[160] Ferguson, T. S. 1967. Mathematical statistics: A decision theoretic approach. Academic Press, New York. Ch. 5, pp. 206-212.

[161] Nikias, C. L. and Shao, M. 1995. Signal processing with alpha-stable distributions and applications. Wiley, Hoboken, NJ. Ch. 2, pp. 13-30.

[162] McCulloch, J. H. 1998. Numerical approximation of the symmetric stable distribution and density. In: Adler, R., Feldman, R., and Taqqu, M. eds. A practical guide to heavy tails: Statistical techniques for analyzing heavy tailed distributions. Birkhauser, Boston. pp. 489-499.

[163] Urazghildiiev, I. R. and Clark, C. W. 2006. Acoustic detection of North Atlantic right whale contact calls using the generalized likelihood ratio test. J. Acous. Soc. Am. 120(4): 1956-1963.

[164] Mellinger, D. K. and Clark, C. W. 2000. Recognizing transient low-frequency whale sounds by spectrogram correlation. J. Acous. Soc. Am. 107(6): 3518-3529.

[165] Urazghildiiev, I. R. and Clark, C. W. 2007. Acoustic detection of North Atlantic right whale contact calls using spectrogram-based statistics. J. Acous. Soc. Am. 122(2): 769-776.

[166] Cichocki, A. and Amari, S. 2002. Adaptive blind signal and image processing. Wiley, Hoboken, NJ. Ch. 3, pp. 87-129.

[167] Blackstock, D. 2000. Fundamentals of physical acoustics. Wiley, New York. Ch. 1, pp. 1-64.

[168] Ljung, L. 1999. System identification. Prentice Hall, Englewood Cliffs, NJ. Ch. 11, pp. 361-398.

[169] Sayed, A. H. 2003. Fundamentals of adaptive filtering. Wiley, Hoboken, NJ. Ch. 5, pp. 212-280.

BIOGRAPHICAL SKETCH OF THE AUTHOR

Berke M. Gür was born in Izmir/Turkey in 1978. He received his B.S. degree in Mechanical Engineering in 1999 from the Middle East Technical University in Ankara/Turkey. From 1999 to 2002, he worked as a HVAC project engineer. He received his M.S. degree, also in Mechanical Engineering, from the University of Southern California in Los Angeles, CA in 2003 with a concentration in Dynamic Systems and Controls. In 2004, he started his Ph.D. studies at the Department of Mechanical and Aerospace Engineering, University of Florida in Gainesville, FL and joined the Smart Structures and Acoustics Laboratory as a research assistant. He later moved to the University of Massachusetts Lowell in Lowell, MA following the relocation of his lab, where he was employed as a research assistant in the Structural Dynamics and Acoustic Systems Laboratory. During his Ph.D. studies, his research was focused on underwater acoustic signal processing, wavelets, and blind source separation. At the University of Massachusetts Lowell, he received the Graduate Scholar Research Award and served on the Executive Committee of the Graduate Student Association. While working towards his Ph.D. degree, he also received his M.B.A. degree from Boğaziçi University at Istanbul/Turkey in 2006. In 2008, he was awarded the first place in the Acoustic Society of America, Signal Processing Student Challenge. He is a member of the ASME, IEEE, ASA, and Sigma-Xi.