Detection, localization, and trajectography of marine mammals

Bénard Frédéric

A dissertation submitted in partial fulfillment
of the requirements for the degree of

Doctor of Philosophy

Université de Sud-Toulon-Var

2011

Program Authorized to Offer Degree: your program here

Université de Sud-Toulon-Var
Graduate School

This is to certify that I have examined this copy of a doctoral dissertation by

Bénard Frédéric

and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final
examining committee have been made.

Chair of the Supervisory Committee:

Name of Chairperson

Reading Committee:

Name of Committee member

Name of Committee member

etc

Date:

Université de Sud-Toulon-Var

Abstract

Detection, localization, and trajectography of marine mammals

Bénard Frédéric

Chair of the Supervisory Committee:
Professor Name of Chairperson
Chair's department

It is difficult to study the behavior and physiology of marine mammals or to understand and mitigate human impact on them because much of their lives are spent underwater. Since sound propagates for long distances in the ocean and since many cetaceans are vocal, passive acoustics is a valuable tool for studying and monitoring their behavior. After a brief introduction to and review of passive acoustic tracking and detection methods, this dissertation develops and applies different methods for detection and trajectography. The detection methods are based on the frequency and energy signal properties using tools such as spectrograms, energy detectors and stochastic matched filter depending on the signal characteristics. The localization method use a widely-spaced (several kilometers) bottom-mounted hydrophone arrays,. and relies on arrival times of direct and surface-reflected paths. It is used to track a sperm whale using 5 hydrophones at the Atlantic Undersea Test and Evaluation Center (AUTEC) and gives position estimates that are accurate to within 10 meters. With such accuracy, the Inter-Click-Interval energy modulation and the Inter-pulse-interval can be estimated and then give information on the individual size and age. Moreover, The trajectography method use a short-base hydrophone array (Nemo array) and allow to determine the position of the whale with a hierarchical particle filter algorithm. Statistics on larger data provide information on migration behavior and group density. The performances of the algorithms are computed and compared to the results on the field.

TABLE OF CONTENTS

LIST OF FIGURES

GLOSSARY

ARGUMENT: replacement text which customizes a LaTeX macro for each particular usage.

ACKNOWLEDGMENTS

The author wishes to express sincere appreciation to University of Washington, where he has had the opportunity to work with the TEX formatting system, and to the author of TEX, Donald Knuth, *il miglior fabbro*.

DEDICATION

to my dog

Chapter 1

INTRODUCTION

The past several decades have seen increased concern and controversy over the impact of human activity on marine mammal welfare [Green et al. 1994; Richardson et al. 1995; Malakoff 2001; NRC 2003; NRC 2005]. Anthropogenic noise in the ocean includes sound from shipping, naval operations, and seismic exploration. In response to noise, marine mammals change vocalizations rates, alter habitat use, move away from the source, lengthen songs, change respiration patterns, and possibly strand [Richardson et al. 1995; Frantzis 1998; Miller et al. 2000; Anon. 2001; Caldwell 2002; Jepson et al. 2003;Gordon et al. 2004; Taylor et al. 2004]. Among other things, the response is influenced by source level and frequency characteristics, sound propagation conditions, and the sensitivity of the animal [Richardson et al. 1995; DeRuiter et al. 2006; DSpain et al. 2006;]. Since sound can propagate for long distances without suffering much attenuation, animals may be affected tens, hundreds, and even thousands of kilometers away from a source [Greene and Richardson 1988; Bowles et al. 1994; Nieukirk et al. 2004; Madsen et al. 2006]. In addition to short-term effects, long-term increases in ocean ambient noise [Curtis et al. 1999; Andrew et al. 2002; McDonald et al. 2006], potentially degrade habitat by masking and interfering with acoustic signals that are used for communication, orientation, navigation, and detection of predators and prey [Payne and Webb 1971; Au 1993; Tyack and Clark 2000; Clark and Ellison 2003]. Methods used to study and monitor marine mammals in the wild include multisensor archival tags, visual surveys, and passive acoustics. Tags can provide detailed information on animal depth, orientation, physiology (including heart rate and breathing), and vocalizations [Schevill and Watkins 1966; Leatherwood and Evans 1979; Mate 1989; Goodyear 1993; Johnson and Tyack 2003]. They have facilitated several major advances in our understanding of the impact of noise on marine mammals as well as in behavioral studies and in bioacoustics [Watkins et al. 1993; Fletcher et al. 1996; Miller et al. 2004; Zimmer et al. 2005a; DeRuiter et al. 2006; Tyack et al. 2006; Watwood et al. 2006; Stimpert et al. 2007]. Disadvantages to tagging include logistical problems (tags are expensive and can be difficult to place on an animal), possibly altered behavior, and limited attachment time [Whitehead et al. 2000]. Furthermore, because they require that an animal be tagged, they cannot be used to detect animals for mitigation purposes. Most mitigation measures rely on trained visual observers aboard vessels who scan the sea surface for the presence of marine mammals [reviewed in Barlow and Gisiner 2006; Weir and Dolman 2007]. Visual methods play a key role in many cetacean studies and include photo identification [first described by Wrsig and Wrsig 1977; reviewed in Hammond et al. 1990], as well as aerial [e.g. Watkins and Schevill 1979; Wrsig et al. 1984, 1993; Scott and Perryman 1991; Mobley et al. 1999; Mobley 2005, 2006], ship/boat-based [e.g. Johnston et al. 2007; Williams and Thomas 2007], and ground based studies [e.g. Wrsig and Wrsig 1979, 1980; Clark and Clark 1980; Tyack 1981; Noad and Cato 2007].

Limitations include sea state, daylight, and the amount of time an animal spends near the surface (as little as 5% of the time for some deep diving species [Barlow 1999]). Passive acoustic monitoring (PAM) methods can be used to detect animals that are submerged at any time of day, in poor weather conditions, and they are used extensively in studies of marine mammal behavior and movement [e.g. Leaper et al. 1992; McDonald et al. 1995; Stafford et al. 1998; Au et al. 2000; VanParijs et al. 2002]. PAM is limited to vocalizing animals and although most cetaceans are vocal, they may be silent for long periods of time and may silence in response to noise [Watkins and Schevill 1975]. PAM is also hindered by the incomplete repertoire representations for some species [Dawbin and Cato 1992; Mellinger et al. 2000]. The complementary nature of tagging, visual, and acoustic methods means that they can be especially useful in combination. For example, Zimmer et al. [2005a] used visual sightings and a towed hydrophone system to estimate echolocation source characteristics from a tagged sperm whale. Vocalizations recorded from tagged whales improve PAM capabilities by adding to the known repertoire [Johnson et al. 2004; Stimpert et al. 2007]. Combining visual and acoustic detection methods can improve tracks and increase the detection probability, although methods to relate acoustic and visual detection statistics to the true population need further development [Ko et al. 1986; Frankel et al. 1995; Noad and Cato 2001, 2007; Clark and Fistrup 1997; Tiemann et al. 2006]. PAM is useful on its own for census efforts and behavioral studies, particularly for continuous, long-term monitoring [Clark and Ellison 1988; McDonald et al. 1995; Stafford et al. 1998; Norris et al. 1999] and in hostile or inaccessible areas [Wartzok et al. 1992]. PAM can also be used for bioacoustics research on free-ranging animals [Au et al. 1974, 1986, 1987, 2002, 2006; Mhl et al. 1990; Thode et al. 2002; Wahlberg 2002; Au and Benoit-Bird 2003; Au and Herzing 2003; Rasmussen et al. 2004; Madsen and Wahlberg 2007]. Critical information regarding the biosonar of marine mammals has been derived from tests with trained or captive animals [reviewed in Au 1993; Thomas et al. 2004], but such studies are limited to smaller species, and biosonar performance might differ for free-ranging animals [Au et al. 1974, 2004; Au and Herzing 2003; Madsen et al. 2004a,b].

Chapter 2

TECHNICAL INTRODUCTION

2.1 Localization overview

Passive acoustic monitoring refers to the use of acoustic signals to detect, classify, and localize calling animals. Detection and classification often require sophisticated techniques, and the development of these methods is an active area of research [Altes 1980; Mellinger and Clark 1993; Potter et al. 1994; Mellinger 2000, 2004; Chesmore 2001; Oswald et al. 2003, 2007; Gillespie 2004; Roch et al. 2007]. Except for the detection algorithm used in the first two papers of this dissertation, detection and classification are not dealt with here. Passive acoustic localization (estimation of position) and tracking (taking positions estimates over time) are also active areas of research, and they are the topic of this dissertation. The use of acoustics to track marine life was pioneered in the 1960s and 1970s. Walker [1963] tracked sources of 20-Hz pulses, apparently from a whale, using three hydrophones. Cummings et al. [1964] found the positions of fish and invertebrates, also using three hydrophones. Watkins and Schevill [1972, 1977] used a three-dimensional array with four hydrophones to track the movements of individual whales. Since that time, many studies have used a variety of passive acoustic localization methods and hydrophone configurations to track marine mammals [e.g. Cummings and Holliday 1985; Mhl et al. 1990; Freitag and Tyack 1993; Clark et al. 1994; Stafford et al. 1998]. A commonly used method of passive acoustic localization is known as the timedifference of arrival (TOAD) method (also known as multilateration or hyperbolic positioning/fixing). TOAD methods are useful in a broad range of applications: civilian and military applications to locate aircraft, submarines, ground vehicles, and stationary sources such as explosions, geophysical applications to monitor seismicity [e.g. Fox et al. 1995], terrestrial biological applications to estimate animal positions [e.g. Mennill et al. 2006]. TOAD methods have been used to track almost every imaginable source of sound, from human speakers to gunshots [Lahr and Fischer 1993; Vermaak and Blake 2001; Bucher and

Misra 2002]. In the TOAD method, a signal reaches two spatially separated receivers at different times because of different propagation path lengths from the source to the receivers. For known receiver positions, the locus of possible source locations is a hyperboloid. A third receiver provides another TOAD measurement, which defines a second hyperboloid and a line of possible source locations is defined by the intersection of these two hyperboloids. A fourth receiver defines a third hyperboloid, with the intersection of all three hyperboloids defining a point, which is the estimated source location. Note that each additional receiver actually adds as many TOADs (hence hyperboloids) as there were receivers, but only one new TOAD is unique. Also, depending on the receiver configuration, the intersection of the three hyperboloids may be two points (or infinitely many points in a degenerate case), in which case a fifth hydrophone is required to localize a source in three dimensions [Tyrrell 1964; Spiesberger 2001]. The first step in TOAD methods is to estimate the signal time delay between each pair of hydrophones. The most commonly used method is correlation, in which the estimate is the time-lag that maximizes the cross-correlations between the received signals [Knapp and Carter 1976]. For marine mammal applications, both filtered waveforms [Clark et al. 1986; Spiesberger and Fristrup 1990; Mitchell and Bower 1995; Janik et al. 2000; Tiemann et al. 2004] and spectrograms [Altes 1980; Clark et al. 1986; Frankel et al. 1995; Janik et al. 2000; Clark and Ellison 2003; Tiemann et al. 2004] of the recorded signals have been used in the cross correlation. TOADs can also be estimated by using a matched-filter approach if a template of the call can be estimated [Stafford et al. 1998]. The best method to use to estimate TOAD varies depending on the signal, noise, and propagation characteristics of the problem. The second step in the TOAD method is to find the point of intersection (or the closest such point if intersection is imperfect) of the hyperboloids. Assuming constant speed of sound propagation, the problem can be expressed as a system of linear equations. For a well-defined problem (not underdetermined/overdetermined by too few/many receivers), a closed form solution to this system gives the source location [e.g., Schmidt 1972; Watkins and Schevill 1971; Schau and Robinson 1987; Delsome et al. 1980; Spiesberger and Fristrup 1990; Wahlberg et al. 2001]. For overdetermined systems, a least-squares approach can be used to give the best source position [Spiesberger and Fristrup 1990; Wahlberg et al. 2001]. Reflections from the bottom and surface can be treated as recordings made by virtual receivers and incorporated into the solution [Urick 1983; Mhl et al. 1990; Aubauer et al. 2000; Wahlberg et al. 2001]. Using reflections improves the accuracy of estimated source positions [Mhl et al 1990; Wahlberg et al. 2001; Thode et al. 2002] and fewer real receivers (as few as one) are required for localization [Aubauer et al. 2000; Tiemann et al. 2006; Laplance 2007]. Error analysis [reviewed in Taylor 1997] can be approached by comparing locations obtained by different receiver subsets, or by comparison with positions determined visually [Cleator and Dueck 1995; Smith et al. 1998; Aubauer et al. 2000; Janik et al. 2000]. Another approach involves linear error propagation and considers uncertainties in sound speed, receiver position, ray bending, and TOAD measurement [Watkins and Schevill 1971; Spiesberger and Fristrup 1990; Wahlberg et al. 2001]. The TOAD method can be used with different hydrophone configurations, each with its own strengths and weaknesses, to meet various monitoring and research requirements. For example, a closely spaced planar array is better for echolocation research [Au et al. 2002; Rasmussen et al. 2002; Au and Herzing 2003], but it is limited to short distances and animal positions that are perpendicular to the plane of the array. A widely spaced array is ideal for tracking in three dimensions over long distances [Stafford et al. 1998; Tiemann et al. 2004], but applications are limited to sufficiently loud calls (so that they can be heard on several hydrophones) and for logistical reasons (e.g. cost, clock synchronization, and hydrophone position uncertainties). As refraction becomes significant at long distances, widely spaced arrays are also more sensitive to assumptions of straight-line sound propagation, and sound propagation models may be required to obtain accurate position estimates [Chapman 2004; Tiemann et al. 2004]. A somewhat different approach, which is a version of beamforming [Johnson and DeGraag 1982], is required to localize a source using a towed line array [Leaper et al. 1992; Gillespie 1997; Barlow and Taylor 2005]. This method estimates the bearing to a sound relative to the tow cable axis from the TOAD measured between two hydrophones spaced a few meters apart by assuming plane wave propagation. Range is estimated via a time-motion analysis of the changes in estimated bearing as the platform moves. The method requires that the speed of the vessel be much greater than the speed of the vocalizing animal, that the animal vocalizes continuously for several minutes, and that individuals vocalizing simultaneously can be distinguished conditions that are not always met [Thode 2005]. Furthermore, this method does not distinguish between horizontal and vertical range. Time of arrival is not the only component of a recorded call that contains information about animal location; phase and sound pressure levels can also be useful. Directional hydrophones can be used to estimate the direction to high-frequency vocalizations [Whitehead and Gordon 1986] and for hydrophones configured less than a wavelength apart, differences in phase can be used to estimate the bearing to animal calls [Clark 1980]. Cato [1998] proposed a method using only the differences in received levels to estimate position, although we note here that this method might not be applicable for animals with high directionality. If the source level of the call is known, then it may be possible to estimate the distance to an animal from the received level. However, data on source levels are scarce, individuals can vary their source levels, and directionality will again confound the problem. Matched-field processing (MFP) uses all available information (timing, sound pressure level, and phase) to estimate source position (among other applications). The underwater acoustics community developed MFP in the 1980s and 1990s for naval purposes. Possibly because implementation can be costly and computationally demanding, MFP has seen only limited application in marine mammal localization problems [Thode et al. 2000; Tiemann et al. 2004; Thode 2005]. Very briefly, localizing a source via MFP involves predicting the receiver response given a source at some candidate source position and then comparing the predicted response with the measured response. This process is repeated for each point in a grid of candidate source positions, and the candidate position giving the best agreement between predictions and measurements is chosen as the estimated source position. Details can be found in reviews by Tolstoy [1993] and Baggeroer et al. [1993] and the references therein. Since MFP was designed to exploit the phase information, it traditionally requires large line arrays and low frequency sources (higher frequencies become incoherent for long distance propagation). Although passive acoustic methods for marine mammal localization yielded much useful information over the past 30 years, there is still a need for improvement. Most effort has focused on short duration calls (such as clicks) because arrival time estimates are relatively simple in these cases, while longer duration calls have reflections overlapping direct arrivals. There is a need to improve methods for localizing long duration calls made by many species (baleen whales in particular). Also, methods that increase the accuracy of estimated animal positions, for example by using more realistic sound propagation models, will facilitate behavioral studies by providing more detailed information on animal movement. Another important goal of passive acoustics is to track multiple individuals simultaneously for insight into marine mammal communication and interaction. Real-time methods that enable continuous, long-term monitoring are the long-term goal.

2.2 Data presentation

2.3 AUTEC

2.4 Summary of the contribution

2.5 Presentation of the studied species (finwhale, spermwhale, click...

The species present in mediterranea are the finwhale, spermwhale, bottlenoise dolphine, globicephalus, Risso dolphin. The frequencies domains of the principal spieces are in Tab.2.2. In this dissertation we are particulary interested in finwhale and sperwhale signals.

Outside of mediterranea, other spieces are of interest:
-the southern right whale which emits sound between 50 and 500Hz and can reach 2,2 kHz.
-Blue whale which emits low frequencies growls (10-40 Hz).

2.6 Strategy for detection

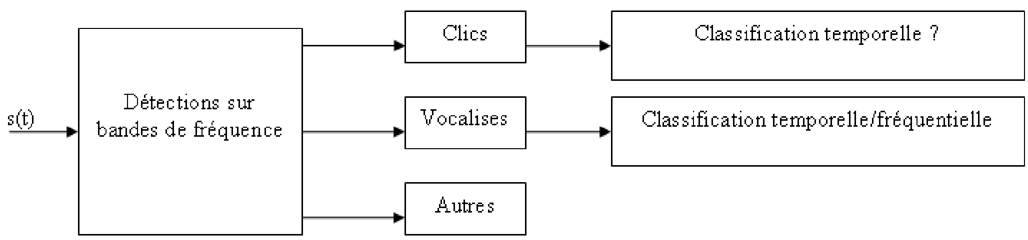

Figure 2.1: Framework strategy

The first step consist in detecting the signals of interest on frequency octave bands. A

Table 2.1: Hydrophones positions: D=Datasets, Dist=Distance to barycenter (m).

D	Hydros	Dist	X (m)	Y (m)	Z (m)
D1	H 1	5428	18501	9494	-1687
	H 2	4620	10447	4244	-1677
	H 3	2514	14119	3034	-1627
	H 4	1536	16179	6294	-1672
	H 5	3126	12557	7471	-1670
	H 6	4423	17691	1975	-1633
D2	H 7	1518	10658	-14953	-1530
	H 8	4314	12788	-11897	-1556
	H 9	2632	14318	-16189	-1553
	H 10	3619	8672	-18064	-1361
	H 11	3186	12007	-19238	-1522

Table 2.2: Sound emmission of the spiecies in mediterranea.

Espce	Domaine de frquence (Hz)
Rorqual commun / Fin whale	30 - 750
Cachalot / Sperm whale	100 - 30 000 (clics)
Globicphale/ Pilot whale	500 - 20 000
Grand dauphin / Bottlenose dolphin	800-24 000 (sifflements)
Dauphin bleu et blanc/ Striped dolphin	6 000 - 24 000 (sifflements)
Grand dauphin / Bottlenose dolphin	110 000-130 000 (clics)
Dauphin de Risso/ Risso's dolphin	65 000 (clics)

first classification permits to clasify between clicks, vacalization et others signals (anthropic, biologic...). Then, a classification step more precise matching the type of signal permits to identify the species (finwhale, spermwhale...).
Finnaly, if an hydrophone array is provided, it is possible to track the individuals present in the data.

Chapter 3

DETECTION AND STRUCTURATION OF SUBMARIN ACOUSTIC SIGNALS

We work on frequency and time frame of variable length according to the frequential resolution. In case of a priori knowledge, the signal is decimated to have a optimal frequency representaiton.

3.1 Detection threshold

The detection thresholds can be computed on linear scales or on octave bands. If no a priori knowledge are avaibale, the firt octave is centered on 5Hz.

3.2 Power Spectral Density (PSD): Welch periodogram

The Welch periodogram is an estimation method of the PSD of a signal. The direct method of the periodogram computing uses the Fast Fourier Transform (FFT) of the signal. This method allows one to compute easily the PSD of a finit duration numeric signal , even if periodic. To reduce the bias and the variance, we mean the Nb FFT of the signal, computed on a sliding window with an overlapping of 50% to avoid the noise correlation on each window. The windows are rectangular and the length is far inferior than the signal length. The estimated PSD is:

$$PSD = \sum_{1}^{Nb} \frac{|fft(S)|^2}{NFFT}.$$

Nb can be equal to 10. NFFT is the number of FFT point.

3.3 Quadratic detection

The quadratic detection is a detector which thresholds are computed thanks to the signal energy on several octaves. The principe is illustrated on the figure 3.1.

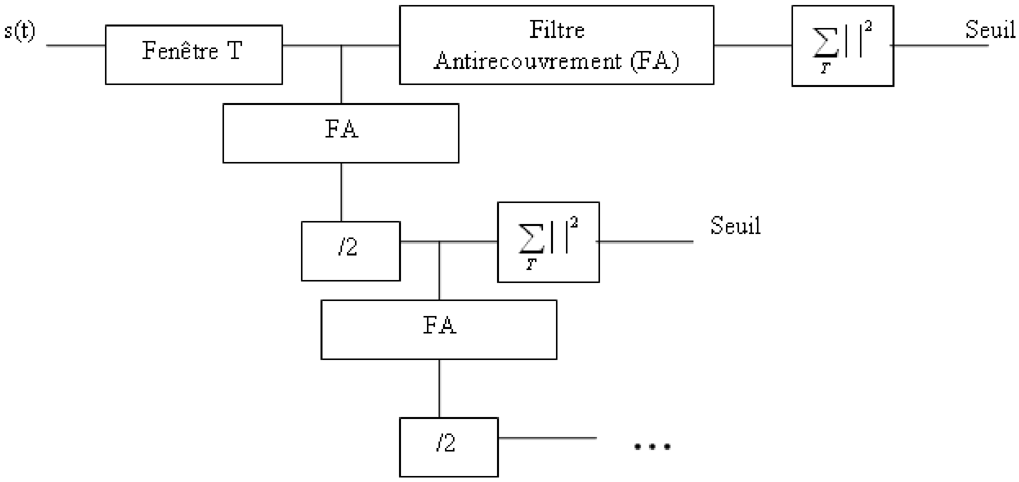

Figure 3.1: Quadratic detector scheme

3.4 Detection on the PSD

The detection can be computed directly on the PSD. First we have to learn the thresholds. Partionning the PSD in several windows, we can computed a threshold pour each of them, which can be done by taking the maximum energy of the window, or a threshold proportionnal to the media, (i.e 110% of the median for exemple). Those thresholds are stored. While on the detection phase, we stored the events over the thresholds, and if positive, the detection of the frequency band is stored.

3.4.1 Evolutive threshold learning

For the detectors above, learning on the threshold is evolutive, i.e when no detection occurs. The thresholds are updated continually during the sigal, and are forbidden when there is a detection

3.5 Performances

The ROC (Receptor Operating Characteristic) are computed on Fig.3.2 for the frequential detector and on figure 3.3 for the quadratic detector.

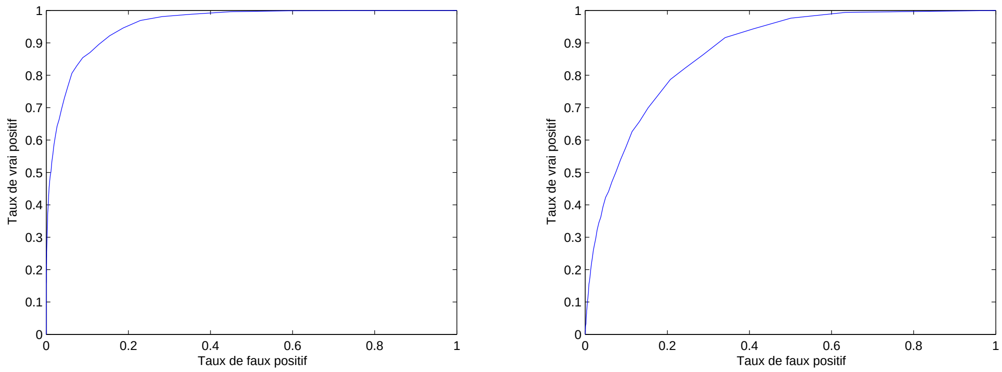

Figure 3.2: On the right (respectively on the left), ROC curve for a Signal to Noise Ratio (SNR) of -10dB (respectively -3dB).

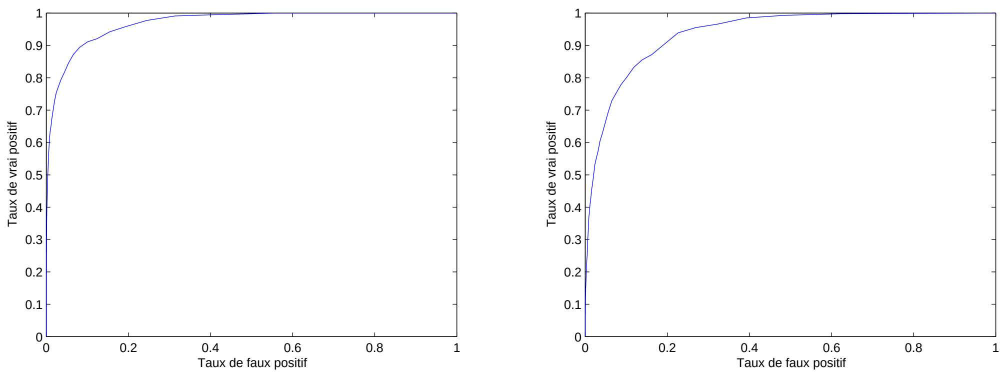

Figure 3.3: On the right (respectively on the left), ROC curve for a SNR of -10dB (respectively -5dB).

Chapter 4
CLASSIFICATION

4.1 Correlation spectrogram

The correlation spectogram [?] is a tool allowing to detect/classify a signal thanks to its signature. It is used in the case of frequency modulated signals , typically low frequency emmission (finwhale...). The method consists in matching a kernel signal (fig.4.1) in the spectrogram of a signal by croscorrelation. The crosscorrelation result is a likelyhood function which is thresholded in order to detect events where the signal of interest appears. First of all, we have to generate a average signature of the vocalization which will be search in the spectrogram. The signature is divided in segments, each segment corresponds to a modulation rate of frequency considered constant. To create the signature, we compute the vocalizations of a known signal, the frequencies and initial and final temporal positions, and the mean band width on a given number of segments. The reference is composed of segments which characteristics are the means of the precedent computed parameters.

The kernel is defined as:

$$x = f - (f_0 + \frac{t}{d}(f_1 - f_0)),$$
$$kernel(t, f) = (1 - \frac{x^2}{\sigma^2})exp(\frac{-x^2}{(2 * \sigma^2)}),$$

with f_0 the initial frequency of the segment, f_1 the final frequency, σ the band width, d duration of the segment. The kernel is here defined as a hat function, The negative parts are used for the rejection of the noise. The threshold of the likelyhood function is manually chosen.

The likelyhood function is:

$$\alpha(t) = \sum_{\tau} \sum_{f} kernel(\tau, f)S(t - \tau, f),$$

with $S(t, f)$ the spectrogram.

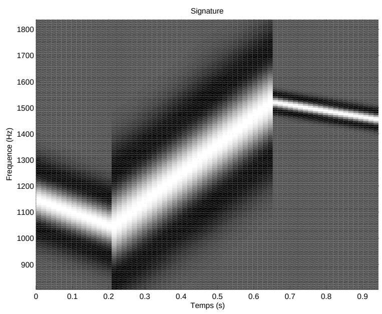

Figure 4.1: Kernel of a right whale vocalization. 3 segments are chosen.

4.1.1 Modification

An evolution of this algorithm consist in re-estimating the signal for each detection. First of all, we binarize the spectrogram after normalization, thresholding this last. We use the 2-3 time the standard deviation of a white noise as a threshold. We compute the signature like above, binarize it (fig.4.3). When the algorithm is running, and after the first detection, a new signature is extracted.

Signature extraction

Le signal d'intrt change au cours du temps et on remarque que souvent, deux missions prises cte--cte, ont des formes proches, mais peuvent aussi avoir des formes exotiques et ne pas bien se corrler une signature statiques. C'est pourquoi changer de signature chaque dtection peut amliorer la probabilit de dtection. Pour ce faire, on utilise l'algorithme Flood Fill (FF), qui consiste rassembler les pixels d'une rgion possdant des niveaux de gris au dessus d'un certain seuil. Le pixel de plus forte intensit est pris comme initialisation de l'algorithme et il est choisi dans la rgion frquentielle des signaux d'intrt (dans notre cas 500-2000Hz) et dans la zone temporelle de la dtection. On laisse ainsi la signature se 'remplir' et dessiner ses contours. Comme rsultat, nous avons alors une nouvelle signature, qui va varier chaque nouvelle dtection (fig. 4.2). Les signatures extraites qui ne rpondent pas certains critres frquentiels et temporels sont rejetes (fig.4.2). Dans le cas o il n'y a pas eu de dtection pendant plus de 10s, la signature initale est de nouveau utilise.

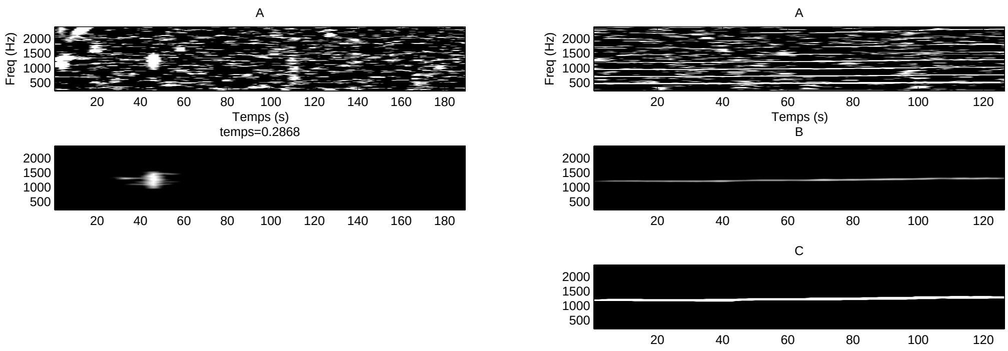

Figure 4.2: A:binarized spectrogram, B: signature after FF, C: extracted signature. On the left, rejected signature, on the right accepted signature

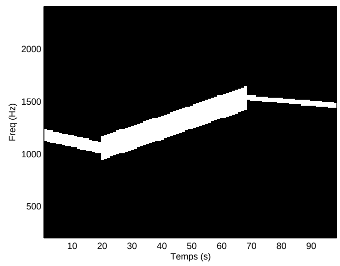

Figure 4.3: Binary vocalize kernel of the right whale. 3 segments are chosen.

4.1.2 Results

The figure 4.4 shows the restulst obtained on a right whale signal of 300s duration. Signature re-estimation are clearly allowing to detect several emission that a static signature did not detect.

4.1.3 Temporal variation

To avoid the temporal variation problem, and enhance the probability of detection, on method consist in correlation the signal with each segment separatelly. Then we have a number of likelyhood function $\alpha_i(t)$ equal to the number of segment. The delays of apparition of each segments in the signature are modelled with probability function $\lambda_i(t)$.

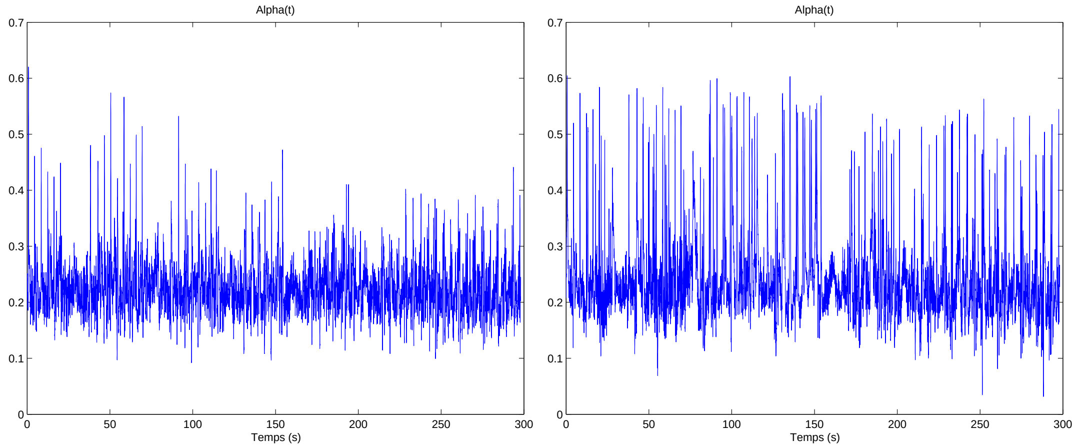

Figure 4.4: On the left, detection results with a static signature, on the right, results with extracted signatures.

Finally we create a function $f(\alpha_i(t), \lambda_i(t))$, and the final score correspond to the maximum of that function.

4.2 Classification for the edge detection on a smoothed spectrogram

This method is used in the case of frequency modulating signals. The spectrogram of the signal is first smoothed with a gaussian kernel before the ege detection The edge detection of the pattern is done by comparison of the smoothed spectrogram $S(t, f)$ with a background measure $B(t, f)$. $B(t, f)$ is evolutive and is calcuted with:

$$B(t, f) = B(t-1, f) + (\frac{S(t, f) - B(t, f)}{\alpha}),$$

with α the time constant for the background updating. The edge is detected if:

$$\frac{S(t, f)}{B(t-1, f)} > \xi,$$

ξ is a manual threshold.

The detected pattern is described with the following parameters:

- Length
- Initial freuqnecy
- Minimum frequency
- sweeping frenquency
- minimal frequency of the temporal position
- maximum frequency of the temporal position
- Maximal band width

Several criterions like the duration of the signal, the initial frequency, the sweeping frequency are applied to eliminate roughly the others signals of no interest. Classification can be done with a discrimating analysis multivariated function on the signals which passed the initial selection

Chapter 5

THE ARRAY CONFIGURATION

5.1 The Cramr-Rao Lower Bound

The CRLB provides the maximum acurracy for the estimation of the position. Considering a constant sound speed profile, the function model of the TDOA is defined:

$$s(\theta) = \frac{1}{c_s}\left\{\sqrt{\sum_{k=1}^{3}(X_{i,k}-\theta_k)^2} - \sqrt{\sum_{k=1}^{3}(X_{j,k}-\theta_k)^2}\right\}.$$

where $X_{\{i,j\}}$ is the vector coordinate of hydrophone $\{i,j\}$, θ is the unknown parameters vector $[x\ y\ z]^T$ and c_s the celerity. Thus, considering the TDOA noise Gaussian and B its variance-covariance matrix, the Fisher Information matrix is:

$$I_\theta = \nabla_\theta s(\theta) B^{-1} \nabla_\theta^T s(\theta).$$

Then, the CRLB is $B_\theta = I_\theta^{-1}$.

5.2 Confidence Regions

The confidence ellipse generalizes the notion of confidence interval in case of gaussian vector in two dimensions. Considering a stochastic gaussian vector $X \sim N(\mu, \Sigma)$ of two dimensions with $\Sigma > 0$. We have to determine the level curves of the gaussian distributionl, i.e. the points x of the ellipse

$$(x-\mu)^*\Sigma^{-1}(x-\mu) = r$$

Being a symmetrical positive matrix, Σ can be spectraly factorized as $\Sigma = PDP^*$ where P is the orthonormal matrix of the eigen vectors and D the diagonal matrix of the eigen values. For each x, the point $\tilde{x}$ is defined with $\tilde{x} = D^{-1/2}Px$, then $|\tilde{x}|^2 = r$, i.e. all the $\tilde{x}$ is the circle centered with radius $r : \tilde{x}(\rho) = r[\cos(\rho)\ \sin(\rho)]^*$ for $0 \leq \rho < 2\pi$. The parametric equation of the ellipse is:

$$x(\rho) = \mu + rP^*D^{1/2}\begin{bmatrix}\cos(\rho)\\ \sin(\rho)\end{bmatrix}, 0 \leq \rho < 2\pi$$

The quadratic norm of the stochastic vector $\tilde{X} = D^{-1/2}PX$ is distributed with the chi-square distribution with 2 degrees of freedom. Putting r = 5.99, then

$$\mathbb{P}(X \in x \in \mathbb{R}^2 : (x-\mu)^*\Sigma^{-1}(x-\mu) \leq r) = 0.95$$

This means the real position is 95% likely to be inside the ellipse.

Chapter 6

DETECTION AND STRUCTURATION OF SUBMARINE ACOUSTIC SIGNALS

6.0.1 Transient signal detection

6.0.2 Kaiser-Mallat Filtering

A sperm whale click is a transient increase of signal energy lasting about 20 ms (Fig.??-a). Therefore, we use the Teager-Kaiser (TK) energy operator on the discrete data:

$$\Psi[x(n)] = x^2(n) - x(n+1)x(n-1), \tag{6.1}$$

where n denotes the sample number. An important property of TK is that it is nearly instantaneous given that only three samples are required in the energy computation at each time instant. Considering the raw signal $s(n)$ as:

$$s(n) = x(n) + u(n),$$

where $x(n)$ is the signal of interest (clicks), $u(n)$ is an additive noise defined as a process realization considered wide sense stationary (WSS) Gaussian during a short time, by applying TK to $s(n)$, $\Psi[s(n)]$ is:

$$\Psi[s(n)] \approx \Psi[x(n)] + w(n),$$

where $w(n)$ is a random gaussian process [?]. The output is dominated by the clicks energy. Then, we reduce the sampling frequency to 480Hz by the mean of 100 adjacent bins to reduce the variance of the noise and the data size. We apply the Mallat's algorithm [?] with the Daubechies wavelet (order 3). We chose this wavelet for its great similarity to the shape of a decimated click [?, ?]. The signal is denoised with a universal thresholding [?] defined as $D(u_k, \lambda) = sgn(u_k)max(0, |u_k| - \lambda)$, with u_k the wavelet coefficients, $\lambda =$

$\sqrt(2log_e(Q))\sigma_N\sigma_{\tilde{N}}$, and Q the length of the signal resolution level to denoise. The noise standard deviation σ_N is calculated on each 10s window on the raw signal with a maximum likelihood criterion. $\sigma_{\tilde{N}}$ is the standard deviation of the wavelet coefficients on a resolution level of a generated, reduced and 0-mean Gaussian noise. This filtering step is very fast without any parameter. Fig.**??**-d is the filtered signal on multiple (Fig.**??**-c) emitting MMs.

Chapter 7

CONTRIBUTION TO LOCALIZATION ON A LONG BASE ARRAY AND PERFORMANCES

7.0.3 Time Delay of Arrival estimation Estimation

Selection of the TDOA and transitivity

First, T estimates are based on MM click realignment only. Every 10s, and for each pair of hydrophones (i, j), the difference between times t_i and t_j of the arrival of a click train on hydrophones i and j is referred as $T(i, j) = t_j - t_i$. Its estimate $\widetilde{T}(i, j)$ is calculated by CC of 10s chunks (2s shifting) of the filtered signal for hydrophones i and j [**?**, **?**]. We keep the 35 highest peaks on each CC to determine the corresponding $\widetilde{T}(i, j)$. The filtered signals give a very fast rough estimate of T (precision $\pm$ 2ms). Fig.(**??**.e) shows the CC with the raw signal and (**??**.f) with the filtered signal. Without filtering, CC generates spurious delays estimates and the tracks are not correct. The maximum $\widetilde{T}$ rank (Fig.**??**) in D1, pitching the source localization, are high among the 35 $\widetilde{T}$ kept in the CC which justifies this number.

7.0.4 TDOA Selection and Localization with a Constant Profile

Each signal shows echoes for each click (Fig.**??** b), maybe due to the reflection of the click train off the ocean surface or bottom or different water layers. We use a method based on autocorrelation [**?**, **?**] to eliminate it. Then, thanks to the $\widetilde{T}$ transitivity system described in [**?**] we keep $\widetilde{T}$ triplets coming from the same source. Finally, thanks to the measured delays and an acoustic model based on a constant sound speed profile, the least squares cost function determines the MM positions using a multiple non linear regression with Gauss-Newton method (Levenberg-Marquardt) [**?**]. The residuals are approximatively following a Chi-square distribution with $Nc - d$ degrees of freedom, noted X^2_{Nc-d}, Nc is the number of hydrophones couples considered and d the number of unknowns, here 3 (x, y, z). The position is accepted if the residual is inferior to a threshold x, That is calculated solving

$P = prob(X^2_{Nc-d} > x)$ with $P = 0.01$ (we keep 99% of the estimates).

7.0.5 Source Localization with a Linear Sound Speed Profile

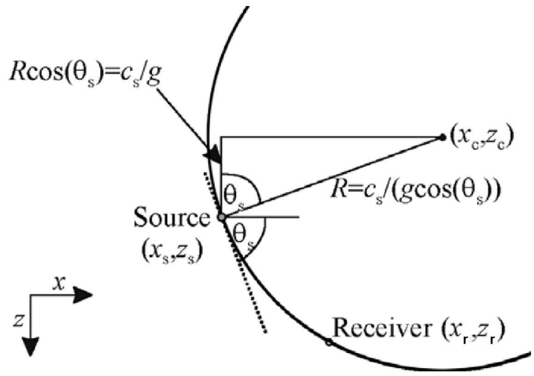

Figure 7.1: Geometry for a source and receiver in a linear profile [?].

It is well known that the ray paths in a medium with linear sound speed profile are arcs of circles and further the radius of the circle can be computed (Fig.7.1). c_s is the sound speed at the source, θ_s is the launch angle of the ray at the source, measured relatively to the horizontal. From the geometry shown in Fig.7.1, the center of the circle, (xc, zc), along which the ray path is an arc, is [?]:

$$
\begin{aligned}
x_c &= \frac{x_s + x_r}{2} + \frac{(z_s - z_r)}{2(x_s - x_r)}(z_r - z_s + \frac{2c_s}{g}), \\
z_c &= z_s - \frac{c_s}{g}.
\end{aligned} \tag{7.1}
$$

For linear sound speed profile the course time τ of the ray is:

$$
\tau = \frac{1}{g}\left\{\log\left(\frac{z_c - z_s}{z_c - z_r}\right) - \log\left(\frac{R + x_c - x_s}{R + x_c - x_r}\right)\right\}. \tag{7.2}
$$

Using Eqs.(7.1)-(7.2) allows one to compute the propagation time from the source to any receiver and then the whale position.

Chapter 8

TRACKING OF SEVERAL TARGETS ON A SHORT BASE ARRAY

8.1 Tracking Method

8.1.1 RBMCDA framework for the data association and AoA estimation

State optimal filtering with the RBMCDA

The question one could ask is 'What if the records produce noisy TDOA measurement?' To tackle this issue, a method consists in using the Rao-Blackwellized Monte Carlo Data Association (RBMCDA). In a nutshell, we solve the tracking part with an EKF (Extended Kalman Filter) filtering [?],[?], and the data association (whale or clutter association) with a particle filter. This can be achieve thanks to the Rao-Blackwell theorem [?], [?]. Developping extensivly this part is not a concern. First, we would need several pages, but also it was treated exhaustivly in article such [?]. We will rather do a brief recall and explain the principle of this method. In multiple target tracking (MTT) we are estimating the states of several targets (whales) through measurements. If we know the targets which produce each measurement the problem reduces to single target tracking and we can use a standard filtering algorithm (e.g. Kalman or Extended Kalman Filter) for estimating the states of the targets independently. Unfortunately, such knowledge is very rarely available in practice, and in many cases some measurements might be also due to clutter, so one is forced to solve the problem of data association. After we have associated measurements to the whale or clutter, we can apply the standard filtering techniques for estimating the target states. In the RBMCDA framework the whale states, data associations and the births and deaths of targets are treated as hidden stochastic processes, which are observed through noisy and indirect measurements. The joint tracking and data association is formulated as a Bayesian estimation problem and the inference is done with Sequential Monte Carlo (SMC) methods (also referred as particle filtering methods) [?],[?],[?], which give Monte Carlo approximations to the posterior distributions.Furthermore, the accuracy and efficiency of the algorithm is enchanced with the application of Rao-Blackwellization, which allows us to integrate over the target states and use SMC only for estimating the data associations.

Optimal filtering of the whale states

The system model for one whale is (by discretization) in the space of states:

$$x_k \sim p(x_k|x_{k-1}), \tag{8.1}$$

$$y_k \sim p(y_k|x_k), \tag{8.2}$$

where x_k is the whale state at step k, in our case, $x_k = (\alpha_k, \phi_k, \dot{\alpha}_k, \dot{\phi}_k)$. y_k is the measure at step k, for example the TDOA. $p(x_k|x_{k-1})$ is the dynamic model of the whale. $p(y_k|x_k)$ is the measure model, which describes how the measures are distributed given the state of the whale. The goal of optimal filtering is to compute the estimate of the current state, using the measures collected untill the current step, i.e, we want to compute recursively the marginal posterior distribution:

$$p(x_k|y_{1:k}). \tag{8.3}$$

The non linear system can be express in our case:

$$\begin{aligned} p(x_k|x_{k-1}) &= N(x_k|A_{k-1}x_{k-1}, Q_{k-1}), \\ p(y_k|x_k) &= N(y_k|h(x_k), R_k). \end{aligned}$$

For the linear form of this system, the common algorithm is the Kalman filter [**?**]. Here we chose to use the EKF thanks to the locally linearization. After the state estimation, it is possible to smooth the states, by calculating the posterior marginal distribution:

$$p(x_k|y_{1:T}),$$

where $T > k$. Those smoothing can be achieved with the Kalman smoother or the RTS (Rauch-Tung-smoother)[**?**],[**?**].

BIBLIOGRAPHY

Appendix A

APPENDIX