Acoustic Detection of Blue Whales

Diplomarbeit

Isabella Biedermann

Betreuer:
Dipl.–Ing. Dr. Thomas Thurner
Begutachter:
A.o. Univ.–Prof. Dipl.–Ing. Dr. Gerhard Graber

Technische Universität Graz

Graz, im April 2014

Contents

Nomenclature

Abbreviations

FFT	Fast Fourier Transform
DFT	Discrete Fourier Transform
FM	Frequency Modulation
ISC	Improved Spectrogram Correlation
SC	Spectrogram Correlation
SNR	Signal to Noise Ratio
MDR	Missed Detection Rate
FAR	False Acceptance Rate
SIO	Scripps Institution of Oceanography
ARP	Acoustic Recording Package
FIR	Finite Impulse Response

Abstract

Blue whales in the North-East pacific call with specific low-frequent down sweeps, mentioned as "B-calls". These utterances have a strong third harmonic and prove to be of different pitch in different years. This thesis aims at evaluating signal processing algorithms for a robust detection and identification of blue whale calls in continually recorded acoustic underwater sounds.

Some of the evaluated algorithms work connected to the signal's shape: A matched filter technique leads to a maximum likelihood estimation down sweep filter bank with variation in sweep rate and mean frequency. In frequency domain, spectrogram correlation is evaluated, as well as, on account of pitch changes, an improved spectrogram correlation method using a maximum likelihood detection for varying sweep mean frequencies. Cepstrum and Delta Cepstrum calculation as used in speech recognition is evaluated for being useful to the specific blue whale signals. Power detection in interesting bands in time domain and two Goertzel-algorithm techniques for calculating DFT points in frequency bands approach the detection over calculating signal to noise ratio in interesting frequency bands.

An evaluation of the presented techniques shows that combined algorithm of the spectrogram correlation method with a maximum-likelihood detection proves to work as a robust detector producing the best results among the evaluated methods for recognizing blue whale B-calls.

Kurzfassung

Der Blauwal als größtes jemals auf der Erde lebende Tier produziert die tiefsten Töne aller Lebewesen. Blauwalschreie im Nord-Ost-Pazifik sind durch eine bestimmte Signalform von tieffrequenten Downsweeps mit einer starken dritten harmonischen Komponente gekennzeichnet, deren Stimmung sich von Jahr zu Jahr ändert. Ziel dieser Diplomarbeit ist die Untersuchung von Signalverarbeitungs-Algorithmen zur robusten Erkennung und Identifikation von Blauwal-Schreien, im Speziellen Blauwal B-Calls in kontinuierlich aufgenommenen akustischen Unterwassersignalen.

Die betrachteten Techniken, denen die Erkennung der Signalform im Zeitbereich zu Grunde liegt, werden durch eine Matched-Filter Methode und einer daraus folgenden Maximum-Likelihood-Schätzung mit Sweeprate und Mittenfrequenz als Parameter repräsentiert. Im Frequenzbereich wird Spektrogrammkorrelation angewendet und daraus eine weitere Methode, die ebenso den Maximum-Likelihood-Ansatz in Bezug auf Variation der Mittenfrequenzen verfolgt, entwickelt. Cepstrum bzw. Delta-Cepstrum-Extraktion wird bezüglich der Anwendbarkeit auf die spezifischen Signale untersucht. Einem anderen Ansatz folgt eine Methode, die durch Filterung im Zeitbereich und Berechnung eines bandbezogenen Rauschabstandes als Kriterium für die Erkennung verwendet. Als Pendant im Frequenzbereich steht dem gegenüber eine Methode, die mit Hilfe des Goertzel-Algorithmus einzelne interessante Bereiche der DFT herausnimmt und ebenso dort den Rauschabstand bewertet.

Die Evaluierung aller untersuchten Methoden zeigt, dass ein gekoppelter Algorithmus, der die Spektrogrammkorrelation mit einem Maximum-Likelihood-Schätzer verbindet, als robuster Detektor funktioniert und sich als beste Methode zur Erkennung für Blauwal B-Calls herausstellt.

Statutory Declaration

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly marked all material which has been quoted either literally or by content from the used sources.

Eidesstattliche Erklärung

Ich erkläre an Eides statt, dass ich die vorliegende Arbeit selbstständig verfasst, andere als die angegebenen Quellen/Hilfsmittel nicht benutzt, und die den benutzten Quellen wörtlich und inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Graz, am ....

(Unterschrift)

Acknowledgements

First of all, I want to thank Dr. Thomas Thurner for helping to "make it", finally. Special Thanks to Prof. Gerhard Graber, who is definitely right in telling that also deadlines have assets. Thanks also to Dr. Daniel Watzenig, who helped me seeing through all those requirements back then in San Diego. Thanks to Dr. Tom Avsic and Prof. Uwe Send I had this opportunity of "whale-listening", which made this thesis possible.
Last but not least: Thanks to my family who helped me throughout my studies, and to my daughter who taught me how to love so much.

Preface

Reading Science Fiction belongs to one of my favourite activities to distract myself from learning and working. Some years ago, when I was studying for my final exams, the book 'Der Schwarm' by Frank Schätzing [20] made my nights I would have needed to relax a time of gleeful dispense. In my mind I was traveling through the ocean, imagining how the several different species living there interact with each other, wondering about their intelligence as well as their life they spend in the depths of the sea, where nearly no human being has ever been before. My interests have always been of an inter-disciplinary nature which was the main reason for starting the audio-engineering course taught at both Graz University of Technology and Graz University of Performing Arts. Therefore I decided to connect some of my different inter-disciplinary interests and ask for an acoustically related Diploma thesis at the 'Leibniz-Institut für Meereswissenschaften' in Kiel, which was mentioned in Schätzing's book as an institution of ocean research. Professor Uwe Send, who had been the head of the acoustic group in Kiel, in the meantime had accepted the invitation to become a professor at Scripps Institution of Oceanography (SIO) in San Diego, where I got my answer from.

Chapter 1

Introduction

Blue Whales are the hugest mammals of the world. As they are members of a species who are living most of their time in the depth of the sea, a fact which makes them hardly tangible to our visual detections, different approaches than only watching them have to be used to gain some deeper knowledge. Acoustic detection makes the opaque deep sea transparent for many sources. Like ultrasonic applications are used in everydaylife situations to make things unseeable objects visible, acoustic whale detection helps tracking the mammals without eye contact.

1.1 Motivation of Listening to Blue Whales

'Why should Blue Whales be surveyed acoustically? Besides, have they not already been surveyed?' someone might ask. The answer for the first question is rather simple as most of the time Blue Whales dive in the depths of the sea. Thus they cannot be observed visually there and that is why an acoustical monitoring method is profitably. Still, there are already several systems which have been monitoring ocean life acoustically (e.g. *HARP* or *ARP* systems which are also used for the *SCORE* project (see Appendix A) . However, a fully automatic tracking system has not yet been introduced. To count, watch and survey the worldwide Blue Whale stock automatically helps to prevent this endangered species from extinction. In order to monitor the salinity, sea state and temperature, a broad spread carpet of so-called floats is installed in the ocean all over the world (Figure 1.1). The floating buoys can stay under water for two weeks in about 1000 m depth, after this time they ascend to the surface and send the collected data via satellite (see [14] for online information)).
The idea is to equip some floats with hydrophones and a processing unit to automatically monitor ocean organisms, in the first trial periode Blue Whales in the North-East-Pacific.

Concerning whale monitoring, the problem has been split into two parts:

- Detection finds out if there is a Blue Whale call present
- Localization demonstrates where the call comes from

The first part deals with evaluating different algorithms to find a Blue Whale call. The second part will not be subject to this thesis.
The possibility of using whale sounds for ocean acoustic tomography will be evaluated within the temporal accuracy of the detection algorithms.

1.1.1 Problem Description

This motivation leads to a more detailed the conceptual formulation for the workload of this thesis:

- Description and evaluation of the recorded signals: The recorded signals are examined concerning sample rate, signal power, bit rate.
- Evaluation of signals and signal parameters: The interesting parameters and unique features of the Blue Whale calls have to be emphasized, to provide a robust identification of the signals.
- Analysis of signal processing algorithms for Blue Whale call detection: A theoretical background for the later used methods should be offered.
- Evaluation of the introduced algorithms: The performance of the algorithms concerning calculation cost, detection ability and accuracy should be evaluated.

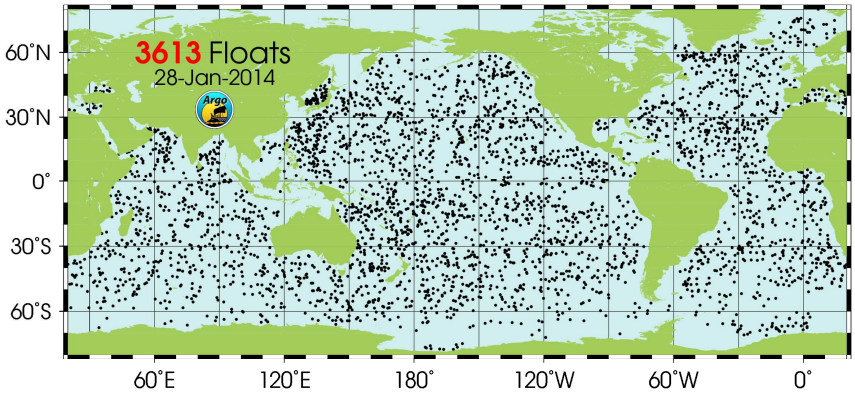


Figure 1.1: Worldwide Float Distribution, ©2014, SIO, see [22]

1.2 Chapter Overview

Succeeding the introduction, chapter 2 offers a basic overview of blue whale sound production as well as underwater sound propagation, and introduces ocean tomography. Chapter 3 provides the theoretical background of the mathematical and signal processing methods used in this thesis. Chapter 4 provides an analysis of the used sound files and data. Chapter 5 offers a detailed signal model of the reviewed Blue Whale calls. In the next part, are the theoretical approaches are confirmed by practicable methods in which every method is evaluated and its outcomes are described: Chapter 5 compares the methods used and interprets the performance of the applied algorithms. The results are shown in Chapter 6 Future work and a conclusion of the results are provided in chapter 7. Detailed information of the used sound files, MATLAB® files and SCORE Project are included in the Appendix.

Chapter 2

Biological and Oceanographic Background

This chapter provides an introductive background knowledge needed to understand the full problem task. Biologic fundamentals of the Blue Whale are explained in section 2.1. The basics of underwater sound propagation and typical noise sources are explained in section 2.3. Finally, section 2.4 briefly describes the functionality of acoustic tomography and evaluates a possible application spectrum.

2.1 The Blue Whale

The Blue Whale, believed to be the biggest animal that has ever lived on earth, is a marine mammal belonging to the suborder of Baleen Whales. These giants are thought to live for at least 70-80 years and possibly longer. Global population estimates range from 5000 to 12000 animals, however, a reliable population estimate does not exist.

2.1.1 Physical Description and Size

The Blue Whale outshines all other animals of the world with its size: Male whales are beaten by their female counterparts and animals in the southern hemisphere are larger than northern hemisphere Blue Whales. Its size is up to approximately 30 m and 80 t to 150 t in weight (the longest whale measured by scientists at the NMML - American National Marine Mammal Laboratory was 29.5 m long). Even the tongue of this mammal is equal or superior to the size of an elephant. The Blue Whale has a long tapering body which appears stretched compared to other Baleen Whales. Its head is U-shaped and flat, the mouth is containing approximately 300 baleen plates which are responsible for the denomination of

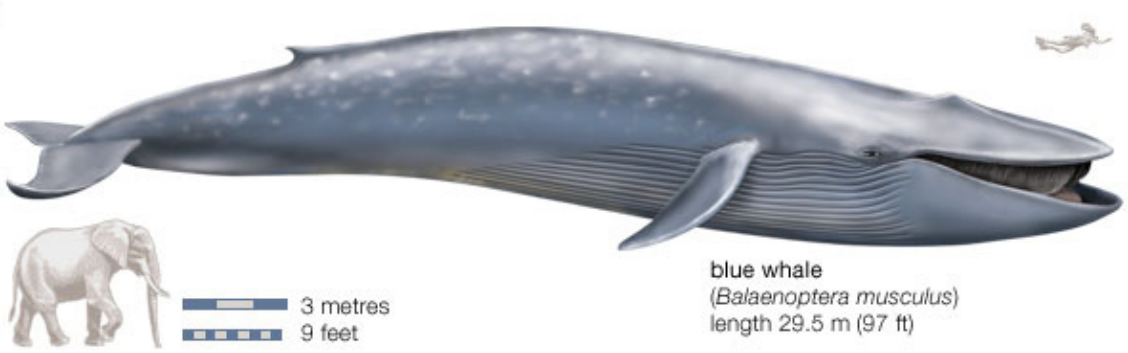


Figure 2.1: Blue Whale compared to elephant and human diver. ©2002, Encyclopedia Britannica Inc [6]

the baleen species. Feeding Blue Whales soak up water and filter it with their baleens so that the krill (marine microorganisms) and crustacea are kept inside the mouth. Their feeding depth depends on where the krill can be found (normally up to 150 m, sometimes up to 200 m) but they appear to eat in lower depths when they are in cold water. The lowest diving-depth has been documented in about 300 m, although their theoretical aerobic dive limit is believed to be in much heavier depths. Blue Whales are estimated to eat about 2 to 4 tons of food per day.
While feeding or traveling, the whales normally cruise at speeds of 2 to 8 km/h, their highest velocity, though, reaches 32 to 36 km/h. The sometimes yellow color of the body arises from little algae which live on the mammal. This color-shift was responsible for the term "Sulphurbottom", the Blue Whale's former name. The Blue Whale manages to dive for as long as 36 minutes. Normally, their diving periods last 5 to 15 minutes after breathing 6 to 20 times at the surface over a period of 1 to 5 minutes. Their lung capacity is 5000 l. When the Blue Whale breathes, a spectacular vertical single column blow is emitted. This reaches sometimes up to 12 m, but more typically 9 m and thus can be seen from a distance of many kilometers. This is a typical visual identification sign of a Blue Whale (see [6]).

2.2 Blue Whale Vocalizations

Sound generation of the Blue Whales is different to Human speech. In particular, Baleen Whales do not possess vocal chords, instead the larynx seems to play a role in sound production. Possibly they also use cranial sinuses to produce sounds. However, this process has not been fully researched yet and remains uncertain. Blue Whales do not have to exhale to produce sound which makes the sound propagation from its origin different to the Human way. The sound pressure level they produce is unique: The Blue Whale is referred to be the loudest animal on earth reaching a source level of 188 dB@1 μPa (the sound pressure level

(SPL) in air is related to 20 μPa, so this equals to a SPL of approximately 162 dB in air) which was described by Cummings and Thompson in 1971 in a paper about a Blue Whale underwater recording project [2]. This maximum level is referred to throughout literature. This sound level allows the biggest mammal of the world to be heard almost all over the world as it enables the sound to travel around almost half the world and still be heard.

2.2.1 Different Blue Whale populations and their vocalizations

Blue Whales inhabit all oceans of the world; similar to birds, they are thought to make seasonal migrations to warmer zones during winter. Hunting them has been very practicable as the Blue Whale represents a large natural resource for producing food or pharmaceuticals. During the first half of the 20^{th} century 325000-360000 animals have been killed in Antarctic waters, which nearly led to their distinction. Since the 1960s the worldwide Blue Whale population seems to have increased slowly. According to the literature, there are seven different Blue Whale populations all over the world. Usually, these populations stay among themselves and even develope their own dialects.

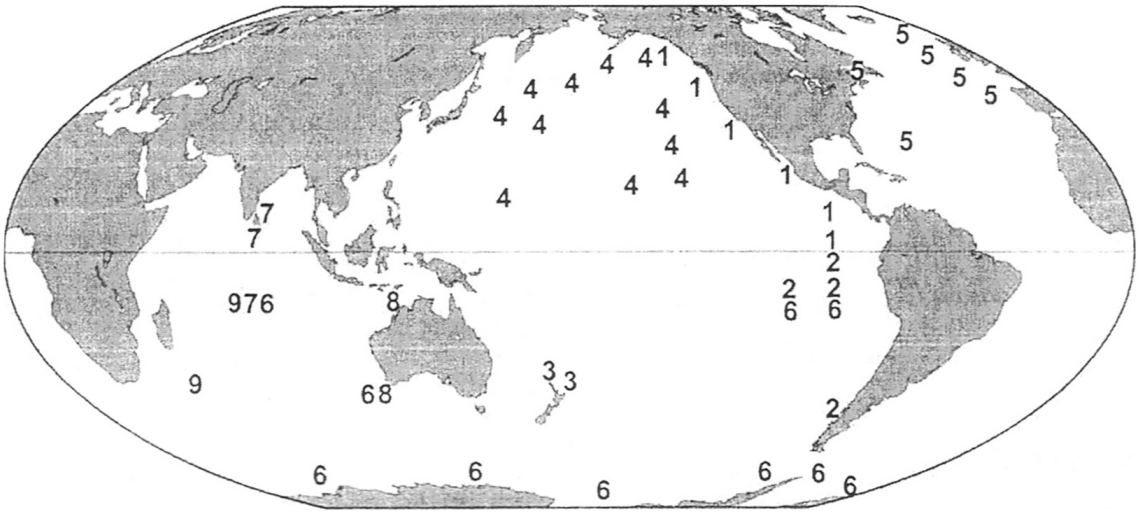


Figure 2.2: Blue Whale populations distinguished by their call behavior (from [11])

All of the populations are able to produce different types of calls, which reside in the low edge of Human frequency perception. The commonly used one is only the D-Call, a single down sweep of a few seconds duration with varying center frequency, which apparently represents a feeding call to display that food is available at the sound-emitting place. The other whale calls form the whale songs which differ from one population to another. They are probably mating songs and are only uttered from the adult male whales as female

whales do not sing.
The Blue Whales off the Californian Coast are the whale population with most underwater data collected, thanks to the geographic position of the NAVY underwater recording stations (see [1]). They are indicated by population number 1 in Figure 2.2). They constitute the healthiest stock of Blue Whales ranging from 1500-3000 animals. Besides the D-Call, these whales produce three well documented types of calls, namely the A, B, and C-Call. The pulsed A-Call contains many non-harmonic components with different time offset. The most frequent type is the B-Call, which consists of a narrow-band down sweep at a fundamental frequency of about 17 Hz and strong harmonic components. The C-Call is a narrow-band noise at a frequency of about 12 Hz, which is sometimes uttered before the A- or B- call. A typical whale song consists of a sequence of whale call patterns. The initiating call is always of an A-type; common phrasing might be for example ABABAB or ABBBABBB.

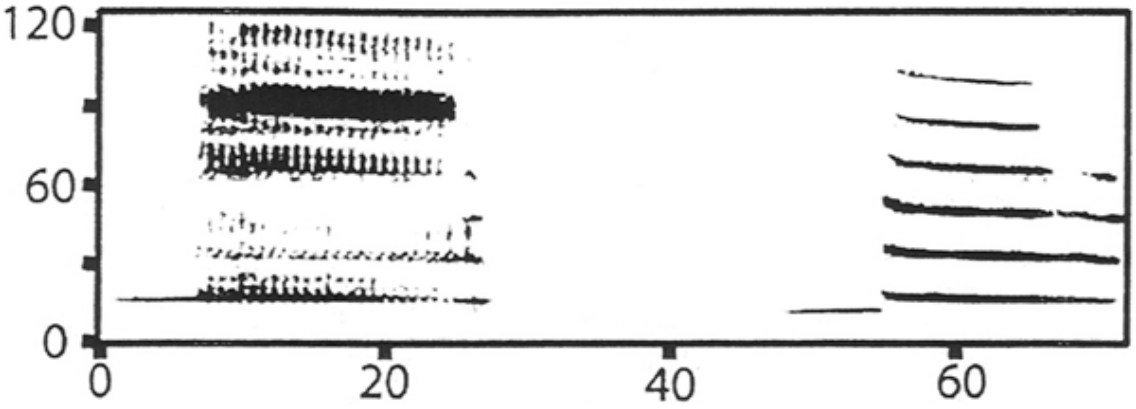


Figure 2.3: Californian Blue Whale song, consisting of the pulsed A-Call and the down sweeping B-Call. The harmonic line at about 12 Hz in front of both calls represents the C-Call (from [19]).

2.2.2 Different Call Types from the North East Pacific Blue Whale

This population's repertoire consists of four types of calls referred to as A-, B-, C-, and D-Call, as described in a study of Blue Whale calls considering gender specific differences [11]. This paper describes A-Calls to have weak tones near 16 Hz combined with pulsed repetitions of a near 90 Hz component. This band is slowly shifted downwards in frequency, varying from whale to whale in its envelope. The amount of the pulsed utterances differs also from only a few to mostly approximately 19 to 23 times with a frequency sweep from 92 to 87 Hz. The average level is calculated to be 178 dB @ 1 μPa for the A-Call. The B-call can be characterized as a long moan at a fundamental frequency of 17 Hz which is shifting downwards in frequency. The call possesses strong harmonics with the most

prominent one around 51 Hz. In [11], the third harmonic was examined to be in average 10.3 dB lower than the basic frequency with a range from 0 to 16 dB lower inm 61 calls. Though, due to the better SNR in this band compared to the fundamental frequency, the effort in recognizing Blue Whale calls is often focused on this frequency band. The overall source level was calculated to be 186 dB @ 1 μPa at a distance of 1 m to the source over the 10 to 100 Hz band in this study, which shows the nearest value to the maximum source level mentioned in [2]. The call duration is at least 10 seconds with sometimes up to 20 seconds. McDonald et al. describe in their paper [11] that the B-Call or a B-Call series is always preceded by an A-Call. A-Calls were never found to be consecutive. The C-Call is a precursor to B-Calls. It lasts several seconds and shows consistency in character. Its envelope is an up-sweep from 10 to 12 Hz. Due to the worse SNR at low frequent content of the recorded data, this precursor is often masked in ocean noise. It does not preceed every B-Call. D-Calls are produced by other populations as well, as Hildebrand describes in [7]. According to McDonald,though, the call is used more frequently among members of the the North-East Pacific Blue Whale population than it might be for others [11]. Considering the different call properties, the B-call has the best repetitive and defined shape, which leads to the usage of this call for acoustic tracking methods.

2.3 Underwater Sound Propagation

Sound in the ocean is sometimes thought to be even more important than sound in the air. Due to the darkness of the deep sea, sound has proved to be an appropriate medium to share information. Many animals such as dolphins use sonar for orientation. The characteristic of sound transmission in the sea is better than visibility in the air as waves can virtually travel round the world and reach the darkest place in the depths of the ocean.

2.3.1 Sound Speed in Water

As a result of the higher density of water compared to the mixture of gases that make up the earth's atmosphere, the speed of sound is much faster in the sea than in air. Knowing the speed of sound is critical to likewise ocean communication and calculating the origin of a sound source. Pressure and temperature have as well impact on sound speed as salinity and composition of the water. In view of the counteractive temperature and pressure effect on sound depending of the depth of the water, different sound speed profiles can be found in different geographic regions.

During the season the speed profile changes. Figure 2.4 shows a maximum sound speed at the dotted line which is at about 1000 m depth. Above this maximum, the sound speed

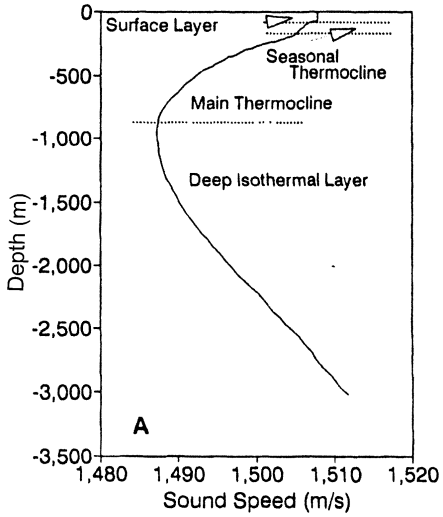


Figure 2.4: Sound speed profile in deep water, Mediterranean profile (from [19]).

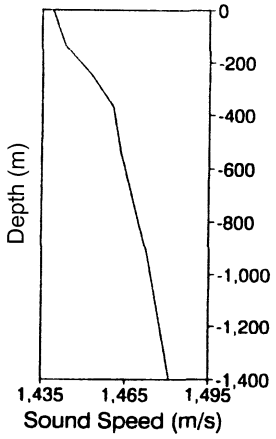


Figure 2.5: Deep water sound speed profile in the arctic sea (from [19]).

rather depends on temperature and below it the pressure effect influences the speed profile. Arctic water shows a greater impact on pressure, for example the maximum of sound speed is near the surface. The higher the water temperature, the deeper the main thermocline is traceable.

2.3.2 Ocean Noise

Ocean noise can be roughly split into three categories:

- Environmental - Physical Noise
- Noise of Ocean Life
- Antropogenic Noise

Environmental noise

Environmental noise has its origin in airborne noise which propagates into the sea, sea state (wind and waves) and noise of seismic activities.
The spectrum of environmental noise has very high peaks in the lower frequency part. Vulcanic eruptions and earthquakes emit high power oszillations in a very low frequency band, sometimes only fractions of Hertz. The noise can mask interesting signals because it has high power emissions. Yet, the endurance of an earthquake or an eruption is much longer lasting and contains a broader spectrum than Blue Whale signals which makes the interesting signals sensible. In general, the environmental noise is responsible for the noise floor. Its power in the ocean is rather high compared to the noise floor in the air. In this process it influences the wanted signals as well but the interesting signal components can be detected at very low SNRs throughout their rather unique frequency shape.

Noise of Ocean Life

Ocean life produces a wide variety of different signals; in the direct neighbourhood of hydrophones even fish produce eating noise. Mammals ar the 'birds of the sea', they vocalize in different frequency bands. Though, the low frequency contained in the Blue Whale B-call is only part of one other whale sound: the fin whale song [28]. These whales produce pulsed down sweeps at about 20 –35 kHz; the durance of the sound is much shorter than the blue whale B-call, whereas the repetition rate is very high. An example of fin whale utterings together with blue whale calls can be seen in Figure 2.6. This shows clearly the different length compared to blue whale sounds 2.2, a fact that is critical for recognization.

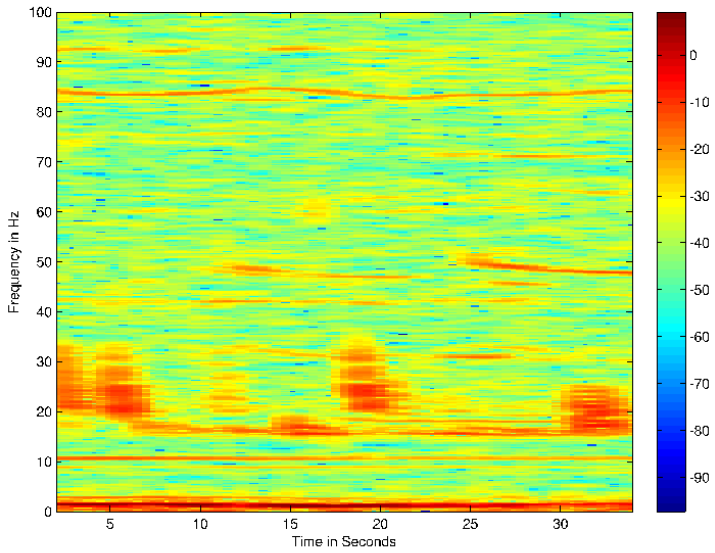


Figure 2.6: Mixed sounds of fin whales and blue whales. The band is the same, though the blue whale B-calls can be clearly seen. Fin whale calls mask some content of the blue whale calls.

Antropogenic Noise

Man made noise represents a strong part of the ocean noise floor. Traffic noise, noise of single ships and industrial activity on shore such as mechanical activity on other kinds, produce ocean noise. For different detection algorithms it is important to limit the time of the expected sound to some maximum, as it might sometimes be the only possibility to distinguish between animal sounds and ship propellers. The latter can produce similar sweeps in a similar frequency band, though the period of time when the signal is present is much longer for the first ones than for a typical whale call. Figure 2.7 provides a view of a spectrogram containing strong ship noise components (for example the flat red line at about 380 Hz).

On the whole, many different influences lead to differences in the ocean noise floor. Explosions evoked by earthquakes, precipation noise from rain and rigid bodies (hail), animal sounds, wind and waves and many more result in a broad spectrum of noise; according to the current situation, the noise floor is always different. In his paper [25], Gordon Wenz drew a conclusion of different noise sources and achieved a formulation of some models for a more simple recognization and calculation of the ocean noise floor. These 'Wenz' curves (see 2.8) are often used as a model for acoustic calculation of sea sound and SNRs.

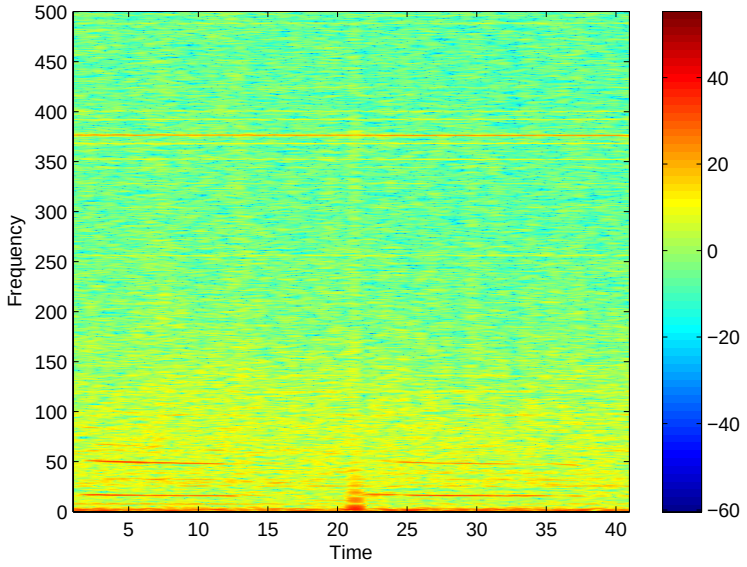


Figure 2.7: Spectrogram of a representative recording with strong ship noise components present.

For personal attempt, signal samples have been taken to average the background noise of the available recordings. In Figure 2.9 it is shown that the shape of the noise floor[1], naturally not of exactly the same shape, seems to be very similar to the shape of the 'Wenz' curves.

[1]dB is related to MATLAB® calculation, the actual gain is different from file to file, see Section 4.1.3.

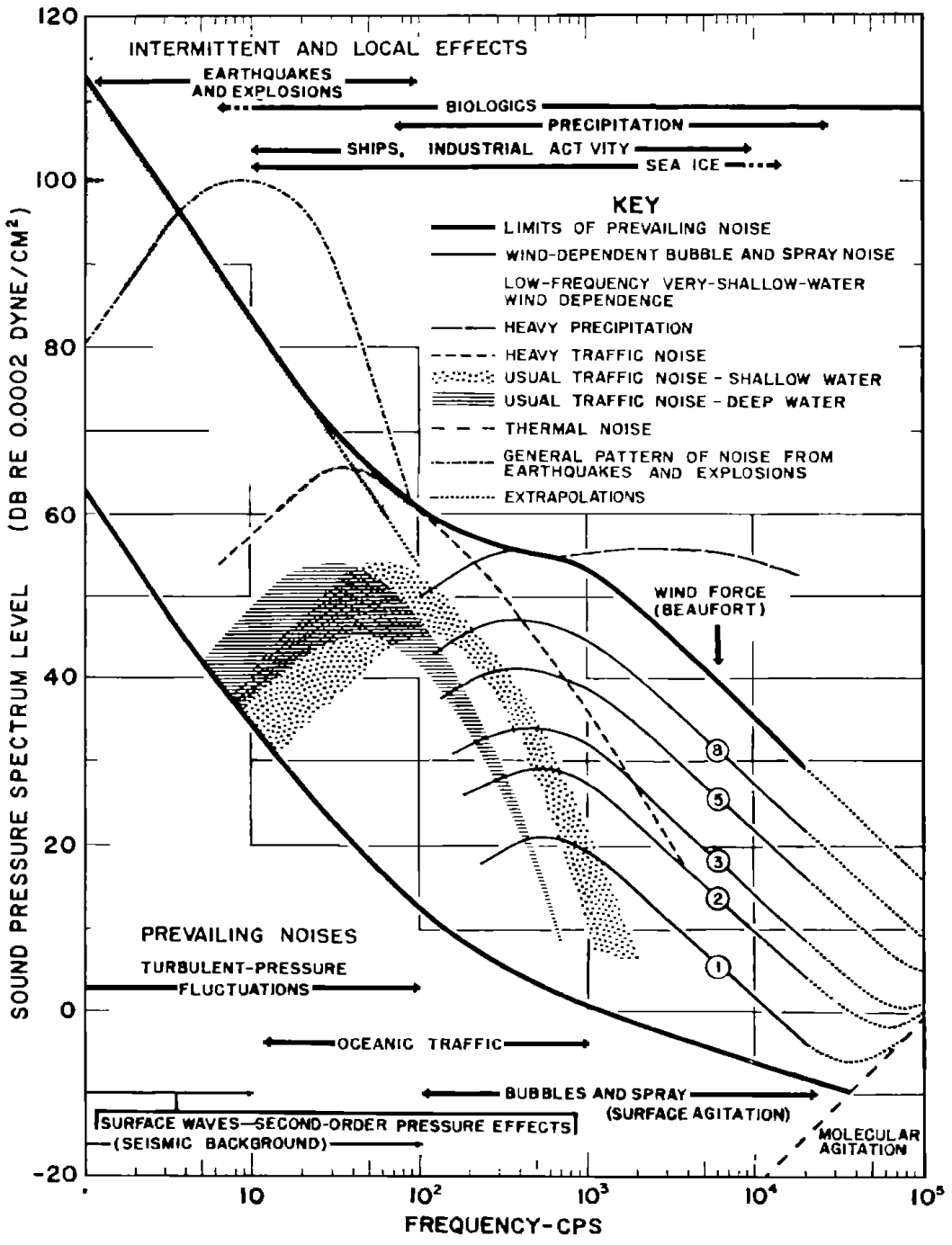


Figure 2.8: Image of different sea-noise types, taken from Wenz' article [25].

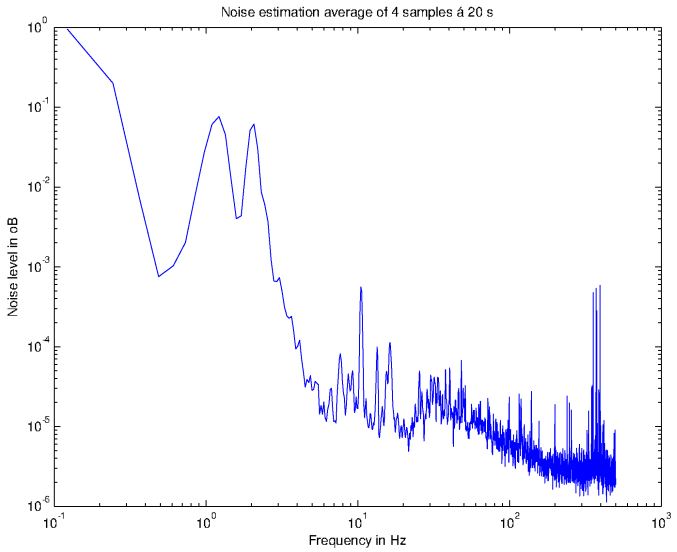


Figure 2.9: Background noise calculation: This image shows an average of several random noise samples from the input data.

2.4 Ocean Acoustic Tomography

Acoustic tomography methods are widely spread in examining different liquids, e.g. oil. As the ocean covers more than two thirds of the earth´s surface, a strong influence on the climate is inevitable. The most critical aspect for air streams and air hybridity is the temperature gradient between the ocean layers on the surface as the resulting eddies and streams affect the weather. Traditional temperature measurements result in high resolution both in time and in temperature dimension. Yet, the measurements are bound to a certain area which is rather small, as averaged punctual measurements cannot guarantee a true average of the enclosed temperature layers. Sound measurements offer a good alternative to get a high distance profile. The method is simple: High preasure and high temperature result in high sound speed (see sound speed profiles in Section 2.3.1). The average water preasure is influenced by the water depth and the persistent time differences for different sound beams (see Section 2.3.1) are related to the temperature gradients. When the transmitted signal is well known in terms of phase and amplitude, the signal response can help calculating the average temperature profile (if the ocean geography is known). A picture, demonstrating how the sound beams look like according to a medium temperature profile between two points, can be seen in Figure 2.10.

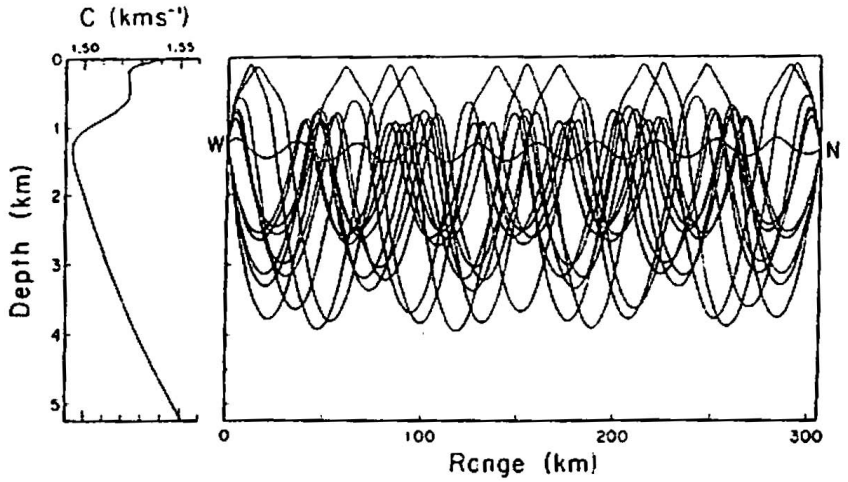


Figure 2.10: Way of sound beams in relation to the temperature profile, picture taken from [19].

A combined sender/receiver can help to cancel effects due to water movement, as both units render the same measurements at the same time but in opposite direction. This also helps to avoid frequency non-linearities caused by the Doppler effect (section 2.3.1). Yet, not every part of the signal reaches the receiver. Beams which do so are called 'Eigen beams'. The fastest incoming waves are the ones which have sometimes traveled the farest way, as the sound speed increases with increasing water depth (see profiles, Section 2.3.1). The later echos which are arriving after a few peaks cannot be calulated that easily by just taking the ray path and the Eigen-model into account. Additionally, positive interference of the later arriving sound results in higher sound intensity. Sound of high indensity arriving after a certain period of time is called cutoff. The sound duct is crucial for the calculation of the temperature profiles, as the shift of the duct caused by the temperature changes the propagation time of the Eigen-beams. Eigen-beams of a lower order (designating the beams travelling near the surface) are used to evaluate temperature zones near to the surface. Eigen-beams traveling through the deep sea provide information about the deeper ocean temperature. An example shows how acoustic ocean tomography works: A model of the ocean geography is used for calculating different sound beams (Figure 2.11. The expected ways of the sound beams are calculated to get an impression of the sound delays. The model compared to the actual recording between two points can be seen in Figure 2.12.
This shows that time resolution is crucial for making ocean acoustic tomography work. The shape of the signal must have a defined starting point both in frequency as in time[12]. Analysis of the time and frequency resolution of the evaluated algorithms will show, if ocean acoustic tomography could be possible with B-calls.

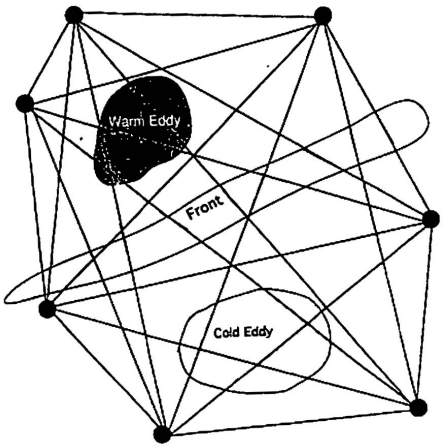


Figure 2.11: Ocean model with calculated sound beams [12].

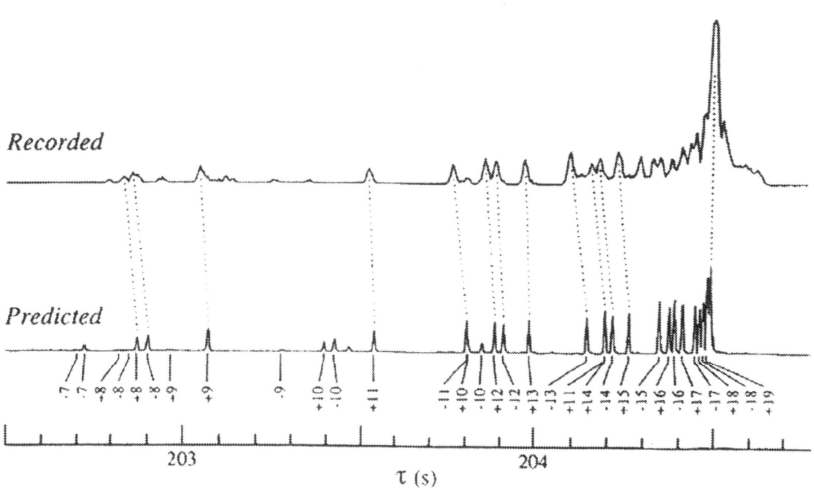


Figure 2.12: Actual recorded sound compared to the predicted model [12].

Chapter 3

Theoretical Backgound of the Evaluated Methods

In this chapter, the theoretical background for the latter calculations follows the formulated approaches towards solving the given problem. The methods are introduced and their mathematical explanations are presented. In section 3.8 a speech communication detector method provides an approach of how to evaluate the algorithms.

3.1 Solutions - An Approach

The Blue Whale signal is very low frequent and of a narrow bandwidth. Different approaches towards detecting a call effectively are now evaluated. The methods implemented and evaluated throughout this thesis are listed as:

- The *spectrogram correlation* method, as already implemented in whale detection is supposed to serve as a standard implementation to the other methods.
- A *maximum likelihood estimation* is as well applied on site in order to gain information about the Blue Whale call. The process includes modeling a filter bank consisting of different sweeps which vary in mean frequency, sweep rate and signal length.
- Detect the Blue Whale calls through a method deriving from Human speech recognition - the *cepstrum* and the Δ-*cepstrum analysis*. Throughout the shape of the different Δ-cepstrum vectors, the variation in the harmonics should be measured to cut the relation between the absolute frequency and the whale detection.
- As a filter bank is very time consuming and storage intensive, a simple band detection different frequency bands should evaluate the possibility to detect a Blue Whale call.

- Even the power detection time cost can be decreased by using DFTs implemented as *Goertzel filters*. The necessary bandwidth and the overall functionality will be analyzed.

3.2 Signal Processing Basic Principles

Some basic signal processing principles from [16] give some fundamental knowledge in the used signal representations: Given an analog signal $x_a(t)$ with the continuous time t, the time-discrete representation of this signal is $x[n]$. The relation between the continuous and the discrete signal is given by

$$x[n] = x_a(nT) \tag{3.1}$$

T is the sampling period. This means, the numerical value of the nth value of $x[n]$ equals the value of the analog signal at the time nT. The time dependent Fourier Transform of an input signal $x[n]$ is given as

$$X[n, \lambda) = \sum_{m=-\infty}^{\infty} x[n+m]w[m]\mathrm{e}^{-j\lambda m} \tag{3.2}$$

with the analog frequency λ and the windowing function $w[m]$. For plotting a *spectrogram*, the frequency of the time dependent Fourier Transform is plot on the y-axis, the time on the x-axis. The color corresponds to the signal amplitude with high values painted red, low values painted blue.

3.3 Spectrogram Correlation

Out in nature, animal sounds must be recognized against background noise of other animals. This results in a perceptual and producing focus on certain features. One of those is the concentration on frequency sweeps. Uttering frequency modulated (FM) signals helps to provide a unique signal against background noise. Spectrogram correlation is a tool frequently used to detect animal sounds working FM and the state of the art tool for Blue Whale detection [13].

Given a spectrogram of a time-series of a sound signal, the method correlates linear sweeps of a certain bandwidth in the spectral domain at defined start and ending frequencies. The correlation kernel for a signal segment can be described by the following formulas

$$s(t, f) = f - (f_0 + \frac{t}{d}(f_1 - f_0)) \tag{3.3}$$

$$k(t, f) = (1 - \frac{s^2}{\sigma^2})\mathrm{e}^{(-\frac{s^2}{2\sigma^2})} \tag{3.4}$$

where $s(t, f)$ represents the distance of this point from the central axis of the segment at time t, σ is the instantaneous bandwidth of the segment at time t, k stands for the resulting correlation kernel value k at a given time and frequency point, d signifies the duration of the sweep segment, using the same unit as time t. f_0 represents the start frequency of the segment, f_1 the end frequency of the segment. The resulting correlation kernel is produced by concatenating all kernel segments for a sweep.

The output of the one-dimensional cross-correlation process for each frame of the spectrogram is given by α, which is calculated by correlating the concatenated kernel segments and the spectrogram of the input signal, which is specified by X.

$$\alpha(t) = \sum_{t_0} \sum_{f} k(t_0, f) X[n - t_0, f] \tag{3.5}$$

A model of the correlation kernel is provided in figure 5.3.

Experiments have shown [13] that this method provides a convenient recognition of whale sounds, in this example bowhead whale songs, compared to matched filter methods, which do not work that well in an environment with harmonic interferences 3.4.

3.4 Matched Filter Design

Basically, the more precisely one knows what to look for, the easier it becomes to find the desired matter. This is also true for signals, especially if they are hidden in the background of white noise. Hence, to catch the desired, the use of filters that sort out the noise and only keep the matching signal seems to be obvious. If the noise does not correlate with the signal, the cross correlation between noisy signal and raw signal will produce the highest output. Naturally, the signal itself will rather be replaced with a model or a template for calculation time and cost reasons. The mathematical derivation of a matched filter starts by looking for a filter h[n] which maximizes the signal to noise ratio (SNR) output of the following convolution

$$y[n] = \sum_{k=-\infty}^{\infty} h[k]x[n - k] \tag{3.6}$$

where $x[n]$ represents the input signal, consisting of the sound of interest ($s[n]$) and assumed additive noise ($v[n]$).

$$x[n] = s[n] + v[n] \tag{3.7}$$

Calculation the SNR needs the introduction of the Expectation of y_s (outcome of the signal of interest through the filter system) and y_v (outcome of the input signal's noise part through the filter system):

$$SNR = \frac{|y_s|^2}{\mathrm{E}\{|y_v{}^2|\}} \tag{3.8}$$

Following [8] leads to a result of the optimal normalized impulse response of

$$h = \frac{1}{\sqrt{s^H R_v^{-1} s}} R_v^{-1} s \tag{3.9}$$

$$R_v = \mathrm{E}\{vv^H\}, \tag{3.10}$$

where R_v represents the covariance matrix of the noise v, v^H represents the conjugate transpose of the noise. In case of v being Gaussian noise, R_v equals the unit matrix which results in $h[n]$ describing the time reversed and complex conjugate version of $x[n]$. If the noise is not white, the knowledge of its covariation matrix is necessary to produce an exact output. Thus, the worse the noise can be described, the worse a matched filter function works. Another disadvantage of matched filtering becomes visible if a time variance between signal and model arises. The more specific the model, the worse the correlation. A typical application area of matched filtering is telecommunications, where discrete signals after being transferred over a transmission path are regained by a matched filter. The outcome of the filter does not describe the signal itself but its autocorrelation, which leads to the introduction of a threshold: Beneath this critical value, the filtering system decides to reject a recognition. For a more detailed explanation, the books [8] and [16] and can provide further information.

3.5 Maximum Likelihood Estimation - MLE

Natural processes always need one or another statistical method to be described in terms of a mathematical model. Maximum likelihood estimators attempt to fit a mathematical model to a set of data. Assuming that the data follow a normal distribution, the MLE method provides a unique solution.
The estimation process needs a defined likelihood function and estimates that its maximum output is most probable to indicate best modal parameters at a certain point.
Different parameters increase the dimension of the MLE. All variations of different estimators are used to calculate the probability function of the corresponding estimator being the most similar to the input signal. The maximum of the probability function proclaims the corresponding estimator as it is shown in the appropriate model description. Mathematical formulae are taken from [8] and [16].

$$\hat{\theta} = argmax L(\theta) \tag{3.11}$$

$$L(\theta) = f_\theta(x_1 ... x_n) \tag{3.12}$$

In equation 3.11 and 3.12, $L(\theta)$ means the likelihood function of n realizations $x_1...x_n$. θ is a (probably multidimensional) parameter. A variation of θ leads to the likelihood function $L(\theta)$, which is maximized to obtain the fitting parameter(s) θ.
In the specific application for modeling a whale song, the sweep parameters mean frequency and sweep rate have been taken as estimator parameters. As a probability function for the estimator fitting best to the signal, the cross-correlation of the signal samples and the model sweep with the two corresponding parameters out of a matrix of sweeps with different mean frequencies and sweep rates are calculated, the result equals a In this case, the cross-correlation the input signal and the estimator with given sweep rate and mean frequency - (see equation 3.13) serves as a probability function to designate the corresponding signal model.

$$y[n] = \sum_{k=-\infty}^{\infty} s[k]x[n+k] \tag{3.13}$$

$y[n]$ is the output of the cross-correlation of the two signals $s[k]$ and x$[n+k]$, with the discrete time points n and k. The higher the correlation value of $y[n]$, the better the estimator fits to the input signal in terms of sweep rate and mean frequency. Compared to equation 3.4, the summation uses $x[n+k]$ instead of $x[n-k]$, which leads to the cross-correlation-function being a convolution with a time-inverse signal.

3.6 Bandpass Based Detection Algorithms

A simple way to realize a signal detection is to examine in which frequency bands the signal might be present. The power in that frequency range where signal components are present, will be higher than in the same frequency range, if the sample is only containing noise. The signal model in 5.2.1 yields to interesting ranges containing possible signal components.

The partitioning of the recorded sample is achieved by using bandfilters. If there is no signal present, bandfiltered frequency ranges *signal low*, *signal mid* and *signal high* (Figure 3.1), contain noise components only. The sum of the signal power divided by the sum of the noise power, represents a characteristic value for the actual SNR. The higher this value, the higher the probability of a Blue whale B-call presence in the recorded sample.

3.6.1 Power Spectral Density Based Detection

The power of a discrete signal over the window length can be calculated as follows:

$$P(x) = \frac{1}{N}\sum_{n=0}^{N-1} x^2[n], \tag{3.14}$$

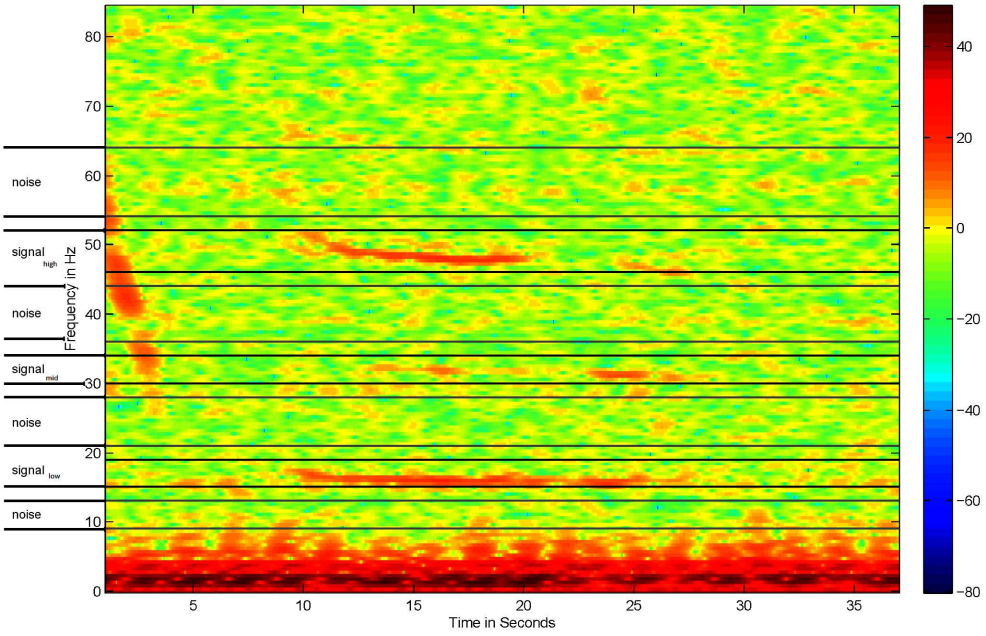


Figure 3.1: Interesting frequency ranges containing signal and noise.

where $x[n]$ signifies the signal at time n, and N is the window length, and $P(x)$ the overall power of signal x during the period of N. The same equation is taken for calculating the power of a noise band.

For a better comparison of different bandwidths, the power was normalized by the bandwidth of the filter to generate the power spectral density for each band as a result. A model of this algorithm can be seen in figure 3.2

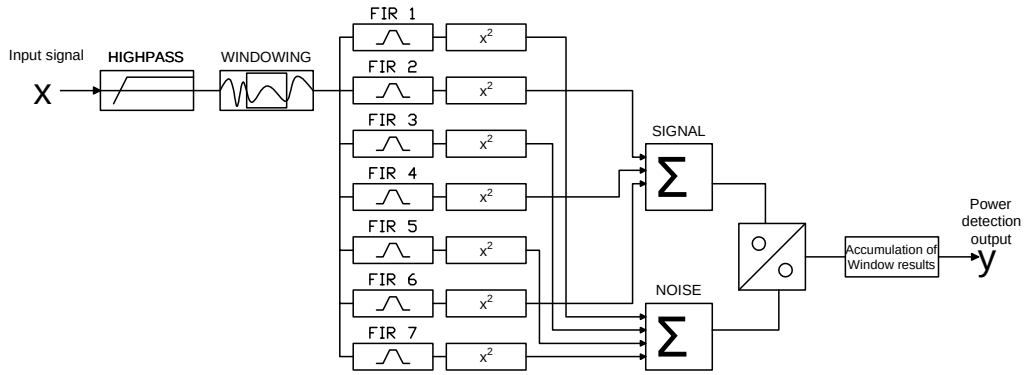


Figure 3.2: Flow Diagram of the power-detection method.

3.6.2 The Goertzel Algorithm

First published by Dr. Gerald Goertzel in 1958, the Goertzel algorithm is a technique in digital signal processing and used for identifying single frequency components of a signal. When utilizing a fast method to detect frequency components in a signal, the first attempt is to calculate the amplitude over the FFT. Yet, this yields as a result of an estimation of the overall spectrum, which is not necessary in some cases. The Goertzel algorithm presents a way to reduce the calculation time and increase the efficiency for such needs. It takes a discrete time input sequence x[n] and calculates the output sequence y[n] given as

$$y[n] = x[n] + 2 \cdot cos\frac{2\omega\pi k}{N} \cdot y[n-1] - y[n-2], \tag{3.15}$$

where $y[-1] = y[-2] = 0$ and ω is the frequency of interest normalized with respect to the sampling frequency. The frequency component at frequency k output is calculated by

$$X[k] = y_k[n] = y[n] - W_n^k \cdot s[n-1] \tag{3.16}$$

with recursive calculation of the weighting function W_n^k- Effectively, this produces an IIR filter of order 2, which is shown by the flow diagram for the filter realization in Figure 3.3. The Goertzel algorithm is advantageous compared to the FFT when

$$M < \frac{5}{6} \cdot log_2(N), \tag{3.17}$$

with the DFT length N and the number of desired pins M. More detailed information can be found in the article [21], and in book [18], where the formulae in this section have been taken from.

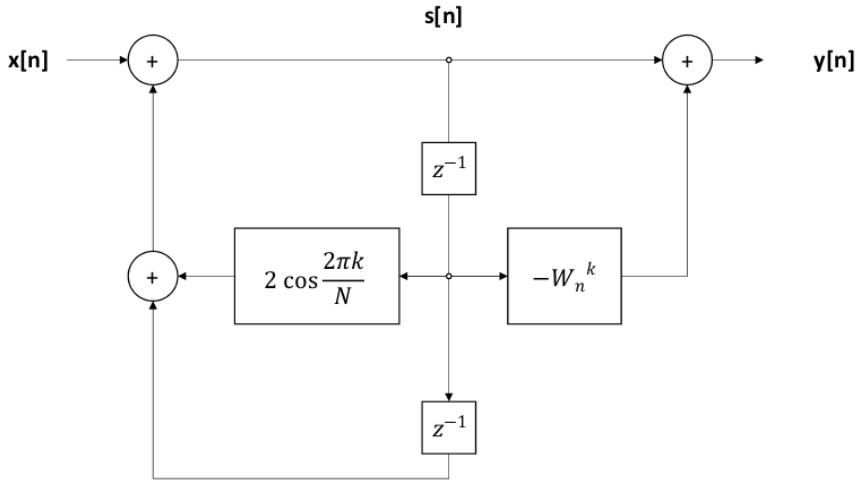


Figure 3.3: Filter realization of a second order Goertzel algorithm.

3.7 Cepstral Analysis

The *cepstrum* of a signal is a homomorphic transformation, meaning that the structural connections are kept throughout the transformation [16]. Analytically, it is accomplished by an inverse Fourier transform of a Fourier transformed signal after taking the logarithm.

$$c_x = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log \left| Y(e^{j\omega}) \right| e^{j\omega n} d\omega \tag{3.18}$$

c_x is called the *complex cepstrum*, which takes the Fourier transform $Y(e^{j\omega})$ at the frequency ω of a signal $y[n]$, and calculates the logarithm. The inverse Fourier transform brings the signal into cepstral domain.

This function separates components of the signal which have been linked by convolution in time domain, $x[n]$ and $h[n]$, since the transformation of two convoluted signals produces two multiplicatively associated functions in frequency domain (Equation 3.7) and the logarithm of two factors equals the sum of their logarithms.

$$\begin{aligned} y[n] &= x[n] \otimes h[n] \\ Y_k &= X_k \cdot H_k \end{aligned} \tag{3.19}$$

$y[n]$ represent the output signal of the convolution in discrete time $[n]$ domain, Y_k, X_k, H_k represent signals in discrete frequency domain $[k]$. The logarithm transforms a multiplication

into an addition and splits the signals X_k and H_k. Due to the logarithmic separation,the harmonic components of the signal should become visual, even if their amplitude is quite low. Yet, as for the power cepstrum (3.20) only the magnitude information of the signal is used, the phase information gets lost which causes an information loss. The signal cannot be completely restored if this is desired but the complex spectrum can be calculated which includes the phase as well in the phase cepstrum part. Yet, the complexity of the algorithm increases but does not justify the benefit of the original phase values in speech applications. The computation of the power cepstrum is done by using the following formula (equation 3.20)[17]:

$$c_x[n] = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log \left| X(e^{j\omega}) \right| e^{j\omega n} d\omega \tag{3.20}$$

$c_x[n]$ represents the real cepstrum of a signal at the discrete time point n, $X(e^{j\omega}$ the Fourier transform of this signal at the frequency ω. The word 'cepstrum' is an anagram of 'spectrum'. In order to avoid confusion between the units of a spectrum of a signal and its cepstrum, the 'quefrency' is analogue to the spectrum's frequency, the 'gamnitude' to its amplitude. Today only the former ones have been accepted. Originally, the cepstrum was invented heuristically to alleviate the finding of echo arrival times, for example in sonar measurements. More detailed information can be found in [5], [16] and the MathWorks® website [10].

3.7.1 The Delta Cepstrum

The Delta(Δ)-cepstrum uses the differentiation of the cepstrum to reject common features. Only changes between the vectors stay and unique events are visible, whereas continuous movement is rejected (equation 3.21, [17]):

$$\Delta c_x[\tau] == \frac{\sum_{l=-L}^{L} l c_x[\tau + l]}{\sum_{l=-L}^{L} l^2} \tag{3.21}$$

L is mostly used from the value of 1 to 3 and signifies the time window of which the derivation is approximated. Δ-cepstrum. $\Delta c_x[\tau]$ finally signifies the first derivative of the cepstral coefficient c_x at the discrete point in time τ, $c_x[\tau + l]$ is the $\tau + l$th cepstral coefficient of a signal at time τ. For $L = 1$ $\Delta c_x[\tau]$ from 3.21 can be written as

$$\Delta c_x[\tau] = c_x(\tau + \delta\tau) - c_x(\tau - \delta\tau), \tag{3.22}$$

where $\delta\tau$ represents a discrete time difference, as also shown in figure 3.4.

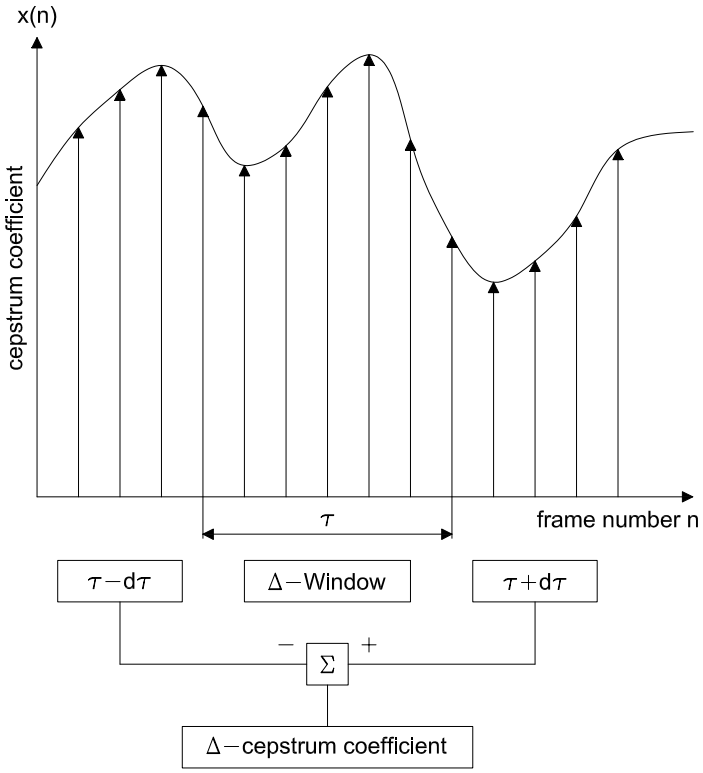


Figure 3.4: Delta cepstrum pinciple.

In cepstrum domain a cepstral coefficient means periodicity of frequency components. If this change is continuous (like in frequency modulation, FM), it will be canceled by differentiation. The idea of using Δ-cepstrum origins in detecting small changes which occur over a certain period of time, e.g. the FM-whale sounds provide an example of. More theoretical background is offered in [17] Though, the low frequent whale sounds are rather difficult sample material, as the resolution is very low. The analysis part explains this issue in detail (see Section 5.4.2).

3.8 False Alarm Rate - Missed Detection Rate

Testing different algorithms means having them to meet common requirements. In case of the evaluated detection algorithms, the critical value is met by overcoming a threshold level, sometimes for a certain span of time.

Reached demands have to be evaluated and for that case two typical errors, as explained in [3] provide a performance rating: False Alarm Rate (FAR - see equation 3.23) and Missed

Detection Rate (MDR - see equation 3.24).

$$\text{FAR} = \frac{\text{number of false detections}}{\text{number of whale calls} + \text{number of false detections}} \cdot 100\% \tag{3.23}$$

$$\text{MDR} = \frac{\text{number of missed calls}}{\text{number of whale calls}} \cdot 100\% \tag{3.24}$$

In evaluating the algorithms, the optimal threshold should be reached, when FAR and MDR are equal. The diagrams of MDR versus FAR show the optimization process using different thresholds. For measuring performance, the distance to the origin is used for evaluation. In that case for threshold optimization of the algorithms, two aspects have to be taken into account:

- The smaller the distance between MDR/FAR ratio to the origin, the better the algorithm should be able to decide between call or not call. This means that MDR and FAR should tend to zero.
- The closer the ratio between MDR and FAR is to 1, the better the overall performance of the threshold should be: The amount of missed and false detections should be balanced.

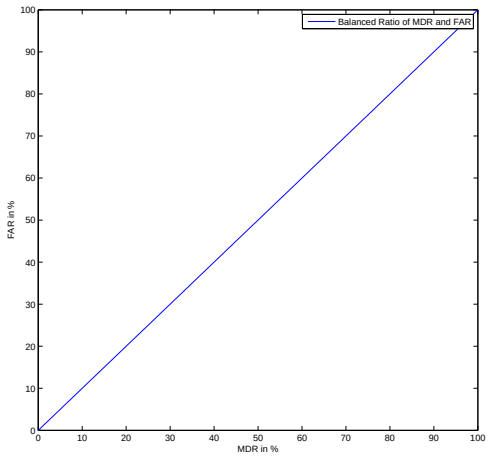


Figure 3.5: Threshold plot with MDR versus FAR. The distance from the origin shows the performance of the specified algorithm (the nearer to origin, the lower the error rate). If the point of MDR vs.FAR is on the diagonal line, the threshold is balanced.

Chapter 4

Characteristics and Modelling of Blue Whale Calls

In this chapter, the characteristics of the Blue Whale B-call as taken from the data and evaluated through literature [9], [11] are evaluated.

4.1 Overall Data Characteristics

This section provides a specification of the used data. The signals were recorded with the Acoustic Recording Packages (AppendixA) from the year 2000 to 2003 with different sample rates (1000 Hz and 500 Hz). 10 signals from SIO from the year 2000, 9 from 2003 which were pre-cut and selected by SIO were used together with data cut from the original ARP-streams, where some signals contain noise only, some noise and several whale calls. The calls are numerally described in Appendix D.

4.1.1 Sampling Rate

The sampling rate f_s of the available data is 500 Hz for data from the year 2003 data and $f_s = 1000$ Hz in the older samples. As the interesting frequencies are in both cases much lower than the particular Nyquist frequency, for some methods the sampling rate has been decimated to 200 Hz. This has been achieved by applying a FIR-low pass-filter with a cut-off frequency f_c of 90 Hz on the data, and afterwards taking the MATLAB® function *downsample* to take take only the samples needed.

4.1.2 Bitrate

The sound signals have a resolution of 16 Bit. Some evaluated recordings are normalized, as the pre-cut files from SIO from the year 2001 and 2003. Throughout the original ARP-data, the gain factors differ from recording to recording. Both facts cause difficulties: If the detection algorithm's threshold is data-dependent and not a ratio, the gain change within the files causes a different correlation result (see section 4.1.3). Figure 4.1 shows an original wave-file with high amplitude low frequency content. As this low frequency noise in the normalized wave-file is several times higher than the signal content, the amplitude resolution of the interesting signal frequencies is sometimes shortened to only some bit.

4.1.3 Signal Power

The mean power of the used whale samples differs to a great extent among the distinct files. The wave files which have been used as test samples in the years 2000 and 2003 are not normalized and sometimes show a highest amplitude value of 10^{-3} times of the maximum possible wave output. The samples include high amplitude low frequent oscillations with less than 1 Hz, which causes problems with signal resolution and signal blasts in the normalized wave-files (see Figure 4.1). They seem to have a common gain factor, at first sight, but a number of power characteristics show clearly that the gain must have been adjusted manually (see the overview of the used samples in Appendix D). For detailed comparison this causes a problem, as the correlation value depends strongly on the signal power. However, the gain factor is unknown and changes throughout the recording situation, absolutely reliable numeral comparisons between the performance, in many different years as well as in one year are hard to realize. The detection might be possible, as the envelope of an algorithm outcome detecting a whale call is similar to one call and another, but the demands of an overall detection threshold cannot be agreed upon data dependent correlation algorithms where no ratio is calculated.

4.2 Characterization of Blue Whale Calls

Providing a detailed signal description fitting to an average Blue Whale B-Call has been a recursive procedure while developing the different algorithms. The signal's features have not entirely been described in literature and additionally, some parameters like durability and mean frequency shift over the years (see Section 4.2.1).

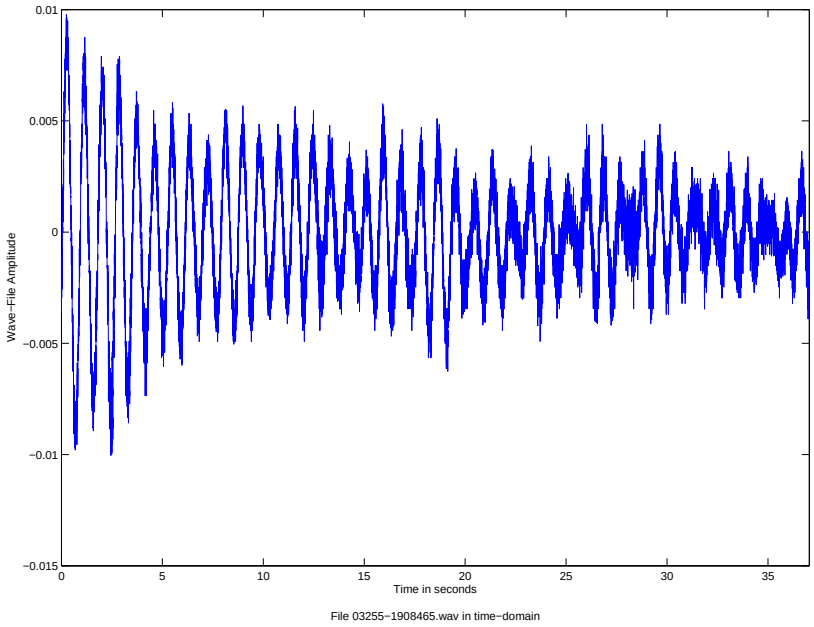


Figure 4.1: Original data wave-file with low frequency content. The amplitude of the low frequency noise is several times higher than the amplitude of the signal content.

4.2.1 Frequency Shift

Blue Whales seem to synchronize the pitch in their calls. Over the last years the calls decreased in mean-frequency, which led to a frequency shift compared to calls a few years ago. In the sixties the mean frequency of the lowest spectral line was about 17 to 18 Hz and in the last years, e.g. the year 2001, the mean frequency was found to be 16.02 Hz [24]. This decreasing pitch leads to detection problems with high-accurate methods, which take a certain mean frequency, as for example the spectrogram correlation method. More details to the frequency shift can be read in [24].

4.3 Extraction of Blue Whale calls

Due to the frequency shift, the developed algorithms have been used to provide a signal description by evaluating the correlation of estimated model filters and true Blue Whale songs. The first idea was to take the whole down sweep as an inverse square-root-function. Though, this was not practicable, as the time variation was too high and the shape of the songs are always different, even from call to call of the same animal. To avoid time problems, the signal has been split into windows, assuming that every window contains a sweep with a specific mean frequency and sweep rate. The window length was set to be

constant over a sweep to minimize calculation time cost. Different window lengths were chosen, as the low frequency signal band provides more inaccuracies in windows that are too short. A reference Blue Whale signal is taken to design a first approach of a signal description. The spectrum analysis of a sample signal containing a Blue Whale B-Call is shown as follows: Figure 4.2 shows the signal approximately beginning at about 9 seconds

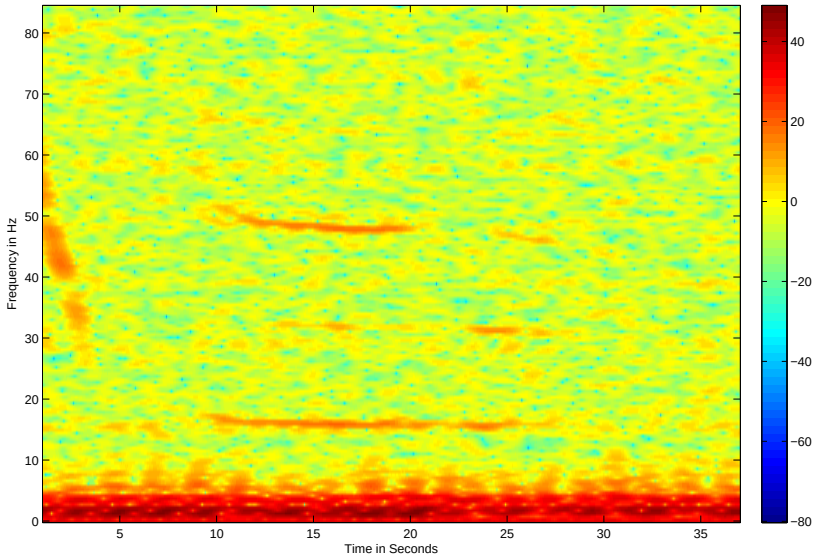


Figure 4.2: Spectrogram of a signal containing Blue Whale B-Call.

after the beginning of the file and lasting for approximately 15 seconds.
To build an appropriate model, the signal has been windowed with different window lengths to examine the best duration. The window lengths were long enough to exactly recognize the low frequent signal contents and short enough to allow the sweep rate to be assumed to be constant. In average, the best results when correlating the different model signals with the originals, to find the right mean frequency and sweep rate were reached with a window length of 3.25 seconds (at a sample rate of 1000 Hz).
In order to guarantee consistency within the samples in different years, the window length was chosen to be 3.25 seconds throughout every sample rate. Though, the time resolution in the newer signals with a sample rate of 500 Hz would not need the same time resolution and could provide a more exact picture of the lower frequencies, which was justified by the argumentation that the lower frequencies of the samples can by neglected due to no signal content.
The signal draft has been calculated by using the output of the later explained maximum likelihood method (Section 5.2). It lasts from 7.5 seconds of the original signal until 22

seconds (shortened spectrogram view because of spectrogram window length).

As a reference, the SIO spectrogram-correlation kernel is depicted below (4.4).

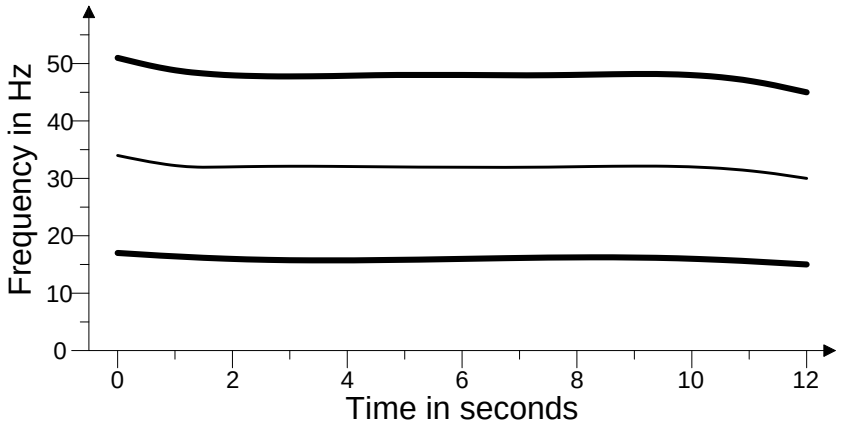


Figure 4.3: Draft of a Blue Whale B-Call with sweeprate and mean frequencies taken from the maximum likelihood detection algorithm.

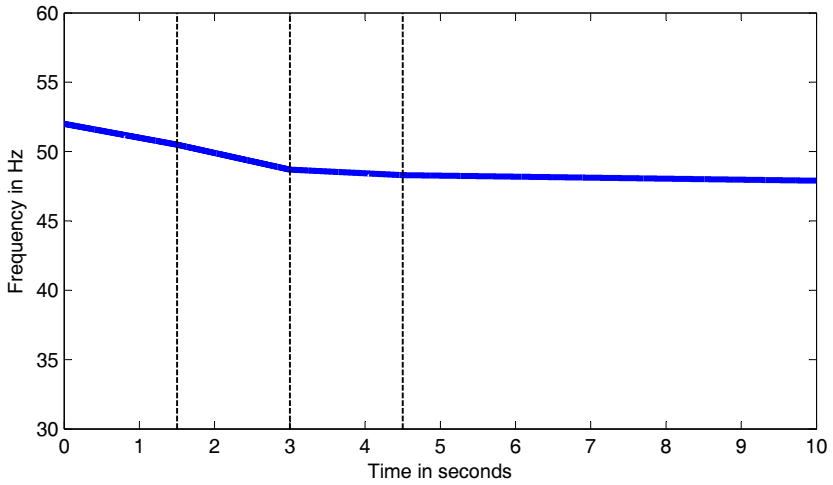


Figure 4.4: Ishmael spectrogram correlation kernel consisting of four segments. This kernel is taken in the applied spectrogram correlation method as a correlator for the input signal.

Chapter 5

Implemented Methods and Performance Analysis

After the theoretical description of the used processing techniques, the implementations of the different algorithms are explained. The spectrogram correlation serves as standard method is presented first (Section 5.1). A maximum likelihood estimation method is evluated in Section 5.2. Bandbass detection algorithms using power spectral density detection and Goertzel filters (Sections 5.3.1) show a similar attempt with different filtering methods. Δ-Cepstrum calculation results are presented in section 5.4. The implementation of the current method of choice used fore Blue Whale B-Call detection, spectrogram correlation, concludes the methods part (Section 5.1). For every method, pictures of the same test-call are provided.

5.1 Spectrogram Correlation

For spectrogram correlation, the algorithm was implemented using formulae of Mellinger's paper [13]. This method serves as a standard method at the SIO for detection of Blue Whale calls. The outcome was compared with the performance of Mellinger's program *Ishmael* (see [23]), which is used in SIO, San Diego, to detect calls automatically. The settings were the same in the MATLAB® file and in the Ishmael program (see table 5.1). In figure 5.2 Ishmaels output of the test call is shown. The re-programmed algorithm output is depicted in figure 5.1. Spectrogram correlation with other test calls has shown that the first 2 seconds differ from the Ishmael output, due to probable windowing and filtering differences in the established program and the method programmed for this thesis, otherwise the contour is similar. The Ishmael program does not show its exact intern arrangement and the MATLAB® program was only related to the mentioned paper's formulae.

Start Time in s	**End Time** in s	**Start Frequency** in Hz	**End Frequency** in Hz
0	1.5	52	50.5
1.5	3	50.5	48.7
3	4.5	48.7	48.3
4.5	10	48.3	47.9

Table 5.1: Ishmael Spectrogram Correlation Settings, contour width = 1.1 Hz (as described in section 3.3).

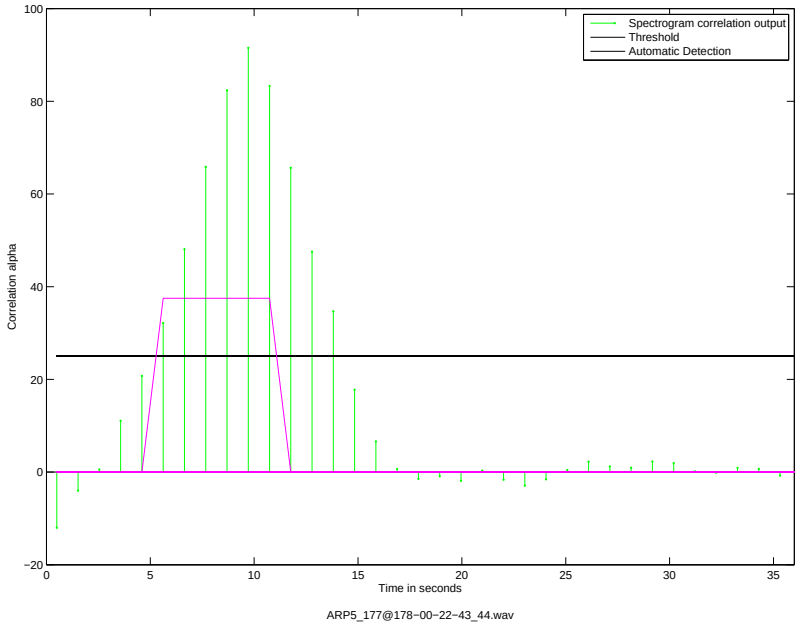


Figure 5.1: MATLAB® spectrogram correlation output with a spectrogram of the test call below.

Figure 5.3 is drawn to depict the correlation kernel values from table 5.1.

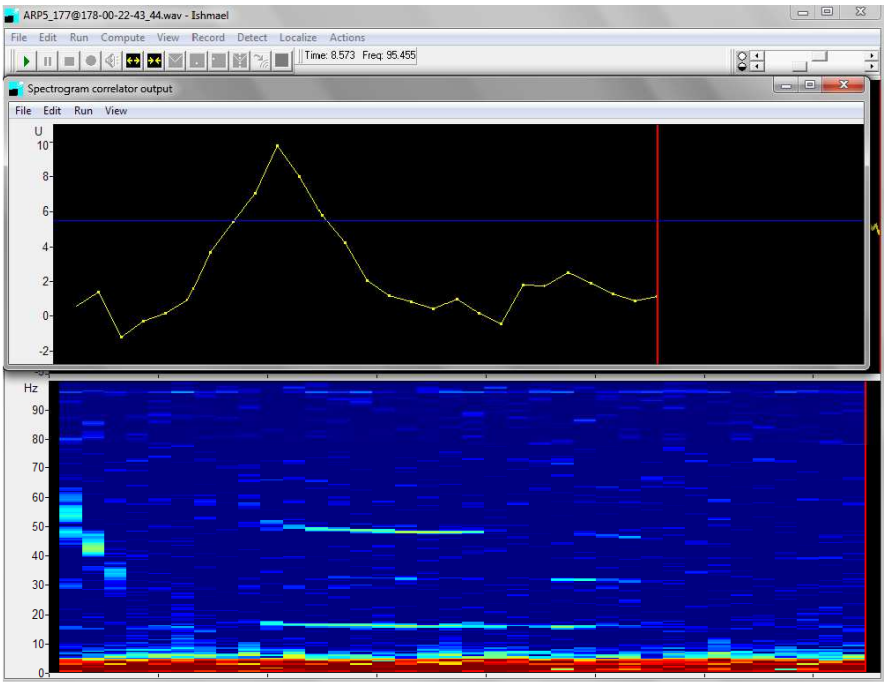


Figure 5.2: Ishmael spectrogram correlation output.

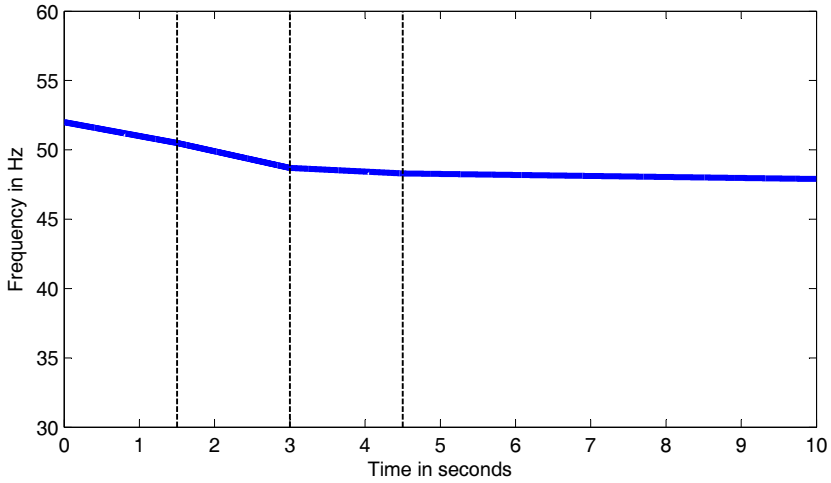


Figure 5.3: Ishmael spectrogram correlation kernel consisting of four segments, seperated by the dashed lines.

5.1.1 Spectrogram Correlation Performance

Spectrogram correlation output of the test signal is shown in figure 5.4. The decision between signal and noise performs well and time resolution is only limited to FFT-window restrictions.

Year	Threshold	Detections	false	missed	FAR [%]	MDR [%]
2000	0.8	11 of 12	0	1	0	8.33
2001	18	27 of 31	11	4	27.19	12.90
2003	0.25	9 of 10	0	1	0	10.00

Table 5.2: Spectrogram Correlation Results evaluating the best Threshold for the Years 2000, 2001, and 2003.

Spectrogram correlation poses a highly exact method to find sweeps in a recording. The

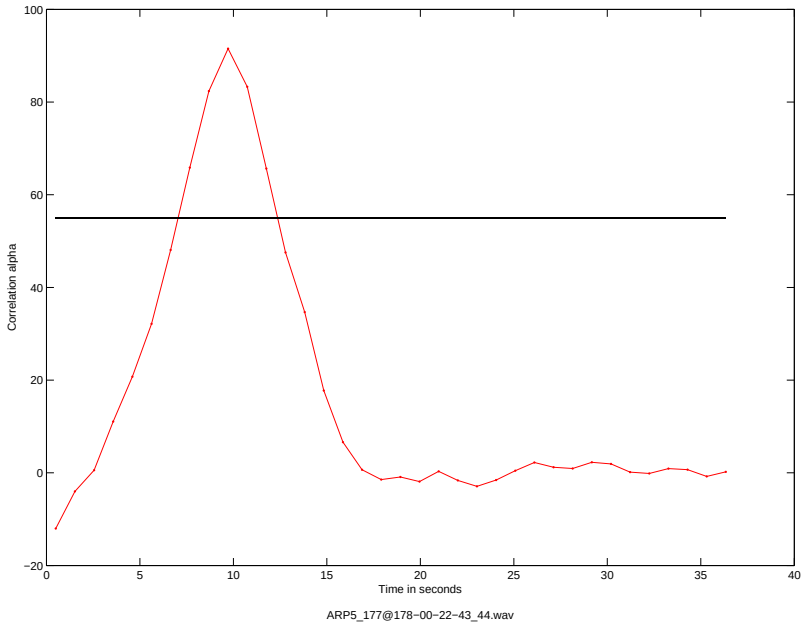


Figure 5.4: Spectrogram Correlation of the test signal

more exactly sweeps are known, the better the performance works. In this case, performance of the algorithm is excellent and proves to be a very good detector reckoning the FAR / MDR criteria as shown in table 5.2 and figure 5.5. Therefore different thresholds have been used as signal power correlates directly with detection output. Yet, pitch shifting through the years is not taken into account. This is marked by a slightly worse MDR when comparing the year 2000 correlation outcome with the one of the signals in 2003.

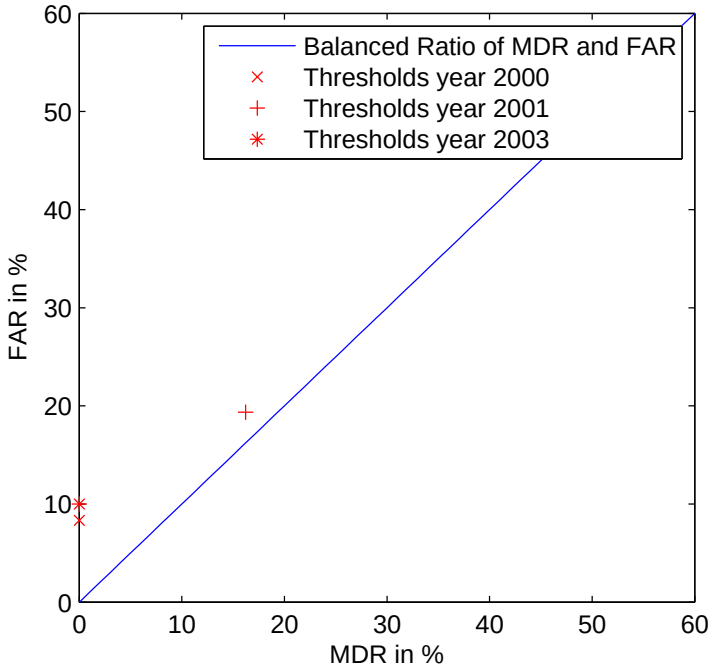


Figure 5.5: Overview of Spectrogram Correlation Results evaluating Different Years, using the figures from Table 5.2.

5.2 Maximum-likelihood signal detection

The Blue Whale B-call consists of harmonically related downsweeps with a basic frequency of about 17 Hz. This mean frequency is decreasing every year (Section 4.2.1) and makes the Blue Whale B-Call difficult to detect with matched filter methods. Though, if not a single matched filter technique is used but a filterbank containing various mean frequencies and sweep rates, a maximum-likelihood-method can help to detect the most probable center frequency and sweep rate of the current signal.
Different sweep shapes which fit the signal best have been evaluated. If the actual sweep is modeled through rather complicate shapes as logarithmic, exponential or root functions, the reproduced signal is more similar to the original (see Section 4). Due to the fact that inter-signal variances are very strong even in the same season, caused by both inter-individual differences and variations in a whale's different vocalizations, a filter bank consisting of appropriate matched filters would contain too many different modes to guarantee an energy and storage efficient usage of a detection algorithm.

5.2.1 Model Design

To detect the sweep rate and the mean frequency, a maximum likelihood estimation was done. Sweeps with different mean frequencies and sweep rates (see Table 5.3)were generated and a cross power calculation between the generated signals and the Blue Whale Call was done. This means that the data was correlated with a number of complex sinusoids shifting down in frequency with a sweep rate from 0 to -1 Hz per second in steps of 0.04 Hz per second. As this MLE's parameters are sweep rate and mean frequency, for every sweep rate step a number of mean frequencies has been applied. As the most interesting part of the blue whale call contains three harmonics, this procedure has been repeated for every harmonic
The maximum correlation for each window and the corresponding sweep rate and mean

	Sweep 3 [Hz]	**Sweep 2** [Hz]	**Sweep 1** [Hz]	**Sweep Rate** [Hz/s]
First value	46.0	46.0/2	46.0/3	0
Last Value	52.1	52.1/2	52.1/3	-1
Step Size	0.1	0.1/2	0.1/3	-0.04

Table 5.3: Different Mean Frequencies and Sweep Rate for the three Sweeps.

frequency were detected. Amongst the different window lengths chosen (1 to 5 seconds) the window length 3.25 seconds raised the highest correlation in the specified signal, which was

the reason for generating the artificial signal model with 3.25 seconds window length (50 % overlap). The different correlation figures for the three spectral lines are shown in Figures 5.6, 5.7, and 5.8[1].

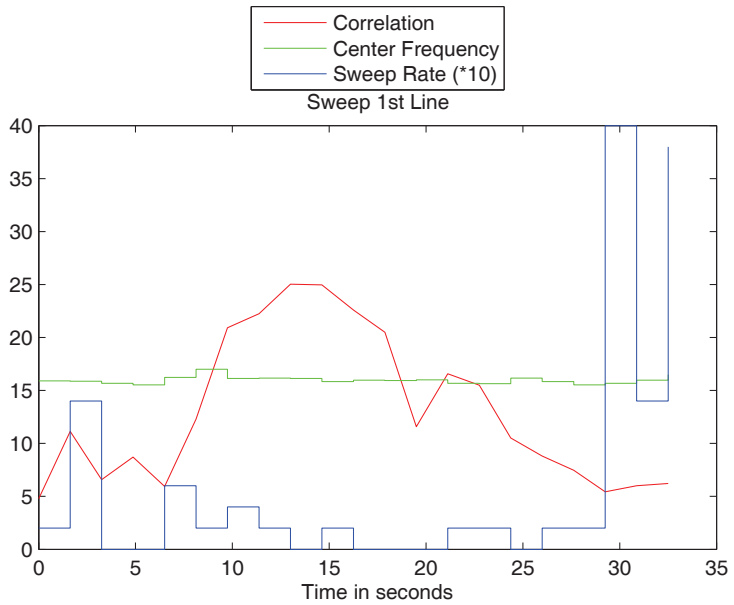


Figure 5.6: 1st Spectral Line. The red line signifies the detection correlation. The green line shows clearly how the mean frequency behaves. The sweep rate, depicted by the blue line, decreases after a short increase.

This shows that the correlation of the second spectral line is not as big as the others, which origins from the lower amplitude of the spectral line. The correlations shown in figures 5.6, 5.7 and 5.8 have shown that the use of the second spectral line is not especially useful for building a filter, as the outcome is not very high.

[1]the Figures' center frequency is in Hz, the sweep rate in Hz per s

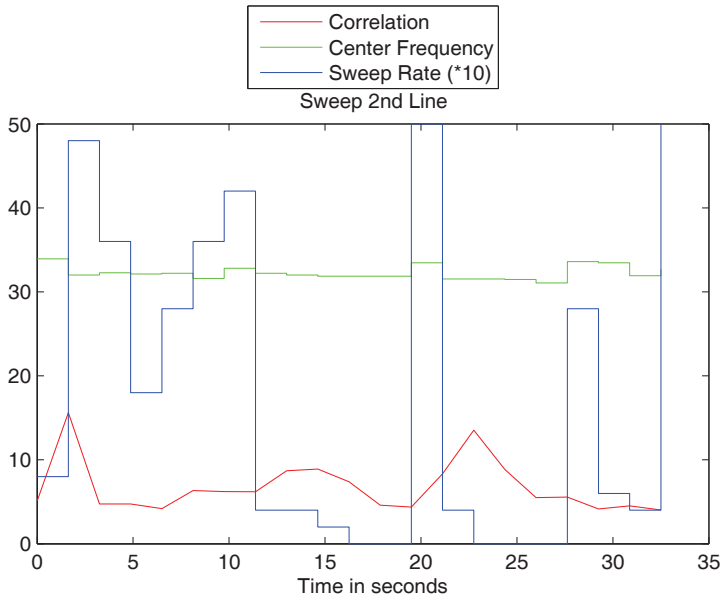


Figure 5.7: 2nd Spectral Line. The detection cannot be verified clearly, as the correlation has similar values in noisy area.

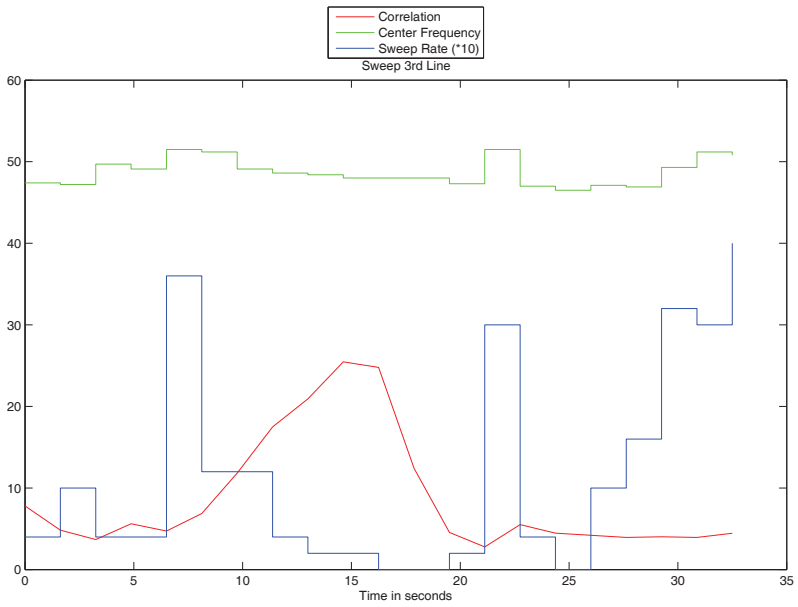


Figure 5.8: 3rd Spectral Line. The correlation here behaves very good, which means that the signal is traceable. The decrease in frequency and sweep rate can be seen clearly, also the increase of the sweep rate at the end of the signal.

5.2.2 Maximum Likelihood Performance

Maximum likelihood detection works in time domain and thus its performance is comparable to matched filtering. The complexity was rather high with building a matrix containing sweeps with different mean frequency and sweep rate. The calculation cost was also the highest of all methods, as correlating in time domain with a sweep matrix took quite some time for all windows. As for matched filtering, calls could be differentiated well when detected. Though, the time window was too long sometimes. In that case, noise which was a disturbance in interesting frequency bands spread in different windows and the decision between call or not call was levelled. Time resolution with 3 s windows works not well, if a method is searched to bear the possibility of acoustic tomography. The evaluation was satisfied with one threshold each year, as the differentiation between signal and noise was rather fluent (see Table 5.4 and Figure 5.9). In reviewing the maximum likelihood method,

Threshold	Year	Detections	false	missed	FAR in %	MDR in %
1.25	2000	5 of 12	6	7	33.33	58.33
1.2	2001	24 of31	8	7	20.51	22.58
1.5	2003	8 of 10	0	2	0	20

Table 5.4: Results of Maximum Likelihood Method evaluating one Threshold for each year.

it proved to be a conducive tool to examine whale calls and searching for mean frequencies and sweep rate. In contrast, real-time automatic detection does not work satisfactorily.

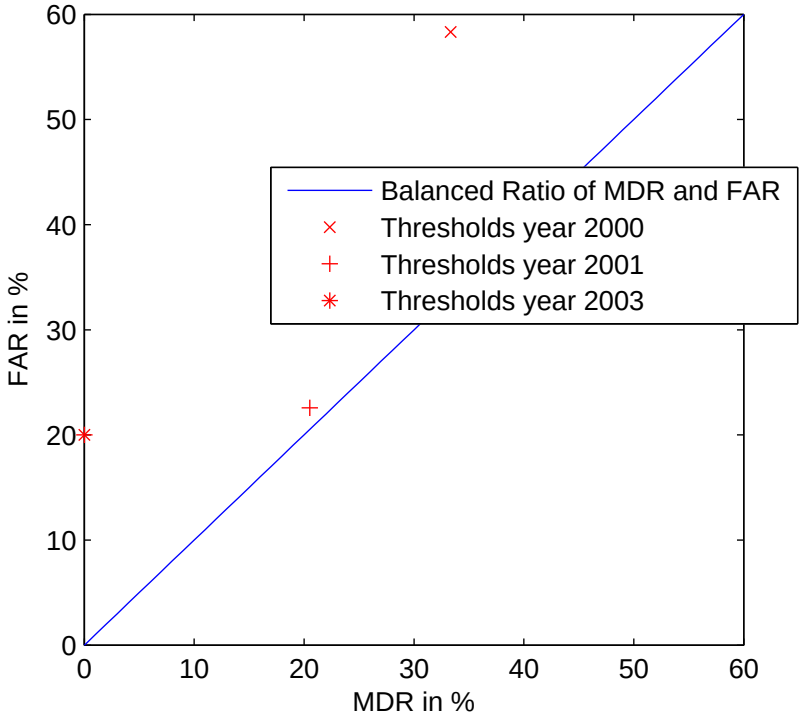


Figure 5.9: Overview of Maximum Likelihood Results evaluating Different Years, using the figures from Table 5.4.

5.3 Bandpass Detection Algorithms

The bandpass detection algorithms calculate the detection rate by using the power in noise and frequency bands for a SNR detection. The spectral power detection algorithm achieves that through FIR filters,

5.3.1 Spectral Power Detection

A power detection method was conducted using 7 bandfilters (see Table 5.5) for extracting interesting frequency bands of the signal as well as specific noise components in between. Some basic theory about power in digital signals can be found in Section 3.6.1.

The power in the different frequency bands was calculated for each window an then the three bands containing signal components were summed up and divided by the noise bands' sum. As window function, a Hamming window has been taken with a length of 2 seconds and an overlap of 90 %.

Lower Bound in Hz	**Upper Bound** in Hz	**Signal or Noise**
9	13	noise
15	19	signal
21	28	noise
30	34	signal
36	44	noise
46	52	signal
54	64	noise

Table 5.5: Powerdetection Bandfilter Boundary Frequencies

The noise bands are very close to the signal bands, so that the effect of environmental broadband noise within the single bands should be neglected.

The flow diagram of the power detection method 3.2 shows an additional highpass which is not necessary when using band filters, but has been used for all signal evaluations.

Powerdetection Performance

If strong noise components are present in the signal bands, the recognition of a whale song can be postulated. This leads to false recognitions. In case of A-Calls there might be some false detections and related to the average detection output of B-Calls, the rate is rather low. The longer a B-Call lasts, the more the frequencies shift downwards into the 'noisy'-area, which leads to a decreased signal to noise proportion in the detection algorithm. It is difficult to discover two calls, when they are close in time, as the detector is not able to find contours or a start-point. The filter output postulates one call, if there are in fact more calls overlapping. White noise detection is negative because of the same power in the noise bands as in the signal bands. Figure 5.10 shows the detector output using the power detection method, processing the test call[2].

For detection, the threshold was set to certain values (see Table 5.6). The detector was set to respond, when the threshold was reached for more than 4 seconds.

The MDR versus FAR graph depicts that the optimal threshold is between 0.8 and 1.0. The distance from origin is about *15* which leads to an overall error rate of 15 % with equality in FAR and MDR.

[2]Picture of test call in Figure 4.2

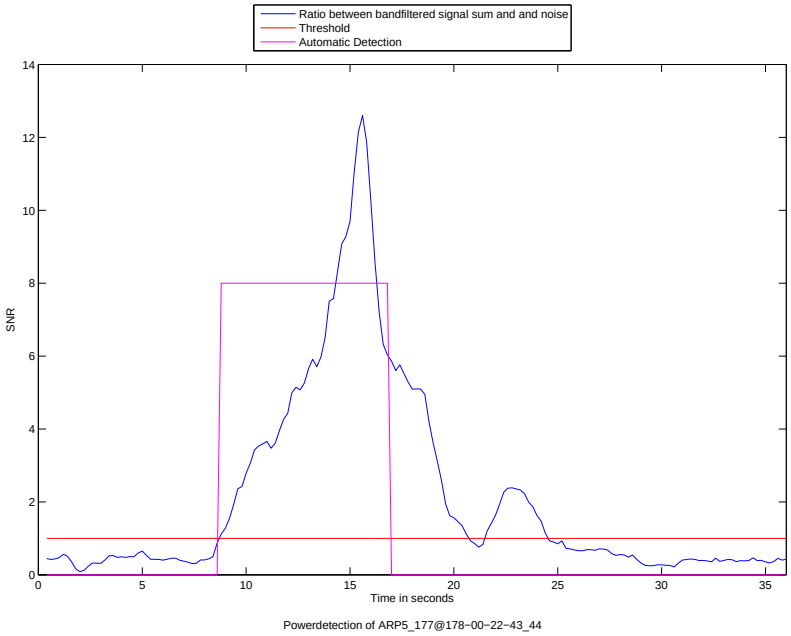


Figure 5.10: Powerdetection of the test signal

5.3.2 Threshold evaluation

The threshold plot 5.11 and the threshold table 5.6 for the power spectral density algorithm show also the iteration process for this method, which is used as an example here of how the threshold adaption has worked. In this method, the optimal threshold differences are not big throughout the years, as band based detection algorithms use a SNR ratio. Correlation methods, such as spectrogram correlation are heavily influenced by the input power 4.1.3.

5.3.3 Band Detection using the Goertzel Algorithm

As for the spectral power detection method, the Goertzel algorithm was used to calculate the proportion of the band-sum with a definite power content if a Blue Whale signal is present. in this process the band-sum did definitely not contain any relevant signal information. The interesting bandwidths are the same as in the power-detection method assuming that the bandwidth of a Goertzel filter at an order of n and a sample rate of f_s Hz is b Hz. The windows in the Goertzel calculation are longer and therefore the time resoultion is much lower, but it results in a much faster calculation. Different approaches were done using this method: First the overlap was bound to 50%. The results were rather inexact, due to the window length. For that reason, the overlap was increased to 95% and the result was

year	Threshold	Detections	false	missed	FAR [%]	MDR [%]
2000	1.5	9 of 12	1	3	7.69	25.00
	1.0	10 of 12	2	2	14.29	16.67
	0.7	11 of 12	2	1	14.29	8.33
2001	1.5	21 of 31	0	6	0.00	22.22
	1.0	25 of 31	2	2	6.90	7.41
	0.7	27 of 31	6	1	18.18	3.70
2003	1.5	5	0	5	0.00	50.00
	1.0	7 of 10	1	3	9.09	30.00
	0.8	8 of 10	3	2	23.08	20.00
	0.7	9 of 10	3	1	23.08	10.00
all years	*1.5*	*35*	*1*	*18*	*1.85*	*33.96*
	1.0	*43*	*4*	*10*	*7.02*	*18.87*
	0.7	*48*	*9*	*5*	*14.52*	*9.43*

Table 5.6: Powerdetection Results evaluating Different Thresholds and Years.

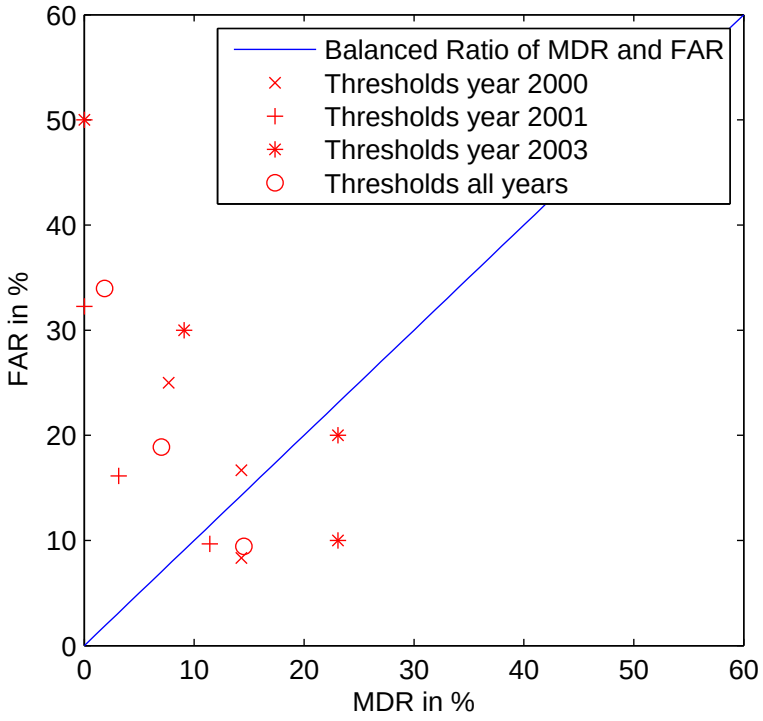


Figure 5.11: Overview of Power Detection Algorithm Outcome evaluating Different Thresholds and Years, using the figures from Table 5.6.

averaged using a filter function producing a running average filtered output signal. As the recordings dating from different years have different sample rates, the average was linked to the actual duration of 30 ms, which results in 30 points at a sampling rate of 1000 Hz and 15 points at a sampling rate of 500 Hz. The running average filter has been calculated by using MATLAB®'s filter function, which is implemented as a direct form II transposed structure. The overall method using the Goertzel algorithm is drawn in Figure 3.3 (see Section 3.6.2), while the mean frequencies and the bandwidth table are shown in table 5.7.

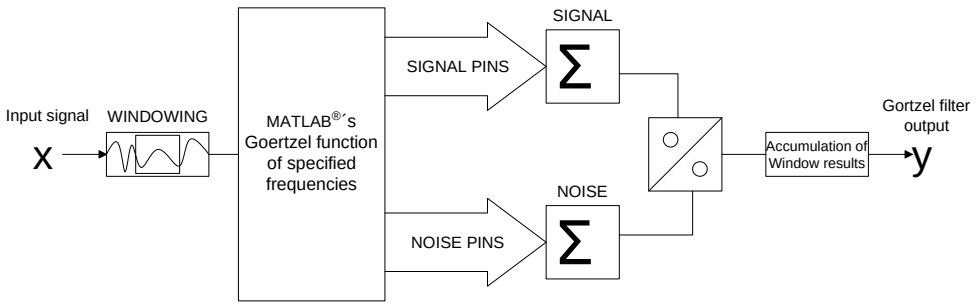


Figure 5.12: Flow Diagram of the Goertzel-detection method.

5.3.4 High Resolution Goertzel DFT-based Detection

The Goertzel algorithm is used for another slightly different SNR calculation as shown in the flow diagram 5.13. Here again, the highpass filtering is not necessary, as band filters are used for detection, but the highpass filtered signal was used for detection. The idea behind this realization is to get a more exact resolution disregarding the parts of the signal which do not contain necessary information, neither signal nor related noise content. The advantage of this approach involves a faster detection with less average calculation time, as only two signal bands are evaluated, and a reduced sample rate is used. Downsampling

Start Frequency f_1 in Hz	End Frequency f_2 in Hz	Number of Frequency Pins	Signal or Noise
10	14	5	Noise
17	19	3	Signal
22	27	6	Noise
32	34	3	Signal
36	47	12	Noise
48	54	7	Signal
59	63	5	Noise

Table 5.7: Goertzel Calculation Table, with the frequencies going from f_1 to f_2, the bandwith of each pin is set to 1 Hz

and punctual frequency content calculation yields a SNR calculated in adjacent frequency pins only. This excludes an interference with narrow band noise which is situated further from the interesting frequency pins.

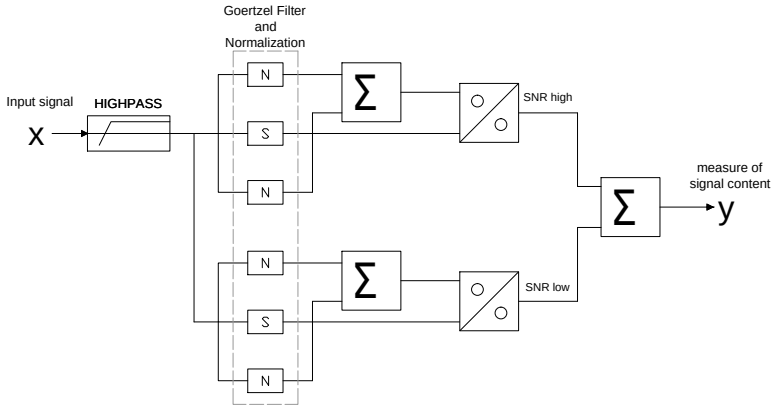


Figure 5.13: Flow Diagram of the second Goertzel-detection method.

Goertzel Filter Performance

Goertzel filter detection outcome has been treated similar to power detection outcome. The time for reaching the threshold was set to 4 s as well, the threshold was only evaluated in two cases (a SNR of 0.7 and 0.6). Figure 5.14 shows the detector's output. Signal bursts

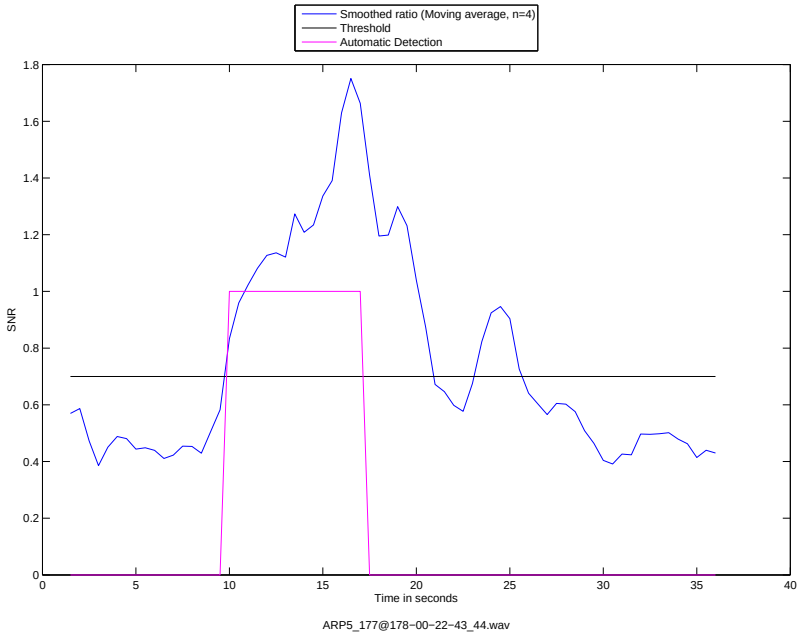


Figure 5.14: Goertzel Detection of the test signal

cause missed detections, if the burst coincidences with a call, as the detector windows are all "infected" with the bursts's impacts. LFlat frequency lines produced by ships passing by around 33 Hz together with other whales' utterances lead to false detections in some cases, as the 33 Hz band is taken as noise identifier (see Table 5.7).
The Goertzel filter performance is comparable to power detection, as the similar approach hypothesizes (see Table 5.8. Though, the MDR versus FAR is slightly better in general, as the needed frequencies seem to be resolved better than in the analogue filter case (see Table 5.15 for the rates).

Threshold	Year	Detections	false	missed	FAR in%	MDR in %
0.6	2000	12 of 12	1	0	7.69	0.00
	2001	28 of 31	5	3	13.89	9.68
	2003	9 of 10	1	1	9.09	10.0

Table 5.8: Results of the Bandpass-based Goertzel Algorithm evaluating Data of Different Years.

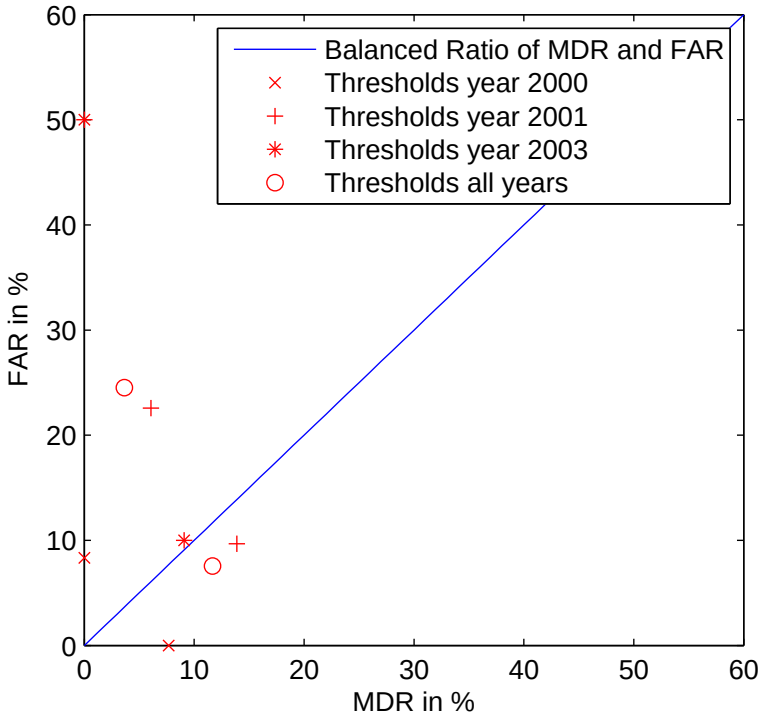


Figure 5.15: Overview of the Bandpass-Goertzel Algorithm Outcome evaluating Different Thresholds and Years, including the Figures for the Best Threshold from Table 5.8.

5.3.5 Performance of the Goertzel-Algorithm Variation

The second variation of the Goertzel Algorithm as proposed in 5.3.4 performs similar to the first one. Limited to a sample rate with 200 Hz, the calculation cost is much lower than in the first variation. When the threshold is set to a low value, the amount of false detections is high, but working as a preliminary filter, this provides almost no missed detections.

Threshold	Year	Detections	false	missed	FAR in %	MDR in %
0.8	2000	12 of 12	1	0	7.69	0.00
	2001	27 of 31	31	4	50.00	12.90
	2003	8 of 10	0	2	0	20
	all years	*47*	*32*	*6*	*37.65*	*11.32*

Table 5.9: Results of the Goertzel Algorithm variation evaluating Different Years.

5.4 Cepstral Analysis

A real cepstrum computation of the input vector is executed with the MATLAB® function *realceps*. The computation is using the formula 3.20 described in section 3.7.

5.4.1 Examples of Cepstrum Representations

Examples of cepstral representations with highpass filter 10 Hz on and off shown in Figure 5.16 and Figure 5.17. Figure 5.18 shows the cepstrum of a resampled signal (new samplerate 200 Hz) with highpass filter. The presentation is built like a spectrogram with quefrecy instead of frequency on the ordinate against time on the abscissa. The color of the cepstral coefficients is a measurement for their size (from blue - low to red - high). Those images do not provide too much recognizable information for detection. This leads to a further calculation of the Δ-cepstrum.

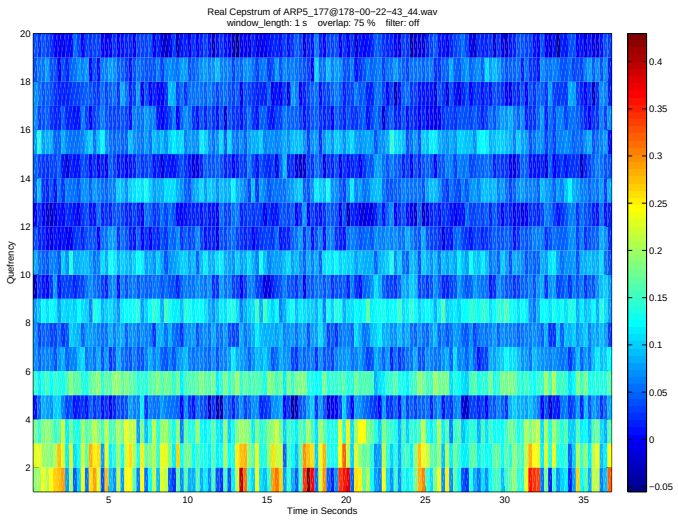


Figure 5.16: Cepstrum plot of input signal with 1 s window length and 50% overlap.

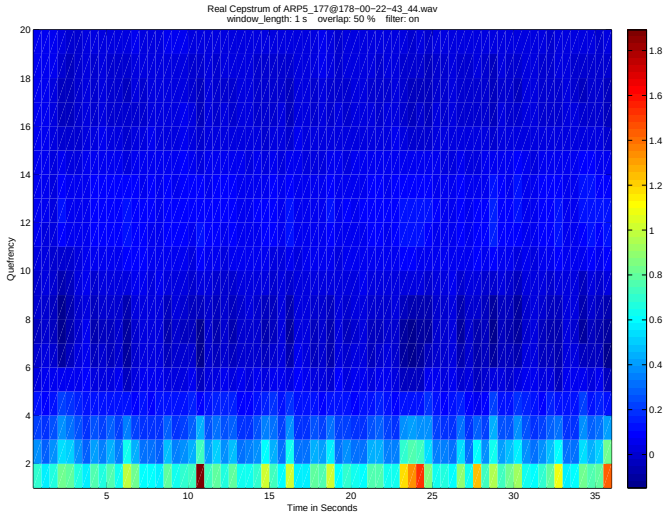


Figure 5.17: Cepstrum plot of input signal with 1 s window length and 50% overlap, highpass filtered

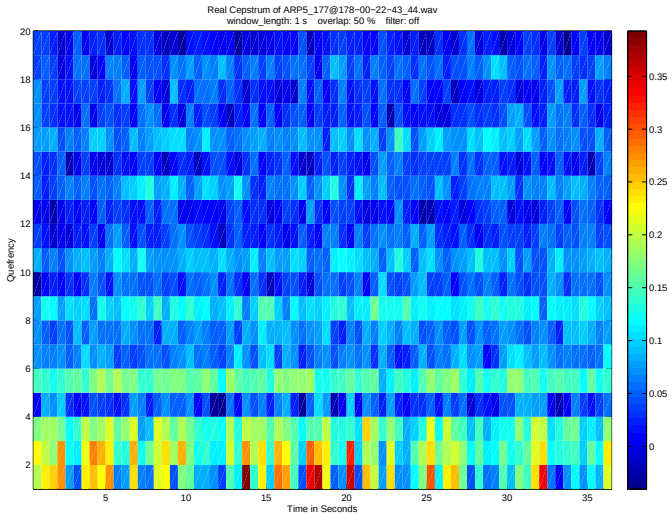


Figure 5.18: Cepstrum plot of input signal with 1 s window length and 50% overlap, highpass filtered, 200 Hz samplerate

The Δ-cepstrum has been evaluated with different $\delta\tau$, according to formula3.22, Figure 5.19 shows a medium $\delta\tau$ of 1 second, Figure 5.20 shows a larger $\delta\tau$ of 2 seconds and Figure 5.21 shows a small $\delta\tau$ of 0.2 seconds. Some peaks in the lower part are recognizable, but

not related to the signal form of the test call in 4.2.

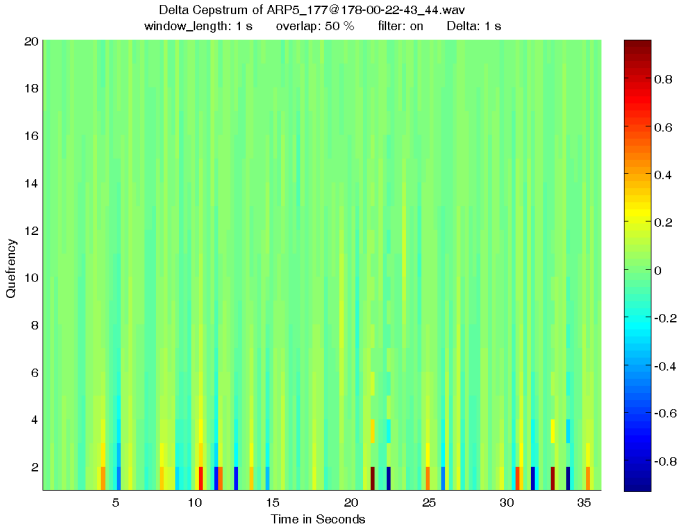


Figure 5.19: Δ-Cepstrum plot of input signal with 1 s window length and 50% overlap, highpass filtered, $\delta\tau$= 1 s

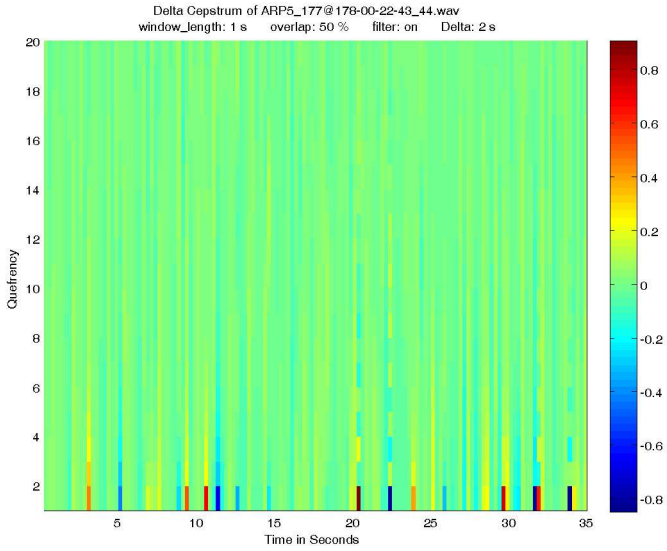


Figure 5.20: Δ-Cepstrum plot of input signal with 1 s window length and 50% overlap, highpass filtered, $\delta\tau$= 2 s

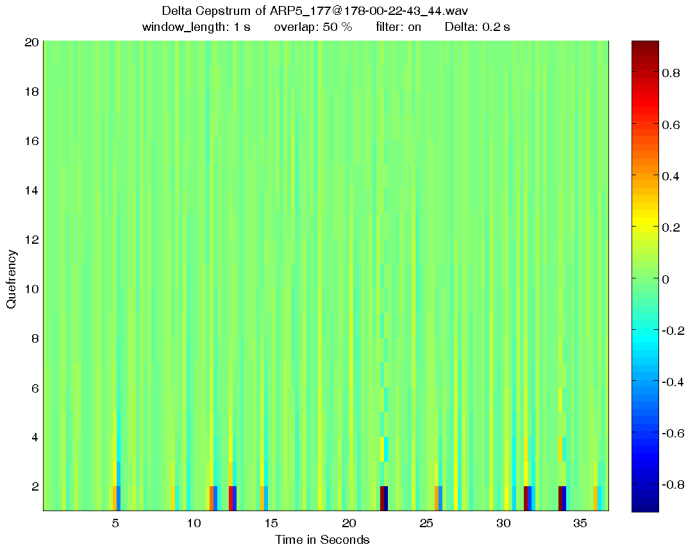


Figure 5.21: Δ-Cepstrum plot of input signal with 1 s window length and 50% overlap, highpass filtered, $\delta\tau$= 0.2 s

5.4.2 Evaluation of the Cepstral Analysis

The fluctuation of the cepstrum vectors is rather small, a unique signature of a signal cannot be evaluated. Filtering and zooming the values evokes visible results. Though, they do not seem to be related to whale calls or well-enough related to develop a procedure for detecting calls. The different Δ-cepstrum results show that the variation of the cepstral vectors is only valuable in the lower pins. The resolution of the variance within the harmonic componentsd is too low and not precise enough for finding a difference within call realizations. The intention was to find a unique fluctuation signature in the Δ-cepstrum, which should have been caused by the harmonic shifting of the downsweeps. Though, this shifting is not constant or precise enough between calls. The periodicity within most of the calls is not too visible, as sometimes only the strong third line of the call is above the noise floor, masking even the basic frequency line. This makes this method not too valuable for Blue Whale tracking. One idea was, that the reason for the failure of the signal tracing through the cepstrum method might probably be the fact that the blue whale signals do not contain any traceable frequency bandwidth above 500 Hz. This was refused by downsampling signal data, which did not lead to valuable results either. As the sweeps are very parallel, a single component can be expected in the cepstrum reflecting the harmonic connection between the sweep lines. The idea was that the sweep rate and its fluctuation could be detected through the Δ-cepstrum. Yet, there were no evident results which might prove a functionality of a cepstrum-detection algorithm for the Blue Whale B-Call (probably due to the overall

noisiness in the signals, as cepstrum analysis is said to be very data dependent [5].

5.5 Improved Spectrogram Correlation

As the spectrogram correlation algorithm in Section 5.1 is a very narrow band correlator, it is very sensitive to pitch change and time variation. This leads to proposing an algorithm using spectrogram correlation in combination with a maximum likelihood estimator. For this estimator, different parameters of modifying the estimator have to be reviewed to be the most promising ones for good results. First of all, a list of possible parameters is compiled to evaluate their relevance for improving detection ability.

- **Variation of the overall kernel window length** The window length variation could help finding Blue Whale calls which are lasting longer, but as the shape is rather constant for the first 10 seconds of the call(see section 2.2), there seems to be no necessity to take this parameter into account for a MLE.
- **Variation of the kernel segments window length** The envelope of the Blue Whale calls stay the same. The variation of the time span within the kernel segments could help improving the performance, and find a more exact model for spectrogram correlation, though. This would optimize the correlation within one year's data, though, frequency shift within the years is not improved by changing this parameter.
- **sweep shape** Sweep shape, which means the usage of non linear sweeps, could help producing a more exact model. The complexity of the algorithm would be increased which leads to an increase of calculation cost. Again, the benefit in catching pitch change is questionable.
- **pitch** Pitch naturally looks like the best parameter for facing change in pitch. The question is, how pitch change can be applied to strain algorithm complexity as little as possible.

According to this evaluation, pitch has been decided to be the parameter of choice to improve performance. The proportion of the different frequencies within one call is assumed to stay constant. This means that the lower the frequency of the call the more it is shifted downwards. This is justified by the logarithmic sensation of sound and the related harmonics, which also prove to be proportional to the basic frequency shift. Recapitulatory, the algorithm for calculating the results of the ISC Method is the same as in the SC method (see section 3.3 and equation 3.3). Moreover, for each window a set of different mean frequencies is evaluated, and the maximum correlation output is taken as the result. This should yield a better detection rate if the mean pitch is lower or higher as the kernel in the original algorithm.

5.6 Implementation of the ISC

Improved spectrogram correlation takes Ishmael's SC settings (see table 5.1) as source for iterating the frequency proportion. Down sampling to 200 Hz also reduces calculation cost. The variation of the sweep's mean frequency is provided through calculating 8 different proportions:

```
prop1 = 1
prop2 = 51/52;
prop3 = 50/52;
prop4 = 49.5/52;
prop5 = 49/52;
prop6 = 48.5/52;
prop7 = 48/52;
prop8 = 47.5/52;
```

Table 5.10: ISC Proportions Table, where 52 Hz is the frequency at the beginning of the original correlation model, 5.1.

These proportions are related to the original settings with 52 Hz as starting point, giving credit to the proposed downward shifting frequency. The downward shifting is calculated in a linear way, taking the same proportion for starting point and endpoint of the sweep. The ISC outcome for the test-call is shown in Figure 5.22 in the results part (chapter 6).

5.6.1 Improved Spectrogram Correlation Performance

Improved Spectrogram Correlation output of the test signal is shown in figure 5.22. The decision between signal and noise, as expected, performs slightly better as the original method, especially for different years. Spectrogram correlation poses a highly exact method to find sweeps in a recording. The more exactly sweeps are known, the better the performance works. In this case, performance of the algorithm is excellent and proves the best possible detector reckoning the FAR / MDR criteria as shown in table 5.11 and figure 5.23.

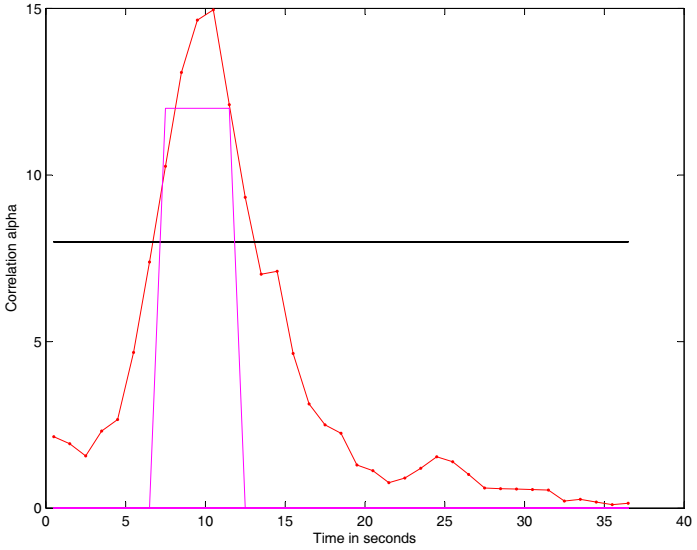


Figure 5.22: MATLAB® improved spectrogram correlation output with the black line to signify the threshold.

Year	Threshold	Detections	false	missed	FAR [%]	MDR [%]
2000	0.15	12 of 12	0	0	0	0
2001	8	29 of 31	3	3	8.82	9.68
2003	0.6	10 of 10	0	0	0	0

Table 5.11: Improved Spectrogram Correlation Results evaluating Different Years.

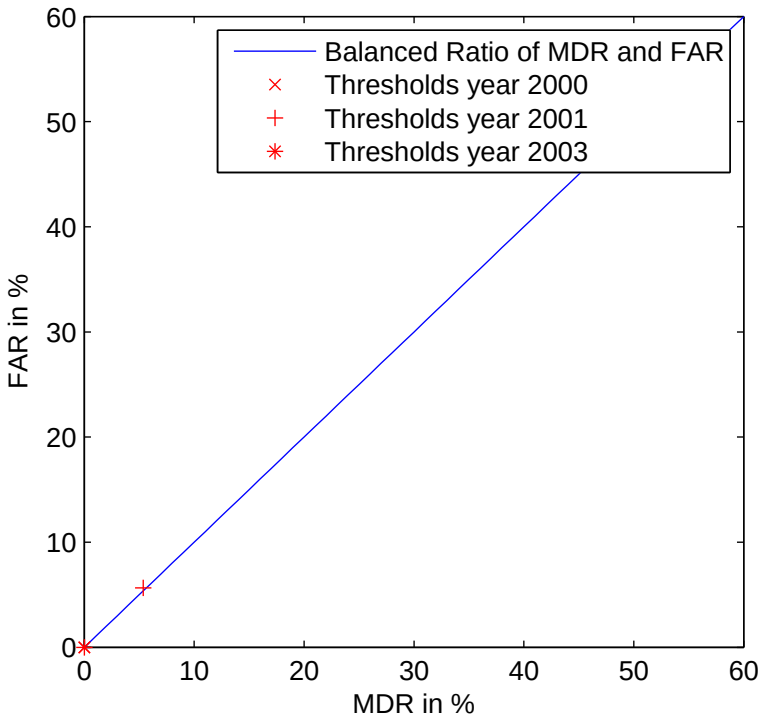


Figure 5.23: Overview of ISC Outcome evaluating Different Years, using the figures from Table 5.11.

Chapter 6

Results and Concluding Comparison

Since every algorithm has now been evaluated individually, a performance analysis containing a comparison of the algorithms including different parameters as time and cost reasons is provided as follows.

6.1 Comparison of evaluated Signal Processing Methods

Comparing all introduced methods, spectrogram correlation proves to be the best detection option, in case it is modified and combined with the benefits of a maximum likelihood detection. The maximum likelihood detection performance falls short of the expectations towards this method. Considering the bandpass based algorithms, DFT based detection works better than power detection, which is quite remarkable, as computing costs are lower. Table 6.1 and table 6.2 shows a final comparison of the different algorithms which have been evaluated in the theses.

Method	Detection Ability	Advantages	Disadvantages
ML Detection	medium	helps with variation in frequency	large computational cost
Power Detection	medium	noise eliminating	rather high computational cost
Goertzel Detection	medium	very fast and noise eliminating	no exact starting point visible
SC	good	good time resolution	works only with exact model
ISC	very good	good time resolution	large computational cost

Table 6.1: Concluding Comparison of the Different Evaluated Algorithms

For the certain evaluations, only a small number of iterations were performed, as the differences between thresholds were rather low. For threshold fitting on a single system,

Algorithm	Detections	false	missed	FAR in%	MDR in %
ML Detection	37	14	16	20.90	30.19
Power Detection	48	9	5	14.52	9.43
Goertzel Detection	49	7	4	11.67	7.55
Improved Goertzel Detection	47	32	6	37.65	11.32
SC	45	5	8	8.62	15.09
ISC	50	3	3	5.36	5.66

Table 6.2: Results evaluating an average detection ratio over different years, taking the Best Threshold for every year, for an overall amount of 53 detectable Blue Whale B-calls.

more iterations with a higher amount of samples should be done to meet the requirements in a statistically firm way. A draft of an optimized algorithm which attempts to consolidate the advantages of some of the evaluated methods is described in Appendix C.

6.2 Evaluation of Ocean Acoustic Tomography Feasibility

Concerning ocean acoustic tomography, a very high time resolution is necessary to be able to perform the method. Even with a synchronization time difference of only 1 ms between the samples, acoustic tomography can be difficult to perform. An even more serious difficulty is provided by the very low frequent Blue Whale utterances. The exact start of a Blue Whale sound is difficult to state with a long window length, but for sea water temperature measurement, the starting time must be stated exactly. Still, with an overlap of 99% or even more, a high time resolution can be provided, Positioning of the floats is another factor of decreasing the chance to perform ocean acoustic tomography with blue whale songs as an input signal. The uncertainty of position cannot be resolved, even after the signal has been sent to the satellite and the way between signal transmissions has been recalculated, as the float is not traceable per satellite in its exposure depth. This traveled path only represents an average estimated way, due to nature forces. A position difference of 1 m produces a time variation of approximately 1 ms in signal arrival (with an estimated average sound speed of 1000 m/s in water). This makes the Blue Whale sounds a difficult input signal in the field of ocean acoustic tomography, as a floating buoy system with unknown individual positioning is used. For the use of ocean acoustic tomography, sender and receiver positions should be fixed, or at least the floating buoys have to be traced and their path must be recorded exactly to recalculate the original position.

Chapter 7

Discussion and Outlook

Throughout this thesis, different processing methods concerning the detection of Blue Whale B-calls have been evaluated. The results concerning the performance of the different algorithms demonstrate that the spectrogram correlation method currently used for Blue Whale B-Call detection proves to be a well established algorithms with a robust and efficient detection ability compared to the other methods. Still, it gains performance in combination with a maximum likelihood algorithm. The performance of the compared methods proved spectrogram-correlation to be very exact and good working in terms of deciding on the time of the whale call start and recognizing a call. Different call shapes of different years have not displayed that great distinction in the test data evaluation.
Still, the improved version is even more promising. If the call shape is different, certain problems may occur with a system working quite a time using the same call pattern. A downward shift in frequency throughout the years cannot be caught completely by the original algorithm, but the proposed method helps providing a wider filter bank to prevent missed detections.
The band filter based methods may cushion the shift, as some frequency components are still detectable within the more low frequent signal. Though, their overall detection rate is not that high and the false detections contradict the use of only frequency related and not time-variant models. Maximum likelihood detection promises an adequate detection if the correlation kernel is exactly matched to the Blue Whale call. However, the performance is, on account to time based evaluation, not that suitable as spectrogram correlation, which works better in an environment that provides no exact starting point and unknown phase.
On balance, the methods have shown that they manage to detect Blue Whale calls in sound data. Future work will be focused on a filter bank like in the improved spectrogram correlation method, which can offer a solution for a better detection ability. The maximum likelihood filtering method could help to find the best fitting sweeps, with a time factor (how long the sweep should last) being added to the method for even better results.

The bandwidth related methods behave appropriate in pre-filtering. A Goertzel-technique could probably decimate calculation cost in spectrogram-correlation, as only a few frequency pins are evaluated. For online detection, the band related filters are exact in postulating the very beginning of a call. These filters might be installed before spectrogram-matrix correlation activates to avoid permanent calculation and decrease the calculation cost, as proposed. In that case, a pre-processed result-stream would help to minimize the data volume by checking only for calls. This should meet the requirements of finding a great amount of whale calls with precise time resolution to enable proper localization.
As a result of this thesis, a detailed signal analysis of Blue Whale B-calls is offered and leads to further ideas to be implemented as a combined method of band based detection and maximum likelihood evaluation. The very good performance of the improved spectrogram correlation method has to be emphasized: When the floating buoys of the system stay in the sea for two years, the frequency shift within the calls through this time can be caught and this gives a high potential for the use of this algorithm in a longer exposure time on site.

Bibliography

[1] Therese C. Moore Ching-Sang Chiu, Christopher W. Miller and Curtis A. Collins. Auto-detection and censusing of blue whale vocalizations along the central california coast using a bottom-lying hydrophone array. *Oceanology International 99 Pacific Rim Conference Singapore*, pages 435–440, April 1999. 7

[2] William C. Cummings. Underwater sounds from the blue whale, balaenoptera musculus. *Journal of the Acoustical Society of America*, 50(4B):1193–1198, 1971. 6, 8

[3] P. Delacourt and C.J. Wellekens. Distbic: A speaker-based segmentation for audio data indexing. *Speech Communication*, 32(1-2):111 – 126, 2000. Accessing Information in Spoken Audio. 26

[4] G. Doblinger. *MATLAB-Programmierung in der digitalen Signalverarbeitung.* J.Schlembach Fachverlag, Weil der Stadt, 2001.

[5] David P. Skinner Donald G. Childers and Robert C. Kemerait. The cepstrum: A guide to processing. *Proceedings of the IEEE*, 65(10):1428–1443, October 1977. 25, 54

[6] John Higgins et. al. Blue whale (balaenoptera musculus), January 2013. *http://www.britannica.com/EBchecked/topic/70418/blue-whale.* 5

[7] C. G. Fox K. M. Stafford and D. S. Clark. Long-range acoustic detection and localization of blue whale calls in the northeast pacific ocean. *Journal of the Acoustical Society of America*, 104(6):3616–3125, December 1998. 8

[8] Steven M. Kay. *Fundamentals of Statistical Signal Processing.* Prentice-Hall PTL, New Jersey, 14^{th} edition, 1998. 20

[9] Enchanted Learning. Blue whale, the loudest animal on earth, September 2011. *http://www.enchantedlearning.com/subjects/whales/species/Bluewhale.shtml*. 28

[10] MathWorks. Cepstrum analysis, March 2014. *http://www.mathworks.de/de/help/signal/ug/cepstrum-analysis.html.* 25

[11] Mark a. McDonald, John Calambokidis, Arthur M. Teranishi, and John a. Hildebrand. The acoustic calls of blue whales off California with gender data. *The Journal of the Acoustical Society of America*, 109(4):1728, 2001. 6, 7, 8, 28

[12] Herman Medwin. *Sounds in the Sea: From Ocean Acoustics to Acoustical Oceanography.* Cambridge University Press, New York, 2005. 15, 16, 67, 71

[13] David K. Mellinger. Recognizing transient low-frequency whale sounds by spectrogram correlation. *Journal of the Acoustical Society of America*, 107(6):3518–3529, 2000. 18, 19, 33

[14] Scripps Institution of Oceanography UCSD. Scripps whale acoustic lab, January 2014. *http://www.cetus.ucsd.edu/.* 1, 66

[15] Erin M. Oleson, John Calambokidis, Jay Barlow, and John a. Hildebrand. Blue Whale Visual and Acoustic Encounter Rates in the Southern California Bight. *Marine Mammal Science*, 23(3):574–597, July 2007.

[16] A. V. Oppenheim and R W. Schafer. *Zeitdiskrete Signalverarbeitung.* R. Oldenburg Verlag, München, 3rd edition, 1999. 18, 20, 24, 25

[17] Beat Pfister and Tobias Kaufmann. *Sprachverarbeitung: Grundlagen Und Methoden Der Sprachsynthese Und Spracherkennung (Springer-Lehrbuch).* Springer Publishing Company, Incorporated, 1 edition, 2008. 25, 26

[18] John G Proakis and Dimitris G Manolakis. *Digital Signal Processing: Principles, Algorithms, and Applications, 4/e.* Pearson Education, 2007. 23

[19] W. John Richardson, Charles R. Greene, Jr, Charles L. Malmen, and Denis H. Thomson. *Marine Mammals And Noise.* Academic Press, San Diego, 1995. 7, 9, 15, 67, 71

[20] Frank Schätzing. *Der Schwarm.* Kiepenheuer & Witsch, Köln, 2004. ix

[21] Petr Sysel and Pavel Rajmic. Goertzel algorithm generalized to non-integer multiples of fundamental frequency. *EURASIP J. Adv. Sig. Proc.*, 2012, 2012. 23

[22] ARGO Project Office UCSD. Argo - part of the integrated global observation strategy, January 2014. *http://www.argo.ucsd.edu/.* 2

[23] Oregon State University. Ishmael, 2010. *http://www.bioacoustics.us/ishmael.html.* 33

[24] San Francisco State University. Blue whales align the pitch of their songs with extreme accuracy, study finds., August 2010. *http://www.sciencedaily.com/releases/2010/08/100802141907.htm.* 30

[25] Gordon M. Wenz. Acoustic ambient noise in the ocean: Spectra and sources. *Journal of the Acoustical Society of America*, 34(12):1936–1956, 1962. 11, 13

[26] Sean Wiggins. Autonomous acoustic recording packages (arps) for long-term monitoring of whale sounds. *Marine Technology Society Journal*, 37(2):13–22, 2003. 66

[27] Sean M. Wiggins and John a. Hildebrand. High-frequency acoustic recording package (harp) for broad-band, long-term marine mammal monitoring. *2007 Symposium on Underwater Technology and Workshop on Scientific Use of Submarine Cables and Related Technologies*, pages 551–557, April 2007. 66

[28] Peter Tyack William A. Watkins and Karen E. Moore. The 20-Hz signals of finback whales (balaenoptera physalus). *Journal of the Acoustical Society of America*, 82(6):1901–1912, 1987. 10

Appendices

Appendix A

The SCORE Project

In the year 2000 the SCORE project was started by the UCSD together with other research institutes to gather acoustic information about marine mammals in Californian sea. *SCORE* stands for *S*outhern *C*alifornia *O*ffshore *R*ange, a region where US Navy frequently conducts operations. The recording environment was specifically developed for this project. *A*utonomic *R*ecording *P*ackages (ARP) (see Figure A.1) have been designed to record ocean sound without being attached to an external energy source. The storage system was capable of about 10 TB data per deployment. Post-processing was done in MATLAB® with a program-environment called *Triton*. The SCORE project is now finished, detailed information can be found under [14] and in Wiggins' paper [27].

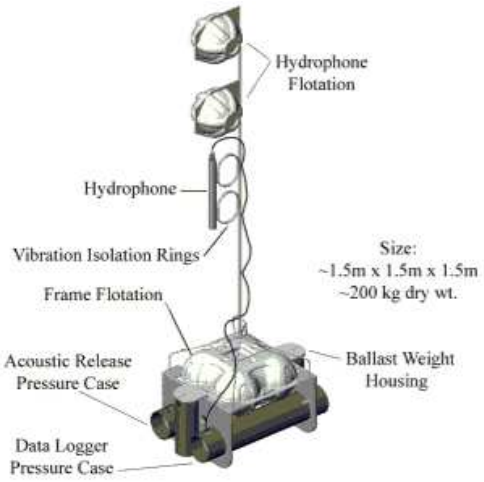


Figure A.1: Acoustic Recording Package,picture taken from [26].

Appendix B

Sound Propagation in the Sea

Sound propagation in the ocean is heavily influenced by different temperature and sound speed profiles explained in 2.3.1. In this chapter some underwater sea sound facts have been taken from [12] and [19] to explain this interesting topic a bit more in depth.

Several calculation methods concerning the sound speed at a specific place exist. Most of them were found heuristically, for example the equation B.1 discovered by Mackenzie in 1987. It claims to have a standard error of 0,07 m/s and works as well below a depth of 1000 m with a broad range of salinities:

$$\begin{aligned} c &= 1448.96 + 4.591T - 5.304 * 10^{-2}T^{-2} + 2.374 * 10^{-4}T^3 \\ &\quad +1.340(S-35) + 1.630 * 10^{-2}z + 1.675 * 10^{-7}z^2 \\ &\quad -1.025 * 10^{-2}T(S-35) - 7.139 * 10^{-13}Tz^3, \end{aligned} \tag{B.1}$$

c signifies the sound speed at a certain place, T stands for the Temperature in degrees centigrade, S means the salinity in parts per thousand of dissolved weight of salts and z represents the depth in meters.

Naturally, the geographic environment influences the oceanic sound field through reflections and refractions, as layers of different sound speed meet in the sea. The wave properties of sound lead to diffraction and interference effects. Currents and eddies can cause Doppler shifts to a certain signal which leads to non-linear frequency distortions. Generally, the sound field in the ocean is a very complex subject. The low sound attenuation in the ocean (Section B.1) denotes a possibly long way for the sound to travel from the source to the receiver. This might be the case, because there are multiple paths for the sound to reach the receiver at a certain time (with a certain phase).
Let the sound come from a point source, the three dimensional wave equation in terms of the acoustic pressure for its radiation is

$$\nabla^2 p = \frac{1}{c^2}\frac{\delta^2 p}{\delta t^2} \tag{B.2}$$

where ∇^2 or Δ relates to the *Laplacian* of the sound pressure p, c is the sound speed, and t means the time in seconds.
Thereby can be seen that the behavior of the sound in water is generally very similar to the one in air. The main difference is the sound speed, which is much higher in water. According to the formula

$$c = \lambda \cdot f, \tag{B.3}$$

where c is the soundspeed in m/s, f the frequency in Hz and λ the wavelength in meters, the wavelength increases with an increase of the soundspeed. A longer wavelength leads to different frequency reaction in inclination with the same geometry as inclination is only relevant at the size of the wavelength (e.g. at an average soundspeed of 1000 m/s 10 m at a frequency of 100 Hz). Moreover, the sound speed layers in the ocean and the low attenuation cause the dimensions of the water to act like a lens for the sound.
In deep water, the way the sound takes in the ocean can be modeled on ray paths that include all reflecting and refracting properties of media with different acoustical density. Sound is partially reflected and partially absorbed at a medium passage, as shown below in Figure B.1 with the ocean floor and the air-water gateway. In case a sound beam hits the

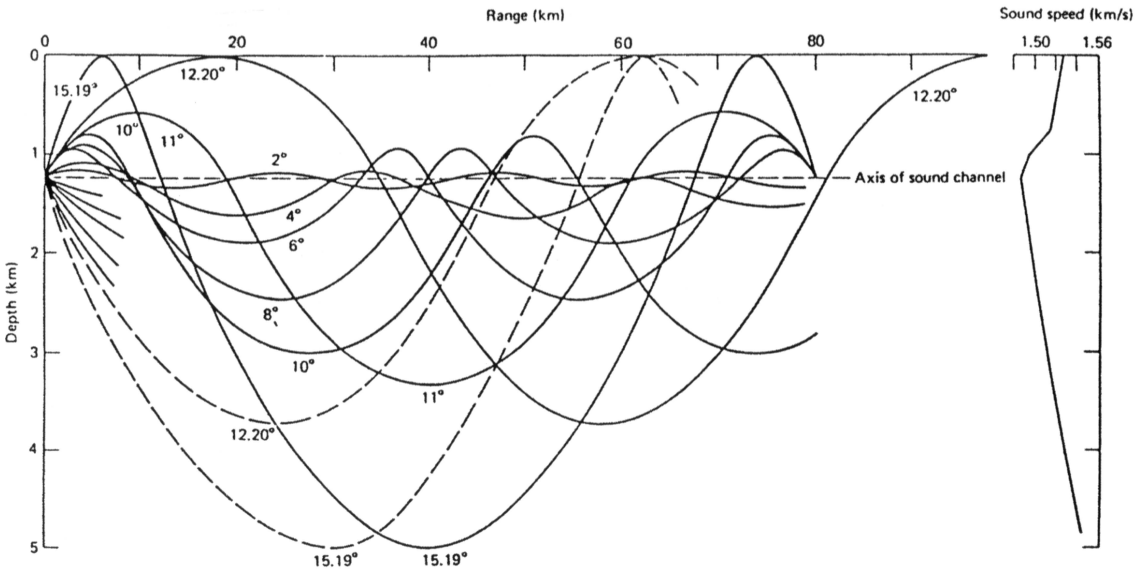
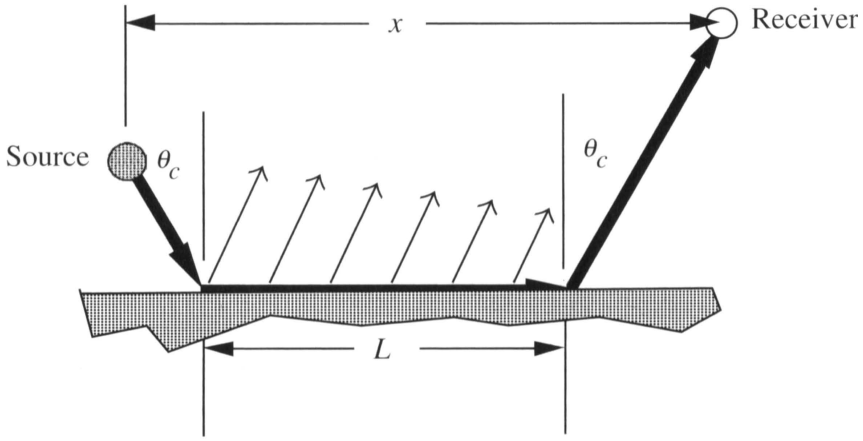


Figure B.1: Sound channel profile in the ocean.

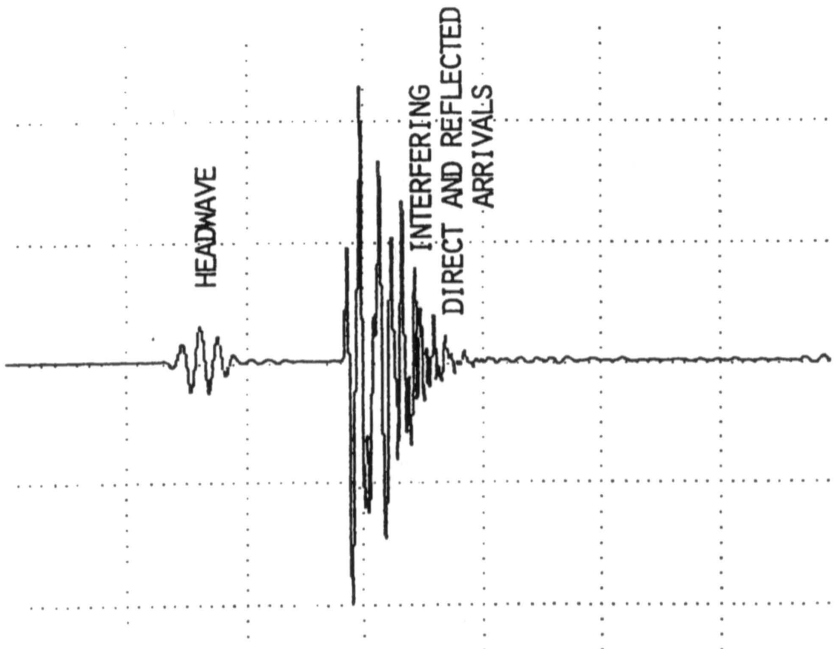
ocean at the bound angle, a "head-wave" starts to propagate along the ocean floor. It starts to radiate at the reflected absolute angle at any point along the floor, assuming every point of the bottom of the ocean being a heuygens-source of an elementary wave (Figure B.2). As the sound is faster in most ocean sediments than in the water itsself, although the path of source through the receiver is longer through the ocean floor, the direct sound is beaten



Head wave propagation through the ocean floor.

Figure B.2: Head wave

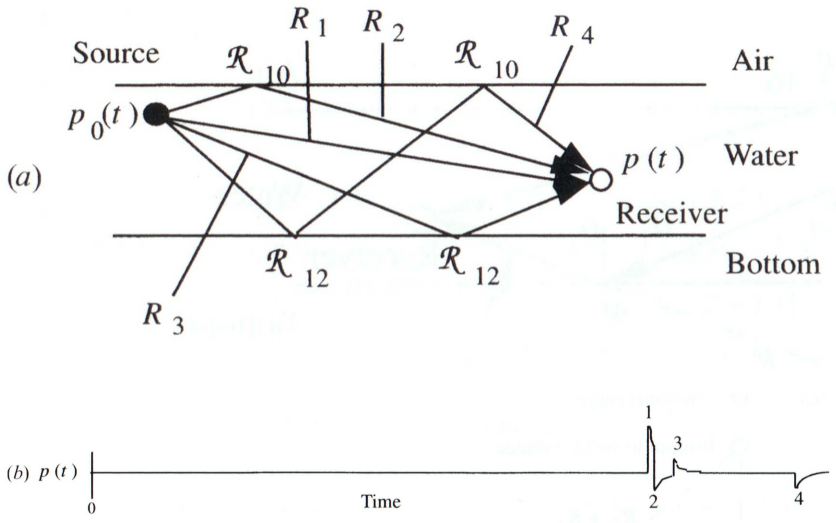
by head wave sound in terms of arrival time. If the minimum of the sound speed in the



Arrival time profile for sound in deep water; the first incoming sound represents the head wave going through the ocean bottom.

Figure B.3: Arrival time profile

ocean is somewhere below the surface (as it is the case in temperature profiles which are at a more moderate climate than the Arctic), the acoustical behaviour of the sound is similar



Ray diagram for an Atlantic Ocean sound channel, the minimum of sound speed is at 1300 m depth.

Figure B.4: Ray Diagram

to the light-wave propagation in optical fibres. If the source is within the channel and the vocalizing angle beyond the critical angle for reflection at different sound speed layers, the sound travels along the axis of the channel. The so called "ray tube" is the reason for the ability of ocean sound to travel such long distances as the attenuation is very small (see Section B.1). Generally it can be found somewhere near to 1000 m depth. Whales are believed to use this ray tube to communicate over far distances, which might also be supported by the observations of whales diving into a depth of a few hundred meters before vocalizing.
If the sound source is near to the surface, a different phenomenon occurs. In this case the ray paths go deep, are reflected at about 4000 m and turn up again. At a range of about 90 km the rays converge and cause a large acoustic pressure in the convergence zone, which is called *caustic* (Figure B.5).

As the velocitiy of some ocean currents, the wave and sound-reflecting objects (e.g. ships) cannot be neglected, the sound-emissions can be Doppler-shifted. These Doppler shifts lead to nonlinear frequency distorsions.
In shallow water, when the water-depth is about the size of the wavelength of the sound, the sound propagation can rather be explained by different vibration modes (e.g. sound that propagates in solid media) than by a mirror-source principle.

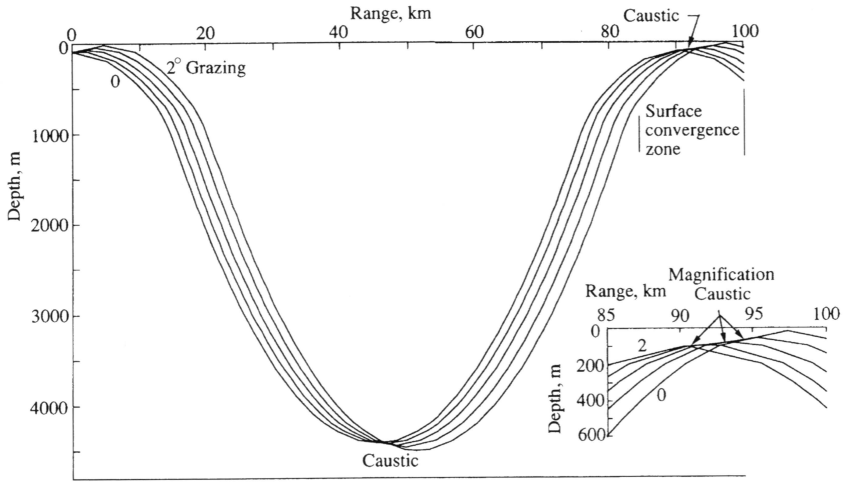


Figure B.5: Model of a convergence zone in the ocean (caustic).

B.1 Sound Level and Attenuation

Sound attenuation in the ocean can be examined in terms of attenuation of airborne noise which radiates into the ocean or ocean sounds which spread directly from their underwater source into the water. Sound in general has a lower attenuation in water than in air.
Absorption of sound energy in water is not only caused by water molecules but also related to various chemicals, which (some in very small amounts) make the transmission loss considerably higher than in fresh water. The two main chemical influences on absorption refer to magnesium sulfate, which increases absorption at 10 - 100 kHz, and boric acid. The latter is responsible for higher sea water attenuation below 5 kHZ, a process which is influenced by other chemicals in sea water as well. Changes in temperature also cause different absorption rates created by shear viscosity, a well known process in air. Detailed formulae can be found in [12]. Overall absorption loss versus frequency range is according to [19]

$$\alpha = 0.0036 \cdot f^{\frac{3}{2}} \, [\mathrm{dB/km}] \,, \tag{B.4}$$

where α is the absorption loss, and f the frequency. The dependency of frequency and attenuation is shown for different types of water (sea and fresh) and different temperatures in Figure B.6.

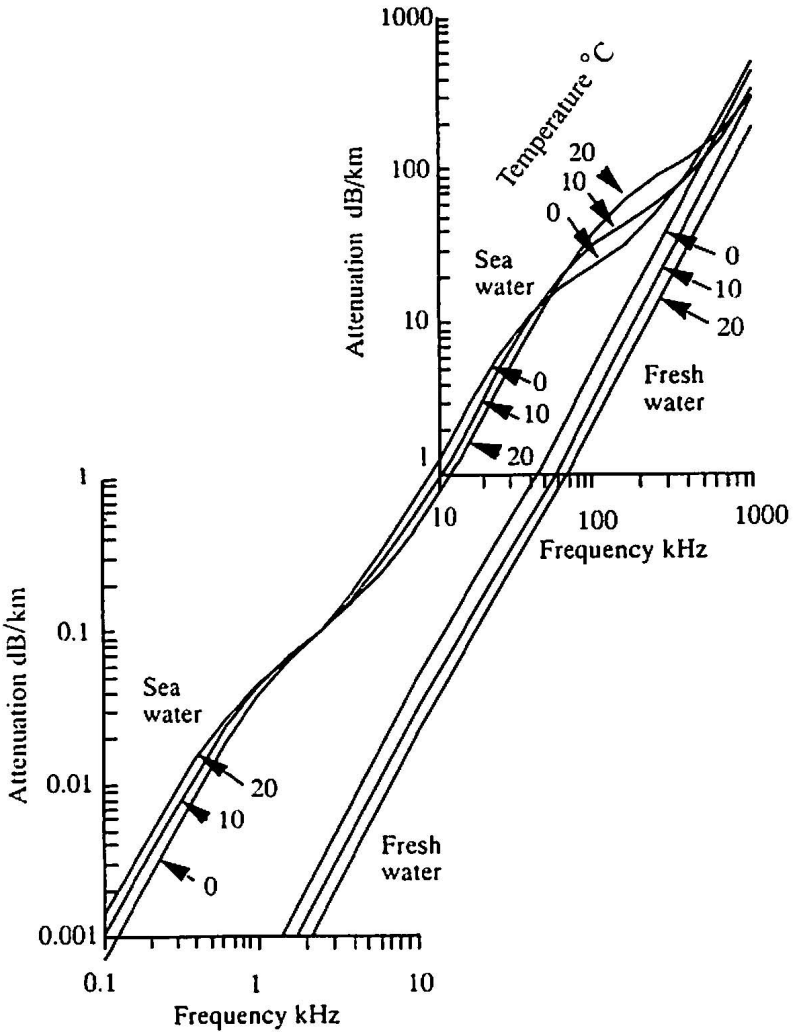


Figure B.6: Sound pressure and attenuation rate in dB/km in fresh and sea water at temperatures 0°, 10°, 20° C, parameters pH of sea water = 8, Salinity = 35 ppt, depth = 0 m.

Appendix C

Further Ideas

The aim of this idea is to include the benefits of a fast working power detection with the improved spectrogram correlation explained in 5.5. This should provide a fast detection in valuable bands, like the proposed goertzel algorithm. This stream should include triggers for all detectable whale songs, which could be achieved by a threshold as low as possible. This leads to false detections, but if the goertzel method is used rather to work as a precleaner than as a detection algorithm, the actual work could be done with another algorithm. For this, improved spectrogram correlation would be the method of choice to provide the actual detection. A specific frequency in the beginning of the signal can serve as a marker for the proposed starting time of the call. This helps automatic tracking with several recording devices simultaniously. Figure C.1 shows a schematic of the contemplated algorithm.

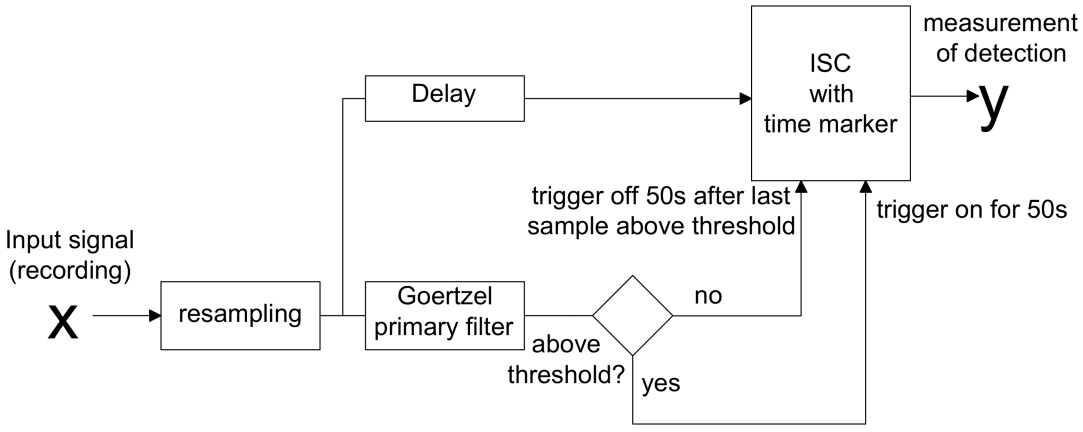


Figure C.1: Schematics of a proposed Combined Power Watch Spectrogram Correlation Algorithm

Appendix D

Sound Files, Data and Program Files

MATLAB® has been used to calculate the results. All program files can be found on a seperate DVD. This chapter proposes a table of the sound data (see D.1). Based on those data the algorithms have been developed. The soundfiles can be found on the DVD, both filtered, downsampled or raw, as well as the results of the algorithms employed on the input data.

Filename	Samplerate	Signal MAX	Signalpower	Filename	Samplerate	Signal MAX	Signalpower
00253-1428068.wav	1000	0.983734131	0.041729865	ARP5_171_22_13_30_78_noise+whales.wav	1000	1	98.2642727
00253-1429011.wav	1000	0.983734131	0.039311024	ARP5_171_22_21_50_78_noise(earthquake)+whales.wav	1000	0.999969482	46.99123959
00253-1433590.wav	1000	0.983703613	0.039358186	ARP5_171_22_40_50_78_noise(earthquake).wav	1000	0.999969482	60.91123774
00253-1434390.wav	1000	0.983703613	0.048263838	ARP5_172_00_00_00_78_noise.wav	1000	0.999969482	76.0607347
00253-1434482.wav	1000	0.983703613	0.048175753	ARP5_177@178-00-22-43_44.wav	1000	1	85.01296772
00254-0703006.wav	1000	0.980926514	0.038298976	ARP5_177@178-00-26-45_44.wav	1000	1	103.0102024
00254-0704235.wav	1000	0.980926514	0.045670486	ARP5_177@178-13-08-10_11.wav	1000	1	130.4491618
00254-0713119.wav	1000	0.980895996	0.063351166	ARP5_177@178-13-47-17_40.wav	1000	1	97.16772662
00254-0715434.wav	1000	0.980895996	0.049179075	ARP5_177@178-13-58-00_40.wav	1000	1	122.3621003
00254-1402100.wav	1000	0.979766846	0.054096403	ARP5_177@178-14-00-10_40.wav	1000	1	83.76211574
03254-1131335.wav	500	0.007415771	0.00038205	ARP5_177@178-14-03-10_40.wav	1000	1	99.10078355
03255-1904284.wav	500	0.015289307	0.012849924	ARP5_177@178-16-25-32_11.wav	1000	1	119.7442431
03255-1906113.wav	500	0.014068604	0.016757292	ARP5_177@178-16-29-22_11.wav	1000	1	155.7135748
03255-1907036.wav	500	0.013153076	0.012039593	ARP5_177@178-16-32-03_11.wav	1000	1	252.9762203
03255-1907542.wav	500	0.025482178	0.031378336	ARP5_177@178-17-04-40_50.wav	1000	1	315.9612177
03255-1908250.wav	500	0.011077881	0.009717885	ARP5_177@178-17-06-54_50.wav	1000	1	339.350371
03255-1908465.wav	500	0.010040283	0.003650096	ARP5_177@178-18-24-35_50.wav	1000	1	131.1984672
03255-1909372.wav	500	0.00894165	0.00381449	noise1.wav	1000	1	405.69195
03255-2324373.wav	500	0.003936768	0.000737113	noise2.wav	1000	0.026824951	0.060617684
ARP5_170@175-17-22-40_78.wav	1000	1	93.84776612	noise3.wav	1000	1	134.490255
ARP5_171_16_58_35_78.wav	1000	1	80.72120371	noise4.wav	1000	1	482.0928245
ARP5_171_17_00_50_78.wav	1000	1	117.7794722	noise5.wav	1000	1	138.5062633
ARP5_171_17_15_45_78.wav	1000	1	143.002486	whitenoise_30s.wav	1000	0.999969482	336.1969906
ARP5_171_19_51_20_78.wav	1000	1	157.4782273				

Table D.1: Samples Overview