DATA COMPRESSION OF BLAINVILLE'S BEAKED WHALE VOCALIZATIONS

LA COMPRESSION DE VOCALIZATIONS DE BALEINE À BEC DE BLAINVILLE

A Thesis Submitted

to the Division of Graduate Studies of the Royal Military College of Canada

by

Jennifer Krista Stamplecoskie, B.Eng.
Captain

In Partial Fulfillment of the Requirements for the Degree of

Master of Applied Science in Signal Processing

April, 2009

Your file Votre référence
ISBN: 978-0-494-52296-7
Our file Notre référence
ISBN: 978-0-494-52296-7

Acknowledgments

I would like to thank my supervisors, Dr. Donald R. McGaughey and Dr. Michael J. Korenberg, for their guidance and enthusiasm toward everything FOS related. Your confidence in my ability to complete this task kept me from feeling too over-whelmed, and your patience when answering the same question for the fifth time was admirable. I would also like to thank Dr Jim Theriault of DRDC Atlantic and Mr Joe Hood, president of Akoostix for providing me with both the data used as a basis for this algorithm, and for their guidance in the area of detection of underwater acoustic marine mammal activity. I would like to thank my partner in Starbucks, Megan Rakoczy. Sharing our pain as we made it through late nights and weeks with no end made the process much more bearable, as did our in-office latte machine. Lastly, I would like to thank my husband Sean for his support during this journey. Thank you for supporting my decision to move away and go to school full-time while you kept up our life at home, I'll be home soon.

Abstract

Stamplecoskie, Jennifer Krista. M.A.Sc. Royal Military College of Canada, April, 2009. *Data Compression Of Blainville's Beaked Whale Vocalizations.* Supervised by Dr. Donald R. McGaughey, Dr. Michael J. Korenberg.

Defence Research and Development Canada Atlantic uses autonomous underwater vehicles and sonobuoys to capture recordings of nearby marine mammal activity. The data are then transmitted via a low bandwidth satellite link. Timely transmission of these data is critical in achieving compliance with the Naval directive to check marine mammal activity within the vicinity of training exercises before their commencement. Currently, the data are not compressed, and timely transmission is not being achieved.

This thesis uses the signal processing technique called the fast orthogonal search to achieve data compression. The fast orthogonal search algorithm has previously been shown capable of providing model expansion representation of both time and frequency domain signals. Furthermore, irregular sampling and domain transformation have been successfully used for data reconstruction. The proposed algorithm, patented by Korenberg et al [1], has previously been used for interferometric spectrometer data.

Two separate versions of the algorithm were used, one for use with sonobuoys, which have more processing capabilities, and one more suitable for autonomous underwater vehicles which have limited on-board processing capability. These data compression techniques allow for a significant reduction in transmission time of the marine activity data while retaining the information necessary for detection.

Resumé

Stamplecoskie, Jennifer Krista. M.Sc.A. Collège militaire royal du Canada, Avril, 2009. *La Compression De Vocalizations De Baleine à Bec de Blainville.* Thèse dirigée par M. Donald R. McGaughey, Ph.D. et M. Michael J. Korenberg

La Recherche et le développement pour la défense du Canada, secteur Atlantique, enregistre les activités des mammifères marins environnants. Les données sont présentement transmises par lien satellite ayant une faible largeur de bande. La transmission de ces données en temps opportun est critique afin de comformer avec la directive maritime qui oblige à localiser les activités des mammifères marins avant de commencer des exercises d'entraînement. Présentement, les données ne sont pas compressées, et la transmission en temps opportun n'est pas atteinte.

Cette thèse utilise la technique de traitement des signaux appellée algorithme de recherche orthogonale rapide pour atteindre la compression de données. L'algorithme de recherche orthogonale rapide a déjà démontré sa capacité à fournir une représentation des signaux du domaine temporel et fréquenciel, utilisant un modèle de développement. De plus, l'échantillonnage irrégulier et la transformation des domaines ont été utilisés avec succès pour reconstituer les données. L'algorithme proposé, breveté par Korenberg et al [1], a été déjà utilisé pour les données de spectromètre interférométrique.

Deux versions de l'algorithme ont été utilisées, l'une est plus applicable aux bouées acoustiques, qui ont plus de capacité de traitement et l'autre est plus attribué aux sous-marins autonomes, qui ont une capacité de traitement plus limitée. Ces techniques de compression de données permettent une réduction significative du temps de transmission de données en gardant les données importantes pour la détection.

Table of Contents

List of Abbreviations

AUV autonomous underwater vehicle

AWGN additive white Gaussian noise

BW bandwidth

CF Canadian Forces

CR compression ratio

DC direct current

DCT discrete cosine transform

DFT discrete Fourier transform

DND Department of National Defence

DRDC Defence Research and Development Canada

ECG electrocardiogram

FFT fast Fourier transform

FN false negative

FOS fast orthogonal search

FP false positive

FPGA field programmable gate array

IDCT inverse discrete cosine transform

JPEG Joint Photographic Experts Group

LOFARgram low frequency analysis record diagram

MPEG Moving Picture Experts Group

MSE mean square error

NPV negative predictive value

SNR signal-to-noise ratio

PCA principal component analysis

PPV positive predictive value

PSD power spectral density

ROC receiver operating characteristic

STD standard deviation

TN true negative

TP true positive

WGN white Gaussian noise

Chapter 1

Introduction

1.1 Problem

Some of the Naval training exercises currently employed have the potential to harm marine mammals within their vicinity. As such it is a policy of the Canadian Forces (CF) Navy to refrain from conducting training activities when there is marine mammal activity within four kilometers of the training area [2]. However, while this policy has been instituted, it is currently enforced in a limited manner because there is no vigorous system in place to verify if there are marine mammals present. The CF Navy relies on historical data and visual clues to determine if marine mammals are present within the training area. A more robust method is desired.

The Department of National Defence (DND) and Defence Research and Development Canada (DRDC) Atlantic have a series of autonomous underwater vehicles (AUVs) and sonobuoys which contain hydrophones. The hydrophones record the underwater acoustic signals, which are then transmitted by the AUVs and sonobuoys. The signals can be received and analyzed to determine the presence of marine mammal activity in the vicinity of the AUVs or sonobuoys.

With some modifications these vehicles could be used in order to detect the presence of marine mammal activity in the vicinity of training missions, allowing the CF Navy to improve its ability to abide by its training policy.

1.2 Objectives

The objective of this work is to improve the efficiency with which data are transmitted from the AUVs and sonobuoys across a low bandwidth (BW) satellite link without significantly degrading the ability to identify marine mammal activity within the signal being transmitted. This is to be measured using the current detection algorithm employed by DRDC Atlantic.

1.3 Current Techniques

Underwater acoustical data have been received from DRDC Atlantic which they collect using three separate platforms. The first platform is a sonobuoy which always has an antenna situated on the ocean surface. This platform is capable of receiving acoustic data and transmitting the recorded data concurrently, and it has 160 GB of memory. The second and third platforms are both AUVs. One is fully mobile, while the other can only rise to the surface and sink using an inflatable balloon.

The AUV platforms must both surface in order to transmit data, and therefore cannot record data while transmitting. They have on-board memory of 16 GB. They also have acoustical detection software on-board, and can currently transmit one of three data sets. The first is a simple text message giving the coordinates and type of detection. The second data set includes the same text message along with 30 seconds of raw data. This can be used to determine whether the detection was a false alarm or not. The third data set is a 5 minute recording. The third data set is preferred for trial use at DRDC Atlantic, and so it is the set which will be used for the purposes of this thesis.

For all the platforms, the acoustic data are recorded at a rate of 96 kHz, producing a 54.93 MB recording, for a 5 minute data capture. The recordings are transmitted over a low BW satellite link to the Iridium satellite system at a rate of 4800 baud, requiring 100 minutes per transmission. Additionally, no data compression scheme is currently implemented by DRDC Atlantic.

1.4 Relevant Work

Both principal component analysis (PCA) and fast orthogonal search (FOS) are techniques which use least squares fitting to construct a functional expansion of a given signal. The performance of both techniques is dependent on the validity of the algorithm's input candidates. The limitation of PCA is that it is restricted to the inclusion of orthogonal candidate sets, where as FOS can include both orthogonal and non-orthogonal candidate sets. In 1996, Oliveira and Romera [3] used PCA to

perform data compression on medical image data. They found that PCA gave a smaller mean square error (MSE) than Joint Photographic Experts Group (JPEG). This work was further confirmed in 2000 by Oliveira and Mazucheli [4]. In 2006, Oliveira [5] extended this research to show that PCA can be used to interpolate data when the data have been irregularly sampled. He has shown this to be possible using both one-dimensional signal and image data. This work indicates that PCA is a viable method of interpolating underwater acoustic data. This hypothesis can be extended to FOS which follows the same principles as PCA.

FOS is similar to the later developed orthogonal matching pursuit [6]. It is also related to, but faster than an algorithm previously published by Desrochers for fitting static nonlinear models where, amongst various differences, the computational and memory storage requirements are proportional to the square of the number of candidates, whereas in FOS they depend linearly on the number of candidates [7].

Korenberg et al [1] have patented an algorithm which uses domain transformation and irregular sampling with reconstruction of the data accomplished through the use of FOS. This technique has previously been used for interferometric spectrometer data [1]. Similarly, Abbas [8] used this technique to compress and reconstruct electrocardiogram (ECG) data. Like marine mammal activity, ECG signals have specific features which must be preserved during compression. It is reasonable to assume that this technique may have similar success with underwater acoustic data.

1.5 Deficiencies

The AUV platforms must surface in order to transmit data. This means that they are unable to record data for the duration of their transmission. Shortening the transmission length would enable these platforms to record underwater acoustic signals for a higher percentage of the time when deployed. For all platforms, data compression would improve the speed at which the data can be received by the operators for post-processing, allowing it to be used for current detection of marine mammal activity, in fulfillment of the naval directive. Furthermore, the AUV platforms have limited processing capabilities and limited battery life.

1.6 Solution

For the sonobuoy, which has a larger on-board processor and fewer power restrictions, a technique was developed in which a time-domain model expansion is performed using FOS. The parameters of this model are then transmitted from the sonobuoy to the receiver station. As long as the receiver station possesses the FOS candidates, the signal can be reconstructed upon receipt. This technique achieves a compression ratio (CR) of approximately 93.1, although the exact compression ratio will depend on the amount of marine mammal activity present as more marine activity causes an increase in the number of model terms required to describe the signal.

The second technique does not achieve as high a compression ratio, however, it is more suitable for the AUVs which have a limited on-board processing capability and more power restrictions and achieves a CR of 5.5. In this technique, domain transformation is performed via discrete cosine transform (DCT) and the resulting data are irregularly sampled. The selected frequency domain samples are transmitted. Upon receipt, reconstruction can be performed using FOS at the receiver station, if the receiver station knows the irregular sampling pattern used to transmit the data.

1.7 Outline

Chapter 2 of this thesis provides the background theory pertaining to the proposed algorithm. JPEG is introduced as a data compression standard for non-text-based data. It is used as a basis for the algorithms developed in this work. The detection algorithm used to assess the quality of the compressed data is also explained.

In Chapter 3 a detailed description of the time-domain and frequency-domain algorithms is given. The time-domain algorithm is used for the sonobuoy platform, and the frequency-domain algorithm is used for the AUV platforms.

Chapter 4 provides the detection algorithm parameters used and a description of the evaluation criteria used to evaluate the algorithms in terms of the detection capabilities.

Chapter 5 provides the results of the simulations performed, including the compression ratios achieved, and the effect of the compression on signal detection.

Finally, chapter 6 provides a summary of the work performed and recommendations for future work in this area.

Chapter 2

Background Theory

2.1 Underwater Acoustic Signals

Not only are acoustic waves better suited to propagation in oceanic waters than electromagnetic waves [9], but they also propagate better in the ocean than they do in the air [10]. For this reason, underwater acoustic data are typically used to increase our understanding of the oceanic environment. The CF Navy has a particular interest in tracking marine mammal activity in the vicinity of their assets. Marine life, particularly mammals, can be detected through the vocalizations that they make. Since sound waves have the best ability to propagate in the water, marine mammals depend on sound as a good way of detecting food. Each species has a distinct set of vocalizations that it makes, however, there are some commonalities between them. This thesis concentrates on the detection of whale vocalizations and uses whale vocalization recordings provided by DRDC Atlantic.

2.1.1 Detection of Marine Mammal Sounds

The ocean is a noisy environment, with a persistent ambient noise level. This may be caused by any number of sources including tides, seismic disturbances, turbulence, ship traffic, as well as some sources which are specific to regions, such as shallow or deep water disturbances or coastal disturbances. In addition to the ambient noise, there will be noise caused by both the hydrophone and the platform in which it is contained [9].

Underwater acoustic data are collected in the temporal domain. An indication that there has been marine mammal activity is a temporary increase in the amplitude of the signal over the average background level. This increase will last anywhere from a few milliseconds to longer than 30 seconds, depending on the type of vocalization and the species of mammal. However, since the ocean is a noisy environment it can

be quite difficult to detect marine mammal activity in the time domain.

In the spectral domain, using a tool such as a low frequency analysis record diagram (LOFARgram), the change in frequencies detected over a period of time can indicate a signal of interest. An example of a LOFARgram is shown in Fig 2.1. It shows the changes over time in the energy at given frequencies. The fast Fourier transform (FFT) has been applied over a set of 1024 samples at a time. There is a 50 % overlap between the sets. Using this technique it is posible to see changes over time in the energy at given frequencies. The energy level is indicated by a change in intensity of the LOFARgram.

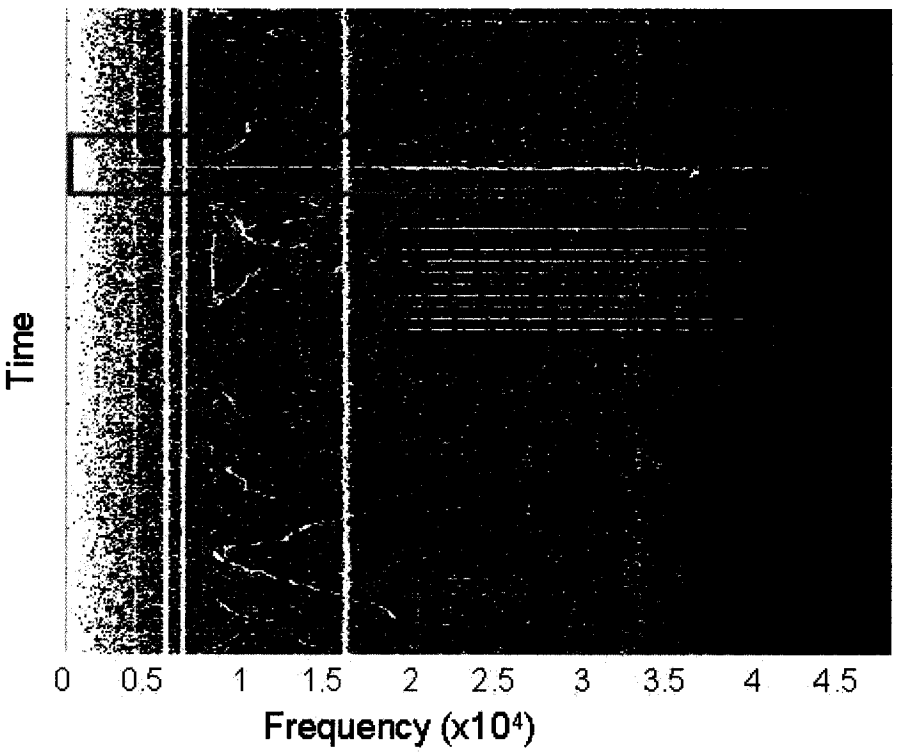

Figure 2.1: This LOFARgram of Blainville's Beaked whales shows a series of buzzes as well as a portion of a whale song. The buzzes can be seen as the bright lines between 20 and 40 kHz. A Humpback whale song is visible between 8 and 18 kHz. The three vertical lines are caused by the recording platform.

Once detected, classification of the signals can be performed in either domain, depending on the technique used. The detection algorithm used in this thesis transforms the signal into the frequency domain.

2.1.2 Characteristics of Whale Vocalizations

Whales produce a wide variety of vocalizations. Each species of marine mammal has a set of features which distinguish it.

The whale which is considered in this thesis is Blainville's Beaked whale. These are the whale calls which were provided by DRDC Atlantic. Blainville's Beaked whales are hard to see when they surface as they do not have visible blows. They also dive for long periods, and so it can be difficult to determine their presence [11]. They produce clicks and buzzes. The clicks last for approximately 0.271 ms with a frequency range between 26 and 51 kHz and a frequency modulated upsweep of approximately 112 kHz/ms [12]. Buzz clicks tend to be shorter, with an average duration of 0.104 ms and a frequency range of 25 to upwards of 80 kHz. They are not known to be frequency modulated [12]. Some recorded buzz clicks have less total energy than others because they are recorded off the body axis of the whale [13]. They tend to have a higher energy content at $30 - 35$ kHz and have less energy beyond this [12].

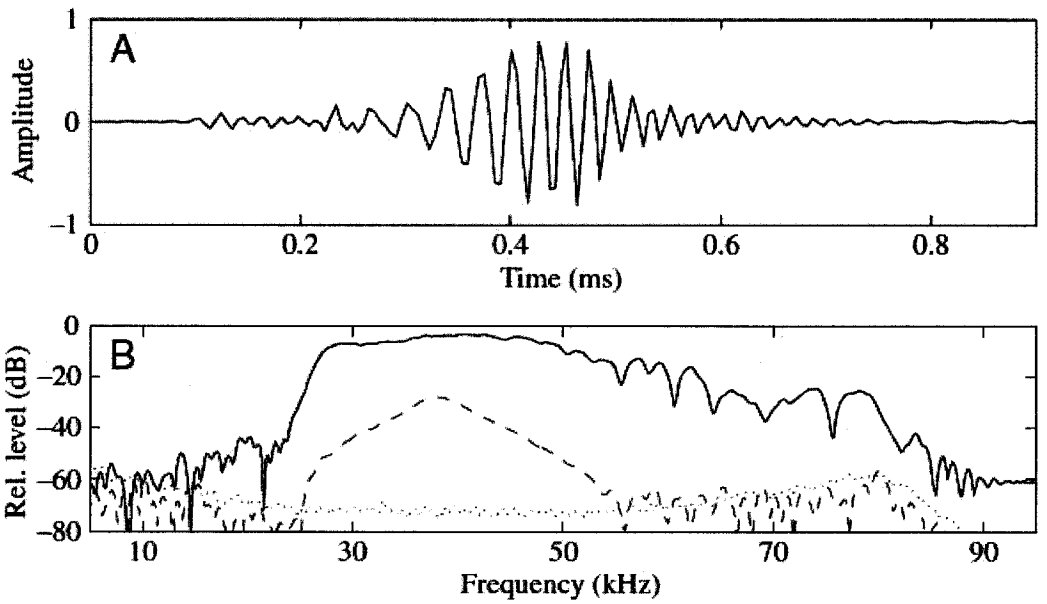

Figure 2.2: In a) a buzz is shown in the time domain. In b) a frequency domain representation is given, where the solid line represents the buzz energy [14].

2.2 Compression of Acoustic Signals

Data compression techniques are commonly compared using the metric CR. The CR is a measure of how much the signal has been compressed. CR is defined as:

$$CR = \frac{Size_{uncompressed}}{Size_{compressed}} \tag{2.1}$$

Data compression standards can either be lossless or lossy, although some compression standards are capable of both types of compression.

2.2.1 Lossless Compression

Lossless compression is data compression whereby none of the values in the signal are changed or removed. Using a lossless technique, a signal can be perfectly reconstructed when decompressed. Lossless compression generally can achieve a CR of 2 or 3 [15]. While necessary for some applications, such as medical imaging, the small CR achieved with this type of compression limits its utility.

2.2.2 Lossy Compression

Lossy compression is a data compression strategy which permits permanent changes to the signal. During lossy compression, some of the information stored in the signal is lost. As a result, when decompressed, the signal cannot be perfectly reconstructed. This is the reason why an image degrades as it is repeatedly re-saved in a lossy format such as JPEG.

The advantage of this type of compression is that it can achieve much higher CRs while producing signals that are of acceptable quality for their intended use. The statistical properties of signals, when combined with the characteristics of the exploiting system are the factors which allow reconstructed signals to be greatly compressed yet appear to be indistinguishable from their originals [16].

Since underwater acoustic signals have a large noise component, absent of useful information, it is reasonable to assume that lossless reconstruction of the signal is not

efficient, nor is it necessary. Thus, so long as the features of interest are retained, lossy compression is a valid compression technique for underwater acoustic signals.

2.3 Compression Standard - MPEG

The first international compression standard for audio data is Moving Picture Experts Group (MPEG) [16]. MPEG audio compression can be performed using an audio sampling rate of 32, 44.1, or 48 kHz [16]. This sampling rate is too low to retain the information of interest for the compression of Blainville's beaked whale vocalizations. Retention of information up to 48 kHz is desired (requiring a sampling rate of 96 kHz). As a result, MPEG was not considered for this application.

Sometimes large data compression can be achieved simply by subsampling an over-sampled record. However, in the present study the sampling rate had not been excessive since the clicks and buzzes only last 7-18 samples, and the frequencies of interest extend up to the Nyquist frequency of 48 kHz.

2.4 Compression Standard - JPEG

A commonly used compression standard for imagery is JPEG [17]. Like imagery, underwater acoustic data can be compressed by taking advantage of the properties of the signal itself. The JPEG compression standard follows a series of steps starting with level-shifting.

2.4.1 Level Shifting

Level shifting is the process of taking the sample or pixel information and shifting it so that it varies about zero. If the pixel values vary between 0 and 255 then, the level-shifted values would be shifted to vary between -128 and $+127$. Following level-shifting, sub-sampling is performed.

2.4.2 Sub-Sampling

Sub-sampling or down-sampling is the process of reducing the number of sample points in a signal. It is an optional step of JPEG.

2.4.3 Block Splitting

Once sub-sampling has been performed, the image is split into blocks. If the number of pixels in the image is not an integer multiple of the block size, additional pixels are appended. These steps are performed to prepare the data for a domain transformation. JPEG uses the DCT to perform domain transformation.

2.4.4 Discrete Cosine Transform

Once the image has been subdivided, the 2D DCT is performed separately on each block. In this step, the image data are transformed from the spatial domain into the spectral domain. Little information is lost, and no compression is achieved during this step. The purpose of this transform is to put the data in a domain where compression is easier. When the 2D DCT is performed on each block of the image, each resulting block consists of integer values corresponding to the coefficients of its DCT candidate functions. There are 64 candidate functions used to represent the blocks in the spectral domain.

After the 2D DCT has been performed on each block of the image, the DCT coefficients represent the values which would be multiplied by their respective DCT candidate function to recreate that block of the image in the spatial domain. The DCT coefficients are the same size as the pixel values (i.e. up to 8 bits per pixel per channel), but fewer of the candidate functions have significant weights. The information lost in this step is partially due to the requirement to maintain integer values which can be represented in 8 bits, and partially because the image may not be fully described in 64 DCT candidates.

2.4.5 Quantization

Quantization is a lossy step in image compression. Typically performed after a transformation, quantization allows the data to be encoded using fewer bits per pixel. It works by removing the data least likely to be noticed. The most basic quantization scheme involves dividing each scalar value by a constant. This strategy reduces the maximum value to be encoded, thereby possibly reducing the number of bits required to represent the value in the encoding scheme. It also means that the scalar values will be reconstructed as approximations of what they were, thus reducing the quality of the image when reconstructed.

2.5 Time Domain Algorithm

The first algorithm used in this thesis is suitable for the sonobuoy platform, which has more processing power than the AUV platforms. It modifies the compression standard for imagery to make it suitable for underwater acoustic data as seen in Fig 2.3.

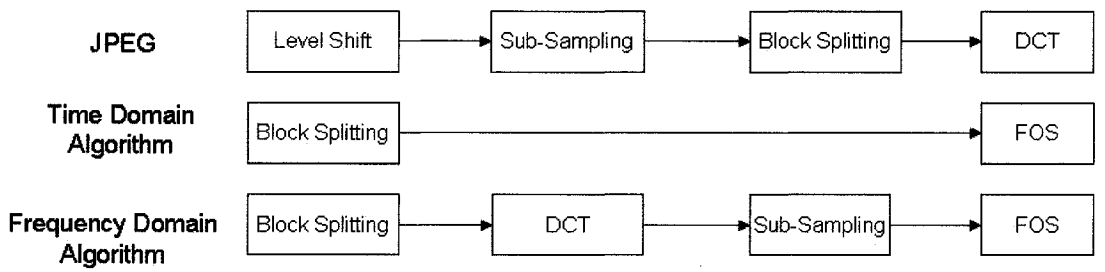

Figure 2.3: Similarities and differences between the JPEG algorithm for image compression, the time domain and frequency domain algorithms for underwater acoustic data compression. Each algorithm transmits only model terms and coefficients, resulting in compression.

The acoustic data are split into segments prior to completion of the remaining steps. For this algorithm, no sub-sampling is performed. Finally, rather than using the DCT to transform the acoustic signal into the spectral domain, FOS is used to create a functional expansion in the temporal domain. Like the DCT, the FOS

functional expansion requires fewer terms to describe each segment than the number of samples in the original segment, resulting in data compression.

2.6 Frequency Domain Algorithm

The frequency domain algorithm does not achieve as high a CR as the time domain algorithm, however it is better suited to the AUV platforms which have a limited processing capability. The first step is the same as the time domain algorithm. The data are split into blocks. Following this, a domain transformation is performed using the DCT. Irregular sampling is then applied to each block in the frequency domain.

The domain transformation prior to the sampling ensures that short-time duration signals will not be entirely removed when irregular sampling is performed. Additionally, the irregular sampling minimizes the loss of high frequency signal components.

Once the irregular sampling has been performed, the remaining samples are transmitted. Data compression is achieved because not every sample is transmitted. At the receiver station, with knowledge of the sampling pattern used, FOS can be used to reconstruct the frequency domain signal. This signal can then be transformed back into the time domain as required.

2.7 Fast Orthogonal Search in Temporal Domain

The FOS algorithm is an alternative to the discrete Fourier transform (DFT) or DCT for performing domain transformation. It can achieve a much higher resolution than the DFT by selecting non-orthogonal candidate functions [18]. It can also detect signals present in noise. Chon [19] used FOS to detect sinusoids at a signal-to-noise ratio (SNR) of -10 dB. Furthermore, Korenberg and Paarmann [20] successfully used FOS to reconstruct signals in noise for irregularly sampled signals.

FOS has been used extensively in the time domain to provide a functional expansion representation of a given signal. The candidate functions successfully used include sinusoidal and exponential candidates. Chon [19] and Korenberg [21] used sinusoidal candidates to detect a sinusoidal signal in noise. Likewise, Korenberg [18],

Korenberg and Paarmann [20] and Korenberg, Brenan and Hunter [22] used sinusoidal candidates to detect a sinusoidal signal with irregular sampling and in noise. Korenberg and Adeney [23] showed the ability of FOS to deal with exponential candidates to fit an exponential signal in noise.

The results of the FOS algorithm are a functional expansion given by [23]:

$$y(n) = \sum_{m=0}^{M-1} a_m p_m(n) + e(n) \tag{2.2}$$

where $y(n)$ is the input signal, a_m are the non-orthogonal weights, $p_m(n)$ are the non-orthogonal candidate functions, and $e(n)$ is the error.

The non-orthogonal functional expansion is determined from an orthogonal function expansion. The orthogonal expansion is given by [20]:

$$y(n) = \sum_{m=0}^{M-1} g_m w_m(n) + e(n) \tag{2.3}$$

where g_m are the orthogonal weights, and $w_m(n)$ are the orthogonal candidates. FOS takes the set of arbitrary candidate functions, $p_m(n)$ and uses them to determine a set of orthogonal functions using the Gram-Schmidt orthogonalization technique

The orthogonal weights, g_m, are given by:

$$g_m = \frac{C(m)}{D(m,m)} \tag{2.4}$$

where $C(m) = \overline{y(n)w_m(n)}$ is the time average (denoted by the overbar) given by:

$$\overline{x} = \frac{1}{N} \sum_{n=0}^{N-1} x(n) \tag{2.5}$$

of the correlation between the orthogonal function and the input function and $D(m,m) = \overline{p_m(n)w_m(n)}$ is the time average of the common energy between the orthogonal and non-orthogonal candidate functions. The common energy between two candidate functions can be determined recursively using [20]:

$$D(m,r) = \overline{p_m(n)p_r(n)} - \sum_{i=0}^{r-1} \alpha_{r,i} D(m,i) \tag{2.6}$$

It can be shown using Eqn 2.6 that $D(m,m) = \overline{w_m^2(n)}$ and the $\alpha_{m,r}$ are the Gram-Schmidt coefficients given by [20]:

$$\alpha_{m,r} = \frac{D(m,r)}{D(r,r)} \tag{2.7}$$

The correlation between an orthogonal function and the input function is determined recursively using:

$$C(m) = \overline{y(n)p_m(n)} - \sum_{i=0}^{m-1} \alpha_{m,i} C(i) \tag{2.8}$$

FOS uses least squares fitting to determine which of the candidate functions are the best match for the signal of interest [20]. The candidate functions are searched for the candidate which will provide the maximum reduction in the remaining MSE. The common energy, Q, given by:

$$Q(m) = g_m^2 D(m,m) \tag{2.9}$$

of the candidate functions not already fit is compared [24]. The candidate function which has the most common energy with the remaining signal is then fit unless:

- the number of candidates grows beyond a predetermined maximum;
- the amount of energy being fit is less than what would be fit for white Gaussian noise (WGN); or,
- the total amount of energy fit is below a predetermined level.

When any of these conditions is met, no more candidate frequencies are fit. These parameters of FOS can be set to improve its performance for a given application.

Once all the candidate frequencies have been fit or any of the exit conditions have been met, the weights of the non-orthogonal set of functions are determined. The non-orthogonal weights, a_m, are determined using an intermediary term v_i, which is given by:

$$v_i = -\sum_{r=m}^{i-1} \alpha_{i,r} v_r \quad ; \quad m < i \leq M \tag{2.10}$$

where $v_m = 1$, and M is the number of terms. Using these results the non-orthogonal weights, a_m, are given by:

$$a_m = \sum_{i=m}^{M} g_i v_i \tag{2.11}$$

2.8 Fast Orthogonal Search in Spectral Domain

In US patent 6005664, Korenberg et al [1] stated in Claim 2 that FOS could be used to form a functional expansion in the spectral domain. The FOS algorithm remains the same as is described in Section 2.7.

2.9 Domain Transformation and Irregular Sampling

2.9.1 Domain Transformation

In order to transform a signal from the temporal domain to the spectral domain, the DCT can be used. The DCT is defined as [25]:

$$X[k] = \frac{w(k)}{\sqrt{N}} \sum_{n=0}^{N-1} x[n] cos(\frac{\pi(2n-1)(k-1)}{2N}) \tag{2.12}$$

where

$$w(k) = \begin{cases} 1 & ; \quad k = 0 \\ \sqrt{2} & ; \quad 1 \leq k \leq N-1 \end{cases} \tag{2.13}$$

where k is the sample number in the frequency domain, n is the sample number in the time domain, and N is the number of samples. Conversely, the inverse discrete cosine transform (IDCT), given by [25]:

$$x[n] = \sum_{n=0}^{N-1} \frac{w(k)}{\sqrt{N}} X[k] cos(\frac{\pi(2n-1)(k-1)}{2N}) \tag{2.14}$$

where

$$w(k) = \begin{cases} 1 & ; \quad k = 0 \\ \sqrt{2} & ; \quad 1 \leq k \leq N-1 \end{cases} \tag{2.15}$$

can be used to transform a signal from the spectral domain to the temporal domain.

The DCT and DFT are closely related. The DCT can alternatively be described as the DFT when performed on real data which have been made symmetric.

As can be seen in Eqs 2.12 and 2.14, each point in a given domain is generated from a weighted summation of all the points in the opposing domain [26]. In conjunction with irregular sampling, this property can be exploited to reduce the number of samples transmitted.

2.9.2 Irregular Sampling

Irregular sampling of the signal reduces the amount of data that are stored or transmitted. When the sampling is performed in the domain opposite the information domain, removing a particular sample will not result in completely removing a sample in the information domain. In Fig 2.4, a signal with three frequency components and additive white Gaussian noise (AWGN) is represented in both the time domain and the frequency domain.

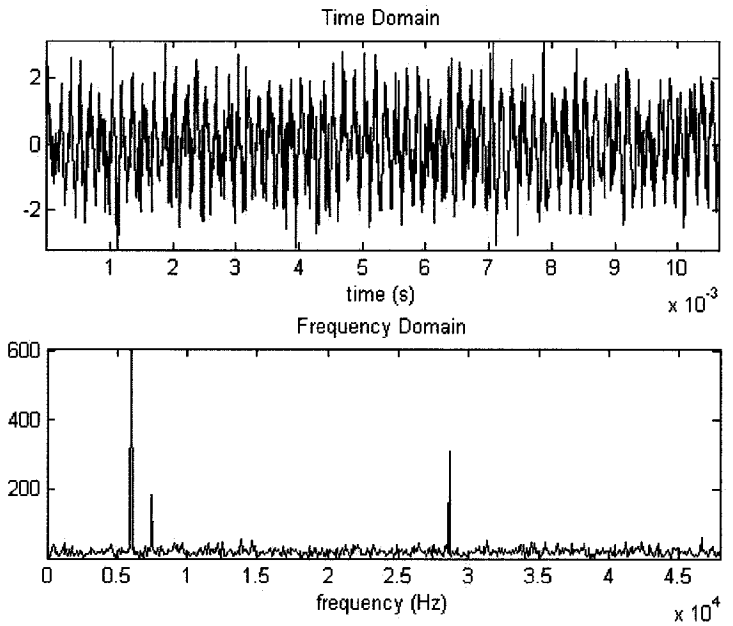

Figure 2.4: In this figure the upper image shows a time domain signal with three frequency components and AWGN. The lower image shows the frequency domain representation of this signal.

The same signal is shown in Fig 2.5 after being irregularly sampled. As the results of the irregular sampling, one quarter of the samples have been removed.

The frequency domain representation shows that no frequency component is lost by combining irregular sampling with domain transformation.

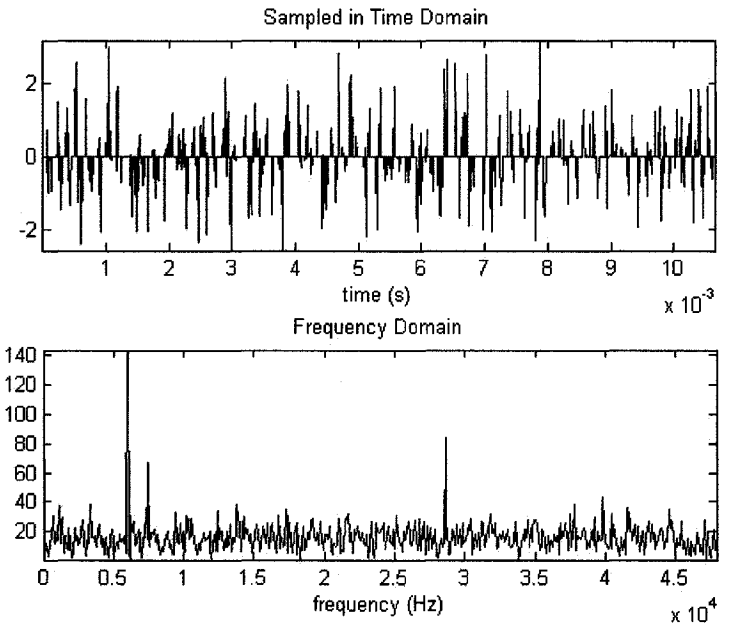

Figure 2.5: In this figure the same time domain signal is shown, however one quarter of the samples have been removed using an irregularly spaced sampling pattern. The lower figure shows the frequency domain representation. The magnitude of the component frequencies have been reduced but no single component has been lost.

The advantage of irregular sampling is that it will not cause a direct reduction in either frequency or time resolution as would regularly spaced sampling [22]. For instance removing every second sample would reduce the Nyquist frequency by $\frac{1}{2}$. However, by sampling irregularly, there will be portions of the sampled signal which still have adjacent samples remaining, thus the higher frequency components will be retained [22].

While the sampling is irregular and can be generated randomly, it is critical to know the sampling pattern when attempting reconstruction as this pattern will be used in the reconstruction of the signal. The transmitted signal will consist of the signal samples remaining after the irregular sampling has been applied. The sampling pattern can be included in the header of the transmitted file or can be known by both

the compression and decompression algorithms a priori. By selecting the sampling pattern a priori and not transmitting it, a measure of security is injected, as the signal can only be accurately reconstructed by receiver stations which have knowledge of the sampling pattern. This also increases the CR achieved.

2.10 Quantization Format

The data recorded by the hydrophones are recorded in signed little Endian format, requiring 16 bits per sample. After data manipulation is performed using the two algorithms, it is stored in IEEE double format, requiring 64 bits per data point. Quantization of the IEEE double format is required to maintain a good CR. There are two options for quantization to 16-bit numbers. The first is to quantize using a fixed point format, such as Q15, and the second is to use half-precision floating point notation, or IEEE Float16.

Q15 notation refers to a 16-bit signed proper fraction in which the first digit is a sign bit, and the remaining 15 bits denote the proper fraction [27]. Q15 notation has a range from -1 to $1-2^{-15}$ and a precision of 2^{-15} [28]. While, the range is generally sufficient to represent the data required, which tends to range between -1 and 1, it is not guaranteed to be sufficient.

IEEE Float16 notation ranges from -2^{14} to 2^{14}, and the precision varies across the range of the data [29]. The smallest magnitude which can be represented is -2^{-14} [29]. This number format is more suitable for data which tends to have a magnitude of less than 1. IEEE Float16 is shown in Fig 2.6.

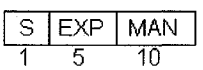

Figure 2.6: Bit allocation for IEEE 16 bit floating point notation. One bit is allocated for the sign, 5 for the exponent and 10 for the mantissa.

The IEEE Float16 format has one bit which represents the sign of the number, 5 bits for the exponent and 10 bits for the mantissa [29]. Numbers outside the range are denoted as $\pm Inf$ as appropriate. Rounding, when equidistant between two

numbers, is typically performed so that the even integer is chosen [29]. This is one of the rounding options available for this format, and it is the one which will be used throughout this thesis.

2.11 Detection Algorithm

The detection algorithm used to evaluate the results of the compression algorithms was developed at DRDC Atlantic [30]. It is a two-stage algorithm which is performed in the frequency domain. The signal is windowed with a Hann window of length $w_{fft} = 128$ using 50% overlapping windows as demonstrated in Fig 2.7.

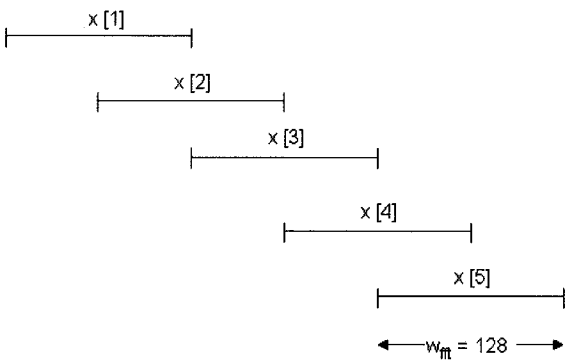

Figure 2.7: The signal is windowed, with a Hann window, using a 50% overlap and divided into segments of length 128 samples.

Each segment of the windowed signal is then transformed into the frequency domain through a 128 point FFT, resulting in a time-frequency representation of the signal, $X_n[f]$, where n is the segment of the signal which has been transformed. The power spectral density (PSD), $E_n[f]$, is then calculated for each segment using [26]:

$$E_n[f] = |X_n[f]|^2 \tag{2.16}$$

Once the signal has been transformed into time-varying segments, signal and noise levels are determined for each segment. The signal level, S_n, is found by summing w_s time windows starting from the time window $(w_s - 1)/2$, prior to window n and finishing at the time window $(w_s - 1)/2$ after window n. The summation is calculated

across the frequency bins known to contain the whale vocalization of interest. This range is shown between f_{low} and f_{high} in Fig 2.8.

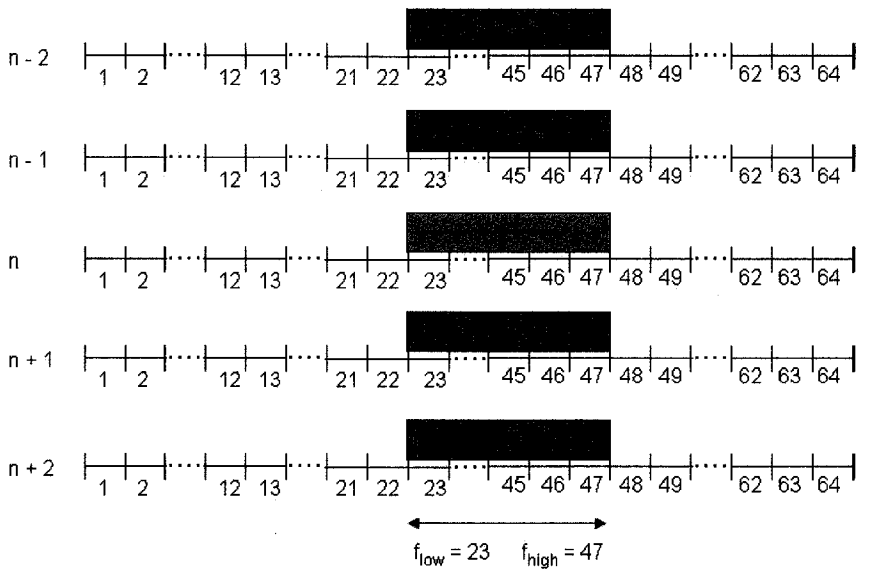

Figure 2.8: The signal level at time n is a summation of the energy across the frequency bins between f_{low} and f_{high} for each segment from $n-(w_s-1)/2$ to $n+(w_s-1)/2$. In this example it is the sum of FFT bins 23 to 47 for segment 1 (shown in green). The noise level at time n is summed across the same frequency bins from $n-(w_n-1)/2$ to $n+(w_n-1)/2$ excluding n. In this example, from $n-2$ to $n+2$ are shown (in grey).

The signal level, S_n, at segment n is calculated by [30]:

$$S_n = \frac{1}{w_s} \sum_{i=-(w_s-1)/2}^{(w_s-1)/2} \left(\sum_{f=f_{low}}^{f_{high}} E_{n+i}[f] \right) \tag{2.17}$$

where f_{low} is the lowest frequency bin in the frequency range which contains the whale vocalization and f_{high} is the highest frequency bin in the frequency range which contains the whale vocalization.

The noise level is calculated in a similar manner, however, the segment(s) used to calculate the signal level are excluded from the segment set which is summed for the noise level. For the noise level, w_n segments are summed, where w_n is odd, and $w_n > w_s$. The noise level, N_n, for segment n is given by [30]:

$$N_n = \frac{1}{w_n - w_s} \sum_{i=-(w_n-1)/2}^{(w_n-1)/2} \left(\sum_{f=f_{low}}^{f_{high}} E_{n+i}[f] - S_n \cdot w_s \right) \tag{2.18}$$

Once a signal and noise level have been determined for a given window, n, the likelihood of a signal in the band of interest, denoted $L1_n$ is determined using [30]:

$$L1_n = \frac{S_n}{N_n} > \tau_1 \tag{2.19}$$

where τ_1 is the threshold above which a detection is generated. Using Eqns 2.17, 2.18, and 2.19, a first stage detection is determined if the energy in a given window is above the average value of the energy in the windows around it. This indicates that there is a short-duration signal with high energy.

The second stage of the detection algorithm was designed to remove false alarms caused by transients such as those caused by the mechanics of the host AUV or sonobuoy. These transients consist of a pulse of energy across the entire frequency spectrum as seen in Fig 2.9.

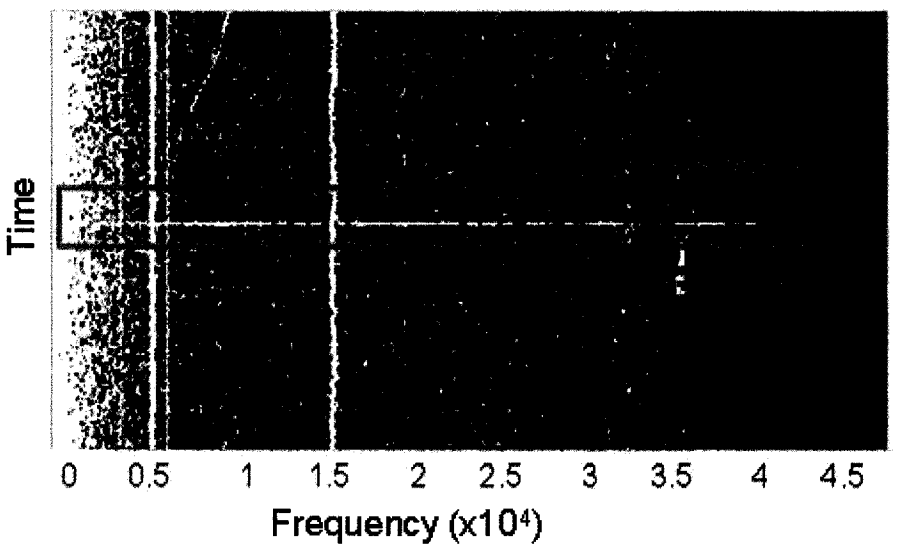

Figure 2.9: A transient pulse has energy starting as low as the DC component unlike the whale vocalizations of interest which tend to have energy components starting at 20 kHz.

Once the first stage detection has been completed, for each detection which was generated, the signal level is compared to the energy level in some of the surrounding bins, or the guard band level. As for the signal level, the guard band level is determined using w_s segments, however different FFT bins are summed. There may be a guard band on either side of the signal, or only on one side. For the case with one

guard band, GB_n, it is calculated using [30]:

$$GB_n = \frac{1}{w_s} \sum_{i=-(w_s-1)/2}^{(w_s-1)/2} \left(\sum_{f=fGB_{low}}^{fGB_{high}} E_{n+i}[f] \right) \tag{2.20}$$

where fGB_{low} and fGB_{high} are the upper and lower limits of the guard band FFT bins as shown in Fig 2.10. Since the vocalizations detected in this thesis tend to be broadband, only the case with one guard band is used. However, in the case where there are two guard bands, the summation is simply calculated over both regions of interest, so that all the bins which are a part of either guard band are summed over the range of segments.

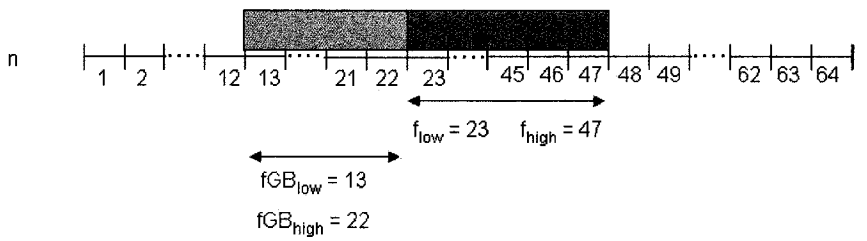

Figure 2.10: The guard band level at time n is a summation of the energy across the frequency bins between fGB_{low} and fGB_{high} for each segment from $n-(w_s-1)/2$ to $n+(w_s-1)/2$. In this example it is the sum of FFT bins 13 to 22 for segment n.

The guard band level is calculated for each segment n where a detection was generated in the first stage of detection.

The second stage then compares the energy in the frequency band of interest to the energy in the guard band. The likelihood of a signal which does not exceed the band of interest, denoted $L2_n$ is determined by [30]:

$$L2_n = \frac{S_n}{GB_n} > \tau_2 \tag{2.21}$$

where τ_2 is the threshold above which it is determined that a signal is present, and below which no signal is present.

Chapter 3

Data and Methods

3.1 Time Domain Algorithm

In the time domain compression algorithm, a time domain functional expansion of the input signal is created. The recorded data are segmented. FOS is then applied to each segment. The input candidates to FOS are complex signals which are models of the whale vocalizations, and are chosen based on their demonstrated ability to aptly represent underwater acoustic signals. These candidates are all normalized to unit energy.

The vocalization candidates are short-duration and are thus cross-correlated with the input segment to reduce the number of time delays of the candidate required as input to FOS.

The FOS operational parameters are set to reduce the model size, and increase the computation speed while ensuring enough information is retained to recreate the marine mammal vocalizations present. Once the significant model expansion terms have been selected, the feature number, along with their quantized weights are transmitted. The time delay is also transmitted when required. Reconstruction is performed at the receiver station using the feature, weight, and if necessary, the time delay. Detection results using the reconstructed signal are compared to the recorded signal.

3.1.1 Signal Segmentation

The selection of the segment length N is constrained by the two contradictory signal processing requirements, the frequency resolution and signal processing time. N is inversely proportional to frequency resolution, and is directly proportional to the increase in processing requirements for FOS. Thus, improvements in frequency resolution will be at the expense of increased signal processing resources. Furthermore,

the selection of N affects the compression ratio which can be achieved, as well as the ability to resolve separate events in time. An N of 1024 was selected for this thesis.

Frequency resolution for the FFT is given by $f_{res} = \frac{f_s}{N}$, where f_s is the sampling frequency. The sampling frequency for these platforms is $f_s = 96$ kHz. Using these parameters, with a $N = 1024$, $f_{res} = 93.75$. However, since FOS has been shown to achieve a frequency resolution of eight times FFT resolution or greater [24], [18], using $N = 1024$ permits a frequency resolution of at least $f_{res} = \frac{93.75}{8} = 11.7$ Hz. Based on the known characteristics of the whale vocalizations, this f_{res} should be sufficient.

The processing requirements for FOS are of order $C \times P \times N$ complexity where C is the number of candidates, P is the number of terms selected, and N is the number of samples in the input signal [31]. An N of 1024 should not increase the signal processing requirements beyond the on-board processing resources of the AUV platform, given that $C = 9872$, as described in Section 3.1.7, and P, is typically less than 20, as described in Section 3.1.6.

The compression ratio is affected by the number of candidates fit for each segment, but the number of segments is also a factor. For instance, for the time domain algorithm, transmitting 20 terms and their associated weights and delays, would not provide a great compression ratio were they replacing 64 samples, as there would be between 40 and 60 values transmitted. The compression ratio is therefore dependent on the length of the segments. Segments of length $N = 1024$ provide transmission savings for both algorithms.

The final factor which must be considered is the ability to resolve events in time. If the N is too large, then it may be difficult to separate individual vocalizations which have an inter-click interval of $0.2 - 0.4$ s [12]. The selection of $N = 1024$ allows complete separation of the short term vocalizations detected in this thesis.

3.1.2 Sinusoidal Candidate Functions

There are two sets of candidate functions which are used as inputs to FOS. The first set is composed of sinusoidal candidates which have previously been useful in recon-

structing a variety of time domain signals in noise [19], [21], [20]. These candidates are provided to the FOS algorithm in pairs, consisting of cosine and sine components so phase can be estimated. By modifying the weighting for each of the candidates in the pair, a different phase delay is achieved as seen in Eqns 3.1, 3.2, and 3.3. The sinusoidal candidates are fitted in pairs by FOS, as opposed to individual candidates which are each analyzed separately.

$$A \cdot \cos(\theta) + B \cdot \sin(\theta) = C \cdot \cos(\theta + \alpha) \tag{3.1}$$

where A and B are the weights of the candidate pair, α is the resultant phase delay, given by:

$$\alpha = -\arctan(\frac{B}{A}) \tag{3.2}$$

and C is the magnitude of the resulting function, given by:

$$C = \sqrt{A^2 + B^2} \tag{3.3}$$

Sinusoidal candidate pairs are generated at $\frac{1}{8}$ the FFT resolution, giving a spacing of 11.71875 up to the Nyquist frequency, $f_N = 48$ kHz. Sinusoidal candidates below 5 kHz have been eliminated due to large noise levels and the absence of signals of interest in this region. The resulting number of sinusoidal candidates used is 7338.

3.1.3 Vocalization Candidates

The vocalization candidates are created based on observed and reported parameters for frequency range, rate of change of frequency and signal length. Johnson et al [12] found that Blainville's Beaked whales use a frequency range of 25 to 80 kHz to create their most common two sounds, the buzz and the click. With $f_s = 96$ kHz, and $f_N = 48$ kHz, Theriault et al [32] determined the frequency range is from 20 kHz to greater than 40 kHz. In Fig 3.1 two sample vocalizations are shown in both the time and frequency domains. These vocalizations have not been filtered, so the noise components of the underwater acoustic signal remain.

For these simulations, a slightly lower range was used, 18 to 38 kHz was deemed an appropriate range based on the observed energy distribution of the PSD in the

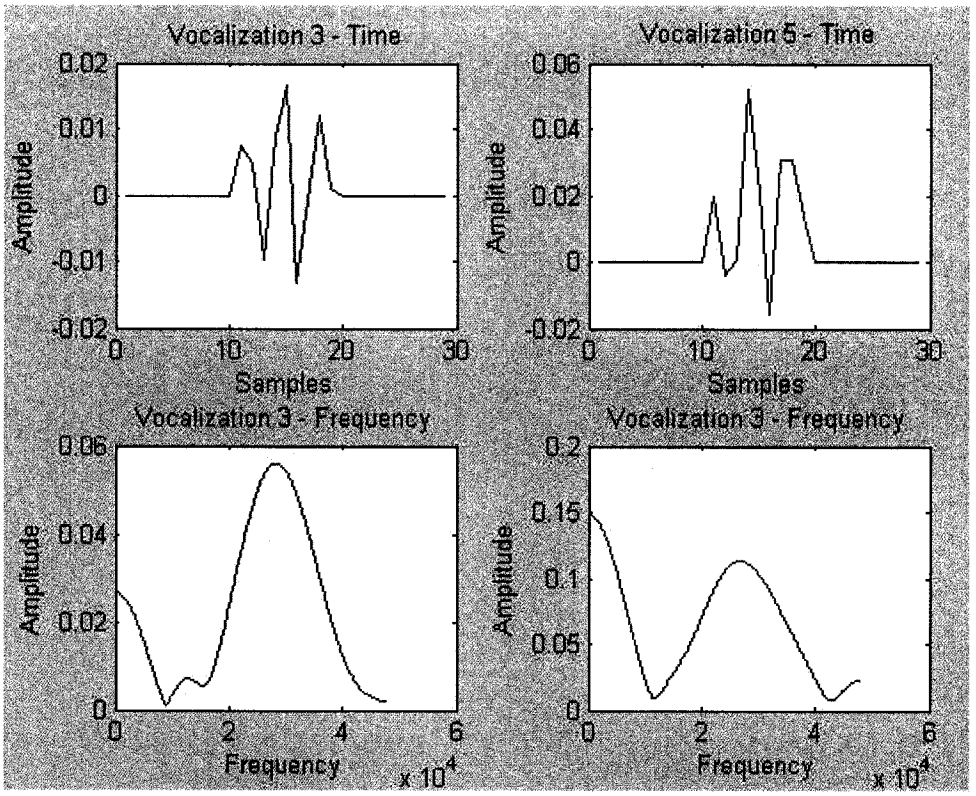

Figure 3.1: Two of the vocalizations from the test signal are shown in both the time and frequency domains. These vocalizations are extracted from the recorded acoustic signal, so there is low frequency noise which has not been removed. This is particularly evident in the frequency domain.

underwater acoustic data. The rate of change for frequency upswept vocalizations, is between 100 and 130 kHz/s [12] according to Au [33]. The length of the clicks and buzzes is 104 - 271 μs, which at a sampling rate of $f_s = 96$ kHz, gives 9 to 26 samples. Based on the observed data, this range was lowered, to give a range between 7 and 18 samples.

The vocalization candidates are separated into two types. The first type is a stable frequency candidate pair. These candidates are windowed using a Gaussian window, and are created using:

$$p_{2m}(n) = e^{-\frac{1}{2}(\frac{2\alpha\cdot(n-n_o)}{N})^2} \cdot \cos(\frac{2\pi f_m}{f_s}(n-n_o)) \tag{3.4}$$

$$p_{2m+1}(n) = e^{-\frac{1}{2}(\frac{2\alpha\cdot(n-n_o)}{N})^2} \cdot \sin(\frac{2\pi f_m}{f_s}(n-n_o)) \tag{3.5}$$

where n_o is $N/2$, f_m is the frequency of the candidate, N is 1024, α is a constant equal to 2.5, and m is the number of the candidate. For each frequency used, the clicks range between $n = 7$ and $n = 18$ samples in length. The frequency, f, ranges between 18 kHz and 38 kHz in increments of 2 kHz. The 132 stable frequency candidate pairs are modeled on the Blainville's Beaked whale's buzz. One of the stable frequency pairs is shown in Fig 3.2. Both time and frequency domain representations are provided. Comparing the stable candidate with the sample vocalizations shown in Fig 3.1, it can be seen that they are similar in the frequency domain.

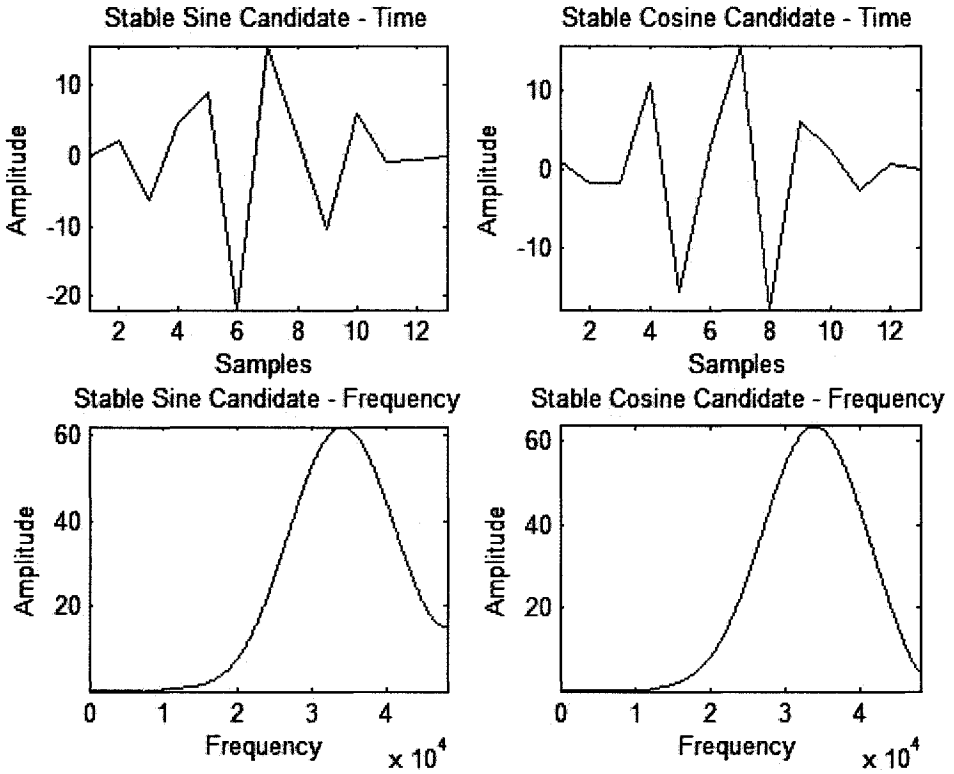

Figure 3.2: A stable candidate pair are shown in both the time domain, and in the frequency domain.

The second type of candidate pair are frequency upswept candidates, which use the same frequency range and sample length as the stable frequency candidate pairs. They also have a frequency modulation rate which varies between 100 and 130 kHz/ms

in increments of 2 kHz/ms. The frequency upswept candidate pairs are given by:

$$p_{2m}(n) = e^{-\frac{1}{2}(\frac{2\alpha\cdot(n-n_o)}{N})^2} \cdot \cos(\Theta((n-n_o))) \tag{3.6}$$

$$p_{2m+1}(n) = e^{-\frac{1}{2}(\frac{2\alpha\cdot(n-n_o)}{N})^2} \cdot \sin(\Theta((n-n_o))) \tag{3.7}$$

where n_o is $N/2$, m is the number of the candidate and $\Theta(n)$ is the phase at sample n, given by:

$$\Theta(n) = \begin{cases} 0 & ; \quad n = 0 \\ \Theta(n-1) + \frac{2\pi f_i(n)}{f_s} & ; \quad otherwise \end{cases} \tag{3.8}$$

where $f_i(n)$ is the instantaneous frequency at the sample n given by:

$$f_i(n) = f_o + v_m(n) \cdot n \tag{3.9}$$

where f_o is the starting frequency, and $v_m(n)$ is the frequency modulation rate.

For each different frequency modulation rate, and for each length n, the starting frequency is varied from 18 kHz in increments of the frequency modulation rate, until the starting frequency is high enough such that the end frequency is 38 kHz. The 1135 frequency upswept candidate pairs are modeled on Blainville's beaked whale clicks. One of the upswept frequency pairs is shown in Fig 3.3. Both time and frequency domain representations are provided. Comparing the frequency upswept candidates with the sample vocalizations shown in Fig 3.1, it can be seen that they are similar in the frequency domain, although this particular candidate pair has a slightly higher frequency range than the vocalization samples.

These parameters create a set of vocalization candidates that contains 1267 candidate pairs, or 2534 candidates.

3.1.4 Normalize Candidates

Each candidate is normalized to have unit energy. Candidate normalization is performed so that FOS is not biased when determining the energy Q as per Eqn 2.9, because bias may occur when a candidate has significantly more energy than the other candidates. Bias may be caused due to numerical precision issues between candidates with large energy and candidates with small energy. In this case, candidates with larger energy will be favoured.

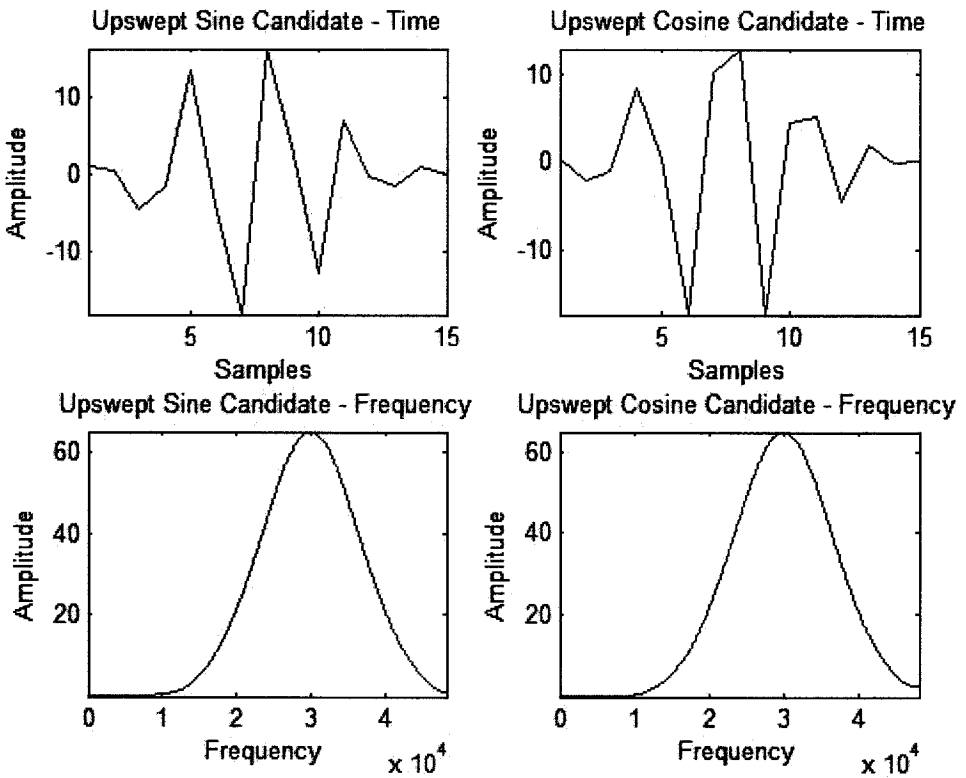

Figure 3.3: A frequency upswept candidate pair are shown in both the time domain, and in the frequency domain.

The first step to achieving normalization is to determine the mean of the candidates. The mean, $\overline{y}$, is determined using [34]:

$$\overline{y} = \frac{1}{N}\sum_{n=1}^{N} y(n) \tag{3.10}$$

where N is the number of samples in y. The candidates are then level-shifted to achieve zero-mean, y_{0m}, using [34]:

$$y_{0m}(n) = y(n) - \overline{y} \tag{3.11}$$

Once the candidates are level-shifted using Eqns 3.10 and 3.11 the result is divided by the standard deviation (STD) of $y(n)$ resulting in a normalized candidate, y_{norm}, using [34]:

$$y_{norm}(n) = \frac{y_{0m}(n)}{\sqrt{\frac{1}{N}\sum_{n=1}^{N}(y(n) - \overline{y})^2}} \tag{3.12}$$

For the vocalization candidates, normalization is done so that only actual vocalization values are changed, not the zero-padded values. This is accomplished by calculating the mean, $\overline{y}$, over the entire length of the click (after zero-padding), and applying Eqns 3.11, and 3.12 to the candidate prior to zero-padding. Thus for y_{0m} and y_{norm}, the values which are modified ranges from 1 to the length of the click, M rather than to N. The normalization must be performed in this manner so that the candidate, once zero-padded, has unit energy.

3.1.5 Cross-correlate Vocalization Candidates

When the signal is segmented, it is possible that a vocalization could start at any sample within a given segment as seen in Fig 3.4. The vocalization candidates range in length from 7 to 18 samples. They are zero-padded to create a candidate which is N, or 1024, samples in length. It is desirable that the vocalization candidate be aligned in time with the vocalization in the signal segment.

In order to achieve a good correlation between the vocalization candidates and a signal segment which contains a vocalization, a cross correlation between the two is done. For each candidate, the cross-correlation is performed for each integral sample time delay. The cross-correlation, $R_{yx}(m)$ is found by [26]:

$$R_{yx}(m) = \frac{1}{N} \sum_{n=0}^{N-m-1} y_{n+m} x_n^* \; ; \; m \geq 0 \tag{3.13}$$

where N is the length of the signal segment, m is the delay under consideration, y is the signal segment, and x is the vocalization candidate. Since the vocalization candidate is a real signal, it is equal to its complex conjugate, x^*. The maximum correlation for each candidate is then selected. It is sufficient to select the maximum correlation for each candidate, rather than to look for multiple maxima, because the segment length is smaller than the inter-click interval for the vocalizations of interest as discussed in Section 3.1.1.

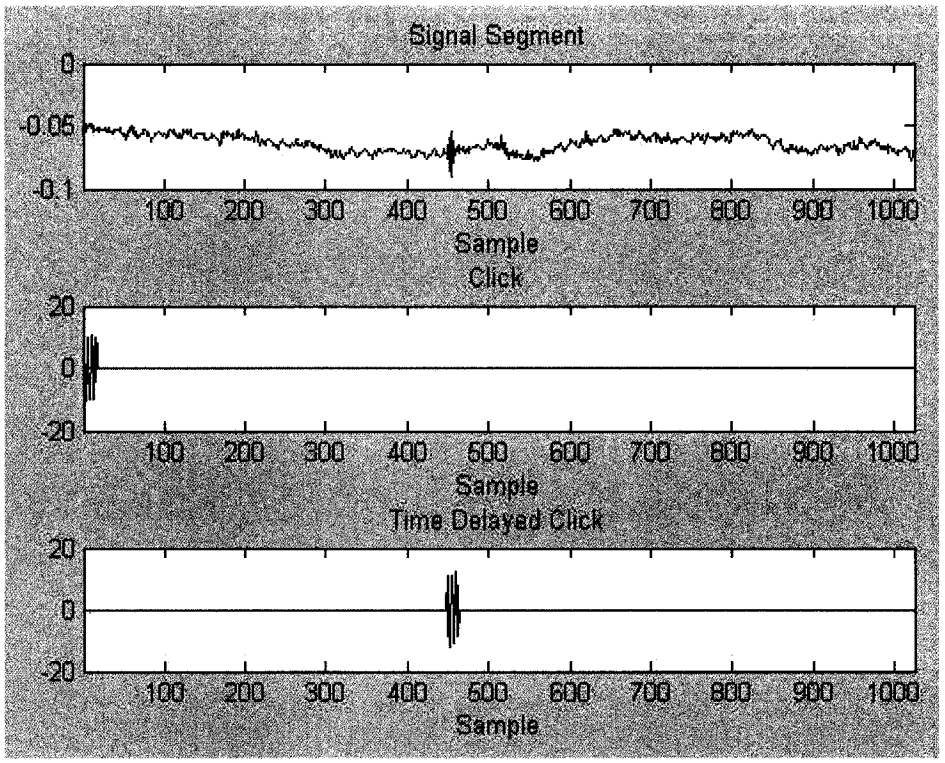

Figure 3.4: The first images shows a segment of the input signal with a vocalization approximately two fifths of the way into the signal. The second image shows a vocalization candidate which has been zero-padded. The third image shows the time-delayed vocalization candidate at the time-delay which will give the best cross-correlation.

3.1.6 Fast Orthogonal Search Operating Parameters

The FOS algorithm has a number of parameters which can be set to change the characteristics of the model expansion it determines for a given input signal. These include allowing candidates to be fit either individually or as part of a group, the inclusion of a direct current (DC) term as the first term fit, a $D(m,m)$ threshold, and a set of exit criteria. The exit criteria include a MSE threshold, a maximum number of terms to be fit, and a threshold for the correlation coefficient between the residual energy in the signal and the energy in the new term.

The FOS algorithm implemented for this thesis fits the candidates in groups. Some of the candidates are fitted by FOS in pairs, while other candidates are fitted individually. The number of terms in the group, n_{group} is used to ensure that each

candidate or pair of candidates is treated similarly. When checking each of the exit conditions, consideration is given to how many candidates are included in the group fit. Candidate groups are based on an index which is sent in to the FOS algorithm ordering the candidate terms as either individual candidates, or as members of a group.

Since neither the time domain nor the frequency domain algorithm includes a level-shifting step, a DC value is force-fit for this algorithm. This effectively adds a level-shift, while providing the amount by which each segment is shifted as a weighted output of the FOS algorithm.

A $D(m,m)$ threshold is set so that, when division by $D(m,m)$ is performed, numerical precision errors are not magnified by a very small value in the denominator [18]. This threshold is set as a safeguard against including candidates which are highly correlated to each other. It is set to 0.05.

The first exit condition for the FOS algorithm is an MSE threshold. The MSE threshold measures the percentage of the initial MSE, MSE_{init}, which is removed by fitting a candidate set. For this application, the percentage of the initial MSE of the segment fit by a given candidate set should not eliminate the candidate set from consideration, therefore the MSE threshold, τ_{MSE} is set to 0 as given in Eqn 3.14.

$$\frac{MSE_{rem}}{MSE_{init}} < \tau_{MSE} = 0 \tag{3.14}$$

where MSE_{rem} is the the MSE remaining after fitting the candidate set in question. This threshold condition may only be met due to numerical precision errors when dealing with very small MSE_{fit} and MSE_{rem} values, resulting in a MSE_{fit} greater than the remaining MSE.

The maximum number of candidates is another criteria by which the FOS algorithm can terminate the process of adding terms to the functional expansion. It was found, in testing experimental data, that FOS typically modeled fewer than 20 model terms per segment. Thus, the maximum number of terms to add was set to 20.

This value is set as low as possible without impacting the normal results. However, if there is a segment of a signal which is not well represented by the candidates

provided, this segment will not be able to fit terms until all of the terms have been fit, in an attempt to model the signal. This case may occur when there the data are corrupted, or when there are high energy disturbances. Attempting to model these cases would reduce the CR as well as increasing the processing time required to prepare the data for transmission without providing additional useful information about the presence of marine mammals in the vicinity of the recording platform.

The threshold based on the STD of WGN measures the correlation between the the candidate group under consideration and the residual in the signal. If it is less than the 95th percentile of the standard normal distribution, which for large N is 1.96, then the reduction in MSE is considered to be less than the amount of reduction in MSE that would be attained by adding a AWGN candidate to the model expansion. Using the normal distribution [35], the threshold is given by:

$$Q_{fit} < \tau_{WGN} = n_{group} \cdot \frac{1.96^2}{N} \cdot MSE_{rem} \tag{3.15}$$

where Q_{fit} is the energy fit by the candidate group, n_{group} is the number of candidates in the candidate group, and MSE_{rem} is the amount of MSE remaining between the input function, and the current functional expansion. If this threshold is not met, the candidate group is not fit and the FOS algorithm will exit as the functional expansion will be considered complete. If the candidate groups are chosen correctly, at this point all that is left of the signal will be noise. Since this thesis is not concerned with faithfully reconstructing the noise, the energy remaining can be ignored. This is the exit condition on which the FOS algorithm usually exits for the simulations performed.

3.1.7 Quantization

There are three different variables which must be transmitted to the receiver station in order to reconstruct the signal. These include: the candidate number for each of the candidates selected; the weight of each candidate selected; and the delay for the vocalization candidates.

The candidate number is an integer which can vary from 1 to the number of candidates. There are 7338 sinusoidal and 2534 vocalization candidates, creating a candidate set of 9872 candidates, which can be represented in binary using 14 bits.

The delay value only needs to be sent for the vocalization candidates, which are the candidates with candidate numbers above 7338. The delay values can vary between 0 and 1023, giving 1024 points, which can be represented in binary using 10 bits. The weight of the terms chosen is calculated in IEEE double precision, however, since this requires 64 bits, quantization was performed. The precision is reduced to IEEE Float16 precision, requiring 16 bits. The samples collected from the original signal use 16-bit signed integers.

To calculate the compression ratio, which was previously described as the original size of the data divided by the compressed size of the data, the following formula is used:

$$CR = \frac{N_{samples} \times 16\ bits}{N_{terms} \times 14\ bits + N_{delays} \times 10\ bits + N_{weights} \times 16\ bits} \tag{3.16}$$

where N_{terms} is the number of terms selected, N_{delays} is the number of delays transmitted, $N_{weights}$ is the number of weights transmitted and $N_{samples}$ is how many samples were collected in the original signal.

3.1.8 Reconstruction

Reconstruction is completed by multiplying the terms chosen, p_m, by the weight transmitted, a_m, as shown in Eqn 3.17. the reconstructed segment, y_{seg}, is given by:

$$y_{seg}(n) = \sum_{m=0}^{M-1} a_m p_m(n) \tag{3.17}$$

For the vocalization candidates, a delay is also transmitted, and is applied to the term chosen, p_m. Each segment is reconstructed separately and the reconstructed segments are then concatenated to form the received signal.

3.1.9 Detection

Whale vocalization detections are determined, by sending the received signal through the DRDC Atlantic detection algorithm [30]. The detection statistics for the reconstructed signal are compared to the detection statistics for the recorded signal.

3.2 Frequency Domain Algorithm

The frequency domain algorithm takes the recorded data and segments them. Each segment is then level-shifted. Once the segments are zero-mean, the DCT is performed separately on each one. The DCT coefficients are then quantized and irregularly sampled. Sampling is based on the probability of a given DCT candidate having a coefficient which is significant after quantization. For instance, DCT coefficients which are typically quantized to zero would be masked in the sampling pattern. The quantized, sampled data are then transmitted. Reconstruction using FOS may be performed at the receiver station with knowledge of the sampling pattern used. The candidates selected for FOS are chosen based on the characteristics of the underwater acoustic signal in the frequency domain. Finally, the detection statistics of the reconstructed signal are compared to the detection statistics of the recorded signal.

3.2.1 Signal Segmentation

Signal segmentation is completed in the same fashion, and for the same reasons as the signal segmentation in the time domain algorithm. A segment length of $N = 1024$ is used.

3.2.2 Domain Transformation

The signal is transformed into the frequency domain from the time domain using Eqn 2.12. Since the time domain signal is a real signal, the transformed signal, $X(f)$, will also be real. The result, is that both phase and magnitude information are contained within the real signal, and the reverse transformation can be completed without any additional information. The energy is distributed through the DCT

candidates in a predictable pattern. Fig 3.5 shows the distribution for a segment of underwater acoustic data which does not contain any marine mammal vocalizations. The coefficients of each DCT candidate vary between successive signal segments, however, the general distribution remains consistent. Each of the marine mammal vocalizations adds energy to a region of the DCT candidates. For instance, Blainville's Beaked whale buzzes add energy between the 400th and 800th DCT candidates for a 1024 point DCT. Knowledge of the general distribution pattern as well as which candidates are affected by the marine mammal vocalizations of interest is used to determine a sampling pattern.

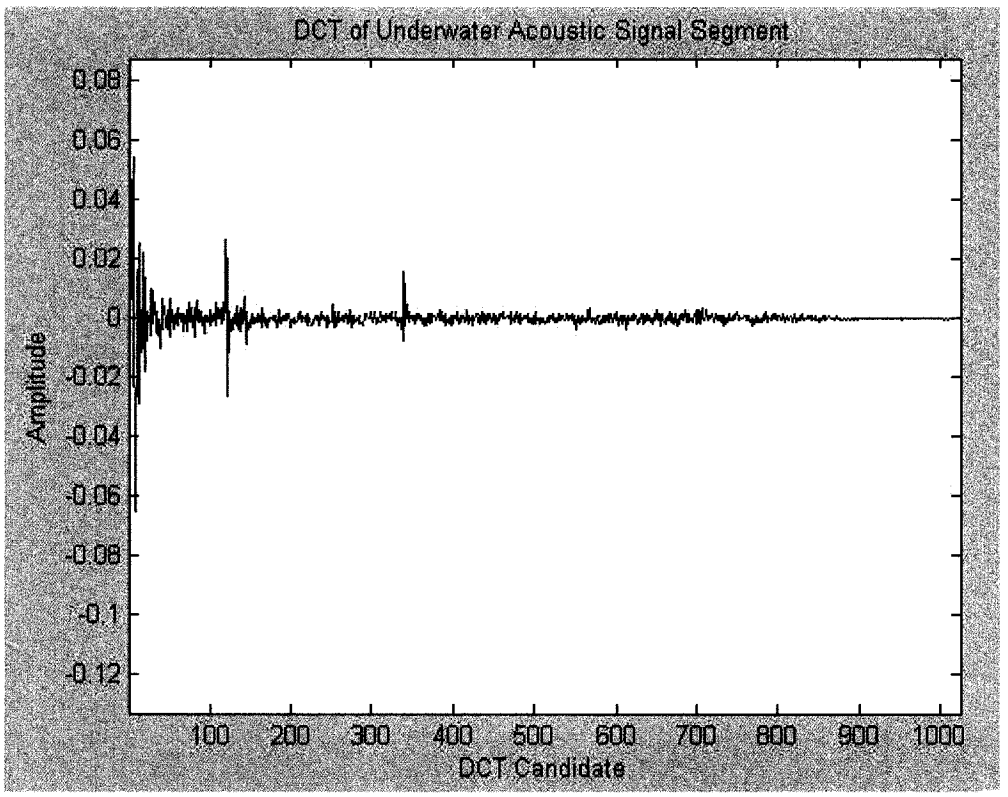

Figure 3.5: A typical distribution of the coefficients for a segment of underwater acoustic data. The first coefficients hold more information than the remainder of the coefficients, although there are regions where there is more information than others as seen by the larger amplitude in these regions.

3.2.3 Quantization

Only the DCT coefficients must be transmitted to the receiver station for signal reconstruction when using the frequency domain algorithm as the sampling pattern is known a priori. Since the level shift and DCT are computed using double precision numbers, and double precision is not required for reconstruction of the signal, the precision is reduced to IEEE Float16 precision. This requires 16 bits rather than 64 bits. Quantization of the DCT coefficients reduces the number of bits required to represent each DCT coefficient.

The compression ratio is calculated using:

$$CR = \frac{N_{samples} \times 16\ bits}{N_{DCTCoeffs} \times 16\ bits} \tag{3.18}$$

where $N_{samples}$ is the length of the original signal and $N_{DCTCoeffs}$ is the number of DCT coefficients transmitted.

3.2.4 Sampling Pattern

The sampling pattern determines which coefficients will be transmitted and which coefficients will be dropped. For irregular sampling, an appropriate pattern must be selected so that the marine mammal vocalizations are retained. The sampling is done in the frequency domain so that individual vocalizations, which can be as short as 7 samples long, are not missed by removing too many adjacent samples. Furthermore, there is a section of the frequency band which is known to contain solely noise. The DCT coefficients corresponding to this section will be entirely removed by the sampling pattern.

To determine whether or not to retain the remaining DCT coefficients, a training set of underwater acoustic data was used. The DCT was taken on each segment of this training set. DCT coefficients below a threshold were zeroed. A histogram of the number of times a given candidate was used over the course of the signal was computed. The coefficients with the smallest occurrence in the histogram were masked in the sampling pattern.

Segments which hold marine mammal vocalizations, and DCT coefficients which are used often in these segments, were chosen in the sampling pattern regardless of their rate of occurrence in the remainder of the training set.

The distribution of the sampling pattern is shown in Fig 3.6.

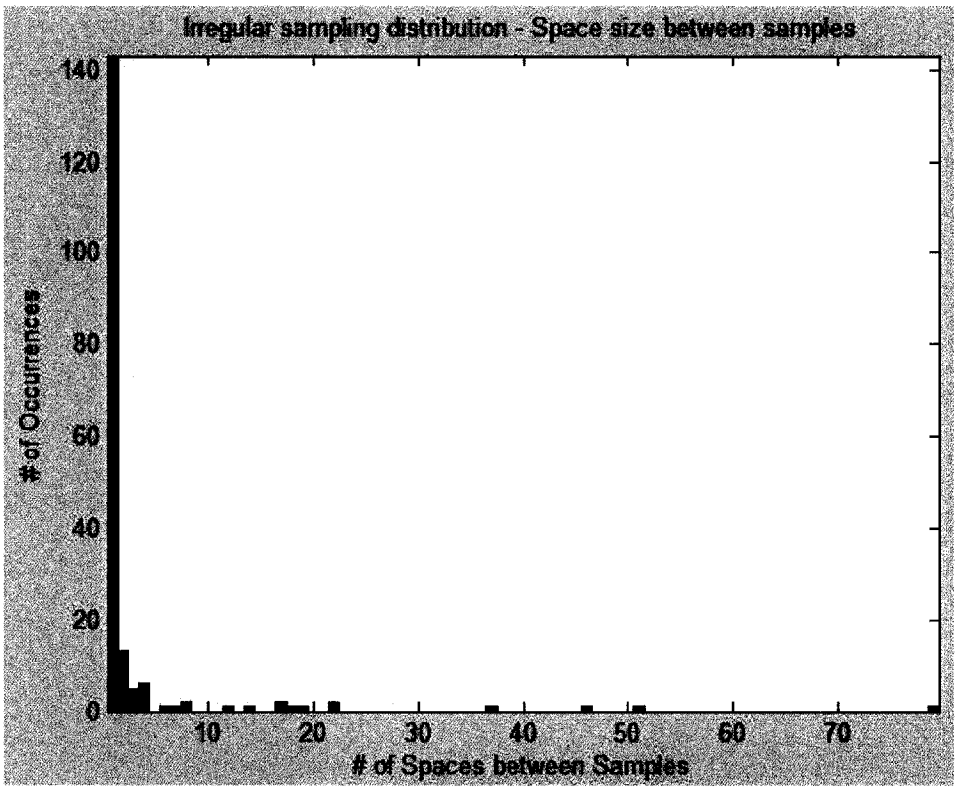

Figure 3.6: A histogram of the number of spaces between samples for the sampling pattern used by the frequency algorithm.

3.2.5 Irregularly Sample

The signal segments are each sampled using the sampling pattern. During reconstruction, the FOS candidates are sampled using the same sampling pattern which is known by both the AUV platform and the receiving station a priori.

3.2.6 Discrete Cosine Transform Candidate Functions

As in the time domain, there are two sets of candidates which are used as inputs to FOS. The first set is composed of DCT candidate functions. These candidates are sent into FOS individually. The DCT candidate functions are generated similarly to the sinusoidal candidates in the time domain. They are generated at $\frac{1}{8}$ the FFT resolution up to the Nyquist frequency, f_N, which is $\frac{1}{2}$ the sampling frequency, f_s or $f_N = 48$ kHz. Candidates below 5.4 kHz have been eliminated because there is a large amount of noise and no signals of interest in this frequency range.

Sinusoidal candidate functions are created in the time domain and the the DCT of these time domain functions is taken to create the frequency domain candidate functions. 3635 DCT candidates are created.

3.2.7 Vocalization Candidates

The vocalization candidates are based on the energy distribution seen in the test signal for the DCT of signal segments which contain marine mammal vocalizations. In Fig 3.7 the segments containing two of the vocalizations are shown as they look after the DCT is applied, upper images, and after the FFT is applied, lower images. The energy caused by noise and by the platform have not been removed.

The DCT candidates are sinusoidal pairs, which are windowed using a rectangular window. The vocalization candidates are given by:

$$p_m(n) = \Pi(n) \cdot \cos(\frac{2\pi f_c}{f_s} n) \tag{3.19}$$

$$p_{m+1}(n) = \Pi(n) \cdot \sin(\frac{2\pi f_c}{f_s} n) \tag{3.20}$$

where $\Pi(n)$ is given by:

$$\Pi(n) = \begin{cases} 1 & ; \quad 200 \leq n \leq 400 \\ 0 & ; \quad otherwise \end{cases} \tag{3.21}$$

and where $n = 400$, which is the length in terms of frequency bins, of the marine mammal vocalizations, and f_c is the frequency of the candidate pair. The frequency, f_c, ranges between 20 kHz and 40 kHz in increments of 2 kHz. The vocalization

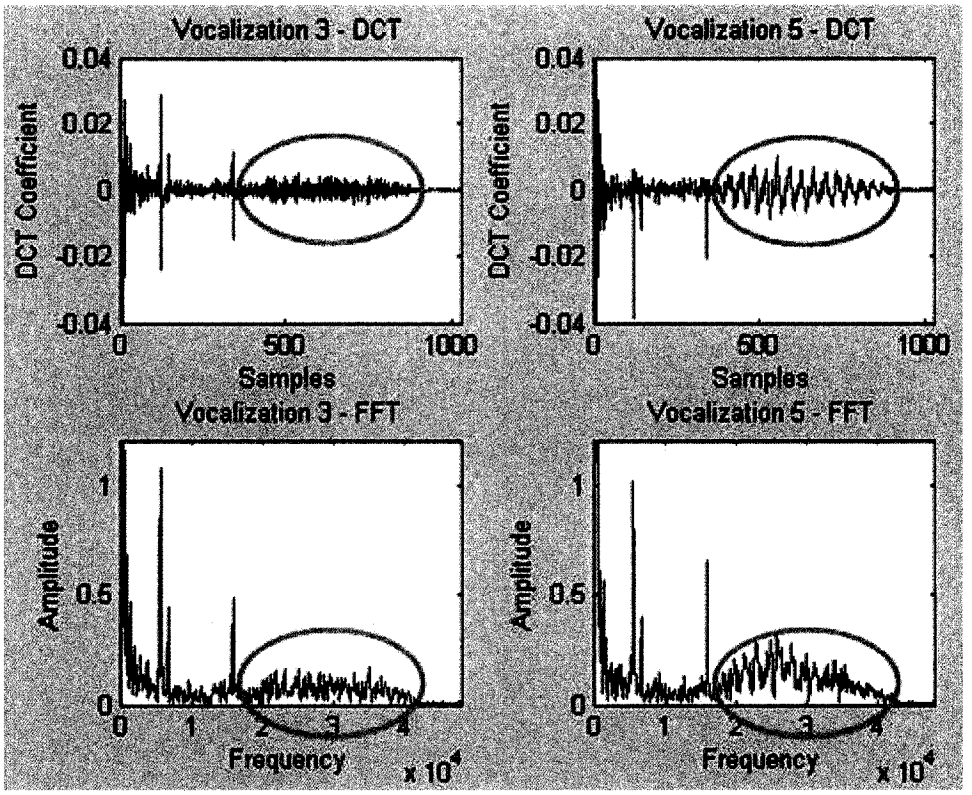

Figure 3.7: Two of the segments containing vocalizations from the test signal are shown in both the DCT and frequency domains. These vocalizations are extracted from the recorded acoustic signal, so there is low frequency noise which has not been removed. The energy of interest has been circled.

candidates are zero-padded to length N by adding 400 zeros prior to the candidate, and 224 zeros after the candidate. They are then normalized in the same manner as the time domain vocalization candidates. This ensures the energy of the vocalization candidates does not spread beyond the frequency range of the marine mammal vocalizations studied.

There are a total of 22 vocalization candidates used by the frequency domain algorithm.

3.2.8 Normalize Candidates

As in the time-domain algorithm, each candidate is normalized to have unit energy. The vocalization candidates are normalized using the same modifications as the time

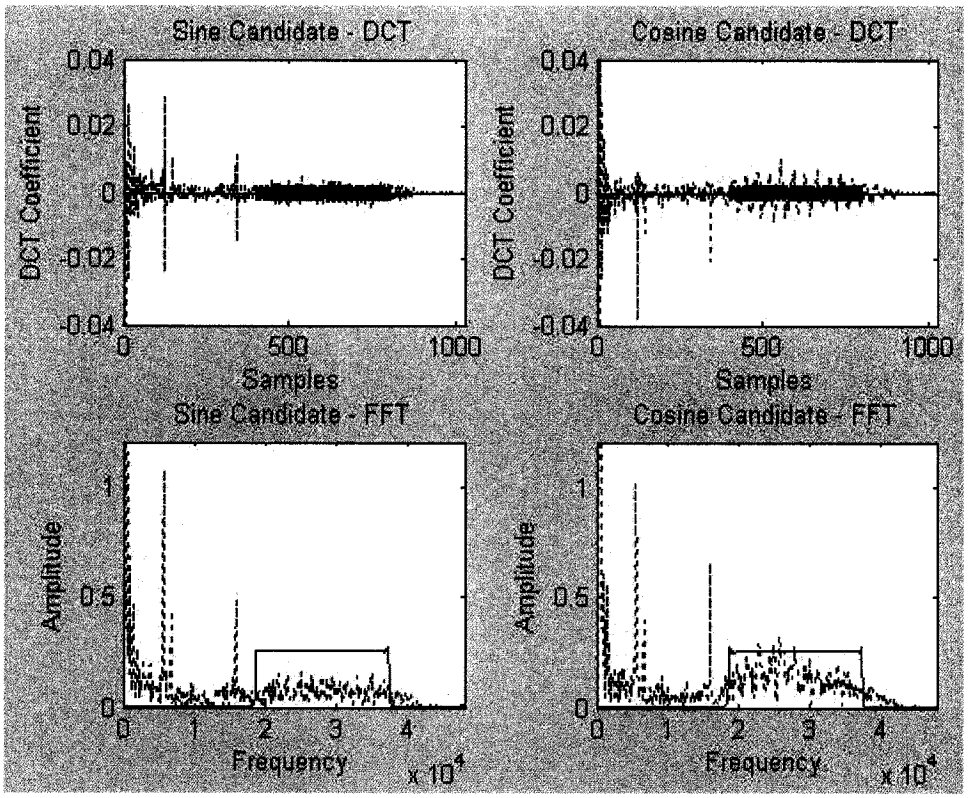

Figure 3.8: The vocalizations from Fig 3.7 are shown in red. They are overlaid by a DCT candidate pair which is shown in blue in both the DCT and FFT domains.

domain vocalization candidates. This eliminates the possibility of adding energy outside the frequency range of interest for the frequency band-limited candidates.

3.2.9 Reconstruction

The DCT term coefficients are received, and the DCT terms are known. The FOS candidates are subjected to the same irregular sampling that was applied to the DCT coefficients prior to transmission. The sampled coefficients, and the sampled candidates are input to FOS. Once the model terms are chosen, reconstruction is performed by applying the weights to the unsampled candidates using Eqn 3.17. The resulting signal is then subjected to the IDCT.

3.2.10 Fast Orthogonal Search Operating Parameters

The FOS algorithm used by the frequency domain compression algorithm is the same as the FOS algorithm used by the time domain compression algorithm. The $D(m,m)$ threshold, MSE threshold, and threshold for the correlation coefficient between the residual energy in the signal and the energy in the new term remain the same as in Section 3.1.6. The maximum number of terms to be fit is increased to 25. In the frequency domain algorithm, the number of candidates is the stopping criteria which is most often chosen.

3.2.11 Detection

The reconstructed signal is transformed into the time domain. This time domain signal is used in the DRDC Atlantic detection algorithm [30]. The detection statistics for the reconstructed signal are compared with the detection statistics for the original signal.

Chapter 4

Performance Metrics

The performance metrics used to evaluate the time domain and frequency domain compression algorithms are derived in this chapter. These tools assist in evaluating how the reconstructed signal compares with the original signal in terms of detection of the marine mammal vocalizations present. The performance metrics include:

- probability of detection;
- receiver operating characteristic curve; and
- significance testing.

The parameters used in the application of the detection algorithm are also defined in this chapter.

4.1 Probability of Detection

The probability of detection for each vocalization as the detection threshold changes will give an indication of an appropriate threshold, τ_1 to select for use in the second stage of the detection algorithm. This curve shows the thresholds at which the individual vocalizations cease to be detected. Differences between the detection capabilities of the signals compressed with the time domain and frequency domain algorithm and the original signal are also highlighted in a probability of detection curve. Since false alarms can be reduced by raising the detection threshold, it is desirable to maintain a high probability of detection as the threshold increases. Ideally the probability of detection would remain at 1 until the threshold has been raised high enough to eliminate false alarms.

4.2 Receiver Operating Characteristic Curve

The receiver operating characteristic (ROC) curve plots the probability of detection versus 1 - (the probability of false alarm) as the threshold of the detection algorithm is varied. The area under the ROC curve measures the probability that a correct identification will be achieved using the detection algorithm and associated detection parameters. By comparing the area under the ROC curve for the original and reconstructed signals, it is possible to evaluate how the compression algorithm affected the signal in terms of detection. The area under the ROC curve should be as close to 1 as possible for best results, with an area of 0.5 being a totally random predictor [36].

4.3 Significance Results

The significance results are measures of the quality of the results. There are a number of different significance results measuring everything from the standard deviation on the ROC curve to the accuracy. For the significance results, it is important that the results be independent. To ensure independence, no overlapping of the windowed segments is performed when using the DRDC Atlantic detection algorithm. Furthermore, a single value for the thresholds, τ_1 and τ_2 is used to calculate the significance results. These values have been provided by DRDC and are discussed in Section 4.4.

4.3.1 Standard Error

The standard error is a measure of STD. It can be used to calculate the error associated with the area under the ROC curve. The standard error, when applied to the area under the ROC curve will give the statistical significance of the results. Standard error, $SE(A)$, is calculated using [36]:

$$SE(A) = \sqrt{\frac{A(1-A) + (n_v - 1)(Q1 - A^2) + (n_n - 1)(Q2 - A^2)}{n_v n_n}} \tag{4.1}$$

where A is the area under the ROC curve, n_v is the number of vocalizations, n_n is the number of non-vocalizations, and $Q1$, and $Q2$ are given by [36]:

$$Q1 = \frac{A}{2 - A} \tag{4.2}$$

$$Q2 = \frac{2A^2}{1 + A} \tag{4.3}$$

This measure is used to calculate how many standard deviations from the area under the ROC curve would be required to get an area of 0.5. The more standard deviations required, the more likely the results are significant. The number of STDs, n_{STD} is calculated using:

$$n_{STD} = \frac{A - 0.5}{SE} \tag{4.4}$$

4.3.2 Curve Comparison

The ROC of the original signal can be compared to the ROCs of the reconstructed data from both the time domain and frequency domain algorithms. The standard error of the difference between the two areas, $SE(A1 - A2)$, is calculated using [37]:

$$SE(A1 - A2) = \sqrt{SE^2(A1) + SE^2(A2) - 2rSE(A1)SE(A2)} \tag{4.5}$$

where $SE(A1)$ is the standard error of the area under the first ROC curve, $SE(A2)$ is the standard error of the area under the second ROC curve and r is the correlation between the curves caused by the fact that they are derived from the same set of data. The term r is derived from a look-up table produced by Hanley and McNeil using A_{av} and r_{av} [37]. A_{av} is the average area under the two curves. It is given by:

$$A_{av} = \frac{A1 + A2}{2} \tag{4.6}$$

The term r_{av} is an average of two r terms. r_{av} is given by:

$$r_{av} = \frac{r_n + r_v}{2} \tag{4.7}$$

The first of these terms, r_n is a measure of the state without marine mammal vocalizations. It is given by:

$$r_n = \frac{(n_{BothTN} + n_{BothFP}) - (n_{Diff})}{n_n} \tag{4.8}$$

where n_{BothTN} is the number of segments where both the original and reconstructed signal have true negative detection states, n_{BothFP} is the number of segments where both the original and reconstructed signal have false positive detection states, and n_{Diff} is the number of segments where one has a false positive and the other has a true negative detection state. The second term, r_v, is a measure of the state with marine mammal vocalizations. It is given by:

$$r_v = \frac{(n_{BothTP} + n_{BothFN}) - (n_{Diff})}{n_v} \tag{4.9}$$

where n_{BothTP} is the number of segments where both the original and reconstructed signal have true positive detection states, n_{BothFN} is the number of segments where both the original and reconstructed signal have false negative detection states, and n_{Diff} is the number of segments where one has a false negative and the other has a true positive detection state.

Knowing the standard error of the difference between the two curves, z, measures how many STDs there are between the two curves. z is calculated using [37]:

$$z = \frac{A1 - A2}{SE(A1 - A2)} \tag{4.10}$$

4.3.3 Prediction Results

There are four states which the results of a prediction may fall into. They are:

- false positive (FP): the number of segments incorrectly identified as containing marine mammal vocalizations;
- true positive (TP): the number of segments containing marine mammal vocalizations which are detected;
- true negative (TN): the number of segments correctly identified as not containing marine mammal vocalizations; and
- false negative (FN): the number of segments containing marine mammal vocalizations which are not detected.

These four states can be used to determine the accuracy, positive predictive value (PPV), negative predictive value (NPV), sensitivity, specificity and phi coefficient of association. In graphical format they they are given by Fig 4.1

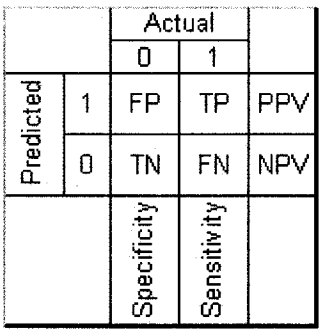

Figure 4.1: The states are given as they pertain to the actual an predicted value. The directionality of PPV, NPV, sensitivity and specificity are also given.

4.3.4 Accuracy

Accuracy is a measure of the percentage of the data which are correctly assigned, as in the number of correct detections or correct non-detections out of the total number of sets. It is given by [38]:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{4.11}$$

When the numbers of TN and TP are very different, the accuracy can be skewed. Since there are 11379 segments without vocalizations and only 21 segments with vocalizations in the test signal, the ability to determine the TN values will have a greater impact on the accuracy than the ability to determine the TP values. In order to provide a measure of accuracy which gives equal weighting to segments segments containing vocalizations and those without vocalizations, a modified accuracy measure is introduced. The modified accuracy, $accuracy_{mod}$, is given by:

$$Accuracy_{mod} = \frac{TP}{2(TP + FN)} + \frac{TN}{2(TN + FP)} \tag{4.12}$$

4.3.5 Positive Predictive Value

The PPV is the measure of the probability that there is a marine mammal vocalization given that the result of the test is positive. It is given by [38]:

$$PPV = \frac{TP}{TP + FP} \tag{4.13}$$

4.3.6 Negative Predictive Value

The NPV is the measure of the probability that there is no vocalization given that the results of the test is negative. It is given by [38]:

$$NPV = \frac{TN}{TN + FN} \tag{4.14}$$

4.3.7 Sensitivity

The sensitivity is the measure of the ability of the detection algorithm to correctly detect occurrences of marine mammal vocalizations. It is given by:

$$Sensitivity = \frac{TP}{TP + FN} \tag{4.15}$$

4.3.8 Specificity

The specificity is the measure of the ability of the detection algorithm to correctly detect non-occurrences of marine mammal vocalizations. It is given by:

$$Specificity = \frac{TN}{TN + FP} \tag{4.16}$$

4.3.9 Phi Coefficient of Association

The Phi test measures how much association there is between two binary variables, in this case, it is the association between a vocalization segment and a segment without a vocalization. It is given by:

$$\phi = \frac{TP \times TN - FP \times FN}{\sqrt{(FP + TP) \times (TN + FN) \times (FP + TN) \times (TP + FN)}} \tag{4.17}$$

This value may vary from -1 to 1 where strong values of Φ are higher than 0.61, between 0.31 and 0.6 is moderate, and anything lower than 0.3 is weak [38].

4.4 Detection Algorithm Parameters

There are a number of parameters which must be set in the DRDC Atlantic detection algorithm. These parameters are set as advised by the scientists at DRDC Atlantic. The windowing length, w_{fft} is set to 128. The length of the signal time window, w_s, is set to 1. The length of the the noise time window, w_n, is set to 313. The frequency bandwidth over which the signal and noise levels are summed ranges from $f_{low} = 17$ kHz to $f_{high} = 35$ kHz. For the guard band, the frequency range is $fGB_{low} = 10$ kHz to $fGB_{high} = 17$ kHz. Finally, the first detection threshold, τ_1 is set to 4, and the second detection threshold, τ_2, is set to 3.

Chapter 5

Results

Trial data containing Blainville's beaked whale vocalizations were provided by DRDC Atlantic. The signal of interest is a 15 second acoustic recording of Blainville's Beaked whales collected in the Alboran Sea in June of 2008, off the coast of Spain as seen in Fig 5.1. The candidates are based primarily on this data set.

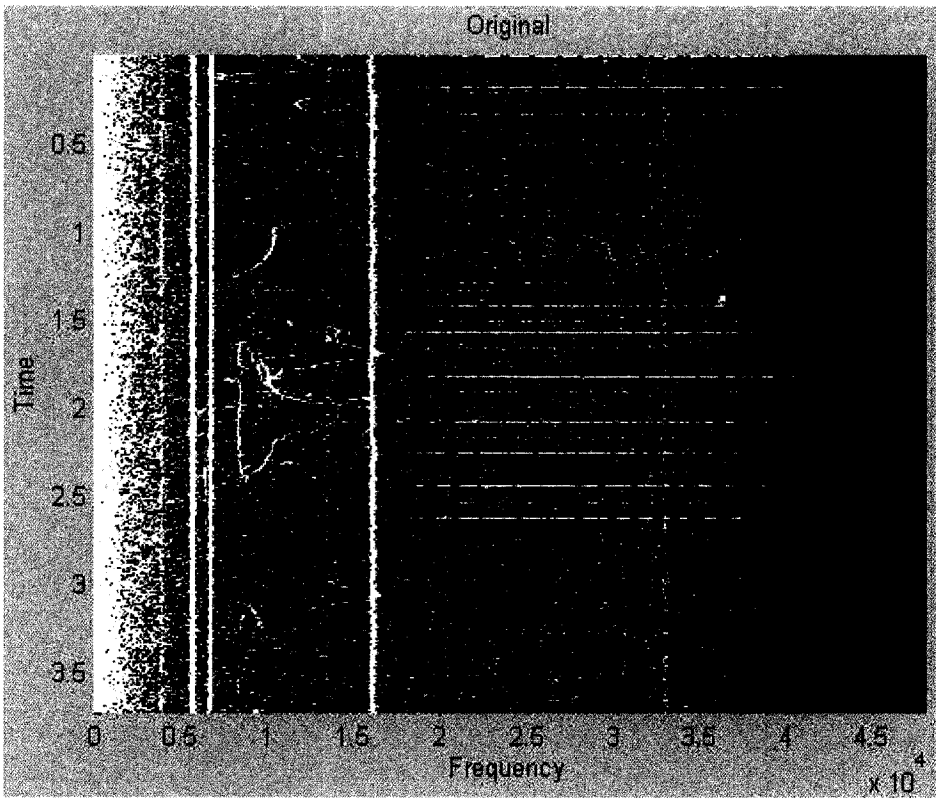

Figure 5.1: A set of buzzes are shown as horizontal lines of increased intensity across the frequency range 20 to 40 kHz. A whale song is visible in the 5 to 15 kHz range, shown as a series of high intensity regions varying in both time and frequency between the 2nd and 3rd vertical lines. The vertical lines are caused by the recording platform.

There are 20 Blainville's Beaked whale vocalizations present in this sample, spread across 21 signal segments. These were determined visually in both the time and

frequency domains. Vocalizations were separated from platform transients, which cause an increase in energy across the entire frequency range. An example of a buzz can be seen in Fig 5.2. The buzz looks like a short duration burst of energy in the time domain. In the frequency domain, the buzz is shown as a wide band of energy.

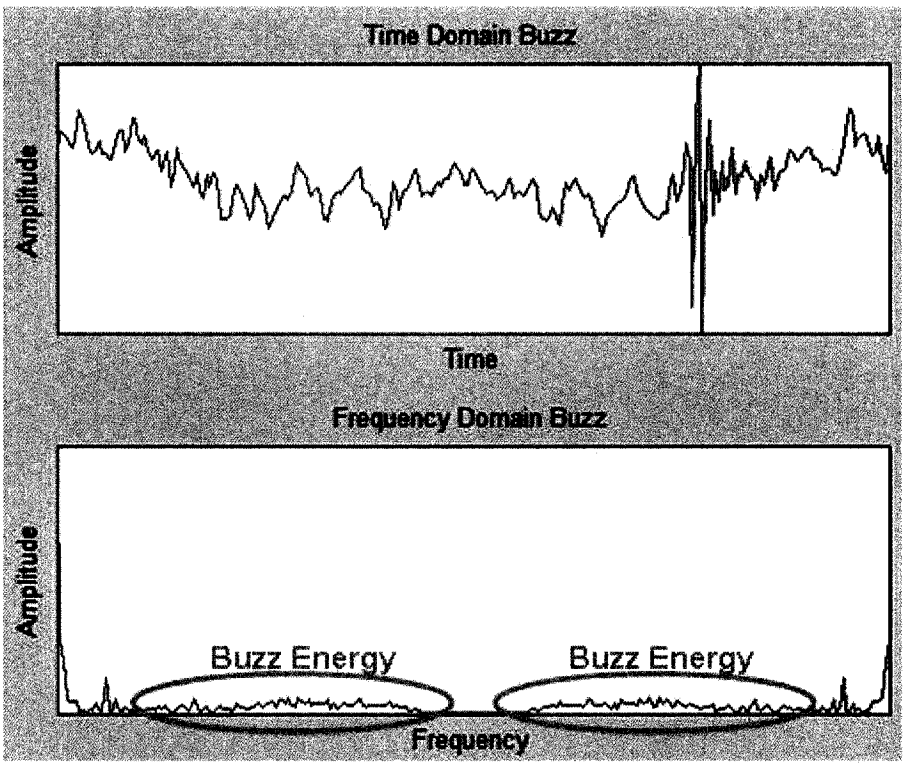

Figure 5.2: A Blainville's Beaked whale buzz is shown in both the time and frequency domains. The buzz has a larger amplitude than the surrounding noise in the time domain view, and causes an increase in energy in across the frequency range 20-40 kHz in the frequency domain view.

Platform transient signals may resemble long burst of buzzes, or they may resemble noise with higher than average energy. An example of a platform transient can be seen in Fig 5.3. In the frequency domain there is an increase in energy across the entire spectrum, not just the frequency range of $20 - 40$ kHz. This property is exploited by the detection algorithm in its second stage to reduce the number of false alarms detected.

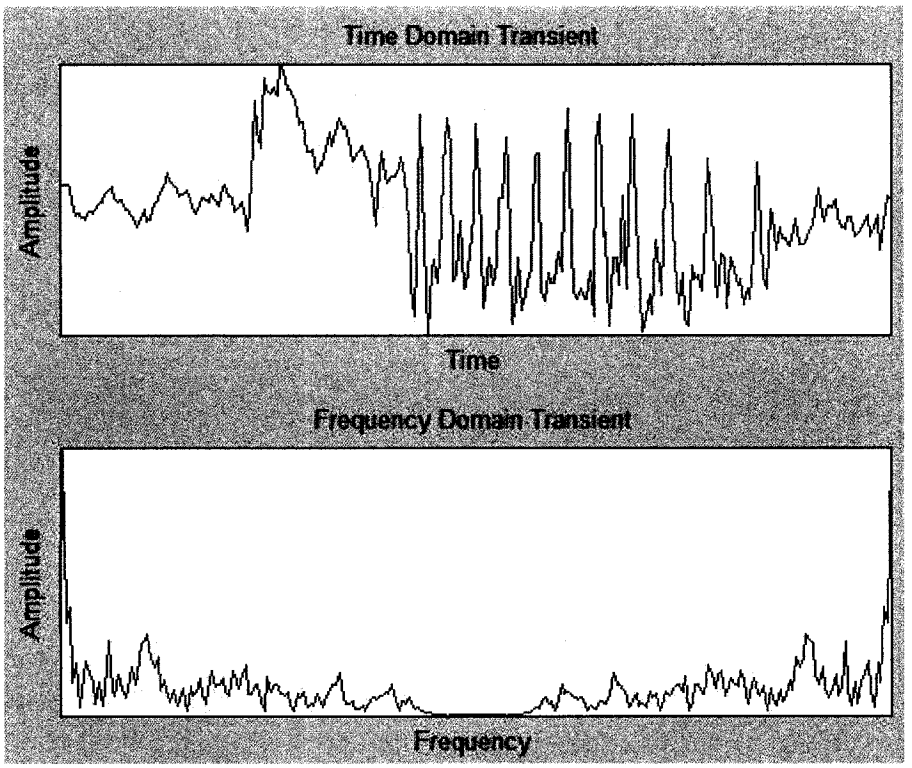

Figure 5.3: A transient signal is shown in both the time and frequency domains. The transient signal also has a larger amplitude than the surrounding noise in the time domain view, although it lasts longer. In the frequency domain view, there is an increase in energy across the entire spectrum.

5.1 Time Domain and Frequency Domain Results

As can be seen in Fig 5.4 the time domain algorithm resulted in a signal which had a visible reduction in noise in the LOFARgram. The Blainville's Beaked whale vocalizations remain visible, as do the transient signals generated by the platform. The whale song is partially visible, and the low-frequency noise has been almost completely eliminated.

As shown in Fig 5.5 the frequency domain algorithm resulted in a visible reduction in noise on the LOFARgram. The Blainville's Beaked whale vocalizations remain visible, as does the energy bursts caused by the platform, although these are no longer easily distinguishable from vocalizations. The whale song is completely visible, and the low-frequency noise has been significantly reduced.

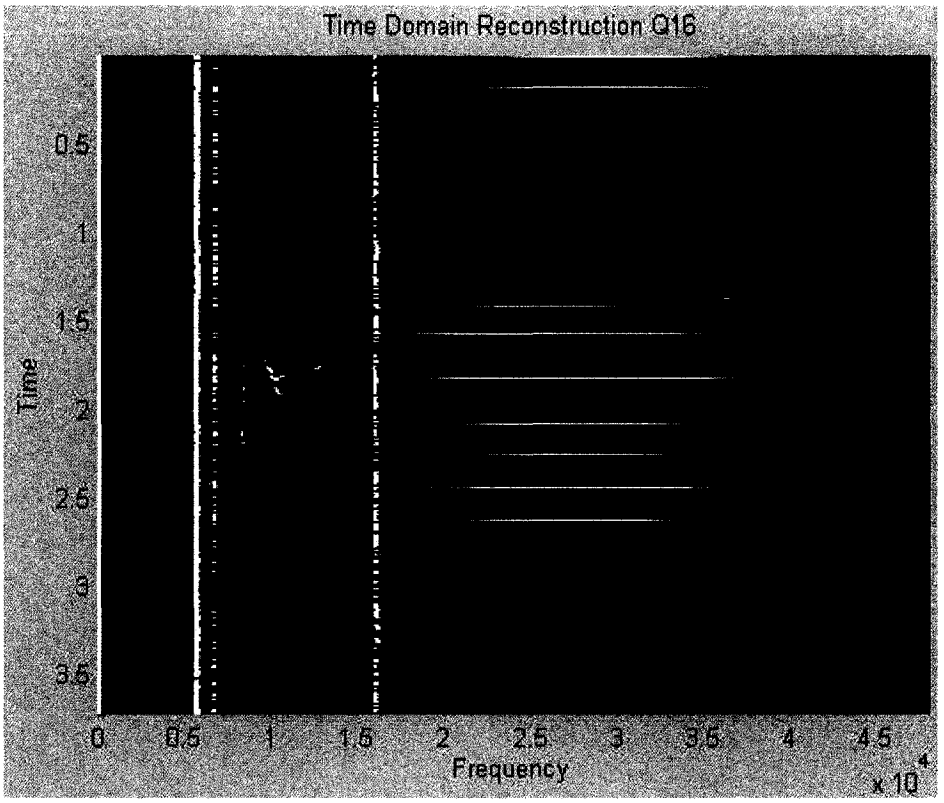

Figure 5.4: The set of buzzes can still be seen, shown as horizontal lines of increased intensity across the frequency range 20 to 40 kHz. Also, a whale song is partially visible in the 5 to 15 kHz range. The energy from the recording platform remains visible, while the remaining energy has been eliminated.

5.2 Analysis

Both algorithms provide a significant gain in terms of speed of transmission due to the CRs achieved. However, the frequency domain compressed signal has poorer detection than the original and time domain compressed signals. The number of false alarms is not reduced enough to provide adequate PPV. Due to the relatively low number of segments which contain vocalizations compared to the number of segments which do not contain vocalizations, any increase in the percentage of false alarms for the same percentage of detections causes a significant decrease in the PPV as well as the Phi coefficient of association.

A summary of the results for the signal of interest, using the detection algorithm

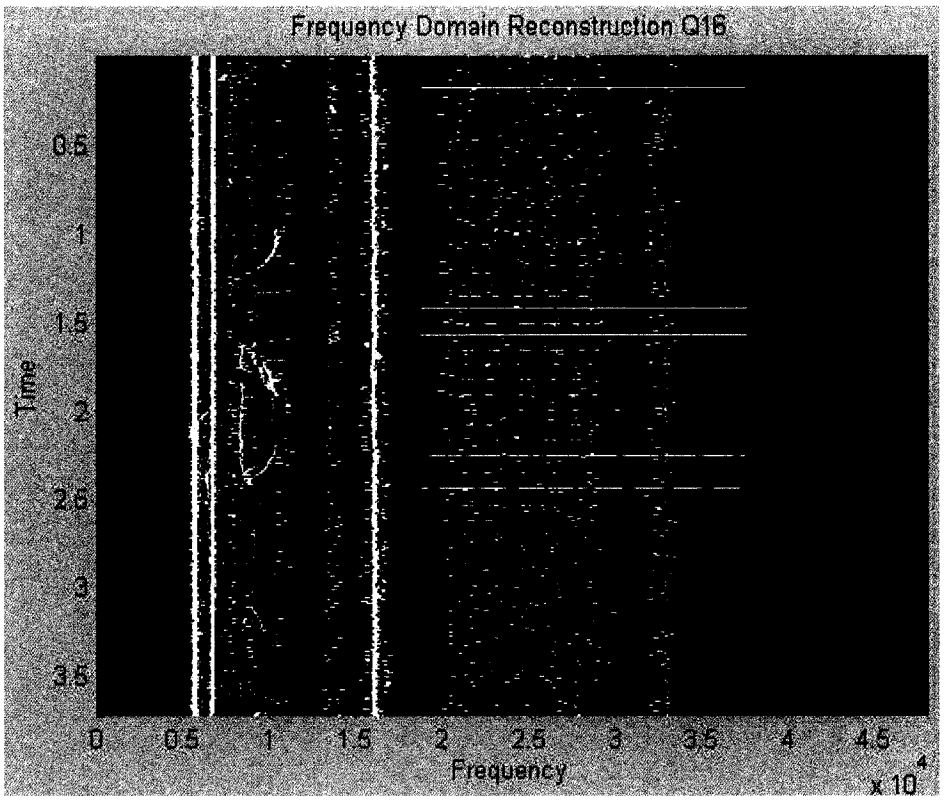

Figure 5.5: The set of buzzes can still be seen, shown as horizontal lines of increased intensity across the frequency range 20 to 40 kHz. Also, a whale song is completely visible in the 5 to 15 kHz range. The energy from the recording platform remains partially visible, while the remaining energy has been reduced at the sampled frequencies.

parameters provided by DRDC is provided in Table 5.1.

5.3 Compression Ratio

There is no compression applied to the original signal. The CR, from Eqn 3.16, of the signal after compression by the time domain algorithm is 93.0597. This was accomplished using IEEE Float16 notation to transmit the FOS coefficients, and integer representation of the delay values and candidate term numbers. This CR means that a 5 minute data capture, which would previously have taken 100 minutes to transmit, can now be transmitted in 64.5 seconds. The CR of the signal after compression by the frequency domain algorithm is 5.5652. This was accomplished using IEEE Float16 notation to transmit the DCT coefficients remaining after irregular sampling. This

Signal		Original	Time D	Freq D
CR		1	93.06	5.5652
# Segments w Voc		21	21	21
ROC	Stage 1	0.8012	0.8125	0.8951
ROC	Stage 2	0.8331	0.8092	0.7555
SE		0.0549	0.0574	0.0616
STD from ROC=0.5		6.0674	5.3868	4.1477
SE between Curves		N/A	0.0196	0.0355
STD between Curves		N/A	1.2939	2.1875
TP		12	12	11
FN		9	9	10
TN		11376	11379	11315
FP		3	0	64
Accuracy		0.9989	0.9992	0.9935
Mod Accuracy		0.7856	0.7857	0.7591
PPV		0.8	1	0.1467
NPV		0.9992	0.9992	0.9991
Sensitivity		0.5714	0.5714	0.5238
Specificity		0.9997	1	0.9944
Phi		0.68	0.76	0.27

Table 5.1: A comparison of the evaluation criteria results for the original signal, and the signals reconstructed after the application of the time and frequency domain algorithms respectively using the detection algorithm parameters provided by DRDC Atlantic.

CR indicates that a 5 minute data capture, which would previously have taken 100 minutes to transmit, can now be transmitted in 17 minutes and 58 seconds. Both algorithms provide significant significant savings in terms of transmission time over the original signal. The time domain algorithm also provides significant transmission time savings over the frequency domain algorithm.

5.4 Probability of Detection

Fig 5.6 shows the probability of detection of the 21 vocalizations as the detection threshold, τ_1, is increased. The threshold at which each vocalization ceases to be detected causes a decrease in the probability of detection at the given detection threshold. The probability of detection for the original signal and for the signal compressed using the time domain algorithm are shown in Fig 5.6. The two curves

cross at $\tau_1 = 1$ and again at $\tau_1 = 7.8$. Prior to $\tau_1 = 1$, the time domain compressed signal has a better probability of detection than the original signal. Between the two thresholds, the original signal has a better probability of detection, and after the second threshold, $\tau_1 = 7.8$, the time domain compressed signal once again has a better probability of detection. At the threshold at which the probability of detection of the original signal is reduced to 0, the probability of detection of the time domain compressed signal remains at 0.33. The threshold chosen will depend on the relative importance of accurate detections with respect to accurate classification of segments without vocalizations. The threshold used by DRDC Atlantic is 4. At this threshold, the original signal has a higher probability of detection which is 0.6667, than the time domain compressed signal which is 0.619. However for the time compressed signal, the values of TP, TN, PPV, NPV, and phi (the coefficient of association or Matthews' correlation coefficient) still equaled or exceeded those for the original signal, as shown in Table 5.1.

The probability of detection for the original signal and for the signal compressed using the frequency domain algorithm are shown in Fig 5.7. The two curves cross at $\tau_1 = 0.5$ and again at $\tau_1 = 1.9$. Prior to $\tau_1 = 0.5$, the original signal has a better probability of detection than the frequency domain compressed signal. Between the two thresholds, the frequency domain compressed signal has a better probability of detection than the original signal, and after the second threshold, $\tau_1 = 1.9$, the original signal once again has a better probability of detection. The probability of detection for the frequency domain compressed signal is reduced to 0 at $\tau_1 = 19.6$. At this threshold, the probability of detection for the original signal is 0.33, and the probability of detection of the time domain compressed signal is 0.52. The threshold used by DRDC Atlantic is 4. At this threshold, the original signal has a higher probability of detection which is 0.6667 than the frequency domain compressed signal which is 0.5238.

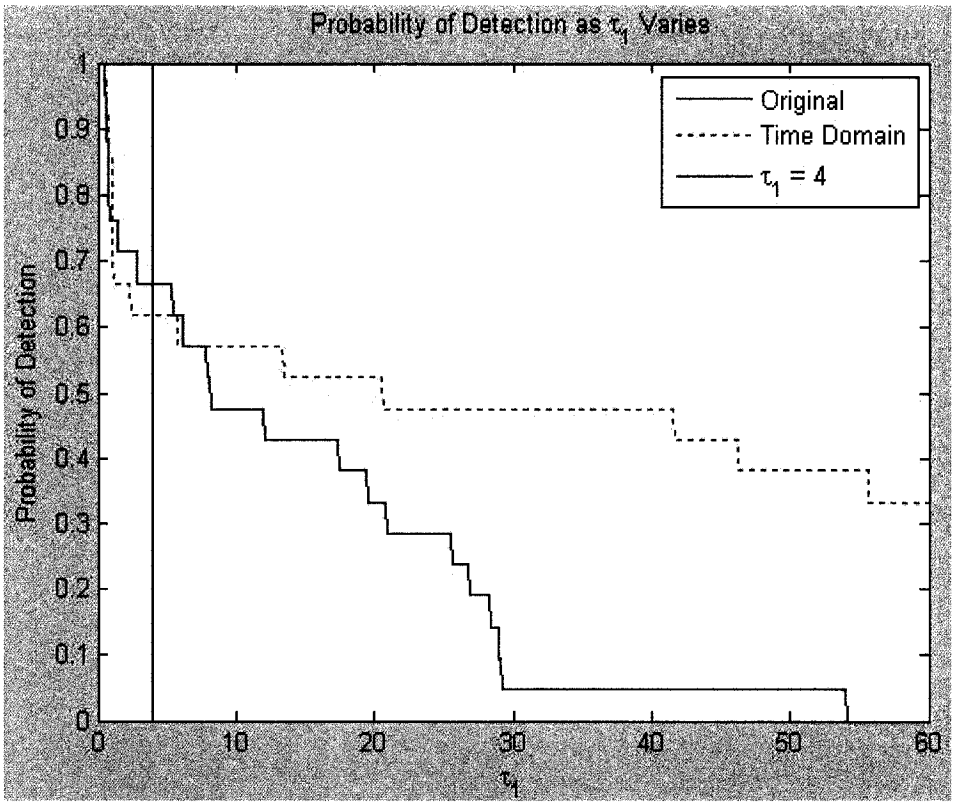

Figure 5.6: A comparison of the probability of detection for the vocalizations in the original signal and the time domain compressed signal. At the DRDC supplied threshold $\tau_1 = 4$, the original signal has a higher probability of detection than the time domain compressed signal.

5.5 Receiver Operating Characteristic Curve

The area under the ROC curve for the original signal after the first stage of detection is 0.8012. From the ROC curve produced by the first stage of the detection algorithm, a threshold, τ_1, of 4 was selected as the basis for the second stage of the detection algorithm. This value was provided as a value typically used by DRDC. After the second stage of detection the area under the ROC curve improves to 0.8331 since the false alarm rate for a given detection level is reduced for the second stage as compared to the first stage. The ROC curves for the original signal are provided in Fig 5.8.

The area under the ROC curve for the time-domain compressed signal is 0.8125 after the first stage and 0.8092 after the second stage. Compared to the original signal, the area under the ROC curve has improved slightly after the first stage of

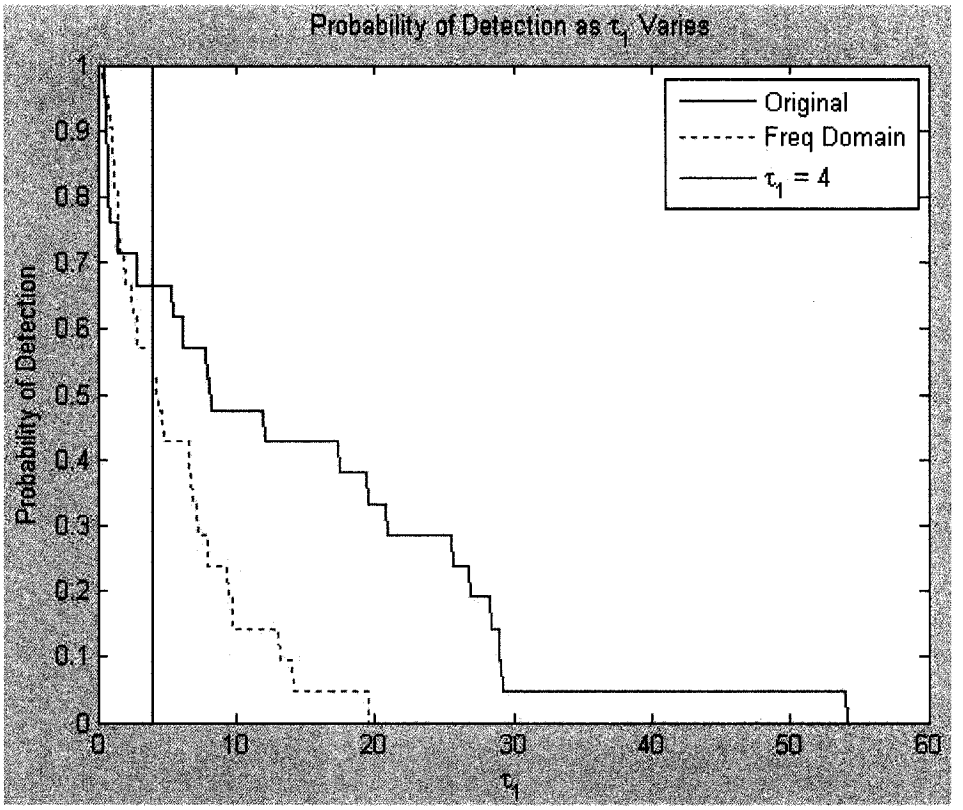

Figure 5.7: A comparison of the probability of detection for the vocalizations in the original signal and the frequency domain compressed signal. At the DRDC supplied threshold $\tau_1 = 4$, the original signal has a higher probability of detection than the frequency domain compressed signal.

detection, but has decreased after the second stage of detection. This indicates that the detection algorithm, given the parameters used, is a better test for the original signal. The ROC curves for the signal subjected to the time domain algorithm are provided in Fig 5.9.

In examining Figs 5.8, and 5.9, it is evident that the first stage of detection for the original signal has a slightly higher probability of detection at a very low probability of false alarm than the time domain compressed signal. However, the point at which the probability of detection reaches 1 for the time domain compressed signal is at a much lower probability of false alarm than for the original signal. Knowing that there are two vocalizations which are not detected in the time domain compressed signal, it is probable that an improvement in the modeling of the vocalization candidates

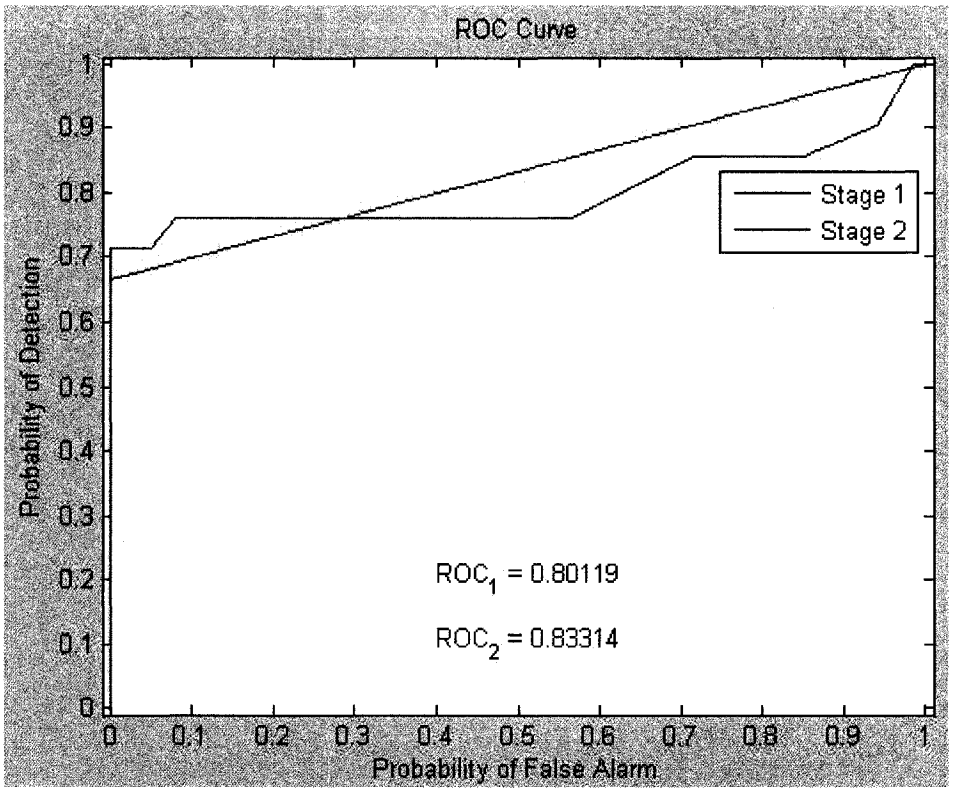

Figure 5.8: The area under the ROC curve measures the ability of the detection algorithm to correctly identify the state of having a marine mammal vocalization or not having a marine mammal vocalization. The blue curve shows the results after stage 1 of the DRDC Atlantic detection algorithm, and the red curve shows the results after stage 2 for the original signal.

would improve the area under the ROC curve for the time domain compressed signal. This may be as simple as providing a larger subset of vocalization candidates from the candidates developed.

The area under the ROC curve for the frequency domain compressed signal is 0.8951 after the first stage and 0.7555 after the second stage. Compared to the original signal, the area under the ROC curve has improved significantly after the first stage of detection, but has decreased significantly after the second stage of detection. This indicates that the detection algorithm, given the parameters used, is a better test for the original signal. The ROC curves for the signal subjected to the frequency domain algorithm are provided in Fig 5.10.

In examining Figs 5.8, and 5.10, it is evident that the first stage of detection for the

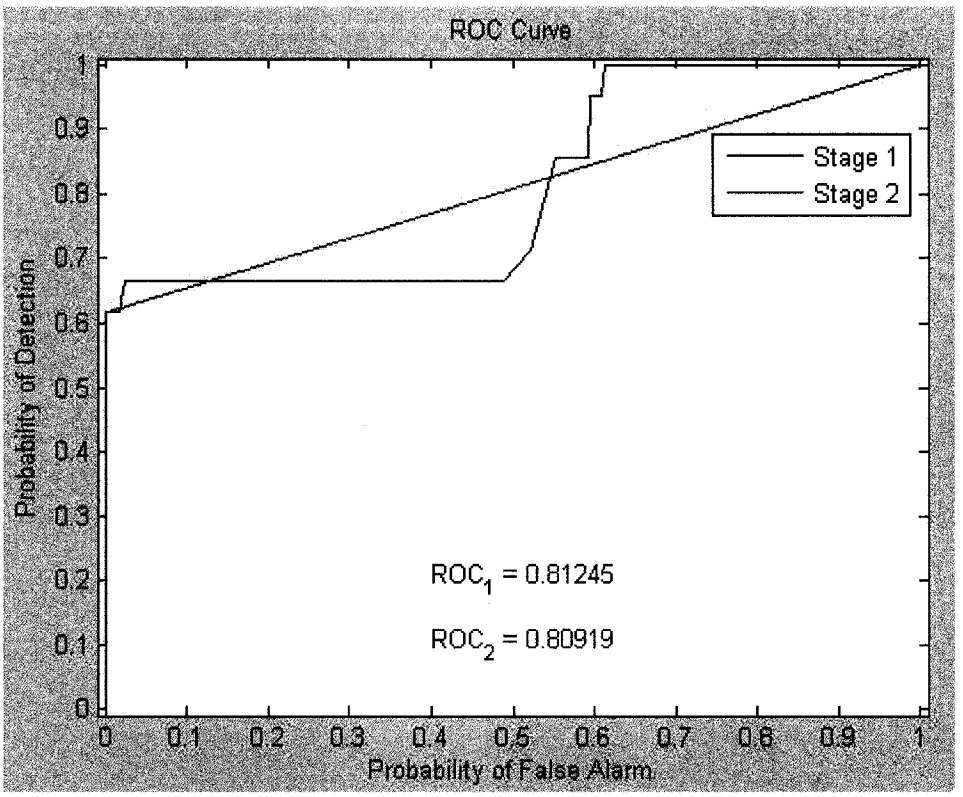

Figure 5.9: The ROC curves for the signal subjected to the time domain compression algorithm are similar to those of the original signal. The results after the first stage are better than the original signal, but after the second stage they are worse.

original signal has a higher probability of detection at a very low probability of false alarm than the time domain compressed signal. However, the probability of detection increases more steadily as the probability of false alarm is increased. The probability of false alarm at which the probability of detection reaches 1 is lower than for the original signal. Improvement in the second stage of the detection algorithm could be achieved by selecting a different threshold, however, the frequency domain algorithm would also benefit from modeling of the platform transient signals. Through examination of the LOFARgrams, it can be noted that the transient signals are modeled using vocalization candidates. The DCT coefficients selected do not adequately show the platform transient energy upon reconstruction, removing the energy in the guard band frequencies. This reduces the ability of the DRDC detection algorithm to differentiate between a true vocalization and a transient. Another possibility to

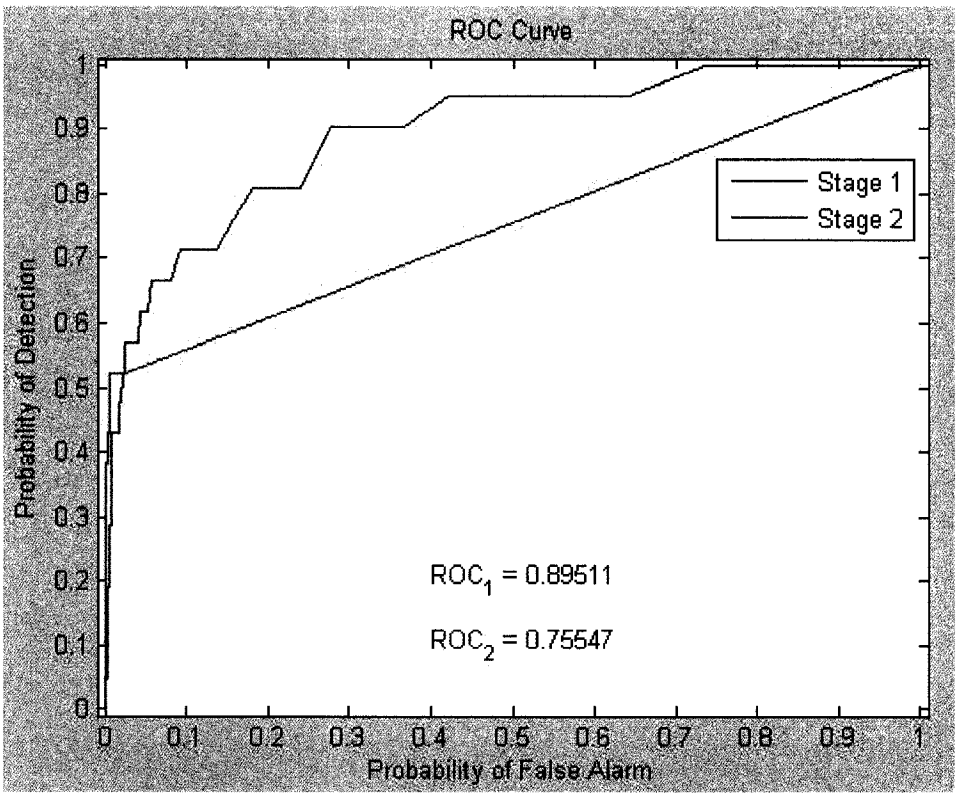

Figure 5.10: The area under the first stage of the ROC curve has improved significantly for the signal which has been compressed with the frequency domain compression algorithm as compared to the original signal. However, the area under the second stage of the ROC curve has deteriorated. In this case, the second stage of detection is actually detrimental to the detection of vocalizations in the reconstructed signal.

improve this defect would be to use Huffman coding rather than irregular sampling. This would keep the energy in the guard bands. This option is discussed in Section 6.2.4.

5.6 Significance Results

The value of τ_1 was set to 4, while τ_2 was set to 3 as specified by the scientists at DRDC. The significance results are calculated using the number of TP, FP, TN and FN resulting from running the detection algorithm at these thresholds. The values of TP, FP, TN and FN for each signal are given in Table 5.1

5.6.1 Standard Error

The standard error of the ROC of the original signal is 0.0549 for an area of 0.8331, meaning that 6.0674 standard deviations are required for the area under the ROC curve to equal 0.5. On a normally distributed curve, 99.994% of all data is within 4 STD of the mean. This standard error result is highly significant.

The standard error of the ROC curve for the signal compressed using the time domain algorithm is 0.0574 for an area of 0.8092, meaning that 5.3868 standard deviations are required for the area under the ROC curve to actually be 0.5. This is slightly worse than for the original signal, although both are highly significant.

The standard error of the ROC curve of the signal compressed using the frequency domain algorithm is 0.0616 for an area of 0.7555, meaning that 4.1477 standard deviations are required for the area under the ROC curve to equal 0.5. This is worse than for the original and time domain compressed signals, although still highly significant.

5.6.2 Curve Comparison

The standard error between the ROC curves of the original signal and the time-domain compressed signal after the second stage of the detection algorithm is 0.0196. As a result, 1.2939 standard deviations are required for the one curve to actually be the other. This indicates that the two curves are correlated, as 1.96 STDs is the standard cut-off for determining correlation.

The standard error between the ROC curve of the original signal and the frequency domain compressed signal is 0.0355. As a result, 2.1875 standard deviations are required for the curves to equal each other. The two curves are not correlated.

5.6.3 Accuracy

The accuracy of the original signal is 0.9989. The accuracy of the signal compressed using the time domain algorithm is 0.9992. This is slightly better than the accuracy of the original signal. The accuracy of the signal compressed using the frequency domain

algorithm is 0.9935. This is slightly worse than the accuracy of the original signal. All three accuracy values are very high, however, since the number of segments without vocalizations is significantly higher than the number of segments with vocalizations, it is necessary to look at further significance results. Using the modified accuracy, which gives equal weighting to the segments with vocalizations and the segments without vocalizations, the accuracies are adjusted to 0.7856, 0.7857, 0.7591 for the original, time domain compressed and frequency domain compressed signals respectively.

5.6.4 Positive Predictive Value

The PPV of the original signal is 0.8. This value is much lower than the accuracy, because the number of segments with vocalizations present is 21, where as the number of segments without vocalizations is 11379. While the percentage of segments which have false alarms, may be small, when the number of FP is compared to the number of TP, even a small fraction of the segments without vocalizations being shown as FP will be large compared to the number of TP. The PPV of the signal compressed using the time domain algorithm is 1. This is significantly better than the PPV of the original signal. The PPV of the signal compressed using the frequency domain algorithm is 0.1467. This is significantly worse than the PPV of the original signal. This value is so low because there are 64 segments which are FPs. In comparison, the original signal has 3 FPs and the time domain compressed signal has 0 FPs. As previously discussed, the frequency domain algorithm needs better transient modeling to avoid this scenario.

5.6.5 Negative Predictive Value

The NPV is an important measure for this experiment. In any typical underwater acoustic signal, the number of segments which have marine mammal vocalizations will be small compared to the number which will have no vocalizations. For this reason, the reduction of false alarms is extremely important, as even a small reduction in the probability of false alarms will have a large impact on the number of detections. The

NPV of the original signal is 0.9992. The NPV of the signal compressed using the time domain algorithm is also 0.9992. The NPV of the signal compressed using the frequency domain algorithm is 0.9991. This is slightly worse than the NPV of the original signal. All three NPV are high.

5.6.6 Sensitivity

The sensitivity of the original signal is 0.5714. The sensitivity of the signal compressed using the time domain algorithm is also 0.5714. The sensitivity of the signal compressed using the frequency domain algorithm is 0.5238. This is slightly worse than the sensitivity of the original signal. These results are all somewhat low because the detection algorithm cannot detect more of the marine mammal vocalizations at the thresholds, τ_1, and τ_2 which were chosen. This indicates that the difference in energy between the segments with marine mammal vocalizations and the segments without is not high enough to register at detection at these thresholds.

5.6.7 Specificity

The specificity of the original signal is 0.9997. The specificity of the signal compressed using the time domain algorithm is 1, which is slightly higher than the specificity of the original signal. The specificity of the signal compressed using the frequency domain algorithm is 0.9944. This is slightly worse than the specificity of the original signal. All three specificity values are high because there are not very many FP relative to the number of TN segments.

5.6.8 Phi Coefficient of Association

The Phi coefficient of association of the original signal is 0.68. Any value of Phi which is greater than 0.60 denotes a strong association. The Phi coefficient of association of the signal compressed using the time domain algorithm is 0.76. This is better than the Phi coefficient of association of the original signal. The Phi coefficient of association of the signal compressed using the frequency domain algorithm is 0.27.

This denotes weak association and is significantly worse than the Phi coefficient of association of the original signal.

5.7 Further Results

The initial tests were all completed using the detection parameters provided by the scientists at DRDC. However, some work was performed using different modifications to the detection algorithm, in order to examine the effects of changing the detection algorithm on the performance by the original, time domain compressed and frequency domain compressed signals. One set of the modified results is provided in Table 5.2. For these results, the only parameters changed were the selection of the detection thresholds, τ_1, and τ_2, which were set to 0.49 and 1.69 respectively. The first threshold, τ_1 was selected because it is the highest value which retains a probability of detection of 1. A threshold of 1.69 was selected for the second threshold, τ_2. This was determined because it is a threshold which has a low number of false alarms, while maintaining a fairly high detection rate.

A second modification of the detection algorithm is performed for the frequency domain algorithm only. The first modification involved replacing the PSD with the magnitude of the FFT. Since the values of the candidates are so small, squaring them makes them even smaller. This process was enough to reduce the area under the ROC curve by about 6%. The second modification was to increase the BW from the range $17 - 35$ kHz to $17 - 38$ kHz. This modification only increased the area under the ROC curve by a few tenths of a percentage point, but the combination of these two modifications allowed the selection of a higher threshold, $\tau_1 = 2.59$ to perform the significance calculations. The second stage of the detection algorithm is removed as it does not improve the area under the ROC curve.

A third set of results is given for both the time and frequency domain algorithms. These results were derived from the reconstruction itself. Each segment which used a vocalization candidate is considered a detection. Since both time and frequency domain algorithms use segment length of $N = 1024$, the segments are longer, and

Signal		Original	Time D	Freq D	Freq D2	FOS TD	FOS FD
CR		1	93.06	5.5652	5.5652	93.06	5.5652
# Segments w Voc		21	21	21	21	20	20
Segment Length		128	128	128	128	1024	1024
ROC	Stage 1	0.8012	0.8125	0.8951	0.9538	N/A	N/A
ROC	Stage 2	0.8922	0.9030	0.8545	N/A	N/A	N/A
SE		0.0465	0.0445	0.0522	0.032	N/A	N/A
STD from ROC=0.5		8.4344	9.0562	6.7912	14.179	N/A	N/A
SE between Curves		N/A	0.0371	0.0271	0.025	N/A	N/A
STD between Curves		N/A	1.8828	1.3195	2.4623	N/A	N/A
TP		15	15	15	13	19	16
FN		6	6	6	8	1	4
TN		11367	11379	10664	11189	1237	1110
FP		12	0	715	190	168	295
Accuracy		0.9984	0.9995	0.9368	0.9826	0.8814	0.7902
Mod Accuracy		0.8566	0.8571	0.8257	0.8012	0.9152	0.795
PPV		0.5556	1	0.0205	0.064	0.1016	0.0514
NPV		0.9995	0.9995	0.9994	0.9993	0.9992	0.9964
Sensitivity		0.7143	0.7143	0.7143	0.619	0.95	0.8
Specificity		0.9989	1	0.9372	0.9833	0.8804	0.79
Phi		0.63	0.84	0.11	0.2	0.29	0.17

Table 5.2: Results for the original, time domain compressed and frequency domain compressed signals using a different set of thresholds. A second set of results for the frequency domain algorithm using a modified detection algorithm are also provided. Finally, there are results derived using the candidates fit by FOS. The segment length for these results is 1024, as this is the length of segment on which the compression algorithms were performed.

there are fewer of them. This gives a coarser detection than is achieved using the detection algorithm, however, the results are comparable.

5.7.1 Receiver Operating Characteristic Curve

The area under the ROC curve for the original signal after the first stage of detection is 0.8012. After the second stage of detection the area under the ROC curve improves to 0.8922 since the false alarm rate for a given detection level is reduced for the second stage as compared to the first stage.

The area under the ROC curve for the time-domain compressed signal is 0.8125 after the first stage and 0.9030 after the second stage. It can be seen that the area under the ROC curve has improved slightly after each stage of detection as compared

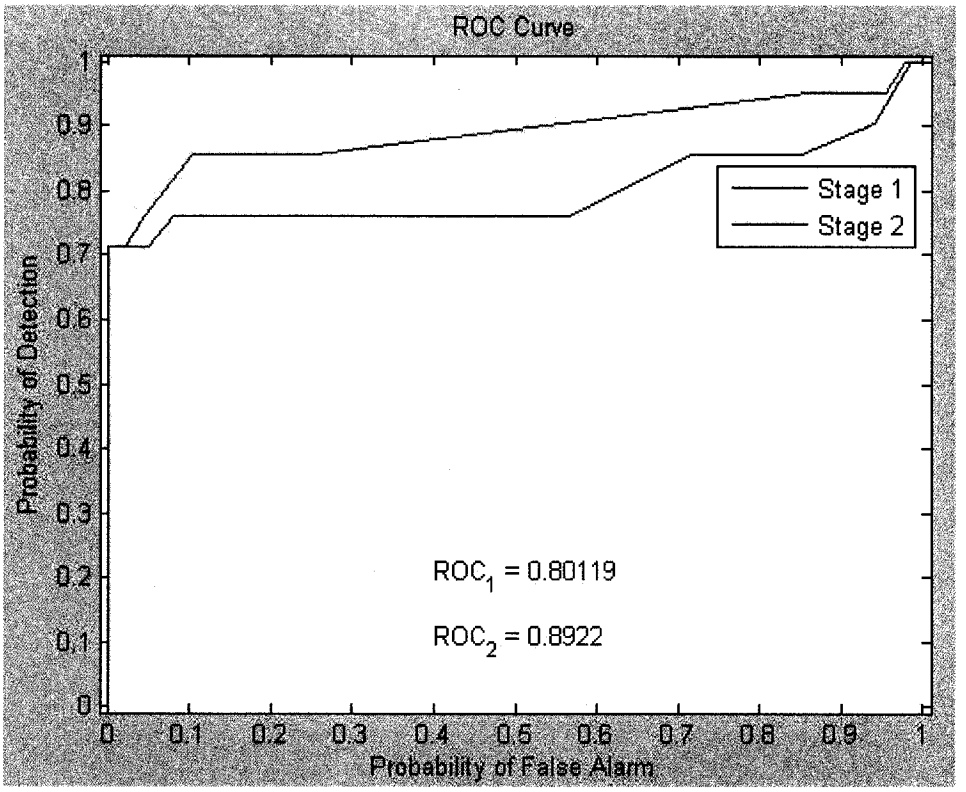

Figure 5.11: The area under the ROC curves for the original signal using the threshold values of $\tau_1 = 0.49$, and $\tau_2 = 1.69$.

to the original signal. This indicates, that despite the signal compression, there has been no deterioration to the signal in terms of ability to detect the marine mammal vocalizations. In fact, the ability to detect these vocalizations has increased slightly.

The area under the ROC curve for the frequency domain compressed signal is 0.8951 after the first stage and 0.8545 after the second stage. It can be seen that the area under the ROC curve has improved significantly after the first stage of detection, however it has been reduced after the second stage. This is because the threshold selected for the second stage of detection is not the optimal threshold for this particular ROC curve. The irregular sampling has modified the properties of the signal in terms of the guard band region. Since a large amount of the random energy has been removed from the reconstructed signal, it has changed the properties in terms of detection.

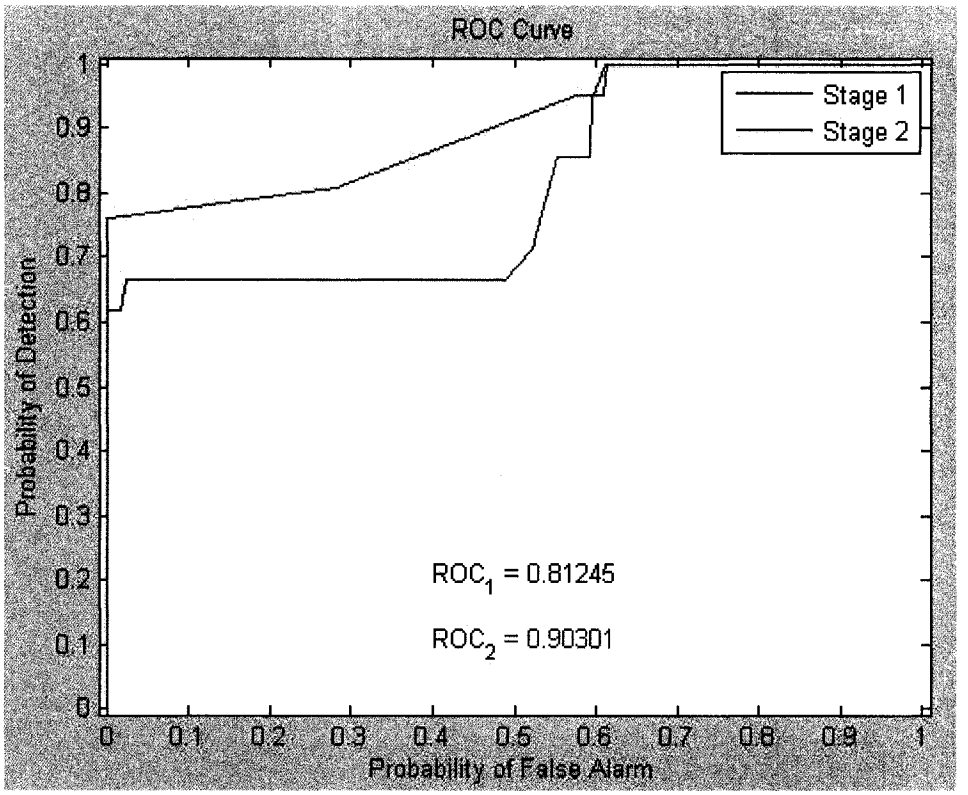

Figure 5.12: The area under both the first and second stages of the ROC curve have improved slightly for the signal which has been compressed with the time domain compression algorithm as compared to the original signal, using the threshold values of $\tau_1 = 0.49$, and $\tau_2 = 1.69$.

Using the additional modifications for the frequency domain compressed signal, the area under the ROC curve is 0.9538 after the first stage. This area is significantly better than for either stage of detection of the original signal. The second stage of detection is not performed in this case because transient signals are not modeled by the frequency domain algorithm. The transient signals are most closely modeled by the vocalization candidates. This means that they are no longer distinguishable from true vocalizations and cannot be eliminated using the second stage of the DRDC Atlantic detection algorithm.

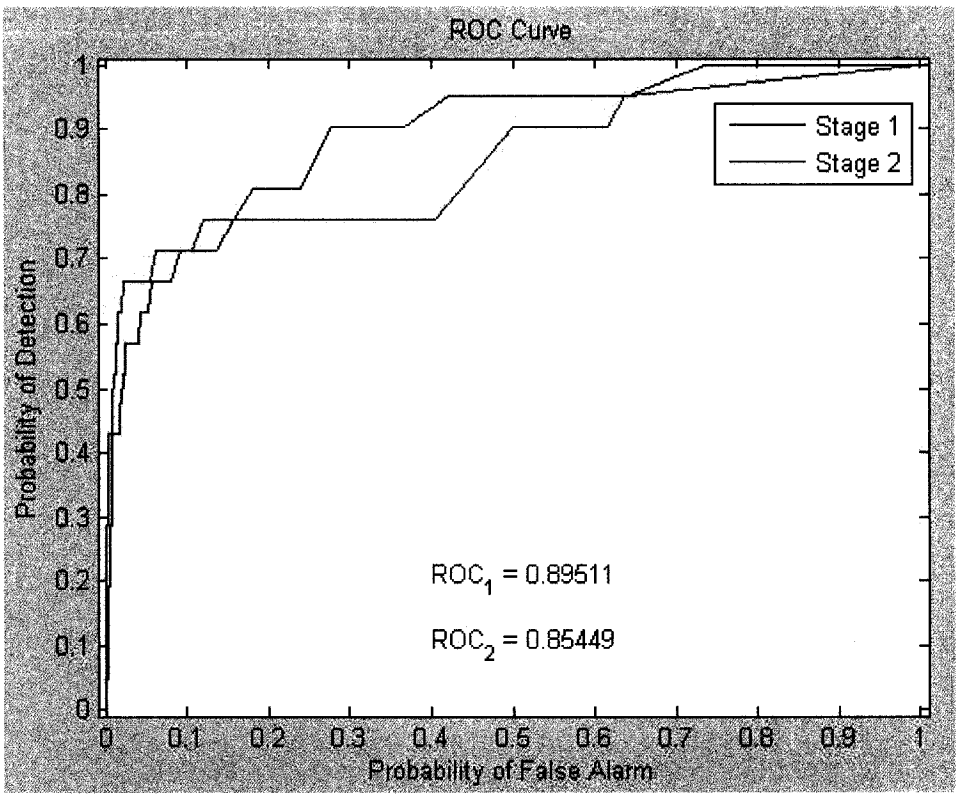

Figure 5.13: The area under the first stage of the ROC curve has improved significantly for the signal which has been compressed with the frequency domain compression algorithm as compared to the original signal. However, the area under the second stage of the ROC curve has deteriorated. In this case, the second stage of detection is actually detrimental to the detection of vocalizations in the reconstructed signal.

5.7.2 Significance Results

The significance results for the modified detection parameters are not significantly different than for the original detection parameters.

The standard error results in a number of STDs between the derived areas and an area of 0.5 are all very high, as indicated in Table 5.2.

The standard error between the time and frequency domain compressed signals and the original signal result in correlated signals with the exception of the frequency domain compressed signal using the second detection modification.

The accuracy of all the signals is extremely high, however, the time domain compressed signal using the modified detection parameters has the highest accuracy of

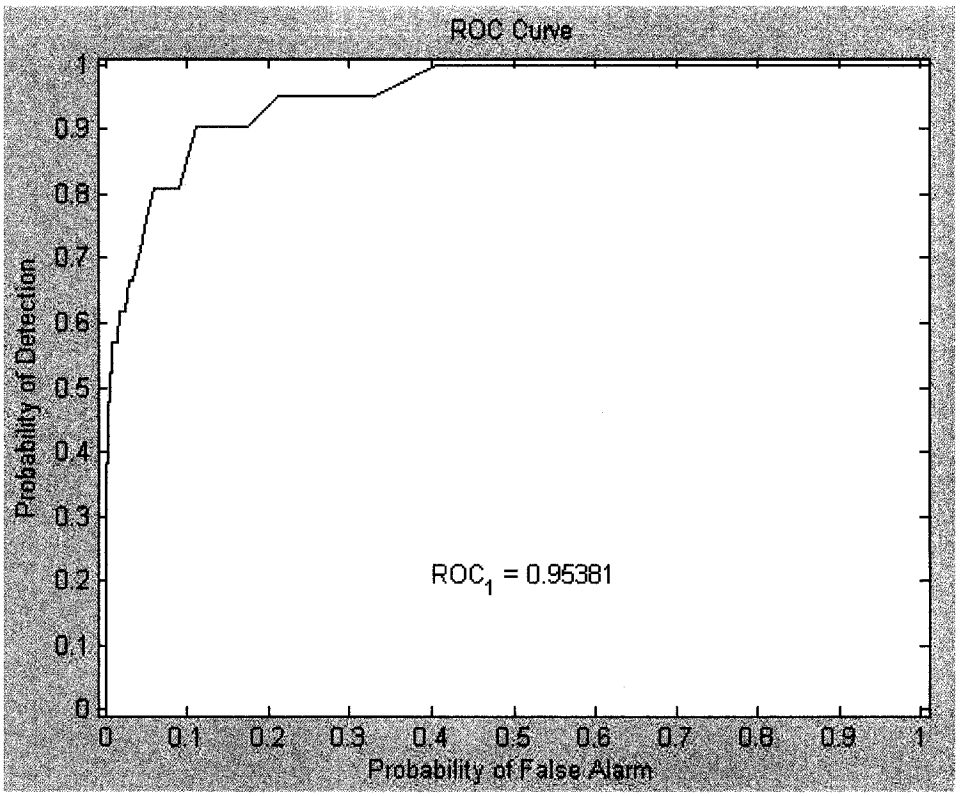

Figure 5.14: The area under the first stage of the ROC curve has improved significantly for the signal which has been compressed with the frequency domain compression algorithm as compared to the original signal. The second stage of the detection algorithm is not run because it is not effective in removing false alarms for the frequency domain, modified signal.

all candidates. Using the modified accuracy, the FOS detector for the time domain compressed signal has the highest value, although all are above 0.79, as indicated in Table 5.2.

The PPV is worse in each case except for the time domain compressed signal. This signal, like with the original detection parameters, has a value of 1.

The NPV is better in all cases, except the FOS detector for the frequency domain compressed signal than any of the NPV values using the original detection parameters.

The sensitivity is better in each case.

The specificity is worse in each case except for the time domain compressed signal. For this case, the specificity is 1 which is better than for the original signal, and equal to the time domain algorithm using the original detection parameters.

The Phi coefficient of association is also worse in each case except for the time domain compressed signal, which is once again better than the Phi coefficient of association achieved for any of the signals using the original detection parameters.

5.7.3 Conclusions

While the modified detection parameters achieve a higher area under the ROC curves of the original, the time domain compressed and the frequency domain compressed signals, the significance tests tend to be worse. The exception is the time domain compressed signal. The modification of the detection thresholds improved the area under the ROC curve as well as all the significance results for this signal. This signal performs better at each test than both the time domain compressed signal using the provided detection parameters, and the original signal, using either set of detection parameters.

This experiment demonstrates the requirement to carefully select the detection algorithm parameters, and the effect it can have on the ability to detect marine mammal vocalizations in the underwater acoustic signal.

Chapter 6

Discussion and Conclusions

6.1 Summary

The detection and classification of marine mammal vocalizations can provide valuable information as to the marine activity in the region in which a sonobuoy or AUV is located. The signal of interest contains a series of features which must remain recognizable after data compression has been applied. The features are dependent on the marine activity present during data acquisition.

The purpose of this thesis was to determine a method of achieving data compression for underwater acoustic data while maintaining the integrity of the marine mammal vocalizations sought. Two algorithms were presented, the time domain algorithm provides a better compression ratio but has greater signal processing requirements, while the frequency domain algorithm is more practical for platforms with less on board processing power such as AUVs.

On the training signal, the time domain algorithm achieved a CR of 93.1, while still detecting these clicks and buzzes. Obviously, these results could not have been achieved simply by retaining every 93rd point because then most of the clicks and buzzes would have been entirely missed and the remainder could each contribute only one point. The frequency domain algorithm achieved a CR of 5.5. The original signal had a CR of 1, since there is no compression applied.

There are performance gains for both algorithms in terms of CR and therefore in the speed of transmission. The time domain algorithm has a higher area under the ROC curve, using the modified detection parameters, and a slightly lower area under the ROC curve using the provided detection parameters. However, in both cases, it is more significant according to the significance tests run. The frequency domain algorithm suffers from a degradation in detection performance when the same thresholds are used. It also performs worse in the significance tests, especially PPV. This

algorithm would likely benefit from additional model terms, and Huffman encoding to replace the irregular sampling.

6.1.1 Lessons Learned

The FOS algorithm, both on its own, and when combined with irregular sampling and domain transformation proved to be a useful tool in order to achieve data compression. However, the ability of both algorithms to provide useful signals after compression is dependent on an appropriate selection of candidate functions as inputs to FOS. As such, an understanding of the operating environment, as well as a priori knowledge of the features of interest is required for these algorithms to perform in an acceptable manner.

6.2 Future Work

There are a number of improvements which could be implemented to make this algorithm more robust for use by DRDC Atlantic. These include hardware implementation of the algorithm, and extension to further vocalization sets, including further modeling of marine mammal vocalizations. Development of the detection algorithm could also be performed to ensure its performance with the reconstructed data set. For the frequency domain algorithm Huffman coding should be considered as a means of increasing compression. Finally, independent verification of the truth sets would be beneficial.

6.2.1 Hardware Implementation

The entire algorithm would benefit from implementation on a hardware platform such as an field programmable gate array (FPGA). This would speed up the time taken to perform the operations required to prepare the data for transmission. Particular effort should be applied to make the candidate terms parametric as inputs to FOS. With equations rather than points as input, the performance of this particular aspect

of the FOS algorithm will be improved because the correlation computations may be performed more quickly.

6.2.2 Extension to Further Vocalization Sets

A limited data set was used during this thesis. Only a few types of vocalizations, from Blainville's Beaked whales were analyzed. Extension of the candidate sets to include other whale vocalizations would expand the usefulness of this algorithm. This should include modeling of the marine mammal vocalizations to determine a set of candidates which accurately represent the specific vocalizations sought.

Modeling of platform transient signals should also be incorporated. There are certain transient signals, produced by the platform, which are known to occur. Their modeling would improve the ability to reduce false alarms during the second stage of the DRDC Atlantic detection algorithm. During the reconstruction of the frequency domain compressed signal, some transients are reconstructed as marine mammal vocalizations because these are the model terms which best fit the energy. This would not be the case if there were candidate functions which modeled the transient signals.

Finally, the modeling of the current vocalizations could be improved. The Gaussian window which was applied to the vocalization candidates for the time domain algorithm was set with the default values, this could likely be improved. The frequency modulation rates were also rather coarse. Further work could be done to find a better model of the vocalizations modeled in this thesis.

6.2.3 Detection Algorithm

Ideally, since FOS uses candidates which were developed to model specific marine mammal vocalizations, their use in a particular segment could be considered a detection. False alarms would then need to be eliminated using a detection algorithm similar to the second stage of the detection algorithm developed by DRDC Atlantic. This could improve the efficacy with which detections are made.

6.2.4 Huffman Coding

The frequency domain algorithm uses irregular sampling to remove the coefficients of the DCT candidates which are statistically insignificant, or which are not in the frequency regions of interest. Another method of achieving data compression would be to use Huffman coding, as is done for the JPEG algorithm. This algorithm encodes zero-run lengths, as well as assigning the most commonly used values to shorter symbols. It is possible, that Huffman coding would increase the compression ratio achieved for the frequency domain algorithm. It would also be a means of retaining energy caused by transient signals, and thereby improving the detection capabilities of the DRDC detection algorithm for the frequency domain compressed signal.

6.2.5 Truth Data

I manually determined the truth data for the data sets provided. Independent verification of the truth sets would improve the veracity of the results provided.

References

[1] M. J. Korenberg, C. J. H. Brenan, and I. W. Hunter, "Nonuniform sampling for spectral and related applications," December 1999. [Online]. Available: http://www.freepatentsonline.com/6005664.html

[2] D. 4-4, "Marine mammal mitigation procedures," apr 2008.

[3] P. R. Oliveira and R. F. Romero, "A comparison between pca neural networks and the jpeg standard for performing image compression," in *Second Workshop on Cybernetic Vision*, Dec. 1996, pp. 112–116.

[4] P. R. Oliveira and J. Mazucheli, "Techniques for image compression: a comparative analysis," in *Sixth Brazilian Symposium on Neural Networks*, Nov. 2000, pp. 249–254.

[5] P. R. Oliveira, "Interpolation of signals with missing data using pca," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 3, May 2006, pp. 828–831.

[6] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *IEEE Computer Society Press*, Los Alamitos, CA, USA, 1993.

[7] A. Desrochers, "On an improved model reduction technique for nonlinear systems," *Automatica*, vol. 17, pp. 407–409, 1981.

[8] H. A. Abbas, "Time series analysis for ecg data compression," in *International Conference on Acoustics, Speech, and Signal Processing*, vol. 3, Phoenix, AZ, USA, Mar. 1999, pp. 1537–1540.

[9] R. J. Urick, *Principles of Underwater Sound*, 3rd ed. USA: McGraw-Hill, 1983.

[10] R. W. B. Stephens, *Underwater Acoustics*. Toronto, ON, CA: Wiley-Interscience, 1970.

[11] J. Barlow, "Trackline detection probability for long-diving whales," in *Marine Mammal Survey and Assessment Methods*, vol. 1, 1999, pp. 209–221.

[12] M. P. Johnson, P. T. Madsen, W. M. Zimmer, N. A. de Soto, and P. L. Tyack, "Foraging blainville's beaked whales (mesoplodon densirostris) produce distinct click types matched to different phases of echolocation," *The Journal of Experimental Biology*, vol. 24, pp. 5038–5050, Dec. 2006.

[13] W. M. X. Zimmer, P. T. Madsen, V. Teloni, M. P. Johnson, and P. L. Tyack, "Off-axis effects on the multipulse structure of sperm whale usual clicks with implications for sound production," *Journal of the Acoustical Society of America*, vol. 118, pp. 3337–3345, Nov. 2005.

[14] M. Johnson, P. T. Madsen, W. M. X. Zimmer, N. A. de Soto, and P. L. Tyack, "Foraging blainville's beaked whales (mesoplodon densirostris) produce distinct click types matched to different phases of echolocation," *The Journal of Experimental Biology*, vol. 209, pp. 5038–5050, Oct. 2006.

[15] K. R. Rao and P. C. Yip, *The Transform and Data Compression Handbook*, ser. The Electrical Engineering and Signal Processing Series. Boca Raton, Fla, USA: CRC Press, 2001.

[16] M. K. Mandal, *Multimedia Signals and Systems.* Boston, MA, USA: Springer, 2003.

[17] Y. Q. Shi, H. Sun, and P. A. Laplante, *Image and Video Compression for Multimedia Engineering: Fundamentals*, ser. Image Processing Series. New York, NY, USA: CRC Press, 2000.

[18] M. J. Korenberg, "A robust orthogonal algorithm for system identification and time-series analysis," *Biological Cybernetics*, vol. 60, pp. 267–276, Feb. 1989.

[19] K. H. Chon, "Accurate identification of periodic oscillations buried in white or colored noise using fast orthogonal search," *IEEE Transactions on Biomedical Engineering*, vol. 48, pp. 622–628, Jun. 2001.

[20] M. J. Korenberg and L. D. Paarmann, "Applications of fast orthogonal search: Time-series analysis and resolution of signals in noise," *Annals of Biomedical Engineering*, vol. 17, pp. 219–231, May 1989.

[21] M. J. Korenberg, "Fast orthogonal identification of nonlinear difference equation and functional expansion models," in *Proceedings of Midwest Symposium on Circuits and Systems*, vol. 1, Aug. 1987, pp. 270–276.

[22] M. J. Korenberg, C. J. H. Brenan, and I. W. Hunter, "Raman spectral estimation via fast orthogonal search," *The Analyst*, vol. 122, pp. 879–882, Sep. 1997.

[23] M. J. Korenberg and K. M. Adeney, "Iterative fast orthogonal search for modeling by a sum of exponential sinusoids," *Annals of Biomedical Engineering*, vol. 26, pp. 315–327, Mar. 1998.

[24] M. J. Korenberg, "Fast orthogonal algorithms for nonlinear system identification and time-series analysis," *Advanced Methods of Physiological System Modeling*, vol. 2, pp. 165–179, 1989.

[25] K. R. Rao and P. Yip, *Discrete Cosine Transform: Algorithms, Advantages, Applications.* Toronto, ON, CA: Academic Press, 1990.

[26] J. Proakis, *Digital Signal Processing*, 4th ed. Upper Saddle River, N.J., USA: Pearson Prentice Hall, 2007.

[27] C. Slivinsky and J. Borninski, "Control system compensation and implementation with the tms32010," *Digital Signal Processing Applications with the TMS320 Family, Theory, Algorithms and Applications*, vol. 1, pp. 691–722, 1989.

[28] H. Choi, "Fixed point arithmetic and q format," Aug. 2005. [Online]. Available: http://cnx.org/content/m10919/2.2/

[29] D. Zuras and M. Cowlishaw, "Ieee standard for floating-point arithmetic," New York ,NY, USA, p. 58, Aug. 2008.

[30] J. D. Hood and D. G. Flogeras, "Improved passive band-limited energy detection for marine mammals."

[31] K. M. Adeney and M. J. Korenberg, "Iterative fast orthogonal search algorithm for mdl-based training of generalized single-layer networks," *Neural Networks*, vol. 13, pp. 787–799, Jun. 2000.

[32] J. A. Theriault, J. Hood, D. Moretti, N. DiMarzio, D. Mosher, and T. Murphy, "Detection of blainville's beaked whale (mesoplodon densirostris) using autonomous underwater gliders."

[33] W. W. L. Au, *The Sonar of Dolphins*. New York, NY, USA: Springer, 1993.

[34] J. M. Wozencraft and I. M. Jacobs, *Principles of Communication Engineering*. Long Grove, Ill, USA: Waveland Press Inc., 1965.

[35] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*. Toronto, ON, CA: McGraw-Hill, Inc, 1991.

[36] J. A. Hanley and B. J. McNeil, "The meaning and use of the area under a receiver operating characteristic (roc) curve," *Radiology*, vol. 143, pp. 29–36, Apr. 1982.

[37] ——, "A method of comparing the areas under receiver operating characteristic curves derived from the same cases," *Radiology*, vol. 148, pp. 839–843, Sep. 1983.

[38] L. Hopley and J. van Schalkwyk, "The magnificent roc," Mar. 2007. [Online]. Available: http://www.anaesthetist.com/mnm/stats/roc/Findex.htm