Chapter 1: Introduction

Sound is known as “a disturbance in pressure that propagates through a compressible medium” [1]. Water which accounts for more than 70% of the Earth surface supports higher propagation speed of sound waves than other forms of radiation. Like the role played by radar and radio waves in atmosphere and space, underwater sound waves are used in the oceans to detect and locate targets, to measure the characterization of marine environments and to transmit signals for communication purposes [2]. Two types of sonar (Sound Navigation and Ranging) can be noticed here: Active sonar that relies on transmitting a characteristic signal and receiving its reflected echo from the target and Passive sonar that depends on the interception of underwater sound coming from the target itself.

The original concept of man using sound as an effective method of underwater sensing dates back to the 15th century when Leonardo da Vinci noted that: “If you cause your ship to stop, and place the head of a long tube in the water and place the outer extremity to your ear, you will hear ships at a great distance from you” [3]. The marine mammals however have used underwater sound for communication with one another for millennia. Bottlenose dolphins use echolocation sounds to detect the presence and location of objects such as prey [4].

1.1 Marine mammals’ vocalisations and hearing

The sirenians (Manatees, dugong, and Steller’s sea cows), carnivore (Pinnipeds, Sea otters, and polar bears), and cetaceans are considered as the orders of Marine mammals. Cetaceans can be divided into Odontoceti (Toothed whales) and Mysticeti (Baleen whales). These creatures use underwater sounds as a primary method for communication [4]. They use different orders of frequency for this purpose. While toothed whales call mainly at moderate to high frequencies (1-20 kHz) and echolocate at high and very high frequency (100 Hz-150 kHz), baleen whales vocalise at low-and moderate frequency sounds (12 Hz to 8 kHz) [4]. Figure 1.1 illustrates examples of vocalisations emitted by toothed whales (Sperm whales and Bottlenose dolphins) and baleen whales (Humpback
and Bowhead whales). These vocalisations are displayed as time-frequency spectrograms with intensity (dB) which is represented by a colour bar.

Figure 1.1: Spectrographic examples of whale’s vocalisations. Note the time and frequency scales are different for each plot. Note that spectrograms are normalized (dB).

Cetacean’s calls can be classified as tonal and impulsive calls. Calls are considered tonal when the fundamental frequency is much stronger than the harmonics [5]. These tonal calls which include whistles and moans range from the high frequency social-purpose whistles emitted by dolphins to the low-frequency tonal calls vocalised by most baleen whales. Figure 1.2 shows the frequency ranges of these moans and whistle sounds.
Figure 1.2: Frequency ranges of tonal sounds—moans and whistles produced by baleen and toothed whales. An asterisk (*) indicates that the upper frequency is unknown because of recording equipment limitations [6].

Impulsive calls are single or repeated short duration broadband pulses commonly produced by toothed whales and dolphins. Such calls are called clicks as labeled by human aural perception. Clicks are used for long-range echolocation including prey location [7]. Figure 1.3 displays the frequency ranges of echolocation clicks produced by toothed whales. Baleen whales do not echolocate.

Figure 1.3: Frequency range of echolocation clicks produced by toothed whales. An asterisk (*) indicates that the upper frequency is unknown because of recording equipment limitations [6].

When the interval between the successive pulses is relatively short (about 5 ms or less) the pulses can not be recognized separately in time despite their click train structure [8]. However, they are perceived as a continuous tonal sound and are called burst- pulse...
sounds which are labeled by human perception as screams, squeals, or moans [8].

The ability to detect sound varies among marine mammals. The absolute hearing threshold is used to measure the lowest sound level that can be heard by the animal in the absence of significant ambient noise [4]. The hearing ability of different species is commonly assessed by using Audiograms which are graphs relating the hearing threshold to frequency. Examples of audiograms based on six species are shown in figure 1.4. This figure shows that the Bottlenose dolphin hearing extends below 100Hz but its sensitivity is quite poor at that frequency. Due to the use of high frequencies for echolocation purposes, the hearing ability becomes better at such frequencies. The best sensitivity of these six species ranged from ~8 to 90 kHz. For example, the killer whale could detect a signal of ~30 dB re 1 µPa at a frequency of 15 kHz [4].

![Figure 1.4: Examples of marine mammal auditory thresholds [9]](image)

1.2 North Atlantic Right whales

The Right whale got its name from being the right whale to hunt by whalers as its dead body floats. According to its distribution in the world oceans, Right whales can be divided into: North Pacific Right whales, South Atlantic Right whales, and North Atlantic Right whales. The North Atlantic Right whale is the target cetacean species in this thesis. It belongs to the suborder Mysticeti (figure 1.5).
During its migration between calving and feeding grounds along the eastern coast of the United States and Canada, the North Atlantic Right whale prefers five main habitats. These include the south east US (the only known calving ground), Cape Cod Bay, Great South Channel, Bay of Fundy and Roseway Basin, see figure 1.6 [10].

Figure 1.5: Taxonomic tree of the North Atlantic Right whale (Eubalaena glacialis)
North Atlantic Right whales interact at the surface in a surface active group (SAG). It is defined as a group of male whales surrounding a focal female [11]. This focal female produces scream calls while the female calves produce warble calls [12]. Adult male Right whales produce brief intense broadband calls known as “gunshots”. These pulsive calls are aurally similar to sounds played by flipper slapping but they are more intensive with energy at high frequencies. They may be used as a productive advertisement [13].

The vocalisation rate of the Right whale is defined as the number of sounds per time unit divided by the number of animals in a group. In the literature, the effects of activity state, group size, and sexual composition of the group on the vocalisation rate was addressed by Clark [11], while Matthews [14] investigated how aggregation size, location, and day time affect such a measure.

1.3 Threats facing marine mammals

There is a range of harmful threats facing marine mammals and potentially resulting in death. This is particularly so for whales [15].
Whaling started in the 10th century. The growing demand for cetacean products such as meat, oil, blubber, and other commodities led to overexploitation of whales which in turn caused their populations to decline [15]. However, the International Whaling Commission (IWC) successfully established moratorium on commercial whaling taking effect from 1986 [16].

A range of anthropogenic activities including fishing, vessel movement in waters where marine mammals live and the use of high power underwater sonar (mostly for military purposes) have been linked with the relatively high level of cetacean mortality recorded in the last few decades.

In pelagic trawl fisheries, cetaceans may become victims of bycatch. The interaction between gill-nets and cetaceans may cause injury or even death to these animals [17]. Thus, to reduce the number of by-caught marine mammals, acoustic alarm devices have been attached to fishing nets in a hope that this will warn animals away from the nets [18]. Management strategies have been proposed to help reduce the number of collisions between whales and ships. These include controlling vessel velocity and moving commercial shipping lanes away from high risk areas [10].

Low frequency active sonar (LFAS) which has been traditionally deployed for submarine detection and tracking is seen as another threat to marine mammal life. Recently, several mass strandings of cetaceans have been temporarily and spatially linked to military sonar exercises [19]. The development of gas-bubble lesions in the internal organs of the stranded cetacean was assumed to be linked to such sonar testing [20]. It has also been suggested that exposure to high power sonar may cause whales to instinctively perform rapid panic dives in their attempt to escape from sonar insonification. This dangerous diving behaviour can result in symptoms similar to decopression sickness in humans [20].

In an attempt to solve this problem, some navies have implemented measures to establish protected areas where no sonar operation is permitted. Alternatively military vessels are required to switch off (or at least ramp-down the sonar source levels) in situations where marine mammals are detected close to such vessels [21].

The proposed mitigation measures require twenty-four hour (automated) monitoring of
such endangered species in order to be able to take measures to protect them from potential harm. This monitoring can be achieved by visual and automated observations as presented in the next sections.

1.4 Marine mammal monitoring

Traditionally, marine mammal monitoring began by observing animals on the sea surface visually; see figure 1.7. From shore, ship board or aircraft, a team of trained marine mammal observers (MMO’s) watch out for surface contacts, note their behaviour and attempt to identify them. In certain applications MMO’s use such information to estimate creature populations within a study area.

![A marine mammal observer (MMO), Image Courtesy of Jonathan Gordon](image)

Figure 1.7: A marine mammal observer (MMO), Image Courtesy of Jonathan Gordon

Traditional visual observations however suffer from a range of drawbacks including [22]:

- Short surfacing intervals of some cetaceans. Most marine mammals spend much of their time submerged.

- Traditional Marine Mammal Observers require extensive training and can only work effectively for short periods of time.

- Visibility is limited by foggy and rainy weather and a rough sea state

- Visual techniques are limited to day light operation.

- The small size of some species such as Minke whales [23] and harbour porpoise makes visual observation very difficult, particularly in rough seas.

To overcome the above drawbacks, passive acoustic monitoring (PAM) has been
proposed as a complementary solution. Listening to the acoustic signals emitted by the targeted animals is the first step toward detecting them. Unlike visual observation that depends on visualising cetaceans in their habitats, PAM relies on recording species calls i.e. cetaceans must be vocalising during the recording time to capture their calls. Visual and passive acoustic techniques can be combined to give a better measure of the abundance of species [24]. The advantages and disadvantages of the PAM technique are listed in the table below [22]:

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>♦ It can monitor all day, even in adverse weather conditions.</td>
<td>♦ Some species do not often vocalise.</td>
</tr>
<tr>
<td>♦ Species can be acoustically detected at long ranges.</td>
<td>♦ In some noisy habitats, calls can be masked by noise sources consequently this affects the ability to detect such calls.</td>
</tr>
<tr>
<td>♦ Recording can be made over long periods of time.</td>
<td>♦ Man-made noise may affect vocal behaviour. For example, vessel noise may reduce the calling rate of targeted species.</td>
</tr>
<tr>
<td>♦ It can monitor at night times</td>
<td>♦ Array limitations including ambiguities and accuracy issues.</td>
</tr>
<tr>
<td>♦ It can detect submerged species</td>
<td></td>
</tr>
</tbody>
</table>

1.5 **Acoustic sensors**

As we use microphones to record airborne sounds, the vocalisations emitted by marine mammals can be captured using a range of underwater acoustic sensors. These can include:

- Cabled hydrophones: these devices consist of bottom-mounted hydrophone arrays connected to land-based processing facilities by underwater communication cables. Historically, the US Navy’s SOSUS (SOund SURveillance System) was operated to detect enemy submarines during the Cold War [25]; see figure 1.8. Although cabled hydrophones can be permanently or semi-permanently installed to provide continuous data in the near-real time, they suffer from the high cost and
the restricted use in small shelf areas [6].

Figure 1.8: An example of cabled hydrophones (SOSUS). Image courtesy Naval Research Laboratory

- Autonomous hydrophones come in a variety of configurations: The first example was developed by the National Oceanic and atmospheric Administration’s (NOAA) Pacific Marine Environmental Laboratory (PMEL). Here the hydrophone is suspended up into the water column (figure 1.9). The second example is the acoustic recording packages (APRs) developed by the Marine Physical Laboratory of the Scripps Institution of Oceanography. Here the hydrophone is moored a few meters above the seabed, as shown in figures 1.9 and 1.10 [26].

Figure 1.9: The deployment configuration of autonomous hydrophones which are (a) PMEL (b) APRs [26]

Figure 1.10: An acoustic recording package [27]

The APRs have advantages over cabled hydrophones because of their low installation cost in distant waters and their long-term deployment [27]. The recorded data has to be
recovered from the sea floor. Due to its low sampling rate (up to 1 kHz), APRs can capture low frequency calls of baleen whales but are not suitable for recording the high frequency calls of toothed whales. To overcome such a problem, Wiggins et al. a High-frequency Acoustic Recording Package (HARP) capable of sample rates up to 200 kHz [28].

A towed hydrophone array deployed in a linear configuration behind a ship [29], [30]; as illustrated in figure 1.11.

![Acoustic data recorded by towed hydrophone array, Image Courtesy of David Mellinger](image)

Figure 1.11: Acoustic data recorded by towed hydrophone array, Image Courtesy of David Mellinger

- Sonobuoys can be deployed either from an aircraft or a ship; see figure 1.13. Each sonobuoy includes a hydrophone and a radio transmitter to send the recorded underwater sounds back to the receiving station. Although this device is relatively cheap and can be deployed close to the emitting source, it is not recommended for long-time recording due to its relatively short life span [29].

![Deployment of passive sonobuoys by air](image)

Figure 1.12: Deployment of passive sonobuoys by air
Since sonobuoys are freely drifting, the change in their position will cause an error of the arrival time on each sensor. This in turn results in inaccurate localisation of the calling species. Because bottom moored recording units are fixed, they can minimize such localization error when used [31].

In spring 2008, new real-time auto-detection buoys were deployed to listen to North Atlantic Right whales in New England waters; see figure 1.13. These listening systems send information to ships in the immediate area to warn them of the presence of whales, giving vessels a plenty of time to slow down or change their route to avoid collisions with whales [32].

1.6 Passive acoustic monitoring (PAM) structure

As mentioned above, in order to ensure their safety marine mammals have to be monitored in high risk habitats. When these creatures can not be detected by visual observation, passive acoustic monitoring has been used to perform such a task by listening to their vocalisations.

Passive acoustic monitoring can be structured into three main stages: Detection, Localisation and Classification, see figure 1.14. The detection stage decides whether the sound of interest is present within the time recording of the data. This normally involves a decision criterion such a threshold which is in turn a trade-off between false alarm and missed calls. However, the simplest form of detection is based on isolating the target call from the background noise within the recording data. In this thesis, an image processing-
based technique was used to perform such a task.

In the next stage, a few descriptive features are extracted from the detected vocalisations. These features have to capture enough information to be used in classification. Based on these features, the classification stage attempts to determine which species the vocalisations belong to. The localisation task aims to estimate the location of the vocalising species or track their movements. This is usually performed using the time difference of arrival of the call at several hydrophones. In this thesis we focus on the detection and characterisation stages of North Atlantic Right whales. The vocalisations are often masked by other noise sources or distorted due to the geometry of the acoustic environments.

1.7 Softwares used in thesis

In this thesis, a number of softwares were used. Some of them are openly available through the internet such as ISHMAEL, KRAKEN, and PAMGUARD. The only commercial software package here is PROSIM. A brief description of these tools is presented

- ISHAMEL, written by David Mellinger [33] of Oregon State University (USA), is a program for acoustic analysis. It is a software package with a variety of acoustic
detection and display functions. It contains a spectrogram viewer, acoustic localisation methods, and methods for automatic call detection, real-time sound recording, a beamformer and a log file annotation feature. The most basic operation in Ishmael is viewing a spectrogram. A spectrogram shows time on one axis (in ISHMAEL, the horizontal axis) and frequency on the other axis. In this work, ISHAMEL was used to estimate the acoustic range using the time difference of arrival (TDOA) between spatially distributed hydrophones.

- Kraken is written by Alec Duncan et al. [34] of Curtin University (Australia). This tool finds the normal modes for the model propagation environment using real arithmetic and estimates the attenuation by a perturbation technique. This method works well for layered fluid seabeds and can handle an elastic lower halfspace (i.e. a halfspace with a significant shear speed). It can’t cope with elastic intermediate layers. The normal modes only account for energy that is trapped in the waveguide and therefore this method is inaccurate at short range where the effect of untrapped energy is significant. This model was used to plot transmission loss versus range in Cape Cod Bay environment.

- PAMGUARD [35] is a recent open source development which provides a flexible, modular software framework with basic application functionality comparable to the existing PAM software (e.g. Ishmael, Rainbow Click). It was implemented in Java modules, is capable of working on multiple operating systems (e.g. Windows/Linux), and has the ability to incorporate new modules as they are developed to include additional detection, classification, localisation, and sound visualisation functionalities. Its versatile software/hardware interface enables flexibility in the configuration of underwater equipment (number of receivers, sensitivities, aperture and geometry). This tool was used to present illustrative examples of PAM techniques (energy sum detector, matched filtering, and spectrogram cross correlation).

- PROSIM [36] is a broadband propagation model. It is based on a model called ORCA. It is based on a layered normal mode approach. Only real eigenvectors are included in the calculations. The actual propagation model requires only an accurate determination of eigenvalues and mode functions for a limited number of discrete frequency components. Between such frequencies the eigenvalues and mode functions are interpolated. This tool
was used in this thesis to investigate the influence of the environmental parameters of a shallow water environment on a linear chirp.

1.8 Ambient noise in shallow water

Ambient noise contributes unwanted signals to the receiving hydrophones. In most shallow waters, the main sources of such noise include commercial shipping, industrial activities, wind noise, wave action, and conflicting biological sounds produced by other marine animals inhabiting the same area [37]. The level and frequency distribution of the ambient noise in the acoustic channel varies from location to location and from time to time [4]. Such variation affects detectability of the marine species especially when their sounds are masked by the ambient noise. This consequently influences the management measures taken to protect them. Shipping traffic is the most likely reason for the low-frequency noise increase in recent years. In deep waters, ambient noise levels increase was reported for frequencies below 100 Hz [38]. In shallow-water environments the local shipping traffic dominates the low-frequency noise [39]. An increase in low-frequency noise in Right whale habitats could potentially effect its calling behaviour. This implies that the Right whale shifts up the minimum frequency of its up call into high frequencies to compensate for the band-limited noise [40]. A recent study suggests that the North Atlantic Right whale responds to the peak frequency of noise (the frequency corresponding to the peak value in the signal spectrogram) rather than the absolute level of environment noise [41].

1.9 Dispersion in Shallow water

Shallow-water environments (<200 m deep) act as waveguides. Sound transmission is dominated by the reflections off the sea surface and the sea bottom. Such reflections cause multipath propagation of the signal. A primary effect of such multiple paths is to cause distortion on the signal received at the hydrophones. The dispersive signal can be seen as multiple modes with different time of arrivals due to the dependence of the mode’s group velocity on the acoustic frequency [37]. Also, these modes are relatively excited according to the depths of the source and receivers in the shallow-water channel. The shallow-water environment also acts as a high pass filter [42]. For instance, signals having a maximum frequency which is less than the cut-off frequency of the first
mode can not propagate in the acoustic channel.

1.10 Motivation for this work

Marine cetaceans are at risk from human activities [15]. Of particular relevance to this thesis is the Northern Right whale which is one of the most endangered whales in the world as its low birth rate does not compensate for its population decline [43]. It is therefore important to be able to detect the presence of such animals in high risk areas in order to activate mitigation measures to protect them from possible harm.

Unlike visual observation which is based on watching the target species at the surface by trained observers, Passive Acoustic Monitoring (PAM) has recently merged as a powerful tool. It is based on listening to vocalising whales and recording their sounds by hydrophones. In this thesis, the real data of North Atlantic Right whales was obtained from the International Fund for Animal Welfare (IFAW) and relates to signals recorded by hydrophones deployed in Cape Cod Bay, the Eastern Coast of the USA, 2001. To represent these sounds in the joint time and frequency domain such as “spectrograms”, a signal processing tool called Short time Fourier Transform was used.

The visual scanning of the data spectrograms indicated that the majority of the Right whale calls are dispersive with multiple modes due to the effects of the shallow water. This inspired the author to investigate the influence of the environmental parameters of the acoustic channels using normal mode modelling. Such dispersive effects influence the capability of the parameters extracted from dispersive received calls to characterise the true vocalisations. Because of the sensitivity of the weak first mode to threshold-based segmentation, we have developed an automatic detection system that is based on region-based active contours to isolate the target calls from the background.

1.11 Thesis organization

The first chapter presents an introduction to marine mammals' vocalisations and environment noise. Also, the passive acoustic monitoring system including the acoustic sensors used to record species vocalisations is reviewed.

Chapter 2 reviews a range of passive acoustic monitoring techniques developed to detect
and classify vocalisations produced by both toothed and baleen whales. Chapter 3 discusses a variety of signal processing techniques used by researchers to represent the spectral energy of marine mammal’s calls in the time-frequency plane. In chapter 4 the theory of the influence of dispersive shallow-water environment on the Right whale up calls is discussed. Such an influence is investigated and evaluated in chapter 5 using both synthesized data and real data recorded in Cape Cod Bay. In chapter 6, the theoretical aspects of the image processing technique (region-based active contour model) is presented. Chapter 7 presents the proposed automatic detection and characterization system and its application on Right whale up calls recorded in Cape Cod Bay. Finally, in chapter 8 the results are discussed and potential future work is outlined.

The thesis also has three appendixes. Appendix A defines the estimation of characteristic impedance of sediment type using the hydrophone configuration and the information obtained from dispersive up calls. Appendix B displays the bathymetry of Cape Cod Bay. Finally appendix C shows the application of the above 2D localisation method on the Right whale calls recorded in Cape Cod Bay using ISHMAEL.
Chapter 2: **Detection and classification of marine mammals’ sounds using passive acoustic monitoring systems**

It is important to be able to detect the presence of such animals in high risk areas in order to activate mitigation measures to protect them from possible harm. A range of different techniques have been employed for this task, from simple human visual observations to complex automated acoustic detection systems.

Recordings can be considered a critical challenge facing an operator either by listening to the recorded data on a headset, or by viewing real-time scrolling spectrograms [30]. In reality, low SNR recording is considered as a critical challenge facing the human’s ability to recognize a target call. Also, tracking specific calls on a computer screen is a time-consuming and tiring work for a trained observer [10]. For this reason, several automated techniques have been proposed to automatically detect and classify the marine mammal calls from the recorded data. Sometimes, both aural detection and visual scanning of the data spectrogram are used to assess the optimization of automated detectors [44].

### 2.1 Detection and classification of marine mammals

For mitigation and monitoring of marine mammals, a whole range of approaches have been developed to detect and classify vocalisations. The principle of any detector or classifier is mainly dependent on the characteristics of the creature’s call and the recording conditions (SNR). Based on the classification of cetacean calls as tonal or pulsed calls (see section 1.1), the techniques of detection and classification of these calls will be presented here.

#### 2.1.1 Tonal calls

Methods involved in detecting and classifying tonal calls will include energy detection, matched filtering, spectrogram correlation, statistical classification, dynamic time warping, and neural networks.

In its simplest form the idea of detection relies on discriminating the signal from the unwanted noise by comparing a signal parameter such as the maximum amplitude
spectrum to a pre-defined threshold to reveal the detection event. Such a simplistic detection concept was utilised to detect the West Indian Manatee [45]; see figure 2.1.

Figure 2.1: Spectral thresholding for use in detection of West Indian Manatee [45]

In energy-sum detectors, a time-varying detection function is computed by summing the signal’s energy within a certain frequency band over the signal’s spectrogram blocks. An empirically selected threshold is applied to that function to indicate the presence of species [10], [33] as illustrated in figure 2.2.

Figure 2.2: Detection of Humpback songs by an energy-sum detector in the frequency range (100-300Hz) indicated by the red dotted lines. (a) Recording spectrogram; (b) Recording waveform; (c) Detection function. Note that the results shown in this figure was produced by PAMGUARD software [35].
Note that the existence of the Humpback songs within the spectrogram (figure 2.2(a)) was revealed once the peaks of the detection function exceed the threshold indicated by the red horizontal line displayed in figure 2.2(c).

Despite being simplistic, the technique of energy-based detection can be used to scan recorded data when the target call is unknown [33].

Another characteristic of the signal used for detection purposes is the entropy of its normalized power spectral density (PSD) that can be treated as a probability distribution. It is computed as the product of the PSD with the logarithm of the PSD. When the instantaneous entropy computed for all successive blocks of the signal surpassed a threshold, a call is detected [46].

Some techniques developed for completely different applications have been applied to the detection of cetacean calls. For example, matched filtering is a common technique used in radar, sonar and communications [47]. It is based on cross-correlating a kernel with recorded data. Peaks of the output function that exceed a specific threshold indicate the detection events.

Different approaches have been proposed to design the matched filtering kernel. Mellinger et al. [22] developed a time-series kernel using Principle Component Analysis (PCA) to recognize the end notes of Bowhead whale songs; see figure 2.3(a). The PCA technique reduces dimensionality while retaining most of the variation in the data. Another time-domain kernel was mathematically synthesized using the average parameters of the down-swept harmonic of the Blue whale calls in the northeast Pacific Ocean for detection purposes [48]; see figure 2.3(b).
Figure 2.3: Diagram of matched filtering techniques used to detect (a) Bowhead whales and (b) Blue whales

Although the two techniques shown in figure 2.3 rely on the same principal of cross correlation, they have different designs for the matched filtering kernel due to the different characterization of each call. Note that the Bowhead song kernel was built using a dimension reduction method while the Blue whale kernel was created using a mathematical technique based on the call’s parameters.

The results of using matched filtering technique to detect northeast Pacific Blue whales are shown in figure 2.4. Note that a time-series template of the call was used as a kernel.
Figure 2.4: Detection of the calls of northeast Pacific blue whales using matched filtering technique. (a) Recording spectrogram; (b) Detection function. Note that the results of this figure were produced by PAMGUARD software [35].

The cross-correlation function displayed in figure 2.4(b) measures the similarity between a template call of the Blue whale and the recording data in the time domain. The high degree of similarity was expressed by the peaks exceeding a given threshold (the red horizontal line) to indicate the detection events.

The technique of matched filtering is optimal for a known call in white Gaussian noise. In reality marine mammal vocalisations vary among individuals and also the ocean noise does not have a flat spectrum. Therefore the matched filtering technique is thus a non-optimal solution [22].

Rather than using the call’s parameters to build a one-dimensional kernel in the time domain [48]. Such parameters were deployed to create a two-dimensional kernel in the time-frequency domain [49]; see figure 2.5(b).

This kernel was built based on the second derivative of a Gaussian distribution with an important property that positive and negative regions sum to zero; see figure 2.5(a). Such
a property causes the cross-correlation of the kernel with a uniform noise background to be zero. The results of using spectrogram correlation to detect the “A calls” of the Atlantic Blue whale are shown in figure 2.6.

![Figure 2.5: Design of two-dimensional kernel (a) Second-derivative of a Gaussian distribution (b) 2D Spectrogram correlation kernel of up-swept call (80-200 Hz)](image)

It is important to notice that cross-correlating the kernel with the spectrogram of the recorded data is performed for each time frame i.e. the detection function is a time series with peaks indicating the presence of the calls; see figure 2.6 (b). This technique was deployed to recognize the calls of Blue whales, Finback whales, Bowhead whales, and Right whales [22],[50],[44].

![Figure 2.6: Detection of Blue whale calls using spectrogram correlation. (a) Recording spectrogram; (b) Detection function. Note that the results of this figure was produced by PAMGUARD software [35].](image)
Statistical methods have been widely used in classification problems. Gillespie [51] separated Right whale calls from the other calls using a discriminant analysis function utilizing parameters obtained from the smoothed spectrogram image. In [30], the same classification function has also been used to classify delphinid species using parameters extracted from the fundamental frequency contour. Clark [52] used Principal Component Analysis to classify Southern Right whale sounds. La Cour et al. [53] recognized non-Gaussian Right whale calls from ambient noise by using Independent Component Analysis (ICA) and applied test statistics to announce detection events for Right whales calls. The maximum likelihood technique was used [54] to estimate the parameter of each chirp representing a segment of the broken frequency contour of Right whale calls.

For certain species’ vocalisations that have the same shape of frequency contour but different duration, the Dynamic Time Warping (DTW) technique is used to align the contours for comparison. This technique is commonly used in speech recognition literature. It performs the contour alignment using nonuniform time dilation by minimizing the total square difference. Previous authors have used this technique to classify the pulsed calls of the Killer whale [55] and to assess the similarity of the frequency contour of Bottlenose dolphin’s signature whistles [56].

For numerous amounts of recorded data with high parameter variation, neural network systems have been adopted to detect and classify species calls. A Back-propagation neural network was used to recognize the calls of Right whales, Bowhead whales [50], [57]. The same type of network was also used to assess the similarity between pulsed calls emitted by killer whales [58]. Murray et al. [59] used an unsupervised, self-organizing neural networks to categorize the repertoire of false killer whale vocalisations.

### 2.1.2 Pulsed calls

A variety of detectors have been proposed to detect the time location of sperm whale clicks in the data recorded in their habits. Based on the click’s features such as their broadband spectrum, high source level (energy measure) and nonstationary nature, detectors can indicate the detection events of the calling species.
The broadband nature of the spectrum of sperm whale clicks was used by [60] to form a detector. When the broadband signal spans more than a certain percentage of the frequency band it is considered as a click.

Based on the signal’s energy, the Rainbow click detector was developed by [61] to detect Sperm whale clicks; see figure 2.7(a). After the rectified input signal is filtered by a high-pass filter with a frequency response given in figure 2.7(b), two thresholds are proposed. The first is used to determine the onset and end of the click signal and the second is applied to the cumulated energy of the signal in a certain frequency range to reveal the detection of clicks; see figure 2.7.

Another energy detector was developed by [7] and [62] to detect sperm whale clicks which are commonly divided into regular clicks and creak clicks. The main difference between these two types of clicks is that the regular clicks have longer Inter-Click Intervals (the interval between two successive clicks (ICI)) and higher source level than...
It is based on Teager-Kaiser energy operator $\psi$ that requires only three samples of the signal $x(n)$ to be computed: $\psi[x(n)] = x(n)^2 - x(n+1)x(n-1)$; see the frequency response of the TK energy filter in figure 2.8(b).

In this detector, a short analysis window containing a few clicks slides along the input signal, see figure 2.8(a). The size of the window is dependent on the previous estimated Inter-Click interval. Since the source level of the creak clicks is much lower than that of regular clicks, a certain energy threshold is applied to recognize them. The application of the filter matching with the impulse response decaying with time enables the detection of the beginning of the regular click regardless of whether the first pulse is weak or strong.

![Block diagram of the proposed detector](image)

![Magnitude of the frequency response](image)

Figure 2.8: (a) Block diagram of the proposed detector [62]; (b) Magnitude of the frequency response of the TK-energy operator filter

To avoid the dependence of the energy detectors mentioned above on signal’s source level (amplitude spectrum), Lopatka et al. [63] relied on the signal’s statistics to develop their detector. This was performed by using the sensitivity of the time varying Schur
coefficients to the non-stationary nature of transient signals as an indicator of the presence of clicks. Also, an energy-independent detector which is based on the slope of the phase spectrum to reflect the positions of the clicks was proposed [64].

Back-propagation neural networks were used to classify Beaked whale’s clicks [65]. Also, Support Vector Machine (SVM) was used for classification of Beaked whale clicks [66] and Humpback whale song units [67].

More review material regards the features extracted from the species’ sound and hence used for classification is presented in chapter 7.

2.2 Chapter summary

This chapter has reviewed a range of automated passive acoustic monitoring techniques developed to detect and classify marine mammal vocalisations. The main goal of such techniques is for mitigation and monitoring of marine mammals in high risk areas.

Since any suggested detector or classifier is mainly dependent on the characteristics of the target call, we grouped the reviewed techniques into two groups: tonal calls and pulsed calls. The results of the application of some detection techniques on examples of cetacean’s calls have been presented using PAMGUARD.

Right whales up calls are tonal calls. The application of the reviewed PAM techniques used to detect and classify tonal calls on dispersive Right whale up calls is limited. In matched filtering approach in addition to the variation in emitted source calls, the shallow water channel contributes to a variation in the received calls at the hydrophones i.e. the existence of multiple modes. This may affect the correlation between the matched filtering kernel and the recording data.

Due to the intermodal dispersion in dispersive shallo-water environments, the likely modes arrive at different time at hydrophones. In the spectrogram correlation technique, if for a single detection a kernel is constructed using only the first mode parameters, the cross correlation between this kernel and the multiple mode signal results in multiple peak in the detection function for a given threshold. This may cause a false detection
event.

The Dynamic Time Warping technique was used to measure the similarity between the frequency contours of dolphin whistles. For dispersive Right whale up calls, the frequency contour extracted from the received call may be different from that of the source call. This is because dependence of mode excitation on the source depth results in a broken frequency contour. See figure 7.18 for illustration.

Neural network system (NSS) is used to assess the variation in numerous amounts of recorded data which is necessary for training the network. Our Cape Cod Bay data only contains 120 calls which is not enough to adopt the NSS.

In the first stage of any detection or classification process, an appropriate signal processing tool has to be chosen to represent the call of interest. A range of such tools will now be reviewed. These include short time Fourier transform, Wavelet transform, Wigner-Ville distribution and Hilbert-Huang transform. In the next chapter, we will introduce a variety of such tools with illustrative examples. We will select the appropriate tool to represent the North Atlantic Right whale calls.
Chapter 3: **Signal processing techniques**

Waveform analysis based on signal processing tools is a necessary step in order to extract information from a signal:

- in passive acoustic monitoring (detection, localization, and classification)
- to correlate recorded animal’s vocalisations with their behavioural patterns
- to investigate the influence of the environment on received sounds in shallow water

In this chapter a variety of signal analysis techniques and a comparison between them were illustrated. These techniques include the Fourier transform, short-time Fourier transform, Wavelet transform, Wigner-Ville distribution, and Hilbert-Huang transform. Two test waveforms including linear and nonlinear (quadratic) chirp signals were used to evaluate such techniques. The reason for choosing such signals is that Right whales produce linear upswept calls. These calls may become nonlinear (at high order modes) when they are received in a dispersive shallow-water environment. These test waveforms have a frequency range of 50-200 Hz and duration of one second. The test waveforms are displayed in figure 3.1.

A Time waveform is the most common way of describing a sound. It expresses the change in the signal’s amplitude over time. In order to obtain simpler patterns of the signal, transforming to the frequency domain can be performed by using the Fourier transform.

### 3.1 Fourier Transform

In the early 19\textsuperscript{th} century, Fourier explored the periodicity of a given signal $s[n]$ by measuring the similarity between the signal and the basis function (complex sinusoidal functions $e^{-j\omega n}$) \[68\]. Fourier transform $S[\omega]$ indicates the spectral content of the signal $s[n]$ at frequency $\omega$. It is defined by \[69\] as:
\[ S[\omega] = \sum_{n=-\infty}^{\infty} s[n] e^{-j\omega n} \] (3.1)

To give an example of the basis function, a 10-Hz complex sinusoidal function is considered here where the waveform and frequency response of its real part are displayed in figure 3.2. Note that the basis function spreads over the entire time domain (figure 3.2(a)) while in the frequency domain the spectrum is concentrated at 10Hz (figure 3.2(b)). Rather than a single spectral line at 10 Hz the finite time duration of the sampled signal results in a frequency spread.

Figure 3.1: Two test waveforms and the corresponding sketches of the frequency sweep type used to evaluate the signal processing tools. Signal intensity is linear.

(a) Test waveform 1 of a linear chirp

(b) Test waveform 2 of a nonlinear quadratic chirp
Applying Fourier transform to the test waveform results in the magnitude spectrum $|S(\omega)|$ against frequency as illustrated in figure 3.3. This result does not indicate how frequency information varies with time since the basis function is not concentrated in time; see figure 3.2(a).

According to the principle of conservation of energy in the time and frequency domains (known as Parseval’s formula), energy spectrum density $|S(\omega)|^2$ is another common tool
used for Fourier analysis to describe the distribution of the signal’s energy in the
frequency domain [69].

In fact, the majority of marine mammal sounds have time-varying spectra. Therefore
using the Fourier transform to represent such calls is insufficient. Thus there is a strong
motivation to determine an optimal way to represent the distribution of the signal’s
energy in both local time and frequency domains simultaneously. A range of joint time-
frequency tools will be presented in the next sections. These include short-time Fourier
transform, Wavelet transform, Wigner-Ville distribution, and Hilbert-Huang transform.

3.2 Short-time Fourier transform (Windowed Fourier transform)

To further develop the Fourier transform and to link the signal’s spectral information to
the time domain, the basis function included in (3.1) has to be localised in both time and
frequency domains simultaneously. This can be achieved by multiplying $e^{-j\omega n}$ by a short-
duration and time-shifted window $w[n-m]$. The resulting transform is called short-time
Fourier transform which is defined by [68] as:

$$
STFT[n, \omega] = \sum_{m=-\infty}^{\infty} s[m]w[n-m]e^{-j\omega n} 
$$

(3.2)

Here two different points of view as to how to interpret the definition of the STFT are
presented. Firstly, the $STFT[n, \omega]$ is basically implemented by applying the Fourier
transform to successively overlapping windowed segments of the signal $s[m]w[n-m]$. The second interpretation historically relies on filter bank analysis where the
$STFT[n, \omega_k]$ is a function of $n$ at a constant frequency $\omega_k$. Then, equation 3.2 can be
rewritten as [70]:

$$
STFT[n, \omega_k] = \sum_{m=-\infty}^{\infty} s[m]e^{-j\omega_k m}w[n-m] = [s[n]e^{-j\omega_k n}] * w[n]
$$

(3.3)

The convolution of $s[n]e^{-j\omega_k n}$ with $w[n]$ in the time domain is equivalent to applying low pass filtering to the spectrum of $s[n]$ after it was shifted by $e^{-j\omega_k n}$ down towards DC; as
shown in figure 3.4(a).

![Diagram](image)

(a) Realization of equation (3.3)

(b) Realization of equation (3.4)

Figure 3.4: Realization the STFT. Note that the window function acts as a L.P.F in (a) or B.P.F in (b)

By making the change of the variable \( l = n - m \) in (3.3) [71]:

\[
STFT[n, \omega_k] = \sum_{l=-\infty}^{\infty} s[n-l]w[l]e^{-j\omega_k(n-l)} = e^{-j\omega_kn} \sum_{l=-\infty}^{\infty} s[n-l]\{w(l)e^{j\omega k}\}
\]

\[
= e^{-j\omega_kn} [s[n] * [w[n]e^{j\omega_kn}]]
\]

An alternative understanding of the above equation implies that the STFT is interpreted as a convolution of a band pass filter (formed by shifting the frequency response of the window \( w[n] \) by \( \omega_k \)) with the signal \( s[n] \) and then followed by a demodulation operation. This is illustrated in figure 3.4(b).

The duration and the bandwidth of the window determine the time and frequency resolutions of the STFT respectively. According to the uncertainty principle, the bandwidth of the window is inversely related to its duration. This implies that achieving fine frequency resolution requires a long-duration window of small bandwidth but this unfortunately results in poor time resolution and vice versa. The trade off between time and frequency resolutions can be summarized as:
Short-duration window \[\Rightarrow\] Good time resolution
Poor frequency resolution

Long-duration window \[\Rightarrow\] Good frequency resolution
Poor time resolution

The type of the window is another critical parameter for the STFT analysis. Various window functions are listed in table 3.1 and displayed in figure 3.5.

Table 3.1: Important frequency-domain characteristics of some window functions: window length M. The main lobe width is measured to the first zero of the frequency response of the window [69]

<table>
<thead>
<tr>
<th>Type of window</th>
<th>Approximate transition width of main lobe</th>
<th>Peak side lobe (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rectangular</td>
<td>$4\pi / M$</td>
<td>-13</td>
</tr>
<tr>
<td>Bartlett</td>
<td>$8\pi / M$</td>
<td>-27</td>
</tr>
<tr>
<td>Hanning</td>
<td>$8\pi / M$</td>
<td>-32</td>
</tr>
<tr>
<td>Hamming</td>
<td>$8\pi / M$</td>
<td>-43</td>
</tr>
<tr>
<td>Blackman</td>
<td>$12\pi / M$</td>
<td>-58</td>
</tr>
</tbody>
</table>

The frequency characteristics of these windows indicate that a trade-off exists between the main lobe bandwidth and side lobe level. For example, bell-shaped windows such as Hanning, Hamming, and Blackman have more suppressed side lobes but wider main lobe than rectangular and Bartlett windows [72], [69].

On the other hand, since each frame of the signal can be considered as a time-limited signal zero-padding such finite-duration signals will interpolate and smooth the signal’s spectral response but will not affect the frequency resolution [73]. $|STFT(t, \omega)|^2$ (the so-called spectrogram) can be displayed as a two-dimensional image (columns and rows represents frequency and time domains respectively). The colour bar indicates the pixels intensity which corresponds to the spectral content of the signal at the coordinates $(t, \omega)$.
Figure 3.5: A range of window functions, Window length M=128.

The spectrograms of the test waveform 1 and 2 are calculated using the parameters listed in table 3.2 and displayed in figure 3.6.

Table 3.2: The values of parameters used to produce the spectrogram displayed in figures 3.6

<table>
<thead>
<tr>
<th>Frame length</th>
<th>Sampling rate</th>
<th>FFT size</th>
<th>Overlap</th>
<th>Window type</th>
</tr>
</thead>
<tbody>
<tr>
<td>128 samples</td>
<td>1 kHz</td>
<td>512 samples</td>
<td>93.75 %</td>
<td>Hamming</td>
</tr>
</tbody>
</table>

35
The spectrogram’s bandwidth is computed for Hamming window as [69]:

$$\Delta f = \frac{8\pi}{\text{frame size}} \cdot \frac{1}{2\pi} f_s = \frac{8\pi}{128} \cdot \frac{1}{2\pi} \cdot 1000 = 31.25 \text{ Hz}.$$  

(a) STFT of test waveform 1  
(b) STFT of test waveform 2  
Figure 3.6: Spectrogram of test waveforms 1 and 2. Note that spectrograms are normalized (Linear). Fs = 1 kHz

### 3.3 Time-scale representation (Wavelet transform)

Based on the fact that scaling in the time domain leads to inverse scaling in the frequency domain (equation (3.5)), a signal’s centre frequency can be adjusted by scaling the time variable.

$$s(t/a) \xrightarrow{\text{FT}} aS(a\omega)$$  \hspace{1cm} (3.5)

If the basis function in the Fourier Transform or the Windowed Fourier Transform is replaced by a scaled and time-shifted basis function $\psi(t-b)/a$, the result is known as the Wavelet transform of the signal. This transform depicts the signal’s spectral content in the time-scale representation and is defined by [68] as:

$$CWT(a,b) = \frac{1}{|a|} \int s(t) \psi(t-b)/a dt$$  \hspace{1cm} (3.6)

where: $s(t)$ = the signal; $\psi(t)$ = the mother wavelet; $a$ = the scaling factor; and $b$ = the
In order to investigate the influence of the scaling factor on the duration, centre frequency and the bandwidth of the mother wavelet, two scaling factors ($a=1$ and $a=0.1$) were used with a complex wavelet namely Morlet which is defined by [74] as:

$$\psi(t) = e^{-t^2/2} e^{j\omega_0 t}$$ \hfill (3.7)

In practice, the angular frequency $\omega_0$ is commonly set to 5.

The waveform and the frequency response of the real component of the Morlet mother wavelet for two scaling factors are produced and displayed in figures 3.7 and 3.8.
The results of figures 3.7 and 3.8 indicates that when the scaling factor is smaller \((a=0.1)\) the wavelet duration becomes shorter while its centre frequency and bandwidth becomes higher and wider respectively. It is worthwhile to note that the ratio between the bandwidth and the centre frequency of the frequency response of the wavelet is constant with the scaling factor. The influence of the scaling factor on the time and frequency resolutions is summarized in table 3.3.

<table>
<thead>
<tr>
<th>Scaling factor</th>
<th>Centre frequency</th>
<th>Duration</th>
<th>Bandwidth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smaller</td>
<td>Higher</td>
<td>Shorter</td>
<td>Wider</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(fine time resolution)</td>
<td>(poor frequency resolution)</td>
</tr>
<tr>
<td>Larger</td>
<td>Lower</td>
<td>Longer</td>
<td>Narrower</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(poor time resolution)</td>
<td>(fine frequency resolution)</td>
</tr>
</tbody>
</table>

To compute the continuous Wavelet transform of a given signal, the similarity between the signal and the scaled mother wavelet for each time shift is measured to depict the signal’s energy at a point of \((a,b)\) in the time-scale plane. To represent the energy in the time-frequency plane, the scaling factor values corresponding to the pseudo frequencies
\( f_a \) lying within the frequency range \([0, \frac{f_c}{2}]\) can be derived as [75]:

\[
a = \frac{f_c}{f_a} f_s
\]  

(3.8)

where \( f_c \) is the centre frequency that maximizes the modulus of the Fourier transform of the wavelet.

The Wavelet transform of the test waveforms 1 and 2 is computed using the Morlet mother wavelet and illustrated in figure 3.9 where the time and frequency resolutions vary with frequency as mentioned in table 3.3.

![Wavelet transform with Morlet mother wavelet. Fs=1 kHz](image)

Figure 3.9: Wavelet transform with Morlet mother wavelet. Fs=1 kHz

As the FT and STFT, and WT use basis functions to analyze signals, we listed a summary of their characteristics in table 3.4.
Table 3.4: Summary of characteristics of FT, STFT, and WT

<table>
<thead>
<tr>
<th>Basis Function</th>
<th>Fourier Transform</th>
<th>Short-time Fourier Transform</th>
<th>Wavelet transform</th>
</tr>
</thead>
<tbody>
<tr>
<td>$e^{-j\omega t}$</td>
<td>$w(\tau - t)e^{-j\omega \tau}$</td>
<td>$\psi\left(\frac{t-a}{a}\right)$</td>
<td></td>
</tr>
</tbody>
</table>

- Uniform but only in frequency domain
- Depends on the number of FFT samples
- Window size dependent
- Uniform in the entire time-frequency plane
- Scaling factor dependent
- Vary with frequency

Both the window function (in STFT) and mother wavelet (in WT) obey the uncertainty principle which puts a limitation on such techniques with applications that require fine time and frequency resolutions. To avoid the uncertainty principle, other techniques including Wigner-Ville distribution and Hilbert-Huang Transform will be presented next.

### 3.4 Wigner-Ville Distribution

The energy spectrum density $|S(\omega)|^2$ (see section 3.2) can also be given as the Fourier transform of the auto-correlation function $R_{ss}(\tau)$ of the signal.

$$|S(\omega)|^2 = \int_{-\infty}^{\infty} R_{ss}(\tau) e^{-j\omega \tau} d\tau \quad (3.9)$$

Where $R_{ss}(\tau)$ is given as:

$$R_{ss}(\tau) = \int s(t)s(t-\tau)dt \quad (3.10)$$

By making the auto-correlation function time-dependent, the Wigner-Ville distribution is derived to depict the signal’s energy distribution in the time-frequency domain as given
The Wigner-Ville Distribution of the test waveforms 1 and 2 was computed using a toolbox developed by [76] and is displayed in figure 3.10.

![Wigner-Ville Distribution](image)

(a) WVD of test waveform 1
(b) WVD of test waveform 2

Figure 3.10: Wigner-Ville distribution with NFFT=1024 points. Note the existence of cross-term interference between components in (b). Note that spectrograms are normalized (Linear). Fs=1 kHz

Although WVD offers better time and frequency resolutions than STFT and WT, it suffers from the so-called cross-term interference that exists with multiple-component signals such as test waveform 2. To illustrate such interference mathematically, consider a two-component signal \( s(t) = s_1(t) + s_2(t) \). The WVD is given by [68]:

\[
WVD_s(t, \omega) = WVD_{s_1}(t, \omega) + WVD_{s_2}(t, \omega) + 2 \text{Re}\{WVD_{s_1s_2}(t, \omega)\}
\]  

(3.12)

The term \( 2 \text{Re}\{WVD_{s_1s_2}(t, \omega)\} \) indicates the cross-term interference.

To obtain an insight into the concept of cross-term interference resulting from applying WVD to a multiple component signal, we will consider here a signal \( s(t) \) consisting of two frequency modulated Gaussian signals concentrated at \( t_1 = -0.25s \) and \( t_2 = +0.25s \) in the time domain and at \( \omega_1 = 2\pi(250)\text{rad.s}^{-1} \) and \( \omega_2 = 2\pi(50)\text{rad.s}^{-1} \) in the frequency domain.
domain as:

\[ s(t) = \left( \frac{a}{\pi} \right)^{\frac{1}{2}} \exp \left\{ -a(t - t_1)^2 + j\omega_1 t \right\} + \left( \frac{a}{\pi} \right)^{\frac{1}{2}} \exp \left\{ -a(t - t_2)^2 + j\omega_2 t \right\}. \]

According to (3.11), the WVD of \( s(t) \) is computed as:

\[
WVD_s(t, \omega) = 2\exp\left\{-a(t-t_1)^2 - \frac{1}{a}(\omega-\omega_1)^2\right\} + 2\exp\left\{-a(t-t_2)^2 - \frac{1}{a}(\omega-\omega_2)^2\right\} + \ldots
\]

\[
+ 4\exp\left\{-a(t-t_1)^2 - \frac{1}{a}(\omega-\omega_1)^2\right\}\cos\left[(\omega-\omega_1)t_d + \omega_d(t-t_0) + \omega_d t_{\mu}\right]
\]

where: \( \omega_\mu = 2\pi \left( \frac{50 + 250}{2} \right) = 2\pi(150) \text{rad.s}^{-1} \)

\( \omega_d = 2\pi(250 - 50) = 2\pi(200) \text{rad.s}^{-1} \)

\( t_d = t_1 - t_2 = 0.5s \)

\( t_{\mu} = \frac{t_1 + t_2}{2} = 0s \)

The above resulting equation describes the WVD of \( s(t) \) as the sum of desired auto-terms centered at \( \omega_1 \) and \( \omega_2 \) and an unwanted cross-term centered at the midway (in time and frequency) between the auto-terms. It is important to note that the cross-term oscillates at 200 Hz which represents the difference between the auto-term frequencies. The realization of this result is illustrated in figure 3.11.
To reduce the cross-term interference, smoothing functions in the time and frequency domains can be used but at the cost of losing useful details such as the high frequency and time resolutions. The resulting transform is called the Smoothed Pseudo Wigner-Ville Distribution (SPWVD) (also called Reduced Interference Distribution) which is defined by [77] as:

\[
SPWVD(t, \omega) = g(t)^* \left[ \int_{-\infty}^{\infty} h(\tau) [s(t + \frac{\tau}{2})^* s(t - \frac{\tau}{2})] e^{-j\omega\tau} d\tau \right] \tag{3.13}
\]

where \( g(t) \) and \( h(\tau) \) are the time smoothing and frequency smoothing windows respectively. Two hamming windows were used to reduce the cross-term interference that appears in figure 3.10(b) and the result is displayed in figure 3.12.
3.5 Hilbert-Huang Transform (HHT)

Unlike STFT and WT, the HHT avoids the uncertainty principle as it does not deal with basis functions. It relies on an algorithm rather than analytical formulation [78]. Conceptually, the application of the Hilbert transform to a signal is necessary to make it analytic. The growth of the signal’s frequency content over time (the instantaneous frequency) is given by the rate of change of instantaneous phase. The Hilbert transform \( v(t) \) of a signal \( u(t) \) is defined by [79] as the convolution of \( u(t) \) with the function \( 1/\pi \):

\[
v(t) = HT(u(t)) = \frac{1}{\pi} P \int_{-\infty}^{\infty} \frac{u(t')}{t'-t} dt'
\]

(3.14)

P indicates the Cauchy Principle Value Integral

The analytic signal of a real signal \( u(t) \) is then given as:

\[
A(t) = u(t) + jv(t) = a(t)e^{j\phi(t)}
\]

(3.15)

Where the instantaneous amplitude and frequency are given respectively:

\[
a(t) = \sqrt{u(t)^2 + v(t)^2}
\]
\[ \omega(t) = \frac{d\varphi(t)}{dt} \quad (3.16) \]

where \( \varphi(t) = \arctan\left(\frac{v(t)}{u(t)}\right) \)

Data are not eligible for having meaningful instantaneous frequency through the Hilbert Transform unless they satisfy the following two conditions [80]:

1. they have to be mono-component functions with narrow bandwidth [81]. According to probability theory, this requires that the number of extrema and zero crossings must equal or differ at most by one

2. the local average of the minimum and maximum envelopes has to be zero in order to avoid meaningless negative frequencies resulting from asymmetric wave forms.

The functions that meet the two above requirements are called Intrinsic Mode Functions (IMFs). These functions can be extracted from the original data by using the empirical mode decomposition (EMD) method which is based on the sifting process. In an attempt to illustrate this method, an envelope fluctuation was created by adding white noise with different SNR to the test waveform 1. Due to its flat spectrum, white noise affects the signal’s frequency content greater than its envelope. For this reason we were unable to decompose data into reasonable IMFs even with high SNR. A low frequency single tone of 15 Hz was subsequently added to the test waveform 1 in order to create a desired envelope fluctuation.

The algorithm of sifting process adopted to decompose the signal into IMFs is described thus:

1. find the cubic splines passing through the local extrema \( e^+(t) \) and minima \( e^-(t) \) of the signal \( s(t) \); see figure 3.13

2. compute the local average envelope \( m(t) = \frac{[e^+(t)+e^-(t)]}{2} \)
3. check if \( z_i(t) = s(t) - m(t) \) satisfies the two conditions of IMF. If “yes” set \( y_1(t) = z_i(t) \), otherwise repeat the loop on \( z_i(t) \) till the fist IMF is obtained

4. Subtract IMF1 from the signal by \( r_1 = s(t) - \text{IMF}_1 \) and repeat the steps (1-3) on the residue \( r_1 \) till the second IMF is extracted then subtract every new IMF from the previous residue till there is no more IMF can be extracted. The resulting IMFs from the original test data are shown in figure 3.14.

![Figure 3.13: The sifting process. \( e^+(t) \) and \( e^-(t) \) represents the upper and the lower envelopes respectively. \( m(t) \) is the mean envelope.](image-url)

Figure 3.13: The sifting process. \( e^+(t) \) and \( e^-(t) \) represents the upper and the lower envelopes respectively. \( m(t) \) is the mean envelope.
Figure 3.14: The resulting EMD components from the sum of test waveform 1 and a 15 Hz single tone. Note that the sifting process extracts the higher components repeatedly. For instance, IMF1 contains higher frequency components than IMF2.

Since the sifting process removes the large-scale details of the signal repeatedly, it can also be considered as low-pass filtering process [82]. The resulting time-frequency representations of the IMFs are illustrated in figure 3.15(a) where the first IMF is represented as a linear chirp while the horizontal line at 0.015fs indicates the single tone.

Figure 3.15: Hilbert-Huang transform. Note that the curve of the instantaneous frequency of the first IMF is enhanced to make the image more illustrative.
Figure 3.16: The time-frequency representation of the sum of the test waveform 2 and a single tone of 125 Hz. Note that the empirical mode decomposition fails to separate these signals. The curves of the IF of the first and second IMFs displayed in (b) were smoothed with an average-moving window of 0.075 s in width. Spectrogram intensity is linear. Fs= 1kHz

According to the results of his work, Adam [83] found that the individual components of multiple-component signals can be separated by using the first IMFs ordered from higher frequencies to lower frequencies. This could not be achieved when the components interfere in the same frequency range. For example, if a single tone of 125 Hz is added to the test waveform 2, due to the sifting process the first IMF will pick up the highest frequencies including parts of both the single tone and the test waveform 2 as illustrated in figure 3.16.

3.6 Time-frequency representations and dispersive linear chirp signals

Due to the dispersive effects of shallow-water environments, the Right whale up calls received at a hydrophone become dispersive. The primary results of such an effect can be seen as multiple modal arrivals. This thus considers the dispersive Right whale call as a multiple-component signal. In order to investigate the most appropriate signal processing tool to represent such dispersive signals, the results obtained from each time-frequency representation including STFT, WVD, SPWVD, WT, and HHT are displayed in figure 3.17.

The test signal in this case is a simulated linear chirp (created using the PROSIM normal
mode simulation model) received in 30 m deep shallow water with a gravelly sand sediment. The range was taken as 10 km. The STFT was computed using the parameters: sampling rate 819.2 Hz, frame size 128 samples, (zero-padded to an FFT size of 512 samples), filter bandwidth 25.6 Hz, frame increment 6.5 ms, and a Hamming window. The results of figure 3.17(a) show that the first and second modes are unresolved at high frequencies.
Although WVD offers improved time and frequency resolutions, it suffers from the existence of cross-term interference between arrival modes (figure 3.17(b)). Such
interference was reduced using SPWVD but unfortunately at the cost of loosing high temporal and spectral resolutions (figure 3.17(c)). The Morlet wavelet transform was not a useful technique as the results of the analysis suffer from poor frequency resolution at higher frequencies (figure 3.17(d)). Finally HTT (see figure 3.17(e)) shows a noisy frequency contour with unclear excitation of the modes.

When adding a white Gaussian noise to the synthesized dispersive chirp used in figure 3.17, the application of the signal processing tools on such a signal is illustrated for two different signal to noise ratios: SNR=10dB (figure 3.18) and SNR=0 dB (figure 3.19).

Also, an example of a dispersive Right whale up call from Cape Cod Bay data is used to evaluate the performance of the above techniques as shown in figure 3.20.
Figure 3.18: Time-frequency representation of the dispersive up-swept chirp buried in white noise (SNR=10 dB):
(a) STFT; (b) WVD; (c) SPWVD; (d) WT; and (e) HHT. Spectrograms are normalized (dB). Fs=819.2 Hz
Figure 3.19: Time-frequency representation of the dispersive up-swept chirp buried in white noise (SNR=0 dB):
(a) STFT; (b) WVD; (c) SPWVD; (d) WT; and (e) HHT. Spectrograms are normalized (dB). Fs=819.2 Hz
From figure 3.20, we can see that of all the given tools only the STFT and SPWVD give a reasonable representation of the synthesized and real data in the time-frequency plane. It
should be added here that the SPWVD could remove the cross-term interference between modes at lower frequencies but not at high frequencies (when modes get close to each other). Also, it suffers from high computational cost.

In addition to the ordinary Fourier transform and its application in time-frequency representation, through the short-time Fourier transform (STFT), there is another generalized transform known as the \( a \)-th order Fractional Fourier transform \( FrFT \) which has been developed and widely used in optics [84].

In order to express the FrFT in terms of signal processing notions, [85] described the FrFT formula in a set of steps as follows:

\[
FrFT \ast \{x(t)\} = \frac{\exp[-j\left(\frac{\pi}{4} \text{ sgn} \varphi - \frac{1}{2} \varphi\right)]}{(2\pi |\sin \varphi|)^{1/2}} \exp\left(\frac{1}{2} j^{2} \cot \varphi \right)
\]

Where the transform order \( 0 < a < 1 \). For \( a=0 \) and \( a=1 \) the results will refer to the identity transform and the ordinary Fourier transform respectively.

The multiplication by chirps with a rate of \( \cot \varphi \) in the time and frequency domain causes the original time-frequency axes to rotate by an angle \( \varphi \) which is a function of the transform order as: \( \varphi = a \frac{\pi}{2} \). In brief words, the optimal transform order required to achieve maximum response is obtained when the rotation angle of the time-frequency axes matches the rate of the linear chirp signal. To extend FrFT applications to time-frequency representation, rather than applying the ordinary Fourier transform to each windowed frame in the STFT, FrFT was used to obtain the short-time fractional Fourier
transform (STFrFT) [86]. An investigation into the applicability of the STFrFT for representing dispersive linear chirps in the time-frequency plane was based on private correspondence with Dr. Chris Capus at Heriot-Watt University, Edinburgh, UK. The author suggested that the applicability of his technique to such signals is limited due to several factors:

♦ The nonlinearity of higher arrival modes

♦ The first and second modes are so close to each other at higher frequencies so that the ability of resolving them is limited.

3.7 Chapter summary

In this chapter, a range of time-frequency representation tools has been investigated to select the most optimal technique to represent Right whale up-calls. These tools include short time Fourier transform, Wavelet transform, Wigner-Ville distribution and Hilbert-Huang transform. The typical Right whale’s up-call can be represented as a linear narrow-band FM signal which is a sine wave modulated by a linear signal. Such a signal becomes dispersive in uniform shallow-water environments where multiple modes are likely to be received. Higher order modes become nonlinear FM signals due to the above mentioned dispersion effects.

These signal processing tools were evaluated using synthesized linear and nonlinear signals buried in Gaussian white noise at different SNR. Also, a dispersive Right whale up call from cape Cod Bay data was used to assess these tools. Results show that the STFT and SPWVD offer a reasonable visibility of the signals energy distribution in the time and frequency domain. The SPWVD suffers from the high computational cost. The HHT and the WT give poor visibility in white noise and ocean noise. The WVD achieves better time and frequency resolutions but suffers from the cross-term interference between the arrival modes.

Because the carrier of the FM Right whale call is a sinusoidal signal, the exponential basis function will be preferred. We thus will use the STFT to produce the spectrogram of the data in this thesis. The spectrogram parameters will be set to optimal values in order
to either achieve better visibility of the calls or obtain good results of further processing work.
Chapter 4: Theoretical analysis of dispersion in a shallow-water environment

Like other marine mammals that use underwater sounds for communication purposes [4], Right whales produce up-sweep calls to stay in contact with each other [11]. Such common calls are characterized as frequency up-sweeps with duration of ~1s and a frequency range from 50 to 200Hz. In deep water environments such vocalisations are received at hydrophones with little distortion. However, the Right whale migrates to feeding grounds such as continental shelves that supports higher concentration of its prey (zooplankton) over the shallow bottom mixed layer [87].

These shallow-water acoustic environments act as waveguides with multiple reflections off the sea surface and the sea bottom resulting in multipath effects [37]. A primary effect of such multipaths in shallow waters is to cause dispersion of whale vocalisations resulting in distorted signals on reception. In the literature, the existence of dispersive Right whales up calls was reported [88] and information from such calls can be used to estimate the acoustic range to the vocalising whale from the receiver [89].

For purposes of classification, variation in the recorded data set assesses the performance of any proposed classifier. However, variations in the parameters of contact calls are not only controlled by the vocalising whale but also by the dispersive shallow water that distorts the received signal. For this reason, a theoretical and analytical investigation of the impact of environmental parameters on the propagating signal will be presented in this chapter.

Section 4.1 presents a review of the normal mode model approach. The dispersion curves that describe the relation between the acoustic frequency and the mode’s group velocity are illustrated in section 4.2.

The impact of water depth on the first mode and time difference of arrival between modes is investigated in section 4.3. Also, the influence of both the sediment type and the compressional wave attenuation of the sea bottom on sound propagation is discussed in section 4.4. Finally, section 4.5 shows how a dispersive call received on a single
hydrophone can be used to estimate the distance between the calling whale and the receiver.

4.1 Normal mode modelling

There is a variety of ocean propagation models currently used in research as aids to understanding sound propagation in the ocean. The models differ primarily in the way they manipulate the wave equation mathematically [42] and in the number of approximations they make. Two key acoustic propagation models are the ray-trace model and the normal mode model. The ray-tracing model is useful for deep water scenarios where only a few rays are significant and acoustic frequencies are in the upper kHz. For low frequency propagation and bottom interactions in shallow water the normal mode model is preferred [90], [91], [4]. Because our real data set contains Right whale calls (low frequency sound) collected in Cape Cod Bay (shallow water), normal mode modelling will be adopted in this thesis.

In normal mode theory, the contributions of the likely propagating modes are weighted according to the source depth and summed up to form the acoustic field at the receiver. This field is traditionally expressed in terms of transmission loss and given for an isovelocity and range-independent medium (see figure 4.1) as [42]:

\[
TL(r,z) = -10 \log \left( \frac{1}{D} \sum_{n=1}^{\infty} \sin(k_m \pi z) \sin(k_m \pi r) \right) \left( H_0^{(1)}(k_m r) \right)^2
\]  

(4.1)

where: \( k_m, k_m \) represent the vertical and horizontal wave numbers respectively.
and \( r \) being source depth, water depth, and range respectively.

\[
H_{0}^{(1)} = \frac{e^{ik_{m}r}}{\sqrt{k_{m}}} \text{ is the Hankel function}
\]

Note that the right hand side of the equation (4.1) includes two terms describing the nature of mode propagation in both range and depth directions. In details, while the term

\[
\frac{e^{ik_{m}r}}{\sqrt{k_{m}}}
\]

(see figure 4.2), the term \( \sin(k_{m}z) \) is responsible for the mode propagation as a standing wave in the depth direction [92]. The standing nature of the acoustic modes will be explained in section 4.2.2.

The number of modes contributing to the acoustic field is influenced by:

- Water depth
- Sediment type
- Acoustic range
- Source frequency

We will focus here on the interference between modes as a function of the source frequency. This can be investigated using KRAKEN (an independent-range ocean acoustic propagation model) [34]. In this example, the transmission of three different single-tone signals 50Hz, 100Hz, and 150Hz through shallow water channels having the same water depth 30m and sediment type namely muddy sand was modelled.

The output of the KRAKEN model is displayed in figure 4.2 by plotting the transmission loss over a range of 5 km. Comparing the figures 4.2(a-c) we see some difference. (a) shows that only the first mode propagates at 50Hz; (b) illustrates the second mode is added to the first mode at 100Hz while all the first three modes join at 150Hz in (c).
Figure 4.2: Transmission loss versus range for three different frequencies in 30 m deep water. (a) at 50 Hz, only the first mode propagates through the channel; (b) at 100 Hz, the first and second modes are joined; (c) the first, second, and the third modes are joined at 150 Hz.

4.2 Dispersion in shallow water

In uniform shallow waters, the environment acts as a waveguide through which the signal becomes trapped and propagates for long ranges due to reflections off the sea surface and the sea bottom see figure 4.3. Such multiple reflections result in multipath propagation which causes a mono-component signal to become a dispersive multi-component signal at the receiver. Such a dispersive signal normally comprises multiple modes, each mode is received at different times with different relative energy.
In order to better understand the modal formation in dispersive shallow water, we will address the dispersion curves and the dependence of mode excitation on source depth in the next subsections.

4.2.1 Dispersion curves

The term “dispersion” is a common term used in waveguides. It refers to the dependence of the mode’s group velocity on its frequency due to channel geometry [93]. Every waveguide is characterized by its “dispersion curves” formed by plotting the group velocity against the acoustic frequencies at which the likely modes propagate.

For illustration purposes, the Pekeris waveguide was considered here where both the water and bottom layers are homogenous and range-independent. According to the acoustic frequency, the interfering plane waves reflected at the waveguide boundaries have to be in phase to produce the propagating modes [42].

The numerical calculations of dispersion curves will be presented as follows:

The vertical wave number of the $m^{th}$ mode is a function of water depth $D$ and the phase shifts at the waveguide boundaries such that [91]:

$$K_{zm} = \frac{1}{D}[(m-1)\pi + \theta_{upper} + \theta_{lower}]$$ (4.2)

$\theta_{upper}$, $\theta_{lower}$ represent the phase shifts at the waveguide boundaries. While the former is considered $\pi/2$ for a free sea surface at which the reflection coefficient $R_{upper} = -1$, the latter is a function of the sea bottom parameters and can be obtained from the reflection
coefficients at the sea floor as follows:

\[
R_{\text{lower}} = \frac{m \cos \theta_1 - i \sqrt{\sin^2 \theta_1 - n^2}}{m \cos \theta_1 + i \sqrt{\sin^2 \theta_1 - n^2}} \Rightarrow \vartheta_{\text{lower}} = \arctan \frac{\sqrt{\sin^2 \theta_1 - n^2}}{m \cos \theta_1}
\] (4.3)

Where: \( \theta_1 \) represents the incident angle to the normal to the plane sea floor surface

\[
m = \frac{\rho_2}{\rho_1}
\]
is the ratio between the densities of sediment \( \rho_2 \) and water \( \rho_1 \).

\[
n = \frac{c_1}{c_2},
\]
the refraction index, represents the ratio between the sound speeds of water \( c_1 \) and the sediment \( c_2 \).

In fact, the equation (4.2) expresses the dependence of the vertical wave number on the environmental parameters such as water depth and the sea bottom type for each incident angle.

After calculating the vertical wave number for a set of incident angles \( \{ \theta = \sin^{-1} \left( \frac{c_1}{c} \right) - 90^\circ \} \), the group velocity \( u_m \) of \( m^{th} \) mode is given as:

\[
u_m = \frac{d\omega}{dk_{rm}}
\] (4.4)

The angular frequency, \( \omega \), and the horizontal wave number, \( k_{rm} \), are expressed respectively as:

\[
\omega = \frac{k_{2m} c_1}{\cos(\theta_1)}
\] (4.5)

\[
k_{rm} = \left[ \left( \frac{\omega}{c_1} \right)^2 - k_{zm}^2 \right]^{1/2}
\] (4.6)

By plotting the group velocity against frequency for the first four modes in the waveguide
described in figure 4.1, the dispersion curves are produced as shown in figure 4.4.

![Dispersion curves in 30 m-deep shallow water with a sediment of muddy sand](image)

**Figure 4.4:** Dispersion curves in 30 m-deep shallow water with a sediment of muddy sand

From figure 4.4, we show that:

1) Two aspects of dispersions can be noted here:

   a) Intermodal dispersion which supports earlier arrival of lower modes than that of higher ones. However in exceptional circumstances when dealing with a complicated sound velocity profile (SVP) structure the first mode will not be the fastest [94], [95].

   b) Intramodal dispersion that shows how higher frequencies within the same mode propagate faster than lower frequencies. This is not always the case. Shallow-water environments with certain SVPs, for example downward-refracting profile, complicate the dispersion structure [94]. The primary result of such a SVP is that higher-order modes support faster propagation of lower frequencies than higher frequencies over a specific frequency band.

2) The group velocity of each mode approaches the sea bottom sound speed at the cut-off frequency below which radiated sound energy propagates through sea bottom rather than
3) Each mode group velocity approaches water sound velocity at high frequencies

Note that according to (2), low-frequency sounds such as 20Hz calls produced by the Blue whales are unable to travel in such environment represented in figure 4.1 because the cut-off frequency of the first mode $f_{\text{cut-off}} = 28.38$ Hz is higher than the maximum frequency of the emitted call.

We mentioned in section 4.1 that the acoustic frequency is one of the factors that control the number of propagating modes in shallow waters. If we revisit figure 4.4, we will see that only the first mode can propagate through the 30m deep water at 50Hz while the first three modes propagate at 150Hz. This agrees with the output of KRAKAN model (refer to figure 4.2).

As noted from figure 4.4, intermodal dispersion causes a group velocity difference between modes at a certain frequency. This helps compute the time difference of arrival between the arrival modes (TDOA) which can be used in estimating the acoustic range, as will be presented later in section 4.5.

4.2.2 Mode excitation

As mentioned before, propagating modes can be considered as standing waves in the depth direction. Every mode has zero crossings (nodes) and its opposites (anti-nodes) along the depth axis. The dependence of the mode’s excitation on depth can be expressed by substituting the vertical wave number $K_{zm}$ calculated in equation (4.2) in the following formula [91]:

$$E_m(z) = \sin(K_{zm}z)$$

(4.7)

As figure 4.4 shows four modes are likely to propagate at 200 Hz through 30m deep water. Equation (4.7) will result in a plot of the excitation functions of four modes versus depth as displayed in figure 4.5.
Figure 4.5: Mode excitation as a function of depth at 200Hz

The depths of both the source and the receiver control the mode excitation [42]. Since the receiver depth is normally known in experiments, we will focus on the influence of the source depth. For example, figure 4.5 indicates that if we had a source at a depth of 17m, the first and third modes would be highly excited, whereas the second and fourth modes would not be excited. No modes are excited at the sea surface ($z=0$).

One implication of the above theory is that by studying the relative excitation of the first and second modes in the dispersive channel, we will be able to assess the likely depth of the source [96]. For example, if the first mode is more excited than the second mode, the source whale is likely to be located at mid water; or alternatively if the second mode is more excited than the first the whale is likely to be close to either the sea surface or the sea bottom.

4.3 The effects of water depth on dispersive up calls

The habitats of Cape Cod Bay, the Bering Sea, and the Bay of Fundy are considered to be favourable for North Atlantic Right whales. Such habitats have different water depths (see table 4.1). This fact led us to investigate the effect of the water depth on the up-swept calls received in geographical locations having different water depths.
Table 4.1: Typical water depths for different Right whales habitats [97], [89], [98]

<table>
<thead>
<tr>
<th>Right whales habitat</th>
<th>Typical water depth</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Bay of Fundy</td>
<td>~150m</td>
</tr>
<tr>
<td>The Bering Sea</td>
<td>~70m</td>
</tr>
<tr>
<td>The Cape Cod Bay</td>
<td>~30m</td>
</tr>
</tbody>
</table>

As we saw in section 4.2.1, the relationship between the mode’s group velocity and its frequency is mainly dependent on the waveguide geometry. Water depth is such an important environmental parameter that its impact on the first mode and group velocity difference between modes will now be investigated in the next two subsections.

### 4.3.1 The frequency contour of the first mode

The peak frequency contour is a common characteristic of the tonal vocalisations. The contour extracted from a dispersive up call does not characterise the source vocalisation due to the formation of multiple modes with different relative excitation at the receiver. The effect of water depth on the frequency contour of the first mode propagating through the shallow waters will be modelled. Right whales produce common calls with variant FM rate and duration in all their habitats [88]. The FM deviation of the received signal’s first mode is not only influenced by the source whale but also by the water depth $D$ which is proportional inversely to the first mode’s cut-off frequency as given by [42]:

$$f_{c1} = \frac{c_w}{4D \sqrt{1 - \left(\frac{c_w}{c_b}\right)^2}} \quad (4.8)$$

The effect of water depth can be demonstrated by producing the dispersion curve of the first modes propagating through shallow waters having different water depths, namely 30, 45, 60, and 90m but with the same sediment type, muddy sand.

From figure 4.6, we see that the group velocity of the first mode in deeper waters is larger than in shallower ones. For example, the first mode travels faster through a 90m deep
water (i.e., arrives earlier) than in 30m deep water. Since the mode’s group velocity approaches water sound speed at high frequencies, its sensitivity to the changing in water depth decreases with increasing frequency.

The number of modes is another aspect influenced by the water depth in shallow waters (section 4.1). Since the number of propagating modes is mainly controlled by the cut-off frequency which is in turn related inversely to the water depth, deeper water channels support higher number of propagating modes than shallower waters in the frequency range of the Right whale.

From equation (4.8), the cut off frequency of higher modes \( \geq 2 \) can be expressed as an odd-integer multiple of the first mode’s cut-off frequency \( f_{c1} \) as:

\[
f_{cm} = (2m - 1) f_{c1} \quad \text{m} = 2, 3, \ldots
\]  

(4.9)

4.3.2 Group velocity difference between modes one and two

Since the group velocity difference between the modes received at a single receiver can
be used to estimate the acoustic range, we investigate here the influence of water depth on such a parameter. The scenario in this example was based on a frequency of 120Hz. The dispersion curves of modes 1 and 2 were produced for uniform shallow waters having the same sediment of muddy sand but with different water depths ranging between 30m and 180m with an increment of 5m; see figure 4.7.

![Figure 4.7: Group velocity difference between the first two modes versus water depth at a frequency of 120 Hz.](image)

It is of interest to note that the group velocity difference decreases steeply with increasing water depth in shallower channels (Cape Cod Bay) while slightly in deeper ones (the Bering Sea and the Bay of Fundy).

### 4.4 Boundaries of the waveguide

The sea surface and the sea bottom constitute the waveguide boundaries in the shallow water environment. Due to the high acoustic impedance mismatch between air and water, the sea surface is essentially considered a perfect reflector in calm conditions [99]. On the other hand, the reflectivity behaviour of the lower boundary varies with the sea sediment type. In this work, we will deal with the reflective sea floors \( c_2 > c_1 \) to insure that sound energy propagates through water. Otherwise, sound energy will penetrate into
the sea floor [42].

The sea bottom types covering the lower boundary of most Right whales’ shallow water habitats differ with geographical location [97], [89], [98]. This inspired us to investigate the effects of the sediment parameters on the transmission loss, the first mode, and the TDOA between the first two modes. Also, we investigated the impact of the compressional wave attenuation on the received up-swept calls.

### 4.4.1 Sediment type and transmission loss

The term “transmission loss” is commonly used in underwater acoustics to describe the change in the propagating signal’s strength over increasing range [4]. Using SAFARI (a range-independent model) to model sound propagation in the Bay of Fundy, it was found that propagation loss is higher over LaHave clay than Scotian Shelves drift [31]. The properties of the above mentioned sediments are listed in the table 4.2.

In this section, the influence of sediment type on the transmission loss of a propagating call in shallow water will be addressed. This can be achieved by modelling the transmission of a single-tone signal of 100Hz over different types of sediment in 30m deep water using KRAKEN model [34].

The results of the above-mentioned scenario are shown in Figure 4.8 where sediments of higher acoustic impedance (density \(\times\) sound speed) supports lower transmission loss at long ranges and consequently this increases the probability of acoustic detection. Note that transmission loss is lower over gravel than that over both silt and sand at 100 Hz for long ranges (>10 km).
Figure 4.8: Transmission loss versus range for different sediment types at 100 Hz, water depth: 30 m, source and receiver depth: 28 m. The geo-acoustic properties of silt, sand, and gravel are listed in table 4.3.

Table 4.2: Sediment parameters in the Bay of Fundy [97]

<table>
<thead>
<tr>
<th>Bottom type</th>
<th>LaHave Clay</th>
<th>Scotian drift</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compressional sound speed km/s</td>
<td>1.261-1.49</td>
<td>1.745-1.92</td>
</tr>
<tr>
<td>Compressional wave attenuation dB/m-kHz</td>
<td>0.023-0.056</td>
<td>0.0065</td>
</tr>
<tr>
<td>Density (g/m$^3$)</td>
<td>1.5-1.54</td>
<td>2.1</td>
</tr>
<tr>
<td>Shear speed km/s</td>
<td>0.0</td>
<td>0.4-0.4</td>
</tr>
<tr>
<td>Shear wave attenuation dB/m-kHz</td>
<td>0.0</td>
<td>10.0</td>
</tr>
</tbody>
</table>

Table 4.3: Geoacoustic properties of typical sediments [42]

<table>
<thead>
<tr>
<th>Bottom type</th>
<th>Density (kg/m$^3$)</th>
<th>Sound speed (m/s)</th>
<th>Compressional wave attenuation (dB/λ)</th>
<th>Shear wave attenuation (dB/λ)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1- Silt</td>
<td>1500</td>
<td>1575</td>
<td>1.0</td>
<td>1.5</td>
</tr>
<tr>
<td>2- Sand</td>
<td>1700</td>
<td>1650</td>
<td>0.8</td>
<td>2.5</td>
</tr>
<tr>
<td>3- Gravel</td>
<td>1900</td>
<td>1800</td>
<td>0.6</td>
<td>1.5</td>
</tr>
</tbody>
</table>

4.4.2 Sediment type and dispersion relations

From equation (4.2), the vertical wave number is a function of the phase shift between
incident and reflected waves at the sea floor. Since the reflectivity behaviour of the sea 
floor varies with sediment types (i.e. sediment parameters), we focus here on the 
influence of sediment type on the dispersion curves characterizing the shallow 
environment. Thus, the dispersion curves of the first mode were produced for different 
channels having the same water depth but with different sediment types; see figure 4.9.

We can see that at very low frequencies (close to the cut-off frequency) both the cut-off 
frequency and group velocity of the first mode decreases with increasing the acoustic 
impedance of the sediment type. For instance, at 80Hz the first mode interacting with 
sediment of silt propagates faster than that interacting with gravel sediment. We can also 
see that at the cut-off frequency, the mode travels at the sound speed of sediments 
covering the bottom of the acoustic channel.

On the other hand, the group velocity difference between the first two modes propagating 
at 120Hz through different channel scenarios having the same water depth of 30m but 
with different sediment types listed in table 4.4 was computed and plotted against 
sediment types as shown in figure 4.10. The results of this figure indicates that the group 
velocity difference is more sensitive to the change in lower acoustic impedance sediments 
than in higher impedance ones.

Figure 4.9: The first mode dispersion curve for the three sediment types. Water depth is 30 m

On the other hand, the group velocity difference between the first two modes propagating 
at 120Hz through different channel scenarios having the same water depth of 30m but 
with different sediment types listed in table 4.4 was computed and plotted against 
sediment types as shown in figure 4.10. The results of this figure indicates that the group 
velocity difference is more sensitive to the change in lower acoustic impedance sediments 
than in higher impedance ones.
Theoretically, we investigated the influence of water depth and sediment type on the up calls received in shallow water. In order to illustrate how water depth controls the influence of sediment type on the received up calls, the transmission of a 120 Hz single tone through uniform shallow water was modelled. This scenario was run for different water depths ranging between 30m and 180m with an increment of 5m and different sediment types ranging between very fine sand to coarse sand. The group velocity difference between modes 1 and 2 was obtained from the modelled dispersion curves for
each different water depth; see figure 4.11.

The results of figure 4.11 indicates that the group velocity difference between the first two modes is sensitive to sediment type change in shallower waters (<60m) while this effect does not seem significant in deeper waters (>60m).

4.4.3 Sediment attenuation coefficient and dispersion curves

The sediments covering the lower boundary of shallow waters are characterized by a range of parameters including density, sound speed and attenuation coefficients. Since the influence of the acoustic impedance of the sea bottom on received up calls has been previously addressed, we investigate here the effect of the attenuation coefficient on the dispersion curve of the first mode. In this work, sediments were assumed to support only compressional wave attenuation.

The compressional wave attenuation appears in the imaginary part of the wave number expression of the sediment layer as given by [42]:

![Figure 4.11: Group velocity difference versus water depth for different sediment types at a frequency of 120 Hz.](image-url)
Where $\alpha > 0$ is given in $\text{neper/m}$. 

The index of refraction is expressed as:

$$ n = \frac{k_2}{k_1} = \frac{c_1}{c_2} (1 + i \frac{\alpha c^2}{\omega}) $$

(4.11)

The attenuation coefficient $\alpha^{(2)}$ in $\text{dB}/\lambda$ and the angular frequency, $\omega$, are given by [42]:

$$ \alpha^{(2)} \approx 8.868 \alpha \lambda $$

(4.12)

and

$$ \omega = 2\pi \frac{c_2}{\lambda} $$

(4.13)

By substituting equations (4.12) and (4.13) in (4.11), the final formula of the refraction index implies

$$ n = \frac{c_1}{c_2} \left(1 + i \frac{\alpha^{(2)}}{2\pi(8.868)}\right) $$

(4.14)

The equation (4.14) has to be substituted in equation (4.3) to produce the dispersion curves for a given attenuation coefficient.

The dispersion curves of the first mode in different water channels having the same water depth (30m) and sediments but with different compressional wave attenuation coefficients ranging between 0.1 and 1.0 $\text{dB}/\lambda$ are shown in figure 4.12.

The influence of compressional wave attenuation of sediment layer on the dispersion
curve of the first mode appears at the cut-off frequency. For example, whereas the first mode travels at the sediment sound speed in the absence of absorption its group velocity becomes less than the water sound speed for high attenuation values such as $0.8 \, dB/\lambda$ and $1.0 \, dB/\lambda$.

![Figure 4.12: The effect of attenuation in sediment layer on the dispersion curve](image)

Due to the weak excitation of modes near the cut-off frequency, less attention would be paid to this part of the dispersion curves. To sum up, the group velocity does not seem to be sensitive to change in the sediment attenuation at frequencies a little above the airy phase (i.e. global minimum). Thus, despite the compressional wave attenuation affects the signal’s transmission loss, its influence on the dispersion characteristics of received up calls will be considered negligible.

### 4.5 Acoustic range estimation using normal mode analysis

The estimation of whale depth and its range from the receiver is considered as the main purposes of localisation techniques used in Passive Acoustic Monitoring (PAM).

The distance between the source and the receiver can be estimated by measuring the time difference of arrival (TDOA) between the propagating modes received at a single
hydrophone. For example, the formula of range estimation using modes 1 and 2 is given by [89]:

\[
R = TDOA_{1,2} \cdot \frac{c_{m1} - c_{m2}}{c_{m1} - c_{m2}}
\]  \hfill (4.15)

Where R is the acoustic range; \( c_{m1} \) and \( c_{m2} \) are the group velocities of the first and second modes respectively at a certain frequency. The impacts of water depth and sediment type on the term \( |c_{m1} - c_{m2}| \) were discussed in sections 4.3.2 and 4.4.2.

![Figure 4.13: The acoustic range versus TDOA between the first two modes at 100 Hz and a water depth of 30m for two sediment types.](image)

The TDOA between the modes is a function of acoustic range, source frequency, and water depth as well as sediment type. This can be given as:

\[
TDOA_{1,2} = R \frac{|c_{m1} - c_{m2}|}{c_{m1} \cdot c_{m2}}
\]  \hfill (4.16)

For a shallow water of uniform water depth, the influence of sediment types on the
TDOA between modes 1 and 2 over increasing ranges can be illustrated by substituting the group velocities of the first and second modes \(c_{m1}\) and \(c_{m2}\) at a specific frequency in equation (4.16); see figure 4.13. As could be seen, the TDOA sensitivity to the sediment type increases with increasing range. When three recording channels exist (such as the case in Cape Cod Bay), the information provided by TDOA between modes received at a single hydrophone and the TDOA between two hydrophones can be used to relatively predict the characteristic impedance of the sediments, see appendix A.

4.6 Chapter Summary

This chapter began by introducing the theoretical principles of normal mode modelling which is preferred to address the propagation of low-frequency signals in shallow-water environments. Also, dispersion curves that refer to the dependence of the mode’s group velocity on its frequency were used to investigate the effects of environmental parameters of the dispersive channel on the received Right whale up calls. This includes the influence of water depth and sediment types on the first mode’s group velocity and the time difference of arrival between modes.

It was found that the number of arrival modes and the group velocity of the first mode increases with increasing water depth. The TDOA between modes 1 and 2 decreases steeply with increasing water depth in shallower waters (<70m); whereas its sensitivity to that water depth change decreases in deeper water (>70 m).

The results of KRAKEN model shows that sediments of higher acoustic impedance increases the ability of acoustic detection as they support lower transmission loss in the propagation channel. Also, increasing the acoustic impedance of sediments decreases the group velocity of the first mode and increases the TDOA between modes 1 and 2.

Water depth and range control the influence of sediment type on the TDOA between modes. It was found that the TDOA is sensitive to sediment type change in shallower water (<60 m) while no significant effect appears in deeper waters (>60 m). Also, the TDOA sensitivity to sediment type increases with increasing range.

After the theoretical aspects of the influence of the dispersive channel on Right whale up
calls, a broadband layered normal mode “PROSIM” is used in the next chapter and the results of its output will be compared with real acoustic data.
Chapter 5: Modelling and real data based evaluation

In the previous chapter, the theory of normal mode modelling and dispersion curves in shallow waters was reviewed and it was noted that shallow-water environmental parameters have an influence on Right whale up-calls received in uniform acoustic channels. To investigate such an influence, a synthetic signal having the same parameters of the common Right whale up-call will be used to feed a normal mode model named \textit{PROSIM}. However, it is important to note that accurate sound propagation modelling for an acoustic channel requires sufficient information about its environmental parameters. These include water depth, sound velocity profile (SVP), sea bottom type, receiver configuration, source depth (unknown in our work), source-receiver distance (to be estimated), and source level.

In this chapter, different scenarios will be explored to investigate the influence of water depth and sediment types on the first mode, group velocity and TDOA between modes. This requires setting the model inputs to the appropriate values of Cape Cod Bay. The model outputs are analyzed to provide results to be correlated with the theoretical principles addressed in the previous chapter.

The normal mode modelling results will be compared with the real acoustic data obtained from IFAW and recorded in Cape Cod Bay. We will highlight the dispersive impacts of the acoustic channel on recorded data. The effect of source depth on mode excitation as well as the influence of sediment type on the TDOA between modes will be investigated. Examples of real dispersive calls will provide an insight as to how the acoustic range influences the TDOA between the first two modes as well as the number of modes.

Finally, the frequency and time resolutions for dispersive calls will be discussed.
5.1 Input data for the model

5.1.1 Receiver Configuration

The data used in this work was obtained from the International Fund for Animal Welfare (IFAW) and relates to signals recorded by triangularly configured hydrophones moored 2m from the seafloor in Cape Cod Bay, the Eastern Coast of the USA, in March 2001; see figure 5.1.

![Figure 5.1: Locations of three pop ups in Cape Cod Bay. Image Courtesy of [101]](image)

The latitude and longitude coordinates of the hydrophones deployed to collect data in Cape Cod Bay experiment are listed in table 5.1. The distances in km between receivers are displayed in figure 5.2.

<table>
<thead>
<tr>
<th>Channel</th>
<th>Latitude</th>
<th>Longitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.(ch1)</td>
<td>41 57 26.28 N</td>
<td>-70 10 0.12 W</td>
</tr>
<tr>
<td>4.(ch0)</td>
<td>41 55 49.80 N</td>
<td>-70 09 55.8 W</td>
</tr>
<tr>
<td>1.(ch2)</td>
<td>41 56 49.92 N</td>
<td>-70 11 35.41 W</td>
</tr>
</tbody>
</table>

Table 5.1: Deployment information for three receivers in the Cape Cod Bay
5.1.2 **Sound Velocity Profile (SVP)**

The sound velocity profile, sound speed (in m/s) versus depth (in m), of the environment of Cape Cod Bay is approximately 1463 m/s [101]. It can be considered to be nearly constant over depth; see figure 5.3.

![Sound Speed Profile in Cape Cod Bay](image)

**Figure 5.3: Sound velocity profile (SVP) for Cape Cod Bay**

5.1.3 **Transmitting signal**

Since Right whales produce up-swept calls that typically lie within a frequency range extending from 50Hz to 200Hz with ~1 second duration, the simulated transmitting signal will be generated as a linear FM chirp with the same duration and frequency range. The time waveform and frequency response of the simulated up call are displayed in figures 5.4 and 5.5 respectively. The source level is set to the average value of 150dB *rms* re 1 $\mu$Pa – m as reported for tonal calls produced by North Atlantic Right whales [12].
5.1.4 Selection of bottom parameters

Cape Cod Bay is a semi-enclosed embayment with the seafloor deepening from south to north as shown in the bathymetry map (see appendix B). The sea bottom type differs between the shallow margins and deep basin of the bay due to the distribution of sedimentary deposition in the bay [98]:

a) Environments of erosion (nondeposition) that occupy the shallow margins (water depth<30m) comprises a sub bottom consisting of bedrock, glacial drift, and sediments ranging from boulder field to gravelly course to medium sands.

b) Environments of deposition are completely covered by muds and muddy fine-to-very
fine sands that were accumulated during calm periods (summer) over the basin floor that ranges in depth from 30m to 60m.

c) Environments of sediment reworking are distributed over transitional slopes connecting margins and the basin between the above environments. Sediments obtained from that area ranges from boulders and gravelly sands to mud.

<table>
<thead>
<tr>
<th>Layer</th>
<th>Muddy sand</th>
<th>Medium sand</th>
<th>Gravelly muddy sand</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sound speed (m/s)</td>
<td>1463</td>
<td>1620</td>
<td>1767</td>
</tr>
<tr>
<td>Density (kg/m³)</td>
<td>1026</td>
<td>1373</td>
<td>1893</td>
</tr>
</tbody>
</table>

The sediment parameters used in the model are summarized in table 5.2. The compressional wave attenuation coefficient was assumed to be negligible in this work.

5.2 Broadband normal mode simulations

In the previous chapter, we have theoretically discussed the influence of water depth and sediment type on received signals in shallow waters. Before progressing to analyse the model data, it is important to note that setting the model parameters to accurate values of the environment of interest is essential to achieve successful analysis. In this section, we shall compare the theoretical principles with the model output data. Such principles include the influence of water depth and sediment type on dispersive received signals.

5.2.1 The effects of water depth on the dispersive received signal

The effects of water depth on received up calls can be considered in term of two aspects:

5.2.1.1 The first mode

This effect was investigated by modelling the transmission of a linear chirp through four water channel scenarios having the same sediment type namely medium sand and different water depths 30, 45, 60, 90m. The range was taken as 10 km. Viewing the
frequency contour of the first mode received at the given long range displays different FM deviations particularly at lower frequencies.

The results of figure 5.6 indicate that the first mode propagating through 60m and 90m deep channels arrives earlier than those traveling through 45m and 30m deep waters. The primary effect of these results is to cause the duration of the first mode to become influenced by the water depth of the acoustic channel. For example, the duration of the first mode is smaller in 30m and 45m deep waters than in 60 and 90m deep channels.

The related question that arises here is “does range control the influence of water depth on the up call duration?” The answer can be illustrated by modelling the transmission a linear chirp having a frequency range between 50 and 200 Hz and a duration of 1 second through different water channel scenarios having the same sediment type namely medium sand and different water depths 30, 45, 60, 90, 120 m. The duration of the first mode is measured at different ranges between 500 m and 15 km with an increment of 0.5 km as illustrated in figure 5.7.
The results of figure 5.7 indicates that the duration of the first mode in shallower waters (<60 m) is very sensitive to range increase while this effect does not seem significant in deeper waters (>60m). However, the first mode duration decreases by about 10% of the source call duration at 5 km.

To compensate for the increased band-limited background noise in Right whale shallow-water habitats, a recent study shows that the minimum frequency of the right whale up calls is shifted up to about 100Hz (see section 1.7). For this reason, we run the above scenario but for a linear chirp having a frequency range between 100 and 250 Hz and duration of 1 second. The results of figure 5.8 show that the influence of range on the first mode duration is very small in deeper waters (90m and 120) over the entire range. In 30 m deep water, the effect of short range (<5 km) does not seem significant.
5.2.1.2 The TDOA between modes one and two

Results for two water depths 30m and 45m with the same sediment of gravelly sand are shown in figure 5.9.

Although 45m deep water supports more propagating modes, the TDOA between the first two modes in 30m deep water is bigger than that in 45m deep channel. This is indicated by the white horizontal arrows drawn at 80Hz.
5.2.2 *Sediment type effects on the dispersive received signal*

The influence of the sea bottom on the dispersive received up calls can be addressed in terms of the acoustic impedance of the sediments. Sediments of higher acoustic impedance support lower cut-off frequencies which in turn increase the number of propagating modes. In contrast with the water depth effect, decreasing the cut-off frequency in this case results in decreasing the mode group velocity.

Three aspects are presented here in order to consider the effects of sediment type on the received signal.

5.2.2.1 The first mode

In section 4.4.2, we observed that low-impedance sediment layers support faster modes at low frequencies (close to airy frequency). Extracting the frequency contour of the first mode for two different sediments indicates that the first mode interacting with muddy sand arrives earlier than that interacting with gravelly muddy sand as displayed in figure 5.10.

![Figure 5.10: The effect of sediment type on the first mode. Water depth=30m and a range of 10 km.](image)

The influence of sediment type change on the first mode appears very small even at lower frequencies at a range of 10 km. With the semi-enclosed embayment in Cape Cod Bay; the limitation of the acoustic range does not support a remarkable effect of sediment type
5.2.2.2 On the TDOA between modes one and two

Results for two different bottom types are shown in figures 5.11 and 5.12. The water depth and range for both results are 30 m and 7 km respectively. It was found that the linear chirp signal propagating over sediments of gravelly sand has a bigger TDOA between the first and second modes than that travelling over sediments of muddy sand at a low frequency such as 100 Hz. The influence of bottom type on the TDOA between the first modes one and two was noticeable in very shallow waters (≤ 60 m) at low frequencies.

The effect of sediment types in deeper environments (≥ 60 m) was found to be negligible for Right whale up-swept calls. This emphasizes the importance of the sediment type at shallow water depths. Figures 5.13 and 5.14 indicate that TDOA between modes one and two is not influenced significantly by the change of sediment type.

Figure 5.11: Spectrogram of a received signal at 7 km; Rd and Sd are at 28 m; Water depth=30 m; sediment type: Muddy sand. Spectrogram is normalized (linear)

Figure 5.12: Spectrogram of a receive signal at 7 km; Rd and Sd are at 28 m; Water depth=30 m; sediment type: Gravelly sand. Spectrogram is normalized (linear)
5.2.2.3 On the number of modes

In shallow waters, the primary effects of sediment type change appear on the mode’s cut-off frequency. This affects the number of propagating modes through the water channel.

The variation in the acoustic impedance of sediments influences the mode’s cut-off frequency which in turn controls the range of acoustic frequencies at which modes can travel. For example, the second mode seems unlikely to propagate over muddy sand sediments in 30 m deep water at 80Hz (figure 5.11); whereas sediments of gravelly muddy sand supports its propagation through the channel at the same frequency (figure 5.12). On the other hand, increasing the number of modes results from decreasing the cut-off frequency of modes. For instance, the number of modes propagating over gravelly muddy sand is larger than that propagating over muddy sand.

5.3 Modelling of mode excitation

In this section, an investigation of theory with regards to the influence of source depth on the mode excitation was undertaken. The receiver depth was set to 28 m as given from the experiment in Cape Cod Bay. Different source depth scenarios (5, 15, 28 m) were made in 30 m deep water having sediments of medium sand. The transmission of a linear chirp was modelled at 7 km from the receiver.
The modelling results, shown in figure 5.15, show that the second mode is more excited.
than the first one when the source was located close to either the sea surface or the sea floor. On the other hand, when the whale moves towards mid water the first mode is the most excited mode.

### 5.4 Acoustic range estimation

In the previous chapter, the acoustic range was expressed as a function of the group velocity and time difference between modal arrivals at a specific frequency. Modelling analysis of a linear chirp received at 5 km from the source in 30m deep water will now be presented here; see figure 5.16.

---

**Figure 5.16:** Spectrogram of a linear chirp received at 5km. colour bar is linear

**Figure 5.17:** The absolute group velocity difference between the first and second modes at 100 Hz

**Figure 5.18:** A zoomed-in spectrogram showing the TDOA between the first and second modes at a frequency of 100Hz
The acoustic range can be estimated from equation (4.15) as:

\[ R = TDOA_{1.2} \cdot \frac{c_{m1} c_{m2}}{|c_{m1} - c_{m2}|} \]

While TDOA can be measured from the spectrogram of the received signal (see the horizontal black arrow shown in figure 5.18), the group velocity is obtained from the dispersion curves of the modeled channel (see the vertical black arrow displayed in figure 5.17)

### Table 5.3: Measurements to estimate the acoustic range at 100Hz

<table>
<thead>
<tr>
<th>Mode number</th>
<th>Group velocity (m/s)</th>
<th>Time of arrival (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1444.5</td>
<td>0.7617</td>
</tr>
<tr>
<td>2</td>
<td>1366.5</td>
<td>0.9631</td>
</tr>
</tbody>
</table>

Substituting the values listed in table 5.3 in equation (4.15) results in \( R = 5.1 \text{km} \). This agrees with the given range with an acceptable error.

#### 5.5 The linear relation between the mode’s time of arrival and range

This section attempts to prove the linear relation between the time of arrival of the first mode and range in shallow waters. Two methods are adopted here: Numerical dispersion curves and modeling by PROSIM in 60m deep shallow water with sediments of muddy sand at different frequencies 80, 150, and 250Hz. In the first method, the group velocity of the first mode was taken at the given frequencies and the time of arrival was referenced to the arrival time at 2 km as the following:

\[ t = \frac{\text{range} - 2000}{c_{m1}} \]  \hspace{1cm} (5.1)

In the second method, the transmission of different single tones through the channel described above was modelled and the time of arrivals was measured. In both cases, the relation between the arrival time of the first mode and range was shown to be linear.
Figure 5.19: The time of arrival of frequencies 80, 150, and 250Hz propagating in 60m deep shallow channel modelled by PROSIM (b, d, f) as well as using numerical dispersion curves (a, c, e). The values on the vertical axis represents the time of arrival referenced to 2 km.
5.6 Dispersive up-swept calls in Cape Cod Bay

In the previous sections synthetic data were used. Attention is now turned to examples of real data. Visual scanning of spectrograms of the recorded acoustic data reveals the existence of dispersive impacts caused by the acoustic environment in Cape Cod Bay (see figure 5.20). This environment can be considered as a waveguide bounded by the ocean surface and the seafloor.

![Spectrogram of a dispersive Right whale up call. Spectrogram is normalized (linear)](image)

**Figure 5.20:** Spectrogram of a dispersive Right whale up call. Spectrogram is normalized (linear)

To illustrate the main features that contribute to the dispersive influence of shallow water on received up calls, figure 5.21 displays the dispersion curves for 30m deep water and an illustrative synthetic dispersive received up call comprising two modes.

![Dispersion curves for 30m deep water and an illustrative synthetic dispersive received up call comprising two modes](image)

**Figure 5.21:** Inter-modal and intramodal dispersion in 30m deep water. Black bars indicate that the time difference of arrival between modes at different frequencies is a function of their group velocity difference.
Figure 5.21 identifies two types of dispersion: inter-modal dispersion that can be recognized as multiple modal arrivals received at a hydrophone at a sufficient range and intra-modal dispersion that exhibits faster propagation of higher frequencies than lower ones within the same mode. The latter dispersion shows smaller TDOA between modes one and two at higher frequency than lower frequency ($\Delta t_2 > \Delta t_1$). Both dispersion aspects can be recognized within the time-frequency representation of a typical real dispersive up call shown in figure 5.20. On the other hand, dispersion curves provide information about the number of modes that are likely to propagate through the channel at a specific acoustic frequency. For instance, only two water-borne modes can propagate at a frequency of 120Hz. Note that it was assumed that all likely modes arrive at the receiver.

To sum up, we have used an example of real data and theory that explains the influence of the dispersive nature of shallow waters on received Right whale up-calls. In the next section, we will investigate the dependence of mode excitation on the depth of the vocalising whale.

5.7 Mode excitation

One implication of the theory of sound propagation with regards to source depth is that from the relative excitation of the acoustic modes in the dispersive channel we will be able to assess the likely depth of the source whale. From previous theory, if the first mode was more excited than the second one the whale is likely to be located at mid water; or alternatively if the second mode is more excited than the first the whale is likely to be close to either the sea surface or the sea bottom.

In order to illustrate how the depth of the calling whale excites the normal modes, two dispersion effects are displayed where mode 1 and then mode 2 dominates respectively in figure 5.22. While the first mode is more excited than the second one (figure 5.22 (a)), the calling whale is expected to be located at mid water. In figure 5.22(b) the second mode of the received signal is more excited than the first one. This suggests that the vocalising whale may be close either to the sea floor or to the sea surface. Note also that in figure 5.22(a) the third mode is minimally excited compared to modes one and two.
Figure 5.22: Mode excitation in Cape Cod Bay data. An illustration of how depth information can be obtained by viewing relative mode excitations. a) Mid water vocalisation; and (b) Close to sea floor / sea bottom vocalization. Spectrogram intensity is normalised (linear)

In order to investigate the sensitivity of mode excitation as a function of source depth two dispersive up calls from the Cape Cod Bay data were located. These vocalisations are displayed in figures 5.22(a) and 5.23. The ranges were calculated using a hyperbolic localisation technique [33] from the three hydrophone receivers deployed in Cape Cod Bay; see figure 5.24.

The localisation results indicate these two calls are received at the same distance (8.8 km) and from the same bearing (from the bay basin where bottom type is muddy sand). Synthetic normal-mode modelling received calls were calculated using initial synthetic two-part upswept call that best fit the first mode, modal dispersion curves, and the
Figure 5.25: The vocalising whale is likely to be (a) located at mid water depth where the second mode is not excited; and (b) moving above or below the mid depth position causes the second mode to become excited. The overlaid white lines are the results from the synthetic normal-mode modelling for the first four modes. The modelling was done for a water depth of 35 m and a range of 8.8 km. Spectrogram intensity is linear.

Figure 5.26: Normalized sound pressure versus depth for the first four modes at 180 Hz. (a) the vocalising whale is likely to be located at mid water depth according to figure 5.25 (a). Moving above or below mid depth position causes the second mode to become excited according to figure 5.25(b).

The normal-mode modelling was made to fit the real data by tuning the water depth parameter. The appropriate value of water depth was set to 35 m. Overlaid on the real mode plot are the results from the synthetic normal-mode modelling for the first four modes.
modes. The modelling emulates the assumed real scenario.

The results of the above normal mode computations illustrates how the second mode is not excited due to the whale depth at mid water (figures 5.25(a) and 5.26(a)) while it becomes excited when the whale moves above or below the mid depth position (figures 5.25(b) and 5.26(b)). This emphasises the significance of that the construction of the mode excitation.

5.8 The influence of sediment type in Cape Cod Bay

As mentioned in subsection 5.1.4, the bottom type in Cape Cod Bay differs between margins (<30m) and basin (30-60m). The distribution of the sediments in the Bay motivated us to investigate the influence of sediment type on received dispersive up calls in 30m deep water using normal mode modelling. Different sediment types including muddy sand, fine sand, and medium sand were used. The detection range was taken as 3.5 km as estimated using [33], see figure 5.27(d).

The results of figure 5.27 shows spectrograms of a dispersive up-call (received at Ch2) overlaid with white lines representing normal mode modelling synthetic data. The sediment distributed in the basin (muddy sand) does not excite the third mode, as shown in figure 5.27(a). Increasing the acoustic impedance of sediment to include fine sand and medium sand shows that a good match exists for sediments, namely medium sand that is assumed to be distributed in the margins of the bay.
5.9  The impact of propagation range on received Right whale calls in shallow water

In previous sections, we investigated the effects of water depth and bottom type on sound propagation through dispersive shallow waters. We now focus on the influence of the distance between the calling whale and the receiver on the received up calls. This includes the likely number of propagating modes and TDOA between them.
Figure 5.28: Three up-swept calls received on three channels in Cape Cod Bay. Note that the vertical red arrows indicating the arrival time of the call show that the ch0 is the furthest channel to the vocalising whale.

Spectrogram intensity is linear

Figure 5.28 shows three Right whale up calls which are assumed to be emitted by the same whale and received on three hydrophones. According to [33], the whale location is estimated to be as shown in figure 5.28(d).

Note that the furthest channel (Ch0) to the calling whale receives the dispersive up call with the biggest TDOA between modes. This suggests that TDOA increases with range and consequently the hydrophone receiving the signal with the smallest TDOA is the closest to the source whale.

To illustrate the effect of acoustic range on the number of modal arrivals, gunshot calls produced by adult male Northern Right whales are displayed as spectrograms (dB) in figure 5.29. Such calls are broadband signals ranging between 20Hz and 20 kHz and are
emitted to be used as a reproductive advertisement [13].

The range to the calling whale from the receivers was estimated using ISHMAEL (figure 5.29(d)). Comparing the figures 5.29(a-c) we see that the dispersive call received at channel Ch0 (the furthest to the calling whale) have the highest number of modes at lower frequencies (<500Hz). The number of arrival modes decreases once the whale gets close to the receiver. For example, Ch1 (the closest to the whale) receives the smallest number of modes. On the other hand, as the gunshot call is a broadband signal we can see that arrival modes can not be easily recognized at higher frequencies (>500Hz). The reason for this is that the group velocities of likely modes approach water sound speed at higher frequency; therefore the TDOA between modes will be very small (recall dispersion curves in section (4.2.1)).
5.10 Estimation of acoustic range in Cape Cod Bay

The distance between the vocalising whale and the receiver can be estimated using equation (4.15). The time difference of arrival and the group velocity of the propagating modes at specific frequency (such as 100 Hz or 150 Hz) are required. While the former is extracted from the spectrogram of the received signal (the white horizontal arrows shown in figure 5.30(a)) the latter is obtained from the dispersion curves of the shallow water through which the signal travels (35 m deep water having sediments of muddy sand), as displayed in figure 5.30(b).

![Figure 5.30: Estimation of acoustic range using Normal Mode Modelling. Spectrogram is normalized (linear)](image)

<table>
<thead>
<tr>
<th>Acoustic frequency (Hz)</th>
<th>Group velocity (m/s)</th>
<th>Arrival time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>1448.1 ( (M_1) ), 1404 ( (M_2) )</td>
<td>1.599 ( (M_1) ), 1.788( (M_2) )</td>
</tr>
<tr>
<td>150</td>
<td>1428.4 ( (M_2) ), 1387.8( (M_3) )</td>
<td>1.968( (M_2) ), 2.151( (M_3) )</td>
</tr>
</tbody>
</table>

From equation (4.15), the acoustic range can be estimated for modes one and two as:

$$R = TDOA_{1,2} \cdot \frac{c_{m1} \cdot c_{m2}}{c_{m1} - c_{m2}} = (1.788 - 1.599) \frac{1448.1 \times 1404}{1448.1 - 1404} = 8.73 \text{km}$$

Or for modes two and three:

$$R = TDOA_{2,3} \cdot \frac{c_{m2} \cdot c_{m3}}{c_{m2} - c_{m3}} = (2.151 - 1.968) \frac{1428.4 \times 1387.8}{1428.4 - 1387.8} = 8.93 \text{km}$$
The range was also calculated using the hyperbolic localisation technique [33], (see appendix C) and it was found to be R=8.65 km. Both the hyperbolic localisation technique and the results from (4.15) give similar results. The hyperbolic localisation technique [33] relies on the TDOA between spatially distributed hydrophones whereas the technique proposed by [89] uses TDOA between modes at a single hydrophone. This result shows an agreement between the outputs of different localisation techniques.

5.11 Frequency and time resolutions for dispersive up-calls

Frequency resolution $\Delta f_{\text{res}}$ can be described as the minimum frequency difference between two sinusoids where two distinct peaks can be recognized in the spectrum. The frequency resolution can be indicated as either good or poor, as shown in spectrograms, figures 5.31 and 5.32 respectively.

![Figure 5.31: Spectrogram of sum of two hanning-weighted sine waves of 100Hz and 150Hz. (Good frequency resolution). Spectrogram intensity is linear.](image1)

![Figure 5.32: Spectrogram of sum of two hanning-weighted sine waves of 130Hz and 150Hz. (Poor frequency resolution). Spectrogram intensity is linear.](image2)

In order to illustrate the frequency resolution for dispersive up calls, figure 5.33 shows a synthetic source call, dispersion curves and the received synthetic dispersive call.
Figure 5.33: Frequency resolution for dispersive calls. (a) Synthetic linear FM call; (b) Dispersion curves for the first 2 modes; (c) Received synthetic dispersive call. Note that 1 and 2 indicate the mode number.

The times required for the first mode frequency \( f_1 \) and the second mode frequency \( f_2 \) to arrive at a range of \( R \) are given as:

\[
\Delta t_{m1} = \frac{R}{c_{m1}} \quad (5.2)
\]

\[
\Delta t_{m2} = \frac{R}{c_{m2}} \quad (5.3)
\]

The two modes arrive at the same time but for different frequencies if:

\[
t_1 + \Delta t_{m2} = t_2 + \Delta t_{m1} = t_2' \quad (5.4)
\]

Where the times \( t_1 \) and \( t_2 \) correspond to the source signal frequencies \( f_1 \) and \( f_2 \).
respectively.

We can say that if:

\[ f_2 - f_1 > \Delta f_{res} \quad \Rightarrow \quad \text{Good frequency resolution} \]

Otherwise \( \Rightarrow \) Poor frequency resolution

From figure 5.33(c), it can be seen that in spectrograms of dispersive calls we have better frequency resolution at lower frequency than higher frequencies where modes can not be resolved.

On the other hand, we have seen that the TDOA between modes at the same frequency is too small at higher frequencies if compared with low frequencies. Thus, to resolve modes approaching each other at higher frequency requires a time-frequency analysis tool with good time and frequency resolutions.

5.12 Beating effects in dispersive Right whale up-calls

According to the dispersion curves, different frequency components of the first and second modes arrive at the receiver at the same time (see the previous section). With increasing frequency, these components become quite close (differences of the order of few hertz). This case scenario leads us to illustrate a phenomenon known in Acoustics as **Beating**. The term “Beating” refers to the interference between two sounds having slightly different frequencies. Such interference can be constructive when the signals are in phase (phase difference=0 degree) or destructive when they are out of phase (phase difference=180 degrees). To illustrate this concept, we consider two sine waves of unit amplitude \((M_1=M_2=1)\) and having a frequency difference of \( \Delta f \) as given:

\[ m_1(t) = M_1 \sin(2\pi f_1 t) \]

\[ m_2(t) = M_2 \sin(2\pi f_2 t) = M_2 \sin(2\pi (f_1 - \Delta f) t) \]
where $\Delta f$ is the frequency difference which is small.

The sum of these signals is given by:

$$m(t) = m_1(t) + m_2(t) = 2\cos(2\pi \frac{f_1 - f_2}{2} t) \sin(2\pi \frac{f_1 + f_2}{2} t)$$

For $f_1 = 126\text{Hz}$ and $f_2 = 120\text{Hz}$, $m(t) = 2\cos(2\pi 3t) \sin(2\pi 123t)$

The resultant signal $m(t)$ can be described as double side band with suppressed carrier signal (DSB-SC) where $m_1(t)$ and $m_2(t)$ constitute the upper side band and the lower side band respectively. The carrier signal of 123 Hz is modulated by a slow signal of 3 Hz constituting the envelope as shown in figure 5.34 where the destructive interference occurs when the envelope $2\cos(2\pi 3t)$ is zero while the signals interfere constructively when the envelope is equal to 2. We have to add here that because the two signals have the same amplitude, the maxima of each cancel the minima of the other and this thus causes the destructive interference, i.e. the resultant signal is zero.

![Figure 5.34: An example of frequency beating with equal amplitude](image)

In fact, the first and second modes do not have the same amplitude according to the
whale’s depth as illustrated in section 5.3. Let us consider the case when the source whale is located close to the sea surface or the sea floor so the second mode is more excited than the first mode \((M_2 > M_1)\). For this reason, the destructive interference does not occur as the maxima of the second mode do not cancel the minima of the first mode. For \(M_1 = 0.4\) and \(M_2 = 0.7\), the resultant signal \(m(t)\) is displayed in figure 5.35.

![Figure 5.35: An example of frequency with different amplitude](image)

**5.13 Chapter conclusion**

A FM chirp signal having the same duration and frequency range of Right whale up calls was simulated. The transmission of this signal in an acoustic channel having the same environmental parameters of Cape Cod Bay was modelled using PROSIM and a synthetic normal mode written in Matlab.

The modelling results shows that for a frequency range (50-200 Hz) of the source up call, the duration of the first mode in shallower water \(<60\, m\) is very sensitive to range change while this effect does not seem significant in deeper water \(>60\, m\). For a frequency range (100-250 Hz), the first mode duration is very small in deep water and this effect is small in shallower water at short ranges.

The dependence of the mode excitation on the source depth of the mode was investigated. The PROSIM output indicates that when the whale is located at mid water, the first mode
is more excited than the second mode or if the whale is located close to the sea surface or the sea bottom the first mode is less excited than the second mode. The implication of these results is that the relative excitation of the acoustic modes will assess the likely depth of the calling whale. This was applied to examples of Cape Cod Bay data to estimate the whale depth.

The good match between normal modelling synthetic data and the modes of a dispersive up call recorded in the margins of Cape Cod Bay indicates that medium sand is distributed in that region.

Using real data and the distances between the hydrophones in Cape Cod Bay, it was found that the TDOA between modes and the number of arrival modes increases with increasing range.

In order to estimate the acoustic range, two localisation techniques were used. The first relies on the TDOA between spatially distributed hydrophones and the second uses the TDOA between the arrival modes of Right whale up call at a single hydrophone. Both techniques show similar results.

Finally, the modal beating between modes 1 and 2 at higher frequency was investigated and found that such a phenomenon results in destructive interference.
Chapter 6: **Region-based segmentation using active contours**

### 6.1 Introduction

Key to this thesis is the detection of North Atlantic Right whale up calls in a dispersive shallow-water channel. This chapter looks at an image processing segmentation tool that will be used to isolate the dispersive Right whale up call from the background noise.

Digital image processing comprises a wide range of processing steps. Normally, the images of interest are acquired by using imaging sensors. In this thesis, the image under investigation is a spectrogram produced by using the short-time Fourier transform (section 3.2) where the value of the pixel intensity indicates the amplitude spectrum of the signal.

![Diagram](image.png)

Figure 6.1: Principal stages of digital image processing

The principal stages in the image processing are:

- After the original image is produced, stage (1) (see figure 6.1) will be used to enhance the image by reducing noise and increasing contrast. This comprises filtering in the spatial and frequency domains [102].

- The next step (stage 2) is referred to as “segmentation”. It aims to partition an input image into constituent objects. Segmentation is an application-dependent process so that it stops once the object of interest is isolated from the background. The output of such process can be a boundary or a complete region according to the type of problem being addressed [102].

- Quantifying such output using appropriate features is an important task to differentiate one class of the objects from another. This is performed in stage (3) “feature extraction”.

110
The final stage (4), recognition and interpretation, uses the information obtained from the extracted features to assign a “label” and then “understanding” to each object [102].

In this chapter, we will focus on the key segmentation stage.

6.2 Detection based on local and global properties

Techniques proposed to perform such a stage are commonly based on local properties (edge detectors that rely on intensity discontinuity) or global properties (region detectors that rely on the uniformity of the sub regions).

The discontinuity of the grey-level of the low-noise image can be detected by searching for the pixels of maximum local gradient in the image [102]. These pixels can also be located by computing the Laplacian operator but at the cost of unacceptable sensitivity to noise. To reduce such sensitivity, the Laplacian of a 2-D Gaussian (LOG) function was alternatively used. The LOG operator offers two main features [102]:

a) Gaussian function smoothes the image to reduce noise but at the cost of blurring the edges. The degree of smoothing is proportional to the standard deviation $\sigma$ of the LOG function.

b) zero-crossing process locates the desired edges by using the Laplacian operator

The convolution of the image with the appropriate masks is the most often used method to detect the desired edges. This operation is accomplished by sliding the mask (K) over the image and the value of the output pixel, at which the kernel is centred, is computed as a weighted sum of neighbouring pixels as illustrated in the example shown figure 6.2.
A range of common gradient masks are displayed in figure 6.3. These masks share the same properties: First, the elements in each mask sum to zero. This guarantees that convolving such mask with a constant value area results in zero. Second, the mask approximates differentiation in order to amplify the slope (the rate of change of the grey level) of the edge [103].

After convolution, thresholding the output image is often desirable to emphasize strong edges and remove weak edges.

Rather than detecting edges as above mentioned, target regions can be extracted directly using the procedure of region growing. It is based on merging pixels or small sub regions into larger regions [102]. This approach is simply implemented by using pixel aggregation that grows regions from seed points. Every pixel that satisfies a certain property is appended to the corresponding seed. In practice, the initial position of the seeds and the application-dependent segmentation criteria arise as fundamental problems with this approach.

Alternatively, the image can be subdivided into disjointed regions and merge and/or split the regions in order to satisfy the segmentation criteria [102].
Due to the discontinuity of the output boundaries resulting from using the traditional gradient-based segmentation and the high computation and complexity of region-based segmentation, active contours are proposed as presented next.

### 6.3 Active contours

In the last decade, the technique of active contours has been used for a range of applications particularly image segmentation and motion tracking. It is based on the evolution of a closed contour that moves within the image to fit the boundary of the object of interest. The two main mathematical approaches the model of active contours is based on are: snakes and level set. While the former moves the initial contour by minimizing the energy functional, the latter performs the task as a particular level of a scalar function (level set function).

The model of snakes was introduced by [104] for use in image segmentation. Its name “snake” came from the shape of the deformable contour during its movement within the underlying image. The closed contour (figure 6.4) is parameterized as:
The classical snake model is based on an energy functional associated with the curve $C$ [104].

\[
E(C) = \alpha \int_0^L \left| C'(s) \right|^2 ds + \beta \int_0^L \left| C''(s) \right|^2 ds - \lambda \int_0^L |\nabla I(C(s))| ds
\]

(6.2)

where $\alpha$, $\beta$, and $\lambda$ are real positive constants.

The basic idea of the model is to find the contour $C$ that locally minimizes $E(C)$. It is important to notice that two main forces control the model functional:

- the internal force consists of the first two terms. Minimizing the first-order term forces the contour to be continuous while minimizing the second-order term enforces smoothing of the contour during the evolution process.

- the external force is responsible for attracting the contour towards the maximum...
gradient of the images by minimizing the third term, i.e. maximizing the gradient of
the image intensity $|\nabla I|$.

To sum up, minimizing the energy functional is equivalent to moving the contour towards
the object boundaries while maintaining its smoothness.

The classic snake model requires the initial contour to be located close to the object
boundaries as it uses the local information of the image along the contour [105]. Also, a
smoothing operation is required to reduce noise in low SNR images. This also smooths
the edges and hence decreases the efficiency of the model. The level set method was
proposed as an alternative approach to overcome the above problems.

### 6.4 Level set method

The level set method is a powerful numerical technique used to track the propagation of
an interface evolving in the normal direction to itself with a known speed [106]. The
basic idea can be illustrated by defining a closed curve $C$ as the zero level set of a level
set function $\phi(x, y): \Omega \to \mathbb{R}$

$$C \equiv \{(x, y): \phi(x, y) = 0\}, \forall (x, y) \in \Omega$$  \hspace{1cm} (6.3)

where $\Omega$ denotes the image plane.

Figure 6.5: The zero level set of the level set function $\phi(x, y) = x^2 + y^2 - 1$
In other words, the function $\phi(x, y)$ is zero if the pixel with coordinates $x$ and $y$ lies on the curve $C$. For example, consider a two-dimensional level set function $\phi(x, y) = x^2 + y^2 - 1 = c$. For different values of $c \{-1, -0.5, 0, 0.5, 1\}$, a family of level sets is produced as a set of circles of different radii centred at the origin $(0,0)$; see figure 6.5. The green circle corresponds to the zero level set $\phi = 0$.

In Computer vision literature, the signed distance function is widely used to define the zero level function as illustrated in the next sub-section.

### 6.4.1 Signed distance function

The signed distance function $\phi$ is defined by [107] as:

$$\forall (x, y) \in \Omega; (x_c, y_c) \in C$$

$$\phi(x, y) = \pm \min \{|(x, y), (x_c, y_c)|\}$$

(6.4)

This function computes the distance between a point $(x, y)$ and the nearest neighbouring pixel $(x_c, y_c)$ lying on the initial contour $C$. The sign of this function alternates depending on whether its variable $(x, y)$ is inside or outside the contour as shown in figure 6.6.

$$\forall (x, y) \in \Omega,$$

$(x, y) \in \text{ outside } C, \phi(x, y) > 0$

$(x, y) \in C, \phi(x, y) = 0$

$(x, y) \in \text{ inside } C, \phi(x, y) < 0$
From figure 6.7, we can see that the sign distance map $\phi(x, y)$ is indicated by the colour bar where the pixel intensity is defined to be positive on the exterior, negative on the interior, and zero on the boundary. The most important property of this function is that its gradient satisfies the so-called Eikonal equation [107]:

$$ |\nabla \phi| = 1 \quad (6.5) $$

This property helps reinitialize the level set function to a signed distance function after some time of interface evolution as presented in section 6.5.3.
The geometric properties of the curve can be easily computed using the partial differential equations of the signed distance function. These properties include the unit normal vector and the mean curvature.

6.4.2 The unit normal vector and curvature

The unit normal vector to the level curve \( C \) through a point \((x, y)\) is defined as the gradient vector of the function defining the curve at this point. For a given curve defined by a two dimensional function \( \phi(x, y) \), the outward directed unit normal vector is given as [106]:

\[
n = \frac{\nabla \phi(x, y)}{\|
abla \phi(x, y)\|} = \left< \frac{\phi_x}{\left(\phi_x^2 + \phi_y^2\right)^{1/2}}, \frac{\phi_y}{\left(\phi_x^2 + \phi_y^2\right)^{1/2}} \right>
\]  
(6.6)

Where \( \phi_x = \partial \phi / \partial x \) and \( \phi_y = \partial \phi / \partial y \) denote the first-order partial derivatives of \( \phi \) with respect to \( x \) and \( y \) respectively.

Figure 6.8: \( k > 0 \) for convex region and \( k < 0 \) for concave region [107]

The mean curvature of the curve \( C \) is obtained by computing the divergence of the unit normal vector \( n \) and given by

\[
k = \text{div}(n) = \frac{\phi_{xx} \phi_y^2 - 2 \phi_x \phi_y \phi_{xy} + \phi_{yy} \phi_x^2}{\left(\phi_x^2 + \phi_y^2\right)^{3/2}}
\]  
(6.7)
where $\phi_{xx} = \frac{\partial^2 \phi}{\partial x^2}$ and $\phi_{yy} = \frac{\partial^2 \phi}{\partial y^2}$ denote the second-order partial derivative of $\phi$ with respect to $x$ and $y$ respectively, while $\phi_{xy} = \frac{\partial^2 \phi}{\partial y \partial x}$ denotes the mixed partial derivative of $\phi$ with respect to $x$ and $y$. The sign of $k$ tells whether the region is convex or concave as shown in figure 6.8.

### 6.4.3 The fundamental equation of level sets

The initial curve $C_0(s)$ describing the path of a point $s$ on the curve is built by a signed distance function $\phi(C(s,t), t = 0) = \phi(x(t), y(t), t = 0) = 0$. The motion of $C_0(s)$ in the direction of its outward normal vector $n$ with speed $V_n$ generates a family of curves $C(s,t)$ over time.

\[
\frac{\partial C}{\partial t} = V_n \cdot n
\]

\[
C_0(s) = C(s,t = 0)
\]

At time $t$, the curve $C(t)$ is given by the zero level set of the evolving function $\phi$ as:

\[
C(t) = \{(x, y) : \phi(x, y, t) = 0\}
\]

Using the chain rule, the time derivatives of both sides are:

\[
\frac{\partial}{\partial t}(\phi(C(s,t), t)) = 0
\]

\[
\frac{\partial \phi}{\partial t} + \frac{\partial \phi}{\partial C} \frac{\partial C}{\partial t} = 0
\]

The second term can be described as the dot product of the gradient of $\phi$ ($\nabla \phi = \langle \frac{\partial \phi}{\partial x}, \frac{\partial \phi}{\partial y} \rangle$) and the normal velocity vector $V_n \cdot n$ therefore; the above equation can be written as:
\[
\frac{\partial \phi}{\partial t} + \nabla \phi V_n \cdot n = 0 \tag{6.11}
\]

By substituting (6.6) in (6.11), the time-dependent level set equation is given:

\[
\frac{\partial \phi}{\partial t} + V_n |\nabla \phi| = 0 \tag{6.12}
\]

In their paper, Osher and Sethian have proposed that the curve evolves in the outward normal direction with the mean curvature-dependent speed \( V_n = -k = -\text{div} \left( \frac{\nabla \phi}{|\nabla \phi|} \right) \) so that the level set evolution equation becomes:

\[
\frac{\partial \phi}{\partial t} = \text{div} \left( \frac{\nabla \phi}{|\nabla \phi|} \right) |\nabla \phi| \tag{6.13}
\]

The evolution of the contour described by (6.13) has to stop when the contour reaches the desired boundaries of the object. Two main stopping functions have been proposed to perform this task. One is based on the edge information and the other relies on the region properties.

### 6.5 Edge-based active contour model

We have seen that the classic snake model depends on the maximum gradient of the image intensity (edge detector) to attract the curve towards the boundary. Such an edge detector (the external force in (6.2)) was replaced by a general edge detection function \( g(|\nabla I|) \) known as a stopping function to stop the evolving curve when it reaches the object boundaries. This function becomes zero at the edges and positive in the homogeneous regions.
\[
\lim_{|\nabla I| \to \infty} g(|\nabla I|) = 0
\]

\[
g(|\nabla I|) = \frac{1}{1 + |\nabla G_\sigma \ast I|^p}; p \geq 1
\]

The denominator of the function contains the term \( G_\sigma \ast I \) that indicates the convolution of the image \( I(x, y) \) with the Gaussian function \( G = \sigma^{-1/2} e^{-(x^2 + y^2)/4\sigma} \) in order to obtain a smoothed version of the image. This stopping function \( g \) with \( p = 2 \) and the mean curvature-based speeds were adopted by [108] to design a geometric active contour model with an evolution equation given by:

\[
\frac{\partial \phi}{\partial t} = g(|\nabla I|) \left( \text{div} \left( \frac{\nabla \phi}{|\nabla \phi|} \right) + v \right) |\nabla \phi|
\]

\[\phi(x, y, t = 0) = \phi_0(x, y)\]

where \( v \) is a positive constant and \( \phi_0(x, y) \) is the zero level set function.

The stopping function \( g \) was also used in the geodesic active contour model proposed by [109] to form the contour evolution equation

\[
\frac{\partial \phi}{\partial t} = \left( \text{div} \left( g(|\nabla I|) \frac{\nabla \phi}{|\nabla \phi|} \right) + v g(|\nabla I|) \right) |\nabla \phi|
\]

\[\phi(x, y, t = 0) = \phi_0(x, y)\]

Two main disadvantages of the edge-based active contour models can be noticed [110]:

1) The stopping function may not vanish (\( g = 0 \)) as the discrete gradients are bounded. This thus causes the curve to cross the boundary of the object.
2) Smoothing Gaussian process is required for noisy images but at the cost of smoothing the edges too.

To overcome the above problems, an alternative stopping function was proposed as illustrated in region-based active contour models.

6.6 Region-based active contour model

Unlike edge-based active contour models that rely on the maximum gradient of the image intensity to enforce the contour toward the boundaries, region-based active contour models are based on searching for the uniformity of the regions within the underlying image to indicate the presence of the boundaries. Also, the regularity of the evolving contour is kept by the curvature-dependent motion motivated by the level set function. The base model, proposed by [111], was based on minimizing the global energy functional $E^{MS}(f,C)$ to find an optimal piece-wise smooth approximation $f$ formed by smooth regions and sharp boundaries.

$$E^{MS}(f,C) = \mu \text{Length}(C) + \lambda \int_{\Omega} |I(x,y) - f(x,y)|^2 \, dx \, dy + \int_{\Omega \setminus C} |\nabla f(x,y)|^2 \, dx \, dy \quad (6.17)$$

The first term is responsible for smoothing the boundaries $C$ (geometric regularity). The second asks that $I$ approximates $f$ while the third term asks that the approximation $f$ varies smoothly within the subsets excluding the edges $\Omega \setminus C$. In this thesis, Chan-Vese active contour model was used. It is based on the particular case of Mumford and Shah Model [111] when $f(x,y) = \text{const.} = \text{average}(I(x,y))$. This model will be illustrated in detail in the next section.

6.6.1 Chan-Vese active contour model

Chan-Vese model is based on a region-based active contour which is represented by a zero-level set function $\phi : \Omega \to R$, i.e. $C = \{(x,y) : \phi(x,y) = 0\}$. The basic idea of this model can be illustrated by assuming an image $I(x,y)$ comprising two uniform regions having different distinct values. The object of interest is represented by a black disk with
a boundary $C_0$ as shown in figure 6.9.

The closed contour $C$ evolves toward the desired object boundary by maximizing the homogeneity of grey level value inside and outside the contour $C$. This can be achieved by minimizing the following fitting functional:

$$F(C) = F_1(C) + F_2(C) = \int_{\text{inside}(C)} |I(x, y) - u_1|^2 \, dx \, dy + \int_{\text{outside}(C)} |I(x, y) - u_2|^2 \, dx \, dy \quad (6.18)$$

Where $u_1$ and $u_2$ denotes the average value of the image intensity inside and outside $C$ respectively. In order to illustrate the relation between the contour location and the estimated value of the energy functional, four main scenarios are given:

a) If the curve $C$ is outside the object, $F_1(C) > 0$ and $F_2(C) \approx 0 \Rightarrow F(C) > 0$

b) If the curve $C$ is inside the object, $F_1(C) \approx 0$ and $F_2(C) > 0 \Rightarrow F(C) > 0$

c) If the curve $C$ is both inside and outside the object, $F_1(C) > 0$ and $F_2(C) > 0 \Rightarrow F(C) > 0$

d) If the curve $C$ is on the boundary of the object, $F_1(C) \approx 0$ and $F_2(C) \approx 0 \Rightarrow F(C) \approx 0$

The final case (d) states that the fitting energy is minimized when $C = C_0$

Chan and Vese have further developed the energy functional by adding a regularizing term that weighs the length of the curve $C$ by $\mu$ as follows:

$$F(u_1, u_2, C) = \mu \cdot \text{Length}(C) + \int_{\text{inside}(C)} |I(x, y) - u_1|^2 \, dx \, dy + \int_{\text{outside}(C)} |I(x, y) - u_2|^2 \, dx \, dy \quad (6.19)$$

The length parameter $\mu$ plays a scaling role in detecting objects with different size. For
example, $\mu$ should be small for detecting small objects while a large value of $\mu$ is required to detect large objects and remove small ones such as those due to noise.

Using the signed distance function, Heaviside function $H$, and Dirac impulse function $\delta_0$, the energy functional can be rewritten:

$$H(\phi) = \begin{cases} 1 & \text{if } \phi \geq 0 \\ 0 & \text{if } \phi < 0 \end{cases} \quad \text{(6.20)}$$

$$\delta_0(\phi) = \frac{d}{d\phi} H(\phi) \quad \text{(6.21)}$$
Since the length of the curve C has to be minimized when C fits the boundary of the
object i.e. \( \text{Length}(C) = \text{Length}( \phi = 0 ) \)

\[
F(C) = \mu \int_{\Omega} \left| \nabla H(\phi(x, y)) \right| dx dy + \int_{\Omega} \left| I(x, y) - u_i \right|^2 H(\phi(x, y)) dx dy + \ldots \tag{6.22}
\]

\[
\int_{\Omega} \left| I(x, y) - u \right|^2 (1 - H(\phi(x, y))) dx dy
\]

The associated Euler-Lagrange equation for \( \phi \):

\[
\frac{\partial \phi}{\partial t} = \delta(\phi) \left\{ \mu \text{div} \left( \frac{\nabla \phi}{\left| \nabla \phi \right|} \right) + (I(x, y) - u_i)^2 - (I(x, y) - u_2)^2 \right\} \tag{6.23}
\]

The constants \( u_i \) and \( u_2 \) are alternately updated with the evolution of the level set function

From the above equation, the evolution of the interface is controlled by two main parts: the first term (including the mean curvature) keeps the regularity of the curvature during the evolution process while the second and third terms search for the uniformity of the regions within the image.

![Figure 6.10: Narrow-band interface](image)

Due to the high computational cost for updating the level sets throughout the entire image space, it was proposed by [112] to compute the evolution of the level sets in the neighbourhood of the interface by defining a narrow tube \( M \) around the interface (figure 6.10) as:
\[ M = \{(x, y) \in \Omega : |\phi(x, y)| < \gamma\} \] (6.24)

### 6.6.2 Stability by the Courant-Friedrichs-Lewy condition (CFL condition)

Stability is an important issue to ensure that the small error resulted from finite difference approximation does not increase over processing time [107]. This requirement can be achieved if the speed of the numerical interface \( \frac{\Delta x}{\Delta t} \) is at least as the same as the physical interface’s speed \( \frac{\partial \phi}{\partial t} \)

\[
\frac{\max(\frac{\partial \phi}{\partial t})}{\Delta t} = \alpha \Rightarrow \Delta t = \frac{\alpha}{\max(\frac{\partial \phi}{\partial t})}; \Delta x = 1
\] (6.25)

The conservative value of the CFL number \( \alpha \) is commonly chosen to be equal to 0.5.

### 6.6.3 Reinitialization

After some time of evolution, the level set of \( \phi \) will be highly deformed so that \( \phi \) will not remain a signed distance function. To avoid this, the level set function is reinitialized during the evolution time to be a signed distance function, i.e. satisfying \( |\nabla \phi| = 1 \). This can be implemented by solving following partial differential equation until convergence [113]:

\[
\frac{\partial \omega}{\partial \tau} = \text{sign}(\phi)[|\nabla \omega| - 1]
\] (6.26)

\[ \omega(C,0) = \phi(C,t) \]

where \( \tau \) is an artificial time.

For better numerical results, \( \text{sign}(\phi) \) which is a sign function is defined as:
\[
\text{sign}(\phi) = \frac{\phi}{\sqrt{\phi^2 + (\Delta \tau)^2}}
\] 

(6.27)

If \( \phi = 0 \) \( \Rightarrow \omega(C, \tau) = \phi(C, t) \)

If \( \phi < 0 \) or \( \phi > 0 \), \( \nabla |\omega| = 1 \)

The signed distance function has to be smooth particularly for sampling and numerical approximation.

Finally, the flow chart of the Chan-Vese model is illustrated in figure 6.11.
Figure 6.11: Flow chart of Chan-Vese active contour model
6.7 Matlab experimental results

Let us present some examples of the proposed region-based active contour model that is based on Chan-Vese model as displayed in figure 6.12. Figure 6.12 (a) shows the initial active contour and the original image of a “black dog in the white snow”.

Figure 6.12: Detection of the “black dog in the white snow” using Chan-Vese active contour model
The figures listed on the right shows the segmented subsets corresponding to the new location of the contour during its evolution. The final stage of contour evolution indicates the detection of the black dog in the white snow; see figure 6.12(d).

Figure 6.13: “Galaxy with a bright core”. The results of region-based segmentation for different values of length parameters.

Figure 6.13 “Galaxy with a bright core” illustrates the influence of the length parameter $\mu$ on the results of region-based active contour segmentation. Figure 6.13(a) shows the initial contour and the original image of galaxy with a bright core while figures 6.13(b,c)
shows the segmented subsets corresponding to the final stage of the contour for two different values of the length parameter. We can see that setting the length parameter to a larger value removes small noise objects (figure 6.13(c)). These noise objects remain in the output segmented image when the length parameter is small (figure 6.13(b)).

### 6.8 Chan-Vese active contour model and spectrogram images

In the two previous examples, images obtained by cameras were used to test the technique of Chan-Vese model. In this thesis, we deal with spectrogram images produced by the short time Fourier transform (STFT). To evaluate the performance of the proposed technique, synthesized signals were evaluated. These signals include linear up chirp, linear down chirp, nonlinear up-down chirp, and nonlinear down-up chirp. After locating the initial square mask at the point corresponding to the maximum energy in the time-frequency plane, the results of the application of the Chan-Vese model on the synthesized signal are shown in figures 6.14-17. In each case the time evolution is from (a) to (d).

![Figure 6.14: Detection of linear up chirp using Chan-Vese model](image-url)
Figure 6.15: Detection of linear down chirp using Chan-Vese model

Figure 6.16: Detection of nonlinear up-down chirp using Chan-Vese model
When adding white Gaussian noise to the synthesized signal with different signal-to-noise ratio (SNR), the efficiency of the model works well even with low SNR signals. This is illustrated in figure 6.18.
To sum up, the figures from 6.14 to 6.18 illustrate that the proposed model works well on
synthesized spectrogram images illustrating various sweep types. The technique is applicable for use with a variety of signal types including nonlinear calls due to the flexibility of contour topology.

6.9 Conclusion

This chapter presents the theoretical aspects of an image segmentation technique. Traditional segmentation techniques are based on convolving a mask with the image followed by thresholding to emphasize the edges indicating the maximum gradient within the image. The quality of thresholding is based on the value of threshold level.

Due to the discontinuity of the output boundaries, active contour models were presented. The model adopted in this thesis is the level set-based active contour model where the contour is represented by zero level set. The level set function used here is the signed distance function. In this model the contour evolves within the image searching for the uniformity within the image to indicate the presence of the object.

The performance of this segmentation technique was evaluated using chirp signals with different sweep types. Results indicate that these signals were segmented well due to the flexibility of the contour topology. Also, the efficiency of the model works well with synthesized signal buried in white noise at low SNR.
In this chapter the discrimination between the Right whale up calls and background noise has been investigated using a classifier namely the Support Vector Machines (SVMs). In section 7.1 the original spectrogram is conditioned before the the region-based active contour segmentation method proposed in chapter 6 is applied. Real data recorded in Cape Cod Bay are used to evaluate the proposed method. In section 7.2 features will be extracted from the segmented spectrogram. The application of discrimination analysis using the SVMs [114] on Cape Cod Bay data is presented in section 7.3. In section 7.4 we show how the shallow-water dispersion effects which cause higher order mode generation causes variation in the parameters used for classification and descriptive statistics. We compare the descriptive statistics of the call duration using both the single mode and the multi-mode approaches. The single mode analysis was performed by extracting the frequency contour of the first mode.

It was found through observing the 120 calls that constitute the Cape Cod Bay acoustic data that 93% of the received Right whale up calls are dispersive. The first mode is less excited than the second mode in about 85% of the dispersive calls having a significant amount of time between modes. In the presence of a weak first mode we will introduce a technique that will use the excitation of the higher modes to help segment the dispersive up call within the average-filtered spectrogram. The region-based active contour segmentation method is used to isolate the dispersive up call from the background spectrogram without losing the information within the weak first mode.

7.1 Detection system

The region-based active contour model presented in chapter 6 is used to isolate the Right whale up calls from the background spectrogram. The detection system, shown in figure 7.1, illustrates the two main steps in this process namely spectrogram enhancement and spectrogram segmentation.
7.1.1 Input spectrogram

The spectrogram of each sound signal, sampled at 2 kHz, was produced using short time Fourier transform (STFT) with a frame size of 300 samples (150 ms). Each frame is weighted by a Hamming window to give a frequency resolution of 26.6 Hz (recall spectrogram bandwidth in section 3.2). In order to interpolate and smooth the signal’s spectral response, each frame was zero-padded to form an extended frame of 1024 samples. Consecutive frames were overlapped by 250 samples to give a time interval of 25 ms between frames; see figure 7.2.

![Figure 7.2: A normalised spectrogram of the original sound file. Spectrogram is normalized (linear)](image)

7.1.2 Spectrogram normalization and equalization

When the presence of noise is observed in recorded data, it is important to separate the signal from the background noise in order to better extract representative features from the subject call. Features have to be independent of background noise and hydrophone dynamics. Prior to feature extraction, a denoising method is normally proposed to
enhance the detection of the target signal and increase signal-to-noise ratio (SNR). The noise in the data may be due to a range of external sound sources that contribute to background noise in shallow waters. These include commercial shipping, industrial activities, wind noise, and biological sounds produced by marine animals sharing the same area [37].

The shallow-water noise varies spatially and temporally. A recent study indicates that the noise spectrum level varies between three different Right whale’s habitats including Georgia, Cape Cod Bay, and Bay of Fundy over different periods [41]; as shown in figure 7.3.

Figure 7.3: Mode spectrum level estimates of ambient noise in the 50-350Hz band. BOF= Bay of Fundy, Canada; CCB=Cape Cod Bay, Massachusetts; and GA=Georgia [41]

The noise sources may not share the same frequency range of the call; therefore using an appropriate filter to remove the unwanted components and pass the signal’s frequency content may be adequate.

On the other hand, if the background noise lies in the signal’s frequency range, using the above filter will not be effective as it emphasizes the frequency information of both the signal and noise; see figure 7.4.
Figure 7.4: Frequency filtering (a) Band-pass filter (75-220Hz); (b) Spectrogram of the filtered signal. Note that all frequency components including signal and noise components are emphasized at the same time. Spectrogram is normalized

As an alternative to frequency filtering, [115] proposed the noise spectrum subtraction technique in the time-frequency plane. This technique involves two main steps. Firstly, the median of initial and terminal quiet spectrogram blocks (which contain only noise) was computed to represent the average noise power spectrum; see figure 7.5.

Figure 7.5: Estimates of the noise power spectrum of the recording used in figure 7.2. This comprise 5% of the sound file. Power spectrum (linear)

Secondly, a multiple of the noise power spectrum was subtracted from each block spectrum and then negative values were set to zero. The multipliers required to implement this process can be fixed (empirically chosen) or adaptive (varying over spectrogram blocks). In the scenario of an adaptive multiplier, a vector of multipliers was computed
by dividing each bin in a block’s spectrum by the corresponding bin in the noise spectrum. Then, these multipliers were sorted and the value corresponding to a given percentile order statistic was chosen.

Because Ocean noise varies over time, initial and terminal blocks may not be adequate to produce an accurate representation of the average noise power spectrum. Also, higher values of the multiplier might remove parts of the weak first mode of dispersive calls. There is always a trade-off between reducing noise and removing useful details of the signal. We are thus motivated to use a time-varying conditioning technique where the value of the current spectrum in the frequency band is dependent on the previous values.

Such a time-varying conditioning technique has been used as a pre-processing stage for conditioning the sound spectrogram to be used in the detection [50] and classification [65] of marine mammals.

In this technique for a given input spectrogram $S(t, f)$, each frequency strip is exponentially smoothed using a low-pass filter to produce the time-averaged spectrogram $M(t, f)$ using the difference equation [50]:

$$M(t_0, f) = S(t_0, f)$$

$$M(t, f) = kS(t, f) + (1 - k)M(t - \Delta t, f) ; \ 0 < k < 1$$  \hspace{1cm} (7.1)

where:

$\Delta t$ = the time interval between two successive spectrogram blocks

$k$ = a smoothing factor $0 < k < 1$.

The transfer function of the infinite impulse response (IIR) filter used to perform this task is given by:
\[ H(z) = \frac{k}{1 - (1-k)z^{-1}} \]  

(7.2)

In this work, \( k \) was set to a very small value (0.001) to avoid losing information from the weak first mode in the next normalization stage.

The normalized spectrogram \( S_N(t, f) \) is computed by subtracting \( M(t, f) \) from \( S(t, f) \).

Finally, the equalized spectrogram \( S_E(t, f) \) is given by [50]:

\[ S_E(t, f) = \max(S_{\text{floor}}, \min(S_{\text{ceiling}}, S_N(t, f))) - S_{\text{floor}} \]  

(7.3)

The values of \( S_{\text{floor}} \) and \( S_{\text{ceiling}} \) are set to 0 and 1 respectively. The reason for setting the minimum value of \( S_E(t, f) \) to zero is to maintain the entire energy of the weak first mode and rectify the negative values. The spectrogram resulting from the equalization stage is shown in figure 7.6.

![Equalized spectrogram](image)

Figure 7.6: Spectrogram equalization and normalization of recording used in figure 7.2. spectrogram is normalized (linear)

Due to initial conditions mentioned in (7.1), normalizing and equalising the spectrogram offers beneficial features such as removing nearly constant noise signals of long duration in the same frequency band; see figure 7.7.
Figure 7.7: Example of removing signals of long duration using the spectrogram conditioning algorithm.
(a) Original spectrogram; (b) conditioned spectrogram. Spectrograms are normalized (linear)

7.1.3 2D average filtering

In acoustics, the beating phenomenon refers to the interference between two sound waves having slightly different frequencies (of the order of a few hertz). This scenario occurs in dispersive Right whale up calls mainly due to the intradispersion of the shallow-water environment which states that higher frequencies within the same mode propagate faster than lower frequencies. This effect causes the first and second modes to get close at higher frequencies and the constructive or destructive interference occurs according to the phase difference between them. The primary result of such an effect is that gaps between the modes are formed in the spectrogram image when the destructive interference occurs. This effect is illustrated in figure 7.8.

Figure 7.8: Zoomed-in image showing the beating effects between the first and second modes.
Spectrogram is normalized (linear)
In order to fill in the gaps; a linear filtering technique is applied to the spectrogram image using a 7×7 moving-average filter which spans 175 ms in time and 14 Hz in frequency. The size of the filter is dependent on the spectrogram’s parameters. For instance, higher percentages of overlap between successive blocks or longer Fourier transform window require a bigger filter size. The results of average filtering the spectrogram of the signal (figure 7.6) is shown in figure 7.9.

The 2D average filtering also helps overcome discontinuities in the calls. For example, the dispersive Right whale calls, displayed in figure 7.10(a) suffers from fading at mid frequencies. If the proposed segmentation method is applied to this spectrogram without 2D filtering, the resulting segmented spectrogram will comprise two objects, see figure 7.10(b). However, if 2D filtering is applied the segmented spectrogram will be similar to the expected target object, see figure 7.10(c).
7.1.4 Region-based active contours segmentation

We have previously found that filtering the spectrogram image with an average filter helps close the holes (gaps) resulted from mode-beating effects and overcome discontinuity on the call. Also, such a low-pass filtering process smoothes the sub-regions in the spectrogram image so that they have nearly uniform intensity. This thus makes it helpful to use the region-based active contour model (see the previous chapter) to segment the spectrogram image into regions and then identify the image region that corresponds to the Right whale’s up-call. The proposed segmentation method begins with creating an initial contour by using a mask centred at the pixel corresponding to the
maximum spectrum value in the time-frequency plane.

Figure 7.11: Detection the up-call with active contours without edge (Chan-Vese model). Left: spectrogram and contour evolution. Right: region-segmented spectrogram. length parameter=1

Having discussed the aspects of the detection system we now turn our attention to the specific application of Right whale vocalisations. We evaluate the performance of the proposed technique using real recorded data. The results of the spectrogram segmentation process associated with the evolution of the contour over time are shown in
The boundaries of the object are not defined by the gradient as in edge detection but are based on the uniformity of the spectrogram. Also, it is not a requirement for the initial contour to surround the entire target object. The initial contour can be anywhere in the image. However, in terms of processing speed locating the initial contour close to the object will accelerate the detection process.

Figure 7.12: Removal of strong medium noise objects

As well as the subject object (upcall), the segmented image may contain other small noise objects. Such objects can be removed by setting the length parameter $\mu$ of the Can-Vese model to a high value (see section 6.5). However, in the worst scenario where strong medium noise objects are detected, an area-based selection process is further applied to pick up the object of maximum area so that noise objects are removed; see figure 7.12.
At this point we have investigated the isolation of the Right whale up calls from background noise within the Cap Cod Bay data using the region-based active contour method. We now start with feature extraction as an essential stage for characterizing marine mammal’s vocalisations.

7.2 Feature extraction

7.2.1 Introduction

The type of features extracted is mainly based on the nature of the species’ sound. For instance, features describing tonal signals differ from those describing pulsed signals. These features should contain as much information as possible on the signal.

In previous works, features based on power spectra have been used for classification. This was performed by constraining the frequency range and time resolution in order to limit the feature space to a certain number of time-frequency pixels. The intensity values of such pixels were extracted from the spectrogram of Bowhead whale songs and Beaked whale clicks to be used as input to a neural network [57], [65]. Features can also be sequenced in time by picking the frequency bin corresponding to the maximum power spectra within a particular time window. This is known as peak frequency contour which is the most common spectral characteristic of tonal calls. The entire frequency contour was used to assess the similarity of Bottlenose dolphin whistles using Dynamic Time Warping (DTW) regardless of the vocalisation duration [56]. A certain number of peak frequency points were extracted from each whistle contour to build a correlation matrix on which Principal Component Analysis is conducted [116]. Based on the factor scores from each data set, K-means cluster analysis was used to group whistles into clusters. The fundamental frequency contour was tracked in Killer whale pulsed calls to develop a DTW classifier [55].

Rather than using the overall whistle contour, a range of specific acoustic parameters can be extracted from the contour in order to develop an appropriate classifier [30]. These parameters include beginning, end, minimum, and maximum frequency; duration; number of inflection points (defined as a change from positive to negative or negative to positive slope); number of steps (defined as a sudden jump in frequency over a short time period) ;
and presence/absence of harmonics. Also, the coefficients of a regression equation were used to parametrically represent the frequency contour of Southern Right whale calls for classification purposes [52].

Both spectral and temporal information can be used to characterize species vocalisations. The peak frequency and the duty cycle within a short-time window were extracted to characterize false Killer whale vocalisations for classification using neural networks [59]. Duty cycle refers to the percentage of time that a signal is “on” versus the total length of the signal. It provides an approximation of the type of waveform whether it is pulsed or continuous (whistles) [8]. Also, the times between consecutive zero crossing are extracted from Beaked whale clicks for classification using a SVM classifier [66]. Cepstral coefficients are extracted from Humpback whale song units to develop a SVM classifier [67].

As the peak frequency contour was commonly used to characterize tonal calls, for pulsed calls, the pulse rate contour indicating the change in the pulse repetition rate (spacing between frequency bands) is used [58]. The pulse-repetition rate is given by the autocovariance value within each time window. Clicks comprising a sequence of pulses spaced overt time were characterized by wavelet coefficients for classification using a nonlinear radial basis function network [117]. Also, clicks are identified using Wavelet transform [118], Hilbert-Huang transform [119], and Matching Pursuit [120].

Other approaches have dealt with the signal spectrogram as an image with pixel intensity indicating the spectral contents of the recording signal in the time and frequency plane. Gillespie [51] extracted the outlines of Northern Right whale sounds from a smoothed spectrogram using a threshold-based edge detector and used parameters obtained from these outlines to describe the sound. Harland et al. [121] converted the recording data to a binary spectrogram using a particular threshold and used a certain connectivity neighbourhood to form the target component within the spectrogram image. Parameters were extracted from these components to be used for classification.

7.2.2 Use of region extrema

After the target call has been isolated, a set of features are selected from the segmented
spectrogram to represent the data. In this work the extrema points of the region illustrated in figure 7.13 are used to indicate the parameters of the call.

Returning to the segmented spectrogram shown (figure 7.11(c)), the extrema points of the segmented region are used to extract the parameters of the upcall as shown in figure 7.14.

A number of calls parameter used for classification are selected for this work. These include minimum frequency, maximum frequency, and call duration, see figure 7.15. These features will be used to feed the classifier.
Figure 7.15: The call’s parameter extracted using the extrema of the segmented region (red points). The blue strip represents the estimated frequency contour of the first mode

7.3 The application of discrimination analysis on Cape Cod Bay data

In the literature of passive acoustic monitoring, several classification applications aim to differentiate between calls produced by different marine mammals or to differentiate between different vocalisations emitted by the same species. In this thesis the classification task will involve the discrimination between Right whale up calls and background noise using the SVM classifier [66].

The SVM is characterized by a high ability of generalization. It provides a global optimum solution to a quadratic programming problem rather than having a number of local minima. Also, the SVM controls the suppression of the outliers [114].

The low dimension of the features does not require a large amount of training data. This is appropriate to our data set which does not include a large number of Right whale calls. The extracted features for ambient noise data do not tend to cluster due to the inherent forces within the algorithm that create totally random shaped objects of noise that are quickly seen as non calls.
Figure 7.16: Scatter plot showing the distribution of (a) duration, minimum frequency, and maximum frequency in 3D space; (b) duration and maximum frequency; (c) duration and minimum frequency and (d) for minimum and maximum frequencies. Note that 120 Right whale up calls and 120 noise samples were used.

Figure 7.16 illustrates a scatter plot of these parameters on a sample of 120 Right whale up calls and 120 non Right whale calls within the Cap Cod Bay data. The blue dots represent the Right whale up calls while the red dots indicate the non Right whale calls. For the Right whale calls, the multiple-mode (dispersive call duration) process was applied on the vocalisations while for the non Right whale sounds the temporal extent of the segmented noise objects is displayed.

Each SVM was trained and tested using randomly selected data. The training data comprise 60 Right whale calls and 60 samples of ambient noise while the test data include the same number of Right whale calls and ambient noise samples (i.e. 60 for each). Note that the test data do not include the training data. The discrimination performance of the SVM was evaluated by the probability of correct classification (Pcc) which is defined as:

\[
Pcc = \frac{\text{number of test RW samples which are correctly classified as from RW class}}{\text{Total number of test samples from RW class}}
\]

This metric (average value) was calculated to be 93%. This discrimination analysis
represented by the block diagram displayed in figure 7.17 illustrates that the proposed technique has adequate discrimination between the Right whale up calls and general background noise in the Cape Cod Bay data set.

![Discrimination system block diagram](image)

**Figure 7.17: Discrimination system block diagram**

### 7.4 Variation of received calls in Cape Cod Bay

The environmental parameters of a shallow-water scenario control the number of the likely arrival modes and the time difference between modes (see sections 4.4 and 5.2). This consequently affects the entire duration of the dispersive received call. Also, the depth of the vocalising whale affects the relative mode excitation (see sections 4.2, 5.3 and 5.7).

The dispersive effect of the shallow-water environment influences the capability of parameters extracted from calls to characterise the species vocalisation. For example, the peak frequency contour that is used to characterise tonal calls is influenced by the depth of the calling whale in the dispersive channel. The extracted peak frequency contour from the dispersive call shown in figure 7.18(a) is presented in figure 7.18(b). Such a frequency contour becomes broken and jumps from one mode to another while tracking the frequency corresponding to the maximum energy of the signal over time. The maximum energy-based contour may follow more than one mode. This thus causes the
peak frequency contour to not only vary with species but also with the whale’s depth and range to the receiver.

Figure 7.18: (a) Spectrogram of a dispersive Right whale up call recorded in Cape Cod Bay, 2001; (b) The peak frequency contour extracted from the spectrogram shown in (a). Note that 1, 2, and 3 indicate the mode number. Spectrogram is normalized (linear)

Vocalisation duration is also influenced by the dispersive shallow-water environment. This parameters was extracted from Northern Right whale up calls by [51] for use in classification, see figure 7.19. The duration was also used for descriptive statistics of North Pacific Right whale calls in the eastern Bering Sea [88].
As illustrated in figure 7.20, the dispersive effect of the shallow water causes the duration $D_R$ of the received signal to be extended in time [88]. This parameter is greater than the duration of the source signal $D_S$. The variation of the duration of the received dispersive calls is mainly dependent on environmental parameters such as water depth, sediment type, and acoustic range.

In order to quantify the variation of the call duration caused by the existence of high-order modes, we thus compare it with the first mode duration. The first mode duration was extracted from the frequency contour corresponding to the points of the left edge of the segmented up-call between the left-top and the top-left points. To reduce the edge effects, a moving-window-average filter for a window size of 5 samples was used to produce the blue smoothed strip as shown in figure 7.15.
For comparison purposes, 120 calls of Cape Cod Bay acoustic data set were used for descriptive analysis. Figure 7.21 shows a comparison of the duration parameter for a single and a multimode analysis. In the single mode analysis we consider the first mode duration in figure 7.15. In the multimode analysis we use the entire duration of the dispersive call. The duration parameter results were consistently larger for the multimode signal analysis compared to the first mode extraction technique.

Figure 7.21: A comparison between the duration parameter extracted from the dispersive up calls and the first mode

This first mode extraction technique also affects the statistical description of the recorded data as would be presented by biologists. Figure 7.22 shows that the median value of the dispersive call duration is larger by approximately 0.15 s than that of the first mode.
In the discrimination analysis performed in section 7.3 the multiple mode process that uses the entire call’s duration was applied to the vocalisations for the Right whale calls. We also applied the single-mode process that uses the first mode duration to the Right whale data and a similar clustering results except for a change in the mean position on the duration axis (as expected from figure 7.22). This slight shift of the call duration does not lead to a significant improvement in the discrimination performance.

The other parameters used in the discrimination analysis were the minimum and maximum frequencies. In Cape Cod Bay (30 m), the cut off frequency of the first mode was computed to be 22 Hz [42] while the minimum frequency of a typical Right whale up call is 50 Hz. This thus shows that the minimum and maximum frequencies of Right whale up calls received in Cape Cod Bay are not influenced by the dispersive shallow-water environment. In other situations where water depth is much smaller for example in
the Georgia habitat of the Right whale where the water depth is about 15 m the modal propagation is limited below 60 Hz [41].

### 7.5 The application of the proposed technique on other recordings

Although this analysis has been specifically applied to Northern Right whale vocalisations recorded in Cape Cod Bay we consider the proposed technique equally applicable to other vocalisation produced by Right whales living in other habitats such as the Bering Sea and the Bay of Fundy. The application on upswept and downswep calls recorded in these environments is displayed in figures 7.(23-25). Also, the application of the proposed approach can be extended to other species such as Humpback whales which favour shallow-water environments.

The segmentation of Humpback recording spectrogram is shown in figures 7.(26-27). The STFT parameters used to produce the spectrograms in figures 7.(23-27) is listed in table 7.1.

Table 7.1: STFT parameters used to produce spectrograms in figures 7.(23-27)

<table>
<thead>
<tr>
<th>Frame length</th>
<th>Sampling rate</th>
<th>FFT size</th>
<th>Overlap</th>
<th>Habitat</th>
</tr>
</thead>
<tbody>
<tr>
<td>128 samples</td>
<td>1200 Hz</td>
<td>512 samples</td>
<td>78.13 %</td>
<td>Bay of fundy</td>
</tr>
<tr>
<td>128 samples</td>
<td>500 Hz</td>
<td>512 samples</td>
<td>78.13 %</td>
<td>Bering Sea</td>
</tr>
<tr>
<td>300 samples</td>
<td>4000 Hz</td>
<td>512 samples</td>
<td>75 %</td>
<td>Hawaii</td>
</tr>
</tbody>
</table>

Figure 7.23: (a) Spectrogram of a dispersive up-swept call produced by the North Pacific Right whale in the Bering Sea, Alaska (b) Segmented spectrogram. Spectrogram is normalized (linear)
Figure 7.24: (a) Spectrogram of a dispersive down-swept call produced by the North Pacific Right whale in the Bering Sea, Alaska (b) Segmented spectrogram. Spectrogram is normalized (linear)

Figure 7.25: (a) Spectrogram of a dispersive up-swept call produced by the North Atlantic Right whale in the Bay of Fundy (b) Segmented spectrogram. Spectrogram is normalized (linear)

Figure 7.26: (a) A spectrogram of a Humpback whale song recorded off the north coast of the island of Kauai, Hawaii (b) the segmented spectrogram. Spectrogram is normalized (linear)
The results of figures 7.23-25 show that the proposed segmentation technique isolates the dispersive up- and down- swept calls produced by North Atlantic Right whales in different habitats. In figures 7.26 the first harmonic of the Humpback song was isolated while three harmonics were discriminated from background noise in figure 7.27.

### 7.6 Conclusion

In this chapter the discrimination of dispersive North Atlantic Right whale from background noise was proposed. This approach is based on a region-based active contour model presented in chapter 6. The detection system comprises two main stages: Spectrogram enhancement and spectrogram segmentation. In the first stage, spectrogram was normalized and equalized in order to remove nearly constant noise signals of long duration in the same frequency band. Next, 2D average filtering was applied to fill in gaps resulting from acoustic beating between modes higher frequencies and to overcome the discontinuity in the calls.

After filtering, the smoothed spectrogram is ready for segmentation using the region-based active contour model. From the segmented spectrogram three features including minimum frequency, maximum frequency and the entire dispersive call duration were extracted from 120 Right whale up calls and 120 non Right whale calls. The scatter plot of the features shows that Right whale data tends to cluster in the feature space while the noise data spread randomly in the 3D space. This is due to the inherent forces within the
algorithm that create totally random shaped objects.

The features of 60 samples of Right whale data and 60 were randomly selected to train the Support Vector Machine classifier (SVM). The same number of data but with different samples was used to test the classifier. Training and testing stages were repeated many times and the average probability of correct discrimination was calculated to be 93%.

In order to quantify the variation of the duration parameter, it was found that the duration parameter was larger for the multimode calls compared to the first mode analysis. Statistically, the median value of the dispersive call duration is larger by about 0.15 seconds than that of the first mode. Biologists should consider the influence of the channel when they statistically describe the extracted parameters from data recorded in dispersive shallow-water environments.

Using the first mode duration instead of the entire dispersive call duration in the discrimination analysis does not lead to a significant improvement. The minimum and maximum frequencies are not influenced by the channel as they are higher than the cut off frequency of the first mode (22 Hz).

It was found that this technique is equally applicable to other marine mammals which favour shallow water environments such as Bowhead whales and Humpback whales.
Chapter 8: Conclusion and future work

8.1 Conclusion

The objective of this thesis was to establish an automatic technique to detect, characterize and discriminate North Atlantic Right whale up calls from background noise in a dispersive shallow-water environment. This work can hopefully be incorporated into existing software used in both mitigation measures to protect such species and in species monitoring. The results of the proposed technique show that variation of the call’s parameters due to the influence of the dispersive channel does not significantly affect the discrimination performance.

(i) The problem

Shallow-water acoustic environments act as a waveguide with multiple reflections off the sea surface and the sea bottom resulting in multipath effects. A primary effect of such multiple paths is to cause distortion on the Right whale vocalisations received at the hydrophones.

The typical Right whale up call is a mono-component signal with a frequency up-sweep from 50 to 200Hz with a duration of 1 second. In Cape Cod Bay, the Northern Right whale feeding ground, a recent study showed that the minimum frequency of up calls has been shifted up to be about 100 Hz due to the presence of band-limited noise. In such environments dispersive up calls are normally received as a multi-component signal comprising multiple modes. Each mode is received at different times with different relative modal energy according to the depth of the vocalising whale.

(ii) Signal processing tools

The first step was to choose an appropriate signal processing tool in order to represent the dispersive up calls in the joint-time frequency domain. A range of tools including short-time Fourier transform (STFT), Wavelet transform (WT), Wigner-Ville distribution (WVD) and Hilbert-Huang transform (HHT) were investigated using simulated signals
based on Right whale up calls. Results indicated that the STFT is the appropriate tool to represent such calls.

(iii) The theoretical study of shallow-water waveguide

In order to illustrate how the shallow-water waveguide affects the signal’s dispersion, the dispersion curves that relate the acoustic frequency to the group velocity of the likely propagating modes were produced. Also, normal mode modelling was selected as an appropriate model to describe low-frequency propagation in shallow water. The model inputs were set to the appropriate values of the environment of Cape Cod Bay. These values include water depth (from Bathymetry), sound velocity profile (SVP), sediment type, and receiver configuration.

The influence of water depth on the Right whale up calls was investigated. Results indicate that deeper water channels support higher number of propagating modes than shallower waters in the frequency range of the Right whale. This is because the shallow water environment acts as a high pass filter with a cut-off frequency that is inversely related to the water depth. The filter of Cape Cod Bay (30 m) allows the entire Right whale up call to pass through as the cut off frequency of the first mode is lower than the minimum frequency of the Right whale call.

Investigation of effect of water depth on the FM deviation and duration of the first mode and the TDOA between modes one and two is discussed. Results show that the influence of water depth on the first mode’s FM deviation is small even at long ranges. Also, the effect of water depth on the first mode duration is controlled by the minimum frequency of the source signal and the acoustic range. Because of the up-shift of the minimum frequency of Right whale calls due to ambient noise in Cape Cod Bay, the influence of water depth on the first mode duration does not seem significant at short range.

As the approach of acoustic range estimation proposed by [89] is mainly based on the time difference of arrival (TDOA) between modes, the influence of water depth on such a parameter was investigated. Results show that shallower channels support bigger TDOA
between modes one and two than in deeper channels.

Next the distribution of the sediments in the Cape Cod Bay motivated us to investigate the influence of sediment type on received dispersive up calls. Results show that firstly sediments of higher acoustic impedance increase the probability of acoustic detection since they support lower propagation loss at long ranges than those of lower acoustic impedance. Secondly, the effect of sediment type on the first mode is very small. Thirdly, water depth controls the influence of sediment type on the group velocity difference between the first two modes. This latter is sensitive to sediment type change in shallower waters (<60m) while this effect does not seem significant in deeper waters (>60m). Also, it was noted that sediments of high acoustic impedance support higher number of propagating modes than that of low acoustic impedance. Finally, the results of using real data and synthetic normal mode modelling in the margin of Cape Cod Bay assumed medium sand to be distributed in the margin.

The focus then shifted to illustrate how the depth of the calling whale excites the individual normal modes of the received up calls. This revealed that if the first mode was more excited than the second one the whale is likely to be located at mid water; or alternatively if the second mode is more excited than the first the whale is likely to be close to either the sea surface or the sea bottom. One implication of such a principle is that the information obtained from the mode excitation can be used to assess the likely depth of the source whale. Examples of synthesized and real data have been presented to illustrate this concept.

Previous work showed that information obtained from dispersion effects of a Right whale shallow-water habitat (the Bering Sea) can be used to estimate the distance between the vocalising whale and the receiver. This requires measurements of the time difference of arrival (TDOA) between modes received at single hydrophone and the group velocity of the arrival modes using the dispersion curves of the target channel. Similarly, we have applied this approach to Cape Cod Bay data. Results agreed with the distance calculated using hyperbolic localisation technique that is based on the TDOA between multiple hydrophones.
After a sound understanding of the effects of the dispersive shallow water on Right whale vocalisations, we turned our attention to the detection of the dispersive up call within the background noise. However, the visual scanning of the majority of the Right whale up calls recorded in Cape Cod Bay indicates that the first mode is less excited than the second mode. Therefore; an appropriate segmentation method was required to detect the call without losing the information within the weak first mode.

**(iv) The novel region based active contour technique**

Prior to detecting the dispersive call within a background spectrogram, the technique of spectrogram normalisation and equalisation was applied. Next, in order to remove the destructive interference effects resulting from the beating between modes one and two at higher frequencies and to overcome discontinuity on calls, a 2D average filtering is performed. This low-pass filtering causes the sub regions of the spectrogram image to have nearly uniform intensity. A region-based active contour technique based on searching for the uniformity of the regions within the underlying image was used. The performance of the segmentation technique was evaluated using both synthesized data based on typical Right whale vocalisations and real recorded data from Cape Cod Bay.

After segmentation, strong noise objects might be detected. Therefore; an area-based selection process is further applied to pick up the object of maximum area so that noise objects are removed.

Next the extrema points of the segmented spectrogram were used to extract the parameters of the Right whale up call. These include the minimum and maximum frequency and entire dispersive call’s duration which will be used to feed the classifier.

**(v) Right whale discrimination analysis**

In this thesis we do not aim to perform classification between calls produced by different marine mammals or to differentiate between different vocalisations emitted by the same species. However, the main classification task involves the discrimination between Right whale calls and background noise using the Support Vector Machine SVM classifier. The
SVM classifier has been trained with randomly selected data including Right whale up calls and background noise. After testing the trained classifier with a test data, the discrimination performance of the SVM was calculated to be 93%. This result indicates that the proposed technique adequately discriminates Right whale up calls from background noise.

(vi) Variation of received calls in Cape Cod Bay data

The dispersive shallow water causes the duration of the received up call to extend in time. This thus suggests that this parameter does not truly represent the vocalisation duration. A comparison between the duration for the dispersive call and the first mode was performed. Results reveal that the duration parameter was larger for the former compared to the latter. The median value of the dispersive call duration is larger by approximately 0.15 s than that of the first mode. The variation in the duration parameter does not improve the discrimination performance.

Due to the dependence of the mode excitation on the whale’s depth, it was found that the peak frequency contour of the received dispersive up call is unable to characterise the source vocalisation. Finally, the cut-off frequency of the high-pass filter of the Cape Cod Bay environments was found to be less than the minimum frequency of the typical Right whale. Therefore the parameters including the minimum and maximum frequencies of the received call can truly represent the whale vocalisation.

8.2 Future work

This thesis has focused on investigating the influence of shallow-water environments on Right whale up calls. This investigation can be extended to include other North Atlantic Right whale shallow-water habitats such as Georgia coast (~15 m) and Bay of Fundy (~150 m) using real recorded data. The dispersion information of these acoustic channels can also be used to estimate the acoustic range after obtaining information about the bathymetry, sound speed profile, and receiver configuration of these habitats.

In chapter 5 the match between the the synthetic normal mode modelling and read data is considered as an inspiration to further develop this piece of work. This can be done by
building a lookup table (LUT) of normal mode models produced for different parameters of water depth, sediment type and range. An automated matched filtering between real data and one of the models will be performed. The results of such matching will reveal an estimation of water depth, sediment type and range.

This work has also developed a novel approach to detect and discriminate the North Atlantic Right whale up calls from the background noise in Cape Cod Bay using region-based active contour segmentation technique. This approach can be further applied to other species vocalising in shallow waters such as Humpback whales.