IMPACTS OF ENVIRONMENT-DEPENDENT ACOUSTIC PROPAGATION ON PASSIVE ACOUSTIC MONITORING OF CETACEANS

by

Carolyn M. Binder

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy

at

Dalhousie University
Halifax, Nova Scotia
July 2017

For my family. I couldn't have done this without your support, dedication, and love.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Significant effort has been made over the last few decades to develop automated passive acoustic monitoring (PAM) systems capable of classifying cetaceans at the species level; however, these systems often require tuning when deployed in different environments. Anecdotal evidence suggests that this requirement to adjust a PAM system's parameters is partially due to differences in the acoustic propagation characteristics. The environment-dependent propagation characteristics create variation in how a cetacean vocalization is distorted after it is emitted. If these difference are not accounted for it could reduce the performance of automated PAM systems. An aural classifier developed at Defence R&D Canada (DRDC) has been used successfully for inter-species discrimination of cetaceans. Accurate results are obtained by using perceptual signal features that model the features employed by the human auditory system. In this thesis, a combination of an at-sea experiment and simulations with modified bowhead and humpback whale vocalizations was conducted to investigate the robustness of the classifier performance to signal distortion as a function of propagation range. It was found that in many environments classification performance degraded with increasing range, largely due to decreased signal-to-noise ratio (SNR); however, in some environments as much as 40 % of the performance reduction was attributed to signal distortion resulting from environment-dependent propagation. It was found that sound speed profiles resulting in considerable boundary interaction were important for producing sufficient signal distortion to affect PAM performance, relative to the impacts of SNR. Therefore, in some environments the ocean acoustic properties should be taken into account when characterizing performance of automated PAM systems. For the environments in which signal-to-noise issues dominate, the use of multi-element arrays is expected to increase the performance of automated recognition systems beyond the minor improvements to be gained from adjusting a PAM system's parameters. Nonetheless, propagation modelling should be used to complement PAM experiments to account for bias in probability of detection estimates resulting from environment-dependent acoustic propagation.

List of Abbreviations and Symbols Used

Abbreviation	Description
AN	Additive noise
AWGN	Additive white Gaussian noise
CFAV	Canadian Forces Auxiliary Vessel
CTD	Conductivity-temperature-depth
CW	Continuous wave
D.R.	Downward refracting
DRDC	Defence Research and Development Canada
EOF	Empirical orthogonal function
FFCPT	Free-falling cone penetrometer
FFP	Fast field program
FFT	Fast Fourier transform
h/p	Hydrophone
LFM	Linear frequency modulated
OASES	Ocean Acoustics and Seismic Exploration Synthesis
OASES-OASP	Pulse propagation module of OASES
ONR	Office of Naval Research
PAM	Passive acoustic monitoring
pdf	Probability density function
PE	Parabolic equation
ROC	Receiver operating characteristic
RMS	Root-mean-square
SAFARI	Seismo-Acoustic Fast field Algorithm for Range-Independent environments
SHARP	Subsurface high-fidelity audio recording package
SSP	Sound speed profile
VIM	Vibration Isolation Module

Abbreviation	Description
WATTCH	WAveform Transmission Through a CHannel

Symbol	Description	Units
AUC	Area under receiver operating characteristic curve	
B	Signal bandwidth	Hz
c	Sound speed	m/s
c_p	Compressional sound speed	m/s
C	Counts	
$\mathcal{C}$	Capacity of a signal transmission channel	dB
CTS	Conversion factor to convert from bits to voltage	dB
D	Decision	
f	Frequency	Hz
fp	False positive	
h	Water depth	m
$h_t(R, z)$	Channel impulse response	
L	Signal length	samples
k	Number of disjoint sets for cross-validation	
m	Number of array elements	
n	Number of samples/signals	
n_t	Noise time series	
n	Negative decision	
N	Total number of negative samples	
N	Number of samples in time window	
NL	Ambient noise level	dB re 1 μPa
p	Positive decision	
$p(R, z, \omega)$	Spatial transfer function of the acoustic field	Pa
P	Total number of positive samples	
r_1	Source-receiver separation	m
R	Range	m, km
R^2	Coefficient of determination for linear regression	

Symbol	Description	Units
RL	Received sound pressure level	dB re 1 µPa
S	Salinity	PSU
$s(t)$	Time series representation of a sinc function	
$\mathcal{S}_H$	Hydrophone sensitivity	dB re 1 V/µPa
$\mathcal{S}_V$	Transmit sensitivity	dB re 1 µPa/V @ 1 m
S	Average signal power	dB
S_p	Power of WATTCH-propagated signal	
$S(\omega)$	Source spectrum	Pa^2/Hz
SL	Source level	dB re 1 µPa @ 1 m
SNR	Signal-to-noise ratio	dB
t	Time	s
T	Temperature	°C
T	Length of time window	s
TL	Transmission loss	dB
TN	Number of correctly identified negative samples	
TP	Number of correctly identified positive samples	
tp	True positive	
TVR	Transmitting voltage response	dB re 1 µPa/V @ 1 m
u	Weight randomly drawn from uniform distribution	
V_{in}	RMS voltage at transducer input	V
V_{p}	Peak voltage	V
V_{RMS}	RMS voltage	V
W	Average white noise power	dB
x_i	The i^{th} signal sample	
x_t	Time series of signal (without transformation by noise or propagation)	
$x_{t,p}$	WATTCH-simulated time series	
$\mathcal{X}$	Continuous variable for decision process	
$y_{t,n}$	Time series comprised of signal and noise	
$y_{t,p}$	Time series comprised of simulated signal	

Symbol	Description	Units
$y_{t,r}$	Time series comprised of simulated signal and noise	
z	Depth	m
z_r	Receiver depth	m
z_s	Source depth	m
α	Compressional attenuation	dB/λ_p
β	Weight used to scale noise vector to achieve desired SNR	
λ	Wavelength	m
λ_p	Wavelength of compressional wave	m
μ_{SL}	Mean source level	dB re 1 µPa @ 1 m
$\mu_{\mathrm{SNR},R}$	Mean SNR for a particular range	dB
ρ	Density	g/cm^3
$\sigma_{accuracy}$	Standard deviation of accuracy	%
σ_{AUC}	Standard deviation of AUC	
σ_n^2	Noise variance	
σ_s^2	Signal variance	
σ_{s+n}^2	Combined signal and noise variance	
σ_{SL}	Standard deviation of source level	dB re 1 µPa @ 1 m
σ_{SNR}	Standard deviation of SNR	dB
$\sigma_{\mathrm{SNR},R}^2$	Variance of SNR at a particular range	dB^2
τ	Decision threshold value	
ω	Angular frequency	rad

ACKNOWLEDGEMENTS

I would first like to extend a heartfelt thank you to my supervisor, Paul Hines. You have been an incredible guiding influence in much of my academic and professional life; not the least of which was the suggestion that I undertake this PhD. Your wisdom even spilled over into my personal life. I always appreciated your ability to recognize the importance of striking an appropriate work-life balance (never more than when my family grew to include a daughter). Your support and guidance have been instrumental in my success as a student and as a scientist. You have always dedicated time in your hectic schedule for me, and I knew that you were always in my corner. I cannot thank you enough. I hope to continue to receive your wisdom as I move on to my next challenge.

I would also like to acknowledge Tetjana Ross who was my co-supervisor for the first three years of my PhD — thank you for the advice and insights you provided. A big thank you to Dave Barclay for taking over as my co-supervisor for the latter part of my PhD; your enthusiasm for ocean acoustics is always inspiring. I truly look forward to future collaborations. I would like to express my gratitude to the remaining members of my supervisory committee — Alex Hay and Christopher Taggart — for the lively discussions, advice, and suggestions for improving my thesis. My external examiner, Gerald D'Spain, deserves special acknowledgment for his generosity in reviewing my thesis, and taking the time to come all the way from Scripps for my thesis defence; thank you so much, Gerald! Your comments and suggestions were very helpful, not only for this thesis, but also for future publications.

Many thanks to my colleagues at DRDC. Dan Hutt, your career guidance and continued support for my PhD program have been invaluable. Without the help of Sean Pecknold, Modelling Guru, I would have been lost many times. Thanks Sean for our many conversations on acoustics modelling, your generosity, and your friendship. I would like to acknowledge Cristina Tollefsen's dual role in providing oceanography/ocean acoustics advice and her initiative for organizing HCC events. Many thanks to the coffee crew for both science and non-science discussions, invariably ending in laughter.

I would be remiss if I didn't acknowledge my funding sources. First and foremost, DRDC, for allowing me time to complete my degree, providing opportunities to attend

conferences, and financial support. Additionally, I would like to recognize funding from the US Office of Naval Research, an NSERC PGS-D, an honorary Killam scholarship, and the 1-year NSGS (for funding in my final year).

On an even more personal note, I could not have undertaken this PhD without the unwavering support of my family. Mom and Dad, you have always encouraged me. Your support went right from the very beginning, and included developing my scientific curiosity, to the present, by taking Lorelei on weekends so that I could focus on finishing my thesis. I hope to be as supportive and encouraging of a parent as you both are. Markus, what can I even say? You have been my everything: my best friend, my encouragement, my support system, and generally the best husband I could hope for. Thank you for your encouraging words whenever I needed to hear them. Your willingness to step up and take on the majority of our household responsibilities these last few months has been such a relief to me — I owe you. I couldn't have done this without you by my side. Finally, to Lorelei, the cutest little distraction, I love you to the moon and back. Your smiles and silliness have been the perfect antidote to thesis-related stress. I wish all the best for you, and hope that your ability to follow your curiosity knows no limits.

CHAPTER 1

INTRODUCTION

1.1 Literature Review

Passive acoustic monitoring (PAM) is widely used to study cetaceans in their natural habitats [1, 2, 3]. The PAM process has become increasingly automated since it has been recognized that there is a requirement for automated detection and classification methods to deal with large volumes of data [1, 3]. The utility of automated PAM systems is dependent on their ability to operate across a wide range of environmental conditions [4]. Since cetaceans are found in all ocean basins, their habitats span diverse underwater environments. For example, PAM has been used successfully in the tropical waters off the Brazilian coast to study humpback whale behaviour [3]; in the temperate waters of the Gulf of Maine for real-time detection of three baleen whale species from ocean gliders [5]; and in the polar waters of the Beaufort Sea to assess the impact of seismic air gun surveys on migrating bowhead whales [6]. PAM of even a single species can be complicated by the numerous environments covered by their distribution range. For example, humpback whales have a truly global distribution — they are found in tropical, temperate, and sub-polar waters worldwide [7]. Humpback whales in the northern hemisphere spend the summer on feeding grounds in the biologically productive areas of higher latitudes and migrate southwards in the winter to sub-tropical and tropical waters to breed and calve [8, 9]; thus, an individual humpback whale will encounter many different ocean regions during its annual migration.

Properties of the ocean environment — such as sound speed profile (SSP), bathymetry, sediment properties, and ambient noise characteristics — can be markedly different

among regions where PAM is used to observe cetaceans. This results in environment-dependent sound propagation characteristics [10, 11, 12, 13] and leads to differences in how a cetacean vocalization is distorted as it propagates along the source-receiver path. Environmental properties also vary over the spatial scales (tens to hundreds of kilometres) [1, 14, 15] at which many cetacean species can be acoustically detected. This means that the acoustic environment must often be considered as range-dependent. In addition to spatial variability, temporal variability also plays an important role in the ocean environment. For example, temporal variability occurs due to diurnal and seasonal changes of the SSP in the upper ocean [10, 11, 13, 16]. All of these factors combine to produce a complex, dynamic environment that affects propagating acoustic signals. This sentiment is well-summarized by Walter Munk's comment, "The very essence of ocean acoustics is its inherent variability [13, p. xiii]."

It is well known that properties of the ocean environment distort sound propagating through the ocean medium [10, 11, 12, 16, 17]. Signal distortion arises due to differences in travel times along different propagation paths between the source and receiver (referred to as *multipath* or time-spreading) [10]. This causes a signal's phase and amplitude to vary in a complex manner [13]. Urick [10] notes that, "it may not be an exaggeration to say that the effects of multipaths in the sea are deleterious to the detection of long-range sonar targets as well as in many short-range applications." In terms of PAM of cetaceans, Helble *et al.* [18] state, "The ocean bottom properties, bathymetry, and temporally varying sound speed act to distort and reduce the energy of the original waveform produced by the marine mammal. In addition, constantly varying ocean noise further influences the detectability of the calls. This ever-changing acoustic environment creates difficulties when comparing marine mammal recordings between sensors, or at the same sensor over time."

Since many PAM systems rely on the time-frequency characteristics of the vocalizations for detection and classification [1, 4, 19, 20, 21, 22, 23, 24], signal distortion has the potential to negatively affect the accuracy of PAM systems [10, 22, 25]; however, little research has been directed towards this problem. Some authors acknowledge that propagation effects likely impact the accuracy of PAM systems, but do not analyze how their system is affected. Thode *et al.* [6] discuss how waveguide effects, resulting from a shallow water environment, substantially alter the time-frequency characteristics of air gun signals and increase the false alarm rate in their automatic detection/classification of bowhead

whale vocalizations. Mouy *et al.* [23] state that the performance of automatic detection and classification methods is dependent on the acoustic properties of the environment. They note that fin and blue whale calls often appear distorted and 'warped' in time and that this degrades the accuracy of detection and classification algorithms. To test the robustness of their processing and classification methods, the authors chose calls that represented different noise and multipath conditions and ensured that data for training and testing the classifier were recorded at different locations and times. While this method provides a preliminary indication of the capability of their method to accommodate signal distortion due to propagation, they do not explicitly consider how changes in the acoustic properties impact their system. At the end of their paper, Zimmer *et al.* [26] conclude that, "Because the performance of passive acoustic detection depends heavily on environmental conditions, it must be implemented carefully to compensate, at least partially, for the effect of these conditions."

One of the few studies that included transmitting signals similar to whale vocalizations was conducted by Mercado *et al.* [27]; however, their research was not focused on PAM, but rather on investigating if a listening whale can use the propagation characteristics of its habitat for information to estimate the range to a vocalizing whale. Using the spectral information of the transmitted signals, they were able to train a neural network to recognize the range ('near' or 'far' and as originating from one of 10 range steps) over which the signals had propagated. Their results indicate that the time-frequency characteristics are substantially changed by propagation effects, which the authors conclude indicates that whales can use the signal distortion as a clue for estimating the range over which the vocalizations travelled. This further suggests that the information content in a vocalization is modified by the ocean acoustic environment, which inevitably will affect features extracted from the vocalizations for automated recognition.

In some cases, simple propagation effects are taken into account during the detection process by incorporating geometrical spreading and/or frequency-dependent attenuation (e.g., in Zimmer *et al.* [26] and Samaran *et al.* [28]); however, little has been done to investigate this issue using more advanced propagation models. When more advanced propagation models are employed it is often when the focus is on density estimation (e.g., in Marques *et al.* [29] and Küssel *et al.* and [30]) or localization/tracking of individual whales (e.g., in Simard and Roy [25], Nosal and Frazer [31], and Chapman [32]). Nonetheless,

there are a few examples of research that implement more advanced propagation models. For instance, a distinct seasonality in baleen whale call detection rates in the Gulf of Alaska was noted by Stafford *et al.* [33]; transmission loss results obtained with a parabolic equation model indicated that the seasonality in call detection rates was a result of seasonal changes to the SSP which affected the detection range. Širović *et al.*'s [4] study is one of the first examples in which a computational propagation model was included to normalize the calling rates of blue and fin whales by the modelled detection area to estimate cetacean abundance. In another case, Helble *et al.* [22] use a parabolic equation propagation model to demonstrate the significant impact of the ocean's environmental properties on the detection stage of PAM. The authors simulated calls recorded on a fixed receiver that propagated from virtual sources placed at various ranges and angles with respect to the receiver. Their results show that the probability of detecting a humpback whale call is a function of the environmental properties and can "vary by factors of 10 or more" among monitoring locations or at the same sensor over time [22]. To date, this research seems to be unique in the marine mammal community as it employs a full wave-field model that allows the transmitted humpback signals to attenuate correctly with respect to frequency, and accounts for phase distortions due to dispersion and multi-path. Furthermore, it comprehensively considers the effects of transmission, ambient noise, and the detection process to develop a realistic model of the probability of detection [34].

1.2 Investigating the Impacts of Environment-Dependent Propagation on an Automated Aural Classifier

From evidence in the literature, presented in the previous section, one can conclude that there ought to be a thorough understanding of how the environment impacts the detection/classification performance in order to develop an automatic recognition system capable of operating effectively under numerous environmental conditions. In spite of this general consensus, there is no study published in the literature that systematically analyzes the impacts of propagation on an automated classifier, using both underwater propagation experiments and complementary simulations. To address this deficiency, this thesis assesses the robustness of an automated aural classifier under various environmental conditions, through a combination of transmission and simulation experiments with modified bowhead

and humpback vocalizations. The aural classifier extracts perceptual signal features to inform the classification decision; the development of these features is motivated by the ability of a human listener to distinguish between sounds with similar temporal and frequency characteristics [35, 36]. Application of the aural classifier assumes a two-stage process: a detection stage followed by a classification stage. First, a general automatic detector is implemented with detection parameters set to achieve a high detection rate, while recognizing that this will likely generate many false detections. Then an automatic classifier is used to considerably reduce the number of false detections and identify the cetacean species [19]. This philosophy for detection/classification is employed to ensure that cetacean presence is noted, which is of particular importance when monitoring for at-risk species, or those that vocalize infrequently. It should be noted that the focus herein is on the impact of the environment on just the classification part of this process, and the effects on the detection stage are ignored.

Throughout this thesis a distinction is made between the two factors which define environment-dependent acoustic propagation characteristics — signal attenuation and signal distortion. Although cetacean vocalizations are modified by both processes as they propagate through the ocean environment, we choose to concentrate on the impacts of distortion resulting from multipath addition. This choice was made because less attention has been given in the literature to the effect of signal distortion in the context of PAM of cetaceans, whereas it is accepted that lower signal-to-noise ratio (resulting from signal attenuation) makes it more difficult to detect and classify sounds.

Preliminary research has shown the potential of the computer-based aural classifier for automatic species recognition [19, 37]. To build on the initial success of the classifier, we now seek to understand how robust the aural classifier is for conducting PAM in various marine environments. An initial investigation by Murphy and Hines [38] considers the temporal robustness of the automated aural classifier for discriminating between active sonar returns; the results of this investigation are extended to PAM of marine mammals and incorporate both an at-sea experiment and complementary simulations. A robust, fully characterized, automated recognition system is beneficial for population monitoring, determining habitat use, and/or conducting surveys that are required to mitigate the impacts of naval exercises and natural resources exploration on marine mammals.

1.2.1 Thesis Outline

Following this introduction, further details on the aural classifier and how its performance is evaluated are provided in Chapter 2. The initial data set, composed of bowhead and humpback vocalizations, is described in Chapter 3. Two types of whale vocalizations — biogenic and synthetic vocalizations — are used as source signals for the experiments and simulations. The development of the synthetic vocalizations is explained, and a comparison with the biogenic calls is conducted. After the signals in the initial data set are presented, details on how the signals were processed for the experiments and detected from experimental recordings are provided. One of the key processing steps included up-sampling signals in the biogenic and synthetic data sets, based on restrictions imposed by the at-sea experiment, such that the signals were transmitted in a higher frequency band than real bowhead and humpback vocalizations.

Chapter 4 describes at-sea propagation experiments that were conducted to gain valuable insight into how real environmental conditions impact the signals as they propagate through the water. Environmental parameters (e.g., SSPs and sediment characteristics) are also analyzed to gain insight into the acoustic environment experienced by the signals. The aural classifier is trained on signals with minimal signal attenuation and distortion and validated on signals that increasingly interacted with the ocean environment (such that increasing attenuation and distortion is expected), as a means of investigating the impact of acoustic propagation on the classifier performance. These experiments provide a baseline for modelling work, since the recorded signals include real propagation effects, resulting from the complex ocean environment.

In Chapter 5 pulse propagation modelling is introduced. A background discussion on the different types of propagation models is provided to motivate the selection of a ray-theoretic pulse propagation model. This model is then used to simulate the Gulf of Mexico experiment. The benefits of the simulation stage of the research are demonstrated by putting bounds on the classifier performance, based on realistic within-environment variability for the Gulf of Mexico experiment.

Further advantages of a simulation approach are demonstrated in Chapter 6, in which the pulse propagation model is applied to determine the relative impact of signal attenuation and distortion on the aural classifier performance. It is widely recognized that increasing signal attenuation decreases the performance of automated PAM systems; however, as

outlined in the preceding literature review (Section 1.1), little research has been directed at analyzing the impact of signal distortion. Hence, a novel method to disentangle the relative impacts of signal attenuation and distortion is introduced, and applied to biogenic bowhead and humpback calls propagated through two environments. The environments were motivated by measurements collected during the Gulf of Mexico experiment.

Finally, the techniques developed in previous chapters are applied in Chapter 7 to investigate the impact of environment-dependent propagation on the automatic recognition of bowhead and humpback vocalizations in their true frequency band (reversing the effects of up-sampling for the Gulf of Mexico experiment). The whale calls were propagated through two simulated environments, one of which was expected to have moderate propagation characteristics, and the second of which was expected to result in considerable propagation-induced signal distortion. The simulation case in the last analysis chapter is intended to provide a connection back to PAM of cetaceans that vocalize in a low frequency band, such that recommendations may be provided on how best to conduct PAM surveys of baleen whales.

CHAPTER 2

AUTOMATED AURAL CLASSIFIER AND PERFORMANCE METRICS

To study the impact of propagation on classification of cetacean vocalizations, a previously developed automated classifier was desired. In this thesis, a computer-based aural classifier developed at Defence Research and Development Canada (DRDC) [35] was employed. Previous results show it can be used to successfully discriminate vocalizations from several cetacean species; notably, the classifier was able to distinguish between bowhead and humpback whale calls with an accuracy of 92 % and an area under the receiver operating characteristic curve (AUC) of 0.97 [19]. The success of the aural classifier is due to the perceptual signal features it employs, which provide powerful discrimination cues for inter-species classification of cetaceans [19]. These features were developed by Young and Hines [35] to take advantage of evidence [36] that experienced sonar operators can often *hear* differences in active sonar returns from man-made metallic objects (targets) and naturally occurring geologic objects (clutter). The perceptual features are different than features obtained using conventional signal processing techniques because they take into account how a human listener perceives sound [19, 35, 38]. For example, conventional signal processing might employ *peak frequency* — the frequency corresponding to the maximum value of the signal spectrum (as in Parks *et al.* [24]). The analogous perceptual feature is referred to as the *peak loudness frequency*, which is the frequency corresponding to the peak in the loudness spectrum (this includes, for example, the bandpass filtering that occurs in the ear canal). This chapter provides a brief overview of the aural classifier, and a discussion on how the performance of the classifier was evaluated; for more detailed discussion on the classifier, the reader is referred to Young and Hines [35], and Binder

and Hines [19] .

2.1 The Aural Classifier

In essence, a classification model is a mapping from data samples to predicted classes [39]. In a PAM context, an automatic classifier can be thought of as a predictive model that is capable of independently identifying a cetacean species based on its vocalizations. Expert marine mammal analysts can often aurally distinguish between cetacean species by listening to recordings [1, 20, 40]. Thus, the aural classifier is a convenient tool that mimics how a human listener distinguishes among cetacean vocalizations. Once trained to recognize discrimination cues for inter-species classification, the assumption is that it will be able to identify novel vocalizations with high accuracy.

The classification process begins with calculating the perceptual features from each of the vocalizations to be classified. To accomplish this, a relatively simple auditory model is first applied to each of the vocalizations to obtain a perceptual representation of the signals [19, 35]. A total of 58 features — defined in Young and Hines [35] — are computed for each signal, 46 of which are one-dimensional time-frequency features (e.g., mean subband decay time), and twelve of which are purely spectral features (e.g., peak loudness frequency). After the perceptual features are determined, the data are split into training and validation subsets; from the training set the most important aural features (those with sample class means that are well-separated relative to the overall variance of the dataset [41]) were identified using the Fisher Linear Discriminant score [19, 38]. These features are then used to train a Bayesian classifier model.

Each of the perceptual features represents an axis of the feature space. So, if all features were used for classification the resulting feature space would be 58-dimensional. A high-dimensional feature space requires many training samples to accurately estimate the underlying patterns in the data; this is referred to as the 'curse of dimensionality' [42]. As described in Chapter 3, the number of samples in the data set was limited. Thus, the first step in training the classifier is to reduce the dimensionality of the feature space. The dimensionality is first reduced by selecting those features with a high Fisher Linear Discriminant score. Discriminant analysis is then used to linearly combine the selected features to further reduce the dimensionality of the feature space [37] by projecting the

data onto a single dimension (throughout the thesis this axis is referred to as the 'Projected Feature Value').

A Gaussian-based Bayesian classifier is then applied to the projected data. A Gaussian probability density function (pdf) is fit to each of the classes; each pdf represents the likelihood for points along the Projected Feature Value axis to belong to one of the two possible classes. Bayesian decision theory is used to combine the likelihood probabilities with prior probabilities (i.e., relative number of samples in each class) to inform the classification decision. Essentially, the classifier makes its decision based on which class has the highest likelihood probability, relative to the number of class samples. A simple Bayesian framework is used for this research rather than one of the many other more sophisticated classifier architectures available (e.g., neural networks [41, 43, 44] or support vector machines [41, 44]) to direct focus on the effectiveness of the perceptual features, rather than the choice of classification algorithm. In the future, it would be possible to implement a more advanced classifier architecture to potentially gain improvements in classifier performance; however, such an investigation is beyond the scope of this research.

In summary, the aural classification process is divided into three phases: first, a relatively simple auditory model is applied; second, the perceptual signal features are computed for each signal; finally, using a linear combination of the features that best discriminate among the classes, a Bayesian classifier is applied to determine to which class each sample belongs [19, 35]. Throughout this process, the classifier is *trained* with a set of signals for which the classifier is provided the class labels. The effectiveness of the classifier is then *validated* with a second set of signals for which the classifier has no direct knowledge of the class labels. Validation is an essential step that ensures the classifier has not been over-trained on the data in the training set, but will maintain good performance when presented with unknown data [45].

2.1.1 Validation of Classifier Performance

An essential step in developing a classifier is to assess its generalization, since we desire a classifier that is not merely tuned to the training set, but will perform well when presented with unknown data [44, 45]. The simplest method to assess the generalization error of a classifier is to randomly split the set of data samples into two parts, with no duplication of samples between the two subsets. One of these sets is used to train the classifier (i.e.,

adjust model parameters in the classifier) and the other, known as the validation set, is used to estimate the generalization error [45]. The classifier is then trained with data for which the classifier is provided a class label. By generalizing the underlying patterns in the training set, predictions can be made about data that do not have a known class label. The effectiveness of the classifier is then validated by imposing the assumptions of the classifier on a dataset for which the classifier has no direct knowledge of the class label. To evaluate the performance of the classifier in generalizing to novel data points, the class label assigned by the classifier to each vocalization in the validation set is compared to the known true label. This process is referred to as cross-validation. The generalization error is low if the classifier maintains similar performance between the training and validation subsets. In brief, cross-validation is an empirical approach to asses classifier performance; once the classifier is trained using cross-validation, the validation error provides an estimate of the classifier performance when it is applied to a novel dataset [45].

In this work, simple validation was used to generate decision regions to visualize the discriminability of the data (see conceptual example in Figure 2.5 and example of experimental results in Figure 3.4). Decision regions are generated by projecting each sample in the training or validation subset onto the 'Projected Feature Value' determined from discriminant analysis, then binning the data to generate class-specific histograms; the decision threshold is set at the point of equal likelihood. One is able to verify classification decisions visually by noting on which side of the decision threshold a data sample lies. This comparison is facilitated by the background colour of the plots; for example, a red bar on the red background represents samples that were correctly classified, whereas a red bar on the blue background indicates incorrectly identified samples. It is also possible to assess the discriminability qualitatively by noting the separation between class means and the overlap between the class distributions (total discriminability is attained when class means are well separated and there is no overlap). This simple validation method is accomplished by randomly splitting the data into training and validation subsets such that half of the signals for each class were placed in the training set and the other half in the validation set. There is no duplication of signals between the training and validation subsets, i.e., these are disjoint sets. This is an important point since it is essential that the validation set not include any of the signals used to train the classifier, resulting in a methodological error referred to as "testing on the training set [45]."

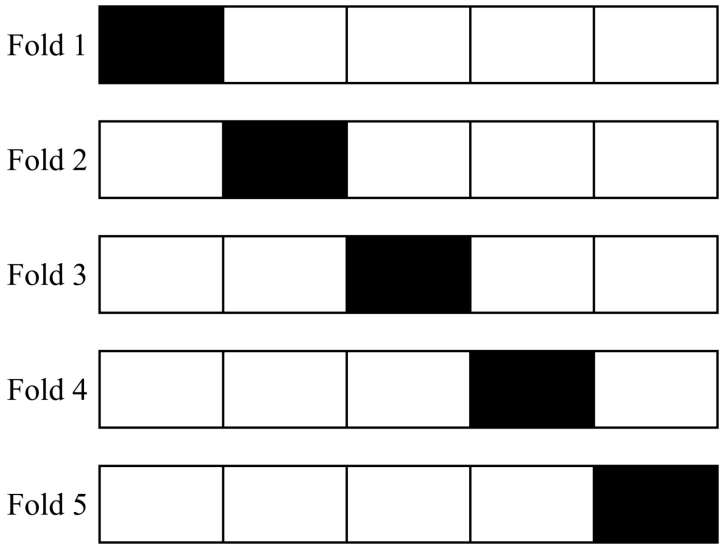


Figure 2.1: Illustration of k-fold cross-validation; the specific case shown is for $k = 5$. Each box represents a disjoint data subset that is $1/k$ the size of the complete dataset. The white boxes represent those used to train the classifier, and the black boxes the validation subset. The classifier is trained and validated for each fold, then the results are averaged together to estimate the performance of the classifier.

A generalization of the simple cross-validation method is known as k-fold cross-validation. In this method, the dataset is divided into k disjoint sets of equal size n/k, where n is the total number of signals in the dataset. The classifier is trained k times, with one subset of the data withheld as the validation set and the union of the remaining $(k-1)$ sets used for training [44, 45]; this technique is illustrated in Figure 2.1, in which each box represents one of the data subsets. The estimated performance of the classifier is then the mean value of the k computed performance metrics [44, 45] and the standard deviation of the metrics is used as an estimate of the generalization error. There is no prescribed rule for choosing the proportion of the data set to be used as the validation subset; however, it is generally accepted that a smaller portion (less than half) of the data set be used for validation than was used to train the classifier. Additionally, k is selected to balance the competing requirements of having enough samples in the training set so as not to bias the performance estimate, while leaving enough samples in the validation set to minimize the variance of the performance estimate. With the exception of generating decision regions, it was found a choice of $k = 5$ resulted in a convenient proportion of samples in the validation subset. By performing k-fold cross-validation, any potential bias associated with having an outlier in either the training or validation set is accounted for.

2.2 Performance Metrics

Since the goal of this thesis is to evaluate how the classifier performance is impacted by propagation effects, the method used to evaluate classification performance must be carefully considered. In evaluating the performance, metrics are desired that are relatively easy to implement, provide useful information, and are simple to interpret. To achieve these goals, two performance metrics were selected: classification accuracy and AUC.

Discriminating between bowhead and humpback whale vocalizations is a dichotomous decision process. The decision is made using a threshold-based rule on a continuous variable, $\mathcal{X}$, that yields the decision, D, as a positive (**p**) or negative (**n**) instance according to,

$$D = \begin{cases} \mathbf{p}, & \text{if } \mathcal{X} \geq \tau \\ \mathbf{n}, & \text{if } \mathcal{X} < \tau \end{cases} , \tag{2.1}$$

where τ is the value of the threshold [46]. Throughout this section, the conventional decision theory terminology — $\mathcal{X}$, positive and negative — will be employed for simplicity and to be consistent with the literature; however, it should be recognized that this is an arbitrary designation. The class labels can be switched from 'positive' and 'negative' to 'bowhead' and 'humpback', and 'Projected Feature Value' may be substituted for $\mathcal{X}$, without any loss in meaning.

Accuracy, which is sometimes referred to as the classification rate, is a performance measure which is defined as the ratio of correctly classified samples to the total number of samples, n, or,

$$\text{accuracy} = \frac{\text{TP} + \text{TN}}{n} , \tag{2.2}$$

where TP and TN are the number of correctly identified positive and negative signals, respectively. Although accuracy is an intuitive metric, it can be limited in its usefulness; namely, it assumes equal risk associated with all incorrect decisions, and it treats all outcomes as being equally likely [46]. Therefore, it is necessary to use an additional measure of classifier performance, such as the AUC, to enhance insight into the classifier performance.

An example of distributions representing the occurrence of positive and negative events is provided in Figure 2.2. By and large, $\mathcal{X}$ is able to discriminate between positive and

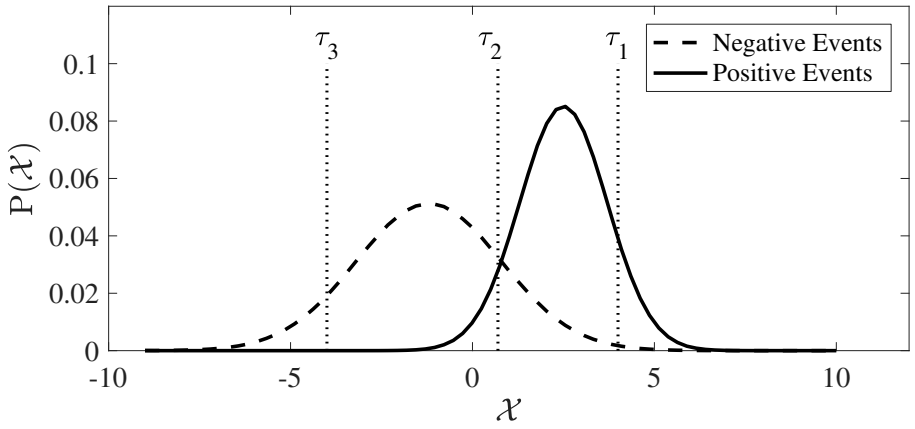


Figure 2.2: Illustration of the occurrence of positive and negative events. Three example thresholds (τ_1, τ_2, or τ_3) are shown — a positive event decision is made for $\mathcal{X} \geq \tau$. The choice of decision threshold determines the true positive and false positive rates.

negative events, since the class means are separated and there is a relatively small amount of overlap between the distributions. Three possible decision thresholds are also depicted in the figure. At threshold τ_1 calling all events with $\mathcal{X} \geq \tau_1$ positive would correctly identify a small proportion of all positive events, but also erroneously label a large proportion of the positive events as negative. The benefit of excluding negatives is made at the cost of missing many positive events. At threshold τ_2 a more even balance is struck in which the majority of positive and negative events would be correctly identified while also incorrectly classifying some negative events. Finally at τ_3, all positive events are correctly identified, but a large proportion of the negative events have been inappropriately labelled as positive. This example highlights how the choice of decision threshold influences classification accuracy.

At each threshold there are four possible outcomes from the decision process; these are tabulated in the confusion matrix in Figure 2.3. The cells along the major diagonal represent the correct decision, and the cells off the major diagonal represent the errors (or confusion) between the classes [39]. For example, a true positive occurs when the classifier correctly identifies a signal as a positive, whereas a false negative is a result of the classifier incorrectly identifying a positive signal as a negative. Receiver operating characteristic (ROC) curves illustrate the tradeoffs between the benefits (true positives) and costs (false positives) of a classifier model [39], by plotting the true positive (tp) rate

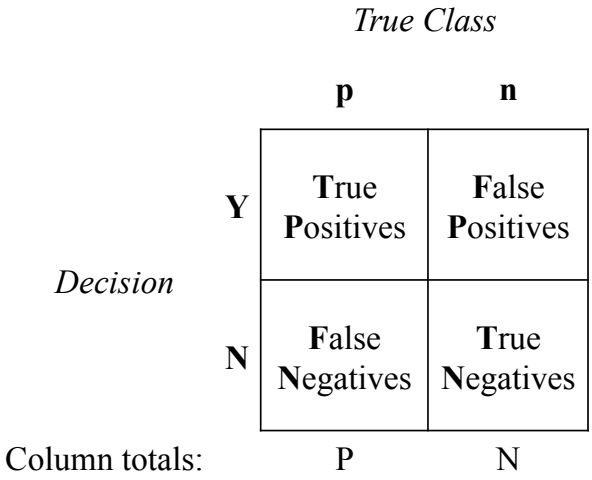


Figure 2.3: Confusion matrix for the two-class classification problem. The symbols **p** and **n** represent the truth value (positive or negative) of a signal, whereas **Y** and **N** represent the decision made by the classifier (after Fawcett [39]).

versus the false positive (fp) rate. These rates are defined as,

$$\text{tp rate} = \frac{TP}{\mathrm{P}} \quad , \tag{2.3}$$

and

$$\text{fp rate} = \frac{FP}{\mathrm{N}} \quad , \tag{2.4}$$

where *TP* is the number of positive signals correctly classified, P is the total number of positive signals, *FP* is the number of negative signals incorrectly classified, and N is the total number of negative signals. A ROC curve is generated by sweeping the decision threshold through values of $\mathcal{X}$ to add points to the curve. Example ROC curves are shown in Figure 2.4. To move away from the chance ROC curve (red solid line) toward the ideal curve (blue dashed line), a classifier must exploit information in the data set [39]. Thus, it can be seen that the ROC curve provides a simple graphical representation of the tradeoffs between true positive and false positive rates, as well as the general discriminating power of $\mathcal{X}$ for distinguishing positive from negative events [46].

Much of the information in a ROC curve can be distilled down to a single value — the area under the ROC curve. The AUC is a non-parametric summary measure, and unlike the accuracy, it is independent of the choice of decision threshold [46]. It also has a convenient statistical interpretation; Green and Swets [47] showed that the AUC can

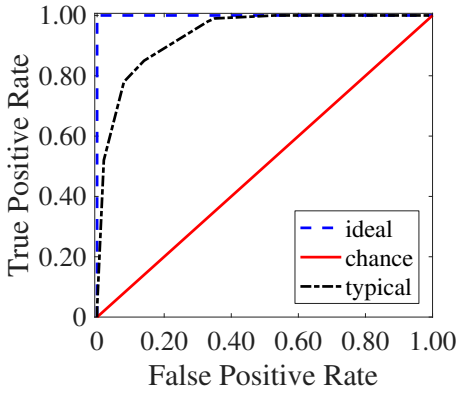


Figure 2.4: Example ROC curves. The blue dashed curve indicates the ideal case in which all samples are correctly identified, the red solid curve represents a useless decision process in which the decision variable has no discriminating power, and the black dotted curve shows a typical result.

be interpreted as the probability that the classifier will correctly discriminate a randomly selected pair of positive and negative samples. It is also an indication of how well separated the class distributions are — a large value of AUC indicates little overlap between the distributions. The AUC varies between 1 (indicative of an ideal classifier) and 0.5 (equivalent to randomly assigning a classification decision). In general, a larger AUC implies better average classifier performance [39]. When assessing classifier performance in terms of the AUC the following general guidelines from the literature [38, 48] may be followed: $\mathrm{AUC} = 0.5$ indicates no discrimination, $0.7 \leq \mathrm{AUC} < 0.8$ is considered acceptable discrimination, $0.8 \leq \mathrm{AUC} < 0.9$ is considered excellent discrimination, and $\mathrm{AUC} \geq 0.9$ is considered outstanding discrimination.

By examining both the AUC and the accuracy, one is able to distinguish between discriminability and decision bias. The former is an inherent property of the classifier, while the later is dependent on the choice of (changeable) threshold [41]. Hence, both metrics are required to interpret fully how propagation impacts classification results [49, 50]. For example, it is possible to achieve an ideal AUC but not attain 100 % accuracy, which may seem counterintuitive; however, one must carefully consider what each performance metric describes. The AUC represents the probability that a randomly selected pair of positive and negative samples will be correctly classified [46]. The accuracy, on the other hand, imposes a threshold and measures the resulting accuracy with respect to that threshold [39]. Clearly the threshold has greater generalization error than the ROC curve in the case

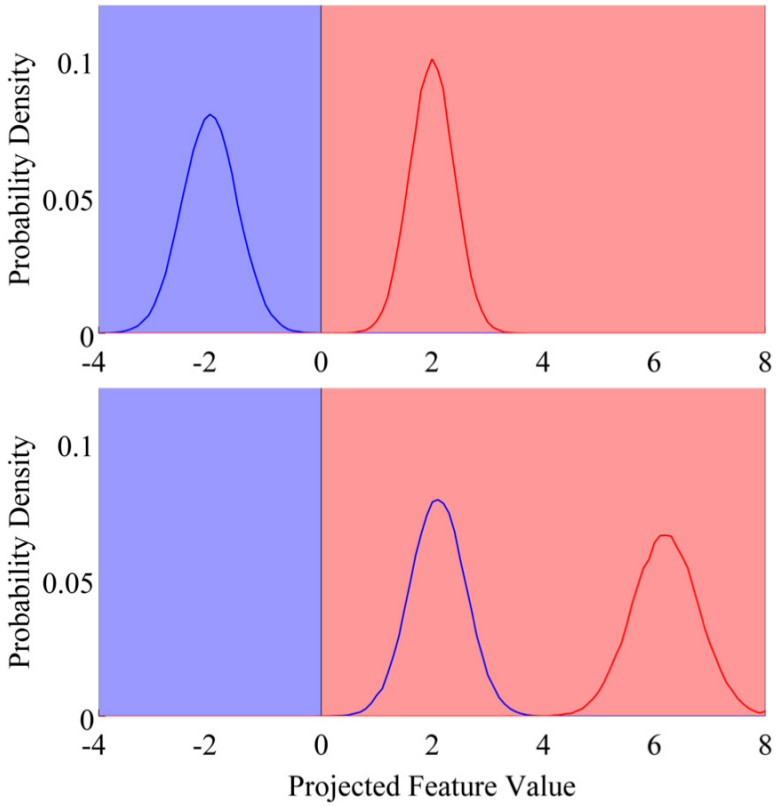


Figure 2.5: Illustration of (top panel) both excellent accuracy and AUC, and (bottom panel) poor accuracy but excellent AUC. Correct classification occurs when the distribution falls on the corresponding background colour.

where the AUC indicates an ideal classifier but the accuracy is less than perfect. The class distributions in Figure 2.5 depicts such a scenario. Ultimately, both the accuracy and AUC are required to assess classifier performance to gain a full understanding of the decision process.

CHAPTER 3

BIOGENIC AND SYNTHETIC VOCALIZATION DATA SET

Two types of signals are used in this thesis: biogenic and synthetic whale calls. Biogenic calls are those derived directly from recorded whale calls, i.e., their origin is purely biological; whereas the synthetic calls were generated artificially to mimic whale calls. The synthetic calls were generated due to the desire to transmit calls that had not been subjected to propagation effects prior to the experiments or simulations. Further details on these two types of signals are found in the following sections. The remainder of this chapter discusses how the signals were conditioned for the experiments and simulations, and briefly describes the detector used to isolate calls from the recordings.

3.1 Biogenic Whale Calls

Bowhead (*Balaena mysticetus*) and humpback (*Megaptera novaeangliae*) whale vocalizations were selected to study the impact of propagation on the performance of the aural classifier. Calls from these species were used in a previous study of the effectiveness of the aural classifier for PAM of marine mammals [19]. Of the whale species considered in Binder and Hines [19], bowhead and humpback vocalizations were the most challenging for the aural classifier to discriminate due to the similar frequency content and duration of many of the calls produced by these species; nonetheless, the classifier was able to discriminate between the vocalizations with 92 % accuracy and an $\mathrm{AUC} = 0.97$ [19]. The bowhead and humpback calls were thought to provide a sufficiently challenging classification scenario to test the robustness of the perceptual features, given the similarity of

Table 3.1: Frequency range and duration of vocalizations in the biogenic data set. Note that these are not representative of the full vocal repertoire of the species, but only for the vocalizations used in this study.

Vocalization Type	Frequency Range (Hz)	Duration (s)
Humpback unit 1	200 to 600	2.5 to 3.0
Humpback unit 2	150 to 700	1.0 to 1.5
Humpback unit 3	100 to 2000	$\sim$1.0
Humpback unit 4	500 to 1300	1.5 to 2.0
Bowhead	50 to 800	2.5 to 3.0

frequency band and duration of the calls.

Both of these species are known to produce song; that is the hierarchical organization of vocalizations used by some baleen whale species [8]. Within a whale song there is a high degree of repetition. As originally described by Payne and McVay [8], the shortest sound that is continuous to a human ear is referred to as a unit; a few units are repeated in a prescribed order to form a phrase, and the phrase is repeated several times within a theme, and so on. Thus, an automatic classifier likely only needs to be trained on a select few song units because they are repeated often enough within a song for the classifier to inform a decision as to what species is/are present.

Recordings of example bowhead and humpback vocalizations were obtained from the MobySound website [51, 52]. The MobySound database contained samples of bowhead song endnotes and complete recordings of humpback songs. A humpback song repertoire contains a large variety of sounds, including a wide range of durations and frequency content [8]. Of the many different sounds (often referred to as 'units'), four song units were selected for the classification research. These sound units were the same as those used in the previous study of the effectiveness of the aural classifier for PAM of cetaceans [19]. They were selected because the frequency band and duration of the units overlapped with the bowhead vocalizations (see Table 3.1). Spectrograms of example bowhead and humpback calls are shown in Figure 3.1. The similarity of calls between these two species makes it difficult for some automatic detectors (e.g., the band-limited energy detector discussed in Bougher *et al.* [21] and Hood *et al.* [53]) to discriminate between the species,

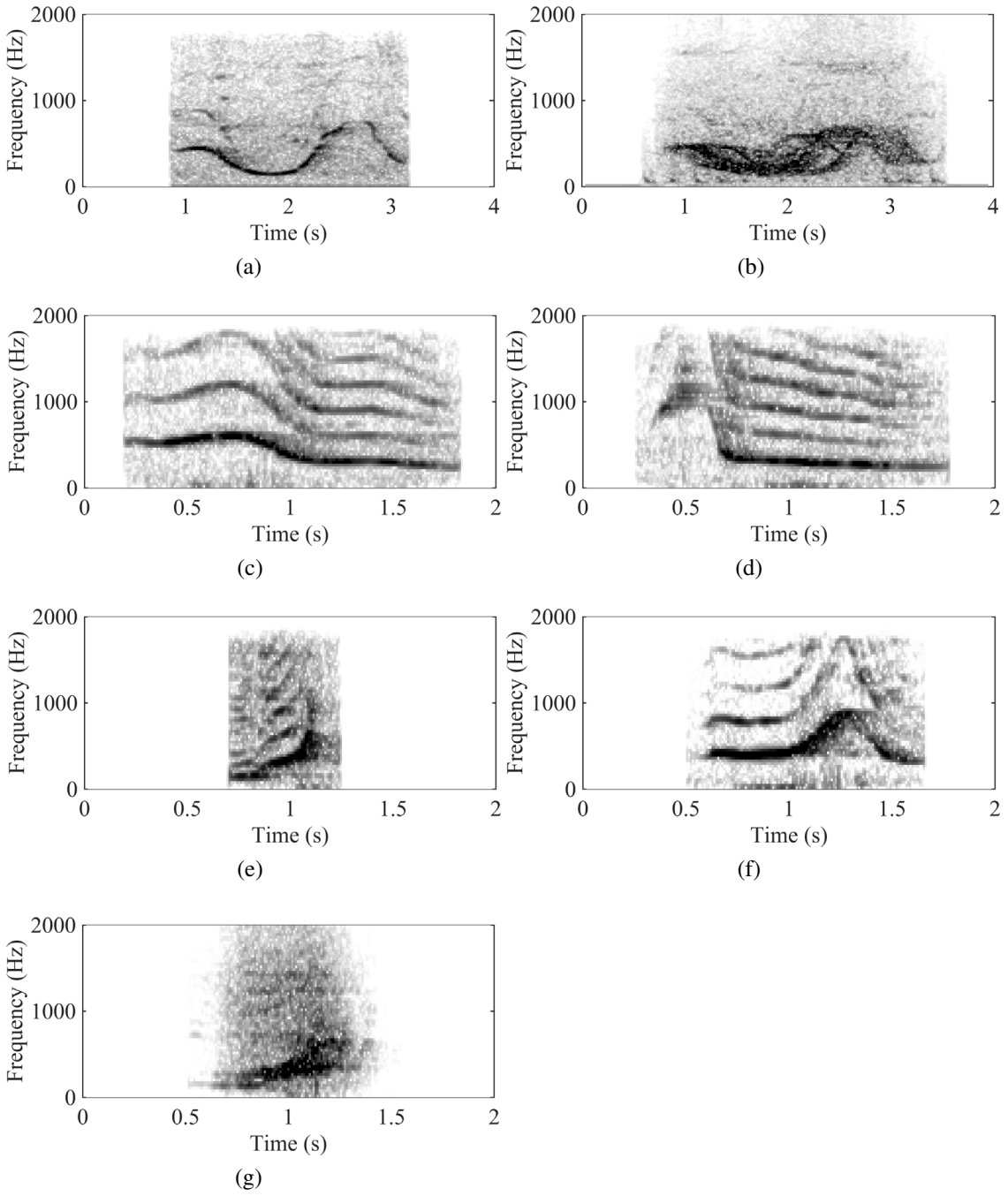


Figure 3.1: Example spectrograms of (a) bowhead, and (b) synthetic bowhead vocalizations; as well as the humpback song units referred to as (c) humpback1, (d) humpback2, (e) humpback3, (f) humpback4, and (g) an example synthetic humpback vocalization. Spectrograms are produced using Hann-weighted windows of length 512 and 256 samples for the bowhead and humpback vocalizations, respectively, and 70 % overlap.

so they were considered a challenging case, in the context of this thesis, to test the robustness of the perceptual features.

Since information on the location of the whales with respect to the recorder and propagation conditions at the time of recordings was not available, a subset of high signal-to-noise ratio (SNR) calls was generated. In doing this, the assumption was made that high SNR signals likely indicated that the vocalizing whale was relatively close to the recording unit. This suggested that fewer propagation effects were contained in the recorded signals, such that after the propagation experiments were conducted, the main source of signal distortion due to propagation would result from the environment at the experimental site. For this reason, 155 of the high SNR vocalizations from the set of bowhead song endnotes and humpback song units used in the previous PAM research [19] were selected for use in the propagation experiments; all vocalizations in the resulting subset had $\mathrm{SNR} > 15$ dB.

3.2 Synthetic Whale Calls

As previously discussed, the bowhead and humpback calls obtained from the MobySound website were subject to unknown propagation effects. Consequently, synthetic signals were developed to provide known starting signals that contained no propagation effects prior to the experiment. These were transmitted in addition to the biogenic bowhead and humpback vocalizations.

The goal in generating the synthetic signals was to produce signals that were similar, in terms of the aural classifier, to the biogenic whale vocalizations. That is, the synthetic calls were considered to be similar to the biogenic calls if the class means and variances of the perceptual features were comparable between the biogenic and synthetic signals for the features that had high Fisher Linear Discriminant scores. Due to the complex nature of the aural classifier algorithm (i.e., it involves several non-linear transformations and yields a set of perceptual features that are inter-related in complex ways), it was decided to generate synthetic calls directly from the biogenic calls rather than assuming a functional form for the calls' time series and/or the temporal envelopes of the frequency subbands, as one might intuitively be tempted to do.

The synthetic signals were based on an example set of high SNR bowhead song end notes and one of the humpback song units (unit 3, as described in Section 3.1). To

generate the synthetic signals, each example call was first denoised using wavelet analysis. The denoising process is a means of recovering the true whale call from the recording and eliminating a considerable portion of the noise [54]. This ensured that when the synthetic signals were generated the noise characteristics of the recording sites were not replicated, and that the synthetic signals were constructed as much as possible from the whale vocalization portion of the signal. Figure 3.2 demonstrates the effect of denoising an example signal. This illustration highlights how denoising with wavelet analysis differs from merely applying a bandpass filter to remove out-of-band noise, which is the commonly used alternative; the distinction is particularly obvious in the 1 to 1.5 s time range, where the frequency band from which the noise has been removed is markedly different than for adjacent time periods. In contrast, a bandpass filter would remove the same frequency content across the entire signal.

After removing the noise from the example recordings, the information in the sample recordings was summarized in terms of the mean signal and the eigenvectors of the covariance matrix. This is similar to methods employed to predict sound speed profiles from a database of historical measurements [55]. More specifically, the mean signal and empirical orthogonal functions (EOFs) [55] were calculated from the wavelet-transformed vocalizations for each species. The wavelet-transformed representation was used to emphasize the importance of the two types of perceptual features (time-frequency and purely spectral) used by the aural classifier. Note that this process was applied separately for the bowhead and humpback signals, that is, a mean signal and EOFs were generated

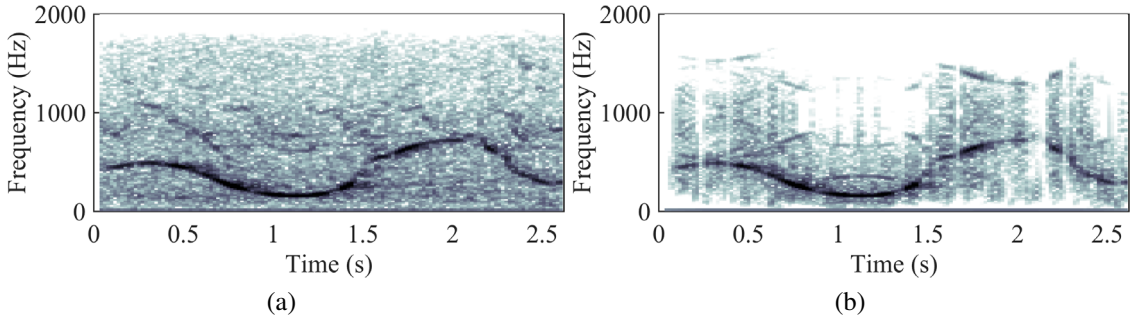


Figure 3.2: Spectrograms of an example bowhead vocalization from the MobySound dataset (a) before, and (b) after it was denoised using wavelet analysis. Spectrograms are produced using Hann-weighted windows of length 512 samples and 70 % overlap.

from the biogenic bowhead calls, and another mean signal and set of EOFs were generated from the biogenic humpback calls.

To generate a single synthetic whale vocalization, random weights were applied to the EOFs that contained 95 % of the variance. The weights for each EOF were generated by randomly drawing numbers from a zero-mean Gaussian distribution with standard deviation given by the square root of the EOF eigenvalue [56]. Then the weighted EOFs were added to the mean signal, and the inverse wavelet transform was performed to obtain the time-domain signal. Applying different randomized weights to the EOFs for each synthetic signal generated a set of 155 signals per species. Finally, a brief analysis was completed to ensure that the resulting collection of synthetic calls were sufficiently representative of whale vocalizations. This included comparing the time-frequency content of the biogenic and synthetic signals, and comparing the aural classifier results (as described below).

3.2.1 Comparison of Synthetic Calls with Biogenic Whale Calls

One of the simplest methods to determine how well the synthetic whale calls approximate the biogenic calls is to compare the frequency and time-frequency content of the signal by way of their pressure spectra and spectrograms. Spectra comparing the biogenic and synthetic signals are presented in Figure 3.3 — these spectra were generated by computing the average of the RMS spectral levels. Note that the synthetic signals were lowpass filtered to facilitate comparison with the biogenic call spectra; this was reasonable because the high frequency portion of the signals was merely noise that was filtered out in later processing. These spectra clearly show that the frequency content of the biogenic and synthetic signals was similar across the majority of the frequency band. There was, however, a noticeable discrepancy at low frequencies for the bowhead vocalizations — fortunately this discrepancy was unimportant since only the 200 to 800 Hz frequency band was transmitted during the experiments (see Section 3.3 for full details). Figure 3.1 shows spectrograms of example biogenic and synthetic vocalizations, for comparison of their time-frequency characteristics. The spectrograms reinforce the similarity of the biogenic and synthetic signals since the shape of the synthetic signals in the spectrograms is consistent with those of the biogenic calls they were mimicking. They were not a perfect match though. The time-frequency 'track' of the humpback3 synthetic call was broader than its biogenic counterpart and the harmonics were not as distinct. The broadening of

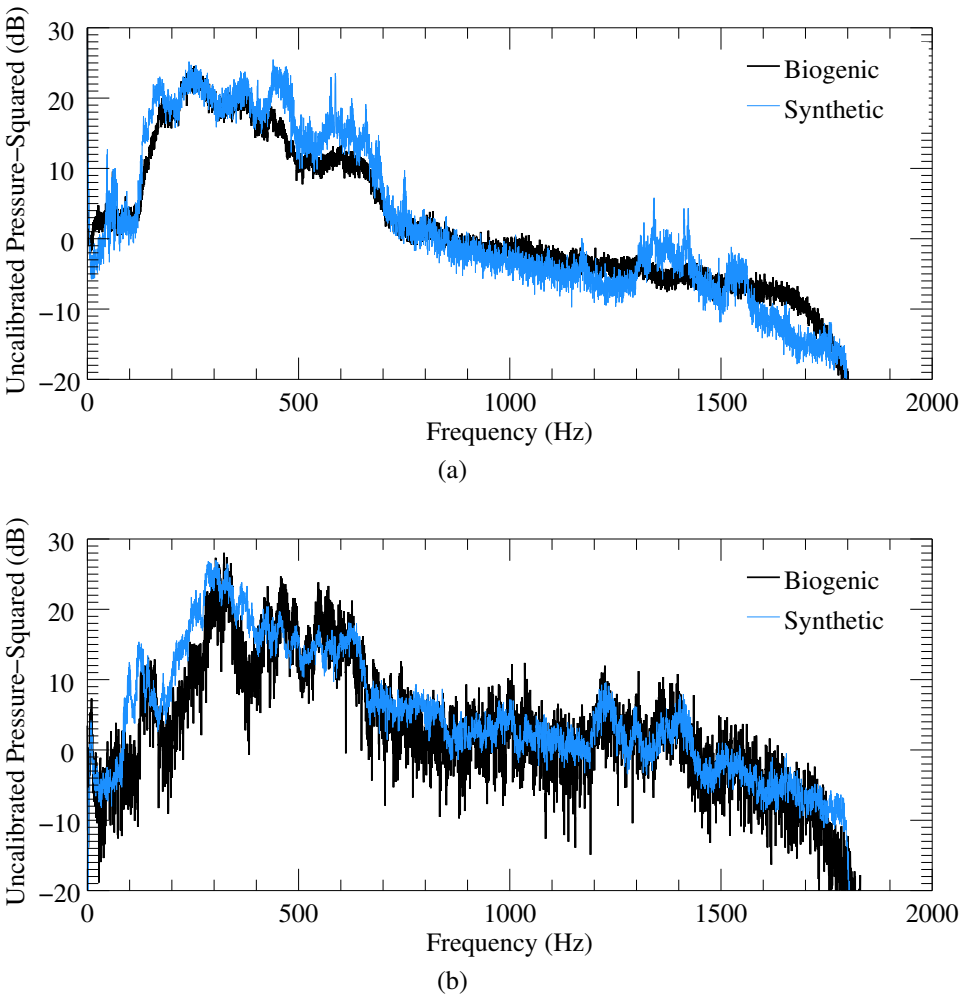


Figure 3.3: Spectra of biogenic and synthetic (a) bowhead, and (b) humpback3 units. Synthetic signals were lowpass filtered prior to determining the spectra to facilitate comparison with the biogenic calls.

the time-frequency 'track' was more noticeable for the bowhead synthetic call, for which it appeared that several bowhead calls were added together. While achieving an approximate match between the time-frequency characteristics of the biogenic and synthetic whale calls is important, it is the similarity of the aural classifier performance and highly-ranked perceptual features that is more important for the purposes of this thesis.

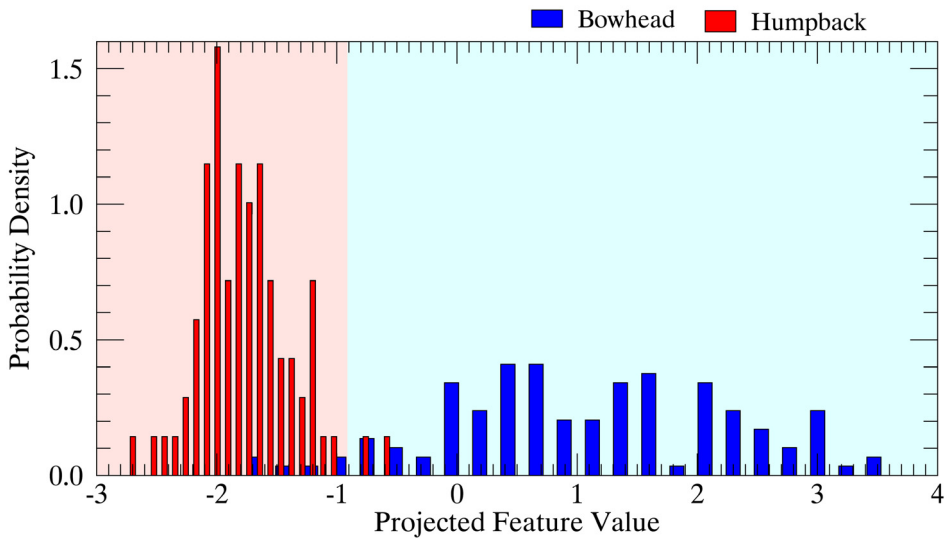
The aural classifier was trained separately on the biogenic and synthetic vocalizations, with the freedom to select the features that best discriminated between the bowhead and humpback calls. The resulting ranking of perceptual features, in terms of their discriminant score [38], was compared for the two cases. Within the highly-ranked perceptual features the biogenic and synthetic calls had five features in common. The five

Table 3.2: The five features that best discriminated between bowhead and humpback whale calls for both the biogenic and synthetic signals.

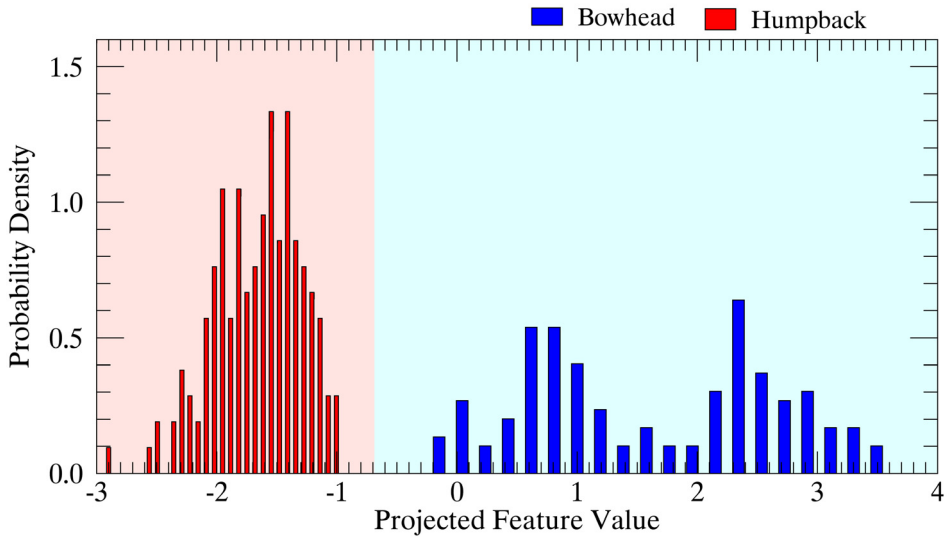
Feature	Description
Duration	Time delay between the start and end of the signal.
Global mean subband decay time	Average time delay between the peak of the temporal envelope and the end of the signal.
Local maximum subband decay time	Maximum time delay between the peak of the temporal envelope and end of the signal; each subband considered individually.
Frequency of global maximum subband attack time	The centre frequency of the filter bank channel which contains the maximum subband attack time (i.e., the time delay between the start of the signal and the peak of the temporal envelope).
Peak loudness value	Maximum value of the perceptual loudness spectrum.

perceptual features that were considered important for discriminating between bowhead and humpback vocalizations (either biogenic or synthetic) are listed in Table 3.2, including a qualitative description of each feature.

To ensure that the classifier results are similar for the biogenic and synthetic calls, the classifier was trained using the five features listed in Table 3.2. The resulting decision regions are presented in Figure 3.4. In general, the decision regions shared many similar characteristics — specifically, the shapes of the bowhead and humpback distributions were much alike. The main difference evident between the two decision regions was the greater overlap between the biogenic bowhead and humpback distributions than between the synthetic vocalization distributions. Notwithstanding this difference, the performance metrics were similar using 5-fold cross-validation: classification of the biogenic calls resulted in an accuracy of $(97 \pm 3)\,\%$ and AUC of 0.99 ± 0.01, and the synthetic vocalization classification produced an accuracy of $100\,\%$ and $\mathrm{AUC} = 1.00$ (note that no uncertainty is provided for the synthetic vocalizations since all five folds produced the same performance metrics). Further to these classification results, the means and standard deviations of the bowhead and humpback classes for each of the individual perceptual features were found to be comparable between the biogenic and synthetic vocalizations. As a result of these three methods for comparing the biogenic and synthetic whale calls, it was determined that the synthetic calls were sufficiently similar to the biogenic calls to be used for the experiments.



(a)



(b)

Figure 3.4: Comparison of aural classifier decision regions for (a) biogenic and (b) synthetic bowhead and humpback whale vocalizations. The aural classifier was trained using the five perceptual features found to be important for classification of bowhead and humpback calls (listed in Table3.2).

3.3 Conditioning and Upsampling of Signals for Experiments

Both the biogenic and synthetic whale vocalizations underwent a conditioning process prior to being transmitted during the experiments. This was largely motivated by the fact that bowhead and humpback whales vocalize at relatively low frequencies and across a wide frequency band. As highlighted in Table 3.1, bowhead and humpback vocalizations included in the biogenic data set had frequency content predominantly in the 50 to 800 Hz and 100 to 2000 Hz range, respectively. No projector was readily available for the experiments that was capable of transmitting signals over the approximately five-octave band of the vocalizations, particularly at these low frequencies. To address these two issues (i.e., low frequency and wide bandwidth) all signals were filtered and scaled for transmission from an ITC-2010 projector. This acoustic source has a two-octave passband from 1 to 4 kHz, based on its transmitting voltage response (TVR) curve (reproduced in Figure 3.5). Pre-processing of the signals was done to take advantage of this two-octave flat frequency band when transmitting the signals. The RMS spectral levels of each species' vocalizations were averaged (results shown in Figure 3.6) to identify a two-octave band that contained sufficient signal content to be representative of the full-bandwidth calls. The energy was integrated for the full-bandwidth signals and compared with the energy in the 200 to 800 Hz range; based on the averaged spectra, this reduced frequency band contained 74 % of the total energy in the bowhead vocalizations and 72 % of the energy in the humpback calls. Thus, a relatively large proportion of the the vocalization energy was contained in the 200 to 800 Hz band.

To test if this reduced frequency band contained sufficient information for aural classification, each of the vocalizations were bandpass filtered and classification was performed on both the full-bandwidth and reduced-bandwidth signals; the results are shown in Figure 3.7. The decision regions were remarkably similar, such that the decision threshold was the same for the two cases. Qualitatively, there was little distinction between the class distributions after the bandwidth was reduced. One would expect a similar amount of variation merely from reshuffling the signals between the training and validation data sets. Using 5-fold cross-validation it was found that the full-bandwidth signals resulted in (92 ± 1) % classification accuracy and an $\mathrm{AUC} = 0.978 \pm 0.004$; when using the reduced-bandwidth

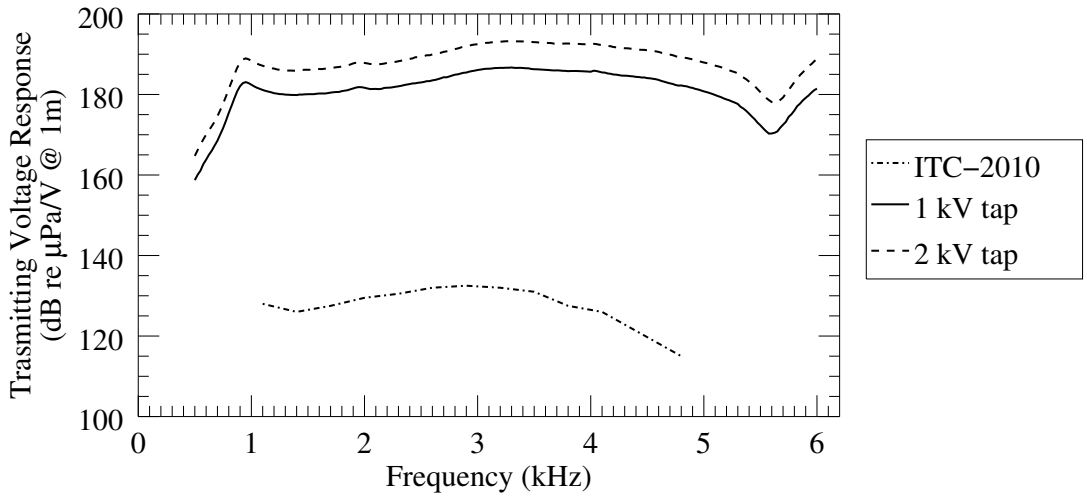


Figure 3.5: Transmitting voltage response (TVR) curves for the ITC-2010 transducer (replicated from manufacturer specifications [57]) and the complete amplifier and ITC-2010 transducer system.

signals the classification performance was not significantly altered (accuracy of (92 ± 4) % and AUC of 0.97 ± 0.02). Many of the same perceptual features were highly ranked discriminators for the full-bandwidth and reduced-bandwidth vocalizations. Taken altogether, this evidence indicates that applying a 200 to 800 Hz bandpass filter to the vocalizations did not remove the essential information required for calculating the important perceptual features and that enough of the signal content was contained in this frequency band to accurately represent both species' vocalizations. Hence, all signals transmitted during the experiments were bandpass filtered to include only frequency content in the 200 to 800 Hz range.

The final step in processing the signals for transmission was to increase the playback speed of each of the filtered signals by a factor of five to scale the signals into the passband of the ITC-2010 source. The results obtained from the frequency-scaled signals were therefore not directly indicative of how propagation impacts the classification of bowhead and humpback whales. Consequently, one must be cautious in applying the numerical results from the higher band signals; however, these signals are useful in establishing general conclusions about how propagation impacts an automated classifier. Simulations in the true frequency band of bowhead and humpback vocalizations (i.e., 200 to 800 Hz) are conducted in Chapter 7 in order to confirm the applicability of the general conclusions to PAM of bowhead and humpback whales.

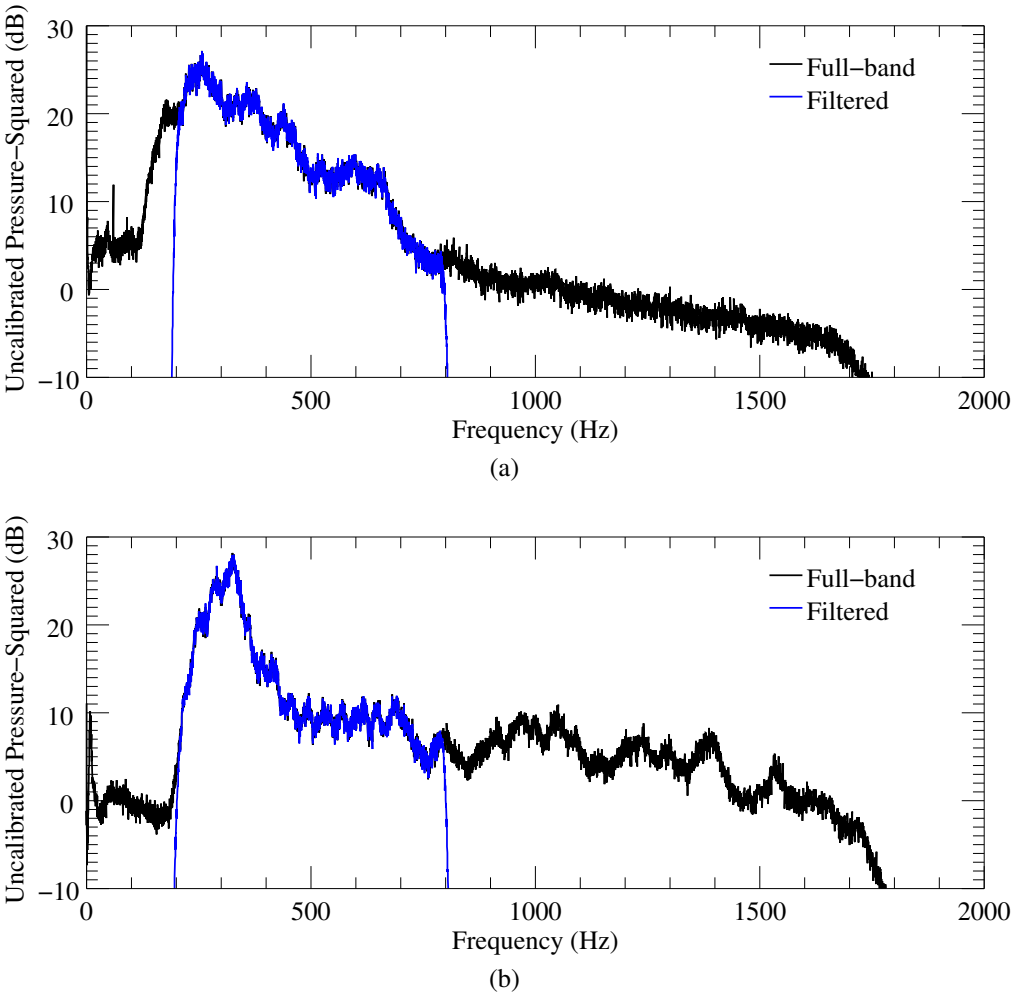


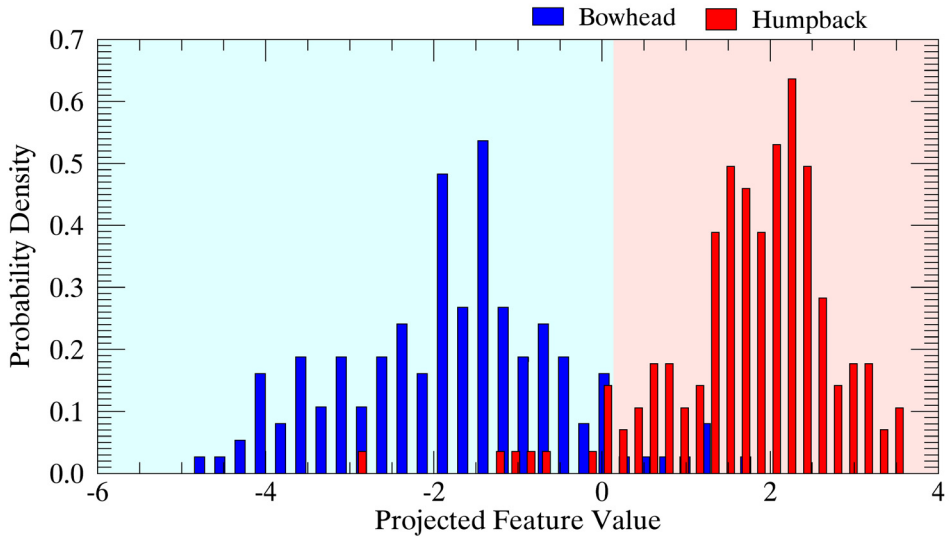
Figure 3.6: Comparison of full- and reduced-bandwidth spectra for (a) bowhead and (b) humpback whale calls. The 200 to 800 Hz bandpass-filtered spectra are also shown to highlight that the bandpass-filtered signals capture the majority of the power unique to each whale call.

Of course, it would have been preferable to transmit the vocalizations in their original frequency band; however, there are several advantages resulting from the necessity of shifting the frequency band of both the biogenic and synthetic vocalizations:

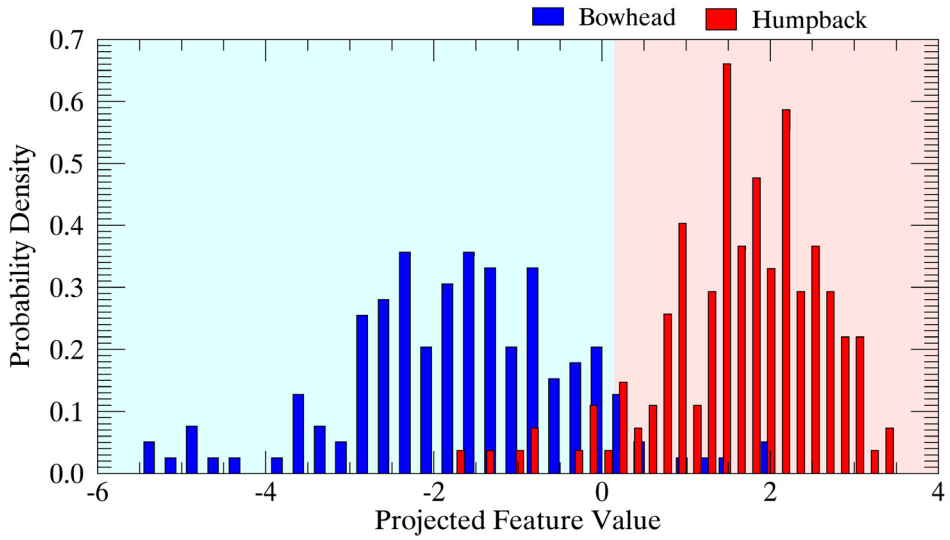
- experiments were approximately five times shorter in duration due to the time compression of the signal sequence;
- by scaling the signals up in frequency it was possible to note propagation effects on the signals at smaller spatial scales [58], allowing for lower source levels;
- the higher frequency signals attenuate more rapidly than lower frequency signals,

so that the spatial extent of the experiments was decreased, thereby reducing the environmental impact, in combination with the lower source levels; and

- the potential of behavioural impact on marine mammals in the area was limited because the signals would not be recognized as real whale vocalizations by marine life.



(a)



(b)

Figure 3.7: Results of training the classifier on biogenic bowhead and humpback vocalizations that were (a) full band and (b) bandpass filtered between 200 to 800 Hz.

In summary, the biogenic and synthetic data set for the experiment was prepared using the following steps: high SNR biogenic signals were selected, and synthetic signals were generated to mimic bowhead and humpback whale calls; all signals were bandpass filtered between 200 to 800 Hz; and then the playback speed was increased by a factor of five so that signals were transmitted in the 1 to 4 kHz passband of the transducer.

3.3.1 Time Series for Transmission

The time series for transmission during the experiments were generated by concatenating the conditioned signals with several marker signals. The same call types were transmitted in groups; that is, all the biogenic bowhead calls were transmitted, then the biogenic humpback vocalizations, the synthetic bowhead signals, and finally the synthetic humpback calls. A 1 s long linear frequency modulated (LFM) chirp, sweeping across the 1 to 4 kHz frequency band, was transmitted at the beginning of each of these groups. The matched-filtered LFMs could be used to accurately locate the beginning of each pulse sequence (as described in Section 3.4), which was particularly useful when the received level was low compared to the ambient noise level. Additionally, every 10^{th} signal in the series was a 1 s long pulsed Continuous Wave (CW) with the tone stepping up in frequency between 1 to 4 kHz. These CWs were useful for determining how far into any particular time series the recording was. Including all the whale calls and marker signals, the biogenic bowhead and humpback time series were each 13.5 minutes long with a vocalization once every 4.0 s and at least 3.6 s of quiet time between calls. Similarly, the time series composed of the synthetic whale calls were each 11.25 minutes long with a vocalization once every 4.0 s and at least 3.0 s of quiet time between calls. When all four time series were played in succession the approximate transmission time was 50 minutes.

3.4 Signal Detection

After the signals were transmitted and recorded, and before applying the classifier, the signals must be identified in the acoustic record. To do this, a matched filter was used to determine the location (in time) of the LFMs that mark the beginning of each signal set. Starting at the detected LFM, a portion of the acoustic record was extracted that included all transmitted signals in the set, taking advantage of the characteristics of the transmitted time series. A bandpass filter with low- and high-frequency cutoffs of 0.1 kHz and 6 kHz,

respectively, was applied to remove the DC-offset, low frequency noise, and high frequency dolphin whistles. The band-limited energy detector, described in Hood *et al.* and Bougher *et al.*, was applied using the following detection parameters: FFT parameters were selected that resulted in a frequency resolution of 15.625 Hz and time resolution of 0.0192 s, the frequency band spanned the full 1 to 4 kHz range of the signals, and the detection threshold (i.e., difference in dB between the signal and noise estimates) was set at 2 dB. These detection parameters were chosen to produce a large number of true detections, and as such there were also many false detections. False detections were removed by comparing the time of the detection with the *a priori* information on the spacing between recorded signals. Thus, for examining the effects of propagation on classification, the experimental dataset comprised the detected signals. This detection process was not applied to the signals which were artificially propagated through the modelled environments.

CHAPTER 4

PROPAGATION EXPERIMENTS

4.1 Introduction

No methodical experiment has been published in the open literature that considers transmitting whale or whale-like signals to assess the impact of environment-dependent propagation on a classifier algorithm designed for PAM of cetaceans. As discussed in the literature review (Section 1.1), some researchers, like Mouy *et al.* [23], attempt to evaluate the robustness of their detection/classification algorithms by selecting recorded whale calls for training and validation sets that incorporate various ambient noise characteristics and multipath propagation. Mercado *et al.* [27] did transmit whale-like signals to evaluate the effects of propagation on the calls; however, their end goal was not to evaluate the impact on a PAM system, rather it was to determine if there is sufficient information in a received call to estimate the range to a conspecific[1]. Although there is a vast body of literature on ocean acoustic propagation experiments with many different applications, little of this has been applied in the context of PAM of cetaceans; therefore, performing a propagation experiment to study the impact on PAM of cetaceans is novel.

An initial investigation of the *temporal* robustness of the aural classifier was performed for the active sonar implementation. Murphy and Hines [38] describe two experiments that were conducted at the same location, separated in time by approximately two years. The experimental procedure of the trials was consistent, so that the difference between the results of the two experiments was driven by substantial changes in ambient noise and sound propagation conditions. The ambient noise changes resulted from a significant difference

[1] A member of the same species

in sea state — Beaufort force 5 to 6 seas were observed during the first experiment and Beaufort force 1 seas throughout the second experiment. The sound speed profiles (SSPs) also underwent significant change; a downward refracting profile was measured in the first year, whereas an approximately isospeed profile was measured in the second year. Even though there were considerable changes in the propagation conditions, the aural classifier maintained excellent discrimination [48] (the AUC was found to be 0.86) when the classifier was trained with data from the first experiment and validated on the data collected two years later. This result indicates that, for the environment considered, propagation effects did not significantly impact the performance of the aural classifier; however, the scope of this investigation was somewhat limited since it only considered changes in ambient noise and SSP. The effect of geophysical properties of the seafloor were not considered. Additionally, the signal types used during the Murphy and Hines [38] experiment were chosen for their relevance to active sonar research. The work presented in this chapter describes an experiment designed explicitly to consider the effects of propagation on the aural classifier using signals relevant to PAM of cetaceans (i.e., the biogenic and synthetic vocalization data set described in Chapter 3).

4.2 Experimental Set-up

A two-day sea trial was conducted in the Gulf of Mexico, approximately 74 km south of Panama City, FL, from 30 April to 1 May 2013. Figure 4.1a contains a map with the trial area delineated by the red box. Two moorings were deployed on each day of the experiment, each with at least two hydrophones at different depths within the water column (refer to Table 4.1 for depths and Appendix A for mooring diagrams). Once the moorings were deployed the complete set of signals (both biogenic and synthetic vocalization sets with marker signals as described in Section 3.3.1) were transmitted to the moored receivers. An ITC-2010 projector deployed from the stern of DRDC's research ship, CFAV[2] QUEST, was used to transmit the signals while the ship drifted. The sound source was deployed at 20 m depth on day 1 and 40 m depth on day 2. In addition to the moored receivers, transmissions were monitored in real-time and recorded using a Reson TC4032 hydrophone deployed from QUEST's forward welldeck. The separation between the sound source and monitor hydrophone was approximately 70 m. A pictorial summary

[2]Canadian Forces Auxiliary Vessel

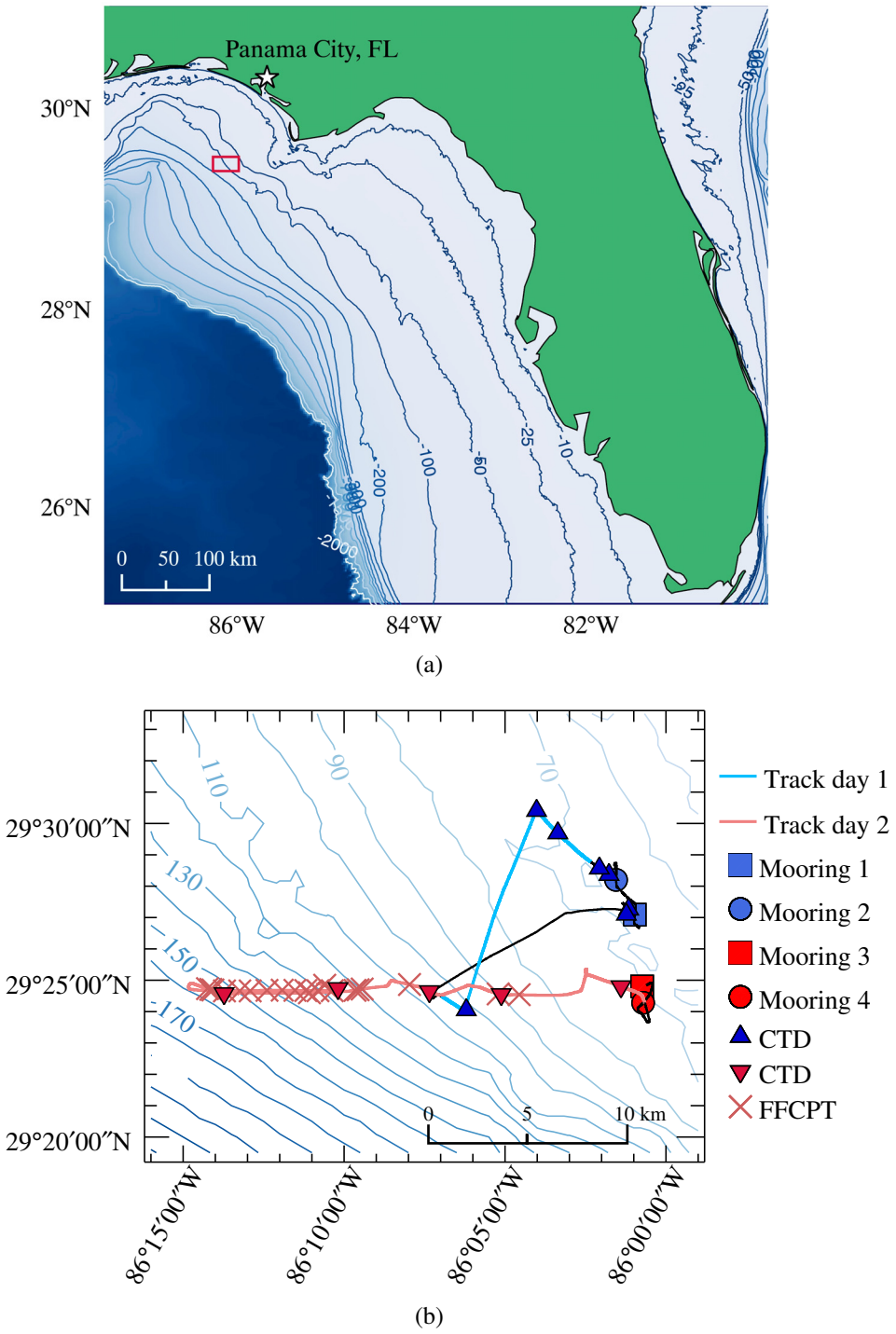


Figure 4.1: (a) Map of the experimental location with Panama City, FL represented by the white star, and the trial location in the area defined by the red box. (b) Map with ship tracks shown for both days of the experiment, including location of the moorings, CTD casts, and free falling cone penetrometer measurements. Elements from the first day and second days of the experiments are represented by blue- and red-coloured symbols, respectively. Contour lines represent isobaths, with depth in meters as labelled. This map is an enlarged and detailed version of the trial area represented by the red box in (a)

Table 4.1: Hydrophones deployed during the Gulf of Mexico sea trial on which the transmitted signals were recorded. Note that due to technical issues, not all of the hydrophones listed here collected data (the recorders which did not collect data are shown in grey text). Simulated results in subsequent chapters are compared with data from representative recorders, which are denoted by an asterisk (*).

Recording device	**Mooring number**	**Depth (m)**	**Notes**
30 April 2013			
Reson TC4032	0	30	Used to monitor transmissions in real-time
SHARP upper	1	40	No data recorded
SHARP lower	1	50	No data recorded
icLISTEN HF	2	36	
Whalesong SM2M	2	45	
1 May 2013			
Reson TC4032	0	35	Used to monitor transmissions in real-time
icLISTEN HF*	3	17	
SHARP upper	3	28.5	
SHARP lower	3	41.5	No data recorded
Whalesong SM2M	4	19	
icLISTEN HF*	4	29	

of the experimental setup is given in Figure 4.2. After the transmissions were completed (about one hour), the range to the moored receivers was increased and the transmissions were repeated. On the first day of the experiment the signal set was transmitted three times, over ranges of approximately 2, 5 and 10 km. The complete signal sequence was transmitted four times on the second day of the experiment, over ranges of approximately 1, 5, 10 and 20 km.

A detailed map of the experimental site, including the ship tracks in relation to the moorings, is provided in Figure 4.1b; precise transmission ranges are given in Table 4.2. The intent was to complete a different experimental geometry on each day: the first with transmission along an isobath, and the second with transmission across isobaths such that the water depth at the transmission locations would increase. Unfortunately, the northward current was much stronger during the first day than anticipated so that the ship drifted farther north than intended by the time the moorings were placed. Mooring 1 was located

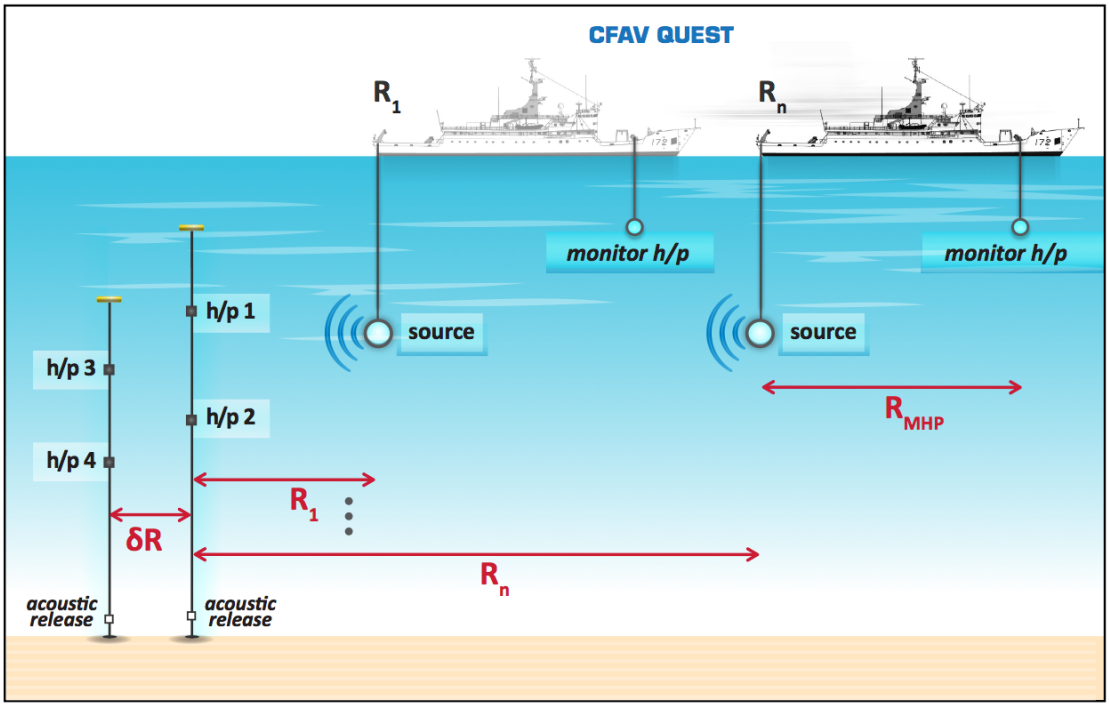


Figure 4.2: Representation of the experimental setup. The ship first deployed two hydrophone (h/p) moorings, moved to the first location and transmitted the full set of signals, then moved farther away from the recorders and retransmitted the signals. R_1 and R_n represent the horizontal ranges the signals propagated from the source to the midpoint between the moorings. Moorings were separated by an amount δR, where $\delta R \ll R_1 < R_n$. A monitor hydrophone was deployed near the bow of the ship (separated from the sound source by horizontal range R_{MHP}) to monitor the transmissions in real time and to provide signals for training the classifier.

at N 29°27′5″, W 86°0′58″ and mooring 2 at N 29°28′12″, W 86°1′33″; both moorings were deployed in 71 m deep water. After the first two transmission sets were completed it was noted that the ship was at the edge of the experimental area. To achieve a longer range transmission the ship then moved to the south-west of the moorings such that the transmission range was approximately 10 km, after which the moorings were recovered. Upon recovery of the moorings, it was noted that no data had been saved on the SHARP recording unit. At the beginning of the second day the moorings were reconfigured and redeployed: mooring 3 at N 29°25′20″, W 86°1′14″ and mooring 4 at N 29°24′23″, W 86°1′16″ in 75 m deep water. The ship moved westwards away from the moorings, sailing downslope so that the water depth at each of the transmission locations was greater than where the moorings were deployed. Although results from day 1 are included in this

Table 4.2: Transmission ranges of signals during the Gulf of Mexico sea trial.

Transmission Set	**Mooring 2 (km)**	
30 April 2013		
1	1.7 to 3.4	
2	3.7 to 5.5	
3	10.8 to 11.0	
	Mooring 3 (km)	**Mooring 4 (km)**
1 May 2013		
1	1.6 to 2.0	0.9 to 2.0
2	5.9 to 8.0	5.7 to 7.9
3	9.6 to 11.6	9.4 to 11.6
4	19.9 to 22.0	19.8 to 21.9

chapter, the focus (particularly for the modelling discussed in subsequent chapters) is on data collected during day 2, since there were fewer technical issues with the recordings and the geometry was more straightforward.

4.2.1 Acoustic Recording Equipment

Four different recording units were used throughout the experiment. As previously stated, a Reson TC4032 hydrophone was deployed from CFAV QUEST to monitor transmissions in real-time; the acoustic data measured by the hydrophone were recorded on a Reach Technologies' Analog Data Recorder [59]. The data recorded using the monitor hydrophone was useful for training the aural classifier, since they underwent the same transform applied by the transmission equipment, as the signals recorded on the moored units, but should have been minimally affected by propagation since the transmission range was two orders of magnitude shorter. The remaining three recording units were used as moored devices — the depths at which the recordings were made are listed in Table 4.1. The first of the moored units were the Subsurface High-fidelity Audio Recording Packages (SHARPs) developed by DRDC, configured to simultaneously record audio from two hydrophones. The remaining recorders are self-contained commercially available recording units with built in hydrophones: Wildlife Acoustics' Song Meter SM2M recording package [60] (referred to throughout this thesis as the Whalesong), and Ocean Sonics' icListen HF recording units [61, 62]. Each recording package had slightly different sampling schemes

Table 4.3: Sampling formats for acoustic recording packages.

Recording Package	**Bit Depth** (bits/sample)	**Sample Rate** (kHz)
Reach	16	44.1
SHARP	24	44.1
Whalesong SM2M	24	48
icListen HF	24	32

due to the configuration options available; configuration parameters for each of these recorders are listed in Table 4.3. Although the sampling formats are inconsistent, it should be noted that this is not likely to impact the results since the Nyquist frequency for each recorder significantly exceeds the maximum frequency of the transmitted signals.

4.3 Ocean Environment Measurements

As a complement to the acoustic measurements, monitoring of the ocean environment was an important component of the propagation experiment, necessary for understanding the propagation conditions at the time of the experiments, as well as for supporting propagation modelling efforts (discussed in Chapter 5). Wind speed information was derived from measurements made onboard CFAV QUEST and the surface conditions were noted in terms of the Beaufort Sea State Code [63]. The ambient ocean noise was examined by determining the spectrum levels in the band of interest during times in which no signals were transmitted. Information on the sediment characteristics was obtained from Free-Falling Cone Penetrometer (FFCPT) casts [64] taken along the ship's track on the second day of the experiment. Conductivity-Temperature-Depth (CTD) casts were performed at each transmission location, represented by the triangles in Figure 4.1b, to characterize the water column properties. The results of each of these environmental measurements are summarized in the remainder of this section.

4.3.1 Wind Speed and Surface Roughness Conditions

During the first day of the experiment the wind was from the east and increased slowly throughout the day from approximately 5 to 8 m/s, as measured by the ship's anemometer. Since some whitecaps (i.e., breaking waves) were observed throughout the day, the

conditions were assessed to be consistent with sea state 2 to 3. During the second day of the experiment there was also an easterly wind, with speed increasing from approximately 9.5 to 14.5 m/s. This resulted in the surface conditions evolving during the course of the experiment from sea state 2 to 3 with a few whitecaps noted, to sea state 4 to 5 with many breaking waves and some spray observed.

4.3.2 Ambient Noise

Samples of the ambient ocean noise were extracted from the acoustic recordings to facilitate analysis of the noise characteristics. An hour-long period of ambient noise recorded by the icListen hydrophone on the first day of the experiment was selected to represent the noise characteristics of day 1. There were no signal transmissions in this data segment. This same period of ambient noise was used as the source of additive noise in Chapter 7. Snippets of experimental noise from the time between recordings of the biogenic whale call transmissions conducted at 18:00 UTC, were used to represent the noise characteristics on day 2. The time period of these extracted snippets was carefully selected to ensure no transmissions were contained in them, and that the received level had returned to ambient conditions. These noise samples were recorded on the icListen hydrophone located at 29 m depth, and were used as additive noise in Sections 5.2 and 5.3.

Ambient noise pressure spectra (Figure 4.3) were generated from the representative data for day 1 and day 2. These were produced using a Hann-weighted window with 65 536 samples and 50 % overlap. The observed noise spectra are compared with noise spectra measured in coastal waters by Piggott [65] for wind speeds of approximately 5 and 8 m/s,. In the frequency bands of interest (200 to 800 Hz and 1 to 4 kHz), the noise spectra corresponded closely to the wind conditions at the time of the ambient noise measurements. The difference in spectral levels between day 1 and day 2 of the experiments was attributed to an increase in the wind speed between the two days. It is also worth noting that the measured ambient noise levels are significantly higher than the estimated system noise spectral levels [3].

[3] Estimates of the system noise spectral density were provided by Ocean Sonics for their icListen HF recording units: the reported levels were 41 dB re 1 μPa^2/Hz at 100 Hz [personal communication], and 30 dB re 1 μPa^2/Hz at 10 kHz [62].

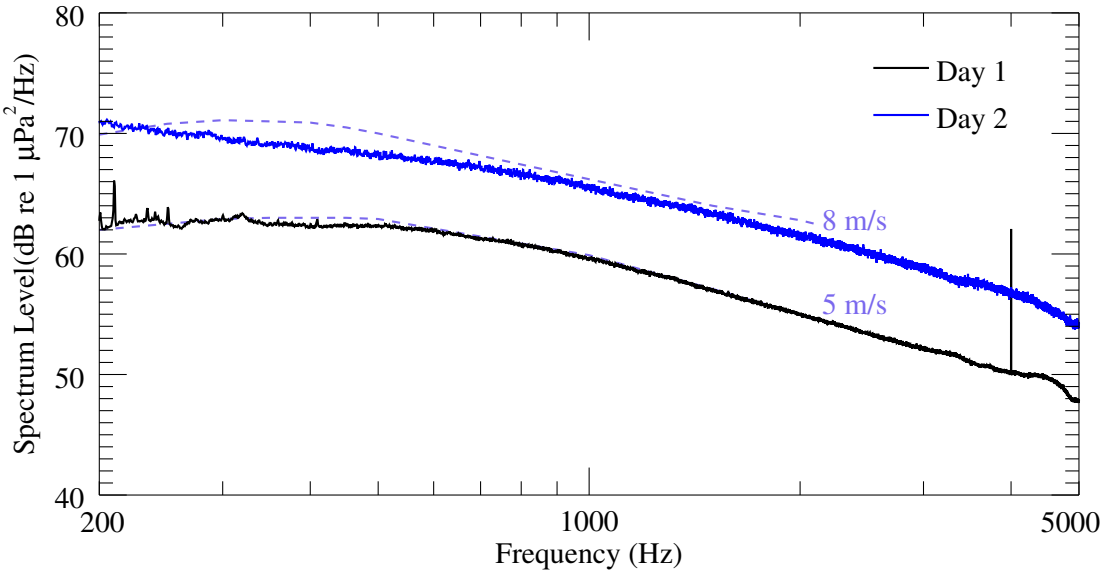


Figure 4.3: Ambient noise pressure spectra (solid lines) typical of noise recorded on day 1 and day 2 of the Gulf of Mexico Experiment. Piggott's [65] ambient noise spectra for coastal waters (dashed lines) with wind speeds of 5 and 8 m/s are included for comparison.

4.3.3 Sediment Properties

Sediment properties were measured using a FFCPT, which is an instrument that collects *in situ* sediment properties by dropping a probe into the sea floor while a ship is underway [66]. The FFCPT measurements allow one to estimate the sediment type (e.g., silty-clay, sand, etc.) based on the probe's acceleration, as a function of penetration depth and the dynamic sediment pore pressure [64]. Geoacoustic parameters of the sediment can then be estimated from the grain size and sediment-type using the Applied Physics Laboratory University of Washington's (APL-UW) ocean environmental acoustic models handbook [67].

Figure 4.4 shows the results of the FFCPT analysis. The along-track range is referenced to the location of mooring 3. FFCPT results were not available along the complete track; however, there was sufficient data collected that it is possible to make general assumptions about the bottom type. The left axis of the plot shows the probe penetration depth — this can be read from the plot by observing the lowest point at which a bin contains colour. The penetration depth indicates how deep measurements were available for the sediment type. The Robertson zone is expressed by the colour of each bin. For reference, the water depth is also plotted (solid black line), with associated values on the right axis. The sediment in this area was predominantly Robertson zones five and six, which are "sand mixtures"

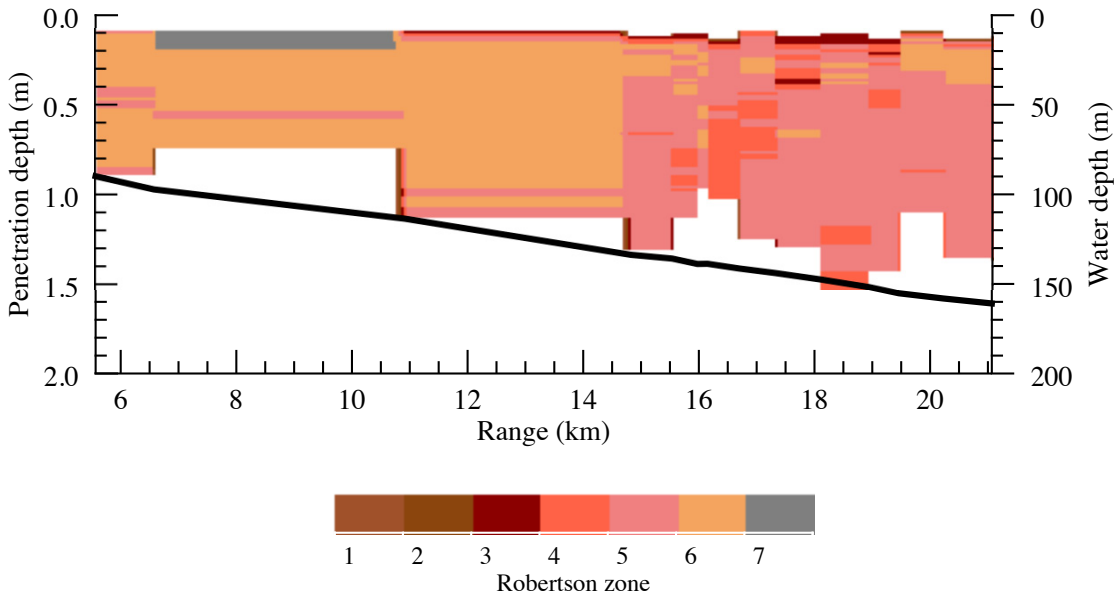


Figure 4.4: Sediment type, in terms of the Robertson zone, obtained from Freefall Cone Penetrometer measurements. In ascending order, the numerical values for the Robertson zones correspond to the qualitative sediment descriptions: fine, organic, clays, silts, sand mix, sands, and gravels. The water depth is indicated by the solid black line. Range is referenced such that $R = 0\,\text{km}$ is the location of mooring 3.

and "sands," with patches of finer grained sediments. A thin layer of clay and, at one drop location, gravel, overlaid the main sediment type. This surface layer was thin enough, compared to the wavelength of the acoustic waves, that it was neglected when performing propagation modelling. Observations of the sediment type were consistent with other measurements in the area, including Steele and Pecknold [66] at an experimental site closer to shore, and Balsam and Beeson's [68] analysis that showed the primary sediment type in the region is quartz sand.

4.3.4 Sound Speed Profiles

Each of the sound speed profiles, resulting from the CTD casts[4] conducted at the locations shown in Figure 4.1b, are shown in Figure 4.5. The three SSPs closest to mooring 2 measured on 30 April showed (what is referred to here as) an "anti-duct," that is, a maximum sound speed at depth that causes acoustic refraction away from the anti-duct axis. This term is used as a contrast to the more commonly used "duct," which has a region

[4]Sound speed values were output by the CastAway CTD [69], which used the Chen-Millero [70] equation to derive sound speed from measured temperature, salinity, and pressure.

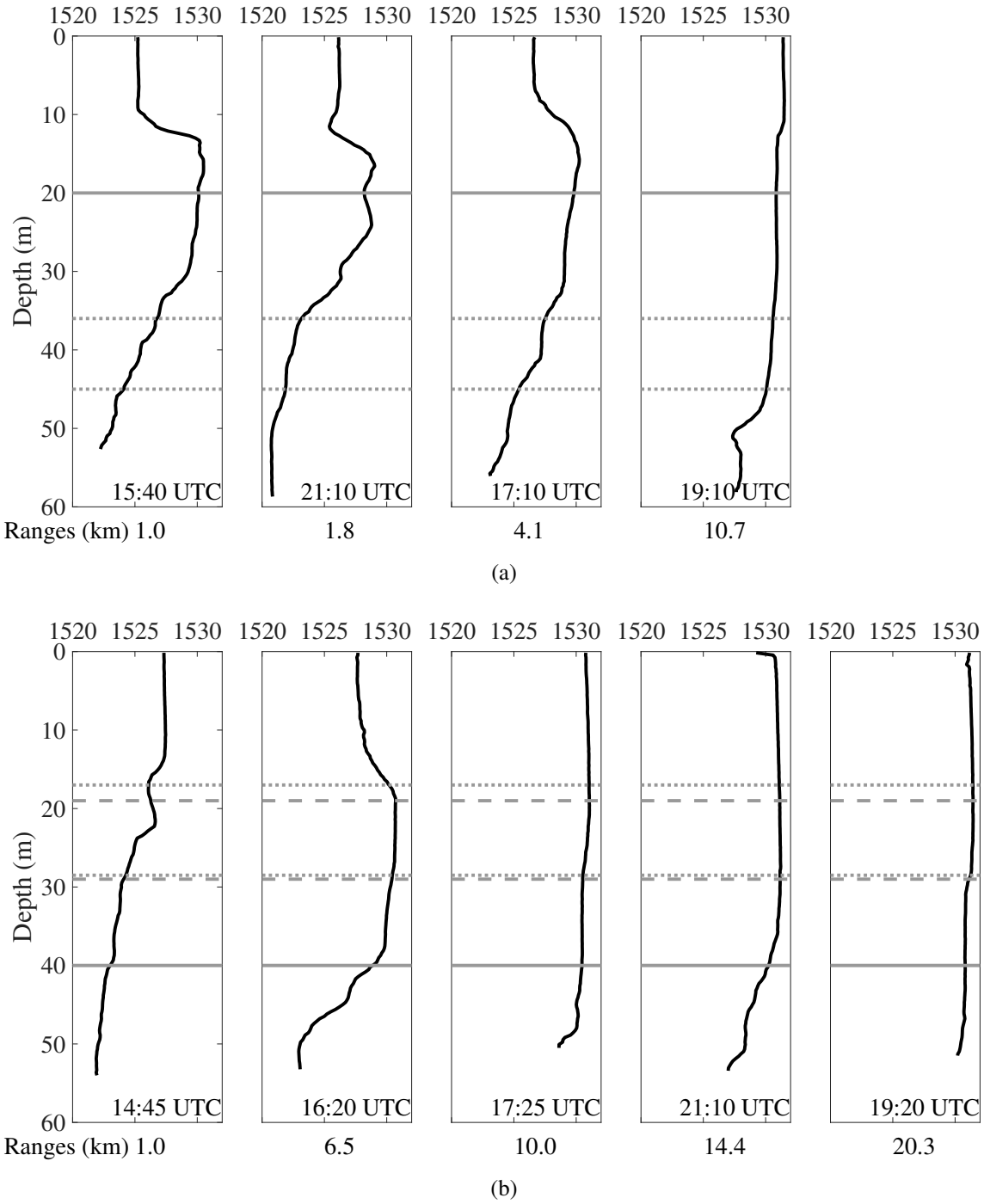


Figure 4.5: Sound speed profiles (SSP) in m/s observed (a) on day 1 and (b) day 2. Locations of the measurements are indicated on the map in Figure 4.1b. The solid grey line (———) indicates the source depth; dotted lines (· · · ·) represent depth of receivers on moorings 2 and 4, and the dashed lines (– – –) represent receivers on mooring 3. Ranges listed below each SSP provide the horizontal range between the location at which the SSP was measured and mooring 2 for day 1 profiles, and mooring 3 for day 2 profiles. The profiles are presented in order of increasing range, such that they are not necessarily ordered chronologically.

of minimum sound speed that causes refraction towards the duct axis. The SSP observed 10.7 km away from the mooring had a structure distinct from the others. It had an approximately isovelocity component to about 45 m depth, then a steep gradient to a sound speed minimum. The first SSP measured on the second day was downward refracting. The second profile was similar to the "anti-duct" profiles measured on the previous day. The last three SSPs developed an isovelocity layer near the surface that deepened during the day. The deepening of the isovelocity layer may be explained by the increased wind/wave height later in the day, leading to mixing in the upper water column.

The two anomalous sound speed profiles (measured at 19:10 on day 1 and 14:45 on day 2) are challenging to explain since both the time and spatial scales are relatively small for many oceanographic processes. Temperature-salinity (T-S) analysis [71] (results in Appendix C) indicated that all SSPs contained similar characteristics, except for the profiles measured at 19:10 on day 1 and 14:45 on day 2. This was a result of each of the profiles containing only 2 of the three water masses incorporated in all other profiles. From the available data, it is not possible to ascertain if this was a purely spatial or temporal phenomenon, in part because this is a dynamic region with significant inter-annual variability during the spring [72]. Nonetheless, based on the SSP observations and T-S analysis, a spatio-temporal event seems to be the most likely explanation. The best indication of this is the spatial separation of the SSP collected at 19:10 on the first day from the other SSPs collected the same day, and the proximity of this SSP with those observed at 16:20 and 17:25 on the second day. These profiles are separated by only a few kilometres but by more than 21 hours. Considering this evidence, a possible explanation for the anomalous SSPs is the passage of an eddy drawing in colder water from the shelf break [72], which was about 130 to 150 km away from the experiment location. Although the cause of the disparate SSPs remains uncertain, this discussion makes it apparent that the sound speed characteristics during the course of the experiment were range-dependent. This is an important point for understanding the propagation characteristics and for selecting an appropriate propagation model.

As previously mentioned, the ocean environment measurements are essential for the propagation modelling that is presented in subsequent chapters of this thesis. An extension of this requirement is that the environment properties were sufficiently sampled to be able to replicate the experimental location as a virtual environment. Unfortunately, this

condition was not adequately met for the sound speed data, as the CTD casts did not extend the full depth of the water column. To develop a comprehensive understanding of the propagation characteristics during the experiments, it was necessary to have a viable estimate of the water properties for depths greater than those reached by the CTD. Jochens *et al.* [72] provides a synthesis of the hydrography of the Northeastern Gulf of Mexico region which was used to produce such an estimate. Based on the data contained in the Jochens *et al.* report [72], at depths greater than 100 m the water characteristics were likely associated with Sargasso Sea Water. With this in mind, estimates for water properties at two depths were generated: at $z = 150$ m a value for temperature of 17 °C and salinity of 36.2 PSU were assumed, and at $z = 200$ m values of $T = 17$ °C and $S = 36.0$ PSU were estimated. The Mackenzie equation [73] was used to convert these temperature and salinity data to sound speed — values of c =1516.7 m/s and c =1517.3 m/s for the 150 m and 200 m depths, respectively, were thus added to the sound speed profiles to ensure accurate representation of the propagation characteristics at depth. It is worth noting that these hydrography-based estimates compare favourably with deep sound velocity measurements obtained using the sound velocity profiler on the FFCPT probe.

4.4 Source Level Estimates

To understand the impacts of propagation on the transmitted signals, one must first have reliable knowledge of the signals which were transmitted. One key aspect of the signals is their source level, SL. Based on the power supplied to the transducer during the course of the experiments, a reasonable estimate of the source level is possible; however, due to the difficulty of calibrating the complete system in advance of the trial, and adjusting operating parameters during the trial, the acoustic power transmitted into the water was somewhat uncertain. Thus, it was desirable to obtain a dependable source level estimate. With this in mind, a small section of this thesis is dedicated to discussing how the SL estimates were conducted, as there are several subtleties. Source level estimates are also useful for conducting follow-on modelling work (Chapter 5), when experimental conditions are replicated.

To corroborate source level estimates, two different methods were used; one method is based on transducer theory and the other on acoustic propagation theory. Full details of both these methods and results are outlined in the following sections. The first method

incorporated knowledge of the drive voltage levels of the transducer and its transmitting voltage response (TVR) to convert these into acoustic power. As there is some uncertainty in the exact voltage drive levels supplied to the transducer and the transmitting voltage response, a second method was used to confirm the transducer-based SL estimates. The second method determined the source level from the signals recorded on the monitor hydrophone deployed from the forward welldeck of the ship (see Section 4.2 and Figure 4.2 for full details on the experimental set-up). Received levels are subject to larger uncertainties because of variability in the path the signals travel between source and receiver due to small-scale fluctuations in ocean properties. For these reasons, it is beneficial to use two different methods to calculate source levels and compare the estimates for consistency. Once SL estimates are obtained using these two methods, they are compared with the maximum possible source levels based on the combined TVR for the amplifier and transducer system.

A few assumptions were made for both SL estimate methods that are worth noting here. First, in making the source level estimates, an omnidirectional source was assumed. This is a reasonable approximation since the the transducer has a toroidal beampattern such that it is omnidirectional in the horizontal. Neglecting the directional vertical component was valid since the mainlobe is quite wide (i.e., the vertical beamwidth is 70° and 90° at the 1 kHZ and 2.5 kHz resonance frequencies, respectively [57]) Secondly, only the LFMs transmitted at the beginning of each signal set (as described in Section 3.3.1) were used to estimate source level. This was done because LFMs are much simpler signals than the whale calls such that the SL was a more intuitive measure of emitted acoustic power for the LFMs than for the whale calls. Additionally, the propagation-based method relies on matched filtering to isolate the acoustic energy associated with the direct path arrival; in this case the LFM provides a clear advantage over the whale calls, as the matched-filter response for the LFM has a narrower peak and lower sidelobes than the matched-filter response of a whale call.

4.4.1 Method 1: Transmit Sensitivity and Transducer Input Voltage

The first method considers the manner in which the signals are generated to estimate the source level, and thus relies on transducer theory. To generate sound in the water a transducer is driven with a signal generator and power amplifier. The transducer may be characterized by the transmit sensitivity [74, p. 132], $\mathcal{S}_V$, typically given in terms of the

source level per volt of drive. Thus, the root-mean-square (RMS) source level, SL, in dB re 1 µPa @ 1 m can be determined using,

$$\mathrm{SL} = \mathcal{S}_V + 20 \log V_{\text{in}} \quad , \tag{4.1}$$

where V_{in} is the RMS voltage at the input of the transducer [14, p. 29] and $\mathcal{S}_V$ is in units of dB re 1 µPa/V @ 1 m.

The TVR from calibration of the ITC 2010 transducer (refer to Figure 3.5) was used to estimate the transmit sensitivity. In the bandwidth of interest (i.e., 1 to 4 kHz) the TVR yielded a transmit sensitivity of $\mathcal{S}_V = (126 \pm 1)$ dB re 1 µPa/V @ 1 m. In equation 4.1 only V_{in} will vary for each signal transmission since $\mathcal{S}_V$ is a constant for the transducer. During the experiments the value of V_{in} was monitored and recorded by the Reach recording unit. During the recording process the V_{in} time series was digitized, thus the recordings must first be converted from counts, C, (as they were recorded in the WAV file) to voltage, to be used in the SL calculation. This is accomplished by noting that 16 bits were used to record the voltage, which was scaled so that the maximum signal extent was ± 5 V peak-to-peak. Thus, the conversion from counts to voltage may be completed as follows,

$$\begin{aligned} 2^{16-1} &= 5V_{\text{p}} \\ 2^{15} &= \frac{5}{\sqrt{2}} V_{\text{RMS}} \quad , \end{aligned} \tag{4.2}$$

where V_{p} is the peak voltage and V_{RMS} is the RMS voltage. Noting that $\frac{5}{\sqrt{2}} = 2^{1.82}$ gives,

$$\begin{aligned} 2^{15} &= 2^{1.82} V_{\text{RMS}} \\ \mathrm{CTS} &= 20 \log \left(\frac{2^{13.18} C}{V_{\text{RMS}}} \right) \\ &= 79.35 \,\text{dB re 1}\, C \,/\, V_{\text{RMS}} \quad , \end{aligned} \tag{4.3}$$

where CTS is a conversion factor, in dB, necessary to convert from bits to voltage. Substituting this conversion into Equation 4.1 gives the following equation to determine the

source level of the transmitted signals,

$$\begin{aligned} \mathrm{SL} &= 20\log(1000C) + \mathcal{S}_V - \mathrm{CTS} \\ &= 20\log(1000C) + 126\,\mathrm{dB\,re\,1\,\mu Pa/V\ @\ 1\,m} \\ &\quad -79.35\,\mathrm{dB\,re\,1}\,C\,/\,V_{\mathrm{RMS}} \quad , \end{aligned} \tag{4.4}$$

where a scaling factor of 1000 is applied to C to take into account that the voltage recording was configured such that 1 mV in the recording represented 1 V supplied to the transducer.

4.4.2 Method 2: Received Signal Level on Monitor Hydrophone

The propagation method makes use of the signals recorded on the Reson monitor hydrophone. Recall that this hydrophone was deployed near the bow of CFAV QUEST, with approximately 65 to 70 m separation from the acoustic source (as in Figure 4.2). The sonar equation [10, 75] was used to estimate the source level in decibels,

$$\mathrm{SL} = \mathrm{RL} + \mathcal{S}_H + \mathrm{TL} \quad , \tag{4.5}$$

where RL is the sound pressure level, in dB re 1 µPa, received at the hydrophone, $\mathcal{S}_H$ is the hydrophone sensitivity in dB re 1 V/µPa, and TL is the one-way transmission loss in dB. Based on calibration of the monitor hydrophone, a value of $\mathcal{S}_H = (167 \pm 1)$ dB re 1 V/µPa was used.

A free-field measurement is typically required when calculating SL from the RL; however, this was not possible given the experimental setup. Thus, one must be careful to include only the signal energy from the direct arrival and to exclude contributions from successive arrivals (e.g., surface and bottom reflections). This is accomplished by matched filtering [16, 76] the hydrophone recording and time-gating the signal around the direct arrival. The time-bandwidth product determines the temporal resolution, Δt, of a matched filter such that $\Delta t = 1/B$, where B is the bandwidth of the matched filter [77]. Thus, the time-gating method was possible since the LFM had a 3 kHz bandwidth, which provided a 0.3 ms temporal resolution. An example of the envelope of the matched-filter output is shown in Figure 4.6. The time scale is defined such that the signal transmission occurred at time $t = 0$ s. The peak at $t = 0$ s is cross-talk from the amplifier and was used to compute the range between the source and receiver; this is discussed in greater detail in Appendix

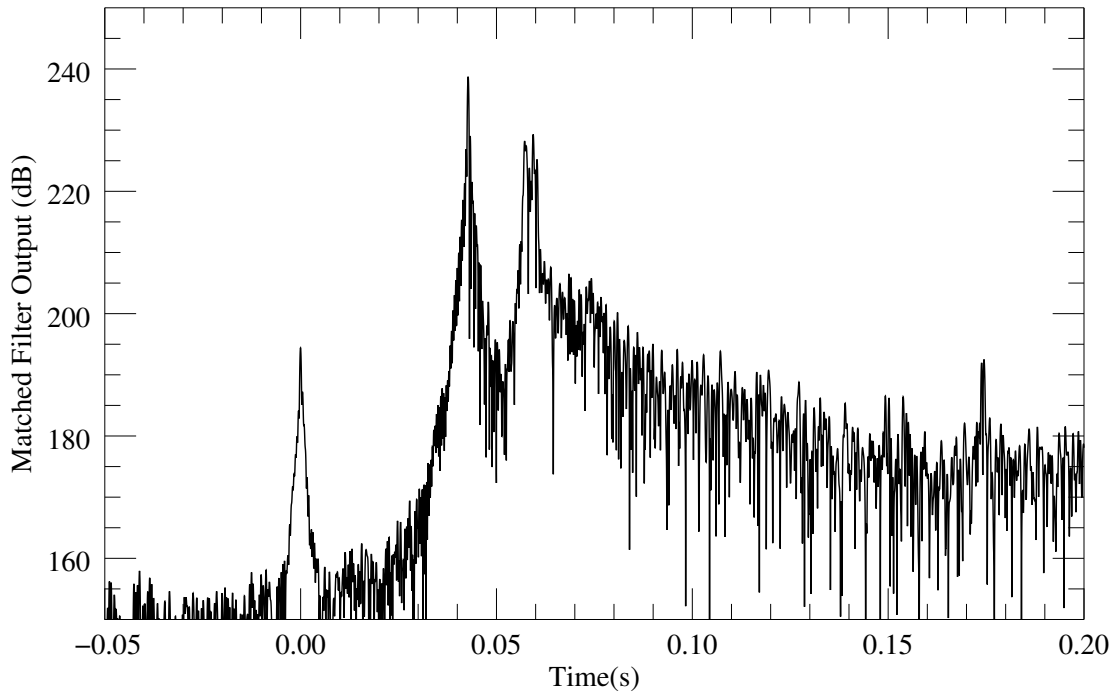


Figure 4.6: Envelope of matched-filter results showing arrival structure of an example LFM recorded on the Reson monitor hydrophone. The time scale is defined such that the signal transmission is at time $t = 0$ s.

D. The arrival structure shows the direct arrival at $t = 0.0428$ s, a strong surface reflection at $t = 0.0594$ s and a weaker bottom reflection at $t = 0.1745$ s. A 0.01 s segment, centred around the peak associated with the direct arrival, was extracted from the matched-filter output, convolved with an LFM that was scaled to the same signal level as the transmitted LFMs, and then the RMS value of the resulting LFM, $C_{\mathrm{MF,RMS}}$, was computed. Thus, the source level was determined from the signals recorded on the monitor hydrophone using,

$$\mathrm{SL} = 10 \log C_{\mathrm{MF,RMS}} - \mathrm{CTS} + \mathcal{S}_H + \mathrm{TL} \quad . \tag{4.6}$$

Bellhop was used to estimate TL with sound speed profiles and bottom properties measured during the experiment as inputs. Of course spherical spreading could have been used to estimate the transmission loss, but because multiple reflections were evident in the arrival structure (Figure 4.6) a ray model was used to obtain a more accurate TL estimate than would have been possible if a free-field environment was assumed. The following input values for the experimental geometry were used for the TL modelling: source depths

of z_{s} =20 m and z_{s} = 40 m were used on 30 April and 1 May, respectively. Likewise, the receiver depths were set at z_{r} =30 m for day 1 and z_{r} = 22 m for day 2. These depth values were obtained from pressure loggers located on the source and the Reson monitor hydrophone. Values for water depth, h, changed throughout the course of the experiment. It was found that TL modelling was not sensitive to changes of a few meters in h, so the following values were used for modelling transmission loss: h = 75 m on 30 April, and h = 110 m for signals transmitted at 17:00 and h = 160 m at 19:00 on 1 May.

The transmission loss value in Equation 4.6 could vary between transmissions because the source-receiver separation was not constant throughout the course of the experiment, due to the relative motion of the source and receiver induced by the drifting ship. It is worth noting that this effect was not large enough to alter the signal coherence during a transmission, rather it just affected the transmission range on a signal-by-signal basis. For each received signal used to estimate SL, time of flight was used to determine the source-receiver separation, r_1,

$$r_1 = c\Delta t_{\mathrm{direct}} \quad , \tag{4.7}$$

where $\Delta t_{\mathrm{direct}}$ is the time delay between the signal transmission and when the direct arrival was received, and c is the sound speed in water assuming an isospeed profile. The value of $\Delta t_{\mathrm{direct}}$ was determined from the time difference between the peaks of the enveloped matched-filter output associated with the transmission and the direct arrival. A value of c = 1530 m/s was assumed, based on sound speed profiles measured *in situ*. The transmission range determined from Equation 4.7 was also used as an input to the Bellhop model to estimate TL.

4.4.3 Source Level Results and Discussion

The first step in estimating the source levels using the received levels from signals recorded by the Reson hydrophone on the well-deck of CFAV QUEST was to determine the distance the signals were transmitted following the method outlined in Section 4.4.2. The one-way incoherent transmission loss between the source and receiver — required to estimate the SL in Equation 4.6 — was determined using Bellhop with the source-receiver separation, r_1, estimated from the time-of-flight (refer to Table D.1 in Appendix D for exact values) rounded to the closest five meter multiple, that is, a value of either 65 or 70 m was used for r_1. Modelling was done using the center frequency of the band of interest, f = 2500 Hz.

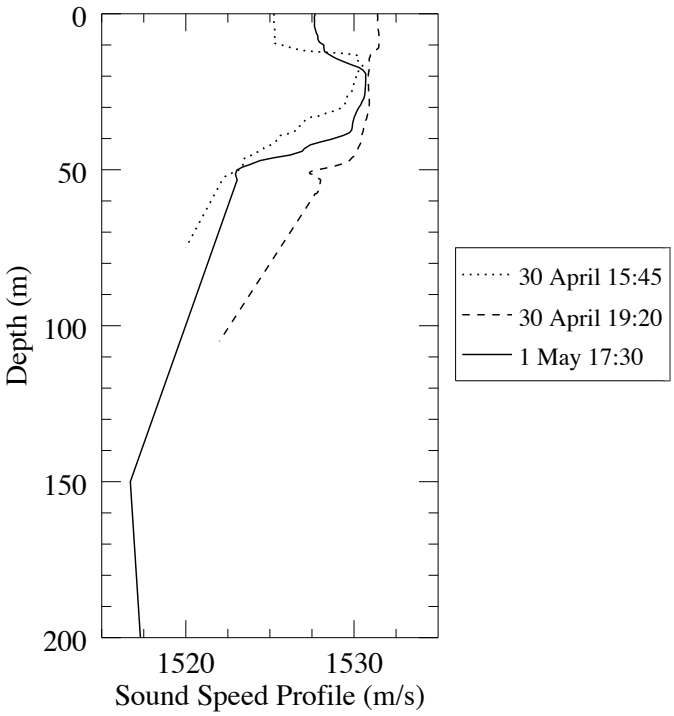


Figure 4.7: Sound speed profiles used to calculate the transmission loss for estimating the source level from signals recorded on the monitor hydrophone.

Three different environment models were used to estimate TL: two for 30 April because of markedly different sound speed profiles, and one for 1 May. The sound speed profiles used for the transmission loss modelling are shown in Figure 4.7. A sandy bottom was used for both days, with a sound speed of 1650 m/s, density of 1.9 g/cm^3, and compressional attenuation of 0.8 dB/λ_p [11, 78]. The Beckmann Spezzichino surface loss option was selected, with the wind speed set at 20 kts. TL results from the Bellhop model are given in Table 4.4.

Table 4.4: Transmission loss values used for estimating the source level of transmitted signals.

Date	**Range (m)**	**TL (dB)**
30 April 2013	65	37.6
	70	38.3
1 May 2013	65	37.6
	70	38.2

Table 4.5: Estimated source level of transmitted signals. All source levels are given in units of dB re 1 µPa @ 1 m. The number of signals, n, used to produce each estimate is also listed.

		Transducer V_{in}		**RL on Monitor Hydrophone**	
Date	$\boldsymbol{n}$	$\boldsymbol{\mu_{\text{SL}}}$	$\boldsymbol{\sigma_{\text{SL}}}$	$\boldsymbol{\mu_{\text{SL}}}$	$\boldsymbol{\sigma_{\text{SL}}}$
30 April 2013	9	175	2	174	1
1 May 2013	6	184.4258	0.0005	184.9	0.1

Finally, the source levels of the signals were estimated for each day of the experiment using Equations 4.4 and 4.6; results are presented in Table 4.5. It can be concluded from these results that on 30 April $\text{SL} = 174$ dB re 1 µPa @ 1 m and on 1 May $\text{SL} = 184$ dB re 1 µPa @ 1 m. The difference in source levels between the two days arose from the fact that the 1000 V tap was used on the amplifier on 30 April, and the 2000 V tap was used on 1 May. Also of note is that σ_{SL} is larger on the first day of the trial, this was because settings (e.g., attenuation in signal generation software, amplifier attenuation) were adjusted more frequently on 30 April to achieve a balance between obtaining maximum source level while avoiding clipping of the signal. On both days σ_{SL} is larger for the second method since the parameters used to calculate SL are subject to greater fluctuations. Some sources of variability include those outlined in Appendix D for estimating Δt_{sb}, as well as fluctuations in properties of the ocean medium, inhomogeneities in boundary layer properties, and change in angle of the source such that it was possible the main beam was not directed towards the receiver.

The amplifier and transducer combination used for this experiment was calibrated approximately a month in advance of the experiments. From the TVRs generated from these calibration results, it was predicted that the maximum source level possible from the 1000 V tap on the amplifier was approximately 180 dB re 1 µPa/V @ 1 m, or 186 dB re 1 µPa/V @ 1 m if the 2000 V amplifier tap was used. Comparing the source level estimates from the experiments with the TVR levels shows that on day 1 the source levels of the transmissions were well below the maximum possible SL, but that the level of signals transmitted on day 2 were closer to the theoretical maximum power. The difference on the first day may be attributed to adjustment of hardware and software settings in the signal generation chain, and that the waveforms supplied to the signal generator were not scaled to the maximum

level of the 16-bit WAVE file (this was corrected by the second day). Determining the SL using these different methods demonstrates the importance of confirming back-of-the-envelope calculations. If only the simple method (i.e. estimating from TVR) to determine SL had been implemented, the estimates could have been off by up to 6 dB re 1 µPa @ 1 m if the gains/attenuations in the signal generation chain had not been included. Because these values were changed frequently, especially on 30 April, there was some uncertainty that they had all been properly documented. This is a general concern when experiments are being conducted at an operational pace and automatic logging cannot be employed. To compensate for this there must be a thorough quality control and analysis of experimental data, as is done here. In addition to providing SL estimates, this section also illustrates the care that one must take when estimating source levels as there are many subtleties that need to be properly addressed.

4.5 Experimental Results and Discussion

Now that all aspects of the experimental setup have been addressed, the acoustic recordings can be examined in the context of investigating the impact of signal attenuation and distortion on the performance of the aural classifier. The acoustic data collected during the experiment were processed as follows: The signals were identified in the recordings using a frequency band-limited energy detector and compared to the known time that signals were transmitted to remove false detections, as described in detail in Section 3.4. After each signal was detected, a 4 s segment of the signal was extracted with the detection located approximately in the centre of the segment. Each extracted detection was saved to a WAVE file and bandpass filtered to remove the DC-offset applied by the recording equipment, as well as some of the low-frequency noise and high-frequency odontocete clicks interspersed through the recordings. The received signals were then processed using the aural classifier algorithm, as described in Chapter 2. Throughout the presentation of the following results, and subsequent discussion, recall that the experimental results are representative of the trends one would expect when performing PAM of bowhead and humpback whales — one must be careful in using the absolute values since the signals were transmitted in a frequency band five times higher than the band of the whale calls.

Table 4.6: Number of detections and classifier performance for biogenic bowhead and humpback calls recorded by the icListen unit on mooring 4. No uncertainty estimates are provided here for the classifier performance results because k-fold cross-validation was not implemented.

	Number of Detected Calls		**Performance Results**	
Transmission Range (km)	**Biogenic Bowhead**	**Biogenic Humpback**	**Accuracy (%)**	**AUC**
0.07	155	155	97	0.99
1	151	151	93	0.97
5	128	132	55	0.64
10	21	82	33	0.56
20	0	0	–	–

4.5.1 Example Decision Regions

The classifier was trained on data recorded by the monitor hydrophone and validated on data transmitted through the water over ranges of 0.9 to 22 km. Examining and comparing the decision regions for signals transmitted over each of the ranges enabled a qualitative analysis of the classifier's robustness to propagation. Example decision regions for the biogenic and synthetic vocalizations recorded on the icListen recorder deployed on mooring 4 are depicted in Figures 4.8 and 4.9. Bear in mind that fewer signals were classified at the longer ranges because only detected signals were provided to the aural classifier. The resulting number of detected calls which were classified are summarized in Tables 4.6 and 4.7. From these tables, the reader will observe that none of the calls transmitted over 20 km were detected as there was insufficient signal level. Thus, none of these calls are included in the subsequent analysis of aural classifier performance. In future experiments, in order to be able to analyze the impact of propagation on signals transmitted over ranges greater than 10 km, the source level should be increased; alternatively, an array of receivers could be used to take advantage of array gain. Tables 4.6 and 4.7 also provide a summary of the classification performance metrics. No uncertainty estimates are provided with these values, as cross-validation was not implemented — instead they are the performance metrics associated with the decision regions in Figures 4.8 and 4.9, for which all the detected signals in each range set were included in either the training or validation data set.

Let us first consider the biogenic decision regions. These were produced using twenty

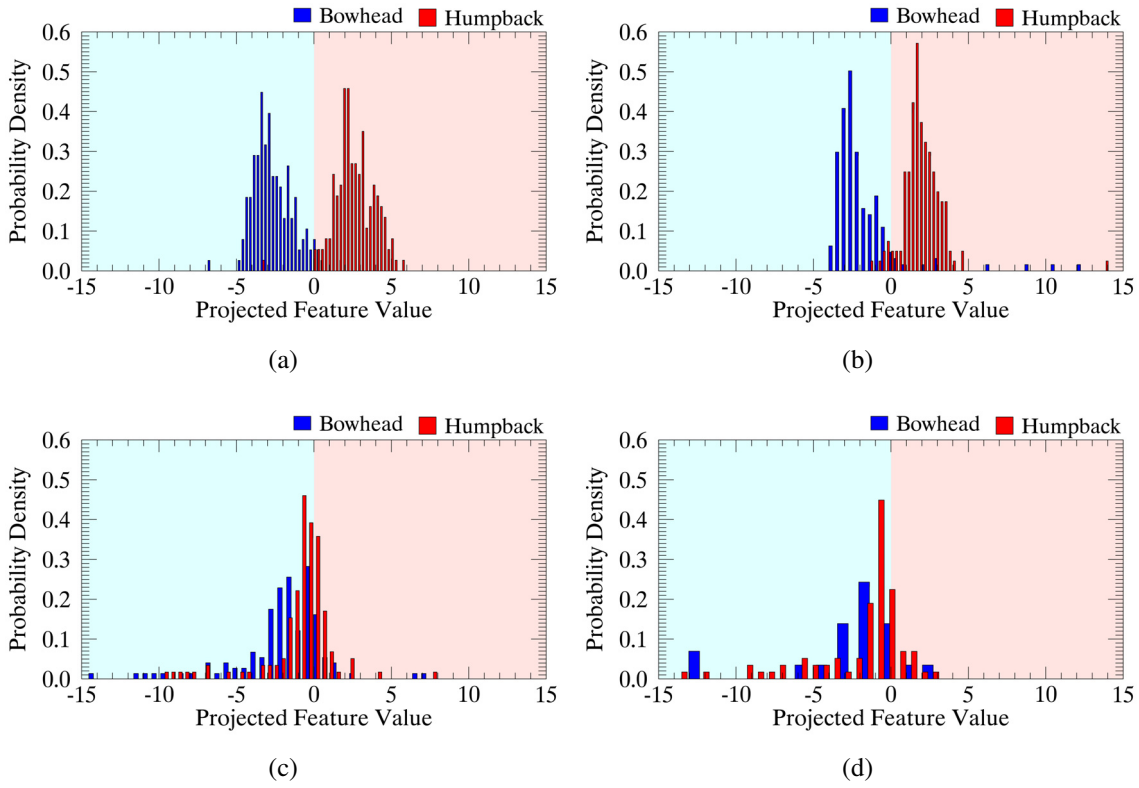


Figure 4.8: Decision Regions generated by training on data from the monitor hydrophone (training results shown in (a)) and testing on data transmitted over approximately (b) 1 km, (c) 5 km, and (d) 10 km. The biogenic bowhead and humpback calls were used that were recorded by the icListen unit on mooring 4. No uncertainty estimates are provided here for the classifier performance results because k-fold cross-validation was not implemented.

of the best perceptual features selected using the discriminant score ranking method with the signals in the training data set. The training decision region is presented in Figure 4.8a; in combination with the performance results in Table 4.6, it can be seen that the classifier produces outstanding discrimination between the biogenic whale vocalizations. Although the same can be said of the classifier performance when the classifier was validated on the signals transmitted over 1 km, it was no longer the case when signals were transmitted over the two longer ranges. The classifier performance was significantly degraded for the vocalizations transmitted over the 5 and 10 km ranges. From the decision regions it can be seen that this was a result of the humpback distribution increasingly shifting towards the bowhead humpback distribution, such that the overlap between the distributions increased. Additionally, there was a trend for the width of the distributions (i.e., within class variance) to increase as well. Both of these factors drove a decrease in accuracy and AUC as a

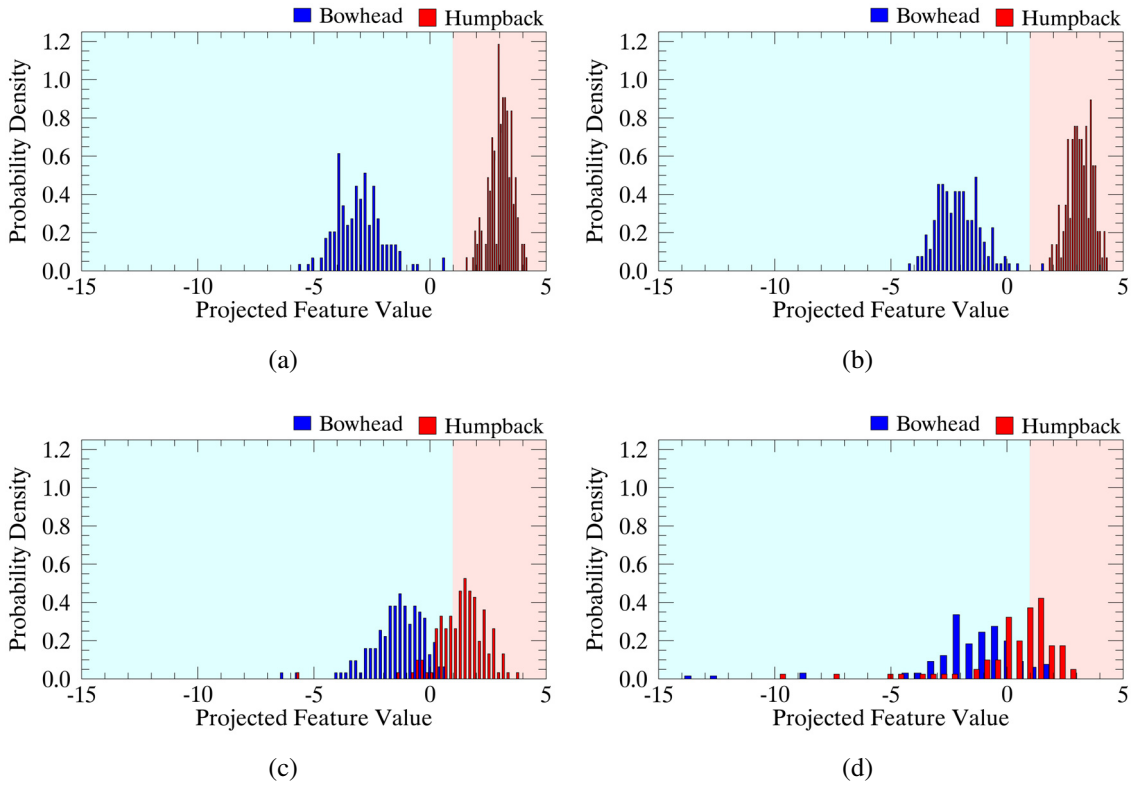


Figure 4.9: Decision Regions generated by training on data from monitor hydrophone (training results shown in (a)) and testing on data transmitted over approximately (b) 1 km, (c) 5 km, and (d) 10 km. The synthetic bowhead and humpback calls were used that were recorded by the icListen unit on mooring 4. No uncertainty estimates are provided here for the classifier performance results because k-fold cross-validation was not implemented.

function of transmission range by reducing the information used to discriminate between the bowhead and humpback classes.

The decision regions and classifier performance results obtained using the synthetic vocalizations displayed a similar trend as noted for the biogenic vocalizations — as the transmission range increased, fewer signals were detected and the classifier performance decreased. There were a few differences, however. A greater number of synthetic calls were detected at the 5 and 10 km ranges than the number of detected biogenic calls at the same ranges. This may have been because there was greater variability in the biogenic calls than the synthetic, such that it was possible to better tune the detector to the synthetic calls. Classification performance for the synthetic training data set and signals transmitted over 1 km was greater than for the biogenic calls. Moreover, classification performance did not degrade as quickly for the synthetic vocalizations — classification was acceptable

Table 4.7: Number of detections and classifier performance for synthetic bowhead and humpback calls recorded by the icListen on mooring 4. No uncertainty estimates are provided here for the classifier performance results because k-fold cross-validation was not implemented.

	Number of Detected Calls		**Performance Results**	
Transmission Range (km)	**Synthetic Bowhead**	**Synthetic Humpback**	**Accuracy (%)**	**AUC**
0.07	155	155	100	1.00
1	148	148	100	1.00
5	148	148	83	0.97
10	119	87	71	0.79
20	0	0	–	–

(following the terminology defined in Section 2.2) at the 10 km range for the synthetic calls, whereas there was almost no discrimination between the bowhead and humpback vocalizations for the biogenic case. The shift of the humpback distribution was more obvious for the biogenic calls than the synthetic calls, and the within class variance increased less for the synthetic calls; for example, for the biogenic case the mean value of the humpback distribution decreased by 3.5 and the within class variance increased by 3.5 from the training case to 5 km range scenario, whereas the mean value only changed by 1.7 and the within class variance by 0.9 for the synthetic humpback calls. This difference between the synthetic and biogenic calls was likely due to the fact that the initial biogenic call data set had higher variability, in particular four humpback units were included in the biogenic data set but only one was used for the humpback synthetic calls. Nevertheless, the key point is that the general trend of decreasing aural classifier performance with increasing transmission range was consistent between the biogenic and synthetic vocalizations.

So far the focus of this discussion has been on the the degradation of the aural classifier performance in general. It is important to realize that this was a result of how the individual perceptual features were altered by the impacts of propagation. For instance, the duration of the biogenic bowhead calls tended to decrease with increasing transmission range, partly because it became difficult to determine the precise beginning and ending of each signal in low SNR circumstances. This is just one particular example of how the perceptual features may be impacted by signal distortion and decreasing SNR. The full details of how the

individual perceptual features are altered are left for future work, since the way in which propagation impacted the individual features was non-trivial. Instead, the focus of the current work is on how the combination of perceptual features used by the aural classifier, as measured by the classifier performance, is impacted by propagation. Regardless, it is important to realize at this point that, although examining the individual features provides insight into how the physics of propagation impacts the aural classifier, it is difficult to link a single physical phenomenon with decreasing classifier performance. Based on the evidence presented thus far it is only possible to discuss performance in generalities; that is, in general, the aural classifier performance degrades with increasing transmission range as a result of increased overlap between the class distributions. Further sections of this thesis will provide more clarity on the mechanism responsible for the observed decrease in classifier performance.

4.5.2 Summary of Performance Results

The previous results exhibited classifier performance for a single recording unit (the icListen on mooring 4). Insight into how the class distributions shifted relative to the distributions used to train the classifier, and relative to each other at longer ranges, was gained from analyzing the decision regions. Now that an understanding of what drives changes in the performance metrics has been developed, the focus is shifted to summarizing classifier performance for all received signals and examining how performance changes as a function of range and signal-to-noise ratio.

Performance results, in terms of classification accuracy and AUC, as a function of transmission range for data recorded on each of the recording units are presented in Figures 4.10 and 4.11 for the biogenic and synthetic calls, respectively. As before, the classifier was trained on data from the monitor hydrophone (represented by the black circle on each plot) and validated on data transmitted over successively longer ranges. Error bars are provided for each data point — the vertical error bars represent one standard deviation of the performance as determined from 5-fold cross-validation and the horizontal error bars indicate the ranges at which the signal transmissions started and ended. Here, the data were plotted for each recording device individually to determine if there was a significant difference in performance that may be attributed to the depth of the recording device.

As expected, the classifier performance for both the biogenic and synthetic vocalizations was found to be range-dependent. In particular, the performance monotonically decreases

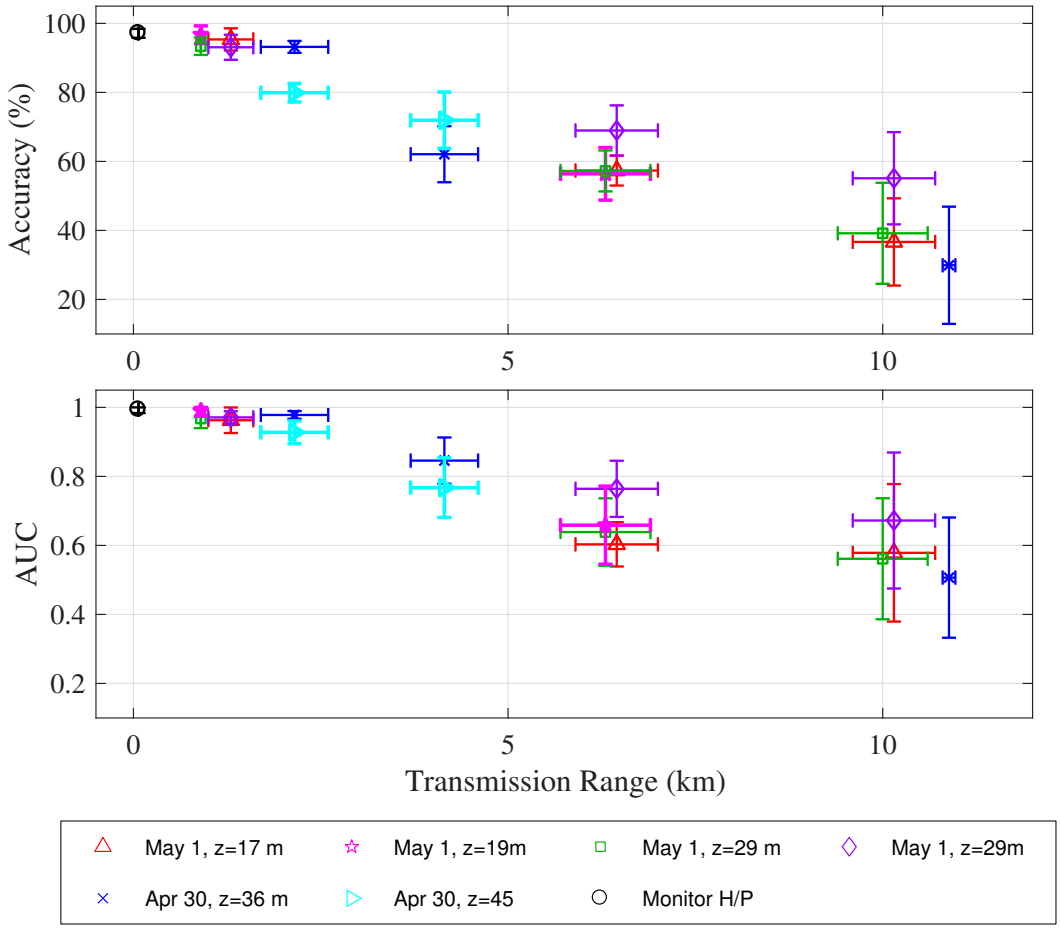


Figure 4.10: Experimental classification performance as a function of range for the biogenic whale calls. The vertical error bars are one standard deviation of the classification performance based on 5-fold cross-validation, and the horizontal error bars define the shortest and longest transmission range for each transmitted signal set.

with increasing transmission range (if one considers the data from each recorder individually). Consider the classifier performance for the biogenic vocalizations: for short transmission ranges the performance was outstanding, as the transmission range increased the performance gradually decreased until it approached the lower limit for performance when signals were transmitted over approximately 10 km. Interestingly, although the performance for the synthetic calls also decreased with increasing range, the decrease was not as gradual as for the biogenic vocalizations. Instead, the AUC still showed excellent discrimination between classes when signals were transmitted over approximately 7 km. As discussed in the previous section, this was likely due to the increased separation between the bowhead and humpback classes and the smaller within class variance for the

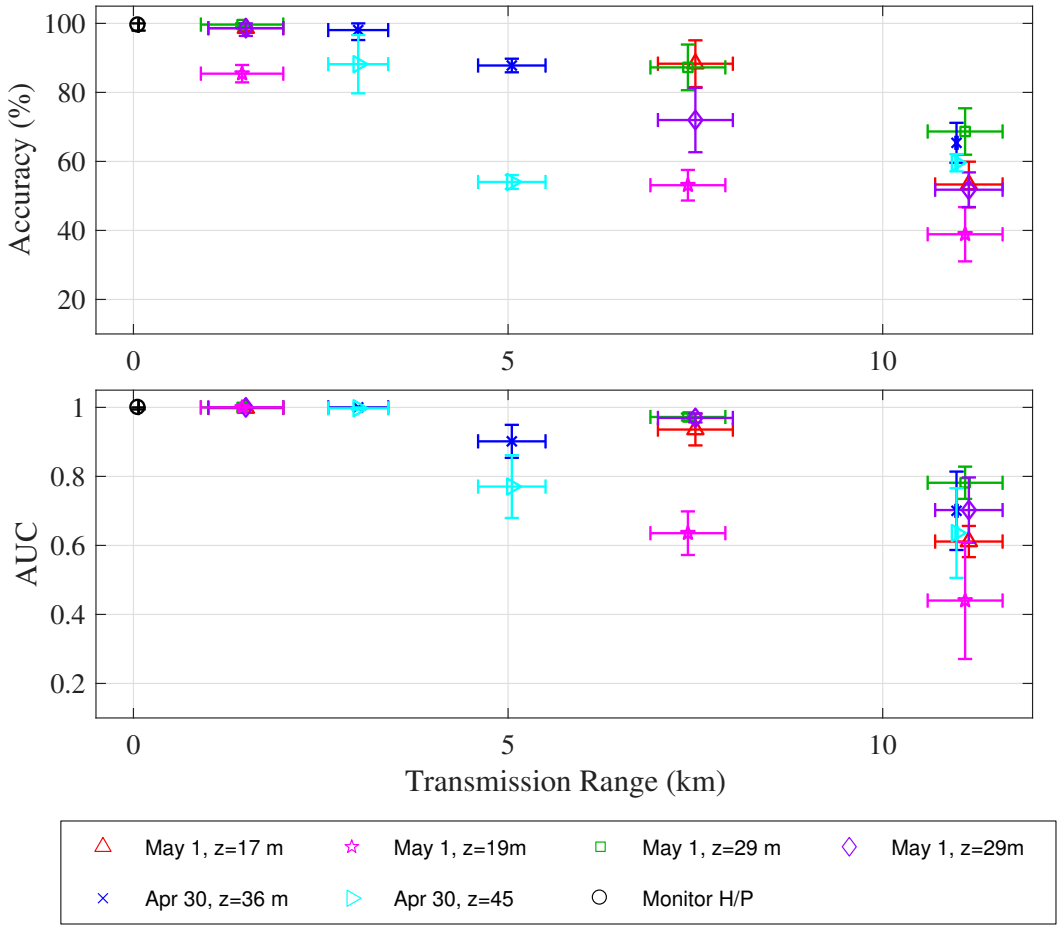


Figure 4.11: Experimental classification performance as a function of range for the synthetic whale calls. The vertical error bars are one standard deviation of the classification performance based on 5-fold cross-validation, and the horizontal error bars define the shortest and longest transmission range for each transmitted signal set.

synthetic calls prior to transmission, relative to the biogenic vocalizations.

The results presented in Figures 4.10 and 4.11 indicate that there was little difference in performance as a function of recorder depth, as there was no obvious trend associated with depth across the complete transmission range. There was, however, greater spread in the performance values at the 10 km range than at closer ranges, both in terms of data from the different recorders, and increased variance in the cross-validation results for data from a single recorder. Based on the performance results (of the synthetic signals in particular) at the longest range, it is tempting to suggest that performance decreases with decreasing recorder depth; however, there are insufficient data points to make such a general conclusion. If this were a real trend, then it is likely an environment-specific one.

One can imagine an environment, for example, that would produce maximum classifier performance mid-watercolumn and decreased performance at greater depths due to the presence of a sound-speed minimum. Thus, the remainder of this discussion is focused on the variation of performance results at the longest range for which there were signals detected (i.e., the 10 km range). There are three factors which may contribute to this increased variation, relative to that of the performance at the closer ranges. First, as signals were transmitted over longer ranges, they were subjected to more small scale oceanographic fluctuations. At the longer ranges these fluctuations may compound so that there was greater variance between each instance of the ocean environment that each signal experienced, leading to an increased variance in classifier performance. The second factor was that there were fewer calls detected at the 10 km range. In some cases there were as few as two bowhead and four humpback synthetic calls included in the validation set. Bear in mind that the training dataset was the same size for all cases. One would expect the variance in performance to increase with a smaller sample size. Finally, the third factor contributing to performance differences between recorders was related to signal-to-noise ratio. Since several different types of recorder were used, each with a different noise floor, there may be a difference in performance results that was recorder-dependent. To clarify, this dependence is not due to differences in propagation characteristics as a function of depth, but rather to which recording unit the signal was recorded on. For example, signals recorded on the Whalesong unit typically had lower SNR than those on the icListen at a similar depth. In short, although signals were transmitted over similar ranges, the SNR of the recorded signal could be different.

As an initial effort to disentangle the effects of transmission range and SNR the performance was plotted with SNR as the independent variable in Figures 4.12 and 4.13 for the biogenic and synthetic calls, respectively. As with the previous plots of performance results, the vertical error bars represent one standard deviation from the mean performance value, as determined from 5-fold cross-validation. From each transmission set, the mean and standard deviation of the SNR was calculated — the horizontal location of the data points on the plots corresponds to the mean SNR and the horizontal error bars represent one standard deviation of the SNR values. In this way there is an equivalence between each data point in Figures 4.10 and 4.11, and Figures 4.12 and 4.13.

A striking feature of these plots is the wide SNR range covered by the uncertainty

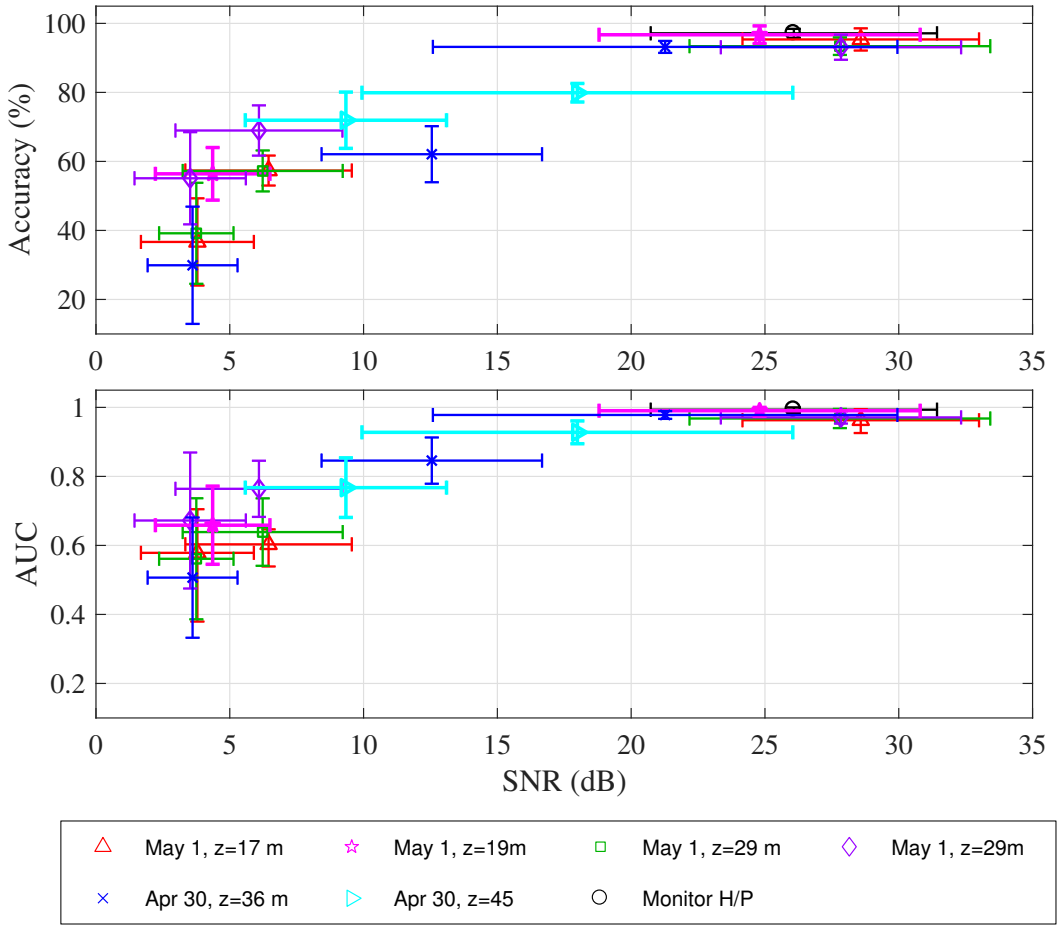


Figure 4.12: Experimental classification performance as a function of SNR for the biogenic whale calls. The vertical error bars are one standard deviation of the classification performance based on 5-fold cross-validation, and the horizontal error bars define one standard deviation of the SNR of the corresponding recorded signal set.

estimates for some of the data points, specifically for the biogenic signals recorded by the WhaleSong and icListen on the first day of the experiments with mean SNR ≥ 8 dB (the blue 'X's and cyan triangles in Figure 4.12). As mentioned in Sections 4.2 and 4.4.3, the signal SL varied throughout the course of the first day of the experiment as the signal generation scheme was optimized. This resulted in a wide range of SNRs for the biogenic calls. This was not as much of an issue with the synthetic calls on the first day of the experiment because they were transmitted after all the biogenic calls, such that the configurations affecting the SL were already locked in by the time the synthetic signals were transmitted.

Considering the extreme ends of the classification performance range, one expects that

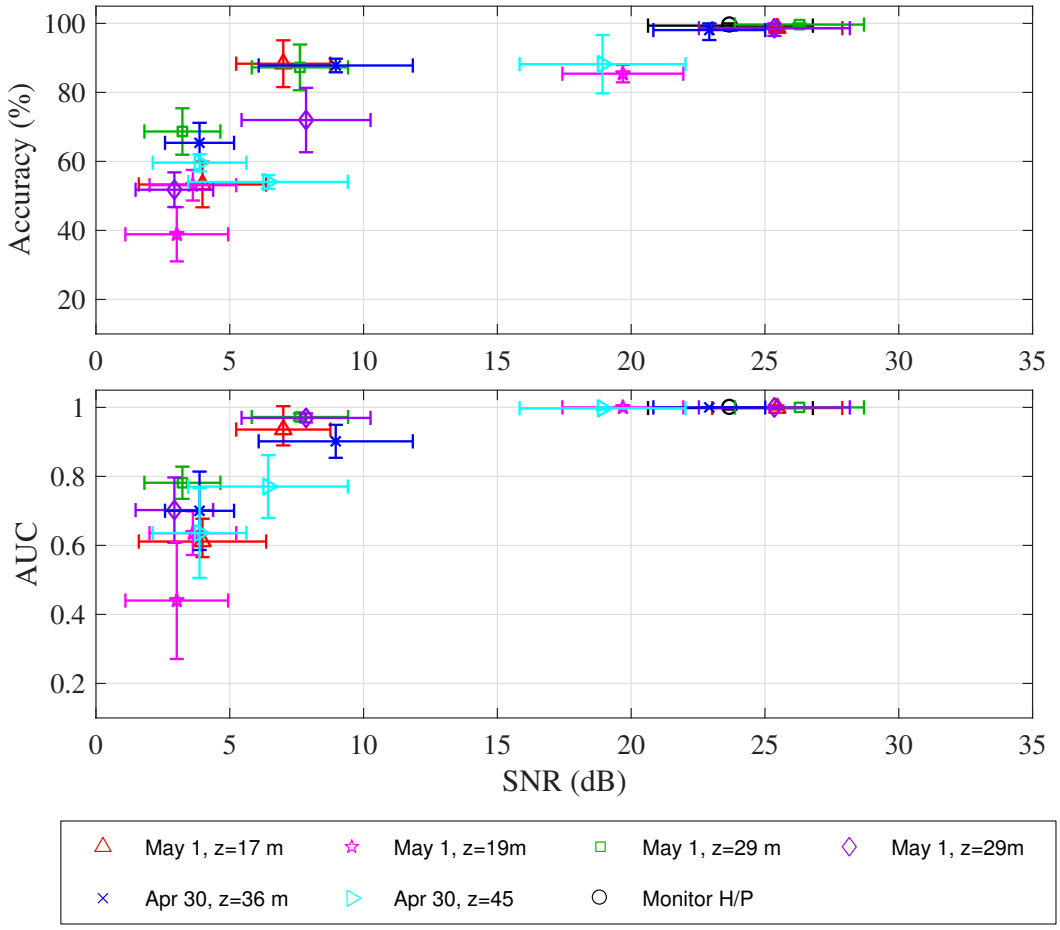


Figure 4.13: Experimental classification performance as a function of SNR for the synthetic whale calls. The vertical error bars are one standard deviation of the classification performance based on 5-fold cross-validation, and the horizontal error bars define one standard deviation of the SNR of the corresponding recorded signal set.

increasing SNR beyond a certain point would not provide any additional information, so that classification should reach a maximum value once a certain SNR threshold is attained. At the other extreme, when SNR is low enough that the signal is no longer distinguishable from the noise, decreasing SNR will not cause the performance to decrease further because it has already achieved the point at which a random classification decision is assigned. Thus, at high SNR the performance should tend to the maximum performance limit, and at low SNR the performance will approach the limit where AUC $= 0.5$ and accuracy $\leq 50\%$. One would expect the performance to monotonically increase as a function of SNR between these two extremes. Figures 4.12 and 4.13 show that the classifier performance determined from the experimental data behaved in exactly the anticipated manner — performance

generally increased with increasing SNR, and reached the upper performance limit at high SNR and the lower performance limit at low SNR values. Murphy and Hines [79] posit a linear relationship between AUC and SNR in the mid-SNR/performance range, based on Shannon's [80] theory that the information capacity of any transmission channel (in this case the ocean) is proportional to SNR (in units of dB). Bear in mind, however, that a linear relationship between AUC and SNR will only occur when SNR effects dominate in an environment; one would anticipate that if propagation-induced distortion has a significant impact on the classifier performance a linear trend will not adequately describe the relationship between SNR and AUC. Unfortunately, there were too few data points to quantify the relationship between SNR and performance in the mid-SNR/performance range, so it is not possible with the experimental data to verify the accuracy of Murphy and Hines' hypothesis. This topic will be addressed in further detail in Chapter 6 with the aid of propagation modelling to increase the SNR-resolution of the data set.

From the plots of performance as a function of range or SNR it was clear that the aural classifier performance was range-dependent. It was also apparent, both from the plots and theory, that SNR was an important contributor to this relationship. However, the question remains: what are the relative contributions to degraded aural classifier performance that arise from decreasing SNR and from propagation-induced signal distortion due to multipath? The data collected during the experiments was insufficient to answer this question; unfortunately, performing additional experiments with sufficient resolution is costly, so we must rely on propagation modelling to augment the experimental data. The results of propagation modelling may also guide what data should be collected in future experiments, and at what ranges to transmit signals, in order to resolve the SNR-performance relationship. Disentangling the relative importance of SNR and other propagation effects is a sufficiently broad subject that it will be addressed in a subsequent chapter (Chapter 6).

4.5.3 Training Set Selection

By training the classifier on data recorded by the shipboard monitor hydrophone it has been established that classifier performance falls off with increasing transmission range. In this section, we consider if high SNR signals recorded close to the sound source are the best for training the aural classifier. Murphy and Hines [79] concluded in their research with active sonar echoes that given two or more datasets with different SNR, the aural

classifier performed better across multiple SNR regimes when the classifier was trained on a low SNR dataset. They ascribed this finding to the perceptual feature subset selected based on the signals contained in the training set — features selected from a high SNR training set did not perform well at low SNRs because the signal characteristics these features depended on became lost in the noise as SNR decreased, whereas features that were selected from the low SNR signals relied on signal characteristics that were still present in the higher SNR signals. Based on these conclusions, the selection of the training set is re-evaluated here to determine if a more robust training strategy could be employed.

To investigate the effect of the transmission range of signals included in the training set, classifier performance matrices (Figures 4.14 and 4.15) were constructed showing average performance. These are similar to the graphics produced by Murphy and Hines [79] to analyze the SNR-dependence of aural classification for active sonar. Performance matrices were constructed by training the classifier on data transmitted over range x and then validating the classifier on data transmitted over range y. The value of each element in the performance matrices represents the average AUC or accuracy determined from performing 5-fold cross-validation on the detected signals, using the twenty features that best discriminated between the bowhead and humpback vocalizations. For example, the square in the fifth column and third row of Figure 4.14b represents a mean $\mathrm{AUC} = 0.80$ determined by training the aural classifier on signals transmitted over 10 km and then validating the classifier on data transmitted over 1 km. These performance matrices provide a comprehensive picture for evaluating the robustness of the aural classifier to propagation effects as a function of range while also addressing the issue that it is unlikely in a real PAM scenario to always have high SNR/close range transmissions with which to train the classifier.

In general, the performance matrices generated using data from each recording unit displayed similar trends — the example results in Figure 4.14 were generated from data measured by the Sharp recorder on the second day of the experiment and are representative of the majority of the results. Only data from one recorder (the icListen unit on mooring 4) exhibited a different trend, and are presented in Figure 4.15 for comparison. Here, the 0 km range refers to the original signals (either biogenic or synthetic) after up-sampling, but not subjected to the signal conditioning designed to flatten the TVR curve of the amplifier/transducer, other transformations not accounted for by the shipboard electronics,

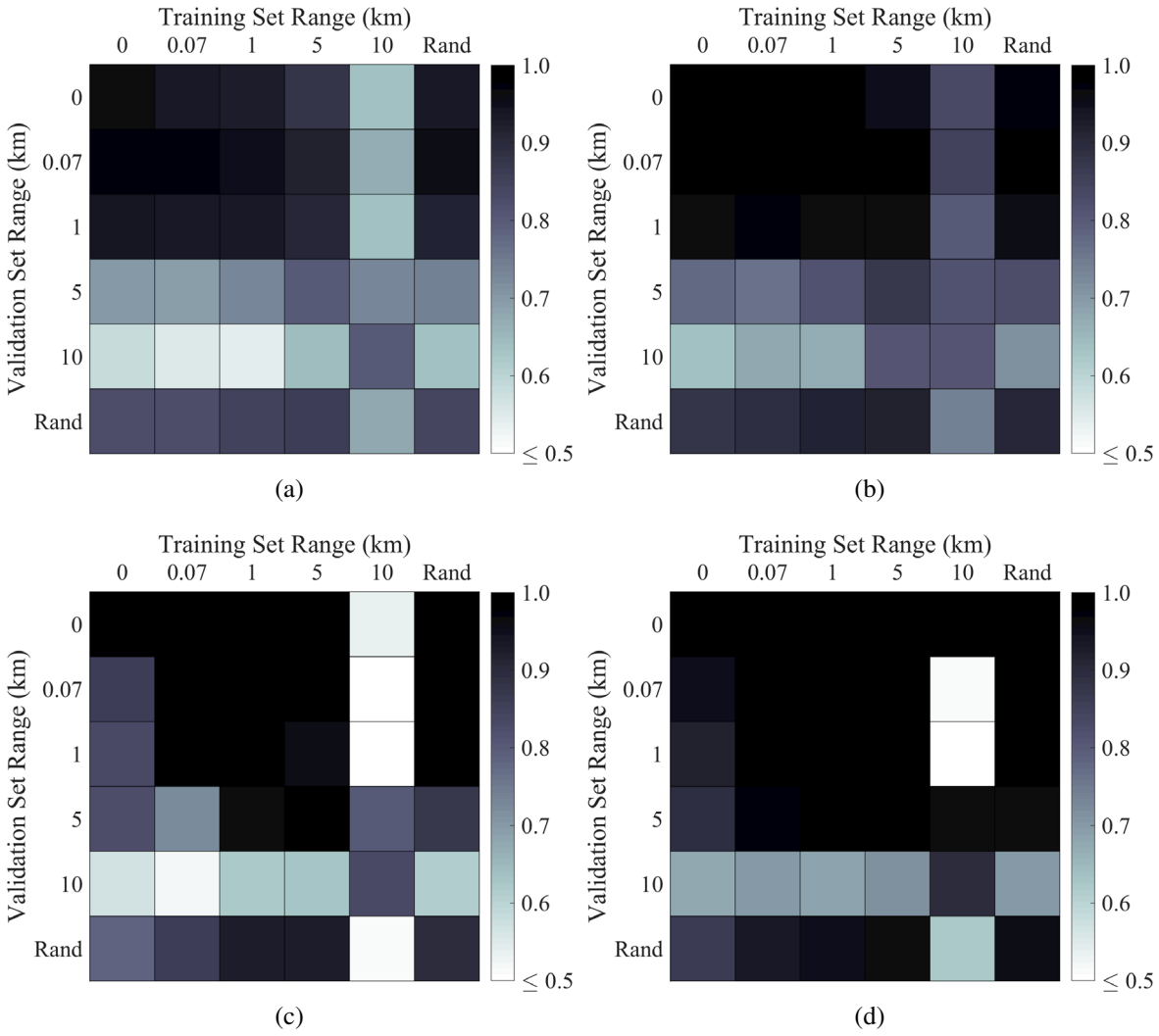


Figure 4.14: Confusion matrices displaying results of how transmission range of training and validation signals impacted the classifier performance. These results were derived from signals recorded by the Sharp unit on day 2, and are typical of five of the six recorders. Classifier performance is given in terms of the (a) accuracy and (b) AUC for the biogenic whale calls, and the (c) accuracy and (d) AUC for the synthetic calls.

nor transmitted into the water. The 70 m range refers to signals which were recorded on the monitor hydrophone. The last column/row, labelled "Rand.", indicates that the corresponding subset of signals was composed of a random selection of transmitted signals from all ranges between 0.07 to 10 km.

First consider the results presented in Figure 4.14, which were representative of the majority of results. These plots confirm the range-dependent component of the classifier performance that was previously noted. The standard deviation associated with these

performance results covered the ranges $\sigma_{accuracy} = 1$ to 10 % and $\sigma_{\mathrm{AUC}} = 0.008$ to 0.2 for the biogenic signals, and $\sigma_{accuracy} = 0$ to 20 % and $\sigma_{\mathrm{AUC}} = 0$ to 0.4 for the synthetic signals. For both the biogenic and synthetic signals the errors were typically $\sigma_{accuracy} \sim \mathcal{O}(1\%)$ and $\sigma_{\mathrm{AUC}} \sim \mathcal{O}(0.01)$, with only a few of the cases with low performance values at the higher end of the ranges listed previously. Generally, the elements along and above the main diagonal (top left corner to bottom right corner) contain better performance than those

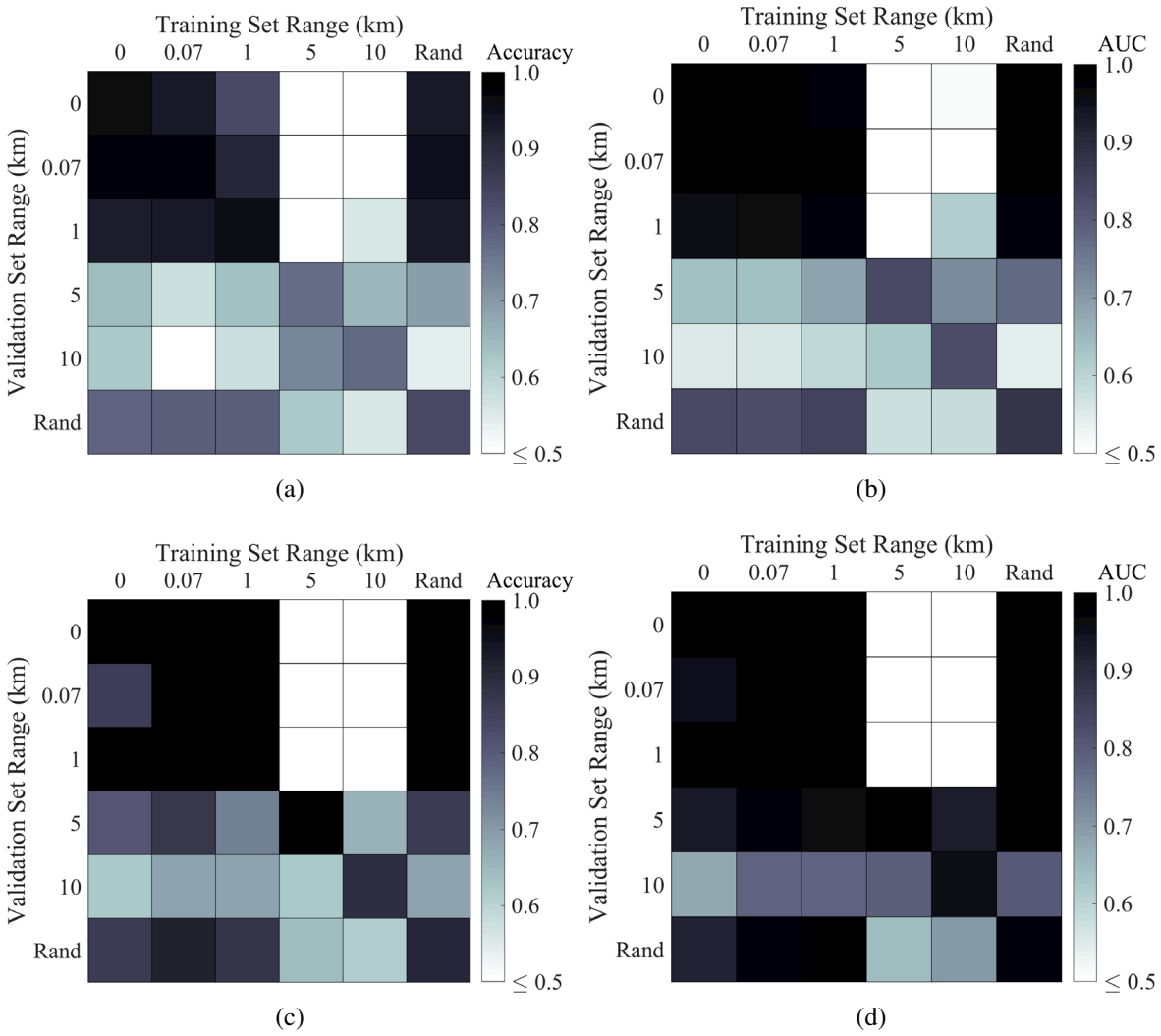


Figure 4.15: Confusion matrices displaying results of how transmission range of training and validation signals impacted the classifier performance. These results were derived from signals recorded by the icListen system on mooring 4 during the second day of the experiment, and are atypical of results from the other five recorders used during the experiment. Classifier performance is given in terms of the (a) accuracy and (b) AUC for the biogenic whale calls, and the (c) accuracy and (d) AUC for the synthetic calls.

below the main diagonal, excluding the row associated with the random-range validation subset. A similar pattern was noted in the SNR-dependence investigation of Murphy and Hines [79], further suggesting that SNR, at least in part, was responsible for driving the observed decrease in classifier performance. Mouy *et al.* [23] also noted that false negative rates increased as SNR decreased. As previously noted, these experimental data were too limited to determine fully the relative importance of complex propagation effects versus decreasing SNR as signals propagated over longer ranges. This issue will be addressed fully in Chapter 6. Also of note was that the classifier generally did not perform as well when trained on the original data. This was especially clear for the synthetic data set. This may be because the signals were somewhat distorted by the transmission process, or merely may be a result of no propagation effects being included in the training data such that the classifier was less able to cope with even the minimal time and frequency spreading applied to signals recorded by the monitor hydrophone.

The performance matrices generated from the data recorded by the icListen on mooring four (Figure 4.15) are included here because they do not agree with Murphy and Hines' [79] conclusion that if a discrepancy in SNRs between the training and validation sets is expected, it is best to train with lower SNR data. Instead, for the 5 and 10 km ranges, the opposite was true — better performance for signals transmitted over longer ranges was obtained by training the classifier on signals which propagated over shorter ranges, or not at all. The cause of this was not clear, especially in light of the evidence that data from all other recorders produced the expected results.

Of particular interest, were the performance results when the classifier was trained with data from a random selection of ranges. For all of the performance matrices shown, the random mixture of signals included in the training set produced the most robust classifier when validated on data across all ranges. One can make sense of this result intuitively, by considering that the classifier was able to select features that distinguished between the bowhead and humpback vocalizations that resulted in good generalization of the classifier. In contrast, when the classifier was trained on data from a single range it was able to learn the signal characteristics well for all data transmitted over a similar range; however, it was not flexible enough to accommodate slight differences in signals transmitted over longer ranges. This can be thought of as *overfitting* the classifier to a specific dataset [41, 81].

In summary, for best classifier performance across a wide variety of transmission

ranges, one ought to train the classifier using signals transmitted across any equally large assortment of ranges. If this is not possible — for instance, when only a limited amount of data is available for training purposes — then it is best to select low SNR calls to be included in the training set, bearing in mind that there must still be sufficient SNR to identify the discriminating features.

4.6 Chapter Summary

In the preceding sections, details of an experiment designed to consider the effects of propagation on the aural classifier using biogenic and synthetic whale calls were presented. The ocean acoustic environment was described using surface conditions derived from wind speed, sound speed profiles obtained at each location which transmissions were made, and sediment properties were inferred from FFCPT measurements made along the ship track on the second day of the experiment. Aural classification results were presented for both days of the experiment. In general, it was shown that classifier performance was range-dependent, such that the performance decreased with increasing transmission range. Generating performance matrices showed that to obtain the best classification performance across a wide variety of transmission ranges it is best to train the classifier on a training set including an equally varied assortment of ranges; however, if there are restrictions on the training data set, it is best to train on relatively low SNR calls.

Examination of the aural classifier results left an outstanding question. When the classifier performance was examined with respect to SNR, it was noted that the classifier performance decreased with decreasing SNR. Together with the range-dependent classifier performance, this evidence suggested that SNR is an important contributor to the degradation of classifier performance. This makes sense intuitively — as the signal-to-noise ratio decreases there is less information unique to the signals of interest available to inform the classifier's decision. In spite of this clear range-dependence, it was not possible to determine definitively, if the classifier performance is also impacted by signal distortion due to acoustic propagation. At this point, it has not been possible to address the relative impacts of SNR and signal distortion on the classifier performance. Propagation modelling is required to adequately address this issue. Nonetheless, the experimental data were shown to be invaluable for analyzing how classifier performance is altered as signals are propagated over increasingly long ranges in a real ocean acoustic environment. These

results provide a baseline for the propagation modelling in subsequent chapters.

CHAPTER 5

PULSE PROPAGATION MODELLING

Propagation modelling is an important component of this thesis that both complements and augments results obtained from the propagation experiments. A pulse propagation model was used to transmit signals through a virtual environment to synthesize how the environment distorts a signal as it propagates through the water. Using a model has several important advantages over sea trials. Modelling is less time-consuming and expensive than performing sea trials, which makes it possible to model different environments, and multiple scenarios for the same environment. The ability to finely control the modelled environment is useful for examining how changes to an environmental variable impact the total signal distortion and the classifier's performance. Additionally, it is possible to increase the range resolution and extend results to longer ranges to provide a better indication of how classifier performance changes with respect to propagation range. By applying propagation models, one is also able to gain a better understanding of experimentally collected data.

5.1 Background

A variety of numerical techniques have been developed to study acoustic propagation in the ocean. These include ray tracing, normal mode, parabolic equation, wavenumber integration, and finite element models [11, 82, 83]. While each has its advantages, no single method is suitable for handling all possible environmental conditions, frequencies, and transmission ranges that are dictated by real-world applications [17]. Thus, one must carefully consider which type of model is most appropriate based on the parameters of a particular research problem. There were *essential* considerations, as well as some *desired*

features, taken into account when selecting the model to be used in this research. The essential features of the propagation model were:

- It must produce an accurate representation of bottom interactions since this research predominantly examines the shallow-water acoustic environment in which PAM of cetaceans typically occurs [1].
- It must be capable of handling a large frequency range; this includes the low frequency range of real bowhead and humpback vocalizations (extending down to approximately 50 Hz) as well as the higher frequencies used for the propagation experiments (1 to 4 kHz).
- It must be accurate over the spatial scales for which baleen whale calls are typically detected (tens of kilometres [1]).
- It must also be capable of performing pulse propagation that can accurately simulate a time-domain representation of a signal after it has been propagated through a modelled environment.

In addition to these required features, it was also desirable for the model to be:

- range-dependent, because shallow-water environments often exhibit range-dependence,
- 'public-domain' or non-proprietary to facilitate possible modifications to the source code and ease of access, and
- reasonably straight-forward to implement.

A model based on ray tracing theory was selected for pulse propagation, that was first validated using a wavenumber integration model (Section 5.1.2.1) because the latter has a larger user community [32, 83, 84, 85, 86, 87, 88]. The following provides a brief overview of some of the benefits and drawbacks of each numerical technique, to illustrate the motivation for choosing this model type. No explicitly time-domain models are considered in the summary since Jensen *et al.* [11] state, "Our experience is that most problems of practical importance in underwater acoustics (long-range propagation of pulses of 'finite' bandwidth) clearly favour the Fourier synthesis technique." The key points of the following discussion are summarized in Table 5.1, at the end of this section.

Ray-tracing models are often an attractive choice because they are computationally efficient, provide a simple pictorial representation, and allow for range-dependent environments. Conversely, a ray description of the acoustic field is not necessarily the most appropriate representation in shallow water environments due to multiple bottom interactions. These multiple boundary interactions provide challenges with respect to accurate representation of the signal phase. It also does not properly handle diffraction and caustics [17]. Additionally, the ray solution to the wave equation is a high-frequency approximation, where a guideline for 'high-frequency' is provided by Etter [82] as,

$$f > \frac{10c}{h} \quad , \tag{5.1}$$

where f is frequency, c is the sound speed, and h is the water/duct depth. For example, using a representative sound speed of 1500 m/s and water depth of 100 m, ray-theory becomes applicable for frequencies above 150 Hz. This simplified calculation shows that, while ray theory may be useful for understanding the propagation experiments in which signals were transmitted in the 1 to 4 kHz band, its applicability may be limited for understanding propagation of real bowhead and humpback vocalizations which have considerable energy below 150 Hz, as discussed in Section 3.3. Nevertheless, Xian [89] used a ray-tracing model (Bellhop) to simulate the effect of the propagation environment on the performance of a Short Time Fourier Transform detector for detecting synthetic North Atlantic right whale calls. For comparison purposes, it is important to note that right whales vocalize in a similar frequency band as the bowhead and humpback whale calls selected for this research. Porter and Liu [90] point out that typical problems suited for ray tracing are those with a high-frequency broadband source in range-dependent environments. Thus, ray models are usually used for active sonar modelling and ocean acoustic tomography. Of particular interest here, is that ray-tracing methods are known to handle broadband problems more efficiently than full wave methods, since many parts of the solution are frequency-independent and so each new frequency does not require a full model run [90].

Normal mode models have been used in underwater acoustics as far back as the 1940s [91] and have undergone significant development since then [11], so that specific models have an established pedigree [83]. Some advantages of this model type are that waveguide dispersion effects are inherently incorporated [82], mode functions do

not have to be calculated for intermediate ranges between source and receiver [17] and can be easily determined for all potential receiver locations (range from the source and depth in the water column), given the frequency and depth of the sourcem which is in contrast with ray models which must be executed for each change in source or receiver depth [82]. Normal mode methods can also handle many phenomena relevant to shallow water propagation (e.g., reverberation [92]). A potential disadvantage is the degree of knowledge about the sediment structure that is required to produce an accurate representation [82]. Normal mode models are primarily suitable for range-independent environments; however, it is possible to extend this method to encompass range-dependent environments, typically at the cost of computational efficiency [11, 17]. Normal mode approaches are best suited to low-frequency (less than 500 Hz) applications because the number of modes increases with frequency, but many numerical implementations treat frequencies up to a few kilohertz [82]. Mercado and Frazer suggest that normal mode theory is well-suited to propagation modelling in the acoustic environments inhabited by humpback whales because bottom properties are only "moderately variable" and the propagation ranges of interest are suitable for a normal mode model.

Parabolic equation (PE) models are the most popular technique for range-dependent ocean acoustic propagation problems [11, 17]. The greatest advantage of PE models is that they are computationally efficient due to an approximating assumption that transforms the elliptic Helmholtz equation — which is solved in the other numerical methods discussed here — to the parabolic equation. Computational advantage is gained because solutions for the entire water column are obtained by marching the solution forward in range, rather than solving for all range-depth pairs [17, 82]. The paraxial approximation that gives rise to the parabolic equation is also the greatest disadvantage of PE models — because of this assumption, solutions are approximate and not as precise as other methods [17], leading to phase errors as a function of angle [11]. PE models are neither applicable nor practical at high frequencies in shallow water [17, 82]. Helble *et al.* [18] used a PE model to propagate humpback whale calls to estimate site-specific probability of detection functions with respect to range, and azimuth. Širović *et al.* [4] also used the PE numerical technique to normalize the number of blue and fin whale call detections by the modelled detection ranges.

Wavenumber integration techniques (sometimes referred to as Fast Field Programs, or

FFP) were first applied to the field of underwater acoustics in 1948 by Pekeris [91]. He considered acoustic propagation in horizontally stratified waveguides using environmental models with only two and three fluid layers [11, 91] (i.e., what is commonly referred to as the 'Pekeris waveguide'). This technique was later applied to 'few'-layer waveguides in the 1950s; however, it was not until the 1980s when efficient and numerically stable implementations of wavenumber integration were developed [93, 94, 95, 96]. The wavenumber integration approach is closely related to normal mode models [11, 17]; however, unlike the normal mode approach, the Green's function solutions provided by the wavenumber integration technique are 'essentially exact' in the far-field and give the full wave solution for the acoustic field [11, 17, 93, 94]. That is, the wavenumber integration approach directly solves the Green's function, with deviations from the exact solution arising only from necessary discretization. This is in contrast to the normal mode technique, which contains inherent approximations. Some of the disadvantages of this method include that solutions are generally restricted to horizontally stratified environments (i.e., range-independent) [11, 17, 93, 94], and although this method "is not too demanding computationally [17]," efficiency can be a concern for pulse propagation [83], in particular for the frequency band in which signals were transmitted during the experiments.

Table 5.1 summarizes the key points of each model type. As a general rule, the division between low and high frequencies is set at 500 Hz, following Etter [82]; he selected this threshold somewhat arbitrarily but it does reflect that above 500 Hz some techniques become computationally intensive, and below it the physics of some ray-tracing methods may be 'questionable,' given the inherent high-frequency approximation of these methods. Given the frequency band and water depths of the Gulf of Mexico experiment, a more appropriate threshold between low and high frequencies, specific to this experiment, is 200 Hz. This threshold was determined using Equation 5.1 with $c = 1530$ m/s and $h = 75$ m. The shallow water regime includes all water depths for which the sound significantly interacts with the bottom boundary — the Gulf of Mexico experiment was conducted in a shallow water environment, and most examples of baleen whale PAM on the continental shelf may be considered to occur in shallow water environments. The column entitled 'Pulse propagation' considers numerical techniques that are known to have already implemented a pulse propagation capability. The column labelled 'non-proprietary' indicates whether there is an available model implementation that is well-established, as

Table 5.1: Summary of the main considerations when selecting a numerical technique for ocean acoustic pulse propagation modelling. Any signal with frequency greater than 500 Hz is considered high frequency for the purposes of this analysis. The definition of shallow water is decided on a case-by-case basis through comparison of the water depth and signal frequency, as in Equation 5.1.

Numerical technique	Frequency range	Suitable for shallow water	Suitable spatial scale	Range-dependent	Pulse propagation	Non-proprietary	Example
Ray-tracing	high	yes*	yes	fully	yes	yes	BELLHOP
Normal mode	low to few kHz	yes	yes	with approximations†	yes	yes	KRAKEN
PE	low	yes	yes	fully	yes	yes	RAM
Wavenumber integration	low to few kHz	yes	yes	with approximations†	yes	yes	OASES

* Given a sufficiently high frequency signal; use Equation. 5.1 as a guide.

† Range-dependence can be implemented using a suitable approximation — either coupled-modes or adiabatic approximation.

well as freely-available. Only a single example implementation is given in this table, although many other implementations exist. The examples listed here are well-known in the ocean acoustics research community, and are freely available on the Ocean Acoustics Library website [88]. From this table it is evident that each numerical technique has many benefits — the primary motivators for selecting a ray-theoretic model were the frequency band of the experiments, the availability of a model (and access to its source code) with pulse propagation implemented, and its implicit range-dependence.

5.1.1 Broadband Modelling with Fourier Synthesis

The pulse propagation models that will be discussed in the following sections generate the expected pressure time series at a receiver for each source signal. They accomplish this with a Fourier approach to broadband modelling, that is, they solve the pulse propagation problem in the frequency-domain by Fourier synthesis of single-frequency (continuous wave, or CW) results [11]. A brief description of the frequency-domain approach to pulse propagation is discussed here, recognizing that it is a general approach which can be implemented using any of the numerical techniques listed in Table 5.1.

The most common approach to broadband modelling, also referred to as pulse propagation, in the ocean acoustics community is to make use of the frequency domain representation. In fact, Jensen *et al.* [11] state that, "Our experience is that most problems of practical importance in underwater acoustics (long-range propagation of pulses of 'finite' bandwidth) clearly favour the Fourier synthesis technique." In the Fourier synthesis approach a propagation model is executed multiple times at discrete frequencies over the frequency band of interest [11, 82, 97]. The resulting time-domain signal is then reconstructed, using Fourier techniques, as

$$p\left(R,z,t\right) = \frac{1}{2\pi}\int_{-\infty}^{\infty} S\left(\omega\right) p\left(R,z,\omega\right) e^{-i\omega t}\,\mathrm{d}\omega \quad , \tag{5.2}$$

where $S(\omega)$ is the source spectrum, and $p(R,z,\omega)$ is the spatial transfer function [11]. It is this spatial transfer function which is generated through repeated executions of a CW propagation model for a discrete number of frequencies, within the frequency band of interest [97]. The evaluation of the integral (Equation 5.2) is then accomplished by application of the Fast Fourier Transform (FFT) at each spatial position (R,z) for which the pulse response is desired [11, 97]; in practice one must truncate the integration interval

such that Equation 5.2 becomes,

$$p\left(R,z,t\right)=\frac{1}{2\pi}\int_{-\omega_{max}}^{\omega_{max}}S\left(\omega\right)p\left(R,z,\omega\right)e^{-i\omega t}\,\mathrm{d}\omega \quad , \tag{5.3}$$

where it is assumed that the source signal emits negligible energy for frequencies greater than ω_{max} [11]. In this way, the frequency domain wave equation, which is valid for a single frequency CW signal, may be extended to treat broadband signals through use of Fourier synthesis of the individual CW solutions [82]. In this description of the Fourier synthesis approach the continuous version of the Fourier transform is employed; in practice Equations 5.2 and 5.3 must be discretized and evaluated via the Discrete Fourier Transform.

This approach to pulse propagation modelling is attractive because of its simplicity and versatility. Any of the single-frequency numerical techniques listed in Table 5.1 may be employed to generate the spatial transfer functions. This means that it is possible to select the most appropriate propagation model based on the parameters of a given problem (e.g., frequency, water depth, spatial scale, etc.) to model the propagation physics, then the pulse result can be generated with little additional effort through application of Fourier synthesis.

Before moving on to discuss the details of the pulse propagation model employed in this research, it is worth making a few comments on how to implement this Fourier synthesis approach for practical problems. If a single source signal is considered, then the task of pulse propagation is straightforward, and accomplished exactly as described above: a propagation model is run at a series of different frequencies corresponding to the Fourier components of the initial waveform, then the resulting waveform is reconstructed through summation of the propagated output with the appropriate weightings. However, when multiple signals must be artificially transmitted through a simulated environment this is not the best approach. Instead, the most efficient means of broadband modelling for a large number of signals is to simulate the propagation of a bandlimited impulse function, which can then be convolved with each of the signals of interest. For example, to generate the bandlimited impulse response of the ocean channel, a sinc function,

$$s(t)=\frac{\sin\left(2\pi f_{max}t\right)}{2\pi f_{max}t} \quad , \tag{5.4}$$

with maximum frequency, f_{max}, can be supplied as the input waveform (the time- and frequency-domain representations of this function are given in Figure 5.1). After the sinc

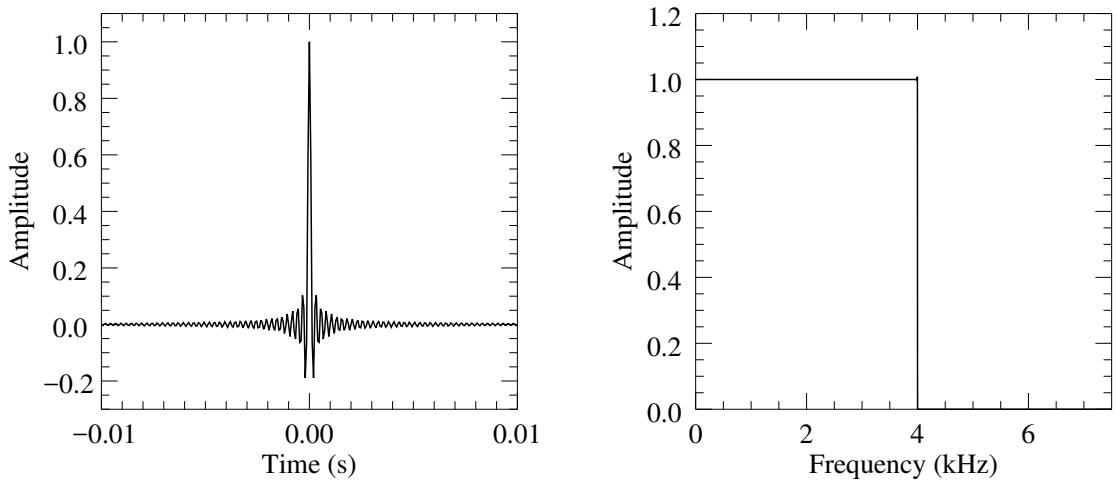


Figure 5.1: Time and frequency response of the sinc function input to WATTCH to generate an environment's channel impulse response. Note that only a narrow time interval is plotted for the temporal representation of the sinc function to demonstrate its structure.

function has undergone the attenuation and distortion predicted by the propagation model, it can be convolved with any number of other signals. In this way, the more computationally intensive pulse propagation modelling needs to be performed only a single time for a large number of input signals [98]. This can be thought of as constructing an environmental transfer function using the broadband model, and then convolving it with the waveforms of interest [99].

5.1.2 Waveform Transmission Through a Channel (WATTCH) Propagation Model and Bellhop

The WATTCH model is a frequency- and phase-dependent (i.e., coherent) pulse propagation model [98, 99] that is essentially an implementation of the general broadband Fourier synthesis methods described above. As such, it simulates the effect of transmitting a waveform through an underwater acoustic environment [98, 99]. In essence, it uses acoustic propagation information supplied by any tonal propagation model to transform an input acoustic waveform; it then outputs the signal expected at a set of receivers, dependent on the range and depth of the receivers [98]. In its current configuration, WATTCH is supplied with the output of a ray-tracing model. For example, the propagation data from ray-tracing models may include source and arrival angles, travel time, cumulative phase changes due to boundary interactions, and amplitude information indicating the frequency-dependent attenuation (Note that the launch and arrival angles are not necessary

for WATTCH to reconstruct the signal because they are not required by Equation 5.3.) [99]. This propagation data may be computed by any ray-tracing model. Here, the relevant data are generated by the Gaussian beam model, Bellhop [100], which is supplied with a model of the acoustic environment, simulation geometry, and the relevant frequency band. As described for the general case in the previous section, each of the frequency components of the input waveform propagate through the environment, then the propagated components are reconstructed to produce the waveform predicted at the receivers [98]. Specifically, given an input signal and the propagation data supplied by Bellhop, WATTCH performs an FFT on the input waveform then, using frequency-interpolated propagation data for FFT magnitudes and phases, integrates the data to calculate the acoustic magnitude and phase spectra at the receiver [99]. By integrating only the phase and amplitude data it is implicitly assumed that arrival times are frequency-independent, further assuming that the medium is non-dispersive.

WATTCH linearly interpolates the magnitude and phase data across frequencies to increase computational efficiency by decreasing the number of times the single-frequency propagation model must be run. In doing this, it is assumed that the magnitude and phase behave linearly between the frequencies for which the propagation model is executed; however, it is necessary to interpolate the data to obtain the required resolution in the time and frequency domains for reconstructing the received waveform. As outlined by Jensen *et al.* [11], if one seeks to determine the acoustic response at a point (R, z) in a time window of length T, the sampling parameters for time, Δt, and frequency, Δf, satisfy the relations,

$$\Delta t = \frac{1}{f_s} \quad \text{, and} \tag{5.5}$$

$$\Delta f = \frac{1}{N\Delta t} \tag{5.6}$$

$$= \frac{1}{T} \quad , \tag{5.7}$$

where f_s is the sampling frequency of the source/output waveform, and N is the number of samples included in the time window. Naturally, N and T are related by $T = N\Delta t$. It is also important to note that the choice of time and frequency sampling interval must satisfy the sampling relation,

$$\Delta t \Delta f = \frac{1}{N} \quad . \tag{5.8}$$

The necessary discretization in frequency to evaluate Equation 5.3 introduces periodicity of T in time. Thus, one must be careful in selecting the appropriate sampling parameters so as to avoid wrap-around (aliasing) of later arrivals to the beginning of the time window. In particular, the choice of T requires a tradeoff between computational efficiency and the elimination of any aliasing — it is desirable to choose T to be as short as possible to decrease computational effort, but one must be careful to include a large enough window to encompass the complete response at each receiver to eliminate any wrap-around. While it is important to select sample parameters to remove aliasing due to wrap-around, one must also take the Nyquist criterion into account. To ensure the entire spectrum of interest is included, f_s should be selected such that,

$$f_s = \frac{1}{\Delta t} > 2 f_{\text{max}} \quad , \tag{5.9}$$

where f_{max} is the maximum frequency of interest [11].

Because a ray-tracing model is used to supply the propagation data to WATTCH, simulations with WATTCH are subject to both the benefits and limits of this propagation model type (refer to Table 5.1). In general, ray-tracing methods are a tradeoff between lower accuracy (relative to a full wave model) and faster processing times for higher frequencies. Bear in mind, however, that the accuracy of ray-tracing solutions improves at higher frequencies [90]. Since many simulations were conducted over the course of this research, the benefits of fast run times greatly outweighed the lower accuracy of the results.

The specific choice of the Bellhop ray model was motivated by its availability, ease of use, and wide use in the ocean acoustics community. Additionally, Bellhop is set up to easily produce the information required for broadband pulse propagation — one need only select an 'arrivals' run type in the model configuration files [101]. The output 'arrivals' file contains amplitude, phase, and time delay data for each ray path [101, 102], which is easily input to the WATTCH convolver. One should be aware that a limitation of the Bellhop model is that it does not model frequency spreading; that is, it does not transfer energy between frequencies. Due to the fact that Bellhop is a widely-known propagation model, it is useful to be aware that the version of Bellhop used throughout this research [102] is a modified version of Porter's freely available Bellhop software [88, 101].

Some of the modifications to Bellhop include the ability to calculate bottom losses for a two-layer fluid bottom with volume attenuation, and Rayleigh reflection coefficients at the interfaces [103].

Overall, the WATTCH model, with Bellhop supplying the acoustic propagation predictions, is a powerful and efficient tool for simulating signals propagating through the ocean medium. The theoretical basis of WATTCH is rooted in the techniques of broadband modelling with Fourier synthesis. Its efficiency stems from the relative computational ease with which Bellhop is run.

5.1.2.1 Model Validation

Since WATTCH has a relatively small user-community, it was necessary to first verify its accuracy. This was accomplished using two methods, to be certain of WATTCH's ability to capture the physical phenomena important to ocean acoustic pulse propagation. A simple approach was first employed to ensure that WATTCH produces the correct levels of signal attenuation in a spherical spreading environment. Then a more complex environment was generated, through which a band-limited impulse was propagated. The results from WATTCH were then compared against results from a benchmark model. The remainder of this section presents full details and results from this validation exercise.

A simple diagnostic tool for verifying a pulse propagation model is to check that it is able to properly estimate transmission loss. To this end, an environment was constructed such that the expected signal attenuation would be close to that of spherical spreading, i.e., little or no interaction of the sound field with the boundaries. A 10 km deep isospeed environment was generated with $c = 1454.25$ m/s. With a source and receiver placed at mid-watercolumn ($z_s = z_r = 5$ km) a 1 to 4 kHz LFM pulse was transmitted over 0.5, 1, 2, 3 and 5 km. The power of each of the received signals was then calculated and used to determine the transmission loss, which could then be compared with the transmission loss expected for spherical spreading. The results of this comparison are depicted in Figure 5.2. The solid line shows the transmission loss for spherical spreading with the addition of Thorp attenuation (about 0.15 dB/km at the LFM centre frequency of 2.5 kHz) [75, 104, 105]. On the whole, the WATTCH model shows good agreement with the analytic results, although WATTCH starts to over-predict TL at longer ranges. This may be because additional loss mechanisms are included in WATTCH (e.g., bottom loss) that are not captured by the simple analytic model.

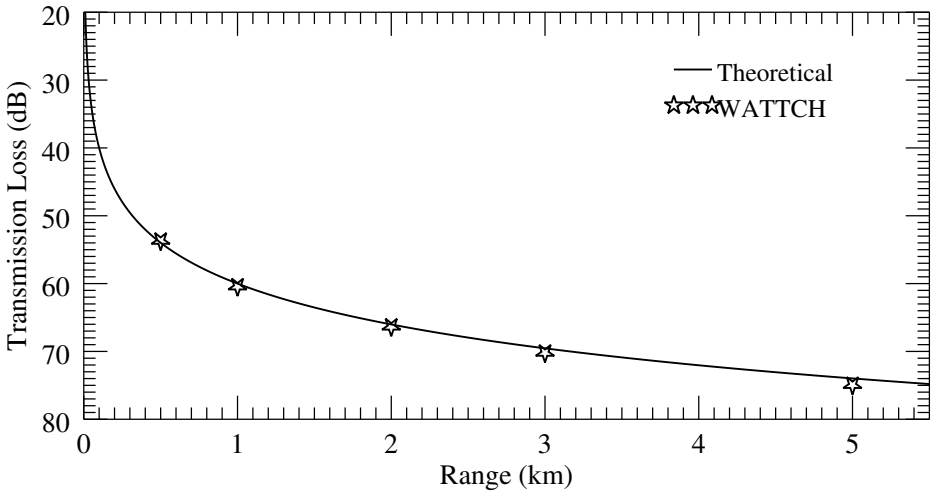


Figure 5.2: Comparison of transmission loss values determined from the WATTCH model and those predicted using spherical spreading with Thorp attenuation.

Although the first validation exercise showed that WATTCH correctly estimates transmission loss — a metric that captures much of the important propagation physics — it did not provide verification of other important features of a pulse propagation model, like accurately simulating timing and relative strength of arrivals. With this in mind, a second investigation was conducted to test this aspect of the WATTCH model. Channel impulse response results generated by WATTCH were compared to those generated by the fast-field model OASES. A more complete description of OASES follows.

The Ocean Acoustics and Seismic Exploration Synthesis (OASES) model uses the wavenumber integration method for modelling seismoacoustic propagation in horizontally stratified waveguides [94]. OASES can determine the acoustic field with multiple sources and receivers simultaneously [17, 94]. It is an up-to-date version of the Seismo-Acoustic Fast field Algorithm for Range-Independent environments (SAFARI) model [83, 94] developed by Schmidt for the SACLANT Undersea Research Centre [93] and has the reputation of being efficient, relative to other wavenumber integration models [17, 83]. Be that as it may, in general, it is considerably less efficient than models based on the other numerical techniques summarized in Table 5.1, since its run-time increases with both maximum transmission range and frequency of interest. The slower run-time was acceptable for producing benchmark results, but would not have been feasible for the main model runs presented in this thesis.

In terms of this research, OASES' strength is its pulse propagation module [94], referred

to as OASES-OASP, which can be used to propagate signals through a user-defined environment. OASES-OASP calculates the acoustic transfer function at each receiver by evaluating the wavenumber integral [94]. The model outputs the received pressure time series for each source signal. Like WATTCH, the output time series is representative of a signal recorded on a hydrophone located at depth z in the water column after the signal was transmitted range R through the ocean environment. OASES has previously been used for many applications, including as a pulse propagation model for baleen whale vocalizations. Chapman[32] used the OASES model in his efforts to localize North Atlantic right whales because he required a model that could properly treat the physical effects typical of a shallow-water waveguide. Most importantly, for a horizontally stratified ocean waveguide the wavenumber integration technique employed by OASES constitutes a benchmark solution against which other approximate solutions, like those generated by ray-based models, can be verified [97].

Similar to Pecknold *et al.* [99], a band-limited impulse waveform (with frequency content equally spread across the 1 to 4 kHz band), was propagated through a constant sound speed environment. The environment was characterized by an isovelocity sound speed profile with $c = 1454.25$ m/s, water depth of 120 m, bottom density of 1.9 g/cm^3 and

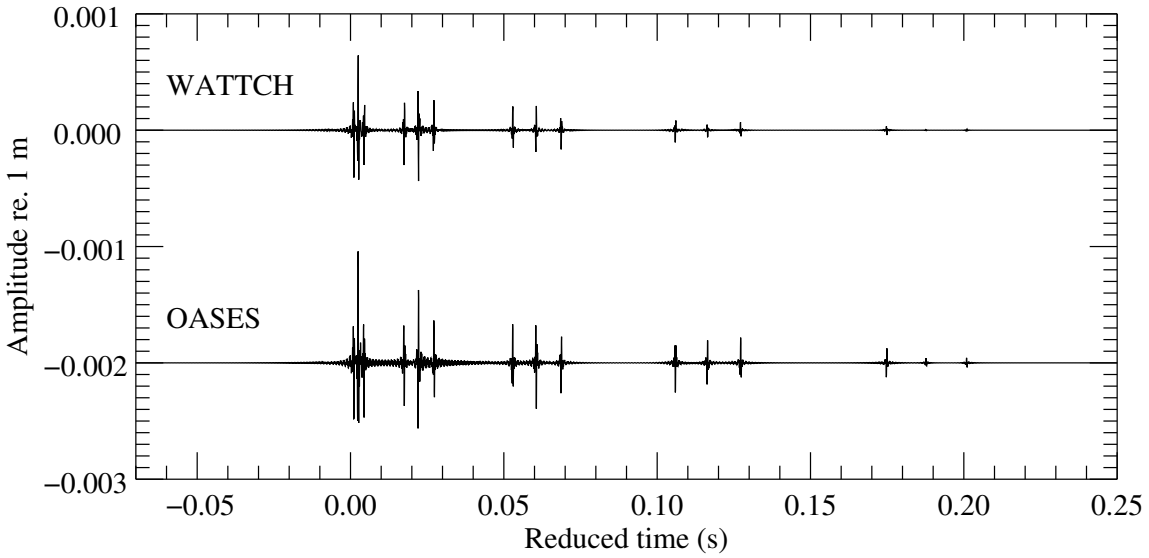


Figure 5.3: Comparison of arrival time series generated by the WATTCH and OASES pulse propagation models for a band-limited impulse function at 2 km distance in a shallow isovelocity ocean environment. Reduced time is defined as $t - R/c$.

compressional sound speed of 1650 m/s. Simulated arrivals were generated for a transmission range of 2 km with a source depth of 20 m and receivers located at 100 m depth, as shown in Figure 5.3. Comparing the timing of the arrivals between the two propagation models shows the arrival times are consistent. The main difference is that the arrivals generated by WATTCH are generally of lower amplitude — this is particularly noticeable for the last two arrival bundles. This is likely because of differences in how boundary interactions are treated between the two models (particularly the ocean-bottom interface). For the purposes, of this research, the absolute levels of the arrivals were not as important as the relative levels; therefore, overall, the models show sufficiently good agreement for the intended use.

5.2 Simulation of Experimental Conditions

To ensure that the WATTCH model was capable of accurately simulating true ocean conditions, the Gulf of Mexico experiment was reconstructed with WATTCH. Comparison of the simulation and experimental results provided a useful baseline and increased confidence in the WATTCH model. In this section the focus is only on experimental results from 1 May 2013 (day 2 of the experiment). This decision was made since there was less variability in source levels than on the first day, and sediment properties were measured along much of the ship's track (which had not been done for the track on day 1). Furthermore, the ship moved away from the moored recorders in a straight line on day 2 which was straightforward to model, whereas the ship's track was more complicated on day 1. The simulation was done only for the shallowest and deepest recorders since experimental results showed little difference in performance as a function of recorder depth. Recall from the discussion of the experimental results that the biogenic and synthetic whale calls produced similar results. Since no additional information was gained from the synthetic signals, the focus in the remainder of this chapter will be on the biogenic whale calls.

5.2.1 Environment and Model Configuration

The experimental geometry and environmental measurements collected during the Gulf of Mexico sea trial (refer to Sections 4.2 and 4.3) were the primary inputs for Bellhop's environment model. The environment model, illustrated in Figure 5.4, consisted of a water column with range-dependent depth varying from 70 to 160 m. The five measured

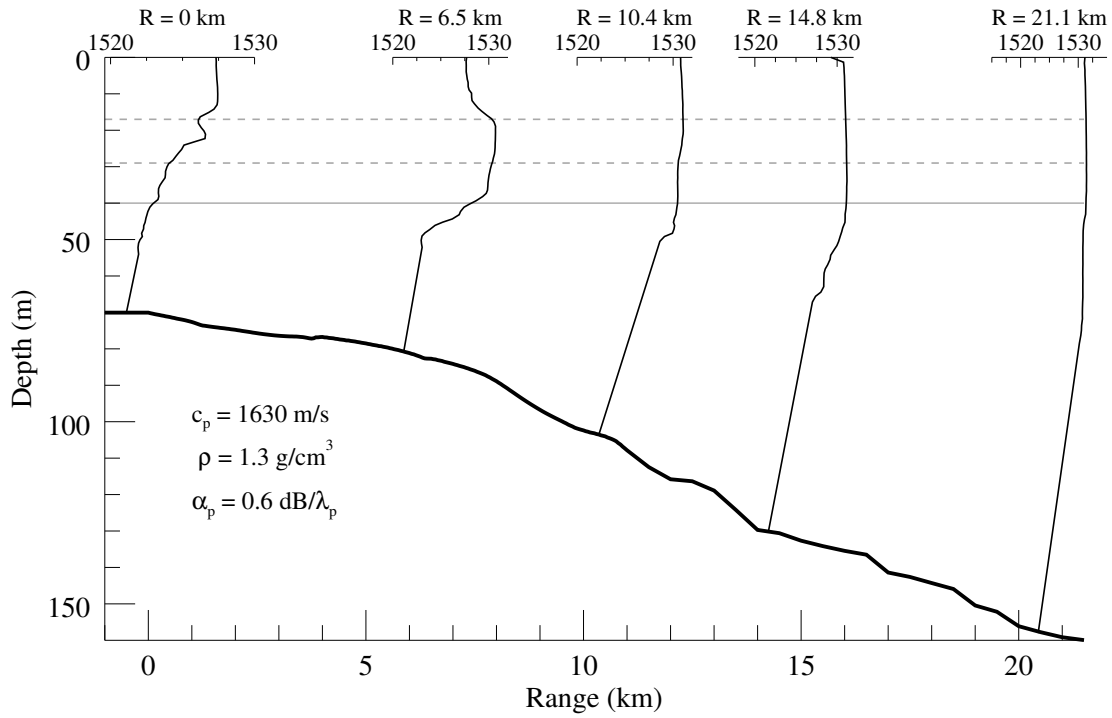


Figure 5.4: Schematic of environment configuration used to simulate the experimental acoustic propagation conditions. The range-dependent sound speed field of the water column is defined in terms of the five sound speed profiles shown (units of m/s). Bathymetry is represented by the thick solid black line. The source and receiver depths are depicted by the horizontal solid and dashed grey lines, respectively. Note the geoacoustic parametrization using the FFCPT data and the APL-UW formulae yielded values of c and ρ lower than those typically associated with sand and sand mixtures; this is discussed further in Appendix E.

sound speed profiles were used to define a range-dependent sound speed field in the ocean medium. A range-independent fluid bottom half space was used with geoacoustic parameters (refer to Figure 5.4 for values) consistent with the sand/sand-mixture bottom type measured by the FFCPT drops; refer to Appendix E for further discussion on the process used to determine the geoacoustic parameters. The source depth was set at 40 m and two receiver depths were modelled at 17 and 29 m. Note that the 17 m deep receiver was on mooring 4 and the 29 m deep receiver on mooring 3, so that the transmission ranges during the experiments were slightly different.

The WATTCH model was used to simulate the received time series by running it for 61 frequencies equally spaced in the 1 to 4 kHz band. These results were then interpolated to provide a 1 Hz frequency spacing and 67 µs temporal spacing, such that when the

Table 5.2: Ranges over which signals were propagated through the WATTCH model for each of the receiver depths. These ranges were based on the mid-ranges over which signals were transmitted during the experiments.

Depth (m)	Ranges (km)
17	1.5, 6.5, 10
29	1, 6.5, 10
30	0.07

inverse Fourier Transform was performed, it produced the acoustic response in a 1 s long time window with a 15 kHz sampling rate. These parameters were selected such that the resulting time window used for the inverse FFT contained all multipath arrival information, while maintaining computational efficiency. A channel impulse response was generated for the ranges relevant to the experimental geometry (refer to Table 5.2), after which the impulse response for each range was convolved with the individual biogenic bowhead and humpback calls. The signals were then centred in their own 3 s long WAVE file — selected to ensure that the output time series was sufficiently long to include the longest duration call with multipath time-spreading and sufficient noise context before and after the signal. Bear in mind, however, that at this point no noise has been added to the simulated signals. Addition of ambient noise is discussed in the following section as there are several subtleties that must be addressed to ensure the correct SNR for signals transmitted over each range.

5.2.2 Noise Addition

One cannot directly account for SNR in a pulse propagation model since the input signal has infinite SNR. To account for the SNR effects, noise was added to the signal at the output of the model. The SNR of experimentally recorded signals which were transmitted over the shortest range, coupled with spreading and absorption predicted by Bellhop, were used to adjust the noise level at each range. It is important to realize that if one were to add noise to the source signal, in an attempt to match the correct SNR at the source, the noise would also propagate along with the signal. This may have unintended consequences by affecting the resulting colouring of the noise. The remainder of this section provides full details of how noise was added to each simulated signal.

Adding ambient noise at the correct level is an important aspect of simulating the signals

which were recorded during the experiments, especially since the experimental results indicated a relationship between SNR and classifier performance. To achieve the most realistic noise characteristics, snippets of experimental noise were added to each of the biogenic bowhead and humpback signals (155 of each) that were propagated through the WATTCH model. These 2.75 s long samples of experimental noise were extracted during the 'quiet' time between recordings of the biogenic whale call transmissions from the icListen recorders located at 17 and 29 m depth. Specificially, 310 noise snippets were extracted corresponding to each range over which the signals were transmitted, for a total of 930 experimental noise samples for each recorder. The duration of the noise samples and time delay after a recorded signal were both selected to ensure that the signals transmitted during the experiments did not contaminate the noise snippets. The same process was followed to extract noise snippets recorded by the monitor hydrophone. These noise samples were then added to the biogenic whale calls which underwent the WATTCH-simulated distortion. This was done in such a way that the resulting SNR for the group of signals transmitted over range R had a similar mean to a modelled estimate of SNR at range R. Thus, the addition of noise was motivated by the need to achieve an accurate representation of the SNR distribution for the simulated signals.

The first step in adding noise at a desired SNR is to determine the power of the WATTCH-propagated signal, S_p, using,

$$S_p = \frac{1}{L}\sum_{i=0}^{L-1} |x_i|^2 \quad , \tag{5.10}$$

where L is the signal length and x_i is the i^{th} signal sample. To do this in practice, the signal variance was used with the assumption that the signal had zero-mean. In order to maintain consistency with the way the aural classifier code calculates the SNR, only a portion of the signal near the peak (i.e., greatest deviation from zero) was used to determine the signal power. The portion of the signal used to estimate the signal power consisted of 75 samples before and 75 samples after the peak (refer to Figure 5.6). Also note that throughout this process, the following definition of SNR (in units of decibels) was employed for consistency with the aural classifier code,

$$\mathrm{SNR}_{dB} = 10\log\left(\frac{\sigma^2_{s+n}}{\sigma^2_n}\right) \quad , \tag{5.11}$$

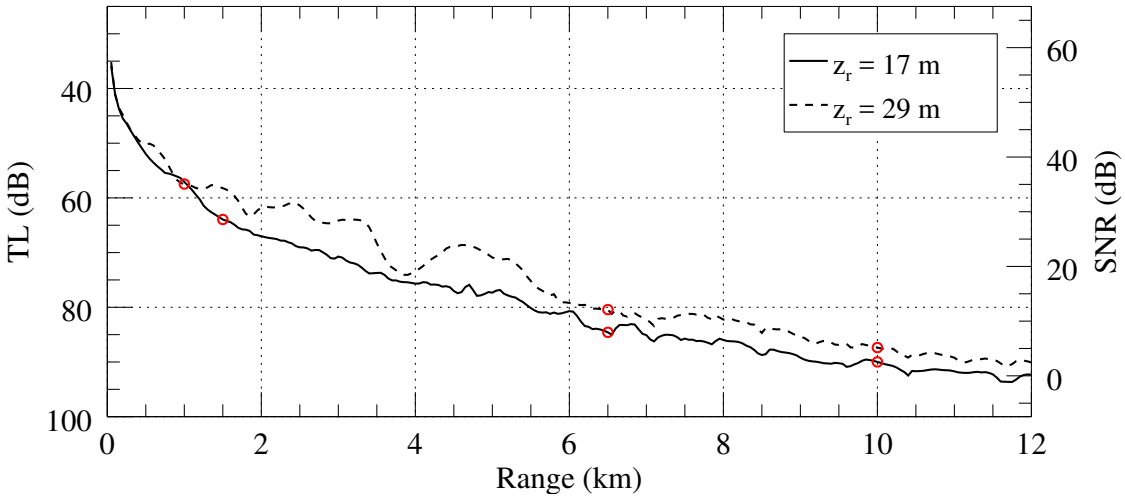


Figure 5.5: Incoherent transmission loss modelled at f =2 kHz using Bellhop for the environment depicted in Figure 5.4. The modelled TL was used to estimate the SNR of signals propagated through the WATTCH model. Red circles represent range-SNR pairs selected for simulation of the experimental results.

where σ^2_{s+n} is the combined signal and noise variance, and σ^2_n is the noise variance.

The desired SNR, SNR_{dB}, of each signal is based on a combination of the SNR statistics from the set of signals transmitted over range R and SNR estimates obtained using Bellhop. In practice, this was done by using Bellhop with the same environment inputs provided to WATTCH (refer to Figure 5.4) to model the TL. The TL curve was then transformed into an SNR estimate by shifting the curve so that the modelled SNR at $R = 1.5$ km matched the mean SNR of the corresponding set of signals recorded on the icListen hydrophone at 17 m depth. This relationship between modelled TL and SNR is demonstrated in Figure 5.5. The TL curves were generated at a frequency of 2 kHz, the geometric mean ($f_m = \sqrt{f_1 f_2}$, where f_1 and f_2 are the lower and upper limits of the frequency band of interest) of the 1 to 4 kHz band, which is a standard approach for estimating the TL of a broadband signal [82]. The SNR curve was then used as a look-up table for determining the mean value of SNR, $\mu_{\mathrm{SNR},R}$, for a particular value of R. This is highlighted in Figure 5.5 by the red circles placed at the range-SNR pairs used to provide an estimate of the desired SNR. Thus, when adding noise to a signal the desired SNR value, SNR_{dB}, is defined as,

$$\mathrm{SNR}_{dB} = \mu_{\mathrm{SNR},R} + u\sigma^2_{\mathrm{SNR},R} \quad , \tag{5.12}$$

where $\sigma^2_{\mathrm{SNR},R}$ is the variance of the SNR of the set of signals transmitted over range R.

Note that a weight, u, randomly drawn from the uniform distribution $\mathcal{U}(-0.5, 0.5)$ is applied to $\sigma^2_{\mathrm{SNR},R}$ to randomize the desired SNR of each signal. In this way the set of simulated signals has similar SNR statistics as the experimental signals transmitted over the same range.

Next, the desired SNR, which is in units of decibels, is converted to linear SNR, SNR_{lin},

$$\mathrm{SNR}_{lin} = 10^{\frac{\mathrm{SNR}_{dB}}{10}} \quad , \tag{5.13}$$

and the weight, β, to be applied to the noise vector to achieve the desired SNR is determined using,

$$\beta = \sqrt{\frac{S_p}{N\mathrm{SNR}_{lin}}} \quad , \tag{5.14}$$

where N is the noise power, estimated as the variance of the noise time series (with the assumption that it is zero-mean).

Finally, the resulting time series, $y_{t,r}$, is obtained simply by adding the noise time series, n_t, with the appropriate weight, β to the WATTCH-simulated signal, $x_{t,p}$,

$$y_{t,r} = x_{t,p} + \beta n_t \quad . \tag{5.15}$$

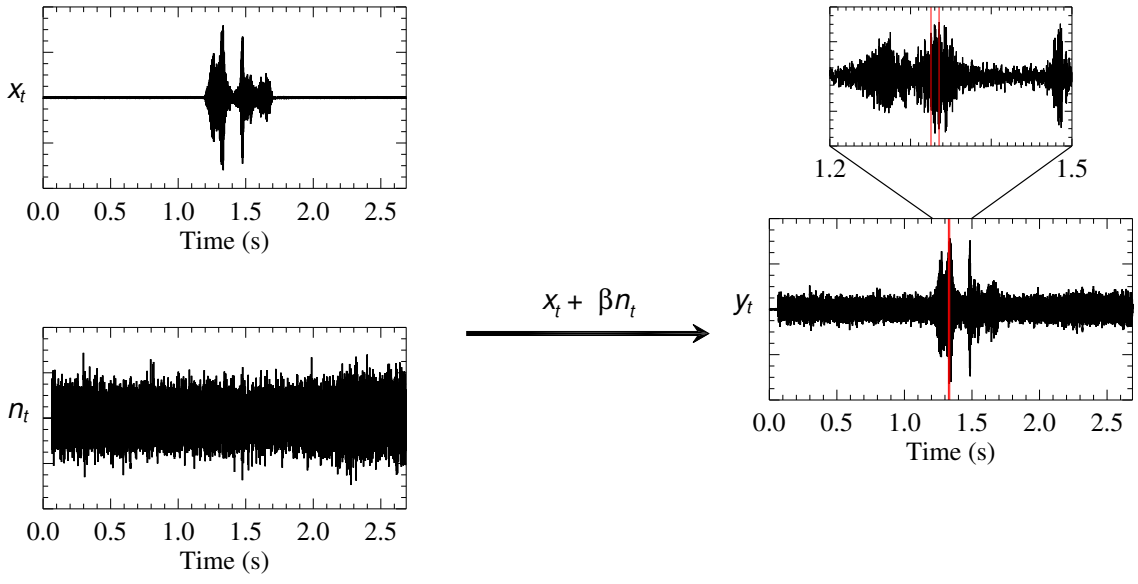


Figure 5.6: Schematic representation of noise addition to a signal propagated through the WATTCH model. The red vertical lines shown on y_t demarcate the region of the signal used to calculate SNR.

This process was repeated for each of the biogenic signals propagated through WATTCH until the appropriate level of noise was added to each signal to produce an ensemble of simulated signals with SNR statistics representative of the experimental signals being simulated. A pictorial summary of this process is provided in Figure 5.6

5.2.3 Results and Discussion

After the signals were propagated through the environment with the WATTCH model, and noise was added to provide the SNR level predicted by Bellhop, each of the signals was processed through the aural classifier algorithm. This process was intended to be analogous to the method employed to process the signals recorded during the experiments; there was, however, one notable exception. The band-limited energy detector (refer to Section 3.4) which was used to extract signals from the experimental recordings was not applied to the simulated signals. The detection step was deemed unnecessary because the SNR statistics were similar to those of the detected calls. Additionally, by not applying the detector there was also a larger set of calls to work with at the longer ranges.

Classifier performance results for the simulated calls as a function of range are presented in Figure 5.7; these results are also compared to the experimental results they are intended to simulate. For the simulated calls, the classifier was trained with signals which were simulated as if they had been recorded on the monitor hydrophone. Similarly, the experimental performance results were generated by training the classifier on signals recorded by the monitor hydrophone and validating the classifier on signals transmitted over the longer ranges. The experimental results are the same as those presented in Figure 4.10. The reader will note that no horizontal error bars have been included for the simulated results — this was because only a single range was used to produce the simulated results as the model was not capable of easily simulating the ship drift experienced during the experiments. In general, the performance results from the simulated calls produce a good match to the experimental results, although at the two longer ranges ($R = 6.5$ and 10 km), the classification accuracies determined for the simulated calls were higher than the experimental results. Overall, a better match between the experimental and simulated results was obtained for the AUC than for the accuracy. This means that the simulation results did well at predicting how the bowhead and humpback class distributions shift relative to each other, but not quite as well at estimating how they shift relative to the decision

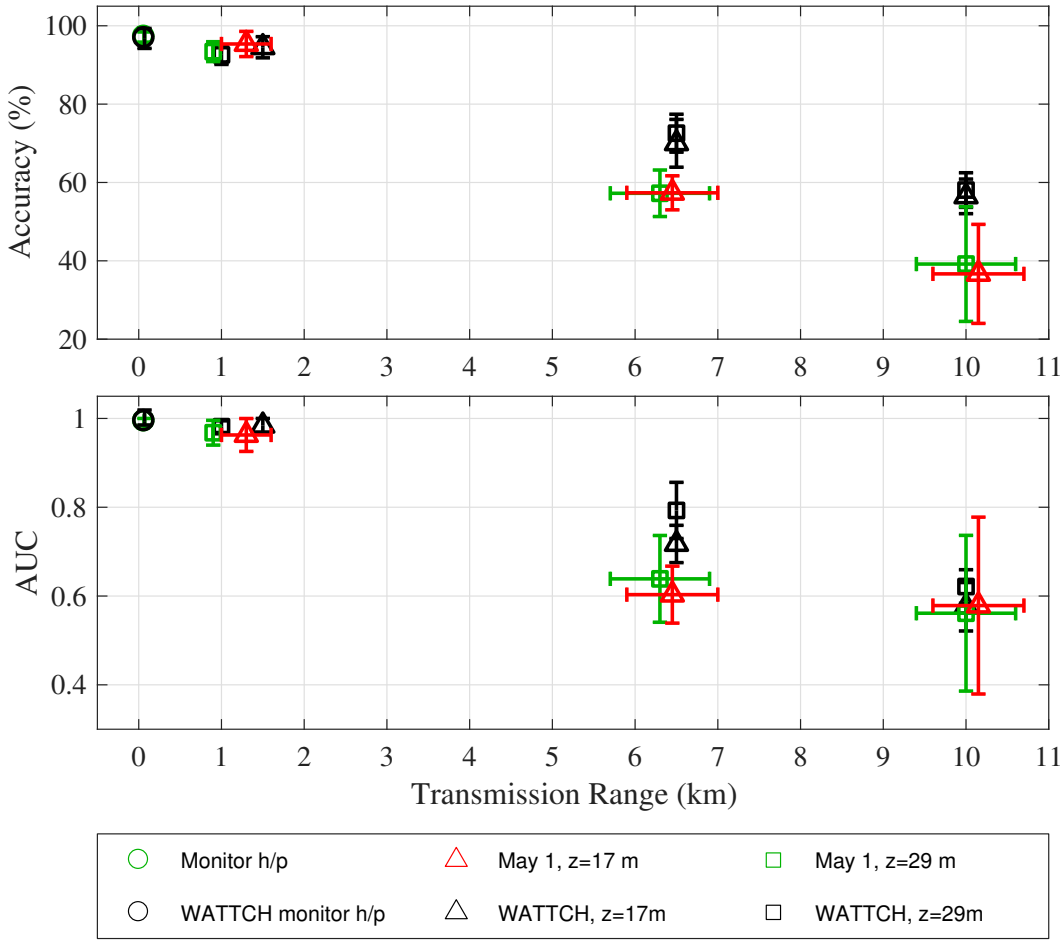


Figure 5.7: Classification performance as a function of range for the simulated biogenic whale calls that were propagated through the WATTCH model, compared with experimental results. The vertical errors bars show $\pm\sigma_{\text{accuracy}}$ and $\pm\sigma_{\text{AUC}}$ based on 5-fold cross-validation, and the horizontal error bars define the shortest and longest transmission range for each transmitted signal set.

threshold defined by the training dataset. This indicates that the decision threshold was more sensitive to environmental uncertainty than the AUC.

Since SNR is assumed to be an important contributor to decreased performance, it is also important that the WATTCH model capture the same relationship between classifier performance and SNR that was noted in the experimental results. As the model is necessary for separating the performance dependence on SNR and acoustic propagation-induced distortion, it is critical that the model be able to reproduce the correct SNR-performance trend. Thus, simulated performance results were compared against the experimental data in Figure 5.8, with SNR as the independent variable. As can be seen from this figure,

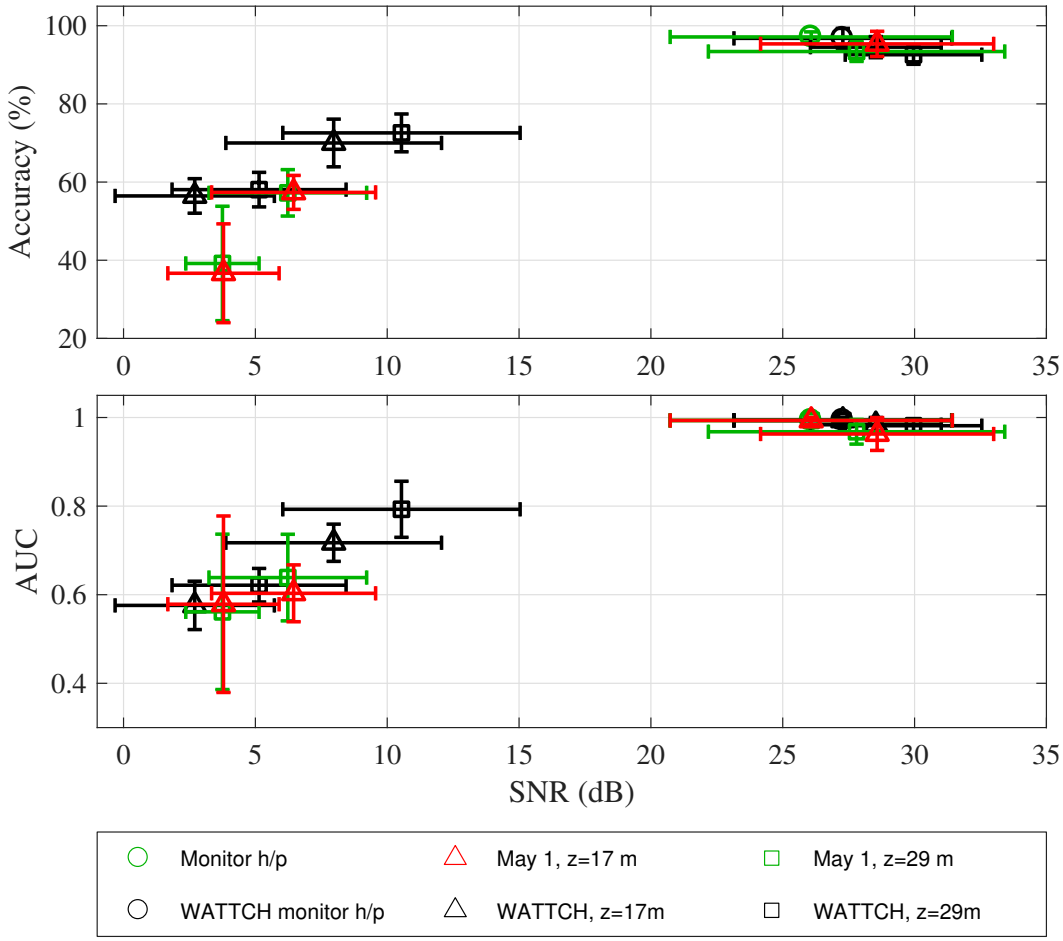


Figure 5.8: Classification performance as a function of SNR for the simulated biogenic whale calls that were propagated through the WATTCH model, compared with experimental results. The vertical errors bars show $\pm\sigma_{\text{accuracy}}$ and $\pm\sigma_{\text{AUC}}$ based on 5-fold cross-validation, and the horizontal error bars define $\pm\sigma_{\text{SNR}}$ of the corresponding recorded signal set.

the simulated calls generally did a good job reproducing the trend established from the experimental results (i.e., increasing SNR caused classifier performance to increase). There are a few features of this plot that must be addressed. First, the modelled SNR for the simulated signals did not provide an exact match to the experimentally-derived SNR values. In general, Bellhop overestimated the SNR at each range, indicating there was some loss mechanism in the measured data that was either not accounted for (e.g., transfer of energy into shear waves in the bottom, scattering in the water column) or underestimated (e.g, scattered energy at the boundaries, bottom loss). Nonetheless, the SNR estimates for the simulated signals and experimental data were within one standard deviation of each other

for matching ranges. Second, given that for the majority of cases the model overestimated the SNR, it is interesting to note that for the signals received at 17 m depth/10 km range, the model estimated an SNR marginally lower than that observed experimentally. The reader will also note that the horizontal error bars are wider for the simulated results than for the experimental results. This was likely a result of a bias in the experimental SNR estimate caused by the detection process that shifted the SNR estimate to a higher value. As a result of the detection process, lower SNR calls were removed from the set of experimental signals, which resulted in an increase in the mean SNR value and a decrease in the standard deviation. Since the simulated signals were not subjected to the detection process, there was no such change in their SNR statistics. Finally, viewing the performance results with respect to SNR suggests that the classifier performance was generally higher at each range step for the simulated whale calls because the mean SNR was also greater than for the experimental results. As has previously been noted, classifier performance has a strong dependence on the SNR of the signals being classified.

In summary, the WATTCH model did a remarkably good job of matching the experimental performance results, both in terms of range (Figure 5.7) and SNR (Figure 5.8), given the highly dynamic ocean environment encountered during the experiments. The degree of model-data mismatch shown in these results was considered acceptable since, as Jensen [97] points out, the performance of a model in replicating experimental results depends on the predictability of ocean acoustic propagation. The ocean is a complex environment with both temporal[1] and spatial variability that cannot always be captured in a deterministic modelling approach [97]. This statement is particularly relevant based on the broad range of ocean environment measurements collected during the experiments. Recall from the discussion in Section 4.3, that significant change of the ocean acoustic environment (e.g., surface roughness and sound speed profiles) occurred throughout the course of the experiment. Consequently, it was difficult to choose model parameters to accurately capture those changes. This was partly because the environment sampling strategy — which is a widely used method — confounded temporal and spatial changes; in future experiments, environmental moorings should be used to complement ship-based environmental measurements. Furthermore, environmental uncertainty accumulated for

[1] In the modelling approach taken here, a 'frozen ocean' was implicitly assumed. That is, this thesis has made use of representative SSPs, when in fact the 'true' SSPs are best represented as functions of both depth and time, $c(z,t) = c(z) + \mathrm{d}c(z,t)$, where $\mathrm{d}c(z,t)$ represents fluctuations in the sound speed that may occur over relatively small values of t.

the longer ranges/later transmission times. For example, when signals were transmitted over the 10 km range, did they experience a water column near the moorings that could be characterized by the SSP shown for the 0 km range in Figure 5.4? This is a difficult question to answer definitively, given that the SSP near the moorings was measured at least three hours prior to the transmissions at the 10 km range. In this way, the model-data mismatch could be a result of sensitivity of the classifier performance to environmental parameters; this issue will be considered in the following section.

Ultimately, the performance results for the simulated calls compared favourably with those derived from the experimental data; hence the model was validated and can be used to explore aspects of the classifier performance in terms of environment-dependent propagation that cannot be addressed merely with the data collected during the Gulf of Mexico experiment.

5.3 Sensitivity to Environmental Parameters

The overarching goal of this thesis is to examine if and how the aural classifier performance is impacted by environment-dependent acoustic propagation. Up to this point, the experimental and model results have indicated that the classifier performance decreased as a function of range and/or SNR. So far, only a single propagation environment has been considered — the one encountered during the experiments. It is difficult to draw general conclusions about the environment-dependence of the classifier without considering other environments, or understanding how certain environment conditions impact the classifier performance. Additionally, propagation modelling can often be sensitive to the choice of input parameters, as alluded to in the previous section. To address this, an exploratory sensitivity analysis is conducted in this section.

Ocean acoustic propagation sensitivity analysis is a broad topic and has been the subject of several conferences (including Refs. [106, 107]) and a U.S. Office of Naval Research (ONR) Directed Research Initiative that culminated in a journal special issue (Ref. [108]). Sensitivity studies typically examine the influence of environmental parameters on propagation in a particular environment. Sensitivity analyses may also be used to identify parameters that have the greatest effect on the acoustic field; these parameters are important to identify since errors or uncertainties in model inputs for these parameters can generate

errors in model predictions [109].

From an extensive review of the literature addressing the sensitivity of acoustic propagation to the ocean environment, Giles [110] concluded that the water column sound speed profile often has the greatest impact on acoustic propagation. Additionally, Dosso *et al.* [109] found that there was a significant sensitivity to the water column sound speed profile and a relative insensitivity to other environmental parameters (e.g., geoacoustic parameters of the sediment) for the environment which they were considering. Furthermore, an environmental parameter may be considered sensitive in an environment where its uncertainty is large, since its variability can significantly affect acoustic fields [12]. Given that there was a large variation in the characteristics of the SSPs measured during the Gulf of Mexico experiments (not just small perturbations from a mean SSP, but distinct changes in the shape of the SSP), one would expect that the uncertainty in the propagation modelling was dominated by the SSP in the ocean medium, and the impact of the other environmental parameters to be small in comparison. Based on these points, the sensitivity analysis discussed here will focus solely on changes in the acoustic field (and thus the classifier performance) due to the SSP input to the WATTCH model. Further analysis of the sensitivity of other environmental parameters was considered to be beyond the scope of this investigation.

The approach for this sensitivity analysis was to examine the impact of SSP on aural classifier performance by comparing the results from the detailed environmental model employed in the previous analysis (Section 5.2) with environmental models for which the water column properties were defined using a range-independent sound speed field. The intent was to focus on environmental measurements collected during the experiment, and to adjust the SSP within the bounds of the observed data. In this way, it is possible to comment on the importance of accurately representing the acoustic environment to achieve an acceptable model-data match.

5.3.1 Environment and Model Configurations

The environment and model configurations for this sensitivity analysis were the same as for the propagation model described in Section 5.2.1, with the exception of the SSP used to define the properties of the ocean medium. Four cases were considered; three range-independent cases, each defined by a single SSP:

1. *Well-mixed*. This refers to the SSP which was approximately isospeed in the upper 60 m of the water column, shown at $R = 21.1$ km in the environment configuration figure (Figure 5.4);

2. *Downward refracting*. This refers to the SSP located at $R = 0$ km in the configuration figure;

3. *Anti-duct*. This refers to the SSP which had a sound speed maximum approximately mid-watercolumn, shown at $R = 6.5$ km in the configuration figure;

and the fully range-dependent case used to simulate the experimental results in Section 5.2. As before, noise was added to each of the signals at an appropriate level based on Bellhop TL modelling. The results of the TL modelling for each environment configuration are shown in Figure 5.9. Note the large variability in TL (e.g., more than a 15 dB difference between the anti-duct and downward refracting cases measured at $R = 6.5$ km and $z = 29$ m), and thus SNR. This was a significant difference, bearing in mind the only difference between the environments was the SSPs. Also of interest were the relatively large differences in the TL values based on receiver depth, especially for the downward refracting sound speed profile. These TL curves demonstrate that the propagation model exhibited a high degree of sensitivity to the input SSP, but the question remains: how does this impact the aural classifier results?

5.3.2 Results and Discussion

Following the same steps outlined in Section 5.2, signals were propagated through the environment with the WATTCH model and noise was added at the correct level for each of the environments under consideration. The classifier was then trained on signals propagated over a 70 m range and validated on signals propagated over the longer ranges. Classifier performance results are exhibited in Figures 5.10 and 5.11 as a function of range and SNR, respectively. There was a large spread in performance values, based on the choice of SSP, for signals transmitted over the 6.5 km and 10 km ranges. The mean performance results were consistently larger for signals propagated through the downward refracting water column than for the other cases. Surprisingly, there was little difference between the well-mixed and anti-duct cases — intuitively one may have presumed that since the characteristics of the SSPs were so different they would have produced significantly

different results. That the opposite was true may have been a result of the source/receiver geometry, particularly for the anti-duct case. If, for example, the source was $\sim$10 m deeper it would have been below the sound speed minimum which would have restricted the amount of energy incident on either receiver, and likely would provide a greater contrast between the anti-duct and well-mixed cases.

Nevertheless, these results showed a clear environment-dependence. The propagation model's sensitivity to the properties of the ocean medium, as characterized by the SSP, was sufficiently large that one may conclude it to be a significant contributor to the model-data mismatch observed in Section 5.2, especially given the wide variation in SSPs measured over the course of the experiment. This is a key result as it reinforces the necessity to characterize the ocean acoustic environment as fully as possible during an experiment, especially if one plans to augment experiments with propagation modelling. This may seem obvious; however, consider if only a single SSP had been measured when the moorings were deployed. One would then necessarily assume that it was applicable for the surrounding medium and thus generate an environment model using the downward refracting SSP. This would have resulted in a significant model-data mismatch with no

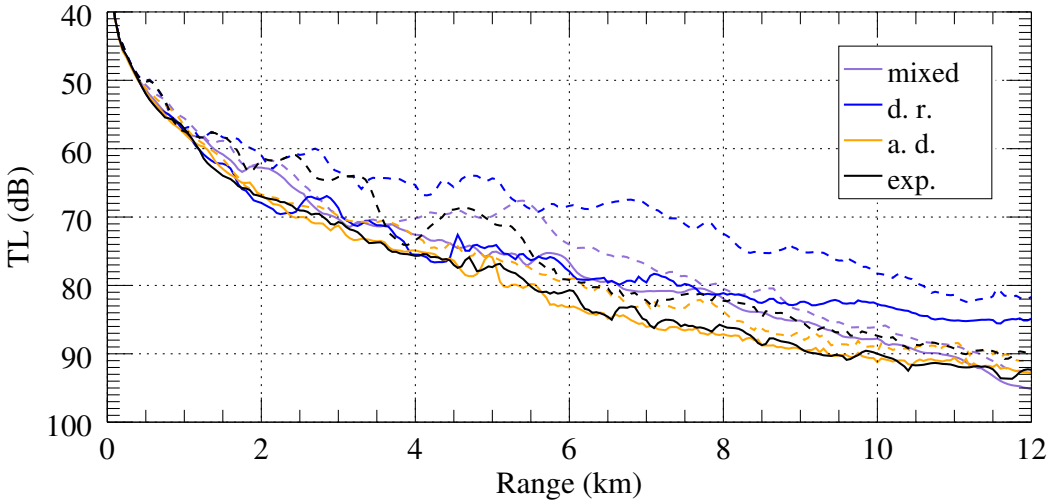


Figure 5.9: Incoherent transmission loss modelled at $f = 2$ kHz using Bellhop for four different sound speed fields. The SSPs used as inputs to the model are shown in Figure 5.4; well-mixed (mixed) SSP at $R = 21.1$ km, downward refracting (d. r.) SSP at $R = 0$ km, anti-duct (a. d.) SSP at $R = 6.5$ km, and the experimental simulation (exp.) sound speed field with all five SSPs (the values of R are referenced to Figure 5.4). For each sound speed field case, the solid line (———) represents TL values for a receiver at 17 m depth, and the dashed line (– – –) for a receiver at 29 m depth.

obvious explanation. Therefore, one must consider the relevant oceanographic scales that may impact acoustic propagation and implement an environment sampling strategy accordingly.

Examining the performance results in terms of SNR provided some explanation for the difference between environments, first noted by examining the performance as a function of range. As with previous results, there was a clear relationship between performance and SNR. An interesting feature of these plots is that the performance values for signals transmitted over 10 km through the downward refracting environment bore the

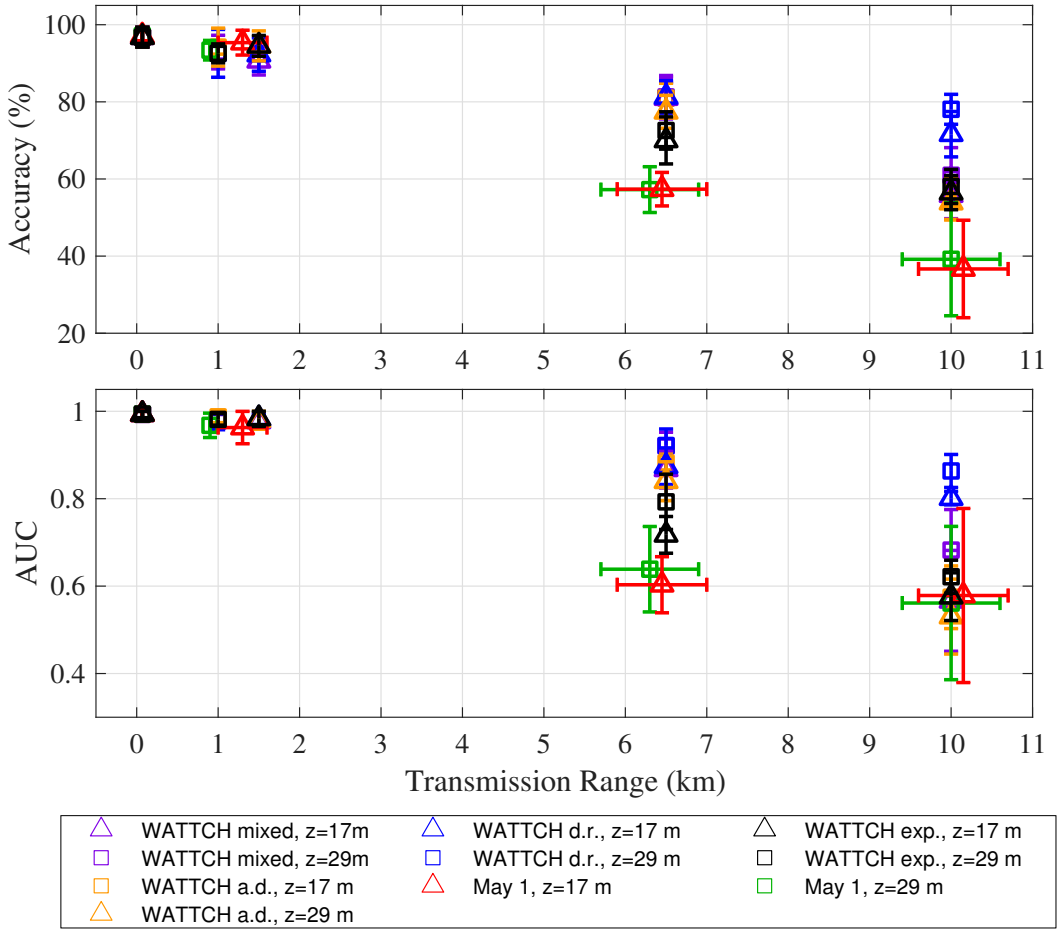


Figure 5.10: Sensitivity of classification performance to choice of sound speed profile as a function of range for the biogenic whale calls propagated through WATTCH model. Experimental results (red triangles and green squares) are included for reference. The vertical errors bars show $\pm\sigma_{\text{accuracy}}$ and $\pm\sigma_{\text{AUC}}$ based on 5-fold cross-validation, and the horizontal error bars on the experimental results define the shortest and longest transmission range for each transmitted signal set.

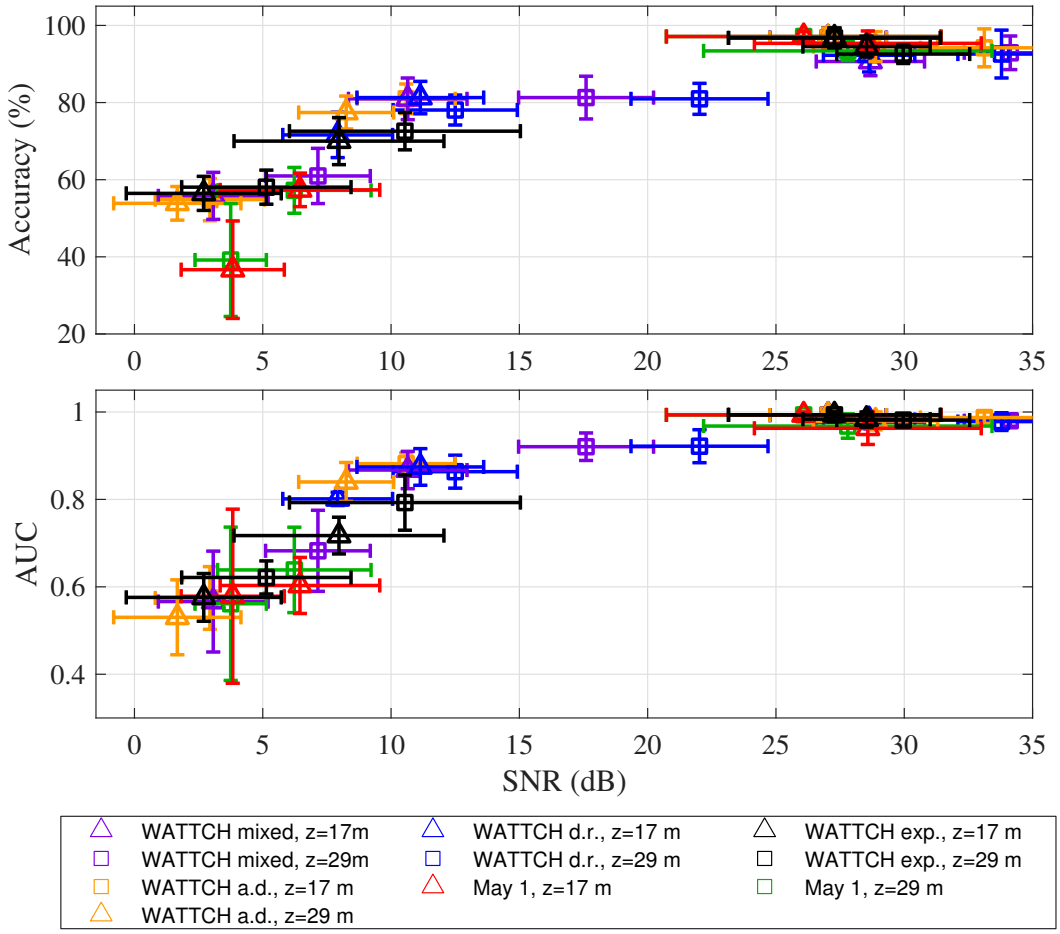


Figure 5.11: Sensitivity of classification performance to choice of sound speed profile as a function of SNR for the biogenic whale calls propagated through WATTCH model. Experimental results are included for reference. The vertical errors bars show $\pm\sigma_{\text{accuracy}}$ and $\pm\sigma_{\text{AUC}}$ based on 5-fold cross-validation, and the horizontal error bars define $\pm\sigma_{\text{SNR}}$ of the corresponding recorded signal set.

most similarity to the performance of signals propagated over only 6.5 km in the other cases — this was likely because the mean SNRs for the downward refracting case at the 10 km range were more similar to the mean SNRs for the other cases at the 6.5 km range (i.e., between 5 to 15 dB). It is therefore tempting to conclude that the sensitivity of the classifier performance was a result of environment-dependent TL (i.e., signal attenuation) and neglect the impact of signal distortion induced by time-spreading. Although there clearly could be a large performance difference between environments, one cannot ignore that there was, in general, an insignificant difference within environments. To clarify, performance results for a particular environment were similar for the two receiver depths, even though in some cases there were relatively large SNR differences for these receivers.

Table 5.3: Example performance values for signals propagated through the WATTCH model over 6.5 km through the downward refracting environment.

Depth (m)	SNR (dB)	Accuracy	AUC
17	11.14 ± 2.47	0.81 ± 0.04	0.87 ± 0.04
29	22.02 ± 2.67	0.81 ± 0.04	0.92 ± 0.04

Consider again the downward refracting case for signals propagated over 6.5 km — despite the $\sim$10 dB difference between the mean SNR values for the two receiver depths, there was no difference in the accuracy and a negligible difference in the AUC values (refer to Table 5.3 for comparison of these values). This indicates that, although SNR explains much of the change in performance, signal distortion may also account for some of the performance degradation.

To recap, regardless of the source mechanism, it was shown that the classifier performance was sensitive to changes in the acoustic propagation field that resulted from differences in the sound speed profiles. Furthermore, differences in environment-dependent attenuation largely contributed to the performance degradation. Nevertheless, there was evidence that suggested signal distortion due to time-spreading also contributed to the environment-dependent differences in classifier performance.

5.4 Chapter Summary

In this chapter, the WATTCH model was introduced and validated for broadband pulse propagation. It was then used to simulate the environmental conditions experienced during the Gulf of Mexico experiments, resulting in an acceptable model-data match and further validating the applicability of the WATTCH model. Since WATTCH is much more flexible and efficient than performing multiple at-sea experiments, it was an ideal tool to conduct an exploratory sensitivity analysis. Variation in classifier performance related to water column properties were explored through changes to the sound speed field. The results clearly showed variability in the classifier performance results based on the properties of the ocean medium. Additionally, it was demonstrated that these differences could largely be explained by disparity in the environment-dependent TL; however, some of

the characteristics of the sensitivity analysis indicated that signal distortion due to time-spreading as a result of multipath addition may also have been important. The combination of propagation model results and sensitivity analysis reinforced the need to study how much of the classifier performance degradation is explained merely by decreasing SNR and how much may be attributed to environment-dependent distortion of the signals. Due to the flexibility of the WATTCH model, it is possible to use it to investigate the relative impacts of SNR and other propagation effects.

CHAPTER 6

RELATIVE IMPACT OF SNR AND OTHER PROPAGATION EFFECTS

6.1 Introduction

Both the experimental results presented in Chapter 4 and the propagation modelling results presented in Chapter 5 suggested that signal-to-noise ratio (SNR) was an important contributor to decreasing classifier performance as signals were transmitted over increasing range. This finding was consistent with previous PAM results, including those of Helble *et al.* [18], for example, which show the probability of detecting humpback whale units changes considerably on both short and long time scales with altering ocean noise characteristics. Mouy *et al.* [23] also noted that false negative rates increased as SNR decreased. Murphy and Hines [79] observed a similar trend of decreasing performance in their investigation of the SNR-dependence of the aural classifier when applied to active sonar echoes. Mellinger and Clark [52] explicitly state that, "The lower the signal-to-noise ratio of a sound, the more difficult it is, in general to detect and classify the sound." Certainly, it is not a new idea to suggest that SNR has an impact on automated detectors and classifiers. The novel elements of this chapter are the consideration given to environment-dependent propagation effects and the methods employed to disentangle the relative impacts of SNR and signal distortion caused by time-spreading.

In discussing the theory of communication over a channel with noise, Shannon [80] distinguishes two cases: distortion and noise. In the first case, a transmitted signal always produces the same received signal, though it is different from the original transmission. The received signal is a definite function of the transmitted signal and the process which

transforms the transmitted signal is referred to as *distortion*. In the second case the signal does not always undergo the same change in transmission, i.e., the received signal is a function of both the transmitted signal and a random variable referred to as *noise* [80]. A convenient way to distinguish these two cases is that, in the case of distortion, it is theoretically possible to predict the received signal precisely if the input signal and transfer function of the channel are known; it is not possible to predict the received signal exactly for a noisy channel since the noise is non-deterministic. In general, signals recorded during ocean acoustic experiments are subject to transformation by both distortion and noise — the distortion is a result of deterministic propagation effects due to time and frequency spreading, and the irreversible changes to the transmission may be attributed to the ambient ocean noise. This chapter of the thesis investigates the relative impacts of distortion and noise on the aural classifier's performance by considering how each individually alters the classifier performance.

The combination of experimental and modelling results presented in Chapters 4 and 5, results presented in the literature, and Shannon's communication theory all indicate that SNR plays a part, and may even drive, the decrease in classifier performance as signals are transmitted over increasingly longer ranges. However, propagation-induced distortion may also impact the performance of an automated classifier. The question therefore becomes, *what are the relative contributions to reduced classifier performance that arise from decreasing SNR and from propagation-induced signal distortion?* While it is complicated to design at-sea experiments to separate the effects of SNR from propagation-induced signal distortion, a pulse propagation model can easily separate the influences of SNR and propagation. Accordingly, a simulation study using the biogenic bowhead and humpback whale calls was performed by individually considering the impacts of decreasing SNR and increasing distortion due to propagation effects, as well as investigating the combination of these factors. The simulations consisted of three test cases:

1. *Noise-only:* This first case simulated the scenario in which only SNR contributes to decreasing the classifier performance. To accomplish this, noise was added to each of the signals to achieve a desired SNR. Initially, noise was added to the signals at levels consistent with modelled TL for the ranges over which signals were transmitted during the experiment (i.e., $R = \{0.07,\ 1,\ 6.5,\ 10\}$ km) to facilitate comparison with experimental results (in Section 6.2). After validation of the methods, the

process was repeated with increased range/SNR resolution (in Section 6.3) for two environments, each defined by a different range-independent SSP.

2. *Propagation-only:* This second case simulated the scenario that only propagation effects are applied to the signal. No noise was added to these signals, such that the SNR was infinite. As described in Chapter 5, the WATTCH pulse propagation model was used to determine the acoustic response at the same points, (R, z), simulated in the noise-only case. After the range-independent impulse responses were determined, they each were convolved with the biogenic bowhead and humpback vocalizations.

3. *Noise and propagation:* This final case is the closest to reality, since both noise and propagation effects were added to the signals. Noise was added to each of the simulated signals, generated for the propagation-only case, to achieve the SNR predicted from the modelled TL.

Further details are given in the following section of how each of these three cases were constructed. Although the use of the WATTCH pulse propagation model was validated in Chapter 5, experimental data are also included throughout this chapter, when appropriate, as a means to validate the methods by ensuring sufficient model fidelity.

6.2 Separating Contributions of Propagation-Induced Distortion and SNR

The first simulation scenario was motivated by experimental measurements and the results of the sensitivity analysis in Chapter 5. Here, the key difference was that the influences of SNR and other propagation effects must be considered separately. Data were simulated as if they had been received by the icListen recorder on mooring 4 (i.e., at a depth of 29 m). The ocean bottom was described using a sand/sand-mixture bottom and the measured bathymetery (refer to Figure 5.4). The water properties were assumed to be range-independent and characterized by the anti-duct SSP. This SSP was selected since the sensitivity analysis conducted in Section 5.3 showed that, in general, signals propagated through the anti-duct SSP environment produced performance results similar to the experimentally observed results. It is worth noting that the signals with both additive

propagation and noise (case three) were the same as those used in the sensitivity analysis in Section 5.3.

Two methods were considered for adding noise to the signals. These consisted of adding either snippets of experimentally recorded noise, or white noise (referred to as additive white Gaussian noise, or AWGN). Both of these methods have been used by other researchers in the past; for example, samples of experimentally recorded noise were added to whale calls that were artificially propagated through an ocean acoustic environment to analyze a generalized power-law detector's performance [111] and to investigate the impact of ambient noise and site-specific propagation effects on a detector's performance [18]. On the other hand, Xian [89] used additive white noise to investigate the effectiveness of a Short Time Fourier Transform detector in different ocean acoustic propagation environments. Both types of additive noise have benefits — the experimental noise samples provide the most accurate representation of the experimental conditions, whereas the statistics of the white noise are well understood, such that it is possible to produce many samples. Both methods are investigated to determine if one is clearly preferable for simulations.

6.2.1 Case 1: Adding Noise

The technical details of how noise was added to the signals are given in Section 5.2. In fact, the additive noise with propagation results shown in Figures 6.1 and 6.2 are the same data points as in Figures 5.10 and 5.11. The same methods were used here with one exception: for this noise-only case, noise was added directly to the biogenic bowhead and humpback vocalizations without any simulation of the propagation effects. Thus, the resulting time series, $y_{t,n}$, was obtained simply by adding the noise time series, n_t, with the appropriate weight, β, to the signal, x_t,

$$y_{t,n} = x_t + \beta n_t \quad . \tag{6.1}$$

This process was repeated for each of the bowhead and humpback biogenic signals until the appropriate level of noise was added to each signal to achieve an ensemble of signals with SNRs similar to that estimated using Bellhop (refer to Figure 5.9). Bear in mind that this process was conducted twice using either additive experimental noise or AWGN.

6.2.2 Case 2: Adding Propagation Effects

The WATTCH model was used to compute the impulse response of the ocean environment based on *in situ* measurements collected on the second day of the Gulf of Mexico experiment. Water depth was defined by the experimentally measured bathymetry, and bottom parameters consistent with the type of sediment measured by the FFCPT were selected: $c = 1630$ m/s, $\rho = 1.3$ g/cm^3, $\alpha = 0.6$ dB/λ_p. The water column properties were characterized by the anti-duct SSP. The acoustic response was determined for a receiver at depth $z_r = 29$ m and transmission ranges of $R = \{1, 6.5, 10\}$ km. Additionally, signals received by the monitor hydrophone were simulated by placing a receiver at 30 m depth and 70 m range.

The channel impulse responses produced by the WATTCH model were convolved with each of the bowhead and humpback biogenic signals. This process is represented succinctly by,

$$y_{t,p} = x_t * h_t\left(R, z\right) \quad , \tag{6.2}$$

where $y_{t,p}$ is the output time series, x_t is one of the biogenic whale calls, and $h_t\left(R, z\right)$ is the channel's range- and depth-dependent impulse response. No noise was added to these so that propagation effects could be examined in the absence of SNR considerations; this has the side effect that the SNR was infinite.

6.2.3 Case 3: Adding Noise and Propagation Effects

This final case combines the methods of the noise-only and propagation-only cases. Noise was added to the signals from case 2 at a level to match the SNR modelled by Bellhop, for each value of R. The resulting time series, $y_{t,r}$, is obtained using,

$$y_{t,r} = x_t * h_t\left(R\right) + \beta n_t \quad . \tag{6.3}$$

This is merely a combination of Equations 6.1 and 6.2.

6.2.4 Results and Discussion

Classifier performance results obtained from training the classifier on data from the 70 m range and validating the classifier on signals propagated over each of the longer ranges are presented in Figure 6.1, for each of the three cases described in Sections 6.2.1 to 6.2.3.

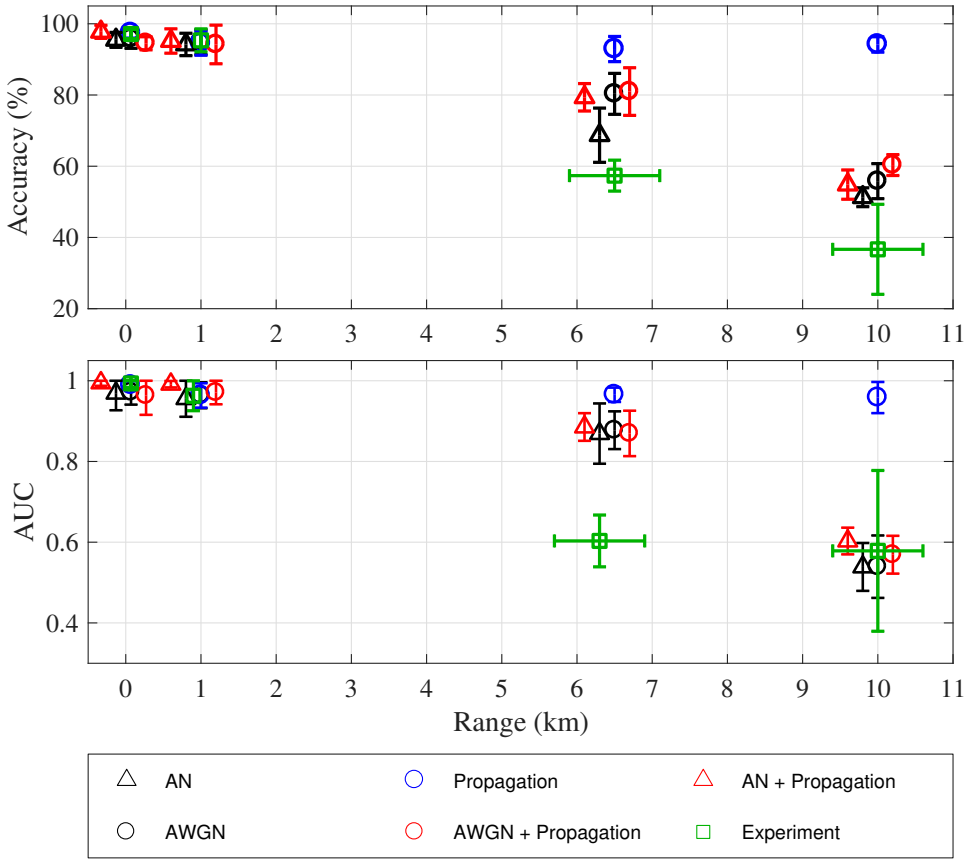


Figure 6.1: Comparison of classifier performance as a function of range for the biogenic whale calls in five different cases: (i) experimental noise (AN), (ii) simulated propagation effects, (iii) both experimental noise and simulated propagation effects, (iv) white Gaussian noise (AWGN), and (v) both AWGN and propagation effects added to the signals. Performance results from these simulated scenarios are also compared with results from experimental data. Note that data points for signals transmitted over the same range have been offset in range to reduce overlap, thereby facilitating comparison of the performance results.

Each data point represents the mean performance value (either in terms of accuracy or AUC) derived from 5-fold cross-validation and the vertical error bars show one standard deviation of the cross-validation results. The same data are plotted in Figure 6.2 as a function of the mean SNR of the signals; the signals to which only propagation effects were added are not included in this plot as the SNR was infinite. Experimental data are also included in both Figures 6.1 and 6.2 to facilitate model-data comparison.

The classifier's performance for the noise-only case behaved as one would expect: performance decreased with increasing range and decreasing SNR. As the SNR decreased, the signal information used by the classifier to discriminate between the vocalizations

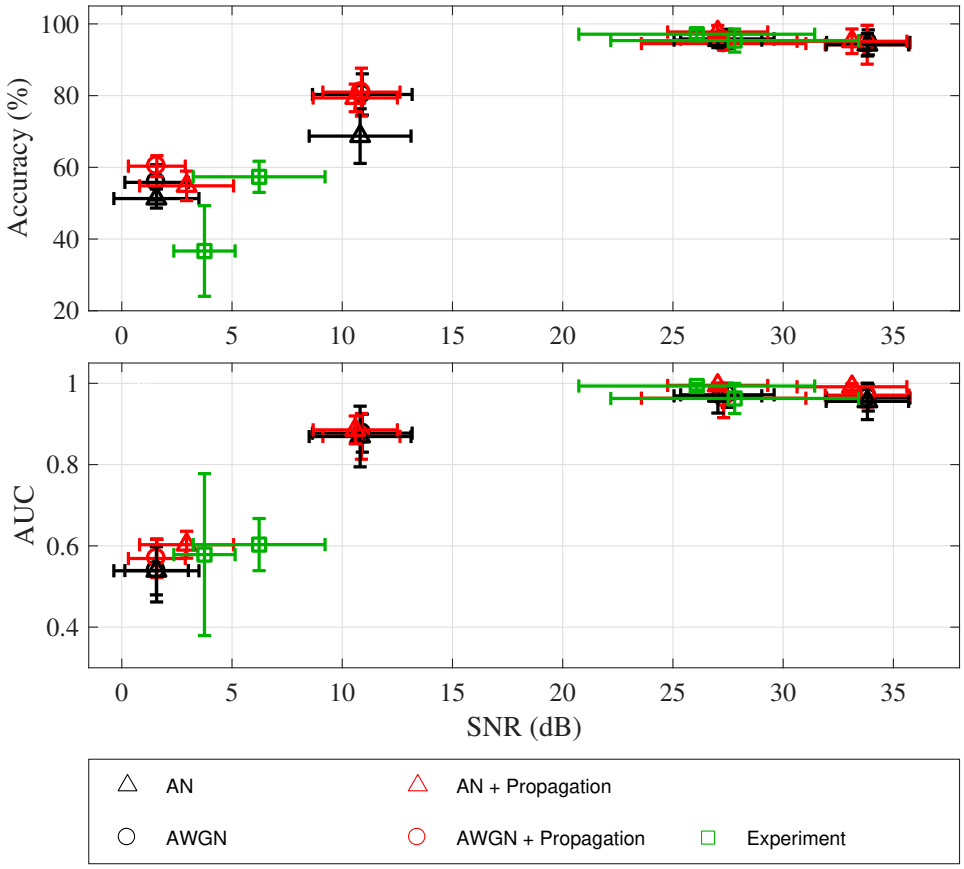


Figure 6.2: Comparison of classifier performance as a function of SNR for the biogenic whale calls in five different cases: (i) experimental noise (AN), (ii) simulated propagation effects, (iii) both experimental noise and simulated propagation effects, (iv) white Gaussian noise (AWGN), and (v) both AWGN and propagation effects added to the signals. Performance results from these simulated scenarios are also compared with results from experimental data.

became buried in the noise to the point where noise dominated and no unique signal information was discernible. At this point the classifier decision was essentially based on chance since there was no information available to differentiate between the signals. Surprisingly, when only propagation effects were applied to the signals, the classifier's performance did not show any significant change as signals were transmitted over longer ranges even though the signal distortion due to multipath effects increased with increasing range. For this environment, and over these ranges, the classifier was robust to signal distortion due to propagation effects. In the third case, in which the impacts of SNR and propagation were considered simultaneously, the performance decreased with increasing range/decreasing SNR. This seemed to be predominantly a result of decreasing SNR

since, in general, the performance results for cases 1 and 3 were within one standard deviation of each other. Also of note, is that the performance results generated from using either additive experimental noise or white noise were consistent with one another (i.e., within one standard deviation) when results are plotted in terms of range and SNR. The conclusion that AN and AWGN produce similar performance results is limited to the specific case examined here (the AN recorded in the 1 to 4 kHz band during the Gulf of Mexico experiment), and is not intended to be a general conclusion. One must reevaluate this result if applying these techniques to different circumstances, particularly for signals in a lower frequency band in which wind is not the dominant noise generation process. Nonetheless, based on the evidence that there was little difference in performance when AN or AWGN was used, it was concluded that little was gained by using additive experimental noise. Bear this in mind when additive noise is discussed in the context of increasing the range resolution of these results.

The SNR-dominated results (i.e., results for cases 1 and 3) were consistent with results obtained in Murphy and Hines [79] in which the authors noted a gradual approach to AUC = 1 at high SNR, an approximately linear relationship between AUC and SNR in the mid-performance range, and predicted a gradual approach to AUC = 0.5 at low SNR. A linear relationship between SNR and AUC suggests that $\mathrm{AUC} \propto \mathrm{SNR}$ (in units of dB). This is consistent with Shannon's theorem [80] that the capacity of a signal transmission channel, $\mathcal{C}$, in the presence of white noise is proportional to,

$$\begin{aligned} \mathcal{C} &\propto \log\left(\frac{S+W}{W}\right) \\ &\propto \mathrm{SNR} \quad , \end{aligned} \tag{6.4}$$

where S is the average signal power, W is the average white noise power, and recognizing that $\mathrm{SNR} = \log\left(\frac{S+W}{W}\right)$ [79, 80] and that the transmission channel here is the ocean medium. If this conclusion is correct, then for transmission of sound through the ocean medium, any deviation from this linear relationship would be attributed to non-SNR impacts — principally propagation effects.

Unfortunately, the range/SNR resolution of the experimental results and these model results simulating the experimental conditions was too coarse (i.e., there was only one data point in the mid-performance region) to be able to draw any conclusions on this account.

Hence, the range resolution must be increased through modelling. Although the simulated classifier results for signals that contained both noise and propagation effects were not in perfect agreement with the experimental data (values of performance metrics at the 6.5 km range were more than one standard deviation apart), the simulated results did capture the general trends in the data. Based on the sensitivity analysis results, this was likely due to using a range-independent SSP. Recall that a range-independent anti-duct SSP was selected for the environmental model — this was done to decrease the WATTCH model run time and because there was uncertainty about how the sound speed field during the experiments evolved over both time and space. Given the uncertainty in the measurements, the environmental model was considered to be sufficiently accurate to draw conclusions from the simulated data, and can be used to obtain higher resolution results.

6.3 Increased Range Resolution

The results of the previous section were useful for developing a preliminary understanding of how to separate the contributions of SNR and propagation-induced distortion; however, they were limited in their ability to describe the fine-scale features of how classifier performance degrades, since there were only a few data points. The intermediate ranges are of the most interest, since one might expect a linear relationship between SNR and performance if SNR drives the decrease in performance as signals are propagated over longer ranges. By comparing how, and at what range/SNR, the classifier performance degrades for the simulated signals from case 3 with the noise-only case, it will be possible to determine the relative importance of SNR and signal distortion resulting from time-spreading. To confirm this hypothesis, increased range/SNR resolution is required.

6.3.1 Methods

Using propagation modelling to increase the range resolution is straightforward — it is simply a matter of running the pulse propagation model with smaller range steps, thereby increasing the number of simulated signals. It is a simple matter to generate signals for the propagation-only case with increased range resolution; however, increasing the range resolution for cases 1 and 3 requires more attention. One must consider the mechanism for adding noise to the signals. When experimental results were simulated, it was natural to use samples of experimental noise; however, there was no clear extension of this when

increasing the range/SNR resolution. Although the recorders contained hours of ambient noise recordings, there was no convenient way to determine which portions to use since the noise characteristics changed over the course of the experiment. Additionally, the experimental noise was quite complex and could contain recordings of marine mammal vocalizations (one must be careful not to include these as they would likely contaminate the bowhead and humpback calls and confuse the classifier). Fortunately, the results presented in Figures 6.1 and 6.2 showed that there was no advantage to using experimental noise, as the performance results for signals with either additive experimental or white noise were in agreement (within one standard deviation). Given the concerns with using experimental noise and the similarity of performance results when using AWGN, white Gaussian noise was added to the signals with increased range resolution.

As in the previous section, the environmental measurements collected during the Gulf of Mexico sea trial motivated the definition of the environment model for WATTCH. Two environment models were considered here. The first was exactly the same as the environment model employed in the previous section (i.e., a sand/sand-mixture bottom, water depth increasing with range, and the anti-duct SSP). The second environment was the same except that the measured downward refracting SSP (refer to Figure 4.5) was used to characterize the properties of the ocean medium. The sensitivity analysis conducted in Section 5.3 showed that these two environment models produced the greatest differences in both the propagation fields and performance results. As such, by considering both environment models, we may be able to put bounds on the relative contributions of SNR and signal distortion to reducing classifier performance.

The incoherent TL curves for both the anti-duct and downward refracting environments were predicted using Bellhop — the resulting curves are depicted in Figure 6.3. The TL curves were transformed into SNR estimates by shifting them so that they matched the mean SNR of the signals transmitted over the 1 km range during the experiments. These curves were used to determine the ranges to which the signals should be transmitted, with the goal of decreasing the SNR by approximately 3 dB for each range step. The red circles (anti-duct) and blue squares (downward refracting) in Figure 6.3 indicate the range-SNR pairs selected for the simulations. The WATTCH model was then used to generate the acoustic response at these ranges for a recorder at 29 m depth. Noise was added at a level to match the modelled SNR (with $\sigma_{\mathrm{SNR}} \approx 1$ for each range set), for the noise-only, and

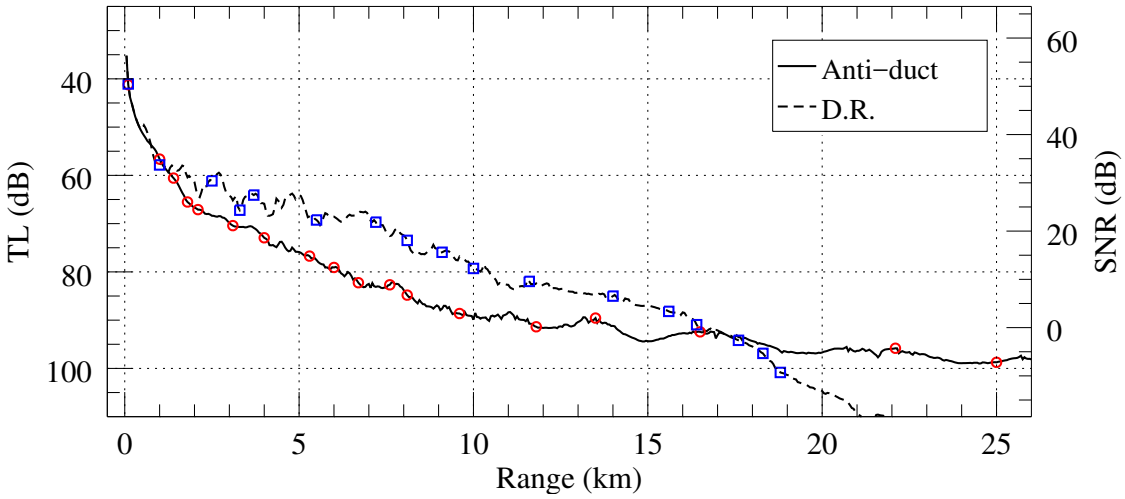


Figure 6.3: Incoherent transmission loss modelled at $f = 2\,\text{kHz}$ using Bellhop for environments defined by either an anti-duct or downward refracting (D.R.) sound speed profile. The modelled TL was used to estimate the SNR of signals propagated through the WATTCH model. Red circles and blue squares represent range-SNR pairs selected for adding noise to the propagated signals for the anti-duct and downward refracting cases, respectively.

propagation with noise cases. Note that when noise was added a different definition of SNR was employed. Here, SNR was calculated using,

$$\text{SNR}_{dB} = 10 \log \left(\frac{\sigma_s^2}{\sigma_n^2} \right) \quad , \tag{6.5}$$

where σ_s^2 is just the signal variance. This is in contrast with the previously employed definition of SNR (given in Equation 5.11), for which the numerator provided an estimate of the combined signal and noise variance. The new definition of SNR was necessary to produce the negative SNR values predicted by the Bellhop model.

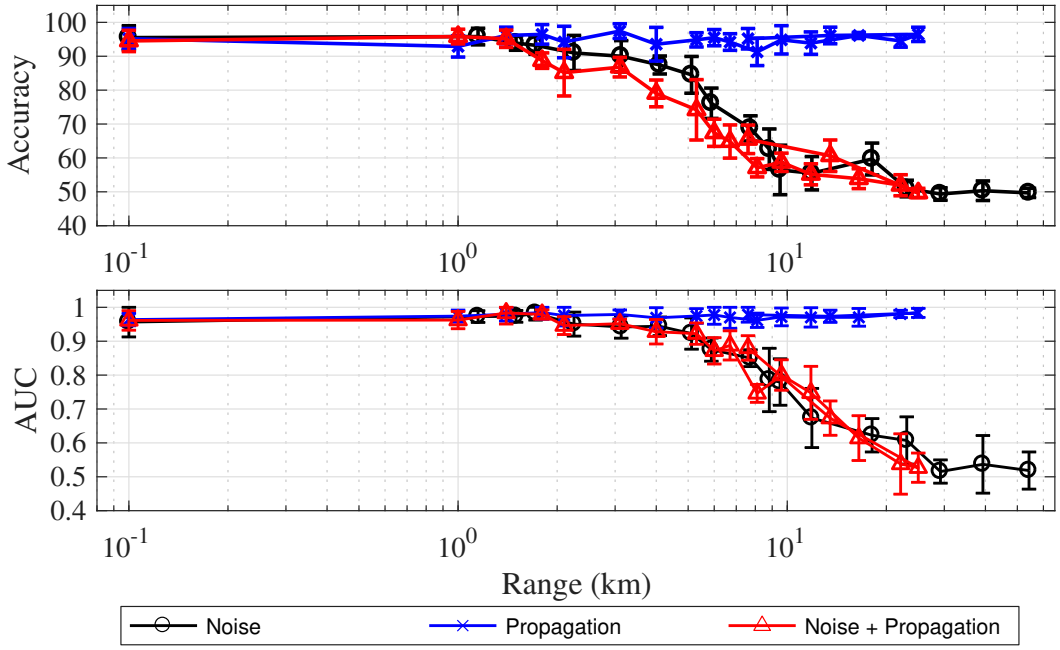
Previous results have shown performance in terms of both range and SNR. From the range plots it was easy to understand intuitively how performance changes as a function of transmission range. Plotting the performance results as a function of SNR qualitatively demonstrated that a large portion of the performance decrease could be explained by decreasing SNR. With increased range resolution it is possible to determine how much of this decrease is attributable to SNR; if SNR is the only contributor then one would expect a linear relationship between SNR and classifier performance (recall Equation 6.4) in the mid-performance region. Deviations from this linear relationship may be attributed to

signal distortion caused by time-spreading as signals propagate through the environment. To quantify this relationship, least-squares linear regression is used to fit a line to the data. The coefficient of determination, R^2, is then used to determine how well a linear equation describes the relationship between SNR and performance. If R^2 is $\sim$1 then one would conclude that SNR is the only significant contributor to decreasing classifier performance.

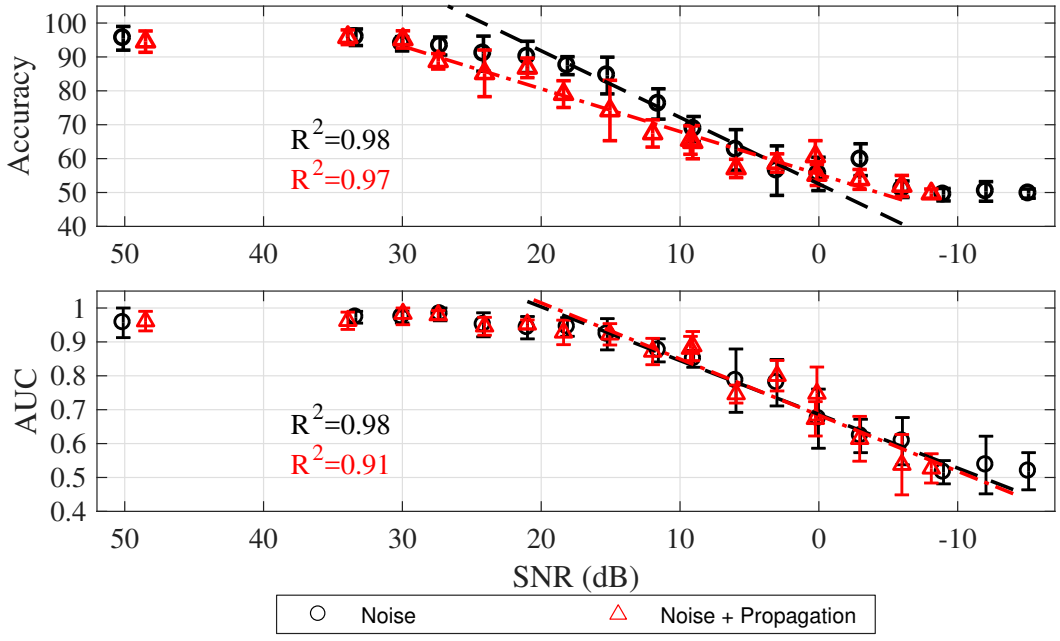
6.3.2 Results and Discussion

Performance results for the anti-duct environment are given in Figure 6.4a in terms of range and in Figure 6.4b in terms of SNR. First consider the results as a function of range; It is once again surprising to note that, even though multipath propagation increasingly distorted the signals with range, there was little change in performance for the propagation-only case. However, when noise was added to these signals, the performance decreased below that of the noise-only case. This indicates that with sufficient SNR, multipath propagation has minimal impact on classifier performance. Lowering the SNR not only decreased the performance, it also amplified the degradation due to signal distortion. This effect was not noted in previous results because the range resolution of the experiments was too coarse. The classification accuracy showed a greater difference between the noise-only case and the noise plus propagation case than was exhibited for the AUC. Recall that when the accuracy is lower than AUC this is indicative of the training data not being general enough, such that the accuracy is low; however, there is still sufficient information available to distinguish between the classes, resulting in acceptable or excellent discrimination. The simulation results, therefore, imply that the separation between class distributions is less sensitive to propagation effects than is the choice of decision threshold.

Now consider the performance results as a function of SNR (as plotted in Figure 6.4b). The noise-only case behaved exactly as predicted: at high values of SNR, near-perfect performance was attained, and at low SNR the threshold of random chance classification was attained. Between these extremes the noise-only case shows a linear relationship, in agreement with Shannon information theory [80]. Linear regression was performed on the data points for which either the accuracy was less than 90 % or $\mathrm{AUC} \leq 0.9$ and to the point at which approximately random classification was attained. Seven data points were included in the fit to the accuracy results and 16 were included in the fit to the AUC results for the noise-only case. The linear fit to both performance metrics resulted in an R^2 value of 0.98, indicating that the relationship between performance and SNR was



(a)



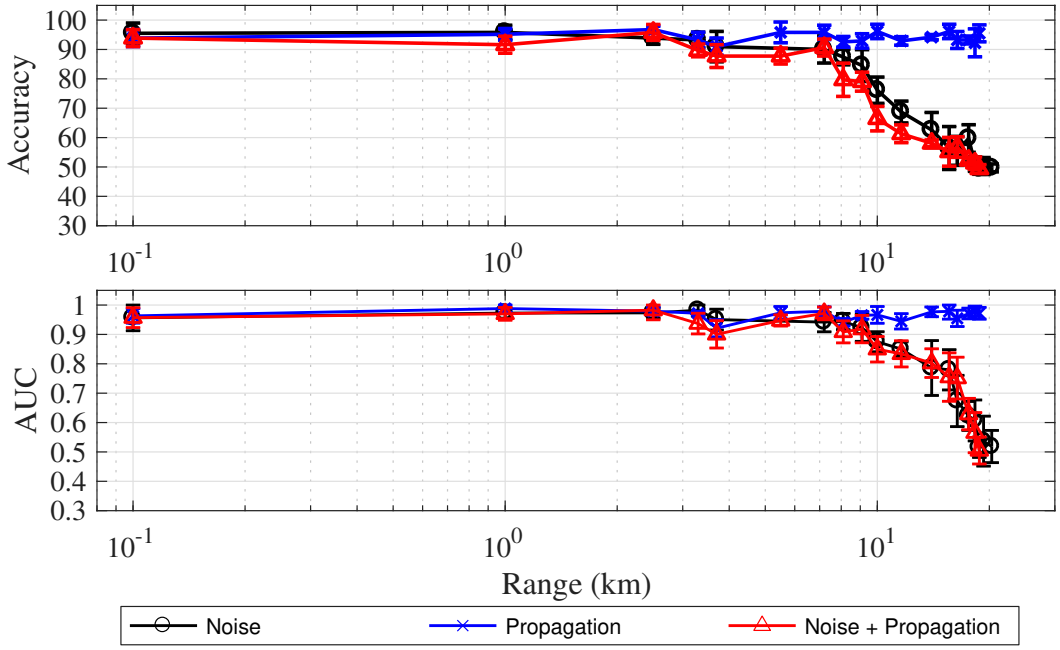
(b)

Figure 6.4: Comparison of classifier performance for the anti-duct environment as a function of (a) range and (b) SNR when either additive white Gaussian noise, or both white noise and propagation effects are added to the biogenic whale calls. Linear regression results are represented by the dashed lines in (b); the R^2 values corresponding to both regressions are also provided. Note that the SNR axis has been reversed, such that SNR decreases towards the right.

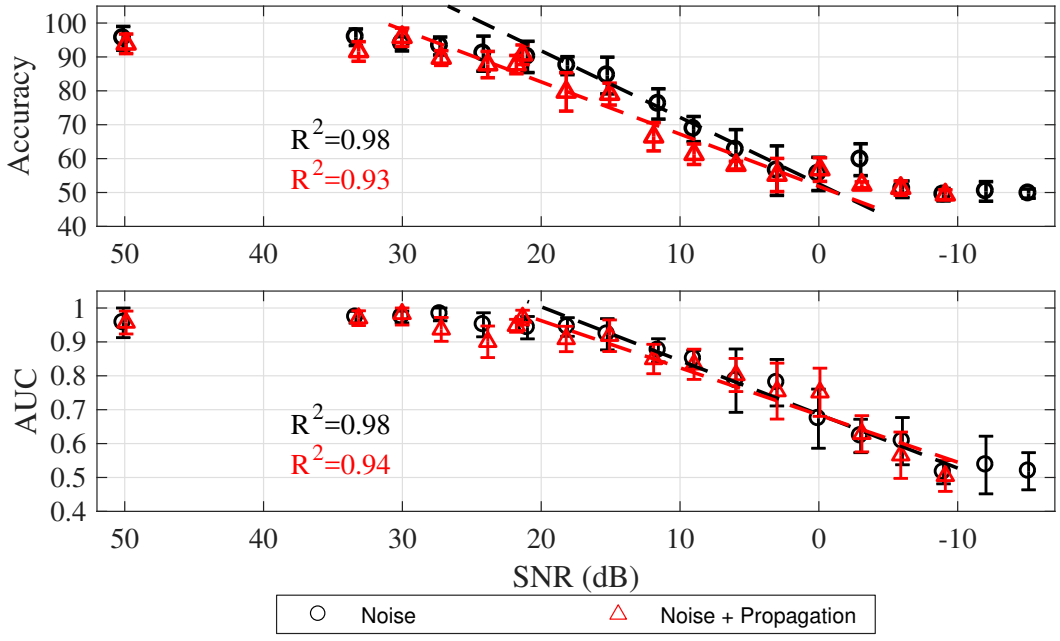
well explained by a linear model. When noise was added to the WATTCH-simulated signals the accuracy decreased at higher values of SNR than for the noise-only case; nonetheless, the relationship between accuracy and SNR was linear with a coefficient of determination ($R^2 = 0.97$) close to that of the noise-only case. While the AUC for the noise and propagation case qualitatively looked similar to the noise-only case (performance drops off at similar values of SNR, similar slope, and chance classification reached at similar SNR value), there was a greater difference between the R^2 values for the two cases than was noted for the accuracy. This suggested that although SNR accounted for the majority of the decrease in the AUC (91 %), a non-negligible portion may be attributed to signal distortion due to time-spreading.

Figure 6.5 contains the performance results for the downward refracting environment. In general, similar conclusions may be drawn from these data; that is, the performance of the classifier was not significantly impacted by modifying signals with increasing levels of propagation-induced distortion. The combination of noise and propagation caused the performance to decrease below the noise-only case — again this was more apparent for the accuracy metric. The most noticeable difference between the anti-duct and downward refracting environments was the range at which the performance began to decrease. Near-perfect performance was maintained for longer ranges in the downward refracting environment — performance started to decrease at about 8 km, compared to about 3 km for the anti-duct environment. This was likely a result of the acoustic propagation environment (refer to Figures 6.3 and 6.6), which caused the received signal level to drop off quicker for $R < 17$ km in the anti-duct environment. This is confirmed by the fact that the change in performance occurred at similar values of SNR for the two environments. Other differences between these environments can be seen in the performance plots as a function of SNR. Although the mid-performance region for the propagation with AWGN and noise-only cases were bounded by similar values of SNR in the downward refracting environment, there was a difference in the R^2 values (relative to the noise-only case) for both the accuracy and AUC.

Incoherent TL fields for both the anti-duct and downward refracting environments were modelled with Bellhop (see Figure 6.6) to examine differences in the acoustic propagation characteristics between environments that could explain performance differences. A relatively large portion of the acoustic energy travels along a narrow collection of ray paths

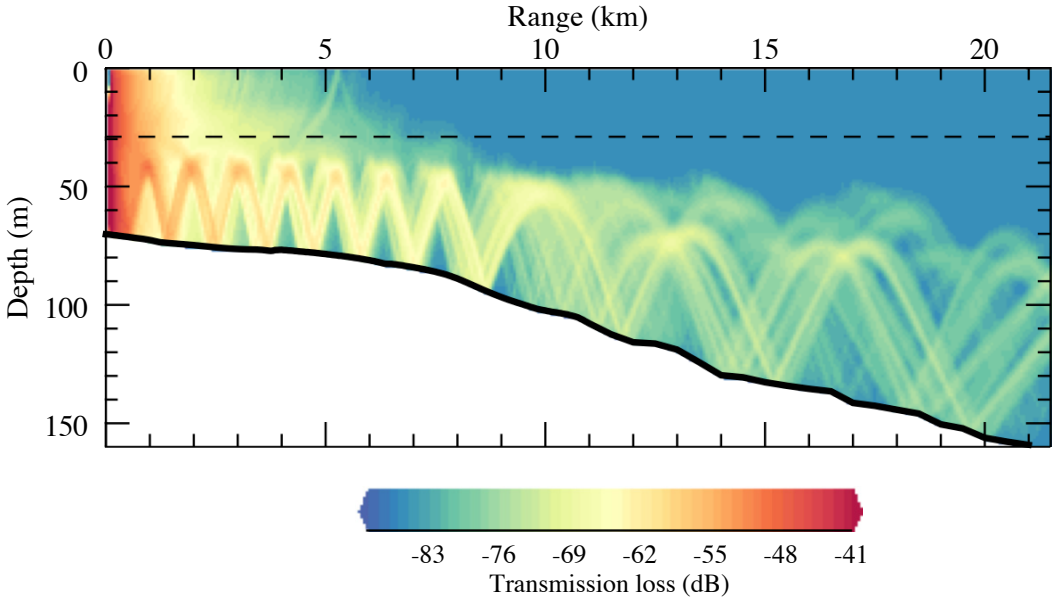


(a)

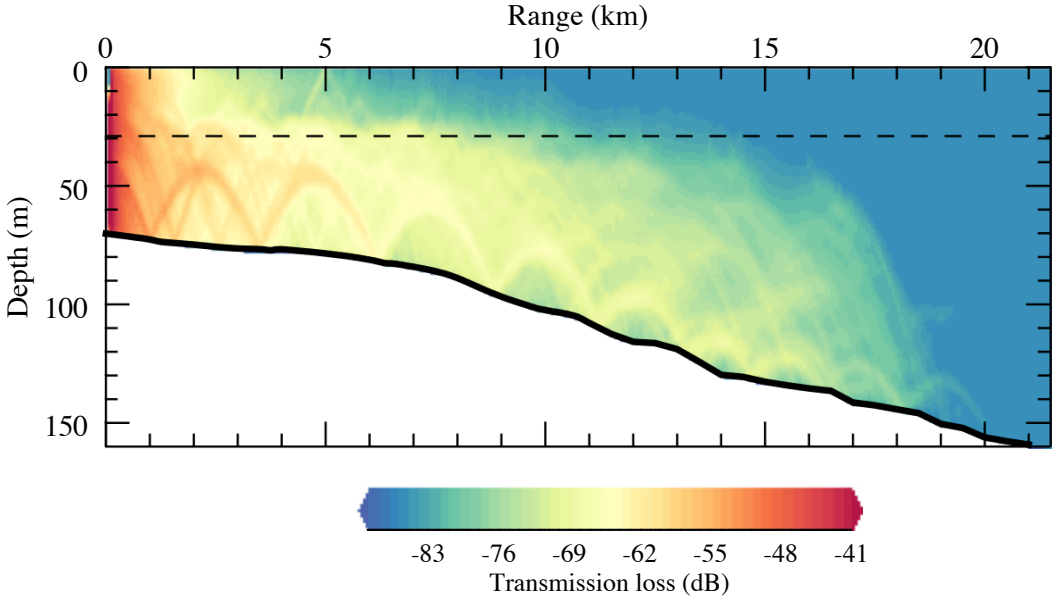


(b)

Figure 6.5: Comparison of classifier performance for the downward refracting environment as a function of (a) range and (b) SNR when either additive white Gaussian noise, or both white noise and propagation effects are added to the biogenic whale calls. Linear regression results are represented by the dashed lines in (b); the R^2 values corresponding to both regressions are also provided. Note that the SNR axis has been reversed, such that SNR decreases towards the right.



(a)



(b)

Figure 6.6: Incoherent transmission loss fields for the (a) anti-duct and (b) downward refracting ocean acoustic environments. The TL fields were modelled at the geometric mean of the signal band ($f = 2\,\text{kHz}$) using Bellhop. The horizontal dashed lines indicate the simulated receiver depth.

in the anti-duct environment. Conversely, the acoustic energy in the downward refracting environment exhibits a more diffuse nature. Consider how this may impact the distortion of received signals — in the anti-duct environment it is more likely that the received signal was composed of a limited collection of rays, especially at long ranges, since rays that strike the bottom at low grazing angles were likely to be refracted back towards the bottom and not reach the depth of the receiver. This is in contrast with the downward refracting environment, in which a received signal may be composed of rays with a broad range of ray path lengths, since it is more likely for both bottom- and surface-interacting rays to be included. Signals containing a broader assortment of ray path lengths — and therefore ray arrival times — will likely result in increased signal distortion due to time-spreading. This may explain why both the accuracy and AUC for the propagation with noise case showed a difference from the linear relationship of the noise-only case.

One positive outcome is that the results suggest that in these environments long range propagation does not inherently (or only minimally) degrade the classifier performance. That is to say, one could substantially enhance long range classifier performance simply by increasing SNR. This could be done with directional receivers such as towed arrays or vector sensors. Urick [10] shows that in the limit of a perfectly coherent signal and incoherent noise, the array gain for an m-element array is $10\log(m)$. For example, an 8-element array would provide a maximum array gain of 9 dB. In the anti-duct environment, an 8-element array may have increased the classification accuracy by as much as 10 % and the AUC by 0.14 at the 12 km range, assuming that SNR was the dominant contributor to decreasing classifier performance.

6.4 Chapter Summary

In this chapter the relative impacts of SNR and propagation-induced distortion on the aural classifier performance were investigated. This investigation was accomplished by considering the impacts of propagation and noise independently, and also the combination of these two effects. The WATTCH simulations proved to be very useful for disentangling the relative importance of distortion and SNR to reduced classifier performance. For the two environments considered, classifier performance was not significantly degraded by signal distortion caused by time-spreading; however, when noise was added to these signals the performance became less than that of the noise-only case. Furthermore, the effects of

propagation can cause performance to drop off at shorter transmission ranges than if SNR was the only contributor to reducing performance.

By employing Shannon information theory — which states that the maximum amount of information contained in a transmission channel is proportional to the SNR — it was found that for these simulations the majority (91 to 94 % $\pm$ 2 %) of the decrease in performance as signals were transmitted over longer ranges may be explained by decreasing SNR. This leaves a relatively small fraction (6 to 9 % $\pm$ 2 %) that may be attributed to propagation-induced distortion. Note that these results were specific to the Gulf of Mexico environment. Further work needs to be done to determine if there are realistic environments in which propagation effects have a greater impact, or if the bowhead and humpback calls in the lower (true) frequency band are more affected because they travel over greater distances with less attenuation. An initial analysis is conducted in Chapter 7 to address this issue, by propagating the lower frequency bowhead and humpback vocalizations through two simple acoustic environments.

CHAPTER 7

IMPACT OF ENVIRONMENT-DEPENDENT PROPAGATION ON AUTOMATIC RECOGNITION OF BOWHEAD AND HUMPBACK VOCALIZATIONS

Now that many aspects of the experimental data have been addressed, it is appropriate to extend the investigation of the impact of environment-dependent acoustic propagation on the PAM of cetaceans to other scenarios. Up to this point, results presented in this thesis have been primarily derived from the upsampled biogenic bowhead and humpback calls. In addition, only environmental properties inspired by the Gulf of Mexico trial location were considered. This was useful for understanding the experimental results; nonetheless we wish to make conclusions about the impact of environment-dependent acoustic propagation on PAM of cetaceans in general. To generalize the findings one must consider the bowhead and humpback whale calls in their true frequency band and consider other types of propagation conditions. Collecting data at sea in many different environments with low-frequency baleen whale calls is not feasible given the time-consuming and costly nature of performing range-dependent propagation experiments. These issues are compounded when signals are transmitted in a lower frequency band because the signals do not attenuate as rapidly with range, thereby increasing the range over which signals must be transmitted. Additionally, the experiments would require more time at each range since signals would not need to be time compressed as they were for the Gulf of Mexico

experiments. Consequently, in this penultimate chapter, the analysis is extended to include low frequency whale calls propagated through different environments, by applying the pulse propagation simulation techniques developed in the previous chapters.

To this end, WATTCH was used to propagate bowhead and humpback whale vocalizations in their original frequency band through two simplified, but realistic, ocean acoustic propagation environments. There is one caveat, however — the frequency band of the vocalizations was limited to 200 to 800 Hz Restricting the frequency band was deemed necessary due to the limitations of ray-tracing theory; consider that for an environment with a 1500 m/s isospeed profile and a water depth of 100 m, Equation 5.1 shows that the application of a ray-tracing model becomes questionable below 150 Hz. Since the full-bandwidth vocalizations include frequencies below the nominal 150 Hz limit (in fact, as discussed in Section 3.1, they contain considerable energy down to approximately 50 Hz) the suitability of employing Bellhop for modelling may be called into question. By removing the lower frequencies from consideration, any issues with the applicability of a ray-tracing model were avoided. This was considered a reasonable approximation since it was shown in Section 3.3 that limiting the frequency band did not result in a large difference in the classifier performance.

7.1 Environment and Model Configurations

The WATTCH model was used to simulate the received time series by running the Bellhop model for 61 frequencies equally spaced in the 200 to 800 Hz band. These results were interpolated to provide a 0.5 Hz frequency spacing and 125 µs temporal spacing and the acoustic response was generated in a 3.85 s long window. A time window longer than that which was used to simulate results in the 1 to 4 kHz frequency band was necessary because the lower frequency signals are longer duration and it was anticipated that there would be stronger multipath arrivals with a broader range of arrival times. A longer time window was therefore necessary to eliminate potential wrap-around of the later arrivals into the beginning of the time window. The channel impulse response was determined for a source placed at mid-water column and for 35 ranges, in range steps of 100 m for the first kilometre, and then for 2 km range steps (starting at 2 km).

Two range-independent acoustic propagation environments were considered, that were

designed to be at (realistic) extremes. Schematics of the two environments are provided in Figure 7.1. The first environment was designed to be relatively benign in terms of acoustic propagation effects. It was defined by a sandy bottom half-space and an isospeed SSP ($c = 1500$ m/s) in the water column. This is a typical environment that one may encounter on a sandy continental shelf, with a well-mixed water column. The second environment was designed to have acute propagation effects (i.e., many boundary interactions and relatively small bottom loss); the water column properties were characterized by a downward refracting SSP, and a sedimentary rock bottom half-space was selected. This environment was inspired by measurements from the Sur Ridge [22]. In both environments a constant water depth of 100 m was employed. The source depth was set at $z_s = 20$ m based on literature reports of the vocalization depths of humpback whales [112, 113].

The incoherent transmission loss fields modelled using Bellhop for these two environments are compared in Figure 7.2. These figures demonstrate the contrast between the environments. In the isospeed environment the acoustic energy decayed with range as one

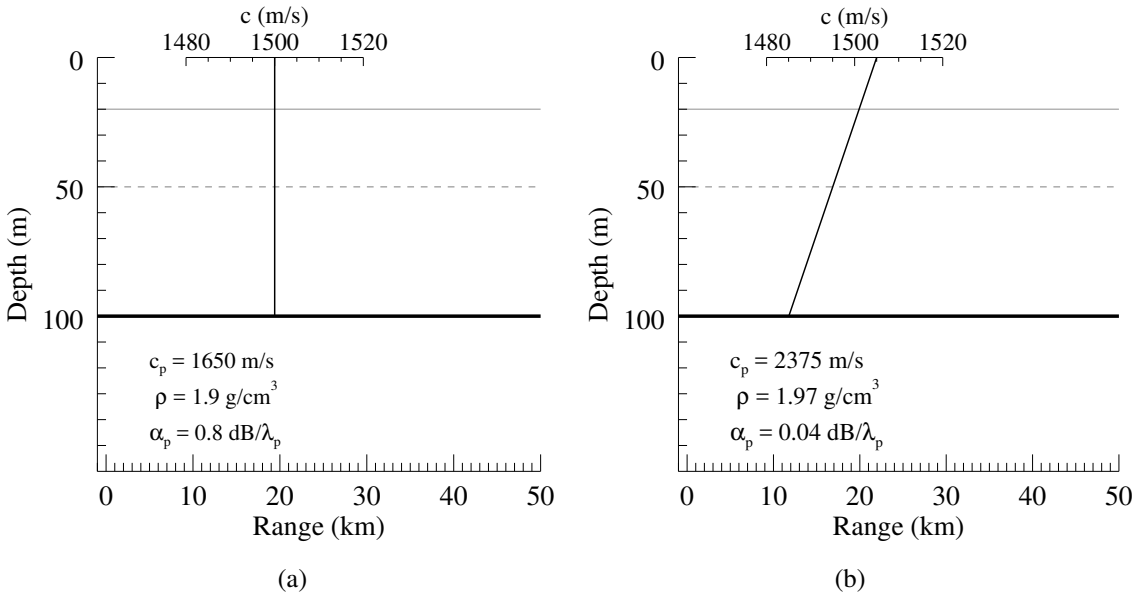
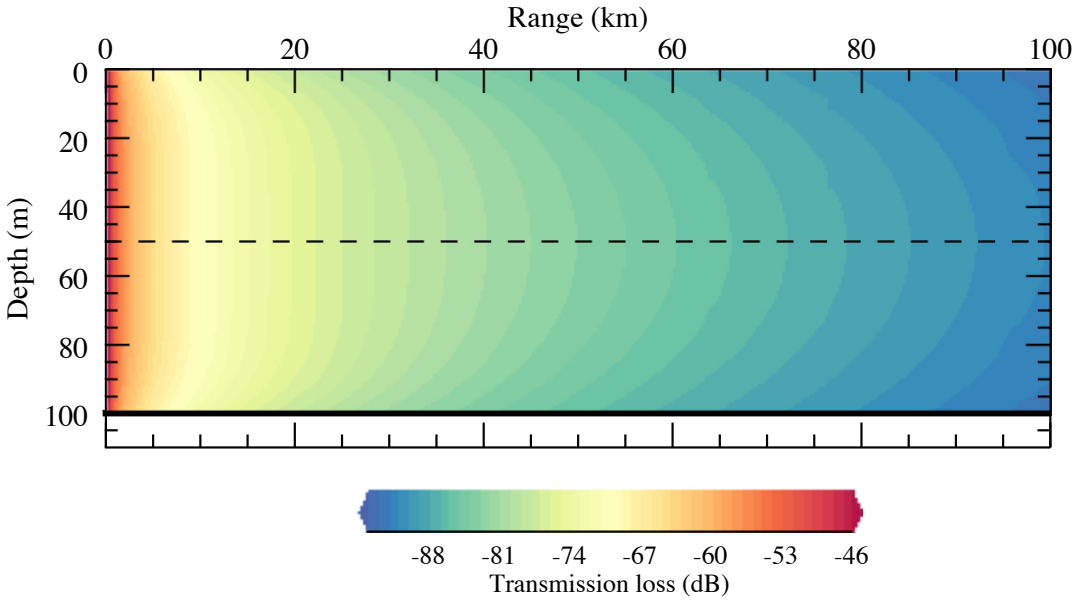
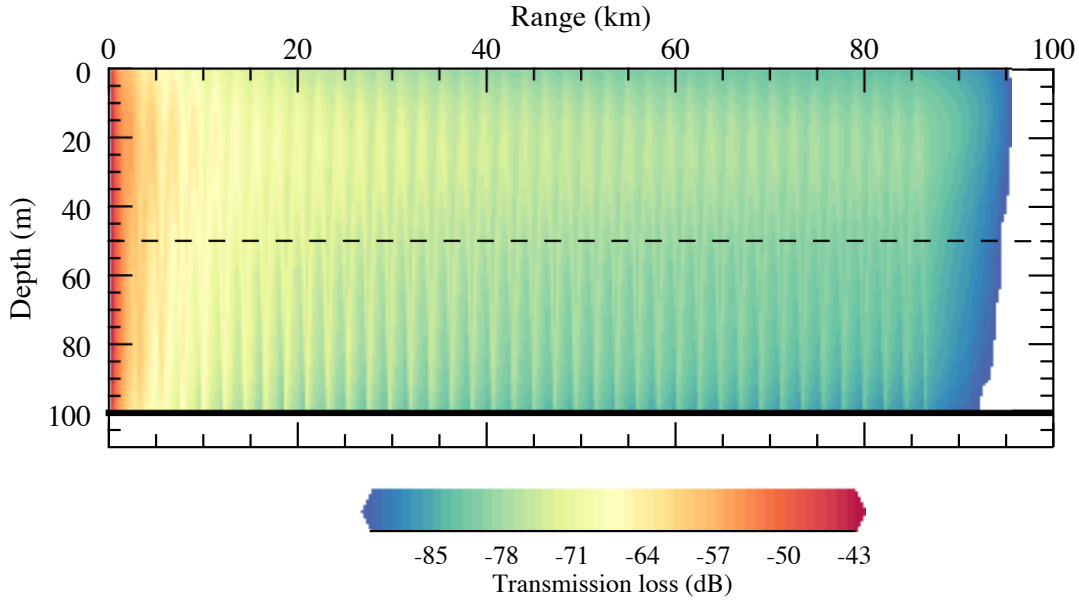


Figure 7.1: Schematics of environment configurations used to simulate propagation of the bowhead and humpback calls in their true frequency band. Two cases were considered: (a) a moderate propagation environment with an isospeed SSP and sandy bottom, and (b) an extreme propagation environment with a downward refracting SSP and sedimentary rock bottom. The solid grey line indicates the source depth, and the dashed grey line indicates the receiver depth.



(a)



(b)

Figure 7.2: Incoherent transmission loss fields for the (a) isospeed and (b) downward refracting ocean acoustic environments. The TL fields were modelled at the geometric mean of the signal band ($f = 400\,\text{Hz}$) using Bellhop. The horizontal dashed line in each figure indicates the receiver depth.

would expect based on a hybrid spherical/cylindrical spreading scenario — the acoustic energy filled the the water column, and there were no obvious preferred ray paths. In contrast, the transmission loss field for the downward refracting environment showed a clear range-dependent interference pattern. Such differences in the transmission loss field were indicative of a disparity in how a signal was distorted as it propagated through each of the environments.

7.1.1 Noise Addition

A strategy for adding noise to the bowhead and humpback whale calls similar to that presented in Section 5.2.2 was employed for the low frequency bowhead and humpback calls, with two notable changes — both the way in which the noise samples were processed and selected, and the transformation from modelled TL to an SNR estimate were different. Subtle changes to the noise addition method were necessary due to the shift in the vocalization frequency band and the removal of the constraint to match source level to experimental results (now the source level should match that of real whale calls).

Experimentally recorded ambient noise was added to the simulated signals rather than the white Gaussian noise used in Chapter 6, to ensure a realistic colouring of the noise spectrum. White noise was an acceptable approximation for the higher frequency vocalizations since the noise spectrum was approximately flat in that frequency band; however, at low frequencies this flat noise spectrum approximation is less applicable. Fortunately, ambient noise recordings were available from the Gulf of Mexico experiments that were sampled during times when no transmissions occurred. An hour-long period of ambient noise recorded by the icListen hydrophone on the first day of the experiment was used as the source of additive noise. The ambient noise was bandpass filtered to include frequency content in only the 200 to 800 Hz band. It was then divided into 423 segments, each 8.5 s long. A random noise snippet was selected to be added to each simulated signal, such that a noise snippet was only used once for signals propagated over the same range.

Estimating the desired SNR is straightforward. Using the passive sonar equation [75], an estimate of the range-dependent SNR, $\mathrm{SNR}(R)$, of the simulated signal may be obtained by,

$$\mathrm{SNR}(R) = \mathrm{SL} - \mathrm{TL}(R) - \mathrm{NL} \quad , \tag{7.1}$$

where $\mathrm{TL}(R)$ is the range-dependent transmission loss modelled by Bellhop, and NL

is an estimate of the ambient noise level. A constant SL of 175 dB re 1 µPa @ 1 m is assumed based on Frazer's [113] report of humpback whale source levels in the 175 to 188 dB re 1 µPa @ 1 m range, and Au *et al.*'s [112] measurement of maximum source levels of humpback song units in the 151 to 173 dB re 1 µPa @ 1 m range. For both the isospeed and downward refracting environments an ambient noise level of 89.8 dB re 1 µPa for the 200 to 800 Hz frequency band is assumed. This noise level is based on Piggott's [65] report of average noise spectrum levels on a continental shelf, which was in agreement with the Merklinger-Stockhausen noise model [114]. Other than the method used to estimate the SNR, all aspects of adding noise to the simulated signals to achieve the correct SNR is the same as described in Section 5.2.2.

Application of Equation 7.1 produced the TL/SNR relationship shown in Figure 7.3, in which the TL curves for the isospeed and downward refracting environments are generated for a receiver at $z_r = 50$ m. The difference between the two environments is evident upon comparison of the curves. In the isospeed environment, the TL monotonically decayed as a function of range, whereas the TL curve for the downward refracting environment exhibited many fluctuations as a result of multipath. At the longer ranges ($R > 20$ km) there was a greater discrepancy between the two curves which was likely a result of the

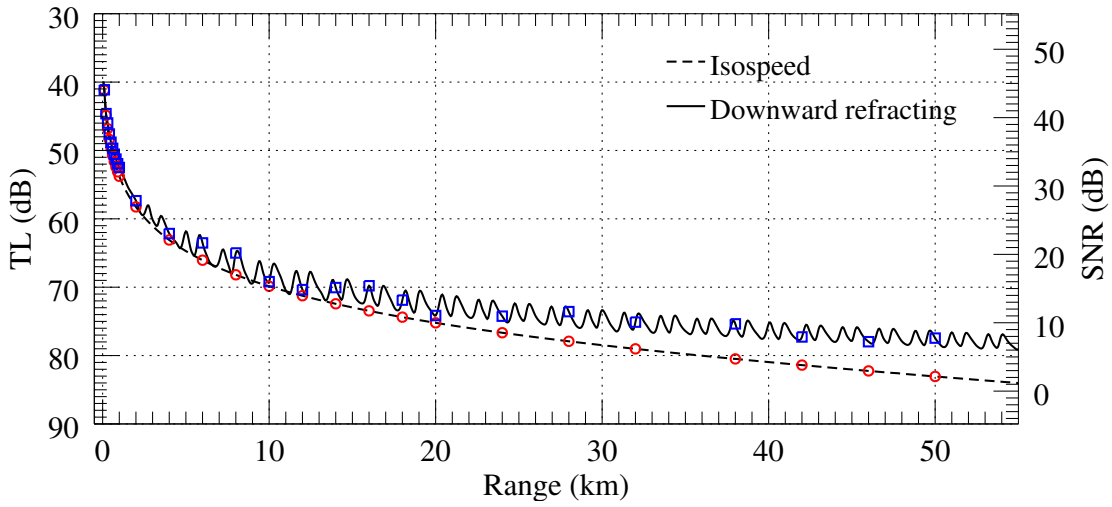


Figure 7.3: Incoherent transmission loss modelled at $f = 400$ Hz using Bellhop for environments defined by either an isospeed or downward refracting sound speed profile. The modelled TL was used to estimate the SNR of signals propagated through the WATTCH model. Red circles and blue squares represent range-SNR pairs selected for adding noise to the propagated signals for the isospeed and downward refracting cases, respectively.

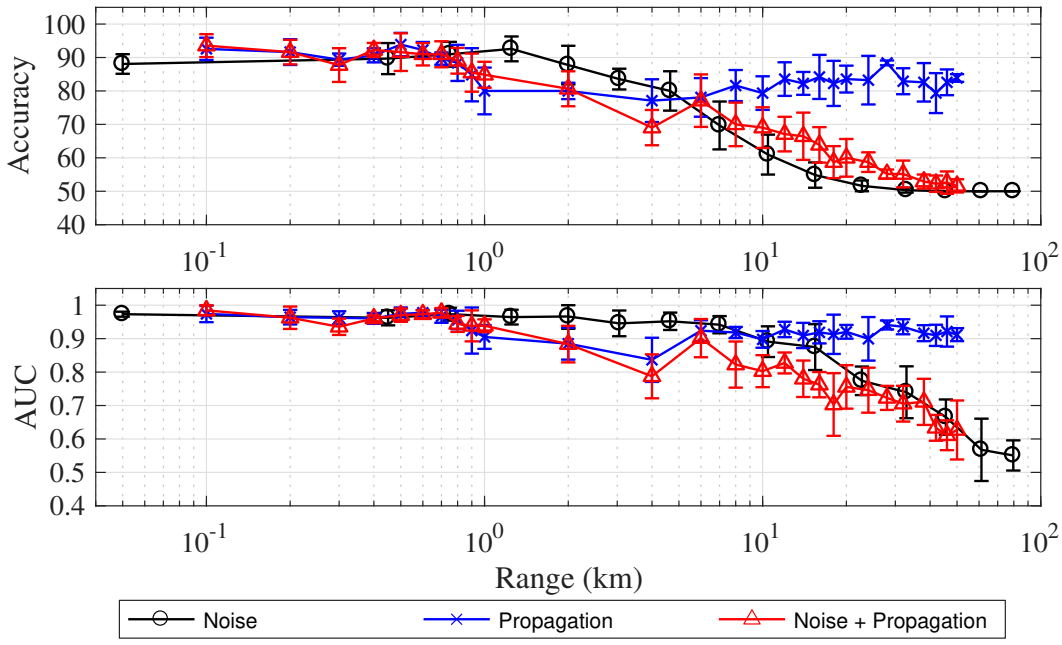
larger bottom loss in the isospeed environment. The range-SNR pairs selected for the simulations are indicated on the TL/SNR curves. It is worth noting that the oscillatory nature of the downward refracting curve was captured by the choice of the simulated ranges; that is, the selected points included (approximately) local maxima, minima, and points in between. Since the variations in the TL curve may be symptomatic of differences in propagation-induced distortion, incorporating simulated signals that lie within varying points of these oscillations should capture differences in the signal distortion. One might then expect to observe greater differences in performance as a result of propagation-induced distortion.

7.2 Results and Discussion

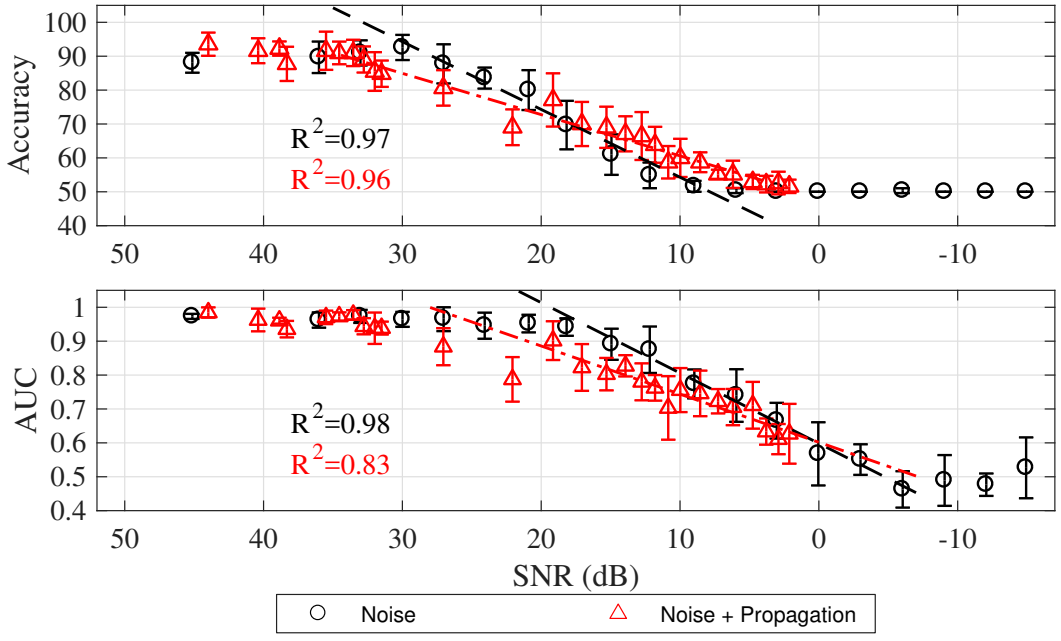
7.2.1 Isospeed Environment

As in Chapter 6, the impacts on classifier performance were considered for SNR effects and propagation-induced distortion independently, and in combination. Results are presented as a function of range in Figure 7.4a for the isospeed environment. The classifier was trained on signals either propagated over the shortest range ($R = 100\,\mathrm{m}$), or with the lowest level of additive noise (SNR = 45 dB). The classifier performance for signals with additive noise behaved as expected based on previous results; i.e., the classification accuracy and AUC decayed smoothly as a function of propagation range. There was a notable deviation from previous results for the WATTCH-simulated signals with no noise added — in this environment there was a small change in both the accuracy and AUC. The classifier performance decreased such that the mean accuracy changed by approximately 10 %, and an approximately 0.13 decrease in the mean AUC was noted. These were small, but significant changes, that occurred for signals propagated between 0.7 to 4 km, after which the performance values remained relatively constant as signals were propagated over longer ranges.

As previously noted, the AUC values for simulated signals to which noise was added, were less than those for the noise-only case (for $9\,\mathrm{km} \leq R \leq 18\,\mathrm{km}$); however, a similar trend was not noted for the performance in terms of accuracy. Instead, for signals propagated over ranges of 10 to 42 km the accuracy was actually greater than for the noise-only case. This is interesting, especially given that the AUC did not mirror this result. Unlike



(a)



(b)

Figure 7.4: Comparison of classifier performance of the whale calls in their true frequency band for the isospeed environment. Performance results are presented as a function of (a) range and (b) SNR when either additive noise, or both noise and propagation effects are added to the biogenic whale calls (in (a) propagation effects are also independently considered). Linear regression results are represented by the dashed lines in (b); the R^2 values corresponding to both regressions are also provided. Note that the SNR axis has been reversed, such that SNR decreases towards the right.

previous results derived from the upsampled biogenic whale calls propagated through environments inspired by the Gulf of Mexico experiment, here we see that the impact of propagation could *enhance* classification accuracy relative to the noise-only case. This suggests that for this scenario, even though there was increasing overlap between class distributions (decreasing AUC) the choice in decision threshold was relatively insensitive to the impacts of distortion for signals transmitted over ranges greater than 10 km.

Consider the performance results in terms of the SNR, as presented in Figure 7.4b. For both performance metrics, for signals to which only noise was added, performance decreased monotonically with decreasing SNR, with a linear relation between the performance extremes ($R^2 = 0.97$ and 0.98 for the accuracy and AUC, respectively), as predicted by Shannon information theory. When propagation impacts were included, the performance began to decrease at higher SNR values. A similar criterion as for the noise-only case was used to select which performance-SNR points to include in the linear regression: data points were included for which the accuracy $\leq 90\,\%$ or $\mathrm{AUC} \leq 0.9$, and for which the minimum expected performance was not reached. This resulted in 19 data points being included in the regression analysis for the accuracy, and 13 points for the AUC. The resulting R^2 values for accuracy and AUC ($R^2 = 0.96$ and 0.75, respectively) showed that a linear model was a good fit for the accuracy metric, but not for the AUC. One may think that including higher SNR data points than for the noise-only case may bias this result; however, if the data selected for the linear regression are limited to an equivalent SNR range, R^2 values of 0.86 and 0.83 result. This outcome does not affect the conclusion that bowhead and humpback whale vocalizations propagated through an isospeed environment were impacted, not only by the effects of decreasing SNR, but also by propagation-induced signal distortion. In conclusion, bowhead and humpback vocalizations were sufficiently distorted as they were propagated through this isospeed environment that between 4 to 14 % of the decrease in classification accuracy, and 17 to 25 % of the decrease in AUC, was explained by time-spreading caused by multipath propagation. This was a surprising result, considering that the acoustic propagation environment was designed to have relatively benign propagation characteristics.

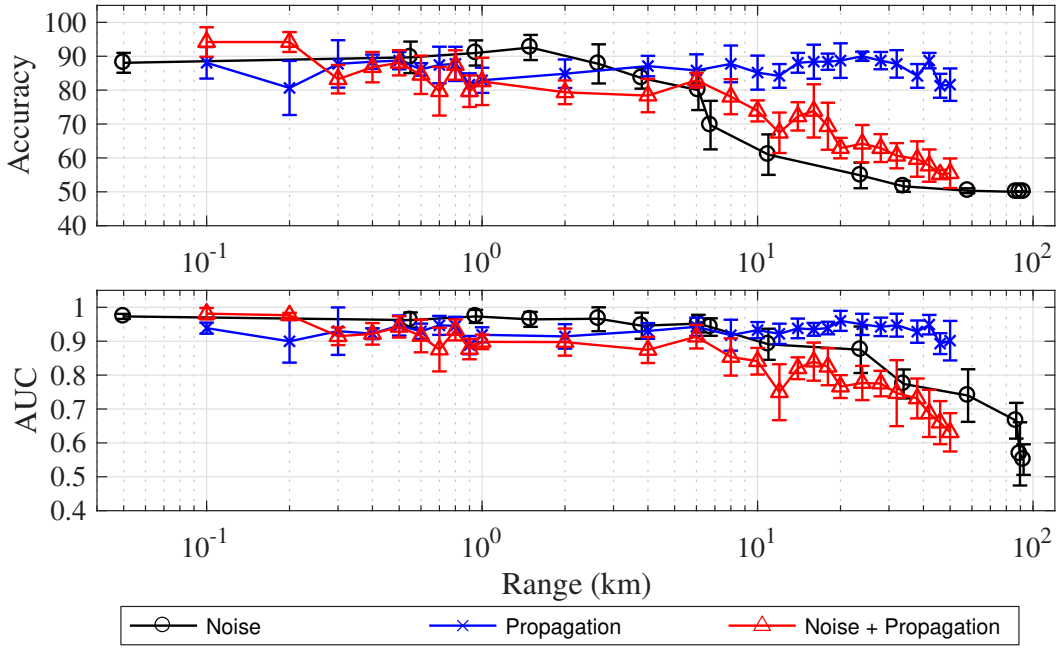
7.2.2 Downward Refracting Environment

Now let us consider the performance results for an environment for which propagation effects are expected to be acute — the downward refracting environment with a hard bottom.

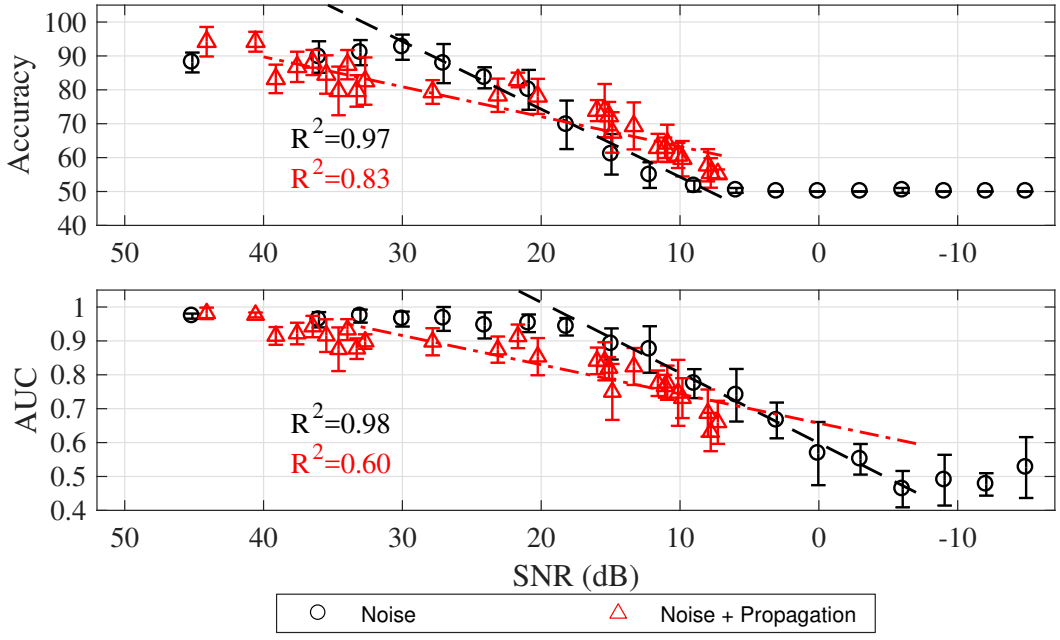
Figure 7.5a shows performance results for signals propagated through the downward refracting environment as a function of range. A surprising result is that there was little change in the performance for the scenario in which the signals were transformed by only propagation effects. A larger impact on the propagation-only case might have been expected based on the relatively extreme nature of this propagation environment. As with the isospeed case, the addition of noise to the signals simulated by WATTCH, caused the AUC to decrease below the noise-only case; however, the accuracy metric did not mirror this result, but instead the accuracy was consistently greater (when $R \geq 8$ km) for this case. Again, the choice of decision threshold was quite robust for this environment.

To determine the relative proportions of changes in performance that may be explained by SNR versus propagation-induced signal distortion, the above results are re-plotted as a function of SNR in Figure 7.5b. These results show that the performance begins to decay at higher values of SNR for the simulated signals with additive noise than for the signals to which only noise was added. As with the isospeed environment, at high SNR, the classification accuracy for signals distorted by multipath addition is less than for signals with just additive noise, but this is reversed for $\mathrm{SNR} \leq 22$ dB. Surprisingly, at low values of SNR, the addition of propagation effects leads to an increase in classification accuracy, relative to the performance for the additive noise signals. The AUC did not follow this trend — instead the AUC for the noise and propagation case was consistently less than that of the noise-only case.

One noteworthy feature of the AUC graph is the vertical scatter of data points with $\mathrm{SNR} \approx 15$ dB. These data points all have similar mean SNR values, but produced a spread of AUC values indicative of the role played by propagation-induced distortion for altering classifier performance. To quantify this role, both performance metrics were fit to a linear model, using the same criteria as outlined for the isospeed environment in the previous section. The linear regression results are also plotted in Figure 7.5b. Not only was there an obvious difference in the slopes of the regression lines between the two cases for both metrics, there was also a relatively large difference between the values of R^2. A linear model only described 83 % of the variance for the accuracy and 60 % of the variance for the AUC. The values of R^2 increase to 0.86 and 0.76 for the accuracy and AUC, respectively, if the regression is performed on data points in the same SNR range as was used for the noise-only case; nonetheless, there was a significant impact of propagation-induced disto-



(a)



(b)

Figure 7.5: Comparison of classifier performance of the whale calls in their true frequency band for the downward refracting environment. Performance results are presented as a function of (a) range and (b) SNR when either additive noise, or both noise and propagation effects are added to the biogenic whale calls (propagation effects are also independently considered in (a)). Linear regression results are represented by the dashed lines in (b); the R^2 values corresponding to both regressions are also provided. Note that the SNR axis has been reversed, such that SNR decreases towards the right.

rtion observed in this environment that could not be explained solely by changes in SNR.

Even though it seemed, once again, that the relationship between SNR and performance was the only significant contributor to decreasing classifier performance, upon closer inspection it was clear that propagation-induced distortion was also a significant contributor. In this downward refracting environment the results indicated that many of the large scale features may be explained by signal attenuation, and the impact of propagation-induced distortion contributed changes to the performance on a finer scale. This was consistent with the TL field depicted in Figure 7.2b — the 'background' TL smoothly decayed as a function of range, much like in the isospeed environment. Superimposed on top of this, however, were oscillations in received level as a function of range that resulted from interference of ray paths. These small scale variations could make it difficult to predict performance if one were only to assume a linear dependence between SNR and noise, thereby discounting the impacts of propagation-induced distortion. If signals were detected at close range, and/or with high SNR, propagation would be unlikely to have any significant impact on classifier performance. On the other hand, if detected signals have intermediate SNR (i.e., they fall within the mid-performance region) there could be an impact on performance due to multipath addition that would be ignored if one assumed a linear relationship between the performance metrics and SNR. Additionally, as a result of signal distortion, the performance begins to decrease at SNR levels for which one would likely assume there was still sufficient SNR to make excellent classification decisions. Realistically, this would have the result of biasing an estimate of the number of vocalizing whales, relative to what one may expect if SNR was the only significant contributor to decreasing classifier performance; furthermore, it would also result in an underestimate of the uncertainty associated with the abundance estimate.

Consider, for instance, the impact on density estimation; Marques *et al.* [29, 115] lay out a framework that has become the standard in the PAM community. A critical step in obtaining an accurate density estimate is determining the detection function. Detection functions are used to map the SNR of a detected vocalization to the probability of detection and correct classification. Many researchers use a simple geometric spreading assumption to approximate this function, which in some environments may not be appropriate for capturing the correct relationship between SNR and detection range. Other researchers, like Küssel *et al.* [30], incorporate acoustic propagation modelling to characterize the

relationship between SNR and the PAM system's performance. Even though application of an acoustic propagation model is necessary for characterizing the relationship between SNR and the detection function, it overlooks a potentially important factor: propagation-induced distortion. As shown through these simulations, propagation-induced signal distortion does impact the performance of automated PAM methods. Therefore, in order to obtain the best possible estimate of the detection function, one should also incorporate the impacts of multipath propagation. This in turn will provide a more accurate density estimate and a better understanding of the associated uncertainty.

7.3 Chapter Summary

In this chapter we have extended the analysis of environment-dependent acoustic propagation effects on the automated classifier performance to consider bowhead and humpback whale vocalizations in their true frequency band. The vocalizations in this lower frequency band were propagated through two simplified, but realistic, environments using WATTCH. On the surface, the performance results were similar to those observed for previous scenarios — the performance decreased with increasing range and decreasing SNR. In spite of this general similarity, there were some unexpected results. The classification accuracy for signals transformed by both propagation and noise was greater than for the noise-only case at long range/low SNR in the isospeed and downward refracting environments. A surprising result was that the propagation-only case showed a decrease in performance with increasing propagation range for the isospeed environment but not for the downward refracting environment. Intuitively one would expect that multipath addition which included ray paths with strong bottom returns would enhance distortion, thereby having a greater impact on performance. However, it was shown that what appeared to be a relatively benign environment can indeed have an impact on performance. Even though the propagation-only case for the downward refracting environment did not show a significant change in performance with increased time-spreading due to multipath addition, it was found that propagation-induced distortion played a relatively large role in determining how the classifier performed on the simulated signals with additive noise. Specifically, in the downward refracting environment it was found that between 60 to 83 % of the change in the mid-performance region was explained by changes in SNR, implying that 17 to 40 % of the performance change was caused by propagation-induced distortion. This is a

non-trivial proportion. Thus, one may conclude that it is necessary to consider the ocean acoustic propagation characteristics for the frequency band and environment of interest in order to determine to what degree multipath addition will impact automated classification of cetacean vocalizations.

CHAPTER 8

CONCLUSIONS

Through the use of an at-sea experiment and simulations, the results presented in this thesis have aided in understanding the impacts of environment-dependent acoustic propagation on passive acoustic monitoring of cetaceans. To ground results in reality, an experiment was designed and conducted in the Gulf of Mexico during the spring of 2013. During the experiment, biogenic and synthetic whale calls were transmitted over ranges of approximately 1 to 10 km. In general, it was found that classifier performance was range-dependent, such that both the classification accuracy and AUC decreased with increasing transmission range. Through the use of performance matrices generated by training and validating the aural classifier on signals transmitted over each range pair, it was found that the best classification performance across a wide variety of transmission ranges was obtained when the classifier was trained on an equally varied assortment of ranges. If a large and varied data set is not available for training the classifier, then it is best to train it on relatively low SNR calls. While the motivation for developing synthetic whale calls was well-founded (the desire to examine signals that had not been previously subjected to propagation effects) it was found that they provided little added value since results were similar to those obtained from the biogenic calls. Therefore, simulations were only conducted with the biogenic calls. Further effort could be made in future to generate synthetic signals using an alternative method.

The WATTCH model was validated and used to perform broadband pulse propagation. WATTCH was first used to simulate experimental conditions encountered during the Gulf of Mexico experiment, resulting in acceptable model-data fidelity. With confidence in the simulation process established, an exploratory sensitivity analysis was conducted. By

varying only the sound speed profile and maintaining all other environmental parameters the same between simulations, it was found that classifier performance was altered by as much as 23 %. It was demonstrated that the differences in performance could largely be explained by disparity in the environment-dependent transmission loss.

As the experimental and simulated results were analyzed it became clear that SNR was an important contributor to decreasing classifier performance as signals propagated over increasingly longer ranges; however, the influence of propagation-induced signal distortion was not clear. Hence, Chapter 6 was devoted to developing a method to disentangle the impacts of signal attenuation and distortion. This was accomplished by considering their impacts separately as well as in combination, for which the WATTCH simulations proved very useful. The environments considered were inspired by the environmental measurements made during the Gulf of Mexico experiment and the results of Chapter 5's sensitivity analysis. Classifier performance was not significantly degraded by the impacts of signal distortion alone; however, when noise was added to the simulated signals, the performance dropped below that for the noise-only case. Additionally, it was found that the effects of propagation caused the performance to begin to degrade at shorter transmission ranges than if SNR had been the only contributor.

Shannon information theory predicts a linear relationship between SNR and classification performance. With this in mind, linear regression was performed on performance data points for attenuated and distorted signals which were plotted as a function of SNR. The coefficient of determination was used as an indicator of the relative proportion of performance change explained by SNR. From this method, it was found that as much as (94 ± 2) %, or as little as (91 ± 2) %, of the decrease in performance could be attributed to attenuation as signals were transmitted through the Gulf of Mexico environment. Therefore, a relatively small proportion — (6 ± 2) % to (9 ± 2) % — was a result of propagation-induced signal distortion. This result indicated that there are likely many environments in which signal-to-noise issues dominate. A positive outcome is that, by employing multi-element arrays, one is able to increase the SNR of recorded vocalizations by taking advantage of array gain, thereby increasing the performance of the classifier with minimal effort. In fact, in SNR-dominated environments, this is expected to produce performance improvements beyond the minor gains from adjusting a PAM system's parameters to account for environment-dependent signal-distortion.

The final scenario considered was that of bowhead and humpback vocalizations in their true frequency band (not the frequency band of the upsampled calls used for the experiments). The bowhead and humpback whale vocalizations were propagated through two simulated environments designed to represent propagation extremes. These consisted of an isospeed water column with a sandy bottom, and a downward refracting sound speed profile with a sedimentary rock bottom. A small change in performance (i.e., difference in accuracy of $\sim$10 %) was noted for the WATTCH-simulated signals without additive noise in the isospeed environment. Although the general characteristics of the performance changes were similar to those noted for the Gulf of Mexico experiment (e.g., decreasing performance with increasing propagation), performance results from the downward refracting environment highlighted that signal distortion can have a significant impact. Specifically, in the downward refracting environment, as much as 40 % of the decay in the mid-performance region was attributed to time-spreading due to multipath addition. Therefore, in at least some environments the ocean acoustic properties which result in signal distortion should be taken into account when characterizing performance of automated PAM systems.

Many of the results in this thesis were derived from bowhead and humpback whale vocalizations scaled to the 1 to 4 kHz band. The frequency scaling was necessary to take advantage of the passband of the source used during the Gulf of Mexico experiment. Unfortunately, this meant that the applicability of the experimental results in Chapter 4, and the simulation results in Chapters 5 and 6, to PAM of bowhead and humpback whales was limited to the general trends. For instance, from the higher frequency band results, it was possible to conclude that classifier performance was range-dependent, the effects of both signal distortion and SNR needed to be accounted for when characterizing classifier performance, and classifier performance can be significantly impacted by changes in the ocean SSP. The lower frequency band results presented in Chapter 7 confirmed the validity of the general conclusions made from the higher frequency band results. It is also important to note that, despite the limitations for the bowhead and humpback vocalizations, the results may be directly applied to other kinds of transient signals in this frequency band (e.g., anthropogenic signals, or other cetacean vocalizations that exist in the 1 to 4 kHz band).

Although the aural classifier was used throughout this thesis, the methods developed

may be applied to any automated PAM system. This could be especially useful for detection/classification systems which are more susceptible to the impact of environment-dependent propagation than the aural classifier, based on the types of features they extract; for example, Baumgartner *et al.* [5] state that the low-frequency detection and classification system (LFCDS) they employ is 'particularly vulnerable' to missing faint calls because it must detect the presence of a sound and estimate the shape of the pitch track from the spectrogram. Baumgartner *et al.* raise the issue of SNR, but neglect the impact of propagation-induced distortion, which can reasonably be expected to alter the shape of a vocalization in a spectrogram. An investigation such as the one conducted in this thesis would be useful for fully characterizing the performance of such systems to develop an understanding of their limitations.

Since propagation-induced signal distortion is a deterministic process, it is theoretically possible to reverse the effects of propagation and recover the source signal (plus noise). With the propagation-induced distortion removed it would be possible to perform classification on the inverted signal. There is a large body of inversion-theory literature that could aid in this. Given that ocean acoustic propagation inversion is relatively complicated and computationally intensive (due to the large number of influencing parameters), and that only a small gain is anticipated, this approach is not recommended in general, since, as we have seen throughout this thesis, the SNR issue tends to swamp the impacts of propagation. It may be worth considering for environments with extreme propagation effects.

Throughout much of this thesis whale vocalizations have been treated as convenient source signals for investigating the impacts of environment-dependent propagation on the performance of automated classifiers, without much regard for the biological aspects of the whale vocalizations. That is to say, the intent behind whale vocalizations has not been considered. In general, there are two possibilities — first, the whales are propagation-agnostic, and second the whales actually make use of the effects of propagation. In the first case, it is reasonable to assume that the whale vocalizations were evolutionarily designed to be robust to the impacts of propagation so that as the calls are transmitted through the ocean medium, the integrity of the encoded information is maintained. In fact, Mercado and Frazer [15] point out that animals that communicate over long distances typically use sound repertoires that remain discriminable even after considerable signal distortion. This could explain why the influence of SNR dominated the decrease in classifier performance

with only a relatively small portion attributed to propagation effects. In the second case, a vocalizing whale may make use of the propagation to encode information about its location. For example, the received call may contain cues as to relative distance between the source and receiving whales. Through simulation, Mercado *et al.* [27] show that there is sufficient information contained in received humpback whale calls to suggest that, "Listening whales should be able to range singers using distance-dependent changes in frequency content." Although two distinct possibilities were initially suggested here, there is a third scenario that is a combination of these two — since whales have a large repertoire of calls it is conceivable that they choose calls based on the desired effect. For instance, if they wish to convey their location to conspecifics they may use a call that will be impacted by propagation. On the other hand, if they need to communicate some information, they may employ a call that is minimally impacted by propagation effects. This makes the design of a PAM scenario somewhat complicated, since one would want to select calls that a whale produces frequently and is robust to the effects of propagation. Otherwise, the impact of environment-dependent propagation should be taken into account when considering the performance of an automated classifier. As a final comment, the methods developed here for PAM of cetaceans may be even more applicable to anthropogenic signals which were not optimized for discriminability after distortion by the ocean medium. The methods employed in this thesis would be equally well-suited to any passive sonar discrimination task.

Future work should be done on a few topics, including considering other environments in which propagation effects drive performance; the waveguide invariant parameter [116, 117] may be implemented in advance of simulations to identify dispersive environments that would result in appreciable propagation-induced distortion. This thesis mainly considered the effects of changing the SSP; the impact of other bathymetry profiles, bottom types, and ambient noise conditions, should also be investigated. Furthermore, the effects of surface roughness were not fully incorporated in simulations. In future, simulations should include frequency-spreading resulting from boundary interactions. After incorporating frequency-spreading, it would be interesting to investigate how propagating signals through an environment with a high sea state and an upward-refracting SSP impacts the performance of the aural classifier over long ranges, since one would expect a lot of signal distortion from this type of environment. Extending the modelling results of Chapter 7 to include

range-dependent propagation environments (e.g., sloping bottom) would also be useful.

Performing additional experiments with high source levels (to allow for longer transmission ranges) and in an environment with more stable properties than observed during the Gulf of Mexico sea trial would be beneficial. As one cannot count on stable ocean properties, future experiments would benefit from more extensive environment measurements; for example, employing both environmental moorings and ship-based measurements to disentangle spatial and temporal information. Conducting a sensitivity analysis in advance of future experiments would be useful to determine which environment properties are critical to sample, and with what required resolution. Such a simulation prior to future experiments or PAM surveys is highly recommended, as the results could also be used to determine the extent to which SNR would be likely to affect the performance of automatic recognition systems. This would help decide if the added cost of a multi-element array is necessary, and at what depth in the water column the recorder should be placed to obtain the best SNR for received vocalizations.

Future work could also be done to investigate the impact of environment-dependent propagation on other cetacean species; for example, humpback and North Atlantic right whales. The time-frequency characteristics of these species' vocalizations are similar, which would provide a challenging test case. Applying the methods employed in this thesis to other species, and different locations, would be useful to assess how well the results presented in this thesis generalize to other typical PAM scenarios.

Finally, it would be interesting to determine how environment-dependent propagation impacts the individual perceptual features employed by the aural classifier. Statistical measures could be used to analyze if individual features are significantly changed, so that general comments and recommendations could be made reflecting the robustness of both the individual perceptual features and the aural classification process as a PAM tool. If it is found that propagation does significantly impact the features, then the next step would be to rank the features in order of their sensitivity to propagation-related effects. A recommendation could be made to remove those features which are especially sensitive to the acoustic environment. Alternatively, it may be found that many of the perceptual features are environment-sensitive and therefore it is unreasonable to exclude all of these features. If this is the case, a strategy may instead be developed to generate training sets for the classifier that take propagation-related signal distortion into account.

APPENDIX A

MOORING DIAGRAMS

Figures A.1 through A.4 show diagrams of the moorings deployed during the Q350 sea trial in the Gulf of Mexico during the spring of 2013. Each mooring was bottom-mounted, with hydrophones attached to the moorings at different depths within the water column (refer to Table 4.1 for hydrophone depths). An acoustic release was used to allow the moorings to be easily recovered after experiments were completed. The Vibration Isolation Modules (VIMs) were used to reduce strumming sounds in the recordings associated with vibrations along the moorings.

Due to recording errors, no acoustic measurements were made with either hydrophone on Mooring 1 (Figure A.1) and data from only a single Reson hydrophone on Mooring 3 (Figure A.3) was recorded. Appendix B discusses which hydrophone's data may have been recorded by the SHARP recording unit and asesses the importance of accurately determining which hydrophone was recorded.

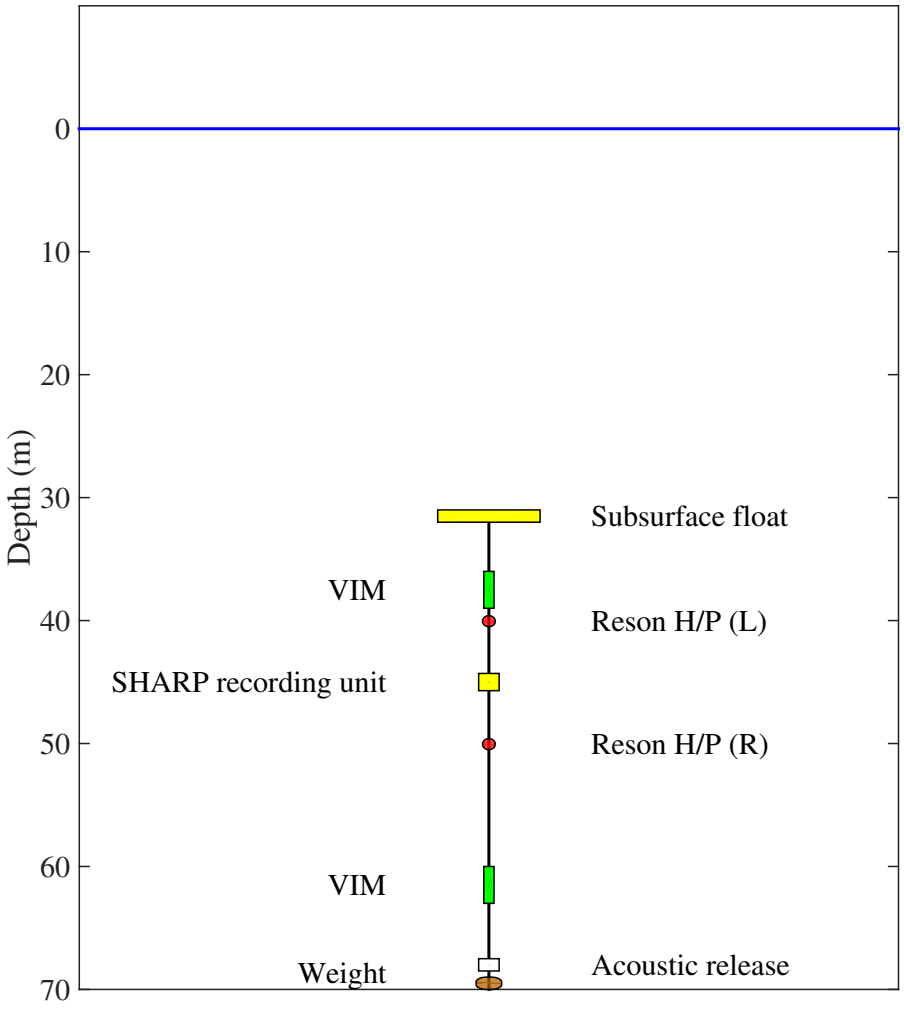


Figure A.1: Diagram of the SHARP mooring deployed on 30 April 2013 (Mooring 1). The acoustic data from both Reson hydrophones (H/P) are digitally recorded on the SHARP recording unit, on either the left (L) or right (R) channel of an audio track.

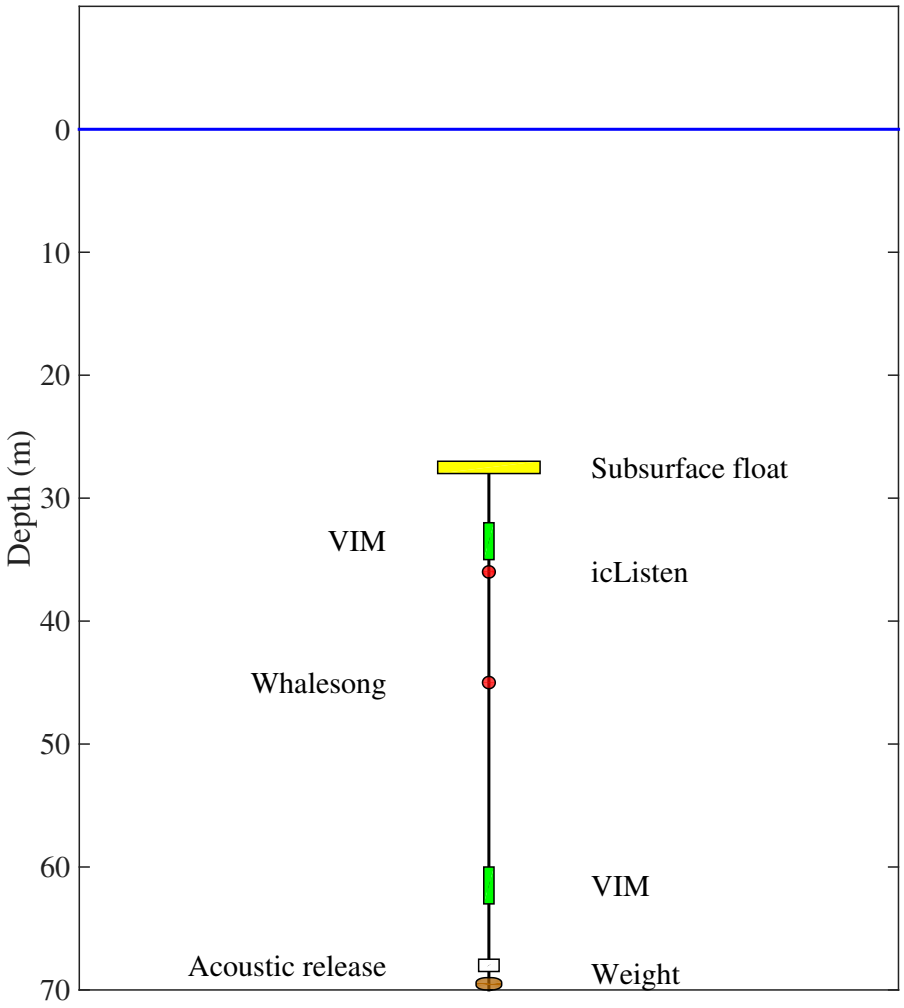


Figure A.2: Diagram of the Whalesong mooring deployed on 30 April 2013 (Mooring 2).

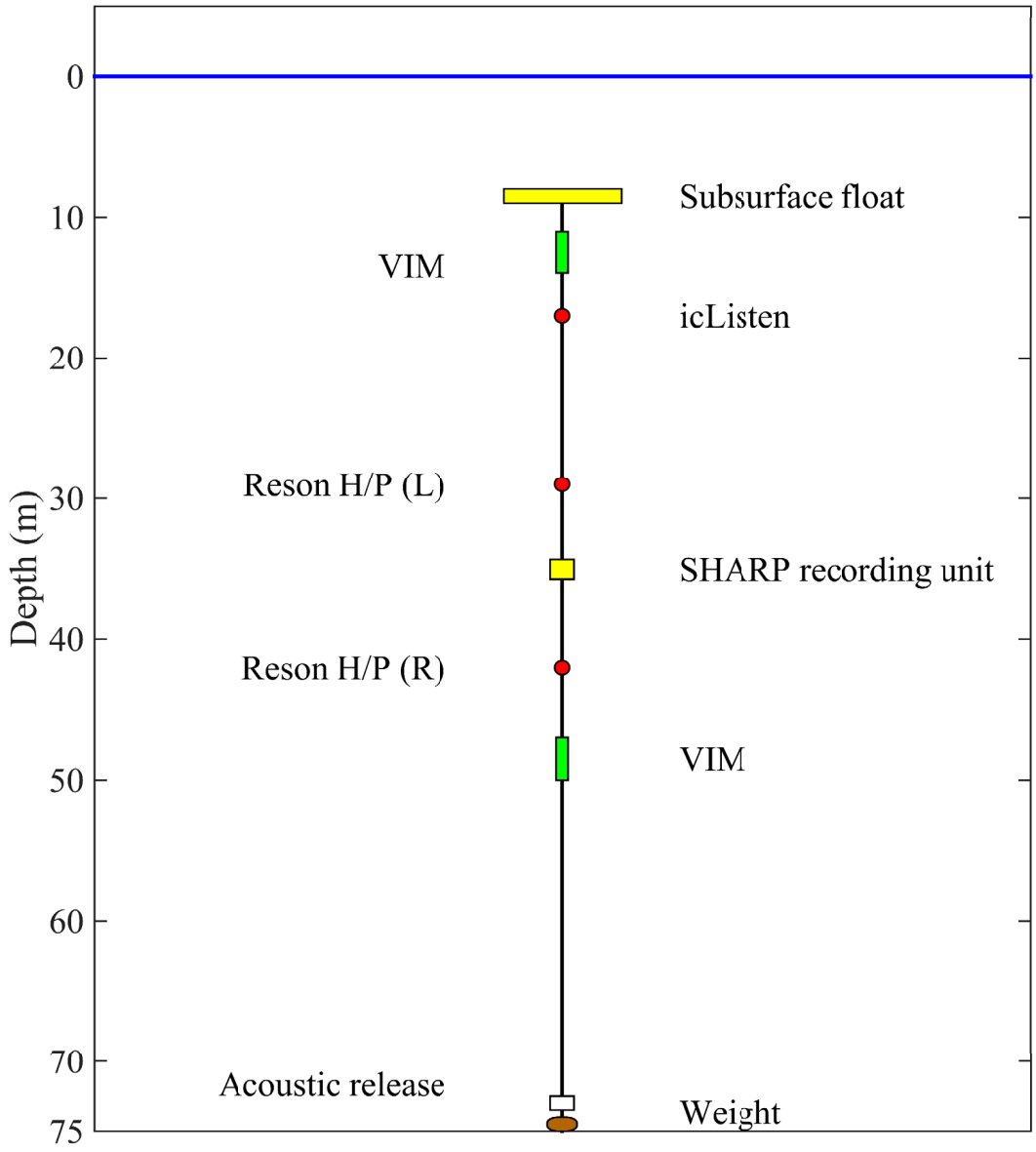


Figure A.3: Diagram of the SHARP mooring deployed on 1 May 2013 (Mooring 3). The acoustic data from both Reson hydrophones (H/P) are digitally recorded on the SHARP recording unit, on either the left (L) or right (R) channel of an audio track.

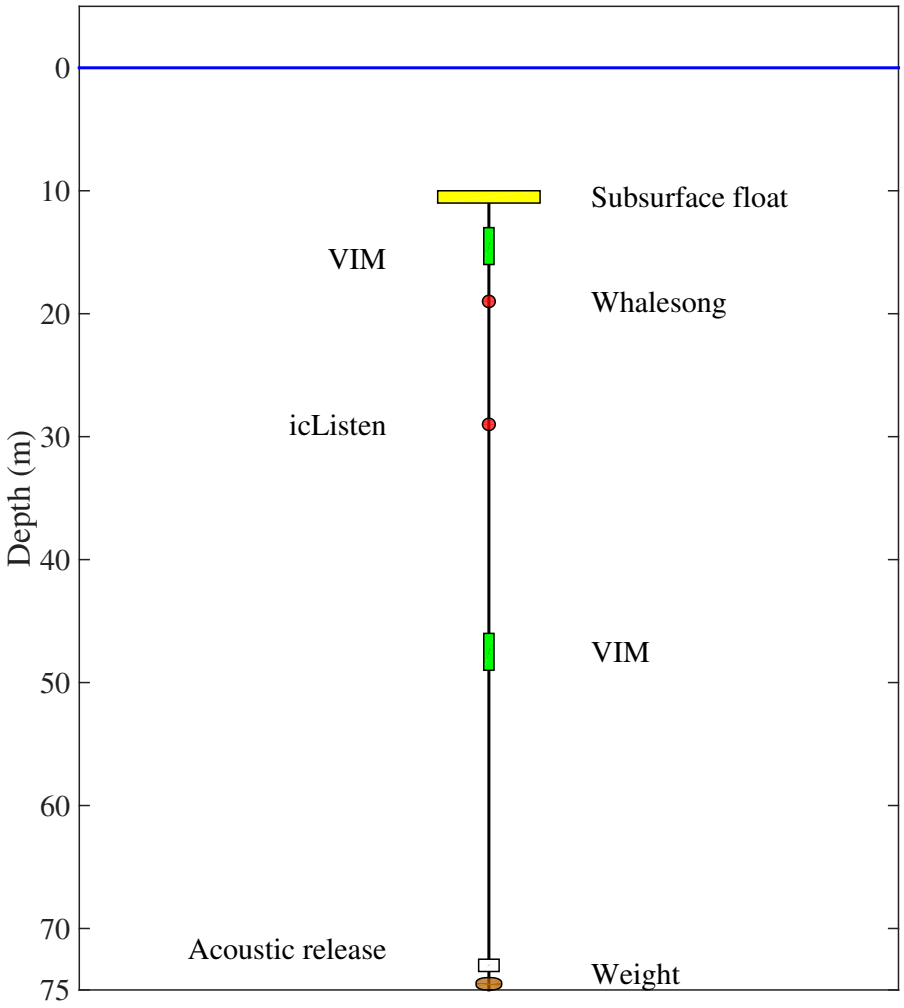


Figure A.4: Diagram of the Whalesong mooring deployed on 1 May 2013 (Mooring 4).

APPENDIX B

DETERMINATION OF HYDROPHONE DEPTH ON SHARP MOORING

B.1 Background

As first mentioned in Section 4.2, due to electronic problems there were almost no data recorded to the right channel of the audio files by the SHARP recording unit on 1 May 2013. Four WAVE files, three approximately 2 hours and 15 minutes and the fourth approximately an hour and 40 minutes long, were created by the SHARP recorder. Of these four files, only one minute and 42 seconds of data were recorded to the right channel in the first file. As shown in the mooring diagram depicted in Figure B.1, the hydrophones were intended to be connected such that the left channel of the audio file would record acoustic signals from the upper hydrophone, and the right channel should have contained recordings from the lower hydrophone.

One potential explanation for only a single channel being recorded is that the audio jack was not correctly connected, such that the left channel on the audio jack aligned with the right channel of the audio port (refer to Figure B.2 for a representation of an audio jack and socket). Due to the design of the audio connector, this would result in the upper hydrophone being recorded on the right channel and the lower hydrophone not being recorded at all. If the internal Sony recorder handles a single input data channel by recording it to the left channel, regardless of where it originated then, this would result in the lower hydrophone appearing to be recorded on the left channel; however, based on post-trial bench experimentation the more likely explanation is that due to the tight fit in the pressure case, a force was applied that caused the connector to bend upwards

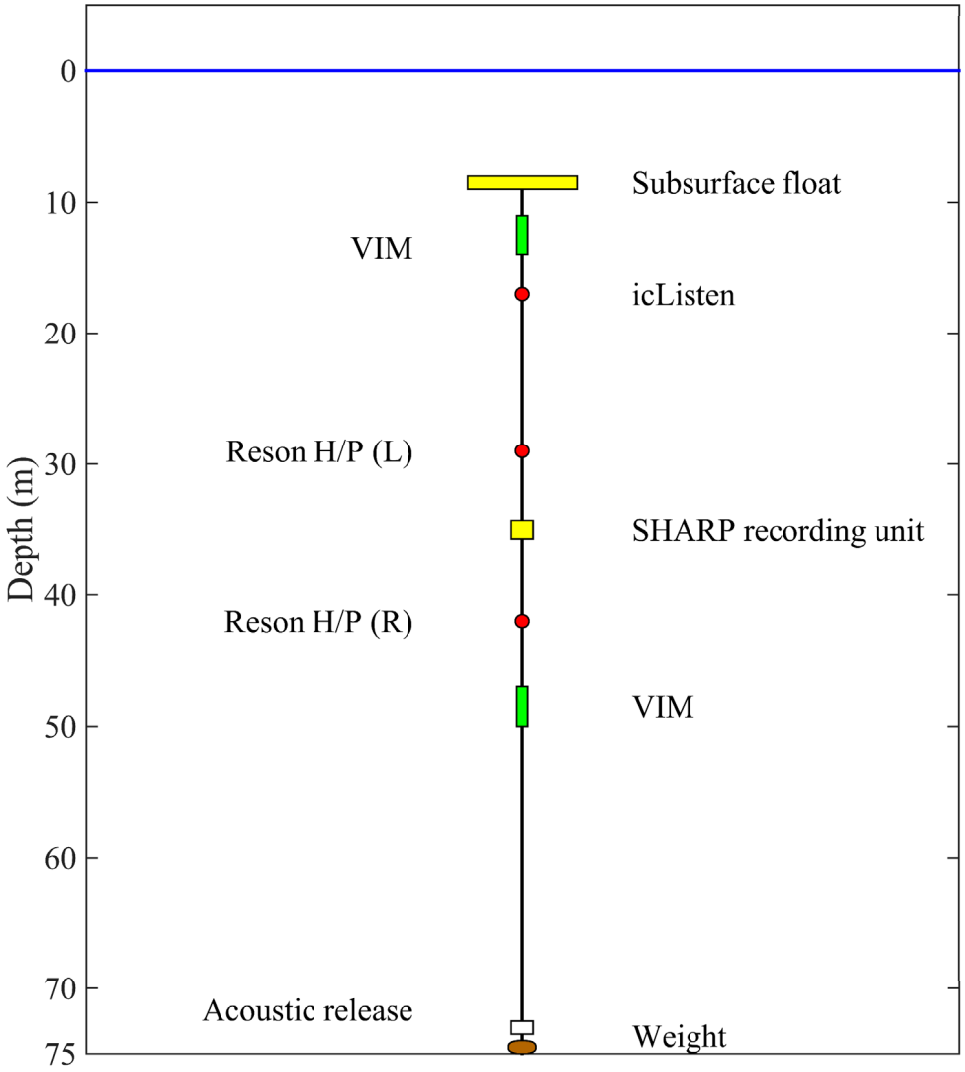


Figure B.1: Diagram of the SHARP mooring deployed on 1 May 2013. The VIMs shown on the diagram are Vibration Isolation Modules. The acoustic data from both Reson hydrophones (H/P) are digitally recorded on the SHARP recording unit, on either the left (L) or right (R) channel of an audio track. This figure is a copy of the diagram shown in Figure A.3.

and disengage the right channel of the audio jack from the port. When this was simulated during bench tests, a recording was made on the left channel, but not on the right channel. Given these possibilities, there is some uncertainty about which of the two hydrophones was actually recorded by the SHARP. In an effort to determine which hydrophone was recorded, the arrival structures of signals transmitted over the shortest range during the experiment were examined. As these results were inconclusive the effect of hydrophone depth on classification performance was examined through simulations to determine if

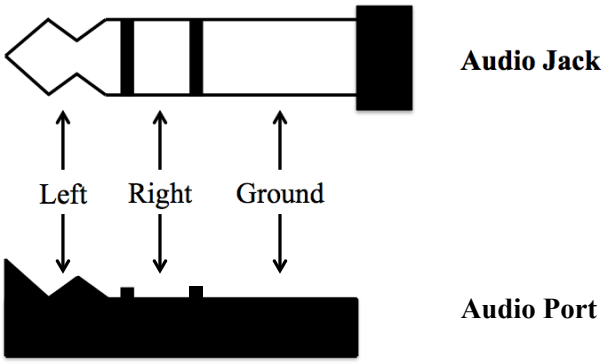


Figure B.2: Schematic of an audio jack and cross section of half an audio port.

it was necessary to know the exact depth at which the signals were recorded or if the hydrophone depth had little impact on the classifier's performance.

B.2 Determining Hydrophone Depth Via Ray Theory

This section discusses how ray theory was applied in an attempt to determine the depth of the hydrophone that was recorded by the SHARP recorder. The possible hydrophone depth was either 28.5 m or 41.5 m, as measured by the pressure loggers attached to the hydrophones on the SHARP mooring.

B.2.1 Method

This analysis makes use of an LFM transmitted at the beginning of the first signal set that was broadcast on day 2 of the experiment. This signal was transmitted at approximately 14:20 UTC when the separation between the source and SHARP mooring was 743.6 m, as determined from the GPS co-ordinates of where the SHARP mooring was deployed and the location of CFAV QUEST at the time of the transmission. Ray-tracing performed by Bellhop was used to estimate the times rays associated with the direct path, and surface and bottom reflections would arrive at the receivers. The centre frequency of the transmission bandwidth ($f = 2500$ Hz) was used for ray-tracing. Environmental measurements collected during the experiments were used as inputs to the Bellhop model. The source depth was set at 40 m, as measured by a pressure logger on the source, the water depth was set at 75 m. Geoacoustic parameters for sand were selected ($c = 1650$ m/s, $\rho = 1.9$ g/cm^3, $\alpha = 0.8$ dB/(m kHz)), and the sound speed profile sampled *in situ* was used and is shown in Figure B.3.

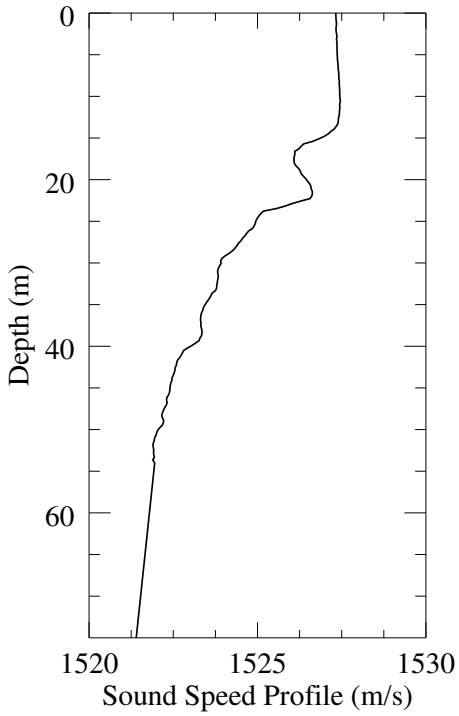


Figure B.3: Sound speed profile measured *in situ* and used as Bellhop model input to calculate ray travel times.

There were 501 rays launched from the source location in a fan between 25° and −25°, measured from the horizontal. Unfortunately, since Bellhop does ray tracing by using a fan of rays rather than eigenrays (i.e., rays that connect the source and receiver) [90] it is not possible to specify a range-depth coordinate at which the ray arrival times are required. Instead one must define a range-depth box and analyze the rays that pass through the box to estimate the arrival times. This introduces some uncertainty as to the precise timing of when the rays arrive at the modelled receivers. The boxes used for determining the arrival times at the hydrophones were defined such that the range was (743 ± 2) m and the depths were (29 ± 1) m or (42 ± 1) m. After the ray arrival times for the direct arrival, surface and bottom reflections were determined, the two modelled results for the possible hydrophone depths were then compared to the signal measured during the experiments.

To determine the timing of the arrivals from the experimental data, the recorded LFM was matched filtered against a replica LFM. The time difference between the direct arrival and surface reflection, $\Delta t_{d,s}$, and between the direct arrival and bottom reflection, $\Delta t_{d,b}$ were then compared against the modelled results to determine the depth at which the

signals were recorded.

B.2.2 Results and Discussion

The envelope of the matched filtered response for the signal recorded on the SHARP is shown in Figure B.4. An inset in the figure shows a zoomed in view of the first cluster of arrivals. This cluster of arrivals contains the direct arrival, surface and bottom reflections. Through comparison with Bellhop results, the first peak was identified as the direct arrival with a high level of confidence. The second peak is likely the surface reflection; however, it becomes difficult to confidently identify which peak is associated with the bottom reflection. Normally one would model the experimental geometry and use the predicted arrival times to identify each peak, whereas in this case the goal is to use the arrival structure to fill in the gap in our knowledge of the geometry. It therefore becomes difficult to identify the peak associated with the bottom reflection since this is an underdetermined problem (i.e., fewer pieces of evidence than unknowns).

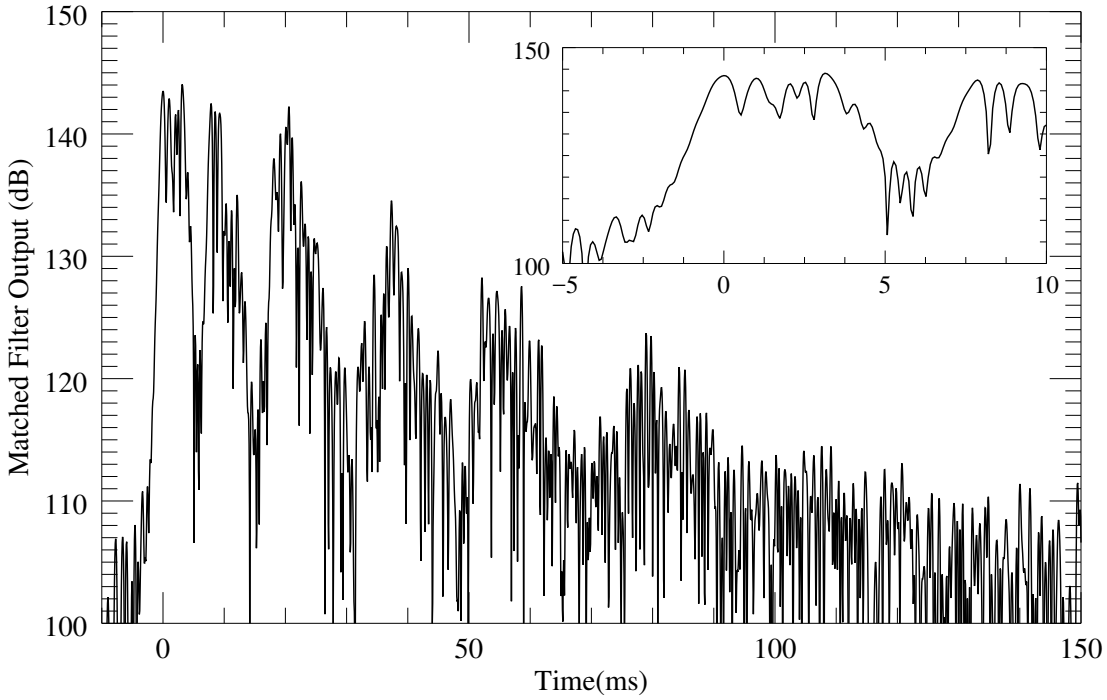


Figure B.4: Envelope of matched filter results showing arrival structure of LFM recorded by the SHARP recording unit. The time scale is defined such that the direct arrival is at time $t = 0$ s. The inset figure shows a zoomed in view of the first cluster of arrivals.

Table B.1: Comparison of the experimental and Bellhop-modelled time differences between the direct arrival, and the surface and bottom reflections. Three possibilities are given for the experimental value of $\Delta t_{d,b}$ since it was unclear from the arrival structure in Figure B.4 which peak resulted from the bottom-reflected ray.

	Experimental	**Modelled**	
		29 m	**42 m**
$\boldsymbol{\Delta t_{d,s}}$ **(ms)**	1.0	1.1	2.0
$\boldsymbol{\Delta t_{d,b}}$ **(ms)**	1.8 2.5 3.1	1.4	2.9

Table B.1 compares the arrival times derived from the recored LFM with the ray-tracing results. Listed values are time differences between the relevant peak and the direct arrival. Note that three values are listed for the experimental $\Delta t_{d,b}$, this is because it was unclear which of the peaks in the arrival cluster resulted from the bottom reflection. Upon comparing the delay times for the experimental arrival structure with the two modelled hydrophone depth possibilities it was still unclear which hydrophone produced the recordings. The experimental result for $\Delta t_{d,s}$ matches the modelled result for the 29 m hydrophone, although none of the possibilities for the experimental value of $\Delta t_{d,b}$ match the modelled result for the 29 m hydrophone but are close to the modelled result for the 42 m hydrophone. Consequently, it is not possible to identify the depth of the hydrophone recorded by the SHARP using these ray-tracing results.

Unfortunately several sources of error compounded such that ray-tracing was not a viable option for identifying the hydrophone depth. One source of uncertainty results from the need to define a range-depth box for the ray-tracing solutions, such that the arrival times could be off by as much as 2.5 ms (assuming a sound speed of 1525 m/s and using a ray passing through the box corner, i.e., furthest away from the centre of the box). Such a large box was nessecary to capture the direct, surface- and bottom-reflected rays; if a smaller box was used one of these rays was not included. There is also uncertainty in the environmental parameters which affects the model fidelity. Additionally, the relative scale of the experimental geometry is such that it is difficult to distinguish arrivals within a few meters of depth; the range over which the signals were transmitted is much larger than the $\sim$10 m depth difference of the hydrophones (i.e., $R \gg \Delta z$), so that the path lengths

for the rays used for the analysis were too similar for the two depths considered. Finally, the sum of the uncertainties in the exact mooring placement, horizontal location of the hydrophone due to currents moving the mooring, and the horizontal location of the source relative to the ship's GPS produce a range uncertainty on the order of 10 to 15 m, which is at least as large as the depth difference to be resolved. Ultimately, it was not possible to use ray-tracing to determine the depth of the hydrophone recorded by the SHARP.

B.3 Determining Importance of Recorder Depth to the Aural Classifier Performance

Since it proved too difficult to determine which hydrophone was recorded by the SHARP, the next obvious question was, "Does the hydrophone depth significantly impact classifier performance?" If the aural classifier performance is not sensitive to the depth of the receiver, for the given experimental geometry and environment conditions, then determining which hydrophone was recorded by the SHARP becomes irrelevant. This section describes the simulation that was conducted to examine the sensitivity of classifier performance to hydrophone depth.

B.3.1 Method

To determine the impact of the hydrophone depth on the classifier performance the WATTCH model was used to simulate bowhead and humpback synthetic signals propagated over three of the ranges signals were transmitted during the experiments ($R =$ 0.75, 7 and 11 km). The model was configured such that signals were "received" at both hydrophone depth possibilities. The aural classifier was then used to discriminate between the synthetic bowhead and humpback calls for each of the range-depth combinations. Then the simulated classifier performance results were compared with each other and the experimental results to determine if classifier performance was significantly impacted by the hydrophone depth.

The WATTCH model was configured to simulate the environment observed on 1 May 2013. Bathymetry and sound speed profiles measured *in situ* (refer to Figure B.5) were used as inputs. A source depth of 40 m and possible receiver depths of 29 and 42 m were used. As described in Section B.2.1, geoacoustic parameters for a sandy bottom were selected. After the signals were propagated using the WATTCH model, snippets of experimental

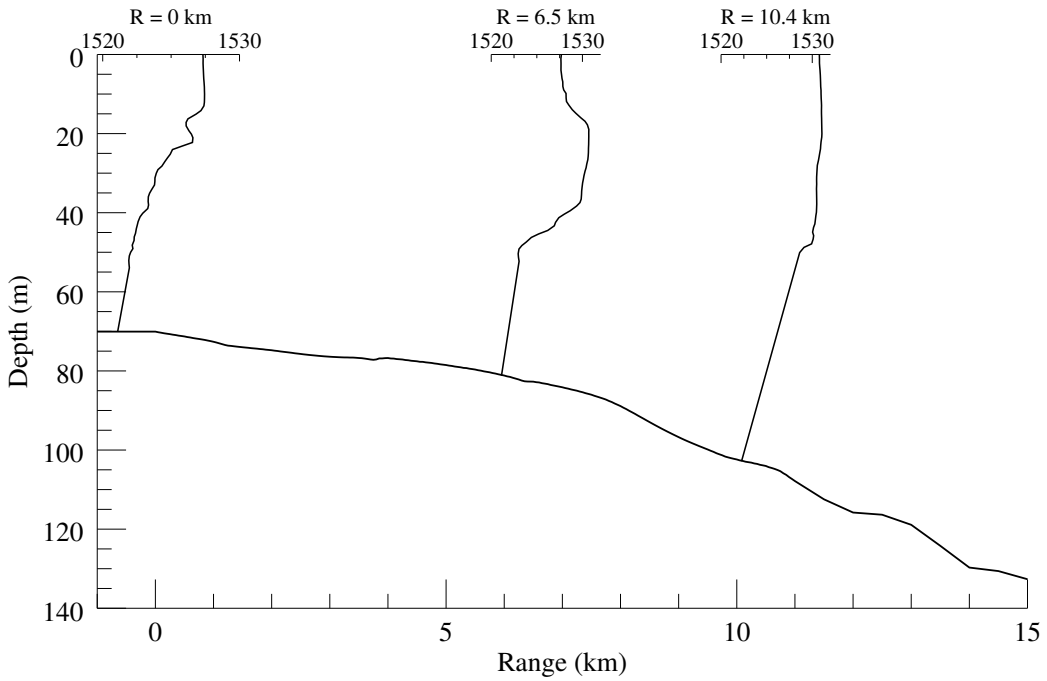


Figure B.5: Bathymetry and sound speed profiles used as environmental inputs to the WATTCH model for simulating the impact of hydrophone depth on the aural classifier performance. The sound speed profiles are given in units of m/s.

noise were added to each signal in such a way as to match the mean and standard deviation of the SNR of the signals recorded by the SHARP unit during the experiment. The method of adding noise to signals propagated through WATTCH is fully described in Section 6.2.1. This process was designed to replicate the signals transmitted during the sea trial, after which the the aural classifier was then applied to discriminate between the bowhead and humpback synthetic signals using 5-fold cross-validation. Performance results in terms of AUC and classification accuracy were analyzed to determine if there is any significant impact on classifier performance associated with the hydrophone depth.

B.3.2 Results and Discussion

Aural classifier performance results are presented in Figure B.6. Each data point represents the average performance and the vertical error bars are one standard deviation, computed using 5-fold cross-validation. There are three obvious groupings of data points along the range axis — within each of these groups the signals were propagated over the same range, the data points are offset along the range axis merely to reduce overlapping data points to facilitate comparison. It is worth noting that error bars on these plots do not include all

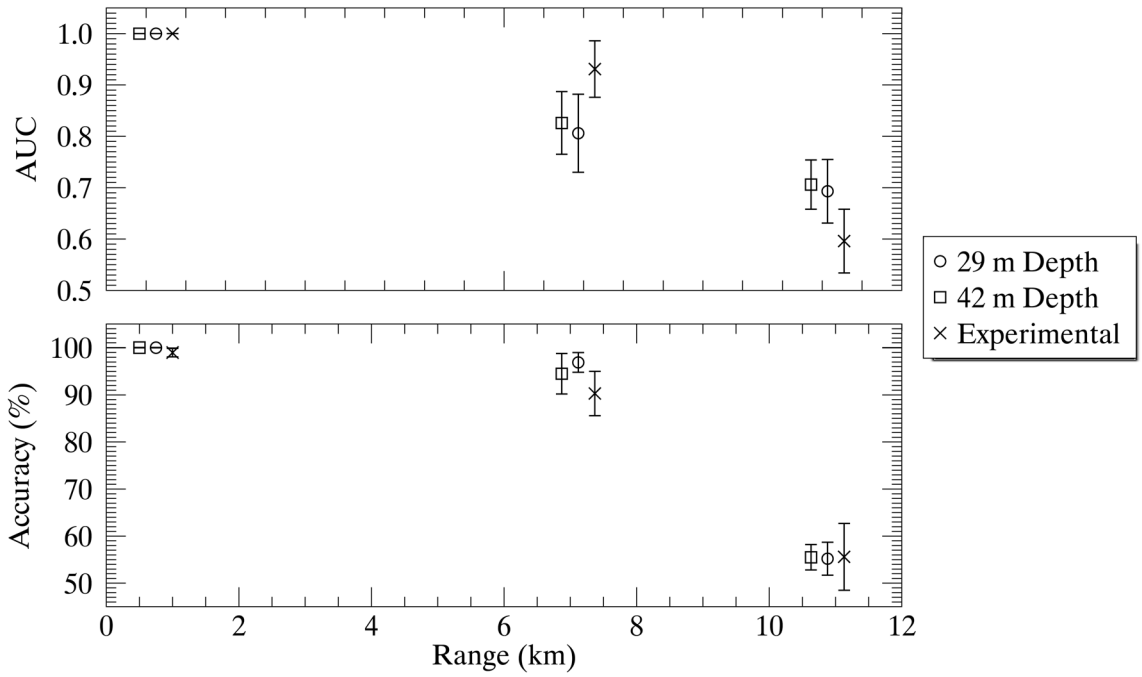


Figure B.6: Effect of SHARP hydrophone depth on classifier performance. Modelled results are compared with results obtained from signals recorded by the SHARP recording unit. Note that groups of data points are slightly offset along the range axis to ease comparison of data obtained at the same range.

sources of error, rather they only capture the classification error. If all errors (e.g., errors in the experimental geometry, errors in sound speed measurement, etc.) were included the errors bars would be larger; therefore, the error bars shown in this plot are an optimistic view of the true error.

From the plots in Figure B.6 it can be seen that the aural classifier's performance in discriminating the simulated signals is statistically similar to its performance in discriminating between the signals transmitted during the experiment, which provides model-data validation. Perhaps more importantly, the difference in performance between the two depth possibilities is minute, and not statistically significant. Thus, it may be concluded that it is inconsequential which of the two hydrophones was actually recorded when interpreting the aural classifier performance results for signals recorded by the SHARP unit. This is an encouraging result, since it was not possible using ray theory (refer to Section B.2) to determine which hydrophone was recorded by the SHARP unit. It also reaffirms the power of the aural classification method — even though the signal waveforms received at each hydrophone depth may look different, the aural characteristics remain the same and so the

aural classifier is not sensitive to relatively small changes in experimental geometry.

It is is important, nonetheless, to realize that the conclusion that hydrophone depth does not impact classifier performance is not a general conclusion, but rather is specific to the experimental geometry and ocean environment simulated. There may in fact be cases in which the hydrophone depth would impact classifier performance; for example, one can imagine a scenario in which both a hydrophone and source are in a surface duct and another hydrophone is below the surface duct. In this scenario the signals recorded on the hydrophone would have much different received levels at the same range from the source, which would be reflected in a difference in classifier performance since it was demonstrated in Chapter 6 that SNR influences the aural classifier performance.

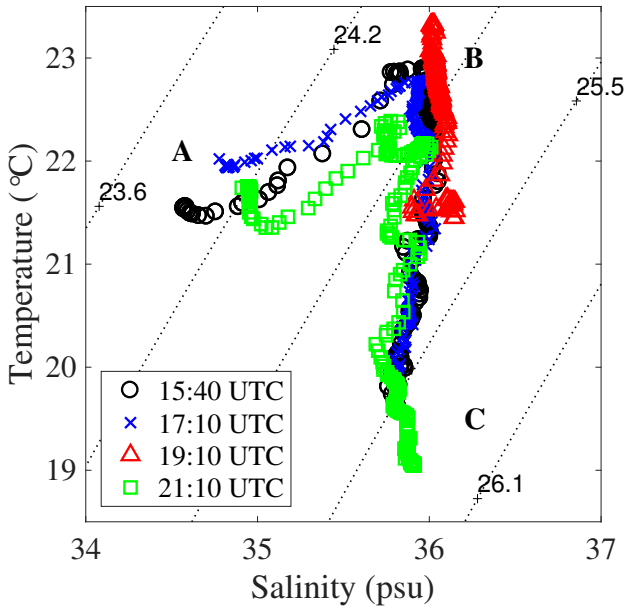
B.4 Summary Remarks

Unfortunately, it was not possible to determine the depth of the hydrophone recorded to the left channel of the SHARP recorder using ray theory due to the relatively large uncertainties in transmission range compared to the relative depths of the hydrophones. Ultimately, this did not matter as it was shown through the use of simulations that classifier performance was not significantly influenced by either hydrophone depth possibility for the acoustic environment encountered in the experiment. With this in mind, and given the results of the post-trial bench tests, it was decided to assume that the left channel recorded signals from the upper hydrophone, as intended.

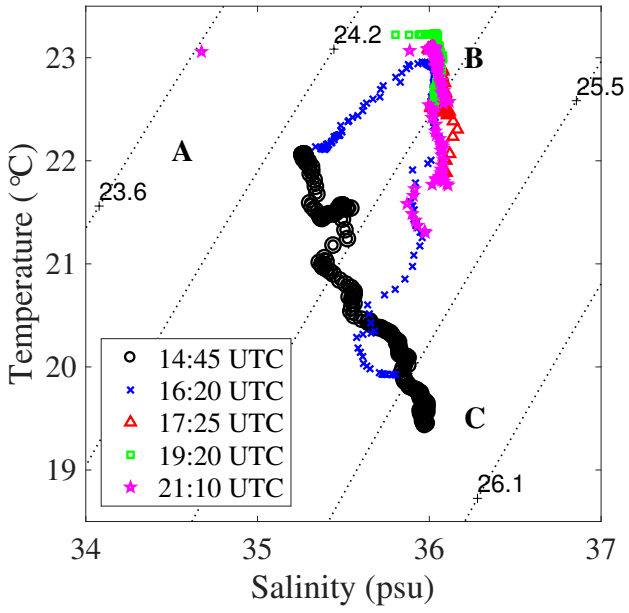
APPENDIX C

TEMPERATURE-SALINITY DIAGRAMS

Temperature-salinity (T-S) diagrams [71] are provided in Figure C.1 — these were generated during analysis of the anomalous sound speed profiles (SSPs) discussed in Section 4.3.4. The data associated with the anomalous SSPs are represented here by the red triangle data series in Figure C.1a and the black circle data series in Figure C.1b. Three regions have been labelled on each diagram, as 'A', 'B', and 'C' — these regions are each associated with a parent water type. The majority of the profiles show mixing lines between 'A'/'B' and 'B'/'C'; however, the anomalous SSPs do not follow this trend. Instead, the profile measured at 19:10 on day 1 is predominantly of the 'B' water type with some mixing with water mass 'C,' but not with the lower salinity water mass 'A.' The profile taken at 14:45 on day 2 is even more distinct from the other profiles when viewed on the T-S diagram. Represented in this way, it is clear that the difference in this profile is a result of the 'B' type water not being included — instead the water sampled at this point in space and time only includes the 'A' and 'C' water masses with a mixing line connecting those two water types directly. No explanation is offered here for the cause of this. For further comments on the hydrography of the Northern Gulf of Mexico region and the impact of the anomalous SSPs on this research, refer to Section 4.3.4.



(a)



(b)

Figure C.1: Temperature-salinity diagrams from (a) day 1 and (b) day 2 of the Gulf of Mexico experiment.

APPENDIX D

ARRIVAL STRUCTURE

While calculating the source levels of signals transmitted during the experiment the matched-filter responses of LFMs were generated to determine the range between the source and receiver, refer to Section 4.4 for full details. A plot showing an example envelope of the matched-filter output was first presented in Figure 4.6 and is reproduced in Figure D.1 for the reader's convenience. The reader should recall that this matched-filter response was the result of matched filtering the signal recorded by the monitor hydrophone with an LFM kernel. When first examining this then, one might assume that the first peak (located at $t = 0\,\mathrm{s}$) corresponds to the direct arrival. Upon closer inspection, the lower level of the first peak compared with the later arrivals calls this conclusion into question, since for short range propagation one would typically expect the direct arrival to have the strongest response in the matched-filter output.

An alternate hypothesis is that the peak at $t = 0\,\mathrm{s}$ is actually a recording of the transducer drive voltage, as a result of cross-talk between channels of the recording system. To confirm this, the delay times predicted by ray theory for the direct arrival, and surface and bottom reflections were compared with the delay times determined from the experimental data. The method for accomplishing this is laid out in the following section. This exercise also has the added benefit that the enveloped matched-filter output provides the opportunity to verify the experimental geometry.

D.1 Method

The experimental geometry used to calculate the path lengths for the direct arrival, and bottom and surface reflections is defined in Figure D.2. The direct arrival travels along the

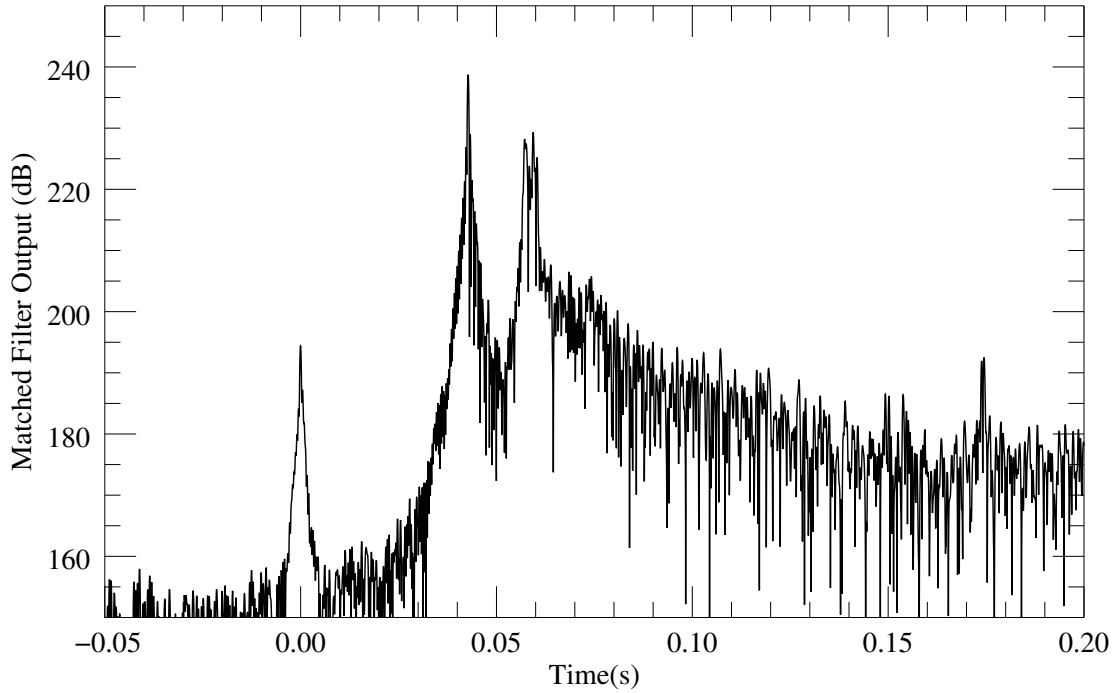


Figure D.1: Envelope of matched-filter results showing arrival structure of an example LFM recorded on Reson monitor hydrophone. The time scale is defined such that the signal transmission is at time $t = 0$ s. This is a reproduction of the Figure 4.6.

path r_1. This path length was determined from the time of flight where,

$$r_1 = c\Delta t_{\text{direct}} \quad , \tag{D.1}$$

Δt_{direct} is the time delay between the signal transmission and when the direct arrival was received, and c is the sound speed in water assuming an isospeed profile. It is expected that r_1 should be approximately 65 to 70 m, since CFAV QUEST's length is 76 m [118].

Let R_{sb} be the path length of the ray representing the surface reflection defined by,

$$R_{\text{sb}} = r_2 + r_3 \quad , \tag{D.2}$$

where r_2 and r_3 are defined in Figure D.2. From the Pythagorean Theorem,

$$R_{\text{sb}} = \sqrt{r_{sb}^2 + z_{\text{s}}^2} + \sqrt{(\Delta r - r_{\text{sb}})^2 + z_{\text{r}}^2} \quad . \tag{D.3}$$

Since the angle of incidence at the surface boundary is equal to the angle of reflection, the

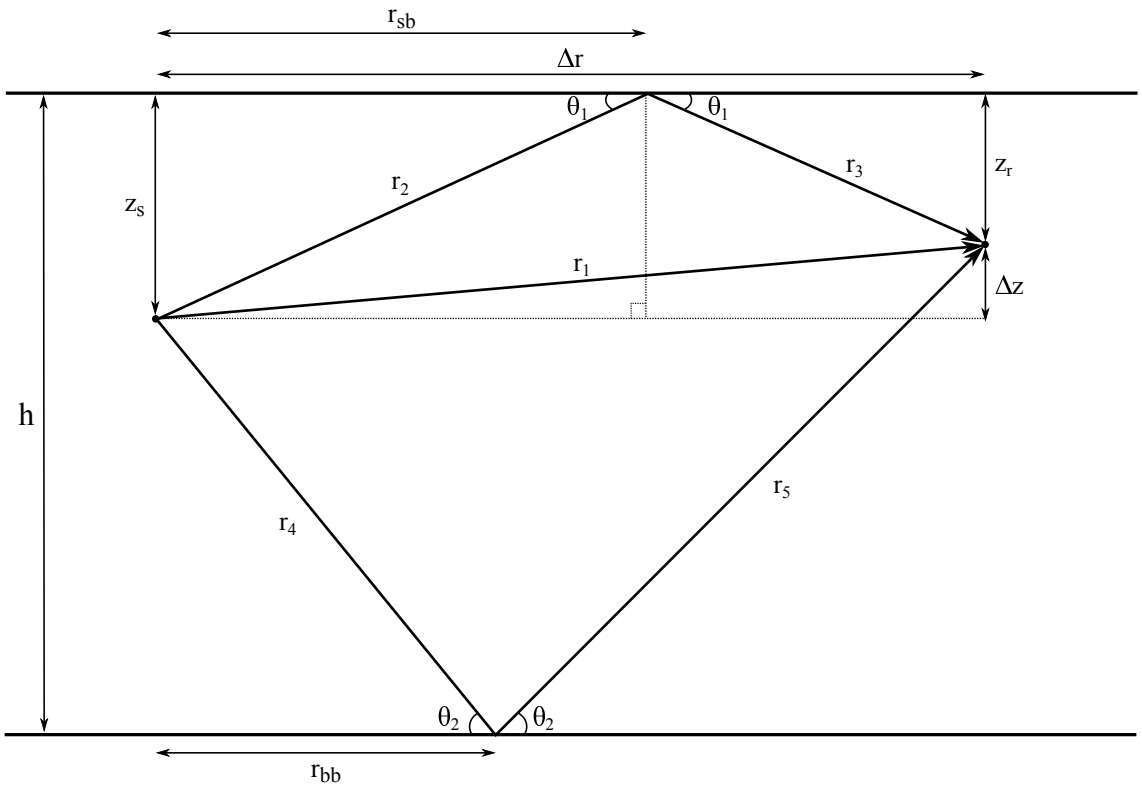


Figure D.2: Experimental geometry used to predict the arrival times of the direct arrival, surface reflection, and bottom reflection. Since an isospeed profile was assumed the rays can be approximated as straight lines.

following relationship is true,

$$\tan\theta_1 = \frac{z_{\mathrm{s}}}{r_{\mathrm{sb}}} = \frac{z_{\mathrm{r}}}{\Delta r - r_{\mathrm{sb}}} \quad , \tag{D.4}$$

where r_{sb} is the horizontal range between the source and where the ray reflects from the surface, Δr is the horizontal source-receiver separation, and z_{s} and z_{r} are the depths of the source and receiver, respectively. After rearranging,

$$r_{\mathrm{sb}} = \frac{z_{\mathrm{s}} + \Delta r}{z_{\mathrm{r}} + z_{\mathrm{s}}} \quad , \tag{D.5}$$

and noting that

$$\Delta r = \sqrt{r_1^2 - \Delta z^2} \quad , \tag{D.6}$$

where Δz is the vertical source-receiver separation. Finally, the time delay between the

signal transmission and reception of the surface reflection is,

$$\Delta t_{sb} = \frac{R_{\mathrm{sb}}}{c} \quad . \tag{D.7}$$

Following a similar method as laid out for the surface reflection, the path length travelled by the ray which reflects off the bottom is,

$$\begin{aligned} R_{\mathrm{bb}} &= r_4 + r_5 \\ &= \sqrt{r_{\mathrm{bb}}^2 + (h - z_{\mathrm{s}})^2} + \sqrt{(\Delta r - r_{\mathrm{bb}})^2 + (h - z_{\mathrm{r}})^2} \quad , \end{aligned} \tag{D.8}$$

where h is the water depth and the horizontal range between the source and the location where the bottom reflection occurs is,

$$r_{\mathrm{bb}} = \frac{h - z_{\mathrm{s}}}{2h - z_{\mathrm{r}} - z_{\mathrm{s}}} \Delta r \quad . \tag{D.9}$$

The time delay between signal transmission and the bottom reflection arrival is,

$$\Delta t_{\mathrm{bb}} = \frac{R_{\mathrm{bb}}}{c} \quad . \tag{D.10}$$

The analysis of the arrival structure described above requires input values for the experimental geometry to set up these calculations. Values for the source depths of z_{s} =20 m and $z_{\mathrm{s}} = 40$ m were used on 30 April and 1 May, respectively. Likewise, the receiver depths were set at z_{r} =30 m for 30 April and $z_{\mathrm{r}} = 22$ m for 1 May. These depth values were obtained from pressure loggers located on the source and the Reson monitor hydrophone. Values for water depth, h, changed throughout the course of the experiment. It was found that the ray calculations were not sensitive to changes of a few meters in h, so the following values were used for modelling transmission loss: $h = 75$ m on 30 April, and $h = 110$ m for signals transmitted at 17:00 and $h = 160$ m at 19:00 on 1 May.

The time delays predicted by Equations D.7 and D.10 are then easily compared with the time delays determined from the time difference between the peak associated with the transmission, and those with the surface and bottom reflections.

Table D.1: Comparison of the predicted and experimental delay times for the surface and bottom reflections. Source-receiver separation, r_1, was determined from Equation 4.7, and Equations D.7 and D.10 were used to predict the time delays of the surface and bottom reflections, respectively.

			Δt_{sb} (ms)		Δt_{bb} (ms)	
Date	**Time**	r_1 **(m)**	**Predicted**	**Experimental**	**Predicted**	**Experimental**
30 April 2013	16:00	67.2	5.86	5.40	7.86	7.69
	16:14	67.4	5.87	5.36	7.87	7.60
	16:28	68.9	5.94	5.42	7.96	7.96
	16:40	67.5	5.87	5.37	7.88	7.97
	17:00	68.4	5.92	5.39	8.11	8.08
	17:14	70.1	6.00	5.49	8.21	8.15
	17:28	69.1	5.95	5.40	8.15	8.21
	17:40	68.5	5.92	5.45	8.12	8.22
Daily Average	–	**68.4**	**5.92**	**5.41**	**8.02**	**7.99**
1 May 2013	17:29	70.9	6.04	5.95	11.68	11.96
	17:42	66.5	5.83	5.93	11.52	11.43
	17:53	65.6	5.78	5.72	11.49	11.69
	18:04	66.3	5.82	5.78	11.51	11.98
	19:23	65.5	5.78	5.94	17.57	17.45
	19:33	66.1	5.81	5.88	17.58	17.51
	19:44	69.9	5.99	5.78	17.67	17.95
Daily Average	–	**67.3**	**5.86**	**5.85**	**14.15**	**14.28**

D.2 Results and Discussion

Table D.1 compares the results of the predicted and experimental time delays for the surface and bottom reflections determined from matched filtering the LFMs recorded on the monitor hydrophone, assuming that the peak at $t = 0$ s corresponds to the transmission. The difference between predicted and experimental values for Δt_{sb} are larger on the first day of the experiment than on the second day. This may be because the isospeed assumption is more applicable on 1 May since the measured sound speed profiles were less complex and had weaker gradients than those measured on 30 April. Additional sources of error include: variability in depth of the source and monitor hydrophone (z_s and z_r), and variability in water depth (h). Nevertheless, the predicted and experimental time delays are sufficiently within agreement, that the peaks in the matched-filter results (e.g., Figure 4.6) were confidently identified as rays associated with the surface and bottom reflections.

Additionally, the values of r_1 determined from the time-of-flight calculations for the first and largest peak in the matched-filter response were within the range expected for the source-receiver separation, so that this peak was positively identified as the direct arrival, and can be used for source level estimates.

APPENDIX E

SELECTION OF SEDIMENT PARAMETERS

The latter chapters of this thesis employed pulse propagation models to study the impacts of signal attenuation and distortion on classifier performance. In general, environmental models are used to define the volumetric properties of the ocean and quantify the boundary conditions at the sea surface and sea floor — in this way the behaviour of the ocean as an acoustic medium is characterized [84]. The environmental models employed in Chapters 5 and 6 were based on measurements collected during the Gulf of Mexico sea trial. This Appendix considers in detail how the geoacoustic parameters used in the environmental models were determined.

Recall from Section 4.3, Ocean Environment Measurements, that *in situ* sediment properties were measured during the the Gulf of Mexico experiment using a FFCPT. Following the method outlined in Pecknold and Osler [103], the Robertson zone sediment types were translated into grain size, from which the geoacoustic parameters were calculated using formulae contained in the APL-UW Environmental Modelling Handbook [67]. For the Gulf of Mexico region, the FFCPT drops indicated a surficial sediment type consistent with "sand mixtures" and "sand," resulting in a mean grain size of $3.24\,\phi$ (or 0.106 mm). Using this mean grain size, the APL-UW formulae produced the geoacoustic parameters listed in the first row of Table E.1. Following the procedure in the APL-UW handbook, the sediments were modelled as a semi-infinite fluid bottom, that was assumed to be statistically homogenous in all directions.

The sound speed and density values obtained using this approach are lower than those typically associated with sand and sand mixtures [16, 78]. For comparison, the canonical

Table E.1: Sediment parameters for geoacoustic modelling.

Bottom Type Method	c **(m/s)**	ρ **(g/cm^3)**	α **(dB/λ_p)**
APL-UW	1630	1.3	0.60
Hamilton — silty-sand	1630	1.8	0.60

Hamilton paper gives values for fine-grained sand of $c = 1749$ m/s and $\rho = 1.9$ g/cm^3, and for silty-sand of $c = 1646$ m/s and $\rho = 1.8$ g/cm^3. The sound speed values from the APL-UW method are comparable with Hamilton's values for silty-sand; however, the density value is considerably lower than that reported by Hamilton. In fact, the author of the APL-UW handbook acknowledges that, for intermediate values of the grain size (between 1 ϕ and 9 ϕ), the sound speed and density values obtained using the formulae contained in the handbook, are known to be lower than those reported by Hamilton.

While it is recognized that the sediment density used in the environment models in Chapters 5 and 6 is lower than typically reported, it is believed that the use of the method to convert the FFCPT data to geoacoustic parameters is justified. The FFCPT is a robust tool that facilitates collection of *in situ* sediment properties; however, the nature of the FFCPT sampling method only characterizes the surficial sediment. To correspond with this sampling method, the geoacoustical parameterization in the APL-UW handbook was employed, due to its emphasis on the surficial sediment layer.

A key difference between the APL-UW handbook's and Hamilton's sediment parameters is the method in which they were measured. Hamilton reports values of c and ρ measured from sediment cores [78], whereas the APL-UW formulae are empirical relationships derived from reflection loss and backscatter data. The physics of the method used by the APL-UW handbook to obtain the geoacoustic parameters is consistent with how the Bellhop propagation model represents bottom interactions. That is, the Bellhop model determines a Rayleigh plane wave reflection coefficient at the water-sediment interface, and does not allow sound to propagate through the sediment. Although the resulting value of density is different from literature reports of the geoacoustic parameters of sand/sand mixtures, employing this combination of geoacoustic parameters should still produce an appropriate reflection coefficient at the bottom boundary. It is the reflection coefficient which is critical for modelling the acoustic field in the ocean medium. Therefore,

employing the smaller surficial sediment geoacoustic parameters, rather than the bulk values reported in other literature, is believed to be an acceptable approach.

E.1 Impact of Sediment Density on Propagation Effects and Classifier Performance

Since there may be some uncertainty about the best choice of sediment density for propagation modelling, it is germane to examine the impact of the two extremes of possible ρ. The two extremes are defined as the value obtained using the APL-UW handbook and a typical value reported in the literature for the sand/sand mixtures bottom type. Recall the sensitivity analysis conducted in Section 5.3, which considered the impact of SSP on the aural classifier performance by using an environmental model inspired by the Gulf of Mexico experiment. Here, one of the same environment models is used to determine if choosing a higher sediment density would impact the propagation or classifier performance results. The water column properties were characterized by the downward refracting SSP measured at $R = 0$ km, and range-dependent water depth measured during the experiment (refer to Figure 5.4). The downward refracting SSP was selected since it ensured the most bottom interaction. A source and receiver were placed at $z_s = 40$ m and $z_r = 29$ m, respectively. The geoacoustic parameters used throughout this thesis were compared with parameters consistent with those reported in Hamilton [78] — these are summarized in Table E.1.

The TL curves obtained using these two bottom parameterizations are shown in Figure E.1. There was little difference in TL over much of the range, indicating that the bottom density value had little effect on signal attenuation for ranges less than about $R = 8$ km. Next, the impact on the aural classifier performance was examined by comparing performance results as a function of range for these two environments (distinguished only be a difference in density). Following the steps outlined in Section 5.2, signals were propagated through the WATTCH model and snippets of experimental noise were added to achieve the estimated SNR, consistent with each environment. Results of training the classifier on signals propagated over a 70 m range and validating on signals propagated over 1, 6.5 and 10 km are shown in Figure E.2. Note that the results for $\rho = 1.3$ g/cm^3 (blue squares) are the same as the results for the downward refracting environment in Figure 5.10. Comparing

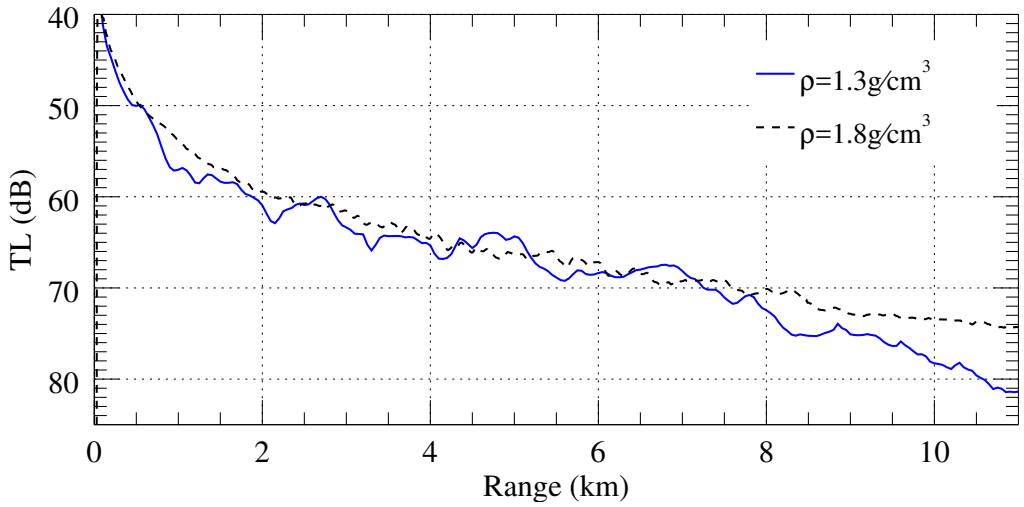


Figure E.1: Incoherent transmission loss modelled at f =2 kHz using Bellhop for sediment densities used for modelling throughout the thesis ($\rho = 1.3$ g/cm^3) and a typical value for silty-sand ($\rho = 1.8$ g/cm^3).

the performance results for each range, between the two sediment densities, shows that there was no significant difference in performance. Therefore, using a lower sediment density than the typical bulk densities for sand reported in the literature did not have a significant impact on the results presented throughout this thesis.

E.2 Summary Remarks

In summary, the geoacoustic model used in Chapters 5 and 6 was based on sediment characteristics measured *in situ* by FFCPT casts taken during the Gulf of Mexico experiment. With the mean grain size obtained from the FFCPT data, formulae in the APL-UW Environmental Modelling Handbook were used to determine the geoacoustic parameters to model the ocean bottom as a fluid half-space. It was noted that, in particular, the sediment density obtained using this process was lower than other literature reports. Use of this lower value was justified by the fact that the FFCPT only samples the surficial sediments, the correspondence between the experimental methods used to obtain the APL-UW formulae and the representation of the ocean bottom in Bellhop, and that parameters calculated from APL-UW's empirical formulae are likely to produce the correct reflection coefficient at the water-sediment interface. Furthermore, there was little difference between the two TL curves for bottoms with densities corresponding to either the APL-UW handbook's

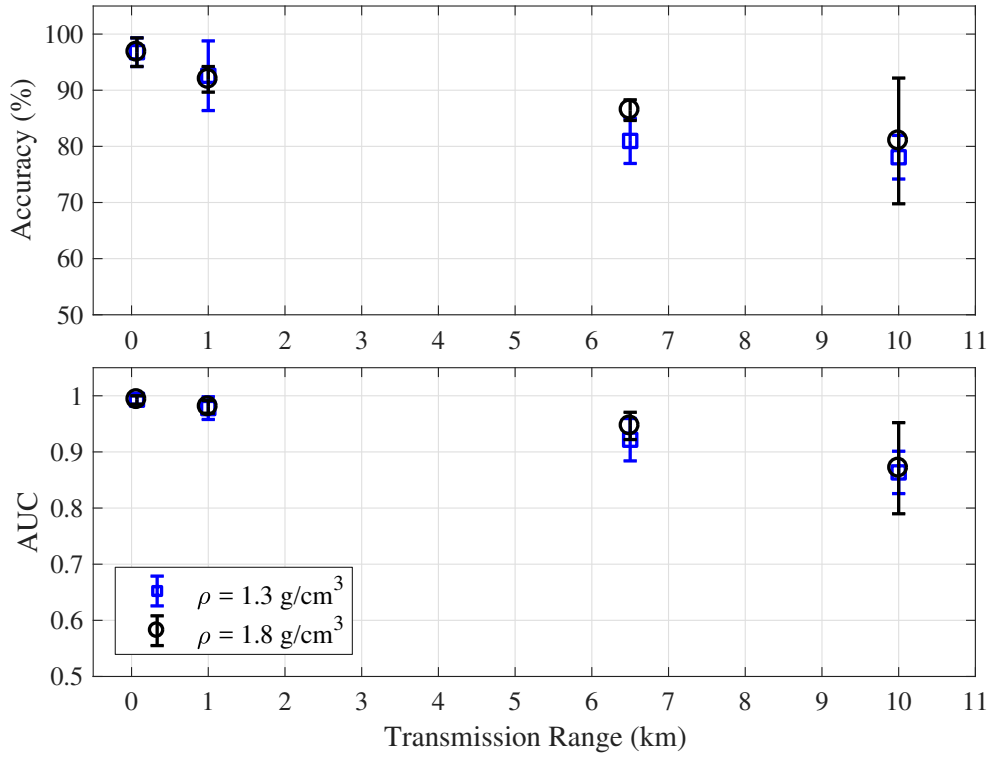


Figure E.2: Comparison of classifier performance as a function of range for sediment densities used for modelling throughout the thesis ($\rho = 1.3$ g/cm^3) and a typical value for silty-sand ($\rho = 1.8$ g/cm^3). The vertical errors bars show $\pm\sigma_{\text{accuracy}}$ and $\pm\sigma_{\text{AUC}}$ based on 5-fold cross-validation.

formula or the silty-sand density reported by Hamilton. A key finding of this Appendix was that no significant difference in the aural classifier performance results was attributed to a difference in sediment density.

BIBLIOGRAPHY

[1] D. K. Mellinger, K. M. Stafford, S. E. Moore, R. P. Dziak, and H. Matsumoto, "An overview of fixed passive acoustic observation methods for cetaceans", Oceanography **20**, 36–45 (2007).

[2] R. S. Sousa-Lima, T. F. Norris, J. N. Oswald, and D. P. Fernandes, "A review and inventory of fixed autonomous recorders for passive acoustic monitoring of marine mammals", Aquatic Mammals **39**, 23–53 (2013).

[3] S. M. V. Parijs, C. W. Clark, R. S. Sousa-Lima, S. E. Parks, S. Rankin, D. Risch, and I. C. V. Opzeeland, "Management and research applications of real-time and archival passive acoustic sensors over varying temporal and spatial scales", Marine Ecology Progress Series **395**, 21–36 (2009).

[4] A. Širović, A. Rice, E. Chou, J. A. Hildebrand, S. M. Wiggins, and M. A. Roch, "Seven years of blue and fin whale call abundance in the Southern California Bight", Endangered Species Research **28**, 61–76 (2015).

[5] M. F. Baumgartner, D. M. Fratantoni, T. P. Hurst, M. W. Brown, T. V. N. Cole, S. M. VanParijs, and M. Johnson, "Real-time reporting of baleen whale passive acoustic detections from ocean gliders", Journal of the Acoustical Society of America **134**, 1814–1823 (2013).

[6] A. M. Thode, K. H. Kim, S. B. Blackwell, C. R. G. Jr., C. S. Nations, T. L. McDonald, and A. M. Macrander, "Automated detections and localization of bowhead whale sounds in the presence of seismic airgun surveys", Journal of the Acoustical Society of America **131**, 3726–3747 (2012).

[7] "Species at risk public registry — species profile (humpback whale)", URL `http://www.registrelep-sararegistry.gc.ca/species/speciesDetails_e.cfm?sid=148#population`.

[8] R. Payne and S. McVay, "Songs of humpback whales", Science **173**, 585–597 (1971).

[9] A. Fleming and J. Jackson, "Global review of humpback whales (*Megaptera novaeangliae*)", NOAA Technical Memorandum NMFS NOAA-TM-NMFS-SWFSC-474, National Oceanic and Atmospheric Administration, National Marine Fisheries Service, Southwest Fisheries Science Center (2011).

[10] R. J. Urick, *Principles of Underwater Sound*, 3rd edition (McGraw-Hill Book Company) (1983).

[11] F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics*, AIP Series in Modern Acoustics and Signal Processing, 1st edition (American Institute of Physics) (1994).

[12] S. E. Dosso, P. M. Giles, G. H. Brooke, D. F. McCammon, S. Pecknold, and P. C. Hines, "Linear and nonlinear measures of ocean acoustic environmental sensitivity (L)", Journal of the Acoustical Society of America **121**, 42–45 (2007).

[13] J. A. Colosi, *Sound Propagation Through the Stochastic Ocean* (Cambridge University Press, New York, NY) (2016).

[14] W. W. Au and M. C. Hastings, *Principles of Marine Bioacoustics*, Modern Acoustics and Signal Processing Series (Springer, New York) (2008).

[15] E. Mercado and L. N. Frazer, "Environmental constraints on sound transmission by humpback whales", Journal of the Acoustical Society of America **106**, 3004–3016 (1999).

[16] C. S. Clay and H. Medwin, *Acoustical Oceanography*, Ocean Engineering: A Wiley Series, 1st edition (John Wiley & Sons, Inc.) (1977).

[17] M. J. Buckingham, "Ocean-acoustic propagation models", Journal Acoustique **43.30**, 233–287 (1992).

[18] T. A. Helble, G. L. D'Spain, G. S. Campbell, and J. A. Hildebrand, "Calibrating passive acoustic monitoring: Correcting humpback whale call detections for site-specific and time-dependent environmental characteristics", Journal of the Acoustical Society of America **134**, EL400–EL406 (2013).

[19] C. M. Binder and P. C. Hines, "Automated aural classification used for inter-species discrimination of cetaceans", Journal of the Acoustical Society of America **135**, 2113–2125 (2014).

[20] M. F. Baumgartner and S. E. Mussoline, "A generalized baleen whale call detection and classification system", Journal of the Acoustical Society of America **129**, 2889–2902 (2011).

[21] B. B. Bougher, J. D. Hood, J. Theriault, and H. Moors, "Generalized marine mammal detection based on improved band-limited processing", in *Proceedings of the 11th European Conference on Underwater Acoustics*, volume 17, 070067 (Edinburgh, UK) (2012).

[22] T. A. Helble, G. L. D'Spain, J. A. Hildebrand, G. S. Campbell, R. L. Campbell, and K. D. Heaney, "Site specific probability of passive acoustic detection of humpback whale calls from single fixed hydrophones", Journal of the Acoustical Society of America **134**, 2556–2570 (2013).

[23] X. Mouy, M. Bahoura, and Y. Simard, "Automatic recognition of fin and blue whale calls for real-time monitoring in the St. Lawrence", Journal of the Acoustical Society of America **126**, 2918–2928 (2009).

[24] S. E. Parks, I. Urazghildiiev, and C. W. Clark, "Variability in ambient noise levels and call parameters of North Atlantic right whales in three habitat areas", Journal of the Acoustical Society of America **125**, 1230–1239 (2009).

[25] Y. Simard and N. Roy, "Detection and localization of blue and fin whales from large-aperture autonomous hydrophone arrays: A case study from the St. Lawrence Estuary", Canadian Acoustics **36**, 104–110 (2008).

[26] W. M. X. Zimmer, J. Harwood, P. L. Tyack, M. P. Johnson, and P. T. Madsen, "Passive acoustic detection of deep-diving beaked whales", Journal of the Acoustical Society of America **124**, 2823–2832 (2008).

[27] E. Mercado, S. R. Green, and J. N. Schneider, "Understanding auditory distance estimation by humpack whales: A computational approach", Behavioural Processes **77**, 231–242 (2008).

[28] F. Samaran, C. Guinet, O. Adam, J.-F. Motsch, and Y. Cansi, "Source level estimation of two blue whale subspecies in southwestern Indian Ocean", Journal of the Acoustical Society of America **127**, 3800–3808 (2010).

[29] T. A. Marques, L. Thomas, S. W. Martin, D. K. Mellinger, J. Ward, D. J. Moretti, D. Harris, and P. L. Tyack, "Estimating animal population density using passive acoustics", Biological Reviews **88**, 287–309 (2013).

[30] E. R. Küssel, D. K. Mellinger, L. Thomas, T. A. Marques, D. Moretti, and J. Ward, "Cetacean population density estimation from single fixed sensors using passive acoustics", Journal of the Acoustical Society of America **129**, 3610–3622 (2011).

[31] E.-M. Nosal and L. N. Frazer, "Track of a sperm whale from delays between direct and surface-reflected clicks", Applied Acoustics **67**, 1187–1201 (206).

[32] D. M. Chapman, "You can't get there from here: Shallow water sound propagation and whale localization", Canadian Acoustics **32**, 167–171 (2004).

[33] K. M. Stafford, D. K. Mellinger, S. E. Moore, and C. G. Fox, "Seasonal variability and detection range modeling of baleen whale calls in the Gulf of Alaska, 1992–2002", Journal of the Acoustical Society of America **122**, 3378–3390 (2007).

[34] T. A. Helble, "Site specific passive acoustic detection and densities of humpback whale calls off the coast of California", Ph.D. thesis, University of California, San Diego (2013).

[35] V. W. Young and P. C. Hines, "Perception-based automatic classification of impulsive-source active sonar echoes", Journal of the Acoustical Society of America **122**, 1502–1517 (2007).

[36] N. Allen, P. C. Hines, and V. W. Young, "Performance of human listeners and an automatic aural classifier in discriminating between sonar target echoes and clutter", Journal of the Acoustical Society of America **130**, 1287–1298 (2011).

[37] C. M. Binder, "Using an aural classifier to discriminate cetacean vocalizations", Master's thesis, Dalhousie University, Halifax, NS (2012).

[38] S. M. Murphy and P. C. Hines, "Examining the robustness of automated aural classification of active sonar echoes", Journal of the Acoustical Society of America **135**, 626–636 (2014).

[39] R. Fawcett, "An introduction to ROC analysis", Pattern Recognition Letters **27**, 861–874 (2006).

[40] A. K. Stimpert, W. W. L. Au, S. E. Parks, T. Hurst, and D. N. Wiley, "Common humpback whale (*Megaptera novaeangliae*) sound types for passive acoustic monitoring", Journal of the Acoustical Society of America **129**, 476–482 (2011).

[41] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd edition (John Wiley & Sons, Inc.) (2001).

[42] R. Bellman, *Adaptive Control Processes: A Guided Tour*, 94 (Princeton University Press, Princeton, NJ) (1961).

[43] J. Ghosh, L. M. Deuser, and S. D. Beck, "A neural network based hybrid system for detection, characterization, and classification of short-duration oceanic signals", IEEE Journal of Oceanic Engineering **17**, 351–363 (1992).

[44] S. B. Kotsiantis, I. D. Zaharakis, and P. E. Pintelas, "Machine learning: A review of classification and combining techniques", Artificial Intelligence Review **26**, 159–190 (2006).

[45] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, chapter Chapter 9.6 Estimating and comparing classifiers, 482–495, 2nd edition (John Wiley & Sons, Inc.) (2001).

[46] C. D. Brown and H. T. Davis, "Receiver operating characteristics curves and related decision measures: A tutorial", Chemometrics and Intelligent Laboratory Systems **80**, 24–38 (2006).

[47] D. M. Green and J. A. Swets, *Signal detection theory and psychophysics*, 45–49 (John Wiley & Sons, Inc., New York, NY) (1966).

[48] D. W. Hosmer and S. Lemeshow, *Applied Logistic Regression*, 162, 2nd edition (Wiley and Sons, New York, NY) (2000).

[49] C. M. Binder and P. C. Hines, "Modelling the impact of ocean environment on automatic aural classification of marine mammals", in *Proceedings of the Underwater Acoustics 2014 Conference* (Underwater Acoustics 2014, Rhodos, GR) (2014).

[50] C. M. Binder, P. C. Hines, S. P. Pecknold, and J. Scrutton, "Robustness of perceptual features used for automatic aural classification to propagation effects", in *Proceedings of Meetings on Acoustics*, 010018 (International Congress on Acoustics, Montreal, QC, CA) (2013).

[51] S. Heimlich, D. Mellinger, and H. Klinck, "The MobySound database for research in automatic recognition of marine mammal calls", (2012), URL `http://www.mobysound.org/`.

[52] D. K. Mellinger and C. W. Clark, "MobySound: A reference archive for studying automatic recognition of marine mammal sounds", Applied Acoustics **67**, 1226–1242 (2006).

[53] J. D. Hood, D. G. Flogeras, and J. A. Theriault, "Improved passive acoustic band-limited energy detection for cetaceans", Applied Acoustics **106**, 36–41 (2016).

[54] A. Graps, "An introduction to wavelets", IEEE Computational Science and Engineering **2**, 50–61 (1995).

[55] L. R. LeBlanc and F. H. Middleton, "An underwater acoustic sound velocity data model", Journal of the Acoustical Society of America **67**, 2055–2062 (1980).

[56] G. R. North, T. L. Bell, R. F. Cahalan, and F. J. Moeng, "Sampling errors in the estimation of empirical orthogonal functions", Monthly Weather Review **110**, 699–706 (1982).

[57] "Model ITC-2010 broadband cylindrical transducer sell sheet", Specifications, International Transducer Corporation (2014).

[58] P. C. Hines, S. M. Murphy, D. A. Abraham, and G. B. Deane, "The dependence of signal coherence on sea-surface roughness for high and low duty cycle sonars in a shallow-water channel", IEEE Journal of Oceanic Engineering 1–21 (2016).

[59] Reach Technologies Inc., Victoria, BC, *Analog Data Recorder User's Manual*, 3.00 edition (2011).

[60] Wildlife Acoustics, Inc., *Song Meter SM2M Marine Recorder: User Manual* (2012).

[61] Ocean Sonics Ltd., Great Village, NS, *icListen HF user guide: icListen HF smart hydrophones* (2015).

[62] Ocean Sonics Ltd., Great Village, NS, *Specification icListen HF 200 kHz: Broadband Smart Hydrophone SB2/35-ETH* (2013).

[63] World Meterological Organization, Geneva, Switzerland, *Manual on Codes: International Codes Volume I.1 Annex II to the WMO Technical Regulations Part A — Alphanumeric Codes*, 2011 edition edition (2016).

[64] J. C. Osler, A. Furlong, H. Christian, and M. Lamplugh, "The integration of the free fall cone penetrometer (FFCPT) with the moving vessel profiler (MVP) for the rapid assessment of seabed characteristics.", International Hydrographic Review **7**, 45–53 (2006).

[65] C. L. Piggott, "Ambient sea noise at low frequencies in shallow water of the Scotian Shelf", Journal of the Acoustical Society of America **36**, 2152–2163 (1964).

[66] S. Steele and S. Pecknold, "Modelling backscatter at mid-frequencies in the northern Gulf of Mexico", Scientific Report DRDC-RDDC-2016-R112, Defence Research and Development Canada (2016).

[67] University of Washington Seattle Applied Physics Lab, "APL-UW high frequency ocean environmental acoustics models handbook", Technical Report APL-UW TR 9407, Applied Physics Laboratory University of Washington, Seattle, Washington (1994).

[68] W. L. Balsam and J. P. Beeson, "Sea-floor sediment distribution in the Gulf of Mexico", Deep Sea Research Part I: Oceanographic Research Papers **50**, 1421–1444 (2003).

[69] "Castaway CTD", URL `www.sontek.com/castaway`.

[70] C.-T. Chen and F. J. Millero, "Speed of sound in seawater at high pressures", Journal of the Acoustical Society of America **62**, 1129–1135 (1977).

[71] J. A. Knauss, *Introduction to Physical Oceanography*, 2nd edition (Prentice Hall, Upper Saddle River, NJ) (1997).

[72] A. E. Jochens, S. F. DiMarco, W. D. Nowlin, R. O. Reid, and M. C. K. II, "Northeastern Gulf of Mexico chemical oceanography and hydrography: Synthesis report", Technical Report MMS 2002-055, U.S. Department of the Interior Minerals Management Service, New Orleans, LA (2002).

[73] K. V. Mackenzie, "Nine-term equation for sound speed in the oceans", Journal of the Acoustical Society of America **70**, 807–812 (1981).

[74] L. E. Kinsler, A. R. Frey, A. B. Coppens, and J. V. Sanders, *Fundamentals of Acoustics*, 4th edition (John Wiley & Sons, Inc.) (1999).

[75] R. J. Urick, "Generalized form of the sonar equations", Journal of the Acoustical Society of America **34**, 547–550 (1962).

[76] C. W. Helstrom, *Statistical Theory of Signal Detection*, chapter IV: Detection of a known signal, 84–128, International Series of Monographs on Electronics and Instrumentation (Pergamon Press) (1960).

[77] C. E. Cook and M. Bernfeld, *Radar Signals: An Introduction to Theory and Application*, chapter 1: The Basic Elements of Matched Filtering and Pulse Compression, 1–17, Electrical Science Series (Academic Press, Inc.) (1967).

[78] E. L. Hamilton, "Geoacoustic modeling of the sea floor", Journal of the Acoustical Society of America **68**, 1313–1340 (1980).

[79] S. M. Murphy and P. C. Hines, "SNR dependence and temporal robustness of an automatic aural classifier", in *Conference Proceedings of the Underwater Acoustics Measuremenets Conference* (Kos, GR) (2011).

[80] C. E. Shannon, "A mathematical theory of communication", The Bell System Technical Journal **27** (1948).

[81] C. M. Bishop, "Neural networks and their applications", Review of Scientific Instruments **65**, 1803–1832 (1994).

[82] P. C. Etter, *Underwater Acoustic Modeling: Principles, Techniques and Applications*, 1st edition (Elsevier Applied Science) (1991).

[83] P. Giles, S. Kilistoff, and G. Brooke, "Portable Acoustic Sensitivity Transmission Evaluation Tool — PASTET", Contract Report CR 2008-299, Defence Research and Development Canada (2009).

[84] P. C. Etter, "Advanced applications for underwater acoustic modeling", Advances in Acoustics and Vibration **2012**, 1–28 (2012).

[85] P. C. Hines, J. C. Osler, J. G. E. Scrutton, and L. J. S. Halloran, "Time-of-flight measurements of acoustic wave speed in a sandy sediment at 0.6–20kHz", IEEE Journal of Oceanic Engineering **35**, 502–515 (2010).

[86] C. W. Holland and C. H. Harrison, "Measurement of the seabed reflection coefficient in shallow water: A comparison of two techniques", in *Theoretical and Computational Acoustics 2003*, edited by A. Tolstoy, Y.-C. Teng, and E. Shang, 178–191 (World Scientific Publishing Co., Singapore) (2003).

[87] S. E. Crocker, J. H. Miller, G. R. Potty, J. C. Osler, and P. C. Hines, "Nonlinear inversion of acoustic scalar and vector field transfer functions", IEEE Journal of Oceanic Engineering **37**, 589–606 (2012).

[88] S. Ivansson and I. Karasalo, "Ocean acoustics library: Wavenumber integration (FFP)", URL `http://oalib.hlsresearch.com/FFP/index.html`.

[89] Y. Xian, "Detection and classification of whale acoustic signals", Ph.D. thesis, Duke University (2016).

[90] M. B. Porter and Y.-C. Liu, "Finite-element ray tracing", in *Proceedings of the International Conference on Theoretical and Computational Acoustics*, edited by D. Lee and M. H. Schultz, 947–956 (1994).

[91] C. L. Pekeris, *Theory of propagation of explosive sound in shallow water*, volume 27 of *Memoir Series*, chapter Part II: Theory — A. Theory of propagation of sound in water underlain by a uniform bottom of different density and sound velocity, 43–64 (The Geological Society of America) (1948).

[92] D. D. Ellis, "A shallow-water normal-mode reverberation model", Journal of the Acoustical Society of America **97**, 2804–2814 (1995).

[93] H. Schmidt, *SAFARI Seismo-Acoustic Fast Field Algorithm for Range-Independent Environements User's Guide*, SACLANT Undersea Research Centre (1988).

[94] H. Schmidt, *OASES Version 3.1 User Guide and Reference Manual*, Department of Ocean Engineering Massachusetts Institute of Technology (2011).

[95] H. Schmidt and F. B. Jensen, "A full wave solution for propagation in multilayered viscoelastic media with application to Gaussian beam reflection at fluid-solid interfaces", Journal of the Acoustical Society of America **77**, 813–825 (1985).

[96] F. DiNapoli and R. Deavenport, "Theoretical and numerical Green's function field solution in a plane multilayered medium", Journal of the Acoustical Society of America **67**, 92–105 (1980).

[97] F. B. Jensen, "Wave theory modeling: A convenient approach to CW and pulse propagation modeling in low-frequency acoustics", IEEE Journal of Oceanic Engineering **13**, 186–197 (1988).

[98] J. A. Theriault and S. Pecknold, "Impulse propagation using WATTCH", External Client Report ECR 20004-248, Defence R&D Canada — Atlantic (2006).

[99] S. Pecknold, J. A. Theriault, D. McGaughey, and J. Collins, "Time-series modeling using the waveform transmission through a channel program", in *Oceans 2005 — Europe*, volume 2, 993–1000 (IEEE) (2005).

[100] M. B. Porter and H. P. Bucker, "Gaussian beam tracing for computing ocean acoustic fields", Journal of the Acoustical Society of America **82**, 1349–1359 (1987).

[101] M. B. Porter, *The BELLHOP manual and user's guide*, Heat, Light, and Sound Research, Inc., La Jolla, CA, USA (2011).

[102] D. F. McCammon, "Users guide to bellhopdrdc-active_v5 — active v5 and updated passive v4a", Contract Report CR 2010-319, Defence R&D Canada — Atlantic (2011).

[103] S. Pecknold and J. C. Osler, "Sensitivity of acoustic propagation to uncertainties in the marine environment as characterized by various rapid environmental assessment methods", Ocean Dynamics **62**, 265–281 (2012).

[104] W. H. Thorp, "Analytic description of the low-frequency attenuation coefficient", Journal of the Acoustical Society of America **42**, 270 (1967).

[105] W. H. Thorp, "Deep-ocean sound attenuation in the sub- and low-kilocycle-per-second region", Journal of the Acoustical Society of America **38**, 648–654 (1964).

[106] J. Potter and A. Warn-Varnas, eds., *Ocean variability & acoustic propagation* (Springer Netherlands) (1991).

[107] N. G. Pace and F. B. Jensen, eds., *Impact of littoral environmental variability on acoustic predictions and sonar performance* (Kluwer Academic Publishers, Dordecht, The Netherlands) (2002).

[108] E. S. Livingston, J. A. Goff, S. Finette, P. Abbot, J. F. Lynch, and W. S. Hodgkiss, "Guest editorial capturing uncertainty in the tactical ocean environment", IEEE Journal of Oceanic Engineering **31**, 245–248 (2006).

[109] S. E. Dosso, M. G. Morley, P. M. Giles, G. H. Brooke, D. F. McCammon, S. Pecknold, and P. C. Hines, "Spatial field shifts in ocean acoustic environmental sensitivity analysis", Journal of the Acoustical Society of America **122**, 2560–2570 (2007).

[110] P. Giles, "Geoacoustic sensitivity study: Phase I: Literature review", Contract Report CR 2006-048, Defence R&D Canada — Atlantic (2006).

[111] T. A. Helble, G. R. Ierley, and G. L. D'Spain, "A generalized power-law detection algorithm for humpback whale vocalizations", Journal of the Acoustical Society of America **131**, 2682–2699 (2012).

[112] W. W. L. Au, A. A. Pack, M. O. Lammers, L. M. Herman, M. H. Deakos, and K. Andrews, "Acoustic properties of humpback whale songs", Journal of the Acoustical Society of America **120**, 1103–1110 (2006).

[113] L. N. Frazer and E. Mercado, "A sonar model for humpback whale song", IEEE Journal of Oceanic Engineering **25**, 160–182 (2000).

[114] H. M. Merklinger and J. H. Stockhausen, "Formulae for estimation of undersea noise spectra", in *Proceedings of the 50th Meeting of the Acoustical Society of America*, volume 65, S88 (1979).

[115] T. A. Marques, L. Thomas, J. Ward, N. DiMarzio, and P. L. Tyack, "Estimating cetacean population density using fixed passive acoustic sensors: An example with Blainville's beaked whales", Journal of the Acoustical Society of America **125**, 1982–1994 (2009).

[116] G. L. D'Spain and W. A. Kuperman, "Application of waveguide invariants to analysis of spectrograms from shallow water environments that vary in range and azimuth", Journal of the Acoustical Society of America **106**, 2454–2468 (1999).

[117] L. M. Brekhovskikh and Y. P. Lysanov, *Fundamentals of Ocean Acoustics*, 140–145, Springer Series on Wave Phenomena, 2nd edition (Springer-Verlag) (1991).

[118] A. J. Collier and A. W. George, "Acoustic research vessel CFAV QUEST", in *IEEE Oceans 81 Boston, MA*, 1038–1043 (IEEE) (1981).