Détection et classification dans un contexte acoustique passive : application à la détection des signaux basse-fréquences des baleines bleues

Léa Bouffaut

UNIVERSITE SCIENCES BRETAGNE DE LA MER LOIRE ET DU LITTORAL

UBO
Université de Bretagne Occidentale

THESE DE DOCTORAT DE

L'UNIVERSITE
DE BRETAGNE OCCIDENTALE
COMUE UNIVERSITE BRETAGNE LOIRE

ECOLE DOCTORALE N° 598
Sciences de la Mer et du littoral
Spécialité : Acoustique sous-marine et Traitement du signal

Par **Léa BOUFFAUT**

Detection and classification in passive acoustic contexts:
Application to blue whale low-frequency signals

Thèse présentée et soutenue à Lanvéoc, le 22 Octobre 2019
Unité de recherche : Institut de Recherche de l'Ecole Navale IRENav – EA 3634

Rapporteurs avant soutenance :

Olivier ADAM,
Professeur des Universités,
Institut ∂'Alembert, Sorbonne Université

Jérôme MARS,
Professeur des Universités,
Gipsa-Lab, Université de Grenoble

Composition du Jury :

Rapporteurs :
Olivier ADAM, Professeur des Universités,
Institut ∂'Alembert, Sorbonne Université

Jérôme MARS, Professeur des Universités,
Gipsa-Lab, Université de Grenoble

Examinateurs :
Jean-Yves ROYER, Directeur de Recherche CNRS, Président du jury,
LGO, Institut Universitaire Européen de la Mer, UBO

Abdeldjalil AISSA-EL-BEY, Professeur,
Lab-STICC, Institut Mines Télécom Atlantique

Holger KLINCK, Directeur de Recherche (Ph. D),
Cornell Lab of Ornithology, Cornell University

Odile GERARD, Ingénieure (Dr.)
DGA Techniques Navales, Toulon

Co-encadrante de thèse :
Valérie LABAT, Maître de Conférences,
IRENAv - Ecole Navale

Directeur de thèse :
Abdel BOUDRAA, Professeur des Universités,
IRENAv - Ecole Navale

École doctorale 598 : Sciences de la mer et du littoral

Thèse de doctorat

pour obtenir le grade de docteur délivré par

l'Université de Bretagne Occidentale

Spécialité doctorale "Acoustique sous-marine et traitement du signal"

présentée et soutenue publiquement par

Léa BOUFFAUT

le Mardi 22 Octobre 2019

Detection and classification in passive acoustic contexts
Application to blue whale low-frequency signals

Jury

Olivier Adam,	Professeur des universités	Rapporteur
Jérôme Mars,	Professeur des universités	Rapporteur
Jean-Yves Royer,	Directeur de recherche CNRS	Examinateur
Abdeldjalil Aissa-El-Bey,	Professeur	Examinateur
Holger Klinck,	Directeur de recherche	Examinateur
Odile Gérard,	Ingénieure-docteure	Examinatrice
Abdel-Ouahab Boudraa,	Professeur des universités	Directeur
Valérie Labat,	Maitre de conférence	Encadrante

Institut de Recherche de l'École Navale EA3634
École Navale / Arts et Métiers ParisTech
BCRM Brest CC600, 29240 Brest Cedex 9, France

Acknowledgments

Après quatre années qui ont filées comme le vent un soir de tempête sur le pont de Recouvrance, me voila à essayer de résumer en quelques lignes à quel point je suis reconnaissante pour tout le soutien que j'ai reçu. Je dois l'admettre, j'ai vraiment attendu jusqu'à la dernière minute; je n'étais pas sûre de trouver les mots. C'est parti…

Pour commencer, un grand merci à mes encadrants Abdel-Ouahab Boudraa et Valérie Labat, pour avoir eu confiance en moi dès le départ, m'avoir suivie et supportée, dans tous les sens du terme, tout au long de cette incroyable odyssée. J'apprécie la liberté que vous m'avez attribuée et le soutien que vous m'avez apporté dans la poursuite de l'ensemble de mes projets, bien sûr scientifiques avec toutes les relectures de dernière minute, mais aussi d'avoir compris et approuvé ma volonté de "voir du pays". Merci de m'avoir accompagnée dans l'élaboration des différents dossier de financement.

Je tiens aussi à sincèrement remercier Jérôme Mars et Olivier Adam pour avoir accepté de lire et d'évaluer mon manuscrit, ainsi que l'ensemble des membres du jury: Jean-Yves Royer, Abdeldjalil Aissa-el-Bey, Holger Klinck et Odile Gérard pour leurs commentaires constructifs et les riches discussions qui ont suivi la soutenance. Merci d'avoir fait le déplacement jusque notre bout du monde !

Ensuite, un immense merci à Richard Dréo, je ne serai pas allée bien loin sans toi ! Déjà, merci d'avoir eu le génie de faire le lien entre les données sismiques du projet RHUM-RUM et les vocalises des baleines. Merci de t'être embarqué avec moi dans ce projet baleinesque, d'avoir remonté les manches et trouvé la trace de Simone ! Je suis aussi vraiment reconnaissante pour ton rôle de *pas si vieux* sage et pour l'aide précieuse que tu m'as apportée. Aussi, c'était une super bonne idée d'aller toquer à toutes les portes de la rade…

Ce qui m'amène à parler de ceux de l'autre rive: Merci à Flore Samaran et Julien Bonnel, de nous avoir ouvert les portes de l'ENSTA et d'avoir prêté oreille à notre chorus. Flore et Manue, nos discussions et vos super explications m'ont permis de mieux comprendre les problématiques et enjeux passés, présents et futurs des baleines de l'Océan Indien (et d'apprendre une flopée de nouveaux mots!). À votre contact j'ai gagné en humilité, vos expériences ont bousculé mes certitudes et m'ont ouvert les yeux sur la complexité de la PAM. Maëlle, merci d'avoir accepté de passer le détecteur de FX sur nos données pour que je puisse faire la comparaison ! Et merci à

tous les membres du très officiel Flore Samaran's lab (Julie, Fabio, Remi . . .) pour le soutien moral lors de ma fusillade présentation au LAB-STICC et les hilarants déjeuners du Flunch. Laurent, Bazile et Florent, merci d'avoir éclairé ma lanterne sur la propa ASM, les mesures de niveaux etc.

Bien qu'éloigné de ne notre terre d'irréductibles Brestois, Guilhem je souhaite te remercier pour nous avoir donné accès à cet incroyable jeu de données et pour tes conseils avisés. Je suis admirative de ta rigueur scientifique, ta réactivité et de ton engagement pour le partage et échanges inter-disciplinaires.

À vous tous, j'espère que nous continueront à collaborer !

Avant de passer en anglais pour dire quelque mots à mes collègues d'outre-Atlantique, il me semble judicieux de remercier ici les financements de mobilité du GdR ISIS et de l'École Navale qui ont rendu ce séjour possible, ainsi que la Fondation de la mer et IEEE OES pour la reconnaissance de ces dits travaux. First, Holger, I could never thank you enough for accepting and welcoming me into your lab. To me, this has been a life-changing experience. It has open my eyes to the meaning and direction I want to give to my work. Shyam, thanks for taking me under your wing, helping me become more accurate in programing and writing. Thanks to both of you for spending a tremendous amount of time editing and improving my English writing skills. I am also grateful that you encouraged me to take a step back from the work to get a better overview and see the whole picture. I would also like to thank each member of the extended CCB community, with a special note to Ana, Michelle, Karolin, David, Nisha, Luciana, Peter, Margherita, Rob, Lauren, Laurel, and Kristin: thanks for the walks in the snow, the enlighten talks about your beautiful country, the good food, drinks and laughs (and pasta making lessons!), game nights and all. I had the most wonderful time with you all. To Shyam, Michelle and Luciana, now that I have some more time, I hope we can get somewhere with our *fancy* project!

De retour en Brittanie, un petit mot quand même sur l'École Navale qui m'a offert un cadre de travail exceptionnel. Les traversées quotidiennes en transrade commencent déjà à me manquer tant pour les magnifiques levers de soleil, les retours pourchassés par les dauphins ou les parties de tarot trop bruyantes de l'étage. Merci pour les moyens qui ont étés mis à ma disposition pendant ces quatre ans. Merci aussi de m'avoir fait confiance pour contribuer à la formation des élèves, où j'ai pu gagner, je crois, en pédagogie et en assurance.

Une antépénultième escale au labo: merci à tous pour ces quatre années. Merci Christophe pour tes conseils sur les bourses et les poursuites de carrière. Gaël et Amélie merci pour l'organisation bien huilée de la soutenance. À mes copains de galère de la première heure, Loïc et Benjamin, merci de m'avoir fait tellement lever les yeux au ciel que j'ai failli m'en décrocher la mâchoire et d'avoir toujours été là ! Bruno, les g.d. (Flo, Benoit ET Arthur pour une sombre histoire de licorne), Kostas (++ pour avoir été un super co-bureau!), Eulalie, Jishen, Goulven, Guillaume, Tom: merci pour les parties de tarot, les conversation sans queue ni tête de midi, la

décompression du Tyr, les BBQ et diners en tout genre. Avec un peu de persévérance, j'ai appris à moins vous détester ;)

À ma famille, j'aime où nous en sommes aujourd'hui. Vous ne savez pas ce que cela signifie pour moi de vous avoir vus tous réunis pour ma soutenance. Maman merci pour tout ce que je ne sais pas exprimer et pour ce buffet somptueux. Thib, tu sais où commander quand ton heure sera venue ! Papa et Claudine, le miel m'a suivi jusqu'à mon nouveau chez moi et rend les soirées thé au coin du feu encore plus douces. Nathalie et Quentin merci énormément d'avoir fait le déplacement !

Pour finir, le plus grand des merci à mon amoureux. Merci de toujours croire en moi et d'être le meilleur des supporters. Merci de toujours prêter une oreille attentive à ce que j'ai à dire, de toujours (ok, *presque*) trouver les mots justes. Merci d'avoir été actionnaire majeur de la SNCF, puis de m'avoir rejoint sur des rives plus septentrionales de l'atlantique et d'être toujours partant pour de nouvelles aventures !

Fortsettelse følger...

Contents

List of Figures

List of Tables

Glossary

ABW Antarctic blue whale. ix, x, xiii, 2, 26, 45–47, 51–54, 56, 62, 72–74, 76, 78–80, 85, 88–91, 93, 97, 98, 100, 105–107, 110–112, 116, 119–122, 124, 127–130

ACF auto-correlation function. 108, 109

BW blue whale. xi, 1–5, 11, 12, 14–16, 18, 19, 21–23, 26–31, 33, 45–47, 49, 53, 61, 64, 67, 101, 105, 108, 118, 124, 125, 127, 129–132

chirpyness rate of frequency change of a signal in Hz/s. x, xiii, 106, 107

complex tone sound wave consisting of sinusoidal components of different frequencies. 48

DCLDE detection, classification, localization and density estimation of marine mammals. 31, 128, 131

DS down-sweep. 47, 48, 107, 119, 121, 122, 124, 125, 129, 130

FDR false discovery rate. x, 92–97, 100, 128

FW Fin whale. ix, x, 45, 49–54, 57, 89, 106, 107, 122, 123, 127

GEP generalized eigenvalue problem. 68, 69, 72, 75, 130

GMM Gaussian mixture model. 120, 121, 124, 125, 129, 131

HPS harmonic product spectrum. 108–112, 114, 115, 117, 129

ICI inter-call interval. ix, 45–50, 53, 131

inharmonic frequencies frequencies which are not rational multiples of each other. 48

ISI inter-series interval. ix, 45–50, 53

IWC International whaling commission. ix, 2, 3

KLE Karhunène-Loève expansion. 67–70, 120

LTS long-term spectrogram. 56

MDR missed detection rate. x, 91, 93, 95–97, 128

MF matched filter. x, xiii, 4, 5, 29–32, 64–68, 74, 78–81, 84, 85, 88, 90–97, 99–101, 127, 128

MPBW Madagascar pygmy blue whale. ix, x, 26, 45, 48, 49, 54, 56, 105–107, 110–112, 116, 119, 121–125, 127, 129, 130

nfft Spectrogram number of sampling points. 47, 49–51, 56, 58, 90

OBS ocean bottom seismometer. ix, xiii, 16, 33, 37–40, 43, 45–51, 53–56, 58, 62, 63, 78, 79, 88–90, 92, 97, 99, 111, 112, 118, 127–130

PAM passive acoustic monitoring. ix, 3, 5, 13–22, 26, 28–30, 32, 33, 38, 49, 63, 64, 66, 76, 98, 124, 125, 127, 132

partial one of a group of frequencies, not necessarily harmonically related to the fundamental, which appear in a complex tone. xiii, 48, 53, 106, 107

PC principal component. xi, 120, 121, 124, 129

PCA principal component analysis. 30, 120, 122, 125

PPV positive predictive values. 91, 93, 113, 128

PSD power signal density. 46–51

pure tone a tone with no *harmonics*. All energy is concentrated at a single frequency. 47, 49, 106

ROC receiver operating characteristics. x, 93, 94, 128

SAR synthetic aperture radar. 67

SAS synthetic aperture sonar. 67

SMF stochastic matched filter. x, xiii, 4, 5, 57, 64, 66–69, 71–81, 83–85, 88, 90–101, 127–130

SNR signal to noise ratio. x, xiii, 4, 29–32, 46, 64–66, 68, 69, 71–73, 75, 76, 78–81, 83–85, 88–92, 94–99, 106, 107, 112–115, 117, 118, 127–130

SOFAR sound fixing and ranging. 5, 13, 17, 130

sonar sound navigation and ranging. 1, 22, 25, 29, 67, 71

sound floor the quietest ocean conditions. 41

STFT short time fourier transform. 75, 120, 121

TDOA$_m$ measured time difference of arrival. 55, 56

TDOA$_{th}$ theoretical time difference of arrival. 55, 56

TF time-frequency. x, 30, 66, 74–76, 84, 90, 98, 105, 109, 111–115, 121, 124, 129, 132

TOA$_m$ measured time of arrival. 56

TOA$_{th}$ theoretical time of arrival. 56

TPR true positive rate. x, 91, 93–97, 99, 100, 113, 128

Ts detection threshold. x, 79, 91, 92, 95, 99, 100

Introduction

Throughout history, the ocean has been a vital source of sustenance, transport, commerce, growth, and, an inspiration for humanity. In our interactions with the ocean and all the life within, underwater acoustics has been playing a central part. Indeed, light does not travel far under the sea surface and, in this realm, acoustic waves became *our eyes*, as they can propagate over long distances. Based on this unique property, the first "utilitarian" use of underwater acoustics dates back to 1912. In the wake of the Titanic disaster and, at the dawn of WWI, Fessenden proved possible the detection of an iceberg up to a range of 2 km. The following underwater acoustic technological advances and research efforts were devoted to the development of military applications such as the development of Anti-Submarine Detection Investigation Committee (or ASDIC) and sound navigation and ranging (sonar). They contributed to the rise of the field of underwater acoustics. The transmission of this newly-developed knowledge, from military to civilian sciences, introduced new ways to observe, measure, and explore the oceans (e.g., echosounder and, the beginning of fishery acoustics).

Among the 20th-century developments for ocean exploration, passive acoustics has been proved valuable for conducting discrete, non-intrusive, long-term and often, cost-effective surveys of the underwater world. Numerous information can be obtained through passive listening of the oceanic sounds. For example, surveillance systems can be wired onboard of warships for threat detection or, on harbor guarding systems to prevent intrusions. Passive acoustics can also be beneficial for understanding the physics of the ocean or oceanic floor, e.g., by listening to seismic events or ice tremors. Contemporary ecological concerns and worldwide awareness about ocean pollution and human impacts on the environment brought the spotlight on conservation issues. As an integrated part of conservation efforts, passive acoustics contributes to the monitoring of these soniferous marine species. As polar bears became synonymous with climate change, whales are now considered as the flagship species of the support of biodiversity conservation in the oceans.

The history between humans and the whale is complicated and, its status has constantly changed: from a mythological creature to a monster of the sea; from an unlimited resource to the one we must protect. It has not always been the case. Over the 18th and early 19th centuries, whales were extensively hunted for their grease by the commercial whaling industry and, almost brought to extinction. First whaling areas were restricted to the Northern Atlantic Ocean. Blue whales (BWs) and rorquals were not, at first, the main target of the expeditions.

They inhabited waters that were too remote; they were going too fast; they were too hard to catch and process (e.g., meat, oil, baleens). But, the decrease of stocks in other preferred species and, the introduction of new equipment (e.g., harpoon cannon and black powder harpoon), drove men to extend their hunting territories to the resourceful southern waters. They started to prey on larger whales. The International whaling commission (IWC) was created in 1946 with the purpose to "provide for the proper conservation of whale stocks and thus make possible the orderly development of the whaling industry"[1]. One of the IWC actions was the designation of sanctuaries or zones where commercial whaling is prohibited. The first, the Indian Ocean Sanctuary, was established in 1979. Still in place, it encompasses all the waters in the Northern Indian Ocean with a southern boundary at 55° S (IWC, 1980). In 1982, the IWC moratorium suspending commercial hunt was signed.

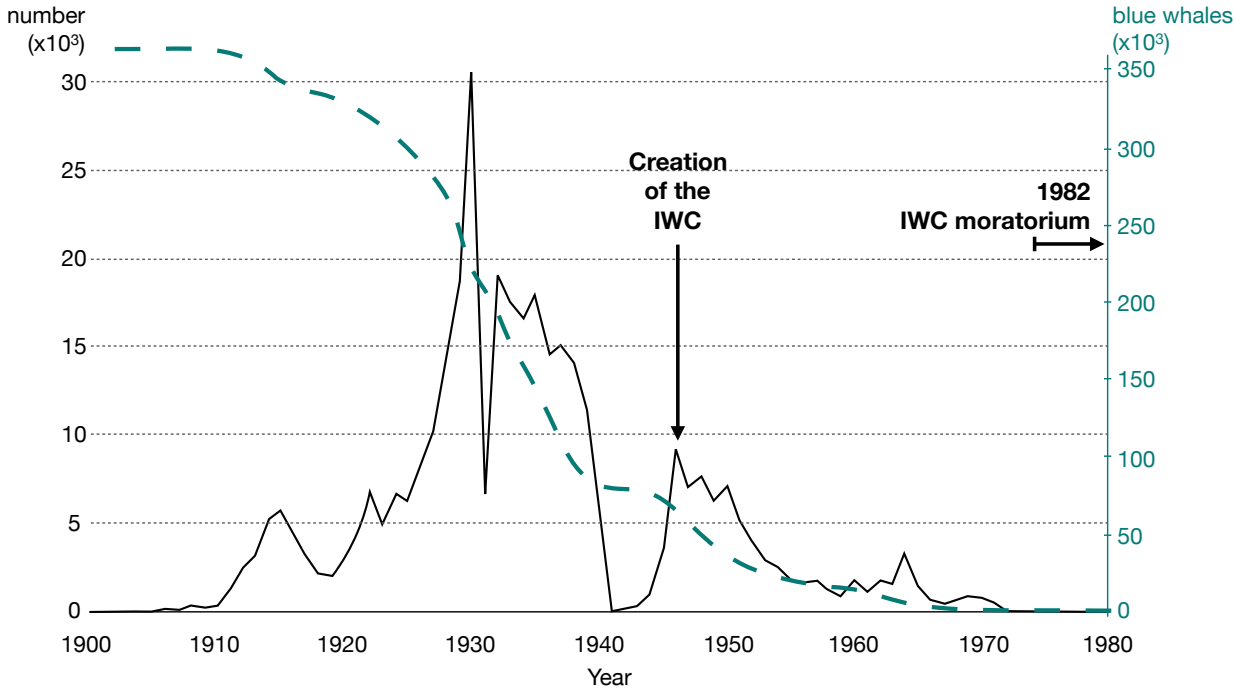

Figure 1: Catch and whale number in the southern Oceans (adapted from (Leroy, 2017), based on IWC data).

No species was brought to extinction but less than 1% of the BWs remained (Figure 1). Nowadays in the process of recovering, they are recorded on the IUCN (International Union for Conservation of Nature) Red List. BW worldwide is considered endangered with stocks estimated between 5000 and 15000 individuals. The Antarctic blue whale (ABW) is considered critically endangered with 3000 individuals, while the pygmy BW is still considered data deficient [2]. Yet, still impaired by the fragility of their stocks, the recovery of BWs is facing threats of the modern world. For example, the intensification of ship traffic changed oceanic soundscapes, and ship strikes are now whales principal cause of mortality. Besides, plastic pollution and over-fishing impact

[1] https://iwc.int/

[2] https://www.iucnredlist.org

their food-sources, and there is still the question of the impact of climate change and oceanic acidification. Taking the appropriate measures to ensure their thrive is crucial to marine health and biodiversity. But, if coastal baleen whale species such as right and humpback whales are well studied, there are still a lot of basic ecological unknowns concerning southern Ocean BWs. It is under these considerations that the IWC southern Ocean Research Partnership was established in 2009. It is described as "an integrated, collaborative consortium for cetacean research, which aims at maximizing conservation-orientated outcomes for southern Ocean cetaceans through an understanding of the post-exploitation status, health, dynamics and environmental linkages of their populations, and the threats they face"[3]. The establishment of surveillance methods is essential to attain these conservation purposes.

Classical monitoring methods such as visual surveys (from a ship or an aircraft) are expensive and hard to deploy in remote open waters subject to heavy weather conditions. Considering the small number of whales compared to the immensity of the ocean, it is like looking for a needle in a haystack. Results of visual observations are too fragmented and therefore, insufficient. On the other hand, BWs species produce loud, low-frequency, regular and regionally distinct calls, propagating for tens to hundreds of kilometers. Eavesdropping, therefore, provides persistent means to conduct short to long-term surveys of the target population(s). These studies often require interdisciplinary efforts, from the understanding of sound production mechanisms and acoustic wave propagation effects to the processing of the information; from the instrumentation to the ecological interpretation. This process is known as passive acoustic monitoring (PAM) (Figure 2).

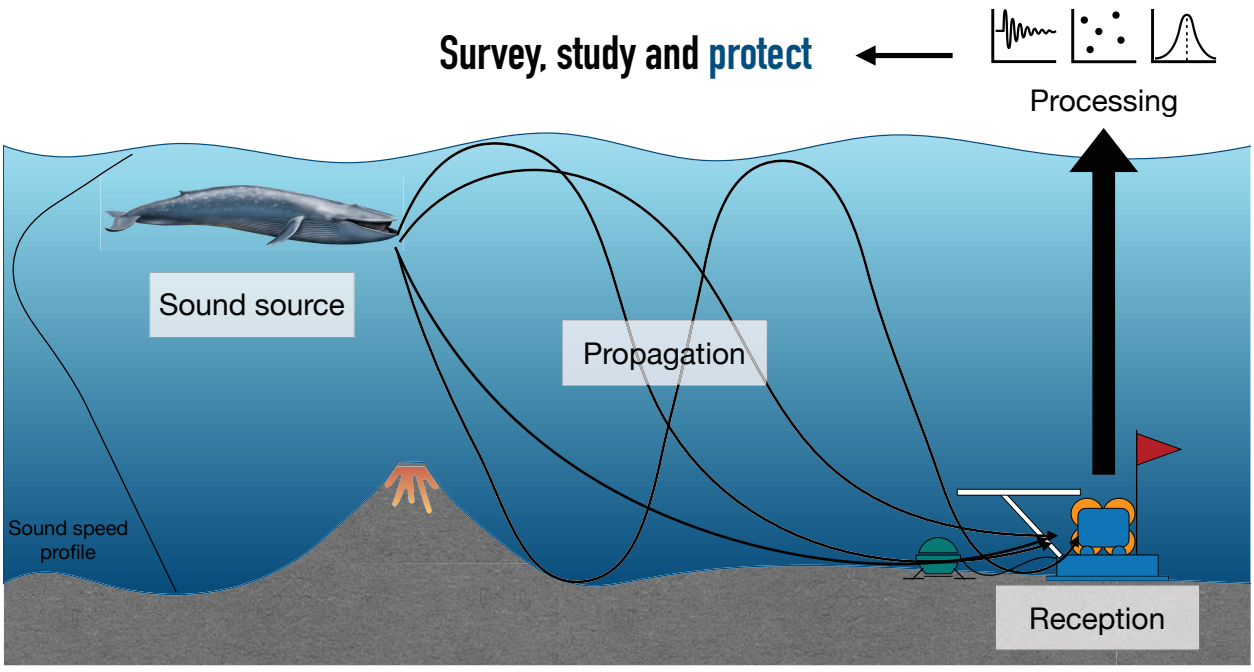

Figure 2: Schematics of PAM.

[3] http://www.marinemammals.gov.au/sorp

Over the last decades, numerous PAM surveys have been conducted all around the globe and the size of collected sound archives is rapidly increasing. The analysis of the large volumes of data resulting from continuous and long-term monitoring efforts unmistakably benefits from the automated detection of target signals. Automatic detection methods must be reliable and robust to gather statistically relevant elements and contribute to answer BW ecological questions. Yet, classical detection methods such as matched filters (MFs) exploit the stereotyped features of BW calls. But, are MFs really adapted to low-frequency passive contexts where, (1) whale sounds can travel across long-distances and are modified by the propagation channel, (2) overlapping noises can interfere and, (3) the contrast between the signal and the background noise is continuously varying (Figure 3)?

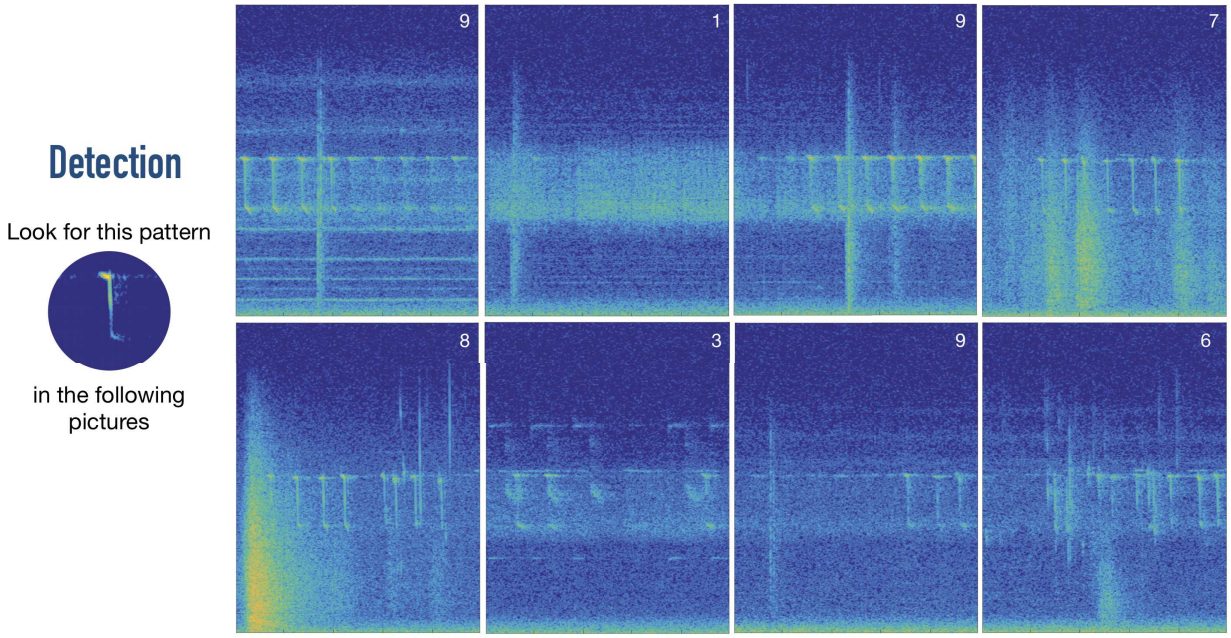

Figure 3: Illustration of the difficulties of detection in a passive context with reverberated attenuated signals, in the presence of sounds (seismic events, ship noise, and other biological sources). The number of occurrences of the target signal is indicated in the top -right corner.

In light of the above, the context of this study is the following: even if emitted signals are stereotyped, when received, they are modified by the propagation channel (echoes, low and varying signal to noise ratios (SNRs)...). Therefore, this thesis aims at addressing the subsequent questions:

1. What are the limitations of the MF?
2. How to improve signal detection in such changing conditions?
3. How to detect and automatically separate concurrently calling species?
4. How to assess the performances and compare the developed methods?

In order to provide some answers to these questions, different strategies are proposed. First, to improve signal detection, the stochastic matched filter (SMF) has been adapted to the passive

context. Then, the separation of concurrently calling species (concurrent and overlapping acoustic sources) is based on a pattern recognition system, with an additional reconstruction step. Performance and limitations of the SMF, MF, and a recently developed detector, the Z-detector (Socheleau et al., 2015), are assessed and compared on annotated recordings.

The dataset used in this thesis was recorded between 2012 and 2013 near La Réunion Island by the RHUM-RUM (Réunion Hotspot and Upper Mantle - Réunions Unterer Mantel) seismological experiment[4], offering countless possibilities for methods development, testing, and, analysis. In addition, as the second long-term BW PAM study in the Western Indian Ocean Sanctuary, this area is a particular source of interest [5]. Recording equipment lying on the bottom of the ocean shed new light on the study of BWs and, more generally, low-frequency sounds in this area.

The thesis is articulated around five chapters. After a review of the different methods for the survey of BWs, Chapter 1 introduces in more details the constituting fields behind PAM. It uncoils the thread connecting the means (recording equipment) and the application (ecology), with the focus of the thesis, underwater acoustics, and signal processing.

Chapter 2 describes the RHUM-RUM experiment and the equipment characteristics. It also summarizes the underwater ambient sounds, their origin, and spectro-temporal characteristics to depict an accurate representation of low-frequency soundscapes from the bottom of the Indian Ocean. Notably, a thorough description of the recorded baleen whale sounds is provided.

Chapter 3 addresses the question of the detection of stereotyped sounds. MFs are introduced from the point of view of noise reduction problem in a passive context. The strategy chosen to compensate for MFs limitations, the SMF, is described along with the improvements to adapt the method to passive contexts, followed by application to ABW call detection and, an illustration for scuba-divers breathing detection.

The comparison of detection methods relies on a fair comparison of their performances in different situations. Therefore, Chapter 4 deals with methods performances evaluation under the specific constraints of the passive context, based on a ground-truth dataset. The proposed detection method, the SMF is compared to the MF and, to the Z-detector (Socheleau et al., 2015).

Finally, multi-class detection and signal reconstruction are applied to the problem of concurrent calling species in Chapter 5. The method is viewed as a pattern recognition system. The first step of signal extraction is based on tonal signal detection. A comparison of tonal signal detectors is carried out and, preliminary results of the complete method are presented.

[4] www.rhum-rum.net

[5] Note that the first BW PAM study in the Western Indian Ocean Sanctuary, the OHASIS-BIO (Observatoire HydroAcoustique de la SISmicité et de la BIOdiversité) observatory has been continuously recording data from the sound fixing and ranging (SOFAR) channel since 2010 (Samaran et al., 2013; Leroy, 2017).

Material developed as a part of this thesis work is freely available online.

Matlab code Matlab codes are shared on GitHub with DOIs under a MIT license:

- Passive SMF Package repository (Chapter 3)
  `https://leabouffaut.github.io/SMF_package/`
  DOI: 10.5281/zenodo.3613788
- Tonal detector comparison code (Chapter 5)
  `https://leabouffaut.github.io/tonal_detectors/`
  DOI: 10.5281/zenodo.3469389

Annotated datasets Datasets annotated for performance evaluation in Chapter 4 (845 ABW calls with varying SNRs) and training of the automatic transcription method in Chapter 5 (more than 4000 BW tonal units) are hosted on Zenodo and can be referred to as:

- Léa Bouffaut. (2020). Western Indian Ocean blue whale dataset (Version v1.0) [Data set]. *Zenodo*. DOI: 10.5281/zenodo.3624145

Publications and communications

Peer-reviewed journal articles

- **L. Bouffaut**, S. Madhusudhana, V. Labat, A. Boudraa and H. Klinck "A performance comparison of tonal detectors for low frequency vocalizations of Antarctic blue whales," accepted for publication in the *J. Acoust. Soc. Am* on Dec 31, 2019,

- R. Dréo, **L. Bouffaut**, E. Leroy, G. Barruol, F. Samaran "Baleen Whale distribution and seasonal occurrence Revealed By An Ocean Bottom Seismometer Network In The Western Indian Ocean," in *Deep Sea Research Part II: Topical Studies in Oceanography*, vol. 161, pp. 132-144 (2019).

- **L. Bouffaut**, R. Dréo, V. Labat, A. Boudraa and G. Barruol "Passive stochastic matched filter for antarctic blue whale call detection," in *J. Acoust. Soc. Am*, vol. 144, no. 2, pp. 955-965 (2018).

Peer-reviewed conference proceedings

- **L. Bouffaut**, S. Madhusudhana, V. Labat, A. Boudraa and H. Klinck, "Transcription automatique des chants de baleines bleues dans différents contextes acoustiques,"in *GRETSI 2019*, France, pp. 2-5 (2019).

- **L. Bouffaut**, R. Dréo, V. Labat and A. Boudraa "Filtrage Adapté Stochastique passif pour la détection de plongeurs," in *GRETSI 2017*, France, pp. 2-5 (2017).

- **L. Bouffaut**, R. Dréo, V. Labat, A. Boudraa and G. Barruol, "Antarctic Blue Whale Calls Detection Based on an Improved Version of the Stochastic Matched Filter," in *EUSIPCO 2017*, Greece, pp. 2283-2387 (2017).

- **L. Bouffaut** and A.-O. Boudraa, "Passive stochastic matched filter : Application to scuba divers detection," in *ICASSP 2017*, USA, pp. 2-3 (2017).

Conference proceedings

- **L. Bouffaut**, S. Madhusudhana, V. Labat, A. Boudraa and H. Klinck, "Automated blue whale song transcription across variable acoustic contexts," in *IEEE OCEANS'19 Marseille*, student poster competition, France, pp. 2-7 (2019).

- **L. Bouffaut**, R. Dréo, V. Labat, A. Boudraa and G. Barruol "Remote blue whale call detection using a passive version of the stochastic Matched Filter", in *UACE 2017*, Greece, pp. 2-7 (2017).
- R. Dréo, **L. Bouffaut**, L. Guillon, V. Labat, G. Barruol and A. Boudraa "Antarctic Blue Whale localization with Ocean Bottom Seismometer in Southern Indian Ocean", in *UACE 2017*, Greece, pp. 2-7 (2017).

Conference presentations

- **L. Bouffaut**, R. Dréo, V. Labat, A. Boudraa and G. Barruol "Performances du Filtrage Adapté Stochastique passif pour la détection des vocalises de baleines bleues Antarctique," in *SERENADE 2018*, Brest, Octobre 2018.
- **L. Bouffaut**, R. Dréo, V. Labat, A. Boudraa and G. Barruol "Passive Stochastic Matched Filter for Antarctic Blue Whale call detection: performance analysis on highly variable SNR ground-truth dataset," in *DCLDE workshop*, Paris, June 2018.
- **R. Dréo**, L. Guillon, L. Bouffaut, V. Labat, "Tracking Antarctic Blue Whales over a mountainous area with an Ocean Bottom Seismometers array: dealing with 3D propagation effects," in *DCLDE workshop*, Paris, June 2018.
- R. Dréo, **L. Guillon** and L. Bouffaut "Mise en évidence d'effets 3D dans la détection de baleines bleues antarctiques par un réseau d'OBS," in *CFA*, Le Havre, Avril 2018.
- **L. Bouffaut** "On the link between demodulated acoustic classification features of small boats and noise generation processes", in *AFPAC 2017*, Marseille, January 2017
- **L. Bouffaut**, R. Dréo, V. Labat, A. Boudraa "Filtrage adapté stochastique appliqué à la détection de plongeurs", in *JJCAB 2016*, Marseille, Novembre 2016.
- **L. Bouffaut**, R. Dréo, V. Labat, A. Boudraa "Filtrage adapté stochastique appliqué à la détection de plongeurs", in *SERENADE 2016*, Brest, Octobre 2016.

Awards

- 2019, *OCEANS student poster competition* First Prize and Norman Miller Award for the poster entitled:
  "Automated blue whale song transcription across variable acoustic contexts"
- 2019, *Fondation de la Mer* PhD scholarship
- 2018, *GdR ISIS* PhD travel grant
- 2018, *Ecole Navale* travel grant
- 2016, winning team collaborator for the first Ocean Hackathon organized in Brest

Chapter 1

Passive acoustic monitoring of blue whales: survey, study and protect the oceans

L'homme et la baleine ne se fréquentent pas. Leurs rencontres sont hantées par la mort - baleines échouées ou scènes de chasse, de dépeçage dans des flots de sang - ou se réduisent à des éclats fugitifs - un souffle, une bosse sur la mer, au mieux un saut, une volte. La vie des baleines se déroule hors de notre vue.

François Garde
La baleine dans tous ses états (2015)

Contents

1.1 Blue whales survey methods

BWs inhabit vast areas across all oceans but are mostly found in the Southern hemisphere. According to global migration patterns (Figure 1.1), the majority of the BW population is feeding in the abundant polar waters during the boreal or austral summer. After the end of the feeding season, they return to their wintering and breeding areas, in tropical-to-temperate waters. Data from the Antarctic waters (first from hunting expeditions, now from explorations from around the world) are available but, only a few information exists on BWs distribution during the austral winter, especially in the Indian Ocean.

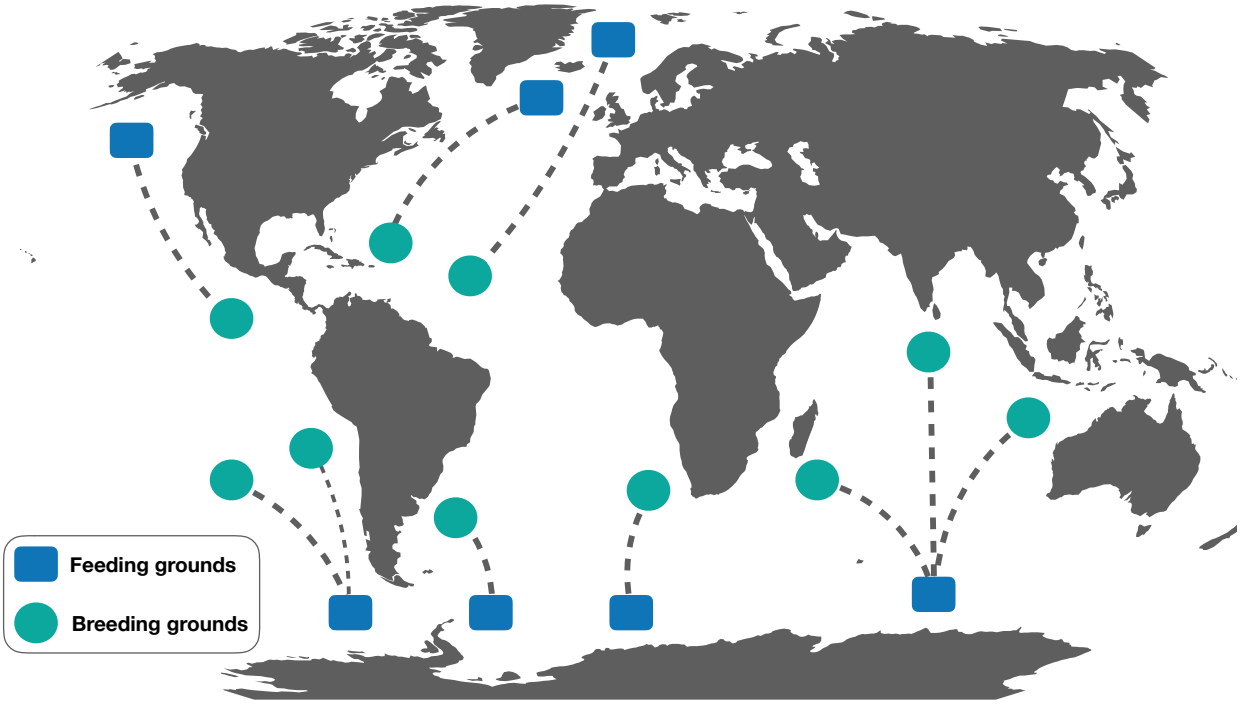

Figure 1.1: Blue whale global migratory map and patterns (adapted from: https://seethewild.org/).

Most research in the past was limited to observation of surface behavior (Zimmer, 2011) but was hindered by of multiple factors. First, the number of individuals is limited. Besides, they are often highly mobile and breach the surface only for a short period, for breathing. In addition, as shown in Figure 1.1, BWs spread over wide and remote areas where severe weather conditions prevail most of the year and, do not facilitate the access for sightings. Under these challenging conditions, how to effectively monitor BWs?

Thankfully, the advent of new autonomous technologies for monitoring helped to make significant advances in the marine mammal behavior field. The following paragraphs briefly introduce different modern techniques for BW surveys.

1.1.1 Sightings

There are multiple ways for sighting whales. From shore or a boat, countless associations and organizations offer whale watching spots and contribute to the local tourism economy. In the

scientific community, surveillance by trained observers scanning the sea surface is still popular and efficient to enumerate, recognize, and follow resident whales or whales that come near shore. Individual marks or specific fin shapes are used for individual identification (e.g, photo identification) (Hammond et al., 1990; Barlow et al., 2018). Sightings can also be completed with mark-recaptures, information (metallic mark with a unique serial number), to help follow whales movements. However, it provides sparse observations and can not be extensively used in the remote area where most BWs live.

Aerial surveys provide an instantaneous measure of abundance and help to measure seasonal occurrence and general distribution of whale species (Gill, 2002). As a part of governmental coastal management organization's duties, different NOAA (National Oceanic and Atmospheric Administration) divisions deploy such surveys regularly. They cover relatively small areas in the presence of critically endangered species such as the North Atlantic Right Whale along the Cape Cod bay coastline or the St Lawrence Gulf, or, to survey critical habitat (i.e., Alaskan coastline) (Cole, 2019). However, these deployments are expensive as they require human resources and flying equipment for only limited range and punctual observations.

Over the past decade, the booming of drones also reached the marine mammal monitoring world. With the benefits from autonomous systems such as reduction of the costs and safety, combined with an appreciated viewing of marine mammals from aerial platforms, this new technique can not be overlooked. Besides, multi-rotor unmanned aerial systems such as the hexacopter in Figure 1.2 are capable of following target individuals. On the downside, autonomous aerial systems are still limited in flight time, although increasing battery power density and rapid advances in charging technology are likely to increase these flight times in the near future (Nowacek et al., 2016). They also require to be launched from the shore or a vessel, limiting the monitoring range.

Figure 1.2: Unmanned aerial system survey of a humpback whale (for reference, the drone is 500x230 mm) source: (Nowacek et al., 2016)).

Recent works showed that whales could even be monitored from space with satellite imagery (Fretwell et al., 2014; Cubaynes et al., 2019). The most recent study (Cubaynes et al., 2019) focused on baleen whales in different oceanic regions. The equipment achieved a maximum spatial resolution of 31 cm, sufficient for body outline and fluke identification and a first step towards automated whale detection methods from satellite imagery. The repercussions of such discoveries enable the studies of whales in remote and inaccessible areas where traditional survey methods are limited or impractical. However, they require considerable satellite coverage and cutting-edge technology that are not yet globally accessible.

All these visual methods share the same limitations: they are less effective in dim light, at dusk or dawn and, are impaired by heavy weather conditions. Independent from their technicality or complexity, the sea surface impedes visual surveys since it can occlude the most significant part of the whale life, underwater.

Nevertheless, the sound generation and auditory systems of marine mammals evolved to accommodated their submarine environment. They base their daily life on acoustics and consequently, eavesdropping is beneficial to collect information from afar. PAM is the label that includes all methods exploiting marine mammals soniferous abilities and is the subject of the following section.

1.1.2 ... and why not use sounds?

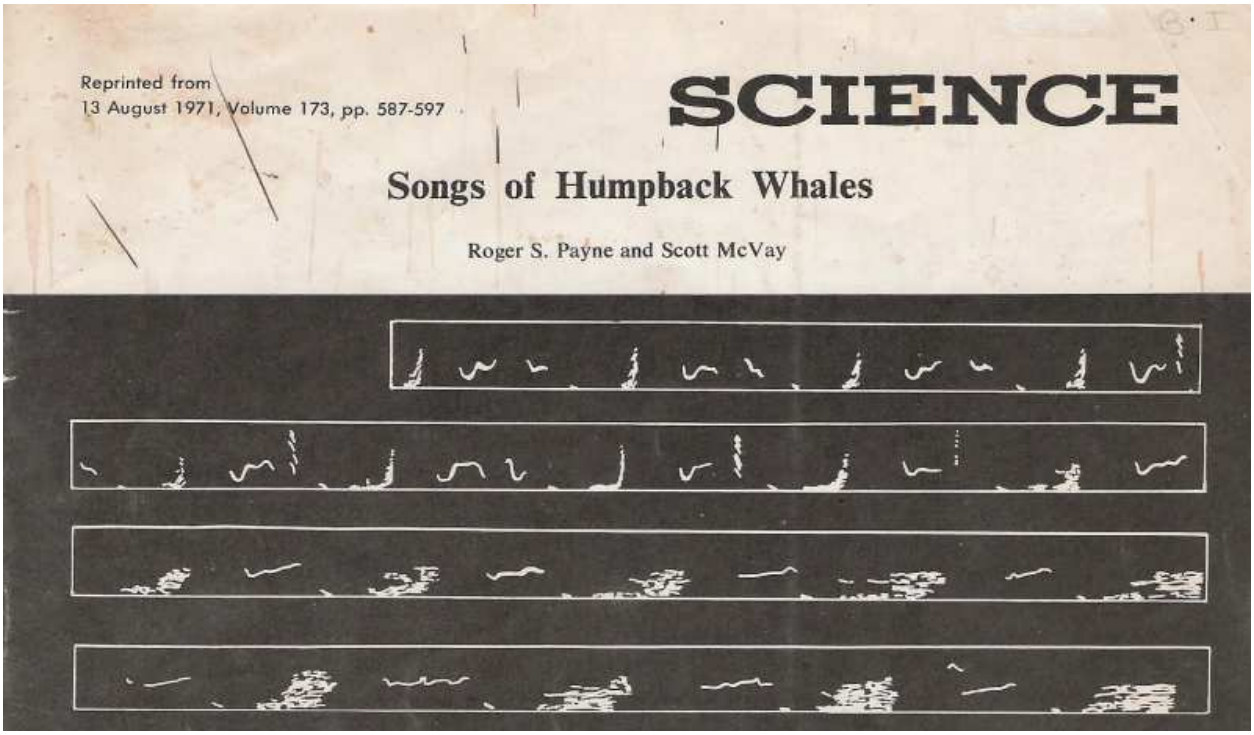

Figure 1.3: Songs of Humpback whales made the cover of the magazine Science of 13 August 1971 (Payne and McVay, 1971).

In the mid-XXth century, the analysis of acoustic recordings from U.S. Navy SOFAR station in Kaneohe Bay, Hawaii led to the following conclusion "*These sounds (...) have a rather musical quality. There is a marked seasonal variation in the production of these sounds (...) coincident with the seasonal variation of whales in the area, and this feature, plus the characteristics of the sounds themselves, has led to the belief that they are produced by whales*" (Schreiber, 1952). According to Payne and McVay (1971), this was the first-ever notification of humpback whales songs. In the following years, the scientific community realized that some of the "*intense, low-frequency, underwater sounds*" that were recorded worldwide were "*apparently of biological origin*" (Walker, 1963). The enthusiasm that followed these first discoveries led to the systematic characterization of cetaceans sounds (Schevill, 1964), with the first description of BW sounds, almost ten years later in Cummings and Thompson (1971). The same year, thanks to the analytical work of Payne and McVay (1971), and their organization into a hierarchical structure, humpback whale sounds made the cover of Science magazine, advertising for the new field of bioacoustics and PAM (Figure 1.3).

As a result of this pioneering work, there is a global recognition in the biology community of the usefulness of PAM for studying cetaceans in their natural environment. In addition, the work summed-up in McDonald et al. (2006) and, updated as a worldwide collaborative effort in Širović et al. (2017) (Figure 1.4) underlines bio-geographic differences in BW species sounds allowing species recognition from any PAM recording.

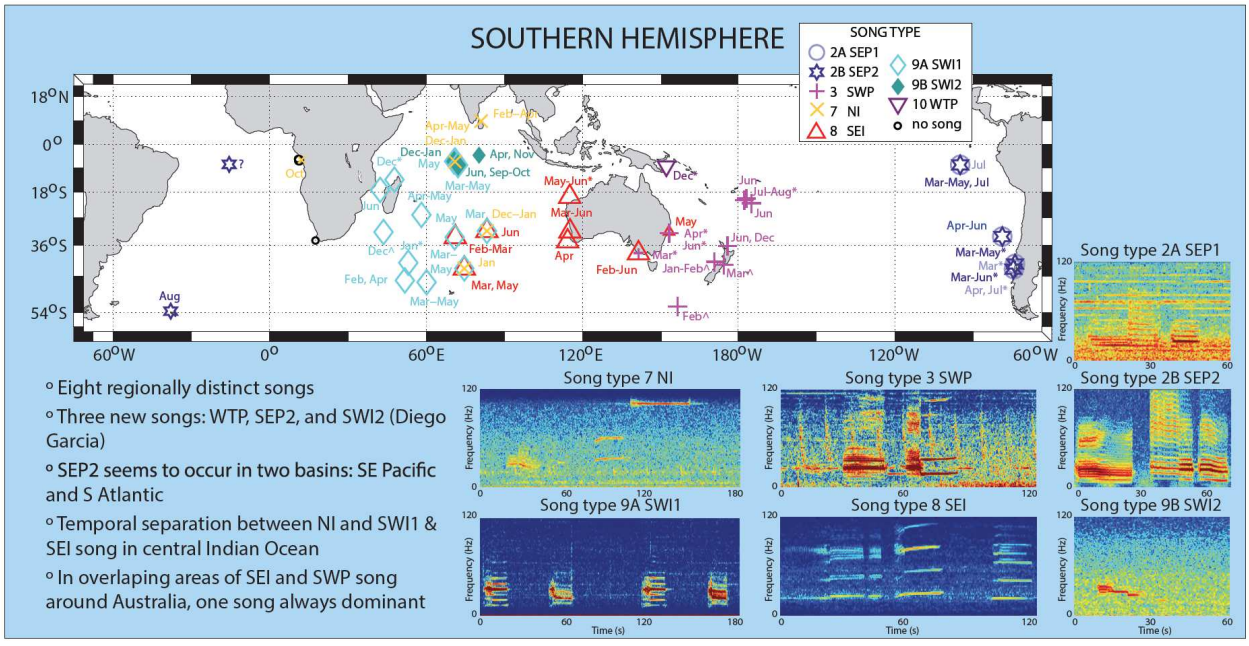

Figure 1.4: Blue whale song diversity, distribution and representation in the Southern hemisphere (Širović et al., 2017).

As heat, light and other forms of electromagnetic energy are severely attenuated in the water, acoustics is thus the most effective method for marine mammals and whales to perform the various functions associated with their life cycle. Consequently, eavesdropping provides

unequaled insights into their underwater activities without interfering at all with their natural behavior. When animals are vocally active, the study of their sounds can favor the estimation of their location, of their movements and even, at a broader scale, help into assessing seasonal distributions and density. In addition and, unlike sighting techniques, PAM is not dependent (much) on the weather or brightness, it can provide and record information all day long and all year round, in various contexts, from open waters to beneath the ice, from coastal to remote areas.

Specifically, BWs produce high intensity and low-frequency vocalizations, probably for long-range acoustic signaling (Payne and Webb, 1971). Their specific signatures can exceed a 100 km range (Payne and Webb, 1971; Frank and Ferris, 2011). These extraordinary vocal abilities are a bargain for PAM: large areas can be monitored with a single sensor.

The tremendous strengths of combining PAM with modern unmanned autonomous systems are unequivocal. The ocean and, therefore, marine mammals can now be acoustically monitored over long time scales. As rightfully said by Zimmer (2011), in an introduction to his book Passive Acoustic Monitoring of Cetaceans: "*As an interdisciplinary subject, successful PAM combines physics, technology, and biology* ". The relationship between these three topics is discussed throughout this chapter. Primarily, an overview typical PAM equipment is given in section 1.2.

1.2 Passive acoustic monitoring equipment

As there are multiple ways to conduct visual surveys of marine mammals, there are also multiple methods and equipment for PAM (Figure 1.5).

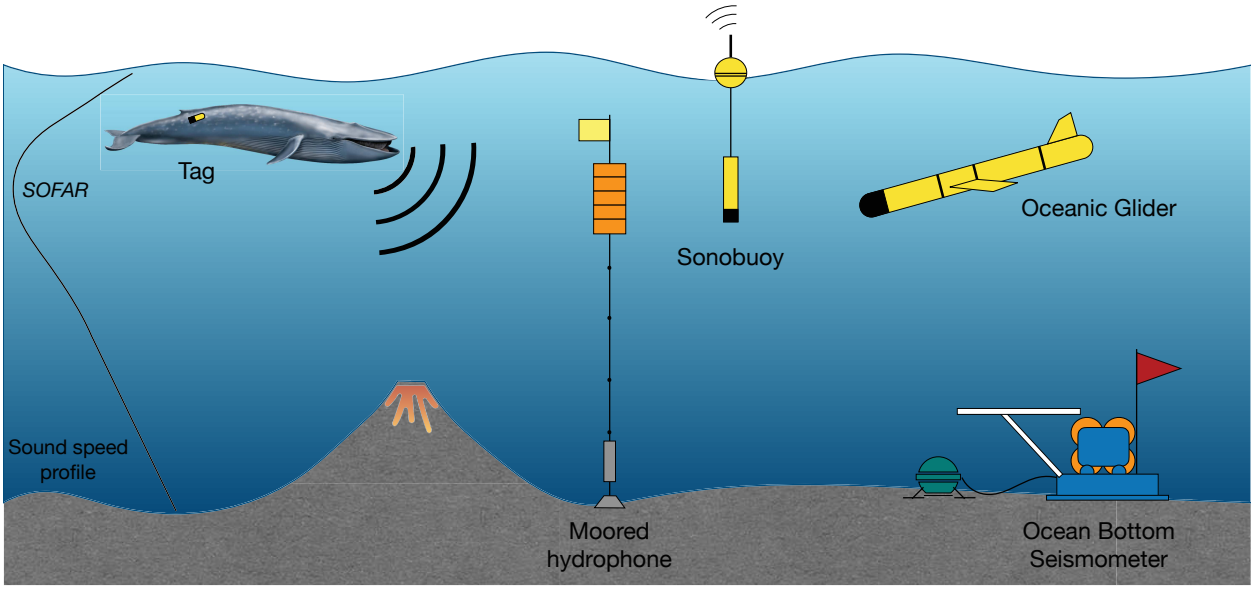

Figure 1.5: Schematics of the different types of PAM recorders.

Hydrophones are the simplest of PAM systems. Lowering a single hydrophone in the water from a boat is the easiest way to listen to whales and dolphins. Easy to implement this approach requires minimal hardware and software. It can be improved by the use of multiple hydrophones such as towed arrays, to ease localization and tracking, and contribute to cover larger areas. These types of "dropped-hydrophone(s)" surveys, efficient in coastal areas are not adapted to BW surveys in remote locations where autonomous systems are preferred. In addition, the low-frequency noise from the towing ship severely limits the detection range, and is the reason why they are rarely used for PAM of BW. The following section gives a brief but non-exhaustive summary of common autonomous instrumentation used today for PAM. A complete review of existing fixed autonomous systems for PAM is presented in Norris et al. (2010) and, a - more recent- full review of real-time instrumentation for PAM is the subject of Baumgartner et al. (2018).

1.2.1 Autonomous fixed recorders

Fixed (moored) and autonomous PAM installations allow for cost-effective long-term monitoring of delimited areas, for extended intervals (e.g., months - years – decades). They are usually separated into two groups depending on the location of the hydrophone in the water column: sound channel-moored hydrophones and bottom-moored hydrophones. When close to shore either systems can be cabled for real-time analysis, overcoming any data storage or limited power supply limits (Ward et al., 2017; Hendricks et al., 2018; Baumgartner et al., 2019).

1.2.1.1 Bottom-moored hydrophones

According to Norris et al. (2010), the first autonomous recorders widely deployed across oceans were ocean bottom seismometers (OBSs). These instruments, typically equipped with a seismometer, a data logger and batteries to power the device (and later an omnidirectional hydrophone) were sunk to the seafloor. Data was processed after recovery. OBSs are usually set to have a maximum sampling rate of 100 Hz. McDonald et al. (1995) was the first to use OBSs for PAM of blue and fin whales.

Yet, OBSs were too expensive to be purchased and deployed by PAM research teams in large quantities and, most bioacoustics team developed their own equipment. Nowadays there is a wide variety of bottom-mounted (or moored) hydrophones covering various frequency ranges such as Cornell Marine Autonomous Recording Units (MARUs - 2 Hz–30 kHz frequency response) (Williams et al., 2014), Rockhoppers [1] or Scripps HARPs (High-frequency Acoustic Recording Package) [2] (Širović et al., 2004, 2007).

Collaborations between the bioacoustics and geophysical scientific communities expanded the possibility of using OBSs for PAM of baleen whales e.g., Dunn and Hernandez (2009); Frank

[1] http://www.birds.cornell.edu/brp/rockhopper/
[2] http://cetus.ucsd.edu/technologies_AutonomousRecorders.html

and Ferris (2011); Harris et al. (2013); Brodie and Dunn (2015); Dréo et al. (2019). Table 1.1 presents typical OBSs characteristics.

Autonomy	8 month to a year depending on the sampling frequency
Depth	maximum 6000 m
Frequency	[0 – 50] Hz

Table 1.1: Example of the RHUM-RUM OBS characteristics (Stähler et al., 2016).

1.2.1.2 Sound channel-moored hydrophones

The deep sound channel is typical of mid-latitude oceans. Physically, in the upper part of the water column, the speed of sound diminishes as the temperature drops. Around depths of 1000 m, when the temperature is constant, the sound speed increases with the hydrostatic pressure (§ 1.4.2). This inflection point provides a well-defined minimum of celerity and, if an acoustic source is placed at the depth of this minimum, the sound will be "trapped" in this low sound speed layer due to the effect of refraction. The sound can then travel very long distances with little attenuation. This property of the SOFAR, discovered by Ewing and Worzel (Worzel et al., 1948), was showed to be applicable to cross-ocean sound propagation by the Heard island experiment (Munk et al., 1994) using a signal near 57 Hz. For that reason, many recorders are positioned in the SOFAR channel, for example as moored hydrophones floating in the water column, e.g., Southwest of Australia (Gavrilov et al., 2011; Gavrilov and McCauley, 2013; Ward et al., 2017) or in the Western Indian Ocean (Samaran et al., 2013; Leroy et al., 2016, 2018) (their characteristics are presented on Table 1.2).

Autonomy	1 to 2 years
Depth	between 500 and 1500 m
Frequency	[0 – 120] Hz

Table 1.2: Example of the OHASIS-BIO (Observatory in the Indian Ocean) network sound-channel hydrophones characteristics (Leroy et al., 2018).

1.2.2 Autonomous moving recorders

Autonomous moving recorders are efficient to acquire short term data. Because they do not often require large scientific vessels to be launched, they can be deployed in areas where access is difficult. Moreover, when launched or retrieved, they can be combined with other non-acoustic techniques such as visual observations (or even bio sampling) for multi-approaches monitoring. Contrary to moored devices, they can provide information as thy move, covering a vast perimeter or by monitoring a specific individual.

1.2.2.1 Sonobuoys

Initially used for military purposes, sonobuoys are one of the most common equipment in the family of acoustic buoy systems for PAM activities (Clark et al., 1986; Miller et al., 2013). The concept is simple: they relay the sound received by a single hydrophone via a radio signal to a nearby receiver (shipboard or airboard surveys). The additional integration of an orthogonally oriented pressure vector sensor and a magnetic compass allows the DIFAR (Directional Frequency Analysis and Ranging) sonobuoy to estimate bearing angles of a sound of interest. A method to estimate the drift direction and speed of a directional sonobuoy is presented in Miller et al. (2018). Cross bearings of two or more sonobuoys allow real-time tracking of target species (Miller et al., 2016; Garcia-Rojas et al., 2018). When batteries run out, the sonobuoy sinks to the bottom of the ocean, in that sense they are disposable equipment. An example of technical sheet is given in Table 1.3.

Autonomy	between 30 min and 8 hours
Depth	between 30 and 300 m
Frequency	calibrated omnidirectional hydrophone [5 – 20k] Hz
	DIFAR [5 – 2.4k] Hz

Table 1.3: Example of the 53-F DIFAR sonobuoys characteristics (Rankin et al., 2019).

1.2.2.2 Gliders

In the early 2000, there was a revolution in ocean science: the advent of autonomous underwater vehicles. PAM was not left behind with, in the summer 2006, the first successful deployment of broadband (5 Hz to 30 kHz) omnidirectional hydrophone seaglider. It recorded calls from BWs and humpback whales as well as odontocete sounds (Moore et al., 2007).

Underwater gliders are cost-effective buoyancy-driven vehicles moving vertically up and down in such a way that, at the same time, they gain horizontal distance. They produce almost no self-noise and minimal low-frequency flow noise (Baumgartner et al., 2018). Thanks to continuous efforts and development, they can now be used for real-time monitoring (Baumgartner et al., 2013), and be deployed in previously inaccessible locations such as the Mariana Trench where baleen whale calls were recently recorded (Nieukirk et al., 2016). The characteristics of the glider used in this experiment are presented in Table 1.4. Their continuous motion is slow compared to most marine mammals movements. Moreover, they can offer good spatial coverage in comparison to fixed sensors and are easier to deploy. These features are the reason why gliders are progressively considered for classical conservation applications such as animal density estimation (Marques et al., 2013).

Autonomy	Technically almost unlimited, virtually months
Depth	between 0 and 1000 m
Frequency	between 15 Hz and 97 kHz

Table 1.4: Example of the Seaglider™ characteristics (Nieukirk et al., 2016).

1.2.2.3 Bio-logging or TAGs

Bio-logging or more commonly "tagging" consists in deploying a high-resolution multi-sensor device directly on an individual in order to document a portion of its life. In addition to the hydrophone, acoustic tags can be equipped with a hydrostatic pressure sensor (depth), an accelerometer, a temperature sensor and, a compass (see models Acousondes [3] and Dtags [4]). Some of these devices can also be equipped with a radio antenna for communication. These methods are designed to be non-invasive, e.g., fixed using suction pads in order to satisfy modern ethical standards and least affect animal behavior.

Figure 1.6: Ari Friedlaender deploys a multi-sensor suction cup tag on a humpback whale in Antarctica's Wilhelmina Bay (Photo by Ari Friedlaender, source: https://www.bates.edu/).

Because they require contact with the whale (Figure 1.6), they are often combined with skin samples for sex or DNA determination. Acoustic tags have provided information on call types, call frequencies, cue production rates, seasonality and sex differences in vocalizations for BWs (Stimpert et al., 2015). More recently, thanks to their combination with different sensors, tags have been used to estimate the depth(s) at which BW sing, evaluated about 30 m (Lewis et al., 2018). Characteristics of the DTAG3 used in Stimpert et al. (2015) is presented on Table 1.5.

[3] http://www.acousonde.com
[4] https://www.soundtags.org/dtags/

Autonomy	up to 3 days depending on the sampling frequency
Depth	500 to 3000 m
Frequency	between 10 Hz and 20 kHz

Table 1.5: Example of the DTAG3 characteristics (Stimpert et al., 2015).

1.2.3 Discussion

Table 1.6 summarizes all the presented visual and acoustic survey methods and, compares their ability to conduct continuous surveys, display common time-scales, if they are communicating and if they can identify an individual or a species. Unique sensors are considered, and tracking opportunities are not yet discussed. In comparison to sightings, PAM offers continuous and

	Method	Continuous	Duration				Real-time	Identification	
			hours	day	months	year		Individual	species
Sightings	Visual survey		■				■	■	■
	Aerial surveys		■				■		■
	Drone		■				■	■	■
	Satellite imagery					■	■		■
PAM	Auto. fixed recorders	■				■	▒		■
	Sono buoy	■	■				■		■
	Glider	■			■		▒		■
	TAG	■		■			▒	■	■

Table 1.6: Baleen whale survey methods recap: white = no, blue = yes and gray = maybe.

multi-time scale opportunities to survey BWs without intruding their environment. However, Table 1.6 also shows that, because sounds are subspecies-specific, it is impossible to use a unique omnidirectional sensor to identify individuals within a species. Besides, to identify new species-sound connections, cross-referencing with, e.g., visual observations or known migration patterns, is essential. This is why all the methods above are complementary, to learn about, survey, and protect efficiently marine mammals using information based multiple time, space, and resolution scales (Nowacek et al., 2016). A brief overview of PAM fall on for animal ecology is given in section 1.3.

1.3 Applications of passive acoustic monitoring

The first and most direct application of PAM is to study animal sounds as part of their communication. Call and song description, as well as the identification of new signatures (Leroy et al., 2017a; Ward et al., 2017), can help to identify connected species or subspecies (McDonald et al., 2006) and, identify vocally active populations (Balcazar et al., 2015; Brodie and Dunn, 2015). Additional information on source levels (Cummings and Thompson, 1971; Samaran et al., 2010c; Širović et al., 2007; Weirathmueller et al., 2013) intervene in the understanding one can have on sound production mechanisms (Adam et al., 2013; Cazau et al., 2016; Adam et al., 2018). Tagging equipment permits the observation and sampling of individual behavior (Lewis et al., 2018). The study of song structure and their variations recently showed gradual synchronization and modification of song rhythms for two Californian fin whales populations over long time scales, indicating some exchanges or visitation among populations (Širović et al., 2017). More recently Jolliffe et al. (2019) demonstrated an increase in song diversity of pygmy BWs, consistent with cultural evolution but, the consequence of (yet) unknown factors. Thorough call analysis also revealed, in addition to seasonal frequency variation, a constant decline of large whale call frequencies over the last decade (McDonald et al., 2009; Leroy et al., 2018) but explanations are yet to be found.

Based on call analyses and pattern recognition algorithms, PAM can be used in animal ecology (Zimmer, 2011):

- to estimate the abundance, the total number of animals in a given area or, equivalently, a population density of the species of interest;
- to mitigate the impact of anthropogenic activity on marine mammals or, to estimate cetacean presence/absence, necessary for risk mitigation.

Note that the method employed for ecology analyses depend on the number of sensors (mono-sensor or multi-sensor), their type (omnidirectional or vectorial) and their spatial distribution. Some configurations enable to localize, track and differentiate individuals.

Habitat analysis is based on the idea that whales are not uniformly distributed throughout the world and that their distribution shows spatial and temporal heterogeneity (Zimmer, 2011). PAM can therefore be used to determine whale-inhabited waters and identify the populations (Delarue et al., 2009; Samaran et al., 2010b; Cerchio et al., 2015; Garcia-Rojas et al., 2018). It can help draw geographical and seasonal patterns across vast areas (Širović et al., 2004; Stafford et al., 2011; Tripovich et al., 2015; Leroy, 2017), refine known migratory corridors (Thomisch, 2017) and, identify and locate species-specific feeding and breeding grounds (Gedamke et al., 2007; Samaran et al., 2013). A better knowledge of whale territories is essential for conservation purposes, as a first step before taking measures on the management of specific areas, for example, to mitigate ship strikes and entanglements (Irvine et al., 2014; Harcourt et al., 2019).

Last but not least, PAM can help to evaluate whale response to global anthropogenic noise (Melcon et al., 2012). Indeed, anthropogenic noise has the potential to mask whale communication and therefore, to limit their communication range. It has serious consequences since, according to Williams et al. (2014), acoustic masking is "a qualitatively similar stressor to habitat loss." For instance, ship traffic can directly affect whale acoustic environment (Redfern et al., 2017) and have been seen to cut down by 87% the communication space of Bryde's whales on routine passages (Putland et al., 2018). The study of transient sounds repercussion on marine life also demonstrated the disturbance of BWs behavior under the effect of mid-frequency military sonar (Goldbogen et al., 2013; Harris et al., 2018). However, the effect of noise on animals and the in-place regulations are controversial subject (Gomez et al., 2016).

To efficiently monitor blue whales, knowledge of fundamental underwater acoustic propagation is essential, for all intended applications. Physical properties such as the spreading of the acoustic wave energy or attenuation need to be taken into account in the estimate of PAM system detection ranges. Underwater acoustics key elements are thus discussed in Section 1.4.

1.4 Underwater acoustics

1.4.1 Geometrical or Modal propagation?

The wave equation showing the relationship between the spatial and temporal variations of the acoustic pressure p is given by

$$\nabla^2 p - \frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} = 0, \tag{1.1}$$

where c represents the speed of sound and where the spatial differences are described by the Laplacian operator ∇^2, the form of which depends in the coordinate system chosen for the application.

In underwater acoustics, as presented in Figure 1.7, there are two approaches to understand and integrate propagation in models as well as in the interpretation of observations: the geometrical and the modal approach. The geometrical approach is generally considered as a "high frequency" approximation derived from the eikonal equation. Under these assumptions the acoustic wavefront is represented by a finite number of rays, each one following their trajectory depending on the position of the source, the emission angle(s), and the sound speed (Urick, 1983). However, other solving methods exist. For example, the modal approach requires an accurate resolution of the acoustic pressure field (same order as the wavelength) and is therefore generally considered as a low-frequency approach.

In underwater acoustics, the modal approach is generally considered when the wavelength and the characteristic dimension (the water column height, H) are of the same order: it considers the standing waves over the vertical ocean. When considering BW sounds in the deep ocean, the maximum wavelength of their sounds in the order of 150 m ($c = 1500$ (m/s) and $f = 10$ Hz)

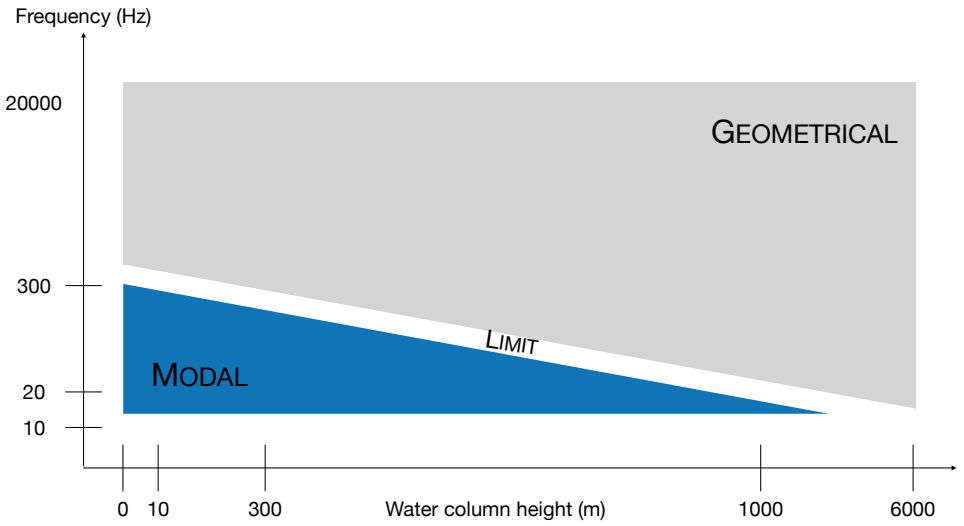

Figure 1.7: Range of validity of modal and geometrical propagation as a function of frequency and watercolumn height (adapted from Josso (2010)).

is more than 20 time smaller than average ocean depth ≃ 3500 m. Under these considerations, geometrical assumptions are used in this study to model the propagation of BW sounds.

1.4.2 Sound speed profile and refraction

1.4.2.1 Sound speed profile

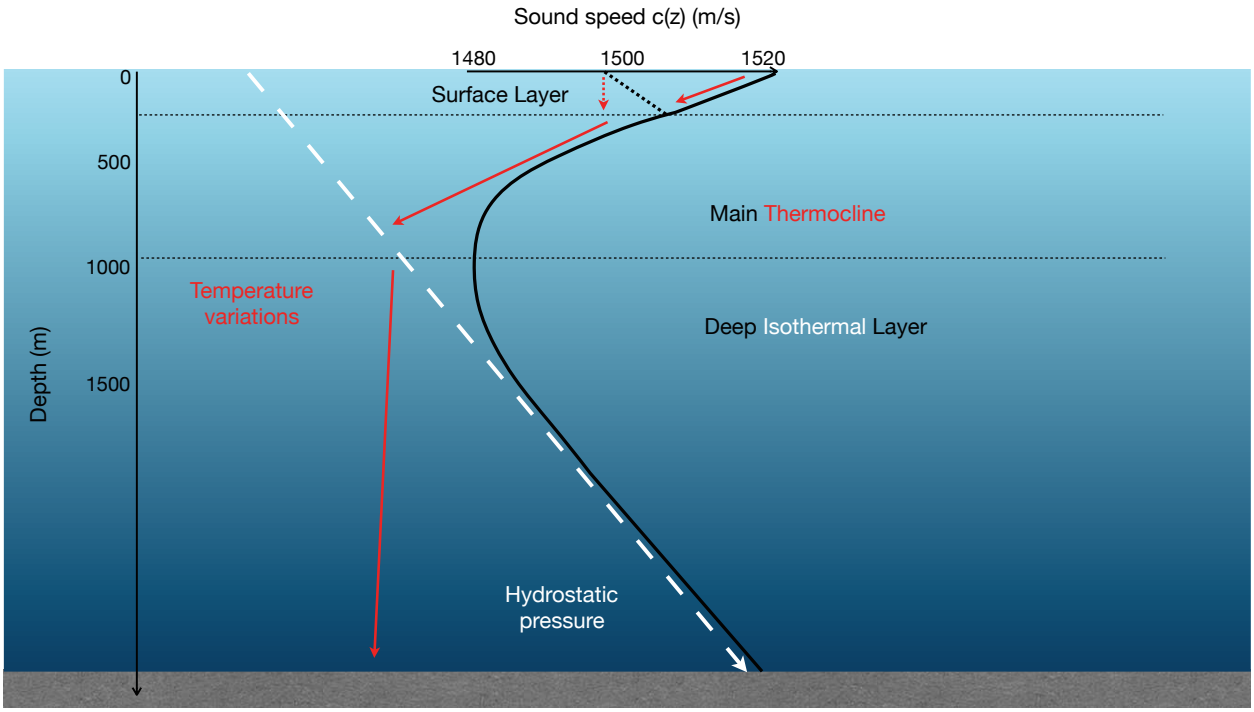

Figure 1.8: Sound speed profile and influential parameters (adapted from Urick (1983)).

One of the most important characteristics of underwater acoustic propagation results from

the fact that the sound speed is not constant. In general, c is determined by a complex relationship between salinity, temperature, and hydrostatic pressure, the last two being the most influential. In a specific area, it mostly varies with depth (therefore denoted $c(z)$), showing a stratification of the ocean (Urick, 1983). An example of typical mid-latitudes deep-ocean profile is depicted in Figure 1.8, showing the two principal characteristics of the sound speed profile: the thermocline where $c(z)$ drops with the temperature and, the isothermal layer, where $c(z)$ increases with the hydrostatic pressure. Variations in surface temperature (due to latitude, seasonal changes or weather) change the sound speed profile.

1.4.2.2 Refraction and the Snell's law

The stratification of the ocean (in depth) induced by variations in the water column parameters and, especially sound speed variations, has an effect on the propagation of the acoustic wave: it is subject to refraction. Acoustic rays are deviated according to the Snell's law

$$\frac{\cos\theta_1}{c_1} = \frac{\cos\theta_2}{c_2}, \tag{1.2}$$

where indices 1 and 2 correspond to two different depths in the water column with $z_1 < z_2$, and θ represents the grazing angle (Figure 1.9). Then considering a simple sound speed profile,

- if $c_1 < c_2$, the grazing angle diminishes ($\theta_2 < \theta_1$) until total refraction and, the ray bends towards the surface, to the minimum sound-speed,
- if $c_1 > c_2$, the grazing angle increases ($\theta_2 > \theta_1$) and, the ray bends towards the ocean floor, again to the minimum sound-speed.

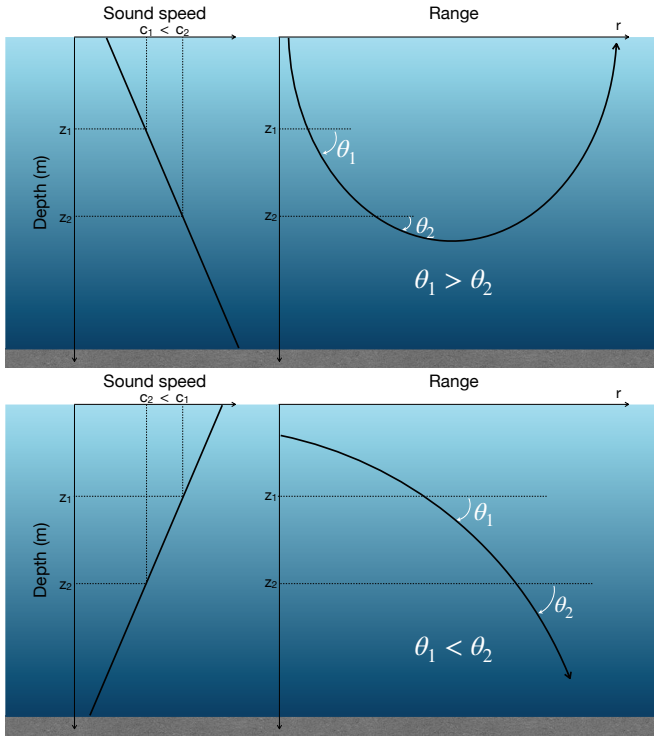

Figure 1.9: Illustration of acoustic ray deviation depending on the sound speed profile.

In other words, the geometrical model of the underwater acoustic propagation is a complex model where, acoustic beams do not propagate in straight lines but, bend towards the minimum sound speed under the effect of refraction. They can reflect at the surface and bottom boundaries, as illustrated in Figure 1.10. Consequently, a receiver can record rays that traveled different paths: direct or with bottom and surface reflection(s). At low frequencies, the sea surface can be seen as a perfect screen where the acoustic wave reflects with only a phase change. However, some energy might be transmitted into the seafloor, depending on the nature of the material.

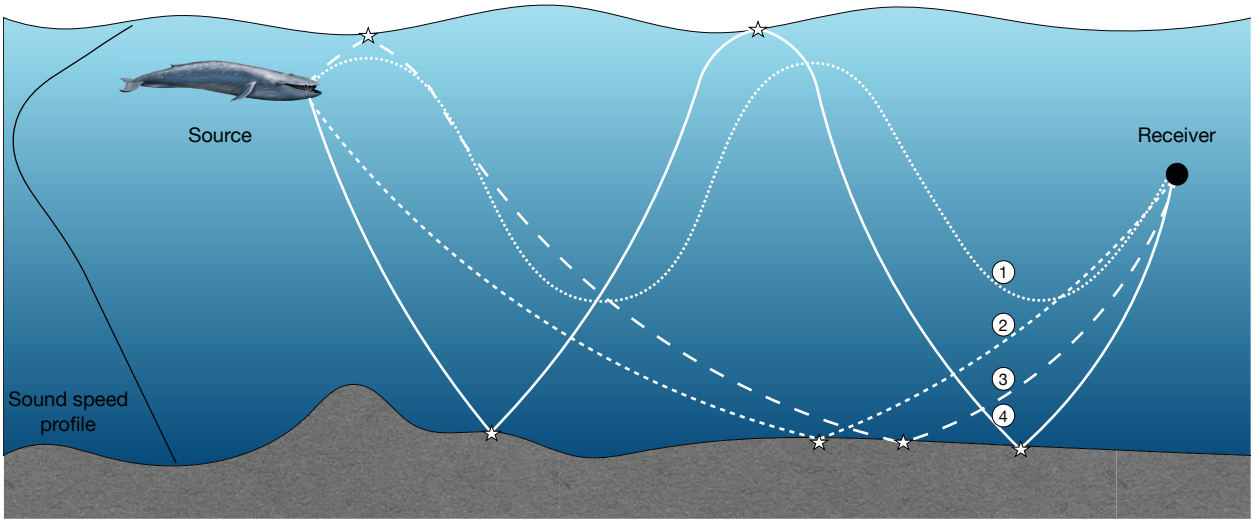

Figure 1.10: Illustration of different acoustic paths: (1) Direct, (2) one bottom reflection (3) one bottom and one surface reflections and, (4) two bottom and one surface reflections.

In order to carry out numerical simulations of the propagation, numerous software have been developed, the most popular being Bellhop (Porter and Bucker, 1987). This program, developed by Porter in the 1980s is freely available online[5]. An example of propagation with a source placed at a 30 m depth and a statistical deep ocean sound-speed profile for the Indian Ocean in May is presented in Figure 1.11. Acoustic rays fill the entire water column with multiple bottom-surface reflections. Ray tracing illustrates 2D propagation space coverage, but, the analysis can be completed with energetic considerations.

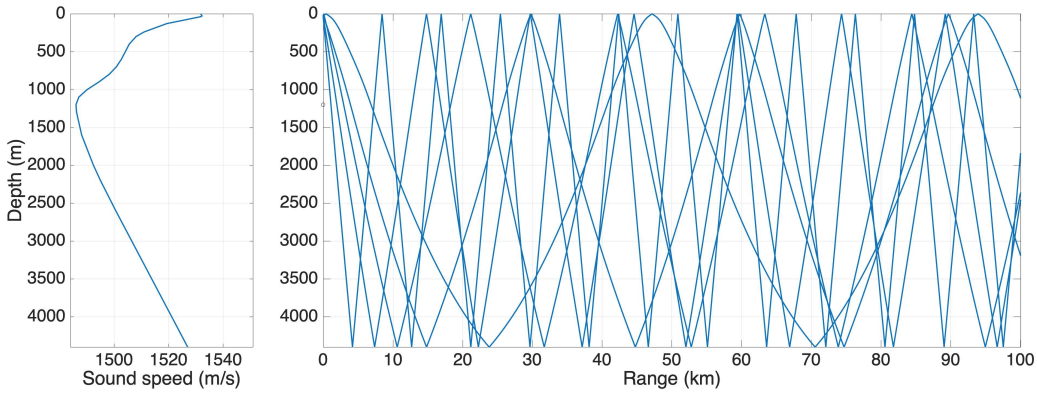

Figure 1.11: Propagation simulation, source depth $z_s = 30$ m, $\theta = [-20;45]°$, propagation range $r = 100$ km.

[5] http://oalib.hlsresearch.com/Rays/

1.4.3 The sonar equation

The conservation of the acoustic energy on the source-receiver path can be seen as a transmission process and, can easily be converted into an equation. In underwater acoustics, the sonar equation in its signal excess form is described as a summation of levels (in dB) such as

$$SL - TL - (NL - PG) \geq DT, \tag{1.3}$$

where

- SL denotes the acoustic source level;
- TL expresses the transmission losses, i.e., the effect of distance and attenuation on the original sound;
- NL is the ambient noise level;
- PG represents the processing gain resulting from, e.g., the equipment gain, directionality, signal processing and;
- DT is the detection threshold.

In PAM, SL corresponds to the level emitted by the studied species. A brief description of the recent breakthrough in the understanding of sound production mechanisms of baleen whale and, estimated SL are presented in § 1.4.3.1 while TL are the subject of § 1.4.3.2. Oceanic ambient noise is the subject of its own chapter (Chapter 2) and, section 1.5 mitigates common signal processing methods for PAM of BWs (both intervene in PG).

1.4.3.1 Sound production

It is only recently that anatomical pieces of evidence were found in the investigation of baleen whales sound production: the whale possesses vocal "chords" - named vocal folds- and vocalizes only in close-circuits when they are underwater (Reidenberg and Laitman, 2007). Baleen whales are equipped with a laryngeal sac, in addition to the lungs, to generate the airflow necessary to excite the folds without air exhalation (Reidenberg and Laitman, 2008). This additional and compressible air container is attached to the folds and allows the whale to cycle air during a diving period. Sound production mechanisms have been modeled for humpback whales and can physically explain the production of tonal and pulsed sounds (Adam et al., 2013; Cazau et al., 2013, 2016). Equivalent work has not yet been published for BWs. It has been shown in Adam et al. (2018) how theses anatomical traits determine the characteristic features of BW sounds such as high acoustic intensity, limited duration, and very low fundamental frequencies. These observations are supported by measurements of different BW subspecies calls source levels, in their specific frequency band such as:

- SL = 189 ± 3 dB ref. 1 µPa @ 1m in the $[25 - 29]$ Hz frequency band for BW (Širović et al., 2007),

- SL = 179 ± 2 dB ref. 1 µPa @ 1m in the [22 – 25] Hz frequency band for Australian pygmy BW (Gavrilov et al., 2011) ,
- SL = 179 ± 5 dB ref. 1 µPa @ 1m in the [17 – 30] Hz frequency band for ABW (Samaran et al., 2010c) and,
- SL = 174 ± 1 dB ref. 1 µPa @ 1m in the [17 – 50] Hz frequency band, Madagascar pygmy blue whale (MPBW) (Samaran et al., 2010c).

These measurements are in the same order as the values estimated initially by Cummings and Thompson (1971). Vocal characteristics of BWs such as the production of low-frequency sounds with strong levels indicate once again the outstanding abilities of these giant species for adapting their communication to long-acoustic ranging in the deep and open ocean (Payne and Webb, 1971).

1.4.3.2 Long range propagation

For long-range underwater acoustics, the solution to the wave equation is rarely considered under Cartesian coordinates (plane wave). Because a generated sound can propagate in all directions, it is often modeled as a spherical wave in the close field. However, when the spreading continues, the wave starts to interact and is constrained with the ocean boundaries (surface and floor) leading to cylindrical waves. Table 1.7 shows the expression of the spatial evolution of the pressure $p(r,t) = p(r)e^{j\omega t}$ (with $\omega = 2\pi f$) for outgoing plane, spherical and cylindrical waves (Urick, 1983).

The losses associated to the geometrical spreading of the acoustic wave can hence be expressed as a logarithmic function of the propagation distance and are presented for their associated wave type in the second column of Table 1.7. The plane wave does not suffer from geometrical spreading. However, losses from spherical spreading are doubled compared to cylindrical spreading (6 versus 3dB when doubling the distance).

	p(r)		**Geometrical losses** (dB)
Plane wave	$\mathrm{A}e^{-jkr}$	(1.4)	None
Spherical wave	$\frac{\mathrm{A}}{r}e^{-jkr}$	(1.5)	$20\log_{10} r$
Cylindrical wave	$\frac{\mathrm{A}}{\sqrt{r}}e^{-jkr}$	(1.6)	$10\log_{10} r$

Table 1.7: Evolution of the acoustic pressure for plane, spherical and cylindrical wave and their associated geometrical spreading losses.

The transmission losses TL are due to the coupled effect of the aforementioned geometrical losses and absorption. In seawater, absorption is due to the combined effect of viscosity, heat

transfers and, molecular relaxation of both magnesium sulfate and boric acid. It is taken into account using a complex wavenumber $k = k_r - jk_i$ in the expression of $p(r,t)$, introducing an exponential decay in the amplitude such as $A_{abs} = Ae^{-k_i r}$ where k_i is the absorption coefficient in Neper/m. Yet, absorption is highly frequency-dependent and does not critically affect low frequencies (less than 0.01 dB/km under a 100 Hz) (Urick, 1983). It is therefore not taken into account in the TL of low-frequency BW sounds that are hence uniquely dependent on the geometrical losses.

Thanks to the loudness of their calls and absence of absorption, BW calls propagate over very long distances. This property also implies that a sensor can record signals of interest as well as various noise sources originating from a large surrounding radius. It is then necessary to sort the content of the recordings. Besides, the size of sound archives related to baleen whales monitoring is rapidly increasing. Efficient analysis of these large volumes of data resulting from such continuous and long-term recordings requires reliable automated detection algorithms. The success of PAM hence relies on well-designed and robust signal processing methods to detect and classify signals of interest in the variety of sounds present in the oceans. A review of the commonly used signal processing methods for the detection of BW sounds is given in Section 1.5.

1.5 Signal processing

Analyzing a PAM experiment, requires to transform raw acoustic data into legible pieces of information. This crucial phase relies on signal processing and can take various forms. Initial manual detection consisting in reviewing multiples hours of recordings is still an essential and non-negligible task. It can help documenting the different encountered sounds, for example, to assess the soundscape diversity (Cerchio et al., 2015; Brodie and Dunn, 2015). However, manual annotation can be tedious, time-consuming and repetitive. Hence, annotation quality is variable and depends on the fatigue of the operator (Leroy et al., 2017b). To deal with the increasing amount of data from continuous and long-term recordings, *robust automatic detection-classification* methods are required. This expression conveys multiple underlying concepts that can briefly be described as (1) detection: the action of finding signals of potential interest, (2) classification: the action of sorting signals of interest into different categories. Moreover, the adjectives can be explained as (3) robust: the ability to detect a signal in noise with a known error rate, (4) automatic: ability to process data with minimum user-set parameters and without supervision. These different elements are discussed in the following paragraphs.

1.5.1 Pattern recognition: detection, classification or both?

Pattern recognition is the global term to designate detection-classification systems and is a branch of artificial intelligence. Generally, pattern recognition systems rely on features extraction to generate a profile of the signal(s) of interest and, classify based on the similarity between

the measured features and those learned from multiple exemplars (Bittle and Duncan, 2013).

A fitting pattern recognition system relies on what makes a particular signal of interest (or call) distinct, from the noise and other transient look-alike signals. The different developed methods, often exploit BW subspecies specificity and the stereotyped features of their vocalizations (§ 2.3). Specifically, they use the characteristics of their short-band frequency modulated signals (found in many bioacoustic signals).

	Pre-processing	Segmentation	Features extraction	Classification
One-class	Pre-filtering, SNR measurement	Signal template and filtering		Threshold
Multi-classes	Filtering, TF analysis	Extraction of signals of interest	Discriminative features	Sort

Table 1.8: One-class and multi-classes pattern recognition layout.

As presented by Table 1.8, various types of algorithms have been developed and used for PAM. Some, uses one-class pattern recognition (§ 1.5.2), others considered as "generalized" detection-classification systems are discussed in § 1.5.3 and § 1.5.4.

1.5.2 The specific case of one-class pattern recognition

One-class pattern recognition is a type of detection, restricted to the class of signals of interest (binary detection). As reported in the sonar equation (1.3), a sensor records a mix of ambient sound (hence considered as background noise) and, the signal of interest with an energy balance indicated by the SNR. The detection problem can then be viewed as a noise reduction problem where the solution consists in finding the optimal filter that maximizes the output SNR. Specifically, it can be shown that the MF is the optimal filter to detect a deterministic signal of interest embedded in white noise (Van Trees, 2002). The MF is referred to as an adapted filter because its impulse response is a reversed version of the signal of interest. In practice, this operation is performed in the time domain by the convolution between the recordings and a signal template (Samaran et al., 2008; Weirathmueller et al., 2017).

The work of Mellinger and Clark (1997, 2000) introduced the PAM community to spectrogram correlation (2D MFs). It has since been implemented in various automatic detection software [6] [7] and is still commonly used (Stafford et al., 2011; Samaran et al., 2013; Balcazar et al., 2015; Shabangu et al., 2017). In Gavrilov and McCauley (2013) a step is added to reduce false alarms

[6] http://ravensoundsoftware.com
[7] https://www.pamguard.org/

(the detection of a non-desired event).

Marginally, a simpler version, the energy detector consists in evaluating energy variations in short frequency bands, tailored to the useful signal. It conveys the same principle in terms of noise reduction but, might be more subject to false detections (Širović et al., 2004; Gedamke et al., 2007; Tripovich et al., 2015).

Other specific detection methods such as MFs, achieve both the detection and classification (one class of signal of interest) of a specific signal or call type simultaneously. For example, the *Z-detector* is a method recently introduced based on the subspace-detection of sigmoidal-frequency signals with unknown time-varying amplitude (Socheleau et al., 2015). This detector has been developed to be able to take into account the main issues in MF methods: frequency variations of BW calls and the presence of other unwanted transient sounds.

In cases where the goal is to detect different types of signals, for example, call types from different species wandering in a common area, running a specific binary detection algorithm can become a bit tedious. This is why other categories of pattern recognition algorithms consider simultaneously multiple classes. The first step is the detection of the potential signals of interests or, *segmentation*.

1.5.3 Segmentation

In order to exploit frequency-modulated specificities of whales calls, signals are typically described based on the time-varying shape of the vocalization in the time-frequency (TF) domain. The first step of most segmentation method is, therefore, often to represent the observation as a spectrogram. Segmentation can be performed from arbitrary samples (i.e., spectrogram temporal bins) or, directly from signals of interest "box" information (minimum and maximum frequencies, begin and end times). However, the first one does not provide temporal coherence and, the second one requires mandatory pre-processing by an operator. Automatic methods are thus preferred.

Once again, various detection techniques can be used to segment BW signals from ambient noise such as the detection of the connected elements (Harland and Armstrong, 2004), pitch-tracking (Baumgartner and Mussoline, 2011), edge detection (Gillespie, 2004), clutter map constant false alarm rate detector (also named energy detector) (Mouy et al., 2008; Thode et al., 2012), spectrogram-based detector (Urazghildiiev et al., 2009), contour detection and Bayesian filtering (Roch et al., 2011; Madhusudhana et al., 2016) or local-max detector (Lin et al., 2013), transient signal detector (Gavrilov et al., 2011).

Following data segmentation comes the characterization phase that aims at extracting specific and discriminative features, to classify the signals. Classical features often express spectro-

temporal variations of the signal such as peak frequency, duration, slope, bandwidth (Gillespie, 2004; Urazghildiiev et al., 2009; Gavrilov et al., 2011; Trygonis et al., 2013). These features can be amplitude-weighted (Baumgartner and Mussoline, 2011) or SNR-weighted (Thode et al., 2012). Some other types of features are based on cepstral coefficients or wavelets (Mouy et al., 2008).

1.5.4 Classification

The classification is the last step of pattern recognition processes and is based on the extracted feature vectors. The outcome of the classification can be binary (seek to determine if a sample is the signal dominant or not) or multi-class. Supervised methods are the most common in PAM as they tend to replace human analysis and are designed for a specific task. Supervised classification maps the attributes into call classes after learning from labeled data. Dimension reduction can be employed, such as principal component analysis (PCA). Classification methods include the use of discriminant analysis (Binder and Hines, 2012), multivariate discriminant analysis (Gillespie, 2004), neural networks (Dugan et al., 2010; Thode et al., 2012; Halkias et al., 2013), Gaussian mixed models (Mouy et al., 2008), Hidden Markov Models (Pace et al., 2012) and, Classification trees (Trygonis et al., 2013).

More recently, a global detection-classification approach based on sparse call representation (directly from the waveform) and dictionary learning has been proposed (Socheleau and Samaran, 2017; Guilment et al., 2018). Unlike MFs, this novel approach has the advantage of accounting for call variability in its sparse models. Besides, these compact representation types enable the building of large dimension dictionaries for better call representation and, dealing with less-stereotyped signals such as BWs D-calls (Guilment et al., 2018).

1.5.5 Discussion

As discussed in the introduction to section 1.5, acoustic detection, and classification can be achieved "manually" by an expert with the appropriate training. Still, automated methods have unequivocal advantages: they can process data faster, more efficiently and, are unbiased (or rather their bias is constant (Mellinger and Clark, 1997)).

On the one hand, because it is specific, easy to implement and has low computational complexity, binary detection, and especially MF (§ 1.5.2) is by far the most common strategy for the detection of BW calls. Besides, it does not require much training (only high-quality signals or models of the signals of interest) and, can reliably detect medium to high SNRs signals (Bouffaut et al., 2018). Nonetheless, the simultaneous or quasi-simultaneous detection of different types of signals (e.g., multiple-species call detection) tends to increase the number of filters (or call templates). The filtering must be performed for all filters, processing data in parallel or, independently. Expanding the number of signals of interest increases the computational complexity of the method: MFs are therefore not well adapted to the detection of multiple call types.

On the other hand, generalized pattern recognition systems (§ 1.5.3-1.5.4) are more adaptable. They can detect and classify different types of signals at once. Besides, even if some *a priori* knowledge on the call types to detect might be helpful to find relevant features to extract, perfect knowledge of useful signals is not mandatory. Furthermore, these types of algorithms can offer more flexibility, for example, to sort a new incoming signal into a rejection class. The quality and performances of the overall method are variable. It depends on the discriminative qualities of the extracted features as well as the quality of the training data. Ideally, in supervised learning, training data should cover all signals of interest in addition to various samples of noise, to design efficient rules of separation. However, only a few of these annotated datasets are freely available for baleen whale sounds, such as the detection, classification, localization and density estimation of marine mammals (DCLDE) low-frequency dataset [8] (where annotated D-calls and fin whale 40-Hz calls are available).

Automatic pattern recognition algorithms can be proved to be affected by long-distance acoustic propagation effects, resulting in the distortion of the received calls, echoes, low SNRs or, various environmental transient noises (Binder and Hines, 2019). It is therefore essential to assess the performances of such detectors, e.g., evaluate the detection and false alarm probability, the miss-detection rate or the quality of the retrieved calls, in order to assess the limitations and range of applications for a given method.

1.5.6 Performance analysis

The evaluation of the performances aims at quantifying in what terms and to what extent a method is reliable in different conditions, e.g., various SNRs, transient sounds or, changing ambient noises. In other words, performance scores are indicators of the confidence one can have in a method, based on a statistical analysis of its results.

There are multiple ways for demonstrating the performances of a method. Formal (analytical) and controlled performance serve as reference in the field of signal processing. However, they are not always available for more practical or data-driven methods. Because any evaluation requires large testing samples to ensure statistical robustness, the solution is often to rely on the use of ad-hoc procedures and extensive Monte Carlo simulations. These types of simulations are often implemented using noise and synthetic signals. They have the advantage of complete control over the SNR. Difficulties are reached when performance assessment requires datasets that are representative of oceanic soundscapes and signals that reflect propagation distortion. The last possibility, to show the effectiveness of algorithms on real recorded signals, is to resort to human inspected data. Thus, a large panel of soundscapes and signals can be covered. However, annotations for performances and testing suffer from the same issues than manual detection and classification: they are subject to the operator instant sensitivity and therefore biased, especially at low SNRs (Leroy et al., 2017b).

[8] http://www.cetus.ucsd.edu/dclde/datasetDocumentation.html

Performance scores can be used to evaluate the detection and classification abilities of a method under a particular setup (e.g., for MFs, if I set the threshold to X value, I can not detect calls with a SNR lower than Y). On the contrary, they can be used to constrain the method to a specific range of applications (if I want a probability of false alarm or probability of detection of X, I need to fix the detection threshold to Y). In cases where the classification is based on extracted features, classification performances can be used to find the best set of discriminative features.

As a statistical evaluation of the outcomes of the method, performance scores are an unbiased way of comparing methods and can be used to make a knowledgeable choice regarding the desired application. Note that it is crucial, for the comparison to be unequivocal, to be "fair" to all methods and compare them under similar setups.

Performances evaluation is essential for most of PAM applications. For example, estimating the detection probability within a specific detection range is essential for call density estimation or, evaluating the precision of the detected frequency is required for precise call description and variation analysis.

1.6 Conclusion

In this chapter, PAM was presented as an autonomous, discrete, low-cost, and, efficient multidisciplinary mean to conduct BW surveys across oceans. The ambient sound, recorded by these autonomous systems depends on multiple environmental factors such as the deployment region, the season, the moment of the day, the weather, human activity, as well as equipment characteristics such as the recorded frequency band, self-noise or depth of the sensor.

The work presented in this thesis is based and illustrated on an opportunistic source of data acquired in the Western Indian Ocean during the RHUM-RUM seismological experiment. OBSs were deployed in an un-monitored part of the Indian Ocean whale sanctuary, providing exclusive seasonal information of local baleen whale activity (Dréo et al., 2019). Chapter 2 presents the multifariousness of the low-frequency soundscapes recorded by the OBSs.

Chapter 2

Low frequency sounds from the bottom of the Indian Ocean

The hypothesis that whale voices could be heard across an ocean was almost too grand to believe. Furthermore, the notion that noise from commercial shipping might be interfering with whale communication seemed far-fetched and was essentially forgotten.

Christopher W. Clark
Testimony for the hearing on "Examining the Threats to the North Atlantic Right Whale" (2019)

Contents

2.1 RHUM-RUM data

2.1.1 RHUM-RUM network

Data used in this work were collected by the RHUM-RUM (Réunion Hotspot and Upper Mantle - Réunions Unterer Mantel) seismic network [1] (Barruol and Sigloch, 2013). This project aimed at imaging the mantle structure beneath the Western Indian Ocean and the dynamics of La Réunion volcanic hotspot. To that extent, 57 OBSs were deployed on the ocean floor, covering an area of 2000 × 2000 km^2 (Lat. 16 – 34°S, Long. 048 – 070°E; Figure 2.1), from October 2012 to November 2013. Each one of these autonomous OBSs was equipped by a three-component seismometer and a hydrophone, recording data continuously [2]. Deployed instruments were from different origins; therefore, sensor types and sampling frequencies were not homogeneous among the 57 OBSs. Paragraph 2.1.2 introduces the characteristics of the OBSs of interest. For more information, the technical description of the network with deployment positions, recording times, sensor types, stations' performances, data quality, and instrumental failures is detailed in RHUM-RUM technical report (Stähler et al., 2016).

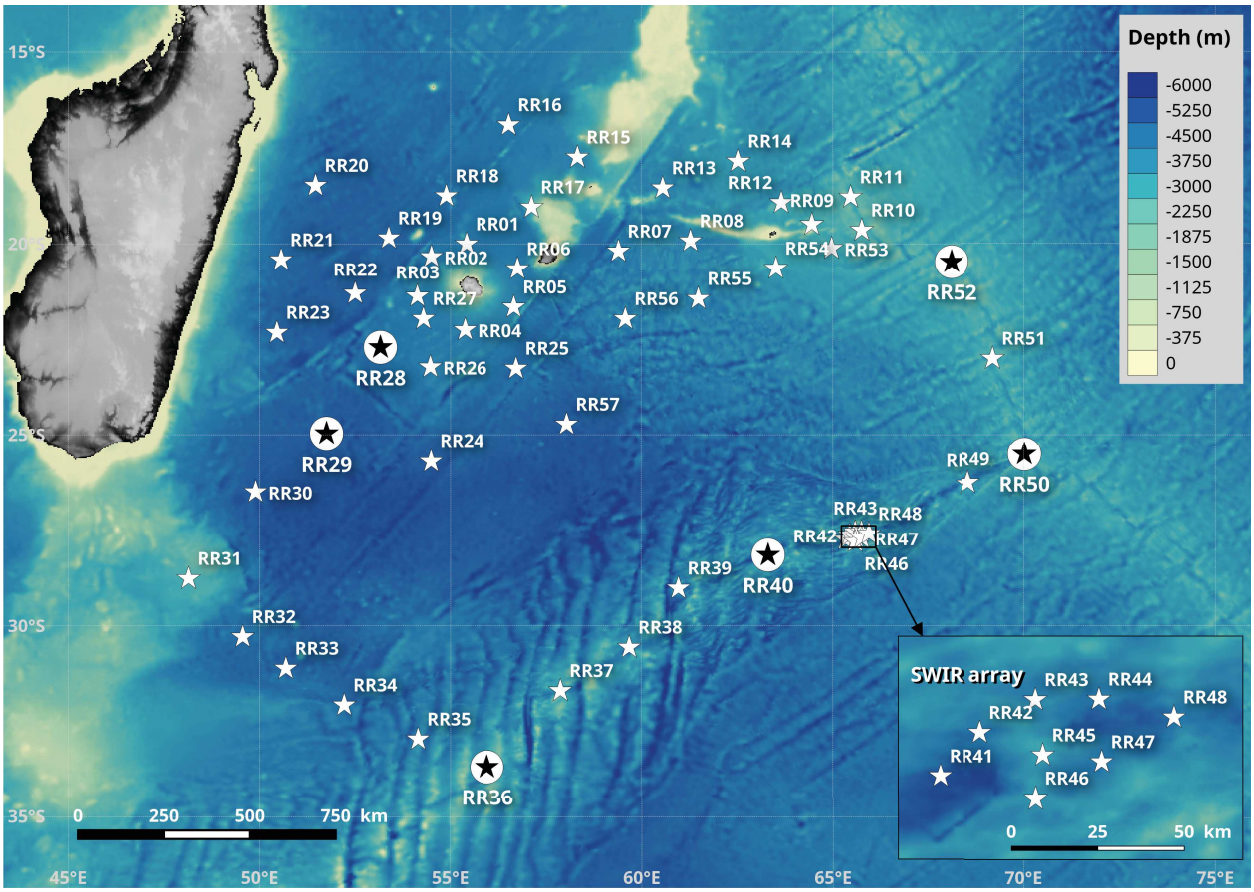

Figure 2.1: RHUM-RUM seismic network of 57 Ocean Bottom Seismometers (OBSs) in the Western Indian Ocean and the denser Southwest Indian Ridge (SWIR) sub-array, deployed on the sea-floor (Dréo et al., 2019).

[1] www.rhum-rum.net

[2] Recorded data are currently hosted on the RESIF French national seismic archive center (Barruol et al., 2017).

The OBSs of the RHUM-RUM project were fully autonomous from deployment to recovery and, despite the wide area covered by the experiment, instruments set up (and then recovery) took less than a month. However, OBS deployment technique raises some technical issues.

- The instruments, deployed from a surface vessel, fall freely into the water. Due to ocean currents and, depending on the water column height, they can drift up to hundreds of meters relative to the surface deployment location: landing positions of the OBSs are not known and, are considered ±500 m from the drop-off location.
- Also, the available amount of battery is limited and, sensor failure might occur before recovery. In addition to lacking part of the data, early failure prevents the synchronization of the internal clocks with the GPS signal that normally occurs immediately after the station recovery.

For the present dataset, clock drift was post-synchronized using the results of multi-component noise cross-correlation presented by Hable et al. (2018). Uncertainties on the OBSs locations are in the process of being solved using ship-generated noise as localized (AIS) source of opportunity (Trabattoni et al., 2020).

2.1.2 SWIR array

In order to characterize the micro-seismicity associated to an active seamount located on the Southwest Indian Ridge (SWIR), the extensive tectonic plate boundary between Africa and Antarctica, RHUM-RUM deployed a local and denser sub-array of 8 OBSs, henceforth denoted as the SWIR array. These instruments belong to the German OBS pool DEPAS (Deutscher Geräte-Pool für Amphibische Seismologie) and are of the LOBSTER (Long-term OBS for Tsunami and Earthquake Research) type (Figure 2.2). The SWIR array covered an area of 70 km×40 km with depth varying from 2822 m at the top of the seamount to 5430 m in the trench. Inter- OBS distances were in the order of 20 km. Their sampling frequency f_s was 100 Hz (except for OBS RR42, f_s was 50 Hz). The hydrophone frequency response for the OBSs is known to be mostly flat from 1 Hz to 90% of the Nyquist frequency (45 Hz) (Stähler et al., 2016). Position, sampling frequency, sensitivity, and sensor recording period are summarized in Table 2.1.

The location and frequency band covered by the RHUM-RUM recordings provided unique and unpredicted opportunities to observe and study the numerous low-frequency sound sources of the Western Indian Ocean. Besides, the SWIR array positioning and inter-sensor distances were ideal for multi-sensor observations, especially of whale sounds. The array arrangement enables the possibility of localization and tracking and, therefore, conduct PAM of baleen whales in the area.

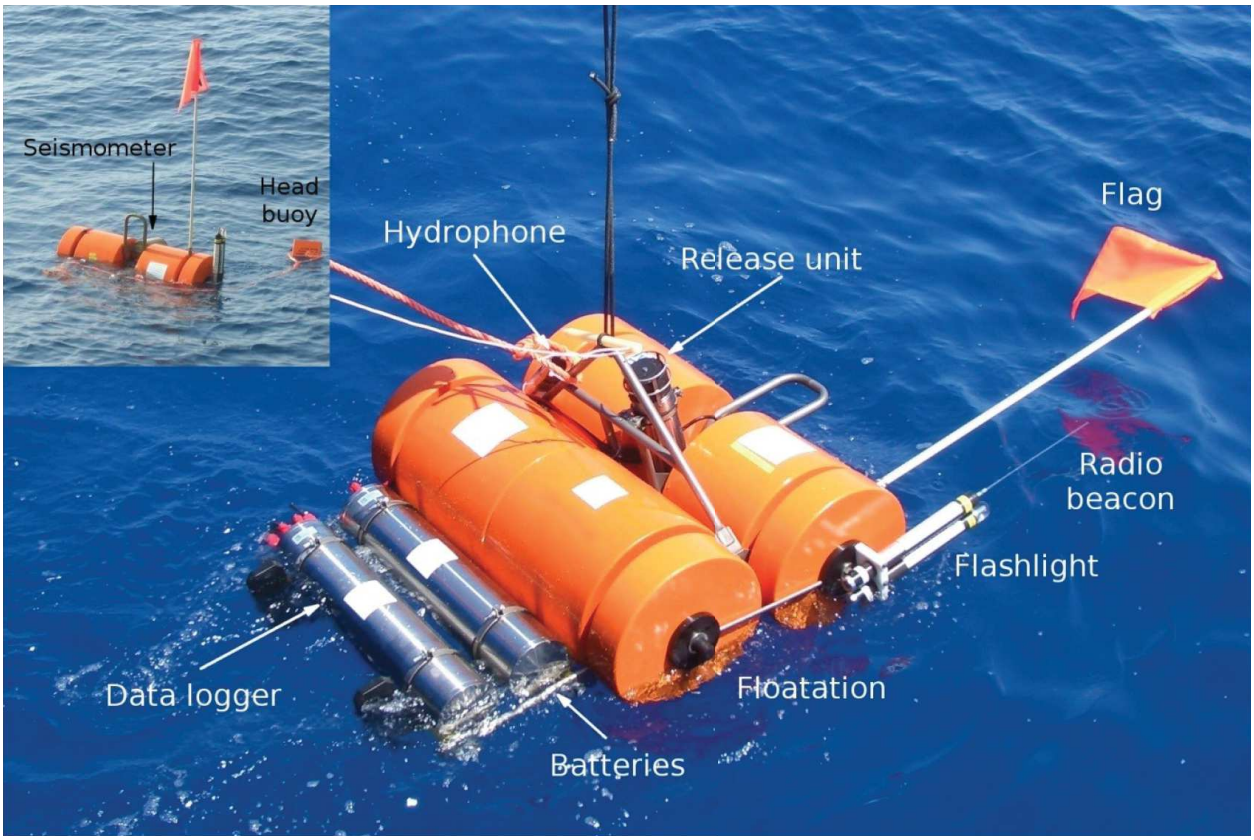

Figure 2.2: Picture of a LOBSTER (Schmid and Schlindwein, 2016).

OBS	Lat. (°)	Long. (°)	z (m)	f_s (Hz)	sensitivity (mV/Pa)	Recording period
RR41	-27.7330	65.3344	5430	100	1.2	16 Oct. 2012 - 17 Jun. 2013
RR42	-27.6192	65.4376	4776	50	1.1	16 Oct. 2012 - 10 Aug. 2013
RR43	-27.5338	65.5826	4264	100	1.1	16 Oct. 2012 - 15 Jun. 2013
RR44	-27.5324	65.7480	4548	100	1.2	16 Oct. 2012 - 03 Jun. 2013
RR45	-27.6581	65.6019	2822	100	1.2	16 Oct. 2012 - 04 Jun. 2013
RR46	-27.7909	65.5835	3640	100	1.1	16 Oct. 2012 - 26 May 2013
RR47	-27.6958	65.7553	4582	100	1.0	16 Oct. 2012 - 22 Jun. 2013
RR48	-27.5792	65.9430	4830	100	1.1	16 Oct .2012 - 10 Jun. 2013

Table 2.1: SWIR array OBSs characteristics: position (Latitude (Lat.), Longitude (Long.), Depth (z)), sampling frequency f_s, sensitivity and recording period (Stähler et al., 2016).

2.2 Low-frequency soundscapes from the bottom of the Indian Ocean

2.2.1 Introduction

Poetically, the concept of soundscape was introduced in 1977 by the composer and environmentalist Murray Schafer, in his book *The tuning of the world* (Schafer, 1980), setting the basis

of acoustic ecology. To this day, his definition of three classes of sound to sort the composing elements of ambient sound are still widely used (Krause, 2008; Pijanowski et al., 2011). They are illustrated in Figure 2.3 and their definition is given as follows:

- the biophony includes all non-human biological sound sources, e.g., baleen whale songs, dolphins whistles, fish sounds, seals thrills or, smaller-scale coral reefs and benthic sounds;
- the geophony regroups the sound sources generated by non-biological natural sources, e.g., meteorological noises, earthquakes and, iceberg tremors; and,
- the anthrophony comprises all sounds generated by humans such as ship or harbor wideband noise, airguns explosions, or drilling noise.

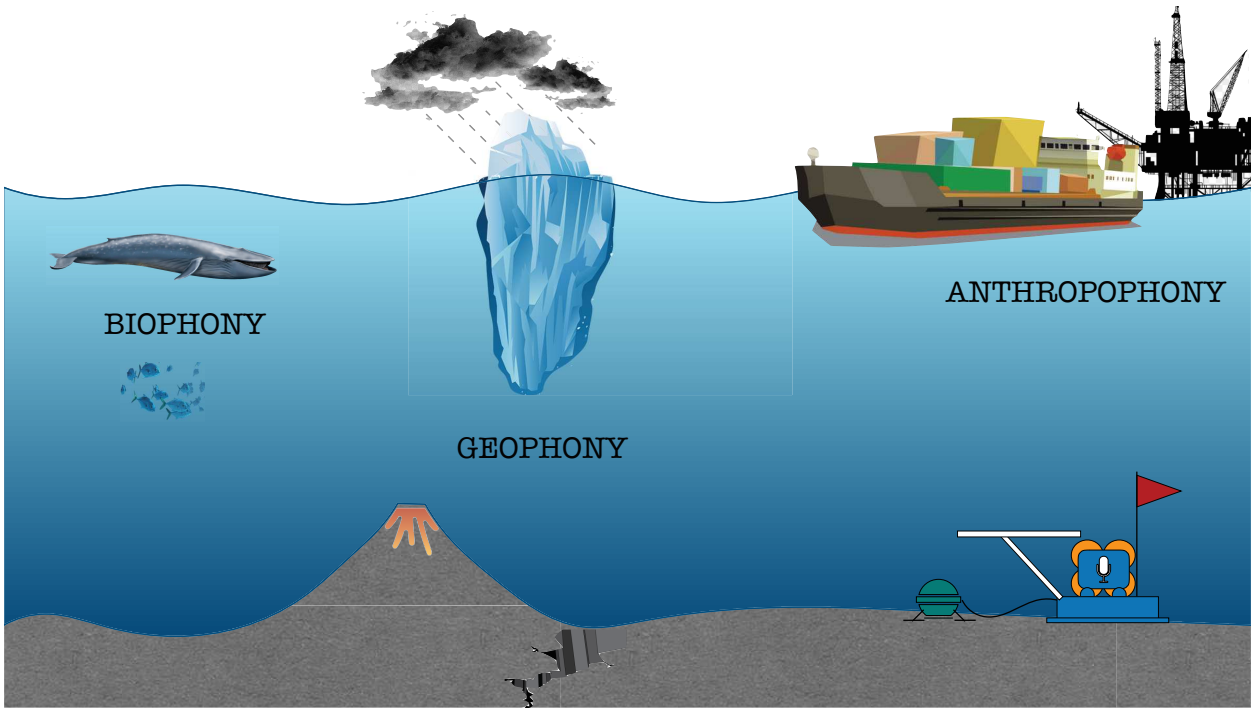

Figure 2.3: Illustration of underwater soundscape categories.

Studying ocean's soundscapes is equivalent to monitoring spectral characteristics (levels and frequency) over time and space in order to point out the different trends in sounds in different frequency bands. The exciting thing is that each one of these specific sounds interests a particular scientific community, that considers all the other sounds as ambient noise.

Empirically tested for decades, Wenz curves (Figure 2.4) specifically take into account the global contribution of each soundscape category and, give an estimate of frequency distribution as well as energy contribution to the oceanic ambient noise level (Wenz, 1962).

Contribution of each soundscape category can be found in the frequency band recorded by the OBSs $[0-50]$ Hz. Notably, because low-frequency sources experience little attenuation, they have a significant potential for long-range propagation. "The low-frequency ambient noise field, therefore, can be a summation of noise across an entire ocean basin" (Hildebrand, 2009).

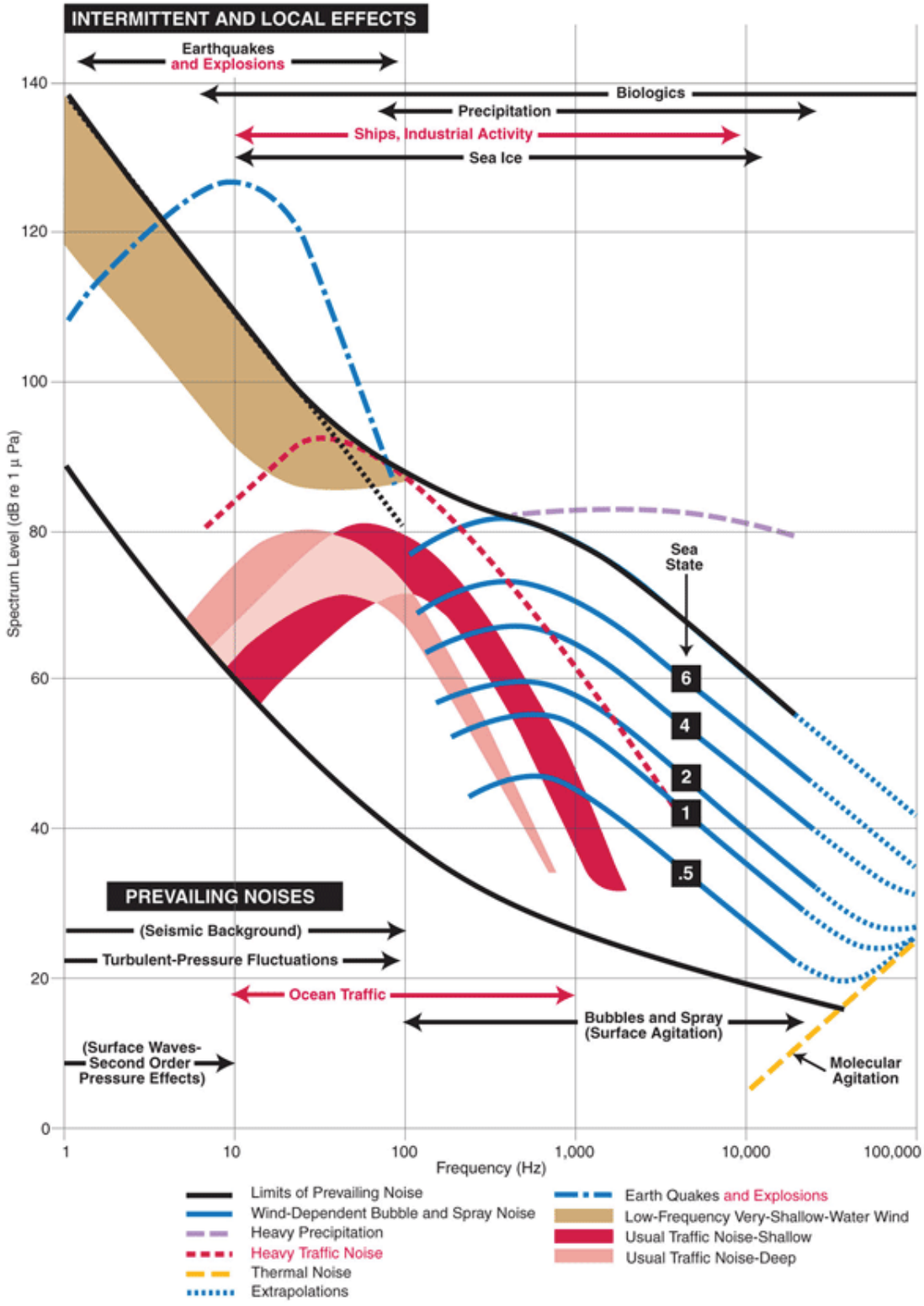

Figure 2.4: Wenz curves: spectra and frequency distribution of underwater sound sources (Miksis-Olds et al., 2013), (reproduction from Wenz (1962)).

2.2.2 Anthrophony

The principal source of anthropogenic noise in the area covered by the SWIR array is ship noise. It is identified as one of the major contributors to the increase of low-frequency sound floor

(quietest ocean conditions) in the Indian Ocean over the 2002-2012 decade, mainly due to an increase of traffic (Miksis-Olds et al., 2013). As an example, Figure 2.5, is an instant capture of the maritime traffic in the Indian Ocean on May 24^{th}, 2019. The SWIR array is located south from the principal lane between the Cape of Good Hope and India/Thailand but is on the route connecting South Africa to Indonesia. The selected tanker is within the vicinity of the area previously covered by the SWIR array. Most of the vessels are cargo or tankers.

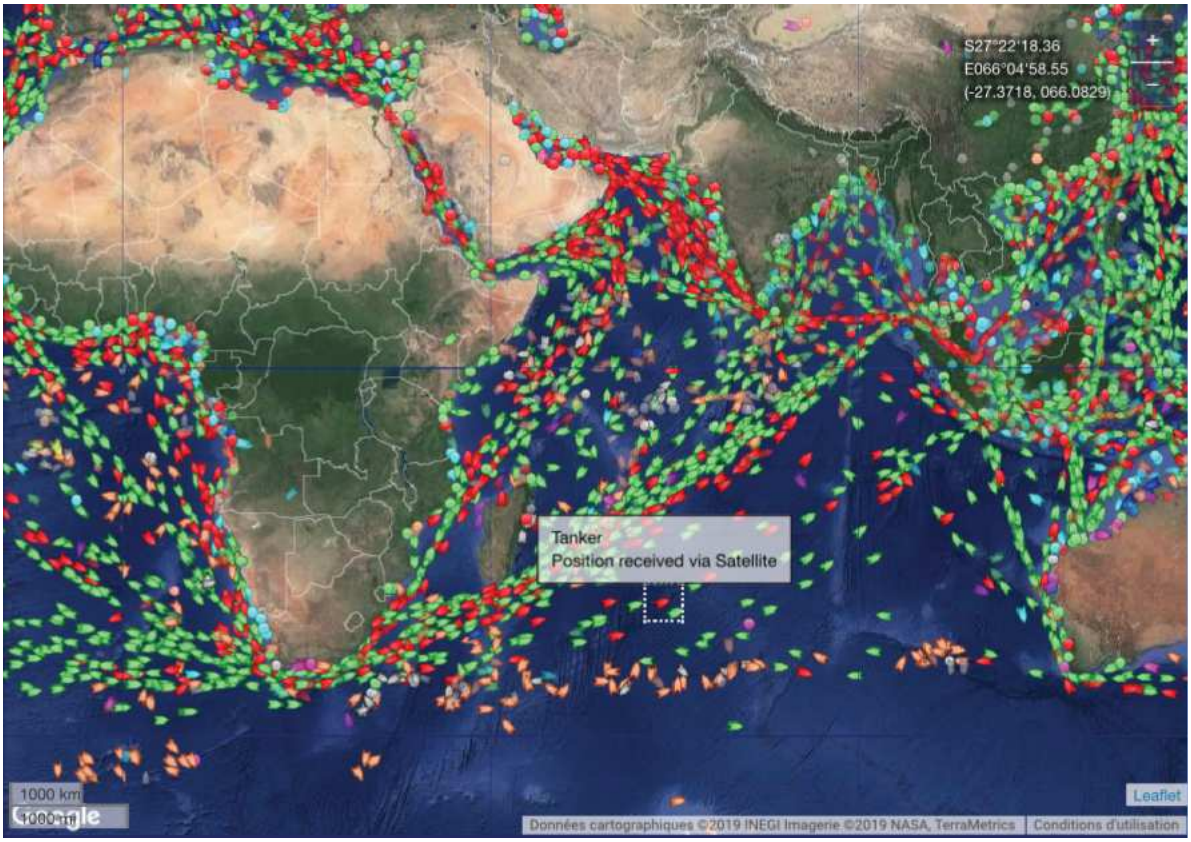

Figure 2.5: Instant capture of Indian Ocean statellite ship traffic on May 24^{th}, 2019 whith a tanker passing on the former SWIR array location. Green = cargo, red = tanker, orange = fishing, blue = tugs and special craft, dots = pleasure crafts (source: `http://www.shiptraffic.net/2001/04/indian-ocean-ship-traffic.html`).

Ship noise, in general, is described as the combination of two physical processes (Ross, 1976) (Figure 2.6): a broadband hydrodynamic noise ($[0-10]$ kHz) generated by cavitation, modulated by low-frequency propulsion noise. Propulsion noise is directly related to the propeller mechanism, frequencies are proportional to the number of blades of the propeller, to the rotation speed of the shaft(s) and to the engine rate. Propulsion-related sources are dominant in the ship's radiated noise at high speeds (Arveson and Vendittis, 2000). This phenomenon is characterized by low-frequency spectral lines and their harmonics < 500 Hz.

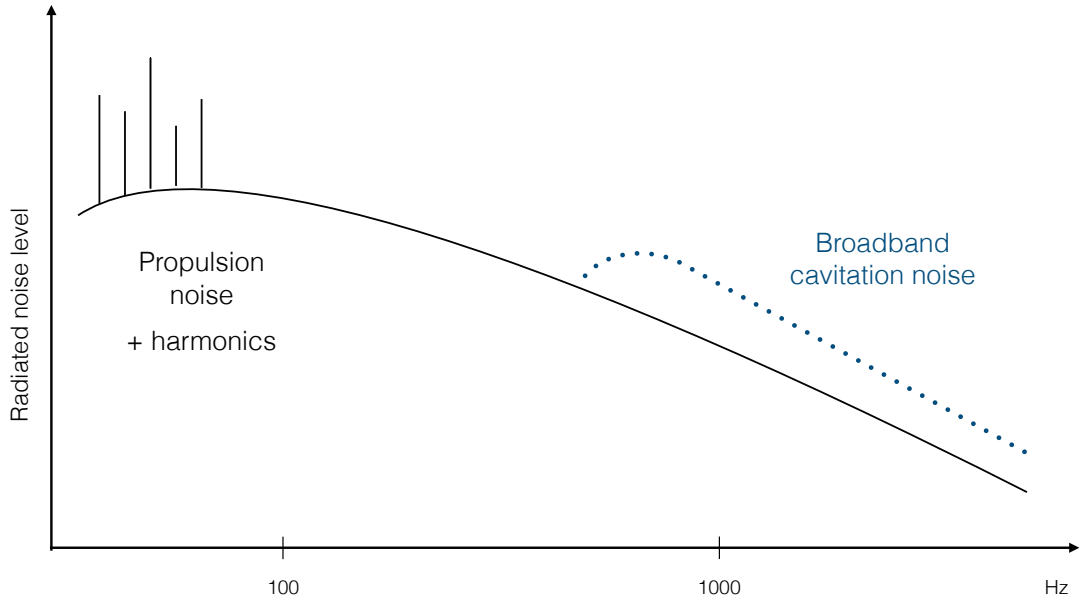

Figure 2.6: Large ship radiated noise spectrum diagram.

The low-frequency band covered by the OBSs provides direct observations of ship radiated noise as spectral lines that can be recorded continuously for multiple hours. As an example Figure 2.7 shows the spectrogram of a ship passing nearby the SWIR array and, recorded for more than ≃ 10 hours (equivalent to a pursuit of more than 300 km at 16 knots).

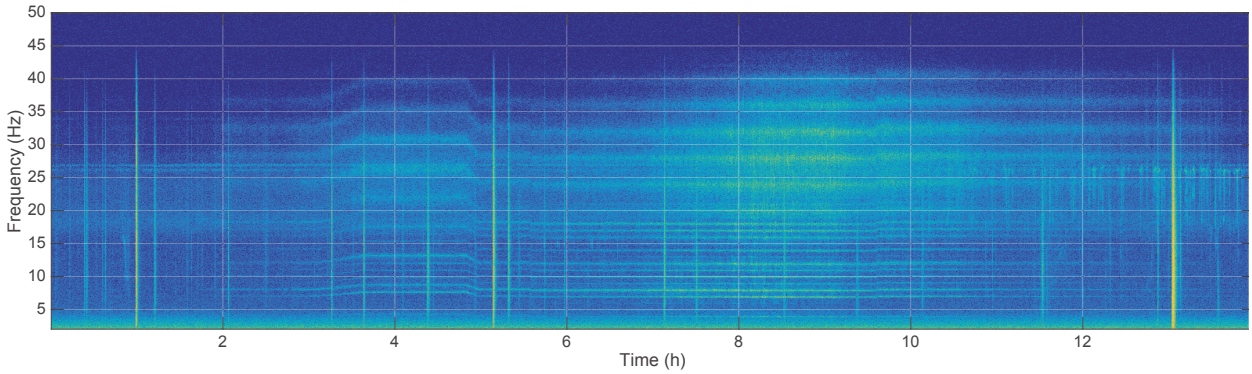

Figure 2.7: Example of the spectrogram of the noise radiated from ship, recorded by an OBS.

Other anthropogenic sources such as airguns for seismic exploration are usually in the 5 to 300 Hz frequency range (Hildebrand, 2009); however, none were recorded during the deployment period.

2.2.3 Geophony

Geophony is by far the soundscape category that regroups the most eclectic range of sources. As a rough approach to the types of sounds that might be recorded by OBSs, they are sorted into two categories: below 2 Hz and broadband [2 -50] Hz.

The ultra-low frequency band, below 2 Hz, is highly energetic in most recordings from the bottom of the ocean. It regroups signals issued from:

- geophysical sound sources such as micro-seismicity (principal micro-seism band [0.1 - 0.2] Hz (Sutton and Barstow, 1990; Schmid and Schlindwein, 2016)),
- tidal cycles that manifests themselves as a 2 Hz "whistle" (Duennebier et al., 1981),
- exceptional events, such as tsunamis, that can also be detected as ultra-low frequency chirps ([1-25] mHz) (Hanson and Bowman, 2005) or even,
- variation of level, temporally correlated with storms, hurricanes or wave height (Sutton and Barstow, 1990).

In the recordings, earthquakes appear as short duration (intermittent) highly energetic broadband noises, lasting from few seconds to ≃ 2 min as presented in Figure 2.8.

Other region-specific sound sources might occur: "In the Southern Hemisphere, the natural sounds associated with the breakup of icebergs represent an important and potentially underappreciated acoustic noise source. Annually tens of thousands of icebergs drift out from Antarctica into the open waters of the Southern Ocean, creating a ubiquitous natural source of sound as they disintegrate" (Matsumoto et al., 2014).

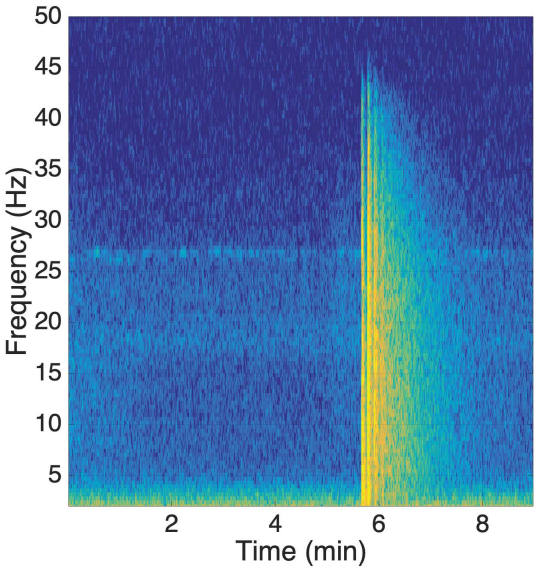

Figure 2.8: Earthquake and whale calls.

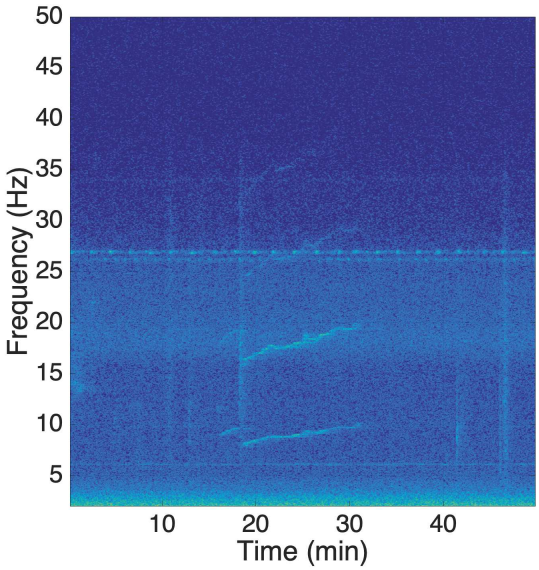

Figure 2.9: Ice tremor and whale calls.

Indeed, Antarctic icebergs generate two types of sounds: long harmonic duration tremors (Figure 2.9) and broadband bursts. Harmonic tremors spectral characteristics are variable. They are described as signals with prominent eigenfrequencies in the 4–7 Hz range, often accompanied by overtones (Talandier et al., 2006). They typically last 2–10 min. They are shown (Chapp et al., 2005) to widely cover the spectrum, up to 100 Hz and, their frequency fluctuate upwards

or downwards by 10–20 %. They are generated by the collision between icebergs (Matsumoto et al., 2014). Iceberg burst signals cover a broader spectrum. They are more common and are associated with iceberg breakup in the open sea (Talandier et al., 2006).

Finally, the full bandwidth can be impacted by geophysical sounds, e.g., the wind increases ambient noise level in the [30 – 800] Hz frequency band (Kewley et al., 1990; Cato and McCauley, 2002), especially in areas unaffected by anthropogenic sound sources (Haver et al., 2017).

2.2.4 Biophony

In the OBS-monitored frequency band, prevalent biological sound sources originate from baleen whales (Wilcock et al., 2014). Bioacoustical studies such as McDonald et al. (2006) reinforced the idea that whale song is an indicator of baleen whale (and especially BW) population structure worldwide. During the recording period, regionally-distinct songs of:

- Antarctic blue whales (ABWs) (*Balaenoptera musculus intermedia* (Leroy et al., 2016),
- Madagascar pygmy blue whales (MPBWs) (*Balaenoptera musculus brevicauda*) (Samaran et al., 2013),
- Fin whales (FWs) (*Balaenoptera physalus*) (Samaran et al., 2010a) and,
- an unknown caller that produces P-calls (Leroy et al., 2017a) (or "spot" call (Ward et al., 2017)),

were recorded across the RHUM-RUM deployment zone. Characteristics of these baleen whale song signatures are described in section 2.3.

2.3 Baleen whale acoustic signatures

Acoustic characteristics of regionally-distinct baleen whale songs recorded during the RHUM-RUM deployment are described in this paragraph for ABW (§ 2.3.1), MPBW (§ 2.3.2), P-calls (§ 2.3.3) and, Fin whale (FW) (§ 2.3.4). These descriptions follow the naming system of McDonald et al. (2006). Analysis results of this section are based on the work we presented in Dréo et al. (2019) characterizing individuals acoustic signature from hand-picked annotations. This study focuses on classical features such as, mean values of (Figure 2.10):

- units characteristic frequencies (Hz), usually the most energetic frequencies sometimes called *peak* frequency,
- units duration (s),
- inter-call interval (ICI) (s), the interval between the beginning of a call and the beginning of the following one and,

- inter-series interval (ISI) (s) the interval between two series,

among the observed panel. These attributes were measured on good quality signals, i.e., calls with high SNRs.

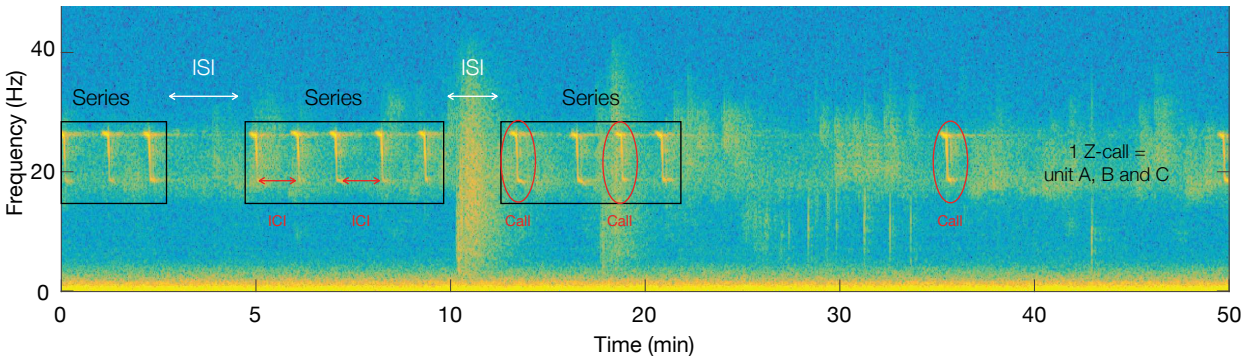

Figure 2.10: Illustration of units, call, ICI, series and ISI definitions on an ABW song.

In the following paragraphs, spectrograms and associated power signal densitys (PSDs) are displayed for each species, from the recording of a whale going through the SWIR array. In order to show the effects of propagation on whales stereotypical songs, a series is observed from two recording locations, close and remote from the whale. Knowledge of the whale location is used to comment on propagation attenuation of whale signals.

2.3.1 Antarctic blue whale

In the Southern Hemisphere, ABW call is the most widely distributed BW vocal signature. It is found near Antarctica (Širović et al., 2004; Thomisch, 2017) up to mid-latitudes of all oceans, including the Indian Ocean (Leroy et al., 2016; Balcazar et al., 2017).

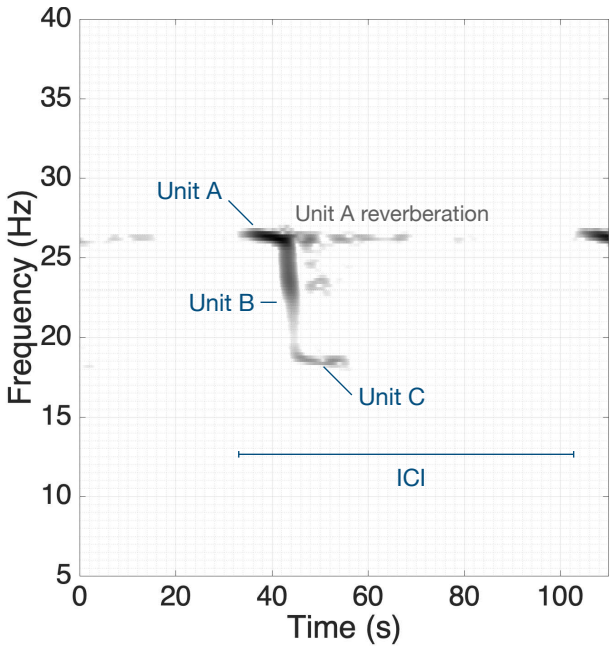

Figure 2.11: Annotated spectrogram of an ABW call recorded on May 31st, 2013 by OBS RR43. Peak frequency and duration of each units are: Unit A 26.2 Hz ($\simeq$ 12 s), unit B joins units A and C (2 s), unit C 18.7 Hz (12.2 s). The ICI is of $\simeq$ 66.4 s.

The ABW call, also commonly named *Z-call* because of its recognizable Z-shape in the time-frequency domain, is constituted of three short successive units (Figure 2.11). Unit A is a slightly modulated pure tone lasting 12 s with a maximum at 26.2 Hz. It is followed by unit B, a 2 s down-sweep (DS) that joins units A and C. Unit C is also a slightly modulated pure tone lasting 12.2 s with a maximum at 18.7 Hz. ABW calls are regularly repeated in series with ICI of 66.4 s. Series are separated by longer intervals of about ISI = 206.4 s corresponding to the breathing of the animal (Dréo et al., 2019).

Series of five ABW calls recorded on May 31st, 2013 at 12:33 by (a) OBS RR43 and (b) OBS RR48, respectively estimated ≃ 2 km and ≃ 36 km away from the location of the ABW are represented in Figure 2.12 (a)&(b). Figure 2.12 (c) represents the PSD of each observation.

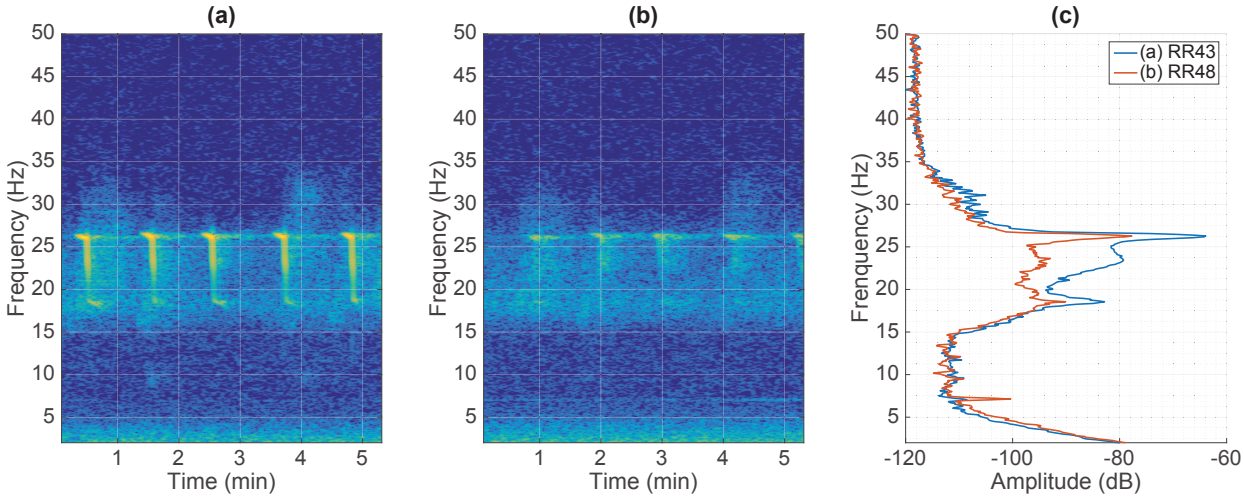

Figure 2.12: Propagation effect on ABW call series of 5 Z-calls recorded on May 31st, 2013 at 12:33 (D151). Spectrogram of the series received by OBSs (a) RR43 at ≃ 2 km and (b) RR48 at ≃ 36 km (nfft = 1024, overlap = 98%) and, comparison of the associated PSD.

On the closest observation (a), the three units appear distinctively, on each repetition of the call. Tonal units A and C affect the PSD (c) by showing peaks on their characteristic frequencies. Unit A peak is ≃ 20 dB higher than unit C. With the distance (b), unit A and unit B "bump" at ≃ 23 Hz, in the first half of the chirp, are attenuated with transmission losses consistent with the area and the distance. Recordings background noises show high energy between 15 and 32 Hz, a larger bandwidth than typical whale chorus (§ 2.3.5). This might be the effect of fin whale calls (§ 2.3.4) or higher frequency-modulated (90 to 25 Hz) "D-calls" which have been attributed to BW feeding activity (Samaran et al., 2010b). Because they exceed sensors frequency boundaries, "D-calls" are not characterized in this work. Probably because of this background noise, the low parts of the call, under 22.8 Hz (second part of unit B and unit C) are no longer visible in Figure 2.12 (b).

2.3.2 Madacascar pygmy blue whale

Out of the four known populations of pygmy BW, three inhabit the Indian Ocean. They are distinguished by their geographic range, morphological and vocal characteristics (Stafford et al., 2011; Gavrilov and McCauley, 2013). They form: the Northern Indian Ocean population (Sri Lanka pygmy BW call), the Eastern Indian Ocean population found west and south of Australia (Australia pygmy BW call) (Gavrilov and McCauley, 2013) and, the Madagascar population found in the Western Indian Ocean (MPBW call or song type 9A (McDonald et al., 2006)). Only calls from this last population are present in our recordings.

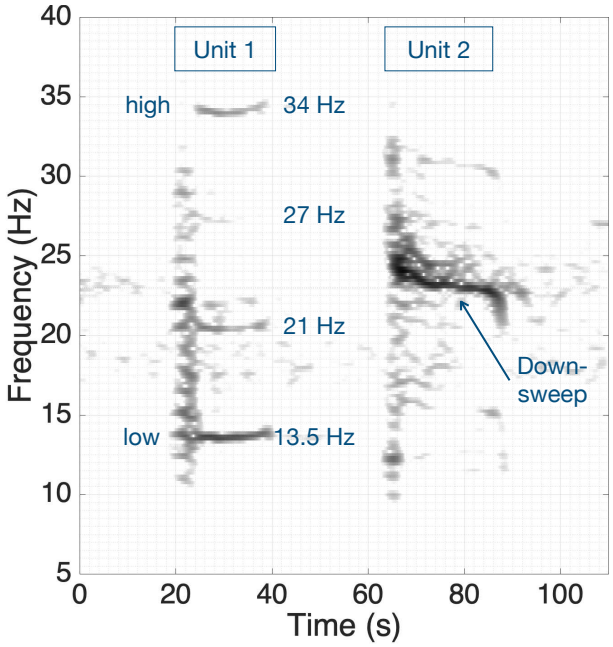

Figure 2.13: Annotated spectrogram of a MPBW call recorded on May 28^{th}, 2013 by OBS RR43. Unit 1 peak frequencies are ≃ 13.5 Hz, 21 Hz, 27 Hz and 34.0 Hz, it has a duration of 27.3 s. Unit 2 down-sweep peak frequency is 23.3 Hz with a duration of 24.4 s. The ICI is of ≃ 103.1 s.

MPBW calls consists of two successive units that are long complex tones (Figure 2.13). The first one, is composed of four partials with inharmonic frequencies spanning from a ≃ 13.5 Hz fundamental (named MPBW unit 1 low) up to 34.0 Hz (named MPBW unit 1 high) with 7 Hz intervals. The 27 Hz partial seems to often have less energy than the others. Unit 1 lasts 27.3 s. Ten to twenty s later follows the second complex tone, named unit 2, where most of the energy is concentrated on a DS that follows three successive laws of frequencies (or three rates of frequency changes) from ≃ 24.4 to 21.6 Hz with a maximum at 23.3 Hz. It is thereafter named MPBW unit 2 DS. Unit 2 lasts 24.4 s. Both units start by a short broadband signal constituted of multiple ≃ 1.4 Hz-spaced low energy partials (see Figure 2.13 at 20 and 65 s). Calls are repeated with ICIs of 103.1 s and series are separated by ISIs of 302.4 s (Dréo et al., 2019).

Series of seven MPBW calls, recorded on May 28^{th}, 2013 at 18:46 by (a) OBS RR47 and (b) OBS RR41, respectively estimated ≃ 2 km and ≃ 48 km away from the location of the MPBW are represented in Figure 2.14 (a)&(b). Figure 2.14 (c) represents the PSD for each observation. On the closest observation (a), all elements that compose the MPBW call appear distinctively, on each repetition of the call. Prominent frequency on the PSD (c) of the close recording are the MPBW unit 1 low (13.5 Hz) and high (34 Hz) partials as well as the MPBW unit 2 DS. A high

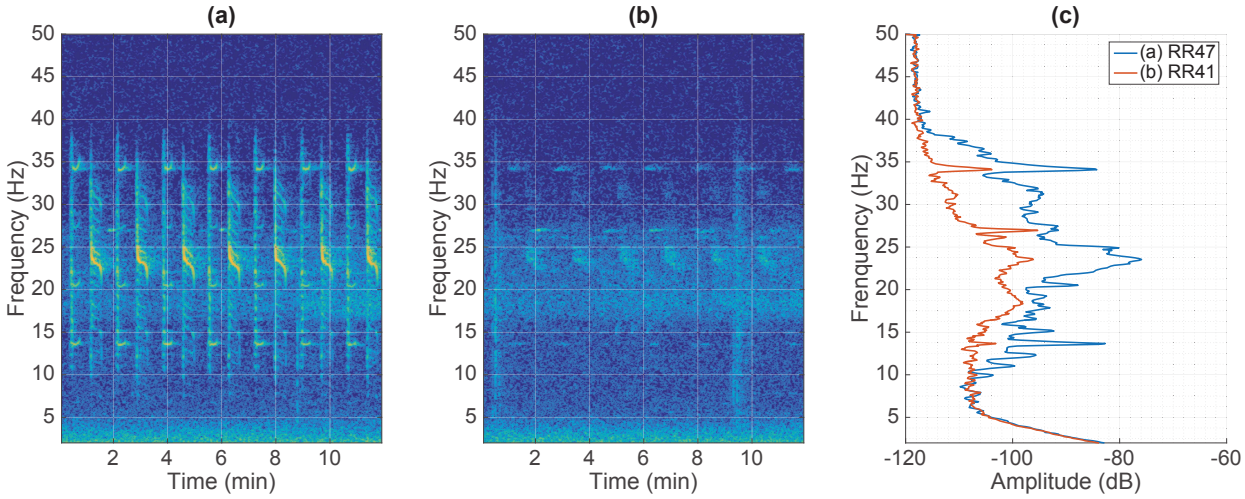

Figure 2.14: Propagation effect on MPBW call series of 7 calls recorded on May 28^{th}, 2013 at 18:46 (D148). Spectrogram of the series received by OBSs (a) RR47 at ≃ 2 km and (b) RR41 at ≃ 48 km (nfft = 1024, overlap = 98%) and, comparison of the associated PSD.

frequency peak at ≃ 27 Hz for RR41 indicates the presence of remote P-calls (§ 2.3.3) in the background noise of both observations: they are not correlated with the recurrence of units 1. Remote observation of the call series (b) highlights that only the prominent frequencies of the call (MPBW unit 1 high and low, unit 2 DS) remain while units short broadband signal and in-between partials with inharmonic frequencies are attenuated, confirmed with the PSD plot.

2.3.3 P-calls

The OBSs network recorded a type of call with similar features to a recently described vocalization called "P-calls" or "Spot-call" (Leroy et al., 2017b; Ward et al., 2017). So far, the source has not yet been identified, but many pieces of evidence suggested the call is produced by a giant whale but not by Antarctic nor pygmy BWs. These recent PAM studies revealed detection of this call type in deep-ocean waters across the Indian Ocean from 26°S to 42°S and from 058°E to 083°E (Leroy et al., 2017a) and in the Southern and Indian Oceans off Australia within 32°S to 38°S and 110°E to 141°E (Ward et al., 2017).

The P-call is a single pure tone at a frequency of 26.7 Hz, with a duration of 14.5 s. It is repeated in series with an ICI of 132.1 s. No recurrent longer intervals, indicating specific ISI were measured. Series of five P-calls, recorded on May 17^{th}, 2013 at 20:30 by (a) OBS RR48 and (b) OBS RR43, respectively at ≃ 20 km and 50 km away from the estimated location of the whale (outside of the SWIR array) are represented in Figure 2.15 (a)&(b). Figure 2.15 (c) represents the PSD for each observation. Figure 2.15 highlights that P-calls frequency seems to limit the upper frequency of whale chorus frequency band 2.3.5. High intensity calls in Figure 2.15 (a) generates strong echoes that are attenuated with the distance. Ship noise is responsible for the spectral lines below 15 Hz and at 30 Hz.

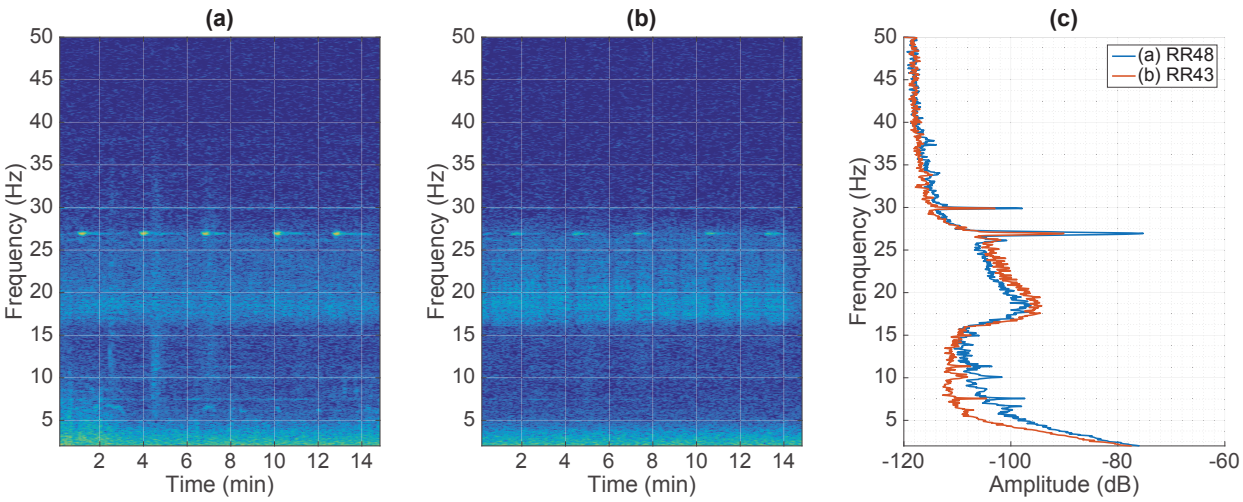

Figure 2.15: Propagation effect on series of 5 P-calls recorded on May 17^{th}, 2013 at 20:30 (D137). Spectrogram of the series received by OBSs (a) RR48 at ≃ 20 km and (b) RR43 at ≃ 50 km (nfft = 2048, overlap = 98%) and, comparison of the associated PSD.

2.3.4 Fin whales

FW habitat spreads across all oceans, however, the structure of their songs varies geographically (Delarue et al., 2009). Typically their songs are composed of sequences of *pulses* centered around 20 Hz (the "20 Hz pulse"), with highly stereotyped repetitions intervals. Geographical variations impact higher frequency components and ICIs (Gedamke, 2009).

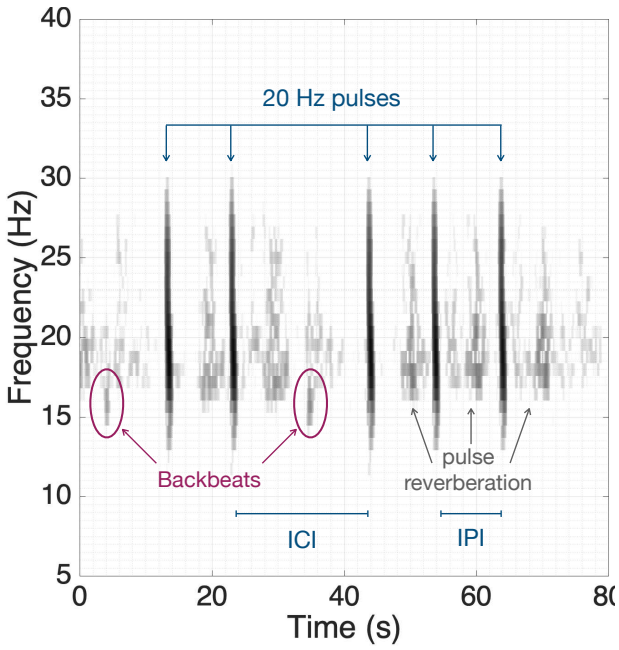

Figure 2.16: Annotated spectrogram of FW pulses and backbeats recorded on May 24^{th}, 2013 by OBS RR41. Pulses occur between 31.5 and 13 Hz and are shorter than a second. Backbeats are visible between 18 and 13 Hz. The interpulse interval (IPI) is of 9.9 s, the ICI of 20.4 s and the ISI of 107.6 s.

Two FW-characteristic units were recorded by the OBSs. The 20 Hz pulse is a short duration (< 1 s) down-sweep between 31.5 and 13 Hz. Most of its energy is concentrated around its maximum at 18.1 Hz, between 16.9 and 20.2 Hz (-3 dB peak width). In our data set, pulses are repeated in groups of 2 to 5, with interpulse intervals of 9.9 s. These groups are separated by ICIs

of 20.4 ± 0.09 s, and repeated in series, with ISI of 107.6 s. The other recorded FW characteristic unit is a small pulse covering a shorter frequency range (≃ 13 to 18 Hz), the "backbeat". It sometimes occurred by itself or between the groups of 20 Hz pulses. Due to the limited sample rate of the recordings, it is not known if these 20 Hz-pulses occurred along with higher frequency components, as observed for instance near the Antarctic Peninsula (Širović et al., 2004).

Two series of 20 Hz pulses and inserted backbeats, recorded on May 24th, 2013 at 23:00 by (a) OBS RR41 and (b) OBS RR48 are represented in Figure 2.15 (a)&(b). The presence of strong echoes on such signals indicates that recording (a) is close (< 5 km) from the emitting FW. The two OBSs are ≃ 70 km apart. Figure 2.15 (c) represents the PSD for each observation.

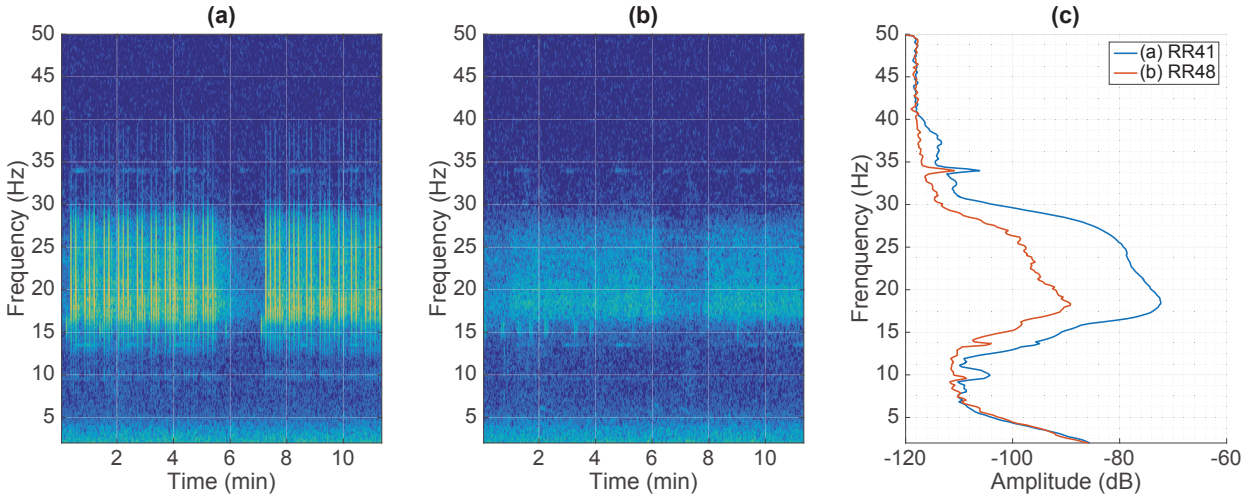

Figure 2.17: Propagation effect on FW call series of 23 calls and incomplete series of 20 calls recorded on May 24th, 2013 at 23:00 (D144). Spectrogram of the series received by OBSs (a) RR41 and (b) RR48 (nfft = 512, overlap = 98%) and, comparison of the associated PSD.

With the distance and due to multipath propagation, clear and distinct pulses of (a) are strongly attenuated. Echoes from all directions are mixed into an unintelligible hubbub. With the same spectrogram parameters, individual pulses are no longer visible. However, global energy variations can be associated with the series. On the PSD, the 20 Hz "bump" looses ≃ 16 dB with the distance. Seven backbeats that were not recorded on OBS RR41 were recorded on RR48. Peaks at ≃ 13.5 Hz and ≃ 34 Hz are due to distant MPBW calls(§ 2.3.2).

2.3.5 Chorus

Animal chorus is generally described as a sustained component of the ambient noise, showing through an increase of noise levels in the specific frequency range of the calling species (Cato and McCauley, 2002; Haver et al., 2018). From a sensor perspective, the limit between individual calls and the unintelligible chorus is often subjective (Leroy et al., 2017b).

Baleen whale chorus manifests itself in the soundscape as an enhanced energy band between 15 and 27 Hz. It has mostly been studied in the SOFAR: hydrophones positioned at these depths

can record chorus energy from long-range whales calls (up to hundreds of km) (McCauley et al., 2018). This is why, the chorus has been used for assessing acoustic whale presence in an area, larger than "clear call" detection radius (Leroy et al., 2016; McCauley et al., 2018). Because they lay on the sea-floor, OBSs are less subject to long-range effects. They still record choruses but, from sources in a supposedly shorter radius. Chorus recorded by the SWIR OBSs change with the types of signals thereof. It is a combination of all whale sounds emitted in the vicinity of the sensor: it might be the effect of remote FW such as Figure 2.17 (b) and ABWs for the upper limit, but the width of its spectrum might be enlarged by the presence of P-calls (Figures 2.15 (b) and 2.14 (b)) or even by "D-calls" Figure 2.12.

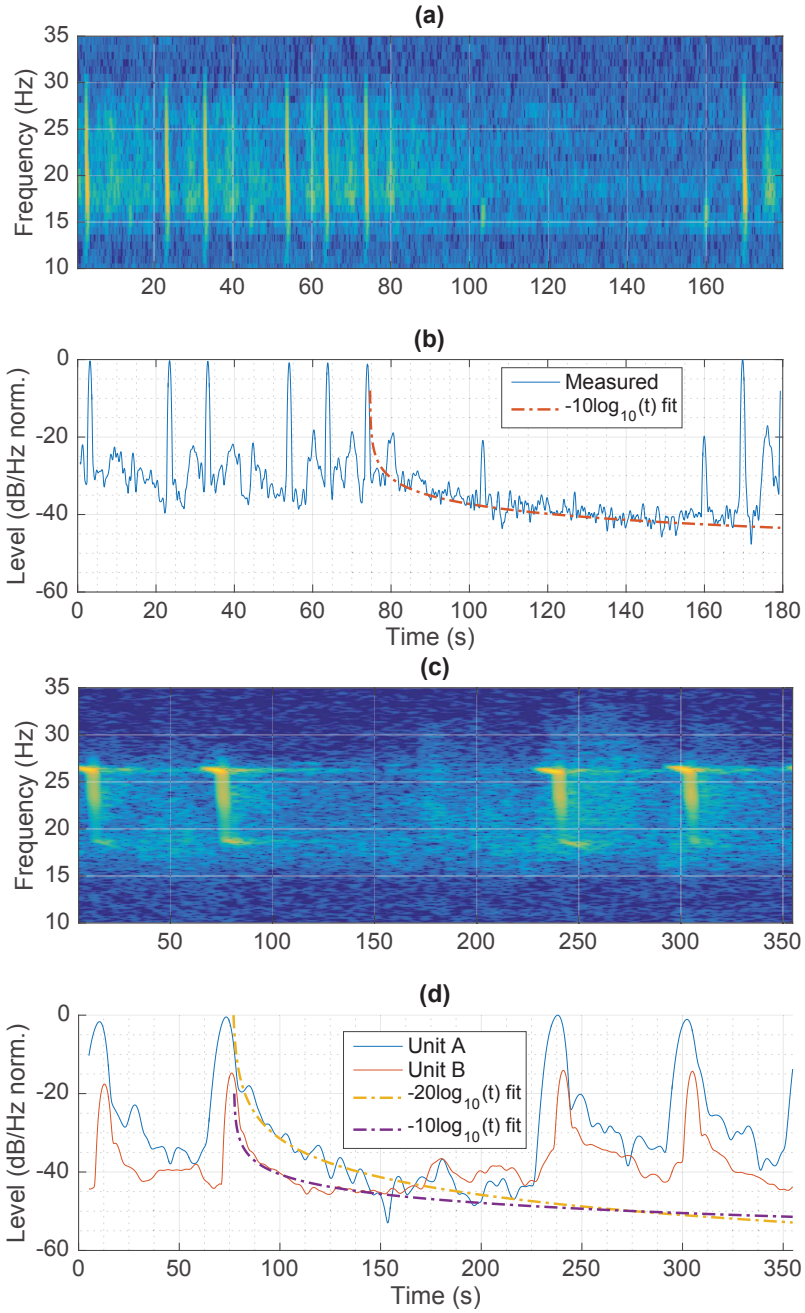

Figure 2.18: Spectrogram and normalized received levels (dB/Hz) of FW pulses (a-b) and ABW calls (c-d). Received levels are measured between 15 and 25 Hz for FW pulses. For the ABW they are measured in the [25.7-26.7] Hz frequency range for unit A and [19.2-25.6] Hz for unit B.

There is a close relationship between chorus and reverberation. The following development illustrates that, from a single sensor point of view, even a close whale can contribute to this sustained component of the ambient noise and that, the reverberation changes regarding the type of signal emitted and the propagation. Figure 2.18 shows examples of the variations of received level per Hz for FW pulses and ABW unit A and C. A logarithmic fitting curve is applied to each song type, starting on the last vocalization of a series.

Reverberation induced by the FW pulse measured in the [15 - 30] Hz as well as ABW unit B, decay in $-10\log t$ dB/Hz while the ABW unit A-induced reverberation decays in $-20\log t$ dB/Hz. The $\times 2$ difference might be due to the duration of unit A: same $-20\log t$ dB/Hz decay is found on unit C (not plotted in Figure 2.18). However, these results highlight that the reverberation of a single close whale also contributes to the global chorus: ICIs are not long enough to retrieve "no whale" ambient noise levels. Building on this thought, ICIs seem to be an interesting indicator relative to the maximum detection range of distinct calls on the OBSs.

2.3.6 Discussion

The focus of section 2.3 is to carefully describe and underline song characteristics of regionally specific species in the Western Indian Ocean. However, it is important to keep in mind that baleen whale share anatomical features, especially when considering sound production (See § 1.4.3.1).

Recent acoustic behavioral studies of BW and FWs were conducted on vocally active tagged animals (Stimpert et al., 2015; Lewis et al., 2018). Emission depths of BWs songs were found between 15 and 30 m (Northeast Pacific BWs) and shallower (<15 m) for FWs (Southern California Bight). Baleen whales seem to dive longer while singing (Stimpert et al., 2015). However, as a marine mammal, they need to breathe at the surface, which brings regular and more extended intervals in songs, associated with ISIs.

Figure 2.19 presents a synthetic scheme of the baleen whale calls recorded by the OBSs, showing their relative frequency spans and intensities. As discussed in paragraph 2.3.5 and highlighted in Figure 2.19, all acoustic signatures overlap and hence, can contribute to the specific chorus bandwidth. However, it appears that each species has at least a characteristic high-intensity unit (or partials) out or at the limit of the chorus span. These specific units appear never to overlap, which makes sense in terms of interspecific competition for auditory space (Pijanowski et al., 2011).

In terms of long-range communication, baleen whale signals seem to satisfy both sides of Naguib and Wiley (2001) paradigm. On the one hand, the tonal parts favor the transmission of the information with minimum degradation through propagation. On the other hand, "pulsed" or vertical parts degradation help receiver to extract information about the signaler's range. The

same observation can be attributed to the difference between FW pulses and backbeats.

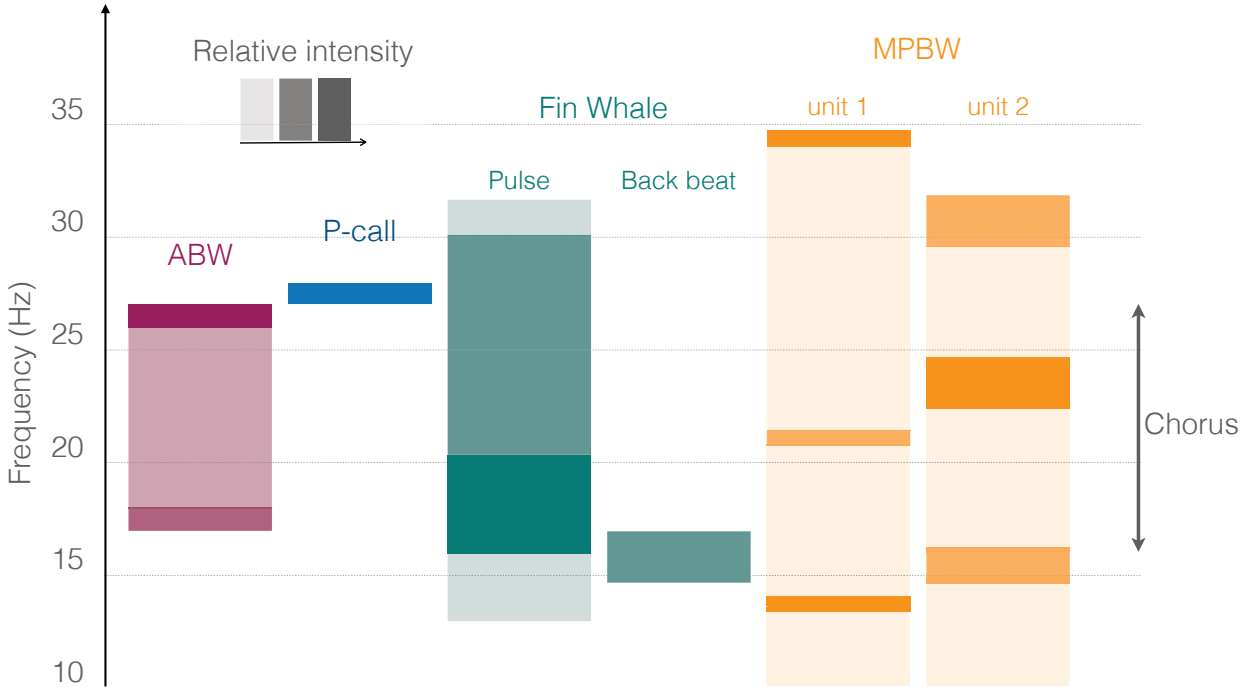

Figure 2.19: Recap of the observed baleen whale calls frequency spans and relative intensities recorded by the RHUM-RUM OBSs in the Western Indian Ocean, between October 2012 and November 2013.

2.4 Multi-sensor observations

2.4.1 Baleen whale seasonal occurrence

The analysis of energy variations in the characteristic frequency bands of the baleen whales species presented in paragraph 2.3 reveals the presence and passage of multiple individuals during the deployment period (Dréo et al., 2019). Six OBSs from the edges of the RHUM-RUM network are processed. The chosen instruments are differentiated in Figure 2.1 by white-circled black stars. The resulting baleen whales spatio-temporal distributions are presented in Figure 2.20.

In the Indian Ocean, the general pattern of baleen whale seasonal migration describes feeding on the *Antarctic feeding grounds* during austral summer and presence in warmer Indian Ocean waters during winter (Samaran et al., 2013). Figure 2.20 supports these observations, with the acoustic activity of all whales recorded during an "extended winter," from February (for the first ABWs) to November (for the last recordings of FW and P-calls.). The observed portion of the Western Indian Ocean might be part of the migration route, breeding area, or winter location of the observed species (Dréo et al., 2019).

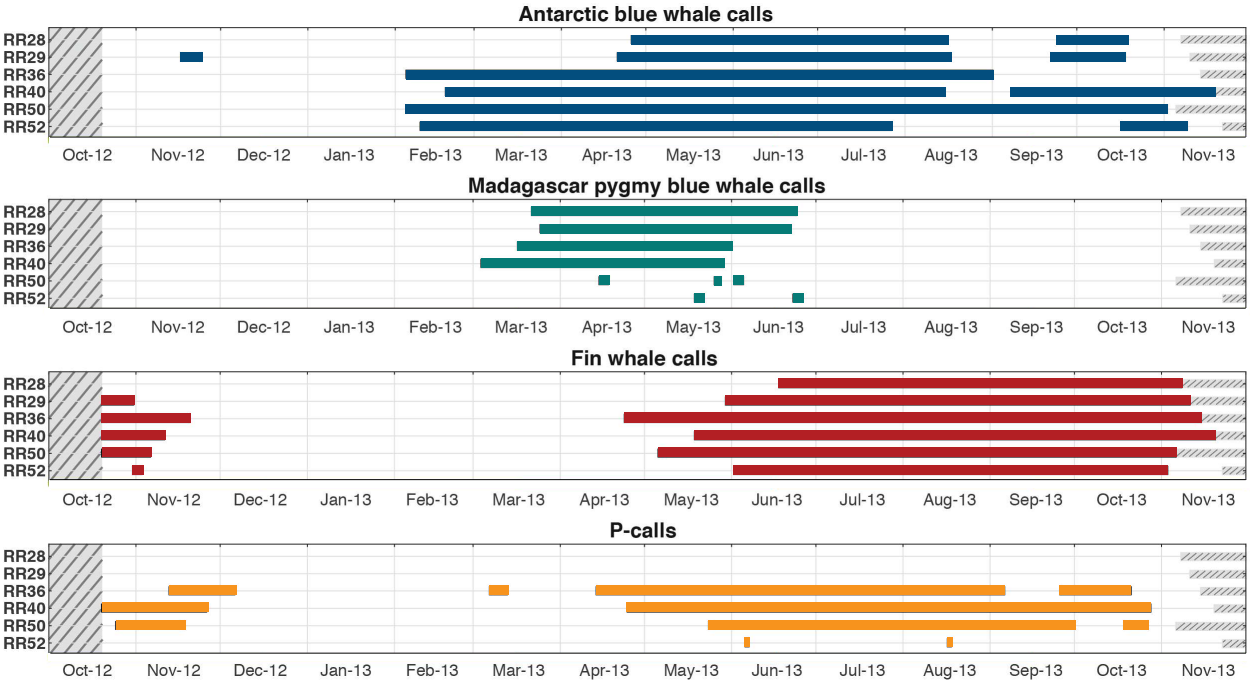

Figure 2.20: Spatio-temporal distribution of baleen whale vocal activity on the RHUM-RUM network, revealed by characteristic frequency band energy variations (Dréo et al., 2019).

MPBW call types were detected mostly by the Western instruments between March and June, confirming their presence only during summer and fall months (January to June) (Samaran et al., 2010c; Stafford et al., 2011; Samaran et al., 2013; Leroy, 2017). The Western Indian Ocean is not a winter or spring location for this population.

P-calls were recorded from April to November confirmed the presence of the species in the Western Indian Ocean with a limit of distribution at 25° S and 056° E.

Figure 2.20 is also used to determine a study period before the SWIR array OBSs failure in ≃ early June (Table 2.1). The closest OBS to the SWIR array is RR40 and, RR40 energy variations reveal an overlapping time period for all species during the month of May 2013 and, therefore, is chosen for training, testing and illustrations in the manuscript.

2.4.2 Localization

The dimension of the SWIR array and the 100 Hz sample frequency of the OBSs are ideal for the multi-sensor observation of whales, and consequently for their localization (Dréo et al., 2019). Vocalizing whales in the vicinity of the array can be localized using a Bayesian inversion method: it generates a probability density function of the location of the acoustic source by comparing some observations to predicted model parameters (Dunn and Hernandez, 2009; Frank and Ferris, 2011).

In the present case, the method compares theoretical time difference of arrival ($TDOA_{th}$) of acoustic rays, simulated using a discretization of the enlarged SWIR surface as a set of possible sound sources to, measured time difference of arrival ($TDOA_m$) from whale emissions. The best match indicates the most likely location of the whale. The method processing steps and succession are illustrated in Figure 2.21.

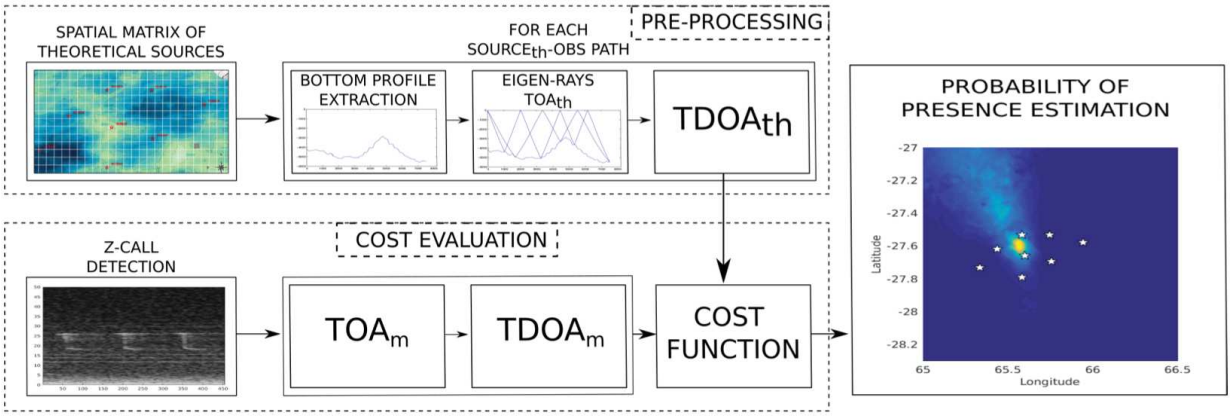

Figure 2.21: Localization method (Dréo et al., 2017; Dréo et al., 2019)

The SWIR surface is split as a spacial matrix of theoretical sources with a maximum precision of 0.01°(about 900 m). It represents a grid of approximatively 20,000 points. Depth of the source is set at 20 m. Acoustic rays are simulated between all possible combinations of sources-OBS eigen-paths using BELLHOP[3], to take into account the acoustic propagation and include environmental characteristics such as the region's complex bathymetry. Bathymetry of the survey area is known from previous experiments (Cannat et al., 2006). Statistical velocity profile of May is used [4]. These simulations estimate theoretical time of arrival (TOA_{th}) then, all $TDOA_{th}$ are calculated.

Measured time of arrival (TOA_m) and $TDOA_m$ ensue from annotated calls. The similarity between each $TDOA_m$ and all $TDOA_{th}$ of the spatial matrix is then measured with a L_1 norm. Results are represented as a *probability of presence* map of the source. Successive locations are connected into a track.

2.5 Conclusion

To close this chapter addressing low-frequency soundscapes recorded by RHUM-RUM OBS and illustrate their diversity, let's look at long-term spectrogram (LTS) representations of the chosen study period, the month of May 2013. LTS representations allow capturing changes in ambient noise as well as soundscapes diversity through time (Curtis et al., 1999; Haver et al., 2017; Miksis-Olds et al., 2013). To represent the contribution of the different source types over

[3]Ocean Acoustics Library web page http://oalib.hlsresearch.com/Rays/
[4]Statistical profile from the SHOM database http://www.shom.fr

the month, LTSs of OBSs RR41 and RR48 are respectively represented in Figure 2.22 (a) and (b). Spectrogram number of sampling points, nfft is fixed to 2^{15}, to draw one point per 10 minutes of recording, with 10% of overlap.

The high energy frequencies, below 2 Hz are attributed to seismic activity (§ 2.2.3). On these representations continuous horizontal lines represent regular acoustic activity at a constant frequency: some characteristic frequencies can be attributed to whale vocal activities, e.g., ABW calls at 26.2 and 18.7 Hz (§ 2.3.5), MPBW at 13.5, 23 and 34 Hz (§ 2.3.2) and, P-calls at 27 Hz (§ 2.3.3). The 15 – 27 Hz frequency band conveys more energy than the rest of the sensor bandwidth. This is attributed to whale chorus, from received activity of remote baleen whale vocalizations (§ 2.3.5). At the passage of a FW, this frequency band is enlarged and amplified (§ 2.3.4). The passage of a ship is revealed by multiple concurrent harmonic frequencies, often covering the entire observed frequency band (§ 2.2.2).

Underwater soundscapes are rich, diverse, and continuously changing. Therefore, the automatic detection of a specific signal among all these overlapping acoustic sources requires dedicated signal processing tools. Regarding that general problem, the main issue faced by current tools is that even signals that are well known at the emission (such as whale calls), change under the constraint of propagation. This issue is tackled in Chapter 3, where the proposed strategy based on the SMF takes into account eventual signal distortion and attenuation and therefore, provides robust detection.

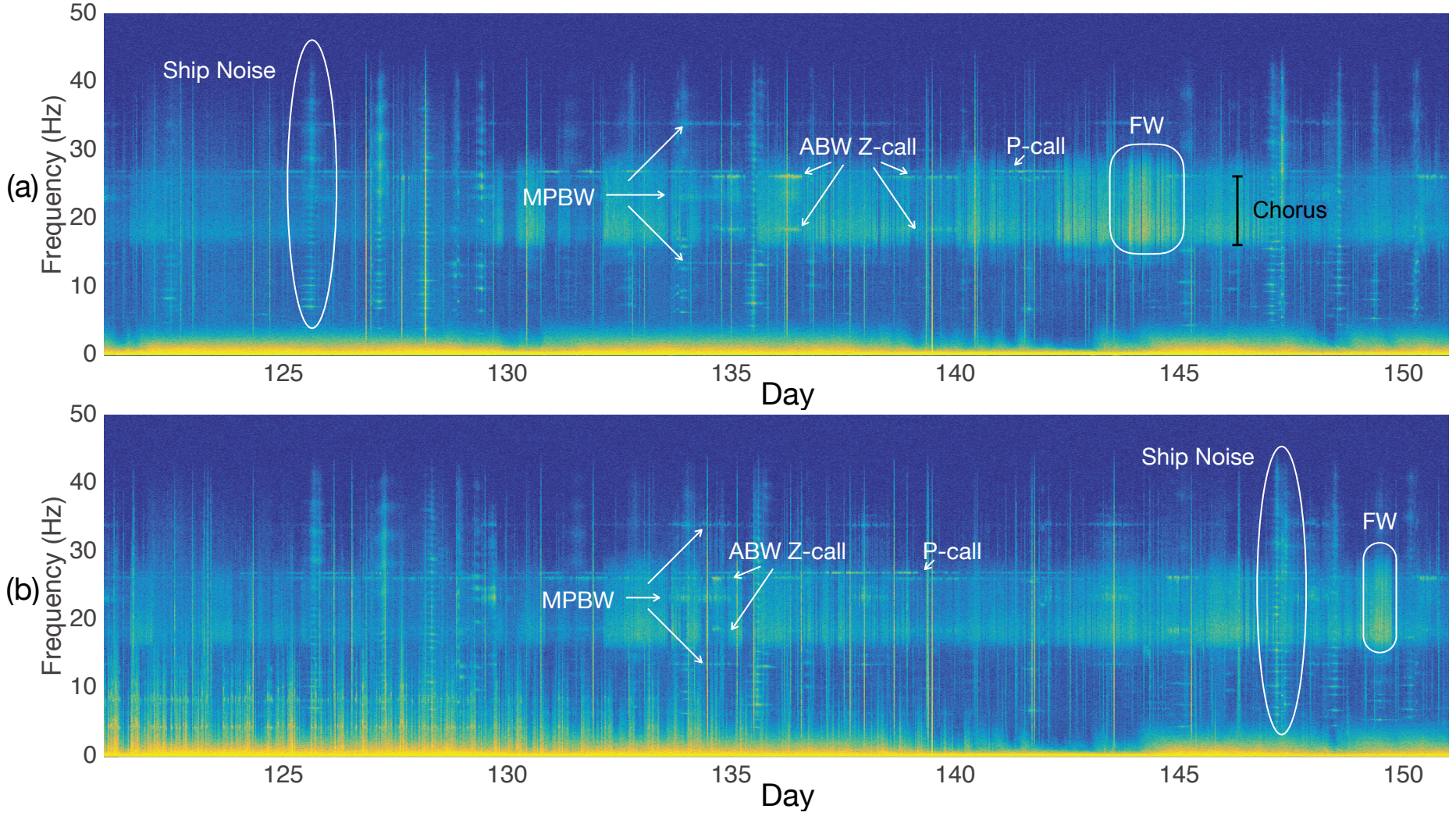

Figure 2.22: Long-term (31 days) spectrograms of May 2013 (year day number 121 to 151) recorded by OBSs RR41 (a) and RR48 (b) (nfft = 2^{15}, overlap = 10%).

Chapter 3

Detection of stereotyped sounds: the Stochastic Matched Filter

Contents

3.1 Are received signals random?

Oceanic physical properties involved in sound propagation, e.g., sound speed profile hence temperature, salinity, depths, or seafloor properties can be season-, region-, range-, depth, and even time-dependent. It leads to differences in how a sound can be modified as it propagates along the source-receiver path and finally affects detection (Binder and Hines, 2019). This section examines if deterministic signals such as BW vocalizations may be considered stochastic after propagation.

3.1.1 Impulse response of the propagation canal

In the time domain, pressure variations of acoustic signals are described as a product, homogeneous to amplitude $\boldsymbol{a}(t)$ and phase $\boldsymbol{\varphi}(t)$ variations. In its complex form, an emitted signal can hence be expressed as $\boldsymbol{s}(t) = \boldsymbol{a}(t)e^{j\boldsymbol{\varphi}(t)}$. Propagation through the underwater medium (see § 1.4) affects the emitted signal (attenuation, geometrical spreading, ambient noise) in a way that the received signal $\boldsymbol{z}(t)$ can be modeled as the convolution between $\boldsymbol{s}(t)$ and the impulse response of the propagation canal $\boldsymbol{h}_{\mathrm{P}}(t)$, embedded in colored ambient noise $\boldsymbol{n}(t)$.

$$\boldsymbol{z}(t) = [\boldsymbol{s}(t) * \boldsymbol{h}_{\mathrm{P}}(t))] + \boldsymbol{n}(t). \tag{3.1}$$

Considering an omni-directional source and using ray-tracing models (§ 1.4.1), the received signal at a given location results from the additive combination of multiple beams, each one representing an acoustic wavefront. Each source-receiver path is referred to as eigenray. The impulse response of the medium along the source-receiver path (respectively located at r_s and r) is expressed in the frequency domain as

$$\mathbf{H}_{\mathrm{P}}(r, r_s, f) = \sum_{i=1}^{\Pi(r)} \mathbf{A}_i(\eta, f)e^{j\omega\tau_i(\eta)}, \tag{3.2}$$

where

- r_s and r are the source and receiver coordinates;
- $\Pi(r)$ is the number of eigenrays reaching r;
- η describes each eigenray trajectory;
- f is the emitted signal's frequency and $\omega = 2\pi f$ its angular frequency;
- $\mathbf{A}_i(\eta, f) = \mathrm{A}_i(\eta)e^{-\beta(f)\eta}$ is the amplitude associated to each eigenray where $\mathrm{A}_i(\eta)$ represents the amplitude along the trajectory and depends on: the soundspeed in the medium, the source-sensor distance, and the emission angle. The other term, $e^{-\beta(f)\eta}$ represents absorption along the trajectory which is a frequency variant parameter and;
- $\tau_i(\eta)$ is the delay (propagation time) associated to each eigenray which depends on sound speed variations along the trajectory (Josso, 2010).

Evaluation of $H_P(r, r_s, f)$ therefore requires accurate knowledge of multiple physical parameters along the source-receiver path(s). Some of these parameters can be acquired along experiments, other can be simulated.

3.1.2 Comparison between real and simulated signals

This paragraph aims at illustrating how complex the propagation can be in a mountainous area such as the SWIR array and, show limitations of classical 2D ray tracing propagation simulations in 3D environments. In order to do so, a comparison is performed between a real recording of an ABW call and the corresponding simulated signals. On May 31st, 2013 at 11:00, an ABW is in the middle of the SWIR array (§ 2.4.2), passing over the central seamount. Measured spectrograms of the same call received on different OBSs are displayed in Figure 3.1 (1-5 a). OBS depths and locations are displayed in Table 2.1.

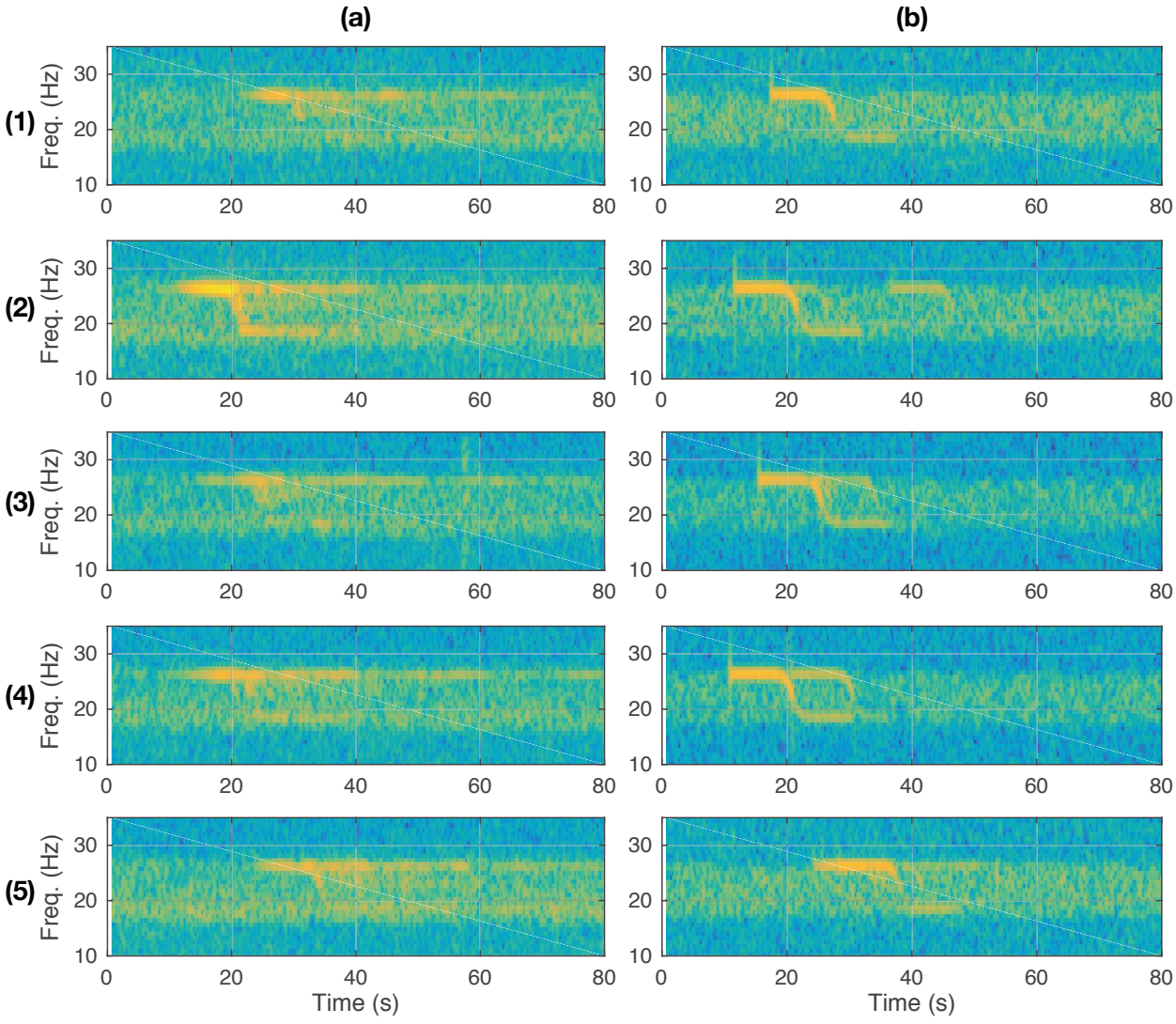

Figure 3.1: Spectrograms of (a) real and (b) simulated received ABW calls on OBS (1) RR41 *13* (2) RR43 *15* (3) RR44 *75* (4) RR47 *5* (5) RR48 *108*, where *italic* value indicates the number of eigenrays. Simulations in 2D are conducted using BELLHOP considering source-OBSs eigenrays (time of arrival - amplitude).

Similarly to the localization process, source-OBS bottom profiles are extracted and May statistical velocity profile is used (cf. Figure 1.11). The sound speed is thus considered range invariant (but still is depth-dependent $c(z)$). Received signals are simulated according to eq. (3.1), as the convolution between a synthetic ABW call (§ 3.5.1.1) and each source-OBS channel-specific impulse response. The impulse response of the medium is estimated in 2D using BELLHOP[1], with a source depth of 30 m. This representation illustrates different source-receiver paths with distinctive distances and bathymetry. Technically, and, from the sensor perspective, this can be understood as the summation of delayed and attenuated synthetic calls corresponding to the different eigenrays. This reverberated signal is drawn into a background noise with higher energy in the $[18-27]$ Hz frequency band. Simulation results are displayed in Figure 3.1 (1-5 b).

In the simulation, the call is emitted by the whale at $t = 0$ s. For visualization and comparison, the time of the recorded data spectrograms are set such as the closest OBSs RR43 (Figure 3.1 (2)) and RR47 (Figure 3.1 (4)) times of arrival of unit B matched the equivalent simulated data. Times of arrival of the most energetic received signal measured on unit B are consistent between simulated and recorded data, except for RR48 (Figure 3.1 (5)). Simulations do not explain the strong unit A reverberations. It may suggest reverberation from the surrounding mountainous sea-floor and therefore, 3D effects that are not considered in the present simulations. Besides, due to the limited discretization of emission angles, simulations underestimate the number of source-receiver eigenrays. Simulations of $\mathrm{H_P}(r, r_s, f)$ provide valuable estimates of most energetic received signal. However, they can not cover the full complexity of the propagation in the SWIR area.

3.1.3 Limitations and position

As described in the section 3.1.2, the estimation of $\mathbf{H}_\mathrm{P}(r, r_s, f)$ requires accurate knowledge of the source-receiver environment which in many cases can not be fulfilled, especially in a passive context. The location of the source is indeed rarely known beforehand and, is often one of the purposes of PAM systems. Without the knowledge of this crucial point, it is difficult to simulate source-receiver paths and hence estimate $\mathbf{H}_\mathrm{P}(r, r_s, f)$. SWIR array uncertainties in OBSs locations also contribute to the randomness of received signals (§ 2.1.1). In addition, the complex bathymetry introduces 3D reverberation that cannot be modeled by bottom-surface reflections only.

In a more global perspective, even if it were possible to model source-receiver paths and received signal accurately, it would still be extremely time-consuming. Most detection methods consider a more practical approach to overcome this issue. It consists in seeing recorded observations as a mixture of the signal of interest and noise, with varying proportions[2]. However, different assumptions can be made on the received signal properties. Instead of considering

[1] Ocean Acoustics Library web page: `http://oalib.hlsresearch.com/Rays/`

[2] The word *observation* is henceforth used according to this definition, while *signal* refers to the useful signal

the received signal as deterministic, it is possible to account for any multi-path propagation by considering the received signal stochastic, with known second-order statistics.

Under these considerations, detection methods have been developed under both signal assumptions (deterministic and stochastic) and rely on noise reduction of the recorded observation. Such a problem can classically be solved through SNR improvement. The following Section (section 3.2) introduces the theoretical elements of the noise reduction problem in the time domain. Two approaches for SNR improvement are considered: the MF that works under assumptions of deterministic signals and white noise (section 3.3) and the SMF derived for stochastic signals and colored noises (section 3.4). Both methods were originally developed or applied in the active signal processing field. Thanks to the reproducibility of BW sounds, MFs have been widely used in the PAM community for detection, but with known limitations (§ 1.5.2). These limitations can be overcome, considering the underlying theoretical assumptions of the SMF. Following the work of Bénard et al. (2011); Caudal and Glotin (2008), the SMF is extended in this thesis to the passive context (Section 3.5). Performances of both methods are presented in Section 4 against a ground-truth dataset.

3.2 Noise reduction problem formulation in the time domain

The foremost goal of the noise reduction problem is to recover the signal of interest (or clean signal) from a noisy observation. The model used through this work, is a superposition of a signal of interest and some noise. Thus, in the discrete time domain, when processing data by blocks of L-samples to ensure stationarity, the observation $\boldsymbol{z}(k)$, the signal of interest $\boldsymbol{s}(k)$ and the noise $\boldsymbol{n}(k)$ vectors are related by

$$\boldsymbol{z}(k) = \boldsymbol{s}(k) + \boldsymbol{n}(k) \tag{3.3}$$

where the signal of interest and noise are zero-mean and uncorrelated. Their respective correlation matrices are given by $\mathbf{R}_s = \mathrm{E}[\boldsymbol{s}(k)\boldsymbol{s}^{\mathrm{T}}(k)]$ and $\mathbf{R}_n = \mathrm{E}[\boldsymbol{n}(k)\boldsymbol{n}^{\mathrm{T}}(k)]$ where $\mathrm{E}[\cdot]$ denotes the mathematical expectation, $(\cdot)^{\mathrm{T}}$ the transpose operation and, where the discrete time index k denotes the center sample of the observation e.g.

$$\boldsymbol{z}(k) = \left\{ z\left[k - \frac{\mathrm{L}-1}{2}\right], \cdots, z[k], \cdots, z\left[k + \frac{\mathrm{L}-1}{2}\right] \right\}. \tag{3.4}$$

Since the signal and noise are uncorrelated, the correlation matrix of the observation $\mathbf{R}_z = \mathrm{E}[\boldsymbol{z}(k)\boldsymbol{z}^{\mathrm{T}}(k)]$ is given by

$$\mathbf{R}_z = \mathbf{R}_s + \mathbf{R}_n. \tag{3.5}$$

The estimation of the signal of interest $\widetilde{\boldsymbol{s}}(k)$ and, therefore, noise reduction is achieved by the application of a linear transformation to $\boldsymbol{z}(k)$ such as

$$\widetilde{\boldsymbol{s}}(k) = \mathbf{H}\boldsymbol{z}(k) \tag{3.6}$$

$$= \mathbf{H}\boldsymbol{s}(k) + \mathbf{H}\boldsymbol{n}(k), \tag{3.7}$$

where $\mathbf{H}$ is a $\mathrm{L} \times \mathrm{L}$ filtering matrix (Benesty et al., 2009). Hence, the correlation matrix of the estimated signal $\mathbf{R}_{\widetilde{s}} = \mathrm{E}[\widetilde{\boldsymbol{s}}(k)\widetilde{\boldsymbol{s}}^{\mathrm{T}}(k)]$ can be expressed as the application of $\mathbf{H}$ respectively to the signal and noise correlation matrices such as

$$\mathbf{R}_{\widetilde{s}} = \mathbf{H}\mathbf{R}_s\mathbf{H}^{\mathrm{T}} + \mathbf{H}\mathbf{R}_n\mathbf{H}^{\mathrm{T}}. \tag{3.8}$$

Finally, with this time-domain formulation, the resolution of the noise reduction problem relies on finding an optimal filter that would attenuate the noise as much as possible while keeping a clean and undistorted signal.

One of the most important measures to noise reduction is the SNR. The input SNR (iSNR) is the ratio of the power of the signal of interest $\sigma^2_{\boldsymbol{s}} = \mathrm{E}[\boldsymbol{s}^2(k)]$ over the power of the background noise $\sigma^2_{\boldsymbol{n}} = \mathrm{E}[\boldsymbol{n}^2(k)]$ such as

$$i\mathrm{SNR} = \frac{\sigma^2_{\boldsymbol{s}}}{\sigma^2_{\boldsymbol{n}}} \tag{3.9}$$

$$= \frac{\mathrm{tr}(\mathbf{R}_s)/\mathrm{L}}{\mathrm{tr}(\mathbf{R}_n)/\mathrm{L}} = \frac{\mathrm{tr}(\mathbf{R}_s)}{\mathrm{tr}(\mathbf{R}_n)}, \tag{3.10}$$

where $\mathrm{tr}(\cdot)$ is the trace operator (sum of the diagonal elements). The power expressed as the operator $\sigma^2_{\cdot}$ also represents the variance of the considered signal.

After the application of the filter and using (3.8), the output SNR (oSNR) can be written such as

$$o\mathrm{SNR}(\mathbf{H}) = \frac{\mathrm{tr}(\mathbf{H}\mathbf{R}_s\mathbf{H}^{\mathrm{T}})}{\mathrm{tr}(\mathbf{H}\mathbf{R}_n\mathbf{H}^{\mathrm{T}})}. \tag{3.11}$$

The noise reduction goal is to find the filtering matrix $\mathbf{H}$ in order to improve the SNR such as $o\mathrm{SNR}(\mathbf{H}) \geq i\mathrm{SNR}$.

3.3 Matched Filter

3.3.1 Theory

To deal with deterministic signals, one of the most popular detection methods is the MF. This process is achieved by cross-correlating a template of the known signal with the incoming recording in order to detect the presence of this template in the observation. In other words, the MF theory shows that the impulse response of the filter optimizing (3.11) for deterministic

signals embedded in white noise, is the reversed signal itself (Van Trees, 2002; Max, 1981). The signal's power is known, consequently (3.11) becomes

$$o\mathrm{SNR}(\mathbf{H}) = \frac{|\boldsymbol{s}(k)\mathbf{H}^{\mathrm{T}}|^2}{\mathbf{H}\mathbf{R}_n\mathbf{H}^{\mathrm{T}}}. \tag{3.12}$$

To determine the oSNR($\mathbf{H}$) upper bound, the Cauchy-Schwarz inquality can be drawn from matrix manipulation of (3.12), such as

$$o\mathrm{SNR}(\mathbf{H}) = \frac{|(\mathbf{H}\mathbf{R}_n^{1/2})^{\mathrm{T}}(\boldsymbol{s}(k)\mathbf{R}_n^{-1/2})|}{(\mathbf{H}\mathbf{R}_n^{1/2})^{\mathrm{T}}(\mathbf{H}\mathbf{R}_n^{1/2})} \leq \boldsymbol{s}(k)\mathbf{R}_n^{-1}\boldsymbol{s}(k)^{\mathrm{T}}. \tag{3.13}$$

The upper-bound is reached if $\mathbf{H}\mathbf{R}_n^{1/2} = \alpha\boldsymbol{s}(k)\mathbf{R}_n^{-1/2}$ where α is a real number. Therefore, the filter can be expressed as

$$\mathbf{H} = \alpha\boldsymbol{s}(k)\mathbf{R}_n^{-1} \tag{3.14}$$

with an impulse response equivalent to the conjugate time reversal of $\boldsymbol{s}(k)$.

3.3.2 Discussion

In practice, templates (and therefore, filters) are usually determined from the recording of a real call, preferably with a high SNR and little distortion or, they are generated synthetically (McCauley et al., 2018). The second option has the advantage of being free of noise and can be adapted, e.g., to long-term variations.

Due to the MF conceptual simplicity and computational efficiency, it is widely used in PAM applications. It has been used for stereotypical call detection directly in the time domain (Samaran et al., 2010a), or in the TF domain as spectrogram correlation to be less sensitive to background noise variations and small frequency shifts (Širović et al., 2004; Balcazar et al., 2015). Besides, it is even implemented as a standard detection tool in bio-acoustic sound analysis software such as PAMGuard and Ishmael.

The definition of the filter requires perfect knowledge of the signal $\boldsymbol{s}(k)$ at all times, as well as white Gaussian background noise to be optimal. These assumptions are hard to meet in experimental conditions and, according to paragraphs 2.2 and 3.1, cannot be satisfied in PAM contexts. The use of the MF in sub-optimal conditions leads to lesser noise-reduction (and detection) performances.

For detection, a more realistic approach consists in considering the signal as *stochastic* where each realization matches a specific source location (i.e., specific impulse response of the propagation channel). However, because the signal is originally deterministic, it is possible to estimate its second-order statistics. This is one of the assumptions for the SMF, presented in section 3.4. Detection under these considerations improves robustness and is less sensitive to long-distance propagation and reverberation (Mori and Gounon, 2000; Bouffaut et al., 2018).

3.4 Stochastic Matched Filter

3.4.1 Introduction

The SMF was presented for the first time by Jean-François Cavassillas (1991). This extension of the MF for stochastic signals embedded in an additive colored noise was rapidly applied to the detection of sonar signals (Cavassillas and Xerri, 1993). In 1997, the SMF was presented as a new formulation for the Karhunène-Loève expansion (KLE) and applied for the first time to 2D signals for image interpolation (Courmontagne and Cavassilas, 1997). As a reminder, a brief note on the KLE theory is presented in § 3.4.2. Based on the KLE approach, dimension-reduction is introduced to the SMF often referred to as Constrained SMF (C-SMF): the number of signal classes (Xerri and Borloz, 2004; Juennard, 2007; Borloz and Xerri, 2011), or the number of filters (Courmontagne et al., 2010; Julien, 2012) can be fixed *a priori*. This last development, allows to describe the SMF as a time-varying linear filter and provides faster *online* data processing and significant representations of the method's filters (Courmontagne et al., 2010). On this theoretical basis, the SMF has been used in multiple applications: in one dimension (in the time domain) for modulated wide-band signal detection in active sonar (Mori and Gounon, 2000; Courmontagne et al., 2010) or audio pattern detection in automatic speech recognition(Bonnal et al., 2010) and, in 2D for synthetic aperture radar (SAR) (Courmontagne, 1999) and synthetic aperture sonar (SAS) (Courmontagne and Chaillan, 2006) imaging improvement.

The SMF is a solution to the problem formulated in (3.11): it maximizes the *o*SNR of stochastic signals embedded in additive colored noises (Cavassillas, 1991). Corresponding theoretical elements, following up on the time-varying filter approach can be found in § 3.4.3. Unlike MF methods and according to section 3.1, the SMF assumptions are rigorously adapted to detection in a passive context. Based on this understanding, the SMF has recently been used for sperm whale click and echo detection (Courmontagne, 2010) that even outperforms Teager-Kaiser-Mallat filter method (Caudal and Glotin, 2008). To pursue this work, an extension of the SMF to the passive context for the detection of BWs calls was developed and is presented in Section 3.5.

3.4.2 The Karhunen–Loève expansion

Let the L × 1 vector $\boldsymbol{x}(k)$ denote a sequence of data drawn from a zero-mean stationary process with a correlation matrix $\mathbf{R}_{\boldsymbol{x}}$ (Benesty et al., 2009). The matrix $\mathbf{R}_{\boldsymbol{x}}$ can be diagonalized as

$$\mathbf{Q}^{\mathrm{T}}\mathbf{R}_{\boldsymbol{x}}\mathbf{Q} = \mathbf{\Lambda} \tag{3.15}$$

where $\mathbf{Q} = [\boldsymbol{q}_1, \boldsymbol{q}_2 \cdots \boldsymbol{q}_{\mathrm{L}}]$ are the orthonormal eigenvectors associated to the eigenvalues $\mathbf{\Lambda} = \mathrm{diag}[\lambda_1, \lambda_2, \cdots, \lambda_{\mathrm{L}}]$ of the matrix $\mathbf{R}_{\boldsymbol{x}}$. Therefore, the KLE of $\boldsymbol{x}(k)$ describes the vector as a pondered combination of the eigenvectors $\mathbf{R}_{\boldsymbol{x}}$ such as

$$\boldsymbol{x}(k) = \sum_{l=1}^{\mathrm{L}} c_{x,l}(k)\boldsymbol{q}_l, \tag{3.16}$$

where the expansion coefficients are

$$c_{x,l}(k) = \boldsymbol{q}_l^{\mathrm{T}} \boldsymbol{x}(k) \text{ and,} \tag{3.17}$$

$l = 1, 2, \cdots, \mathrm{L}$ are the sub-band indices. In a sense, the KLE can be seen as analogous to a Fourier transform but with random variables instead of fixed coefficients and an expansion basis derived from $\mathbf{R}_{\boldsymbol{x}}$ instead of sinusoidal functions.

A few properties can be derived from equations (3.16) and (3.17):

- coefficients $c_{x,l}(k)$ are zero-mean on each sub-band such as

$$\mathrm{E}[c_{x,l}(k)] = 0, \quad l = 1, 2, \cdots, \mathrm{L}, \tag{3.18}$$

- the covariance matrix of the coefficients is null, except on diagonal terms such as

$$\mathrm{E}[c_{x,i}(k) c_{x,j}(k)] = \begin{cases} \lambda_i, & i = j \\ 0, & i \neq j \end{cases}, \tag{3.19}$$

- the KLE allows energy conservation such as

$$\sum_{l=1}^{\mathrm{L}} c_{x,l}^2(k) = \|\boldsymbol{x}(k)\|_2^2 \tag{3.20}$$

  where the operator $\|.\|_2$ is the Euclidean norm.

3.4.3 Theory

Due to the propagation, the model of the signal of interest is never perfectly known. However, it can be considered as a random signal with known second-order statistics. Both signal and noise lie in unknown subspaces. In that context, the SMF was introduced as an extension of the MF, in the sense of optimal filtering and SNR maximization. The process uses a new formulation of the observation input SNR, expressed in the form of a Rayleigh quotient. The optimization of this ratio leads to the construction of a set of sub-band filters, designed to maximize the oSNR. The filters separate the signal and noise when projected onto an optimal subspace. The optimal filters are hence, the set eigenvectors, associated with the greatest eigenvalues of the generalized eigenvalue problem (GEP) involving the signal and noise covariance matrices. They maximize the oSNR.

Note: For additional details on the SMF theory, the reader can refer to Courmontagne (2010) where the complete method is thoroughly described.

3.4.3.1 Output SNR improvement formulation

In the discrete time domain, the received observation follows (3.3), where $\boldsymbol{s}(k)$ and $\boldsymbol{n}(k)$ are realizations of random functions whose second order statistics are known as covariance matrices.

Using the signal of interest and noise variances, respectively $\sigma_{\boldsymbol{s}}^2 = \mathrm{E}[\boldsymbol{s}^2(k)]$ and $\sigma_{\boldsymbol{n}}^2 = \mathrm{E}[\boldsymbol{n}^2(k)]$, the observation is re-written as

$$\boldsymbol{z}(k) = \boldsymbol{s}(k) + \boldsymbol{n}(k) \tag{3.21}$$

$$= \sigma_{\boldsymbol{s}} \boldsymbol{s}_0(k) + \sigma_{\boldsymbol{n}} \boldsymbol{n}_0(k), \tag{3.22}$$

with $\mathrm{E}[\boldsymbol{s}_0^2] = 1$ and $\mathrm{E}[\boldsymbol{n}_0^2] = 1$. Recall that these vectors correspond to L successive samples. The reduced signal $\boldsymbol{s}_0(k)$ and noise $\boldsymbol{n}_0(k)$ are both assumed centered, second-order stationary and mutually independent (Courmontagne et al., 2010). Under these assumptions, it is easy to show the relationship between the correlation matrix of $\boldsymbol{s}(k)$ and the covariance matrix of $\boldsymbol{s}_0(k)$ such as

$$\mathbf{R}_{\boldsymbol{s}} = \sigma_{\boldsymbol{s}}^2 \boldsymbol{\Gamma}_{\boldsymbol{s}_0}, \tag{3.23}$$

with $\boldsymbol{\Gamma}_{\boldsymbol{s}_0}$ the covariance matrix of the reduced signal. The same observation can be shown for the reduced noise, with covariance matrix denoted by $\boldsymbol{\Gamma}_{\boldsymbol{n}_0}$. Covariance matrices are assumed diagonalized and, to avoid any confusion, $\mathbf{H}$ is replaced by $\boldsymbol{\Phi}$. Therefore, with these new notations, (3.11) becomes

$$o\mathrm{SNR} = \frac{\mathrm{tr}(\boldsymbol{\Phi}\boldsymbol{\Gamma}_{\boldsymbol{s}_0}\boldsymbol{\Phi}^{\mathrm{T}})}{\mathrm{tr}(\boldsymbol{\Phi}\boldsymbol{\Gamma}_{\boldsymbol{n}_0}\boldsymbol{\Phi}^{\mathrm{T}})} \tag{3.24}$$

$$= \frac{\sigma_{\boldsymbol{s}}^2}{\sigma_{\boldsymbol{n}}^2} \frac{\sum_{l=1}^{\mathrm{L}} \boldsymbol{\Phi}_l \boldsymbol{\Gamma}_{\boldsymbol{s}_0} \boldsymbol{\Phi}_l^{\mathrm{T}}}{\sum_{l=1}^{\mathrm{L}} \boldsymbol{\Phi}_l \boldsymbol{\Gamma}_{\boldsymbol{n}_0} \boldsymbol{\Phi}_l^{\mathrm{T}}}, \tag{3.25}$$

where the ratio $\frac{\sigma_{\boldsymbol{s}}^2}{\sigma_{\boldsymbol{n}}^2}$ is the iSNR (3.9). Using matricial notation, the SMF oSNR optimization can be written in the form of a *generalized Rayleigh quotient* such as

$$o\mathrm{SNR} = \frac{\sigma_{\boldsymbol{s}}^2}{\sigma_{\boldsymbol{n}}^2} \frac{\boldsymbol{\Phi}\boldsymbol{\Gamma}_{\boldsymbol{s}_0}\boldsymbol{\Phi}^{\mathrm{T}}}{\boldsymbol{\Phi}\boldsymbol{\Gamma}_{\boldsymbol{n}_0}\boldsymbol{\Phi}^{\mathrm{T}}}, \tag{3.26}$$

The oSNR improvement relies on finding $\boldsymbol{\Phi}$ such as the $\frac{\boldsymbol{\Phi}\boldsymbol{\Gamma}_{\boldsymbol{s}_0}\boldsymbol{\Phi}^{\mathrm{T}}}{\boldsymbol{\Phi}\boldsymbol{\Gamma}_{\boldsymbol{n}_0}\boldsymbol{\Phi}^{\mathrm{T}}} \geq 1$. In this well known optimization problem, oSNR is maximum if $\boldsymbol{\Phi}$ is the eigenvector associated to the greatest eigenvalue of $\mathbf{C} = \boldsymbol{\Gamma}_{\boldsymbol{n}_0}^{-1}\boldsymbol{\Gamma}_{\boldsymbol{s}_0}$, if the noise matrix is invertible. In other words, the basis $\{\boldsymbol{\Phi}_l\}_{l=1,2,\cdots \mathrm{L}}$ (of L-dimensional deterministic vectors) that ensures the maximization of the SNR is solution to the GEP that links the respective signal and noise reduced covariance matrices such as

$$\boldsymbol{\Gamma}_{\boldsymbol{s}_0}\boldsymbol{\Phi}_l = \lambda_l \boldsymbol{\Gamma}_{\boldsymbol{n}_0}\boldsymbol{\Phi}_l, \tag{3.27}$$

with λ_l and $\boldsymbol{\Phi}_l$ the eigenvalues and associated eigenvectors of $\mathbf{C}$. The $\boldsymbol{\Phi}_i$ vectors are normalized in order to satisfy $\boldsymbol{\Phi}_l^{\mathrm{T}}\boldsymbol{\Gamma}_{\boldsymbol{n}_0}\boldsymbol{\Phi}_l = 1$. It is also possible to derive another L-dimensional basis $\{\boldsymbol{\Psi}_l\}_{l=1,2,\cdots \mathrm{L}}$ formed with the eigenvalues of $\mathbf{C}^{\mathrm{T}}$ such as

$$\boldsymbol{\Psi} = \boldsymbol{\Gamma}_{\boldsymbol{n}_0}\boldsymbol{\Phi}, \tag{3.28}$$

if $i \neq j$, $\boldsymbol{\Phi}_i$ and $\boldsymbol{\Psi}_j$ are orthogonal.

3.4.3.2 KLE and signal estimation

Relying on the KLE, the observation vector $\boldsymbol{z}(k)$ can be decomposed into a sum of known deterministic functions (vectors) weighted by uncorrelated random coefficients $z_{l,k}$ (3.17). The SMF theory (Cavassillas and Xerri, 1993) demonstrates that theses functions are determined by the $\boldsymbol{\Psi}$ basis such as

$$\boldsymbol{z}(k) = \sum_{l=1}^{\mathrm{L}} z_{l,k} \boldsymbol{\Psi}_l. \tag{3.29}$$

The random coefficients $z_{l,k}$ are determined using the scalar product between $\boldsymbol{z}(k)$ and the deterministic vectors Φ_l such as

$$z_{l,k} = \boldsymbol{z}(k)^{\mathrm{T}} \boldsymbol{\Phi}_l. \tag{3.30}$$

The noise, decomposed on the $\boldsymbol{\Psi}$ basis gives uncorrelated coefficients with unit power while the signal expanded on $\boldsymbol{\Psi}$ has uncorrelated coefficients of power equal to λ_l. In other words, the new representational orthogonal space determined by $\boldsymbol{\Psi}$ and $\boldsymbol{\Phi}$ allows the differentiation between observation coefficients ($z_{l,k}$) that are carrying more signal than noise. Since the development is performed on reduced values, it makes sense that the distinction between the signal and noise occurs at a threshold Q representing the number product between the eigenvalues λ_l and oSNR that exceed one (Courmontagne, 2010).

The dimension reduction properties of the KLE can, therefore, be applied. The signal is estimated from the observation by a reconstruction up to the order Q. Then, (3.29) becomes

$$\widetilde{\boldsymbol{s}}(k) = \sum_{l=1}^{\mathrm{Q} \leq \mathrm{L}} z_{l,k} \boldsymbol{\Psi}_l. \tag{3.31}$$

The order Q is the dimension of the basis $\boldsymbol{\Psi}$ that minimizes the mean square error between the signal of interest and its approximation.

3.4.3.3 Time-varying linear filter

In practice, to overcome the stationary issue of the noise, the L-sample block of the observation presented by (3.4) is considered as an odd sliding window, centered on the k^{th} sample. As opposed to the full observed window still denoted as $\boldsymbol{z}(k)$, variables depending on the specific sample k are denoted with brackets, e.g., $z[k]$. The iSNR is also a time-dependent parameter, the iSNR of the k^{th} sample is noted $\rho[k]$.

An approximation of the signal of interest is then reconstructed by keeping only Q$[k]$ components associated to the eigenvalues greater than a given threshold such as

$$\widetilde{s}_{\mathrm{Q}[k]}[k] = \sum_{l=1}^{\mathrm{Q}[k]} z_{l,k} \boldsymbol{\Psi}_l[\tfrac{\mathrm{L}+1}{2}], \tag{3.32}$$

with $z_{l,k} = \boldsymbol{z}(k)^{\mathrm{T}} \boldsymbol{\Phi}_l$ and, where $\boldsymbol{\Psi}_l[\frac{\mathrm{L}+1}{2}]$ denotes the center sample of the vector $\boldsymbol{\Psi}_l$. The order Q$[k]$ is then mathematically expressed as

$$\mathrm{Q}[k] = \#\left(\rho[k] \times \lambda_l \geq 1\right) \tag{3.33}$$

where #(.) is used as an abbreviation for *the number of*.

In the form of a linear transformation (3.6), the approximation of the signal of interest is then reconstructed for the central sample of the sliding window, using the product of the observation $\boldsymbol{z}(k)$ and a Q[k]-dimensioned filter $\boldsymbol{h}_{\mathrm{Q}[k]}$ such as

$$\widetilde{\boldsymbol{s}}_{\mathbf{Q}[k]}(k) = \boldsymbol{z}(k)^{\mathrm{T}} \boldsymbol{h}_{\mathrm{Q}[k]}, \tag{3.34}$$

where the filter being expressed as

$$\boldsymbol{h}_{\mathrm{Q}[k]} = \sum_{l=1}^{\mathrm{Q}[k]} \boldsymbol{\Psi}_l [\tfrac{\mathrm{L}+1}{2}] \boldsymbol{\Phi}_l \tag{3.35}$$

and where $\{\boldsymbol{\Psi}_l\}$ and $\{\boldsymbol{\Phi}_l\}$ are the SMF L-dimensional basis determined using equations (3.27) and (3.28).

The order of the filter Q[k], takes values between 1 and L, limiting (3.35) to compute a maximum of L vectors filters ($\boldsymbol{h}_{\mathrm{Q}[k]}$ with $1 \leq \mathrm{Q}[k] \leq \mathrm{L}$). Each one of these filters can be seen as a sub-band filter. The first filter $\boldsymbol{h}_1$ ensures the maximization of the SNR. A $\mathrm{Q}max \leq \mathrm{L}$ can be set, and, then $\boldsymbol{h}_{\mathrm{Q}max}$ is given by the superposition of all filters up to $\mathrm{Q}max$, covering the entire useful signal bandwidth (Courmontagne, 2010). All eigenvectors associated with eigenvalues greater than 1 contribute into improving the SNR; therefore, choosing a $\mathrm{Q}max \leq \mathrm{L}$ only helps to reduce the size of the filter bank (Cavassilas et al., 1997). The different filters representation in the frequency domain and comparison to the signal's spectrum are presented in § 3.5.1.2.

3.4.3.4 Discussion

Theoretical elements presented in this Section are general in the sense that they are valid for both active and passive acoustic signal detection. In this development, equation (3.35) presents the SMF in the form of a classical noise-reduction problem, where the signal is estimated by filtering the observation. This filter, or more exactly, this filter *bank* to consider all sub-band filters in the appellation, is seen to be directly dependent on the *a priori* knowledge of the signal and noise covariances matrices (equations (3.27) and (3.28)). The filter bank is adapted to the signal class (Juennard, 2007). An underlying advantage of the time-varying filter approach of the SMF is to separate the processing into two steps, summarized in Figure 3.2. The first step aims at determining the filter bank and, can be seen as an *offline* process. The second one is the application of the time-varying filter to the observation, to preserve and reconstruct the signal only. Therefore, it is referred to as an *online* process. The terms *online* and *offline* are used as opposites. The *offline* process is not time-dependent and can be computed beforehand and stored. The *online* process is time-dependent and, except for differences accounted for by the sliding window of size L, can be achieved in real-time.

In active sonar applications, because they are essential to the experiments, the emitted signal characteristics, e.g., exact time and duration of the emission, signal spectro-temporal evolution

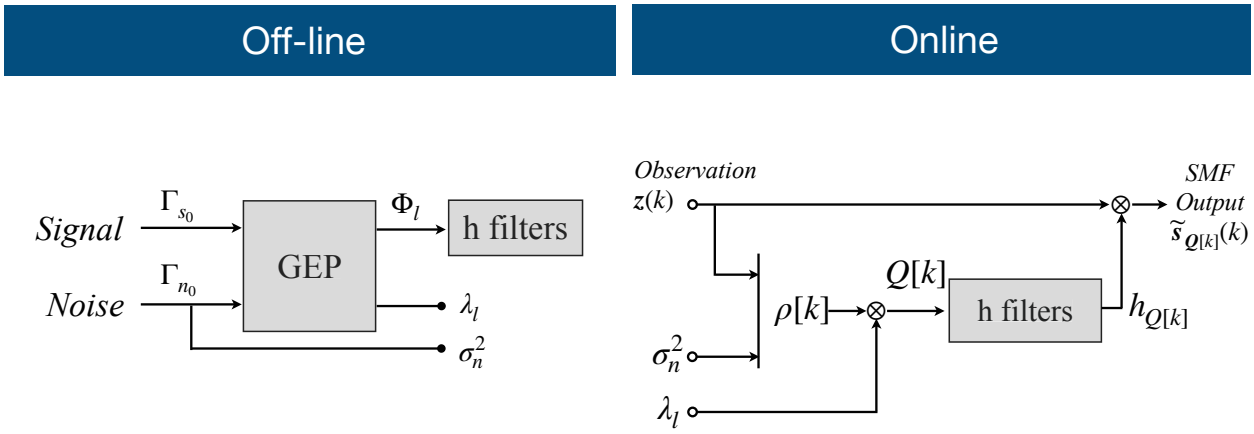

Figure 3.2: Scheme of classical SMF processing, as a succession of *offline* and *online* processes.

are known (and so is $\mathbf{\Gamma}_{s_0}$). In these circumstances, it is quite easy to determine a signal-free portion of recordings, i.e., as a background noise measurement before starting the experiment. Applying the principle of ergodicity, it is hence, possible to estimate the reduced noise covariance matrix $\mathbf{\Gamma}_{n_0}$ and the noise power $\sigma^2_{\boldsymbol{n}}$. Therefore, the time-dependent *i*SNR $\boldsymbol{\rho}(k)$ is described as the ratio between the residual power of the observation after removing the noise power, over the noise power (Julien, 2012)

$$\boldsymbol{\rho}(k) = \frac{\sigma^2_{\boldsymbol{z}(k)} - \sigma^2_{\boldsymbol{n}}}{\sigma^2_{\boldsymbol{n}}}. \tag{3.36}$$

The noise covariance matrix $\mathbf{\Gamma}_{n_0}$ and $\boldsymbol{h}_l$ can be updated as often as required.

The application of the SMF to the passive context is not as straightforward as in active applications. As discussed in § 3.3.2, the advantage of stereotypical baleen whales calls is that it is possible to determine their templates and then estimate the signal covariance matrix (§ 3.5.1.1). Classical SMF noise estimation is built on the strong assumptions that a portion of the observation can be signal-free. In the passive context, there is no *a priori* knowledge on signal times of occurrence; a signal of interest can occur anytime. One solution to that particular issue would be noise annotation in datasets. Although efficient, this solution diverges from the principle of fully automatic detectors, designed for large dataset analysis. The problem of background noise estimation is addressed in § 3.5.2.1. Last, due to the plurality of sound sources overlapping in close frequency bands (paragraph 2.2), SNR estimation is a widely acknowledge challenge in the passive context (Mellinger and Clark, 2006). Ratio as expressed by (3.36) are indeed not sufficient to describe SNR variations. Keeping in mind the idea behind (3.36), $\boldsymbol{\rho}(k)$ estimation issue is tackled in § 3.5.2.2 for the passive context.

3.5 Extension of the Stochastic Matched Filter to the passive context

The extension of the SMF to the passive context is decomposed in two processes: first *offline* § 3.5.1, then *online* § 3.5.2. This development is illustrated for the detection of ABW calls.

3.5.1 Offline SMF

The *offline* SMF aim at generating a filter bank, matching the signal and adaptable to a large class of noises. To solve the GEP (3.27), the signal covariance matrix is required (§ 3.5.1.1). The noise covariance matrix is calculated over a classical simulated underwater colored noise, with higher energy at low frequencies. The resulting filter bank is presented in § 3.5.1.2.

3.5.1.1 Signal's covariance matrix estimation

The SMF requires an accurate knowledge of the signal's covariance matrix. Computing $\mathbf{\Gamma}_{s_0}$ requires the signal's probability density function. This can be obtained either by the estimation of an acoustic pattern from a dataset of multiple high SNR representative signals (Stafford et al., 2004; Balcazar et al., 2015) or spectrum modeling using, for example, a mixture Gaussian models (Samaran et al., 2008, 2010c).

For a signal with a known instantaneous frequency such as the ABW call, it is possible to find a corresponding parametric model that describes the time variations of the amplitude and phase (Socheleau et al., 2015; Leroy et al., 2016; Bouffaut et al., 2018). Such parametric models are based on the complex form of an amplitude modulated and frequency modulated acoustic signal $s(n) = a(n)e^{j\varphi(n)}$, with $a(n)$ the time-varying amplitude and $\varphi(n)$ the time-varying phase. From the definition of the instantaneous frequency and its parametric expression as a function of the (continuous) time

$$f(t) = f_c + \frac{1}{2\pi}\frac{\mathrm{d}\varphi(t)}{\mathrm{d}t} = f_c + \mathrm{L} + \frac{\mathrm{U}-\mathrm{L}}{1+e^{\alpha(t-\mathrm{M})}}, \tag{3.37}$$

it is possible to derive the expression of the time-varying phase $\varphi(n)$, where n denotes the discrete time, as

$$\varphi(n) = 2\pi\left(\mathrm{L}\frac{n}{f_s} + \frac{\mathrm{U}-\mathrm{L}}{\alpha}\ln\left(\frac{1+e^{-\alpha\mathrm{M}}}{1+e^{\alpha(\frac{n}{f_s}-\mathrm{M})}}\right)\right) + \varphi_0, \tag{3.38}$$

where $f_c = 22.6$ Hz is the central frequency in the $[15-30]$ Hz bandwidth, L and U are respectively linked to the lower and upper asymptotes of the Z-call, M represents the time shift and α the grow rate. The amplitude $a(n)$, is set to vary in accordance with the energetic difference between unit A and C (Section 2.3.1). To compensate observed annual and seasonal frequency variations within the call, the signal is built as a summation of multiple frequency modulated signals with the following parameters: $f_s = 100$ Hz, $\mathrm{T}_{Zcall} = 20$ s, $\mathrm{L} = [-4.5; -4; -3.5]$ Hz, $\mathrm{U} = [3.2; 3.6; 4]$ Hz, $\mathrm{M} = [\frac{\mathrm{T}_{Zcall}}{2}; \frac{(\mathrm{T}_{Zcall}+0.5)}{2}; \frac{(\mathrm{T}_{Zcall}+1)}{2}]$ s and $\alpha = 1.8$. This temporal signature is then used to compute the signal's reduced covariance matrix $\mathbf{\Gamma}_{s_0}$. The dimension of the signal is also chosen to be the size of the $\boldsymbol{z}(k)$ sliding window: therefore $\mathrm{L} = 2001$ bins.

3.5.1.2 Filter bank

The earlier discussion in § 3.4.3.4 led to a key element of the extension of the SMF to the passive context: it is essential to design a filter bank that is applicable at all time with no *a priori* knowledge on the instantaneous variations of the noise. The considered approach is to compute the filter bank using a simple low-frequency synthetic sea-colored noise (white noise with -6 dB/decade) to estimate $\mathbf{\Gamma}_{\boldsymbol{n}_0}$. All required elements are then gathered to solve $\mathbf{\Psi}_l\left[\frac{L+1}{2}\right]\mathbf{\Phi}_l$ from (3.34). Filters are saved and stored for further processing in the *online* application of the SMF. Observation noise-dependent parameters such as $\rho[k]$ and and λ_l are estimated *online* after a pre-processing of the observation (§ 3.5.2.1).

Figure 3.3 illustrates the frequency response of filters $\boldsymbol{h}_1$, $\boldsymbol{h}_{10}$ and $\boldsymbol{h}_{Qmax}$. Symbol $\boldsymbol{h}_Q$ denotes the superposition of the Q-first filters of the optimal linear filter bank for ABW call detection and capital $\mathbf{H}_Q$ denotes their spectral representation. They are compared to the spectrum of the simulated reference signal (§ 3.5.1.1) considered as the optimal filter, $\mathbf{H}Opt$. Note that $\mathbf{H}Opt$ is also the Z-call MF filter. The reduction of the matrix dimension of the filter bank is performed by setting the order $Qmax$ to select only the 10^{th} percentile of the eigenvalues ($Qmax = 201$). The $L \times L$ filter bank matrix is then reduced to $L \times Qmax$.

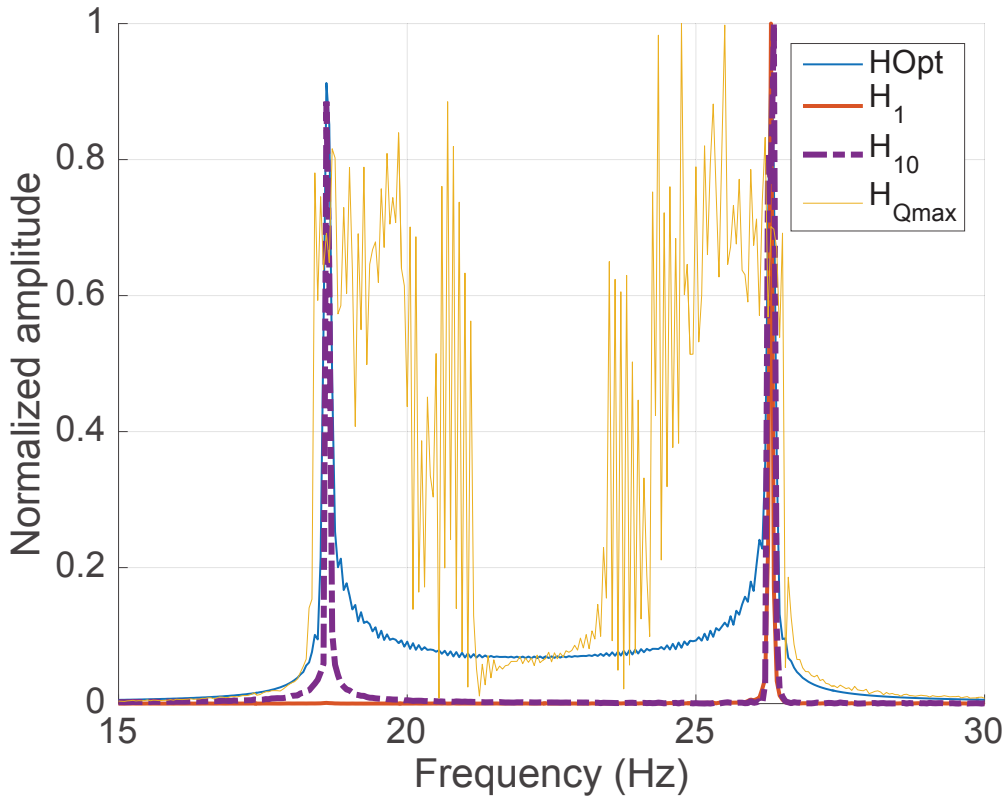

Figure 3.3: Spectrum of three filters (H_1, H_{10}, H_{Qmax}) of the permanent filter bank compared to the spectral representation of the reference signal (HOpt).

To maximize *o*SNR, the first filter H_1 is a short-band filter centered on the most energetic component of the call, unit A at ~ 26.3 Hz. This filter is applied when the estimation of $\rho[k]$ indicates that there is no signal and achieves "noise cancellation." The superposition of the

first 10 filters H_{10} leads to two slightly larger band-pass filters, respectively centered on units A (~ 26.3 Hz) and C (~ 18.6 Hz) of the call. H_{Qmax} represents the superposition of the maximum number of filters that are applied when the estimated input $\rho[k]$ (Section 3.5.2.2) is high enough. The filter's pattern is then close to HOpt, but when applied, it band-pass filters the observation in the signal frequency band: it reconstructs the entire signal.

3.5.2 Online SMF

The outcome of the *online* step of the SMF is the detected signals. To perform detection, the observation backgound noise is first estimated using a TF analysis (Section 3.5.2.1). This estimation provides both the real noise covariance matrix $\mathbf{\Gamma}_{\boldsymbol{n}_0}$ and a noise reference for *online* input SNR, i.e., $\boldsymbol{\rho}(k)$ estimation (Section 3.5.2.2). Knowing all the parameters of the GEP for the observation (3.27), eigenvalues λ_l are evaluated. According to (3.33), for each sample of the observation, it is then possible to evaluate the corresponding $Q[k]$, and apply the appropriate number of filters corresponding to the current observation.

3.5.2.1 Online noise's covariance matrix estimation

The SMF has been used for the passive acoustic detection of whales clicks but, the noise covariance matrix is estimated on annotated signal-free samples (Courmontagne, 2010; Bénard et al., 2011). Although it is quite convenient for supervised detection on relatively short records, it becomes tedious and unpractical for automatic detection on multiple-hours-long passive acoustic monitoring datasets with highly varying background noise. Consequently, the following development deals with finding a way to blindly take into account noise variations despite frequency dependence and high energy events occurrence, aiming at performing accurate estimation of $\mathbf{\Gamma}_{\boldsymbol{n}_0}$ even in the presence of the signal of interest $\boldsymbol{s}(k)$ (Bouffaut et al., 2017b).

To that extent, the proposed method relies on a TF analysis: in these representations, transient signals are concentrated around a few TF bins (Socheleau et al., 2015). The noise statistics can hence be estimated after the application of a filter aiming at attenuating the contrast between transient signals and the noise. A three-step procedure is presented.

First, the observation is converted into the TF domain by a short time fourier transform (STFT) noted $\boldsymbol{\gamma}_{\boldsymbol{z}}(k', f)$. The STFT is a complex transformation, its squared modulus, represents the energy of the observation. The STFT of the observation is calculated using weighted overlapping windows and, the associated parameters (window size and type, overlap rate, number of samples for the spectral analysis) are related to the sampling frequency, the duration of the observation and the duration of the useful signal.

Then, the STFT modulus, $|\boldsymbol{\gamma}_{\boldsymbol{z}}(k', f)|$ is used to filter transient signals. The attenuation of the transient signal to noise contrast is performed using median absolute deviation (Rousseeuw and Croux, 1993; Khalil et al., 2008). It is realized by the application of an odd-lengthen median

filter through time, on each frequency canal (or sub-band) of $\boldsymbol{\gamma}_{\boldsymbol{z}}(k', f)$. The length of the median filter is chosen greater than a useful signal duration in the TF representation to benefit from its outliers smoothing property. Transient signals in filter window are then replaced with the median of neighboring entries (in time), without any effect on the overall amplitude. With an adequate filter length, the filter smooths out the presence of whale calls or short seismic events and provide a TF estimate of the background noise, denoted $|\gamma_{\tilde{\boldsymbol{n}}}(k', f)|$.

Finally, the Wiener-Khintchine theorem (Max, 1981) that links $|\boldsymbol{\gamma}_{\boldsymbol{z}}(k', f)|^2$ and the autocorrelation of the signal is applied, to estimate *online* the noise covariance matrix $\boldsymbol{\Gamma}_{\boldsymbol{n}_0}$. It is used to solve *online* the GEP and calculate the eigenvalues λ_l.

3.5.2.2 Time-varying SNR estimation

The estimation of the current input SNR of the observation $\rho[k]$, is essential for the *online* application of the SMF. This time-varying SNR has a strong impact on the calculation of $Q[k]$ (3.33) giving the number of filters to be applied to the k^{th} sample. The classical definition of the SNR is the power ratio between the signal and the noise (3.9). In practice, as the power of the received signal of interest is unknown, it is usually estimated as the difference between the instantaneous power of the observation and the noise's as in (3.36) (Courmontagne et al., 2010). However, PAM records are noisy (section 2.2) and therefore, there is a need for a more accurate $\rho[k]$ estimation.

The strategy to sharpen $\rho[k]$ estimation when dealing with real underwater noises and decrease the false alarm rate, is again to use the TF representation of the observation. The matrix $|\boldsymbol{\gamma}_{\boldsymbol{z}}(k', f)|$ is compared $|\gamma_{\tilde{\boldsymbol{n}}}(k', f)|$ to provide valuable information of the energetic variations in the known signal's bandwidth. Yet, other signals of no interest might occur in the same bandwidth (other biological sources, distant ship noise or transient noise), so the previous calculation is compared to the nearest frequency bands (Gavrilov and McCauley, 2013).

For ABW call detection investigated in this work, the $\rho[k]$ estimation strategy results in 3 steps, using the previously computed TF estimate of $|\boldsymbol{\gamma}_{\boldsymbol{z}}(k', f)|$ and the noise's $|\gamma_{\tilde{\boldsymbol{n}}}(k', f)|$ (Section 3.5.2.1) described by the following algorithm on each time segment (denoted by k'):

↦ **Step 1:** The signal presence is evaluated in the ABW call unit A frequency band (A = [25.5 – 26.5] Hz) by the absolute ratio between the maximum value of the observation's absolute spectrum and the mean value of the estimated background noise

$$zcall[k'] = \frac{\max\limits_{f}\{|\boldsymbol{\gamma}_{\boldsymbol{z}}(k', f)|\}}{\overline{|\gamma_{\tilde{\boldsymbol{n}}}(k', f)|}}, f \in \text{A} \tag{3.39}$$

A ratio greater than 1, indicates the presence of a short-duration signal in unit A frequency band but does not differentiate signal from "non-signal" short duration events.

↦ **Step 2:** The false alarm due to energetic transient wide band noises is estimated by the ratio between the observation absolute and the estimated background noise, outside the

ABW call frequency band ($f \in [0, 15[\cup]27, f_s/2[$).

$$trans[k'] = \max_f \left| \frac{\Upsilon_z(k', f)}{\Upsilon_{\tilde{n}}(k', f)} \right|. \tag{3.40}$$

↦ **Step 3:** The time-varying SNR $\rho[k]$, is then determined in dB using the ratio:

$$\rho[k'] = 20\log\left(\frac{zcall[k']}{trans[k']}\right). \tag{3.41}$$

To take into account global changes in the acoustic environment, mostly due to continuous sounds, $\rho[k]$ is shifted by a value β such as $\rho[k] \leftarrow \rho[k] - \beta$ for the entire observation duration, depending on the background noise estimation as

$$\beta = \begin{cases} 0 & \text{if } M > 0 \\ M & \text{elsewise,} \end{cases} \tag{3.42}$$

with $M = \frac{1}{L}\sum_{l=1}^{L}(zcall[k] - trans[k])$. To go back to the time domain k, data are interpolated. Only positive values of $\rho[k]$ trigger the reconstruction of the observation (3.33). An application of this improved SMF is presented in section 3.6.

3.6 Results

The SMF described for passive application in section 3.5 is applied for ABW call detection in paragraph 3.6.1[3]. To show other potential applications of such a method in the passive context, the SMF is applied in paragraph 3.6.2 for the detection of high frequency (frequency of the signal of interest > 5 kHz) and less deterministic anthropogenic signals: scuba-divers breathing.

3.6.1 Antarctic blue whale call detection

The two detection strategies, based on the MF and, the SMF adapted to a passive context, are applied to two simultaneous observations of the same ABW call series but, recorded from different locations in order to observe SNR variations. To that extent, the detectors are applied to 10-min-long observations from May 31st, 2013 recorded at 12.30 by OBSs RR43 (Figure 3.4) and RR44 (Figure 3.5). At this moment (Figure 4.1), the whale is estimated $\simeq$ 2 km away from the closest sensor (RR43) and $\simeq$ 22 km away from RR44.

The observation input waveform is bandpass-filtered between 15 and 30 Hz (denoted $\boldsymbol{z}_{[15-30]}(k)$), before the MF application, where the filter is the signal presented in § 3.5.1.1. The SMF is applied directly to $\boldsymbol{z}(k)$. Its output, the estimation of the useful signal $\widetilde{\boldsymbol{s}}_{\mathbf{Q}[k]}(k)$ is, by itself, an audio vector with all its inherent properties. For a fair comparison with the MF, this reconstructed signal is also correlated with the reference signal. This complete process is denoted SMF + MF.

Results on high and low SNR recordings of ABW calls are reported in Figures 3.4 and 3.5: (a) the spectrogram of the observation is plotted, (b) $\boldsymbol{z}_{[15-30]}(k)$ is presented along with the noise-reduced SMF output $\widetilde{\boldsymbol{s}}_{\mathbf{Q}[k]}(k)$. Waveforms are normalized by the maximum absolute value of $\boldsymbol{z}_{15-30}(k)$. The envelop of the MF and SMF + MF result are respectively displayed in (c) and(d). Calls are manually red-marked from (b) to (d).

The observation is composed of 8 ABW calls and two seismic events near 5 and 8.5 min. Time shifts between the two OBSs confirm that the whale is closer to RR43, in accordance with ABW calls estimated SNR presented in Table 3.1. Time shifts also indicate that the seismic event occurs closer to RR44.

Call number	1	2	3	4	5	6	7	8
SNR on RR43	8.1	14.8	14.4	15.0	17.5	17.5	4.5	15.2
SNR on RR44	-	5.9	7.1	2.2	8.9	9.5	7.0	5.0

Table 3.1: ABW call SNRs (dB) estimated according to § 3.5.2.2 on recordings from OBSs RR43 and RR44 on May 31st, 2013 at 12.30.

[3]The developed passive SMF Matlab package is available online: https://leabouffaut.github.io/SMF_package/

3.6.1.1 High SNR observation

Figure 3.4 (b) shows that ABW calls are entirely reconstructed by the SMF. On the waveform, seismic noise seems to be completely filtered out. The second to last call is reconstructed on a shorter duration, probably due to the concurrent noise (also indicated by a smaller SNR). For these high SNR calls, the MF (c) and SMF + MF (d) perform similarly: there is a detection peak of equivalent amplitude for each call. However, the MF output presents a little bump at 5 min that might be related to seismic noise. The amplitude of this bump is no higher than some of the amplitudes of the MF-detected call echoes and, depending on the Ts choice might trigger false positives.

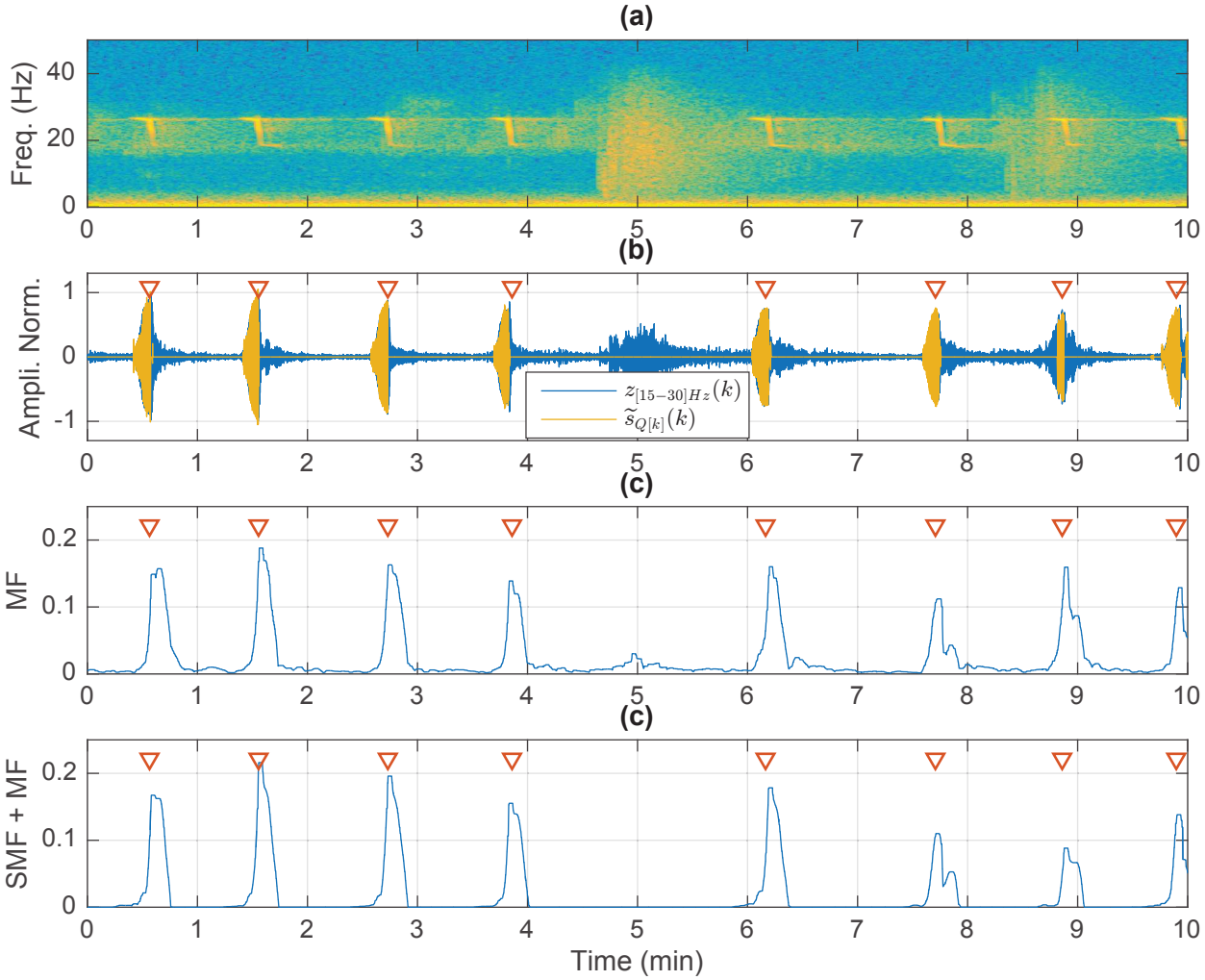

Figure 3.4: Comparison between the SMF + MF and MF on a high SNR recording of ABW calls (a) observation spectrogram (b) observation filtered waveform and output of the SMF (estimated signal) (c) MF (d) SMF + MF.

3.6.1.2 Lower SNR observation

The recording of OBS RR44 depicted in Figure 3.5 (a), shows the same call series but from a further distance. They hence have a lower SNR. On this observation, the most energetic event

is the seismic event at 5 min. Because of the normalization used for the waveform plot, it compresses the ABW calls amplitudes on (b). However, the waveform of the signal reconstructed by the SMF reconstructs only ABW calls. The MF amplitudes (c) are strongly attenuated: the maximum correlation amplitudes on calls decreases from ≃ 0.19 on the close observation to 0.07. One of the strongest amplitudes is due to the first seismic event. It is hard to differentiate the second to last call from the second seismic bump. The SMF + MF is not affected by the seismic events. Despite the distance, the highest peak is in the same order as previously (≃ 0.19). The other peaks are attenuated (lowest peak at 0.9) but are in general greater than the ones of the MF.

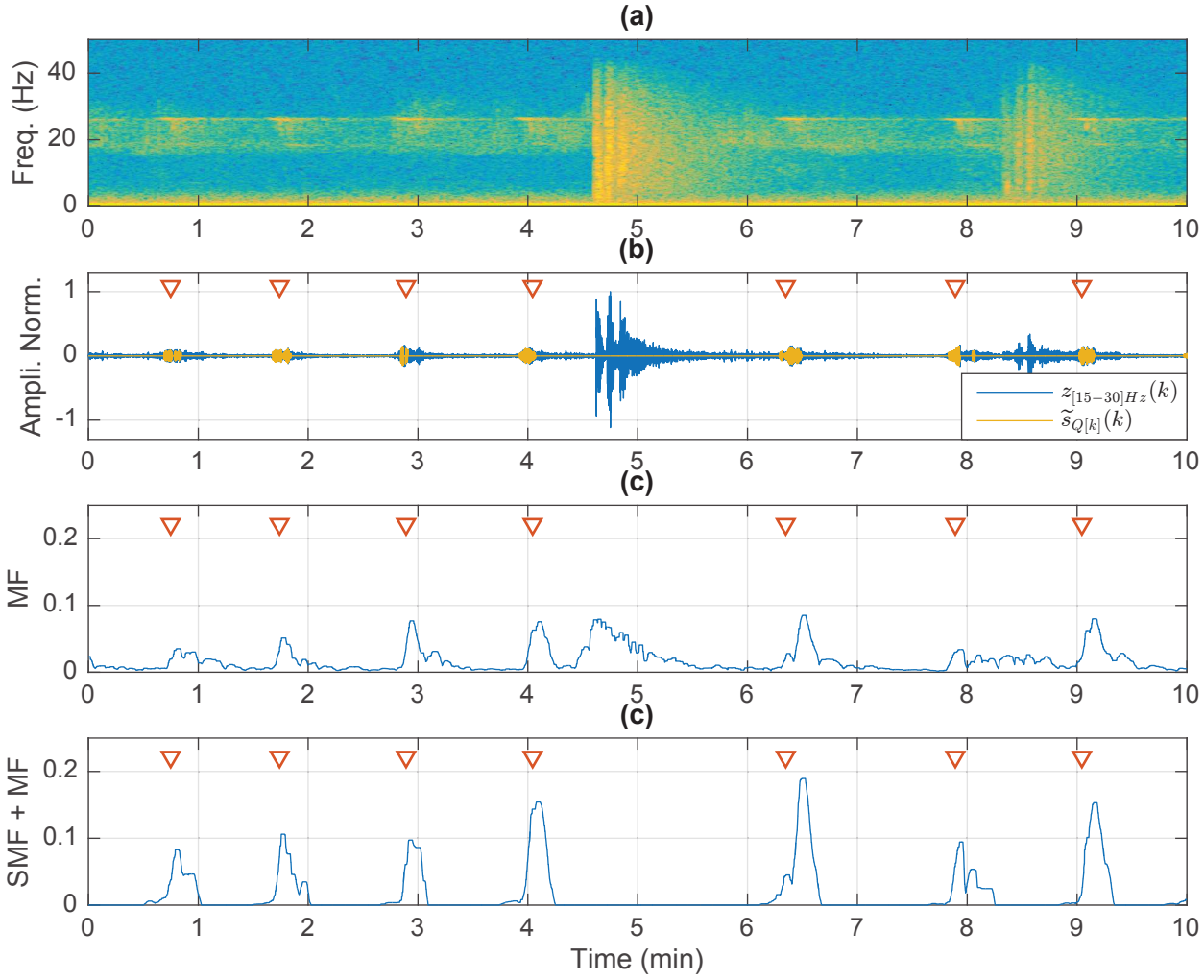

Figure 3.5: Comparison between the SMF + MF and MF on a low SNR recording of ABW calls (a) observation spectrogram (b) observation filtered waveform and output of the SMF (estimated signal) (c) MF (d) SMF + MF.

3.6.1.3 Discussion

Results presented in § 3.6.1.1 and 3.6.1.2 compare the MF and the SMF + MF on high and low SNR recordings of the same series of calls. Both perform similarly on high SNR calls. However, when the SNR decreases, as discussed in paragraph 3.3.2, the MF reaches its limits: the detection

peaks are attenuated in amplitude and spread wider. Besides, the MF is also sensitive to other broadband transient noise that might introduce false detections. The comparison between the two methods identifies the SMF as an excellent asset for the detection of calls with low SNRs. The noise reduction properties clean out the data, and consequently, the SMF + MF drastically reduces the probability of false alarm in comparison with the MF.

3.6.2 Scuba-divers breathing detection

To show the benefits of using the SMF on another type of sounds, paragraph 3.6.2 is dedicated to scuba-diver breathing detection.

There are no reasons for passive acoustics to be limited to environmental applications. It has been used since the cold war by armies for defense purposes and threat detection. Nowadays, new applications such as harbor surveillance and protection emerge, especially for intruders detection such as small speed boats and combat divers. The problem is that the range of passive acoustic surveillance systems remains considerably below other known methods (optics, electromagnetic detection, active sonar...) (Stolkin et al., 2006). To be able to develop efficient devices, enabling instantaneous decision and reaction facing external threats, there is a need for robust and reliable signal processing tools.

Acoustical methods for scuba divers detection mostly exploit periodicity in the breathing using band-pass filtering and envelope detection (Stolkin et al., 2006; Labat and Daré, 2014; Slămnoiu et al., 2017). This method is called DEMON for Detection of Envelope Modulation On Noise. Autonomous diver source level is estimated in the $[1-16]$ kHz frequency band at an average value of 116 dB ref. 1 µPa @ 1m (Slamnoiu et al., 2016). Logically, due to attenuation and noisy harbor conditions in this frequency band, detection ranges are often limited to tens of meters (Lennartsson et al., 2009). In order to expand the detection range and improve detection probability, the SMF, as presented in section 3.5 is applied to scuba-divers breathing detection.

3.6.2.1 Scuba-divers breathing signal and associated filters

During a diver breathing cycle, due to the turbulent decompression of the gas in the high-pressure regulator, inhalation phases are characterized by broadband noises that are louder and more broadband than the bubble noise (Fillinger et al., 2012). In this process, gas relaxation (from pressurized breathing gas to ambient pressure gas) provided by the diving regulator modulates the noises in frequency. Modulation signatures can be found in open and closed circuits (Donskoy et al., 2008). Figure 3.6 presents diver breathing frequency-modulated signals recorded in an acoustic tank using a Reson TC 4034 hydrophone with a sampling frequency of $f_s = 44.1$ kHz. Two energetic frequency bands can be identified at $\simeq 6$ kHz and $\simeq 11$ kHz. This characteristic is used to reproduce scuba-divers acoustic signature (using modulated white noise) and hence, estimate the matrix $\mathbf{\Gamma}_{s_0}$ (Bouffaut et al., 2017a; Bouffaut and Boudraa, 2017).

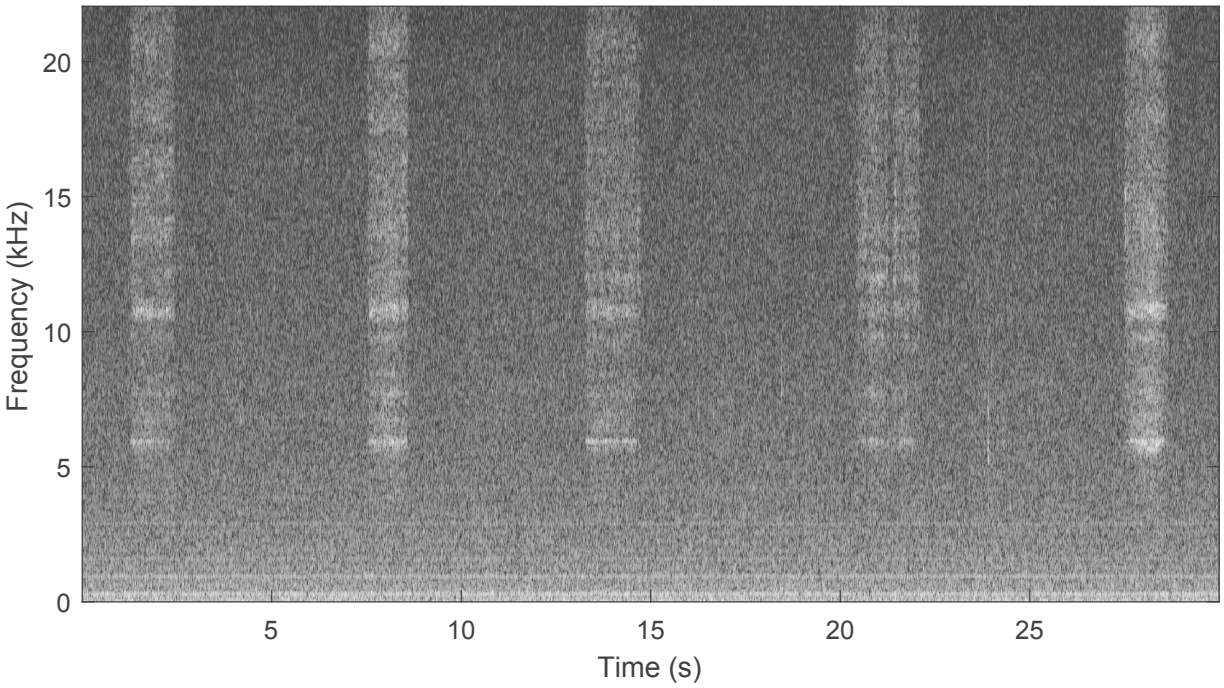

Figure 3.6: Scuba-diver breathing spectrogram recorded in a tank, $f_s = 44.1$ kHz.

The filter bank is then designed as in § 3.5.1.2 for this new signal. Figure 3.7 depicts the frequency response of filters $\boldsymbol{h}_1$, $\boldsymbol{h}_{10}$ and $\boldsymbol{h}_{Qmax}$ against the spectrum of the reference signal HOpt. The first filter H_1 is a short-band filtered centered on ~ 11 kHz. The superposition of the 10 first filters H_{10} is a slightly larger band-pass filter in the same frequency band. H_{Qmax} represents the superposition of the maximum number of filters and is applied when $\rho[k]$ is high enough. Notice that H_{Qmax} covers the same frequency band as HOpt.

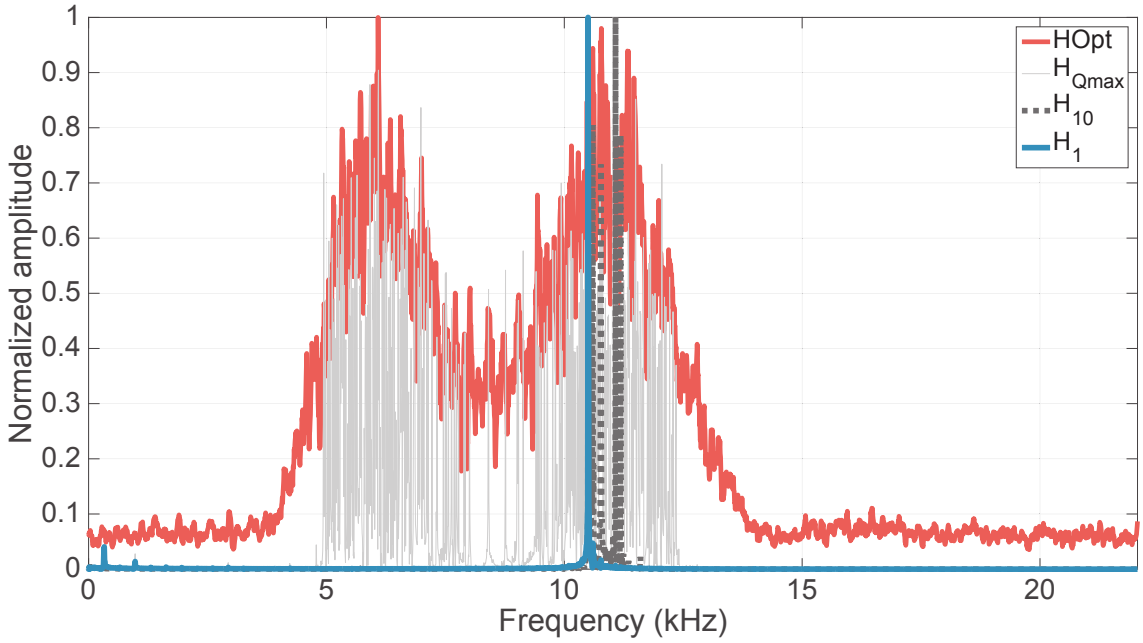

Figure 3.7: Scuba-divers: spectrum of three filters (H_1, H_{10}, H_{Qmax}) of the permanent filter bank compared to the spectral representation of the reference signal (HOpt).

3.6.2.2 Online SNR estimation

The, *online* SNR estimation is the key parameter that determines the number of filters to apply to the observation. It is derived to correspond to scuba-divers signals. For this application, (3.41) becomes

$$\rho[k] = 20\log\left(\frac{1}{\mathrm{J}}\sum_{j=1,\dots,\mathrm{J}}\left|\frac{\max\limits_{f\in \mathrm{A}_j}\gamma_{\boldsymbol{z}}(k',f)}{\frac{1}{\mathrm{K}}\sum\limits_{k'=1}^{\mathrm{K}}\gamma_{\tilde{\boldsymbol{n}}}(k',f')}\right|\right), f'\in \mathrm{A}_j \tag{3.43}$$

with J the number of modulated energy band considered and A_j their frequency interval.

3.6.2.3 Application

The SMF is applied a the recording of two scuba-divers in a swimming pool. Figure 3.8 presents (a) the spectrogram of the recording sampled at $f_s = 96$ kHz, (b) the estimated input SNR $\rho[k]$ and, (c) the observation $\boldsymbol{z}(k)$ and the noise-reduced SMF output $\tilde{\boldsymbol{s}}_{\mathbf{Q}[k]}(k)$ waveforms. Compared

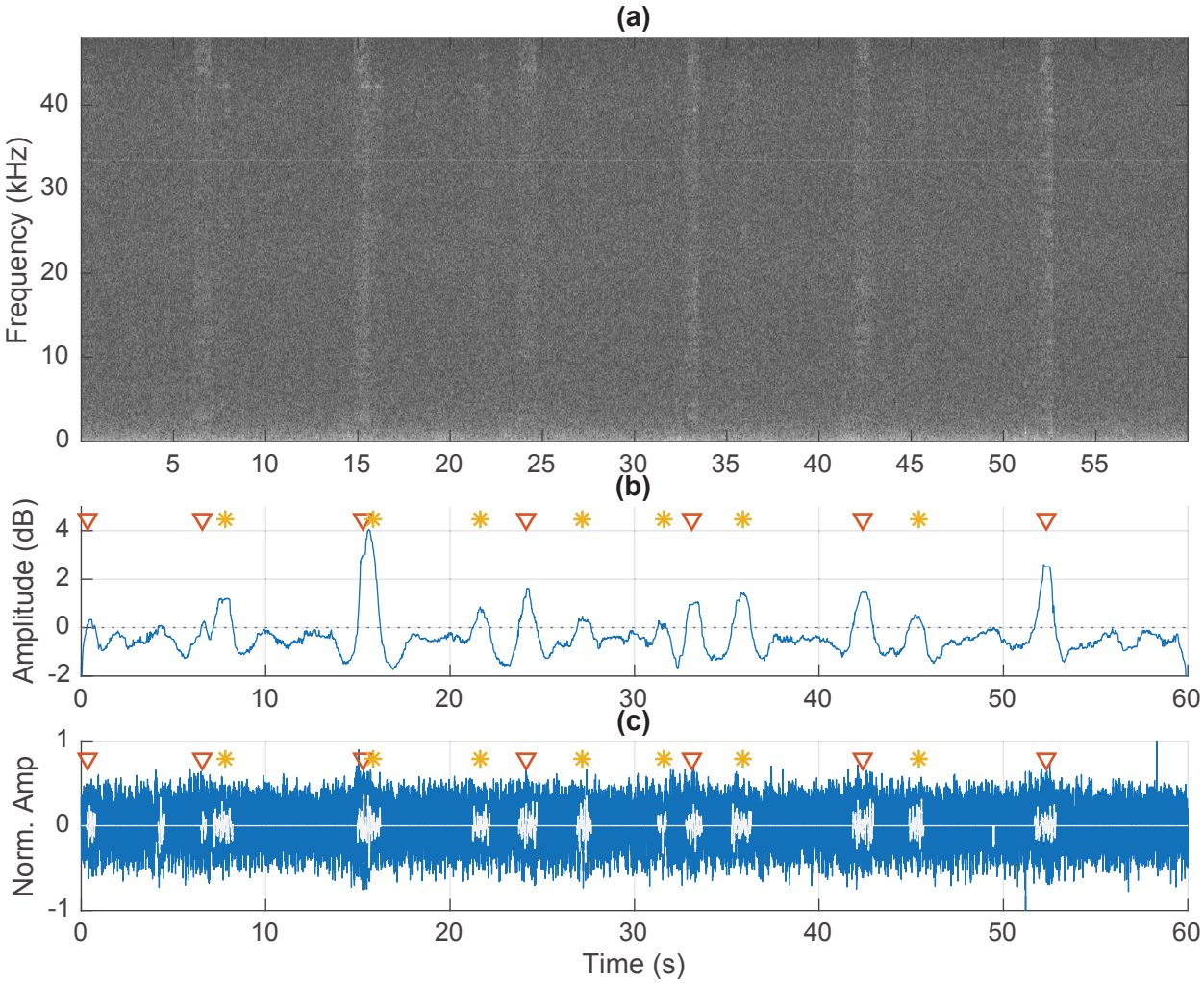

Figure 3.8: Application of the SMF to the recording of two scuba-divers in a swimming pool ($f_s = 96$ kHz). (a) Spectrogram of the observation, (b) Estimated input SNR $\rho[k]$, (c) SMF input (blue) and output (white) waveforms. Annotated divers 1 ▽ and 2 ∗ inhalations.

to Diver 2, Diver 1 is closer to the hydrophone and its breathing stand out in the spectrogram (a). The acoustic presence of the second diver is almost undetectable, only a few inhalations are visible, e.g., at 22 et 26 s. It is possible to notice all breathing from $\rho[k]$ estimation. Comparing $\rho[k]$ to the noise-reduced SMF output $\widetilde{\boldsymbol{s}}_{\mathbf{Q}[k]}(k)$ in (c), highlights that only positive values of $\rho[k]$ trigger signal reconstruction. The noise is strongly attenuated (filtered) for negatives values of the iSNR. It is now possible to recognize two breathing cycles on $\widetilde{\boldsymbol{s}}_{\mathbf{Q}[k]}(k)$.

The SMF applied to recordings taken in the swimming pool, extended the detection range from 20 m (on the spectrogram) to 25 m, which was the maximum diver-hydrophone distance possible in the configuration. The same strategy was applied to noisy observations (sea and loud electrical noises), recorded at sea in front of the Ecole Navale. Visually (on the spectrogram) inhalations are visible up to $\simeq 3$ m. The SMF extended the detection range up to 20 m.

3.6.2.4 Discussion

This new application demonstrates the efficiency of the SMF to detect signals, different from the class of whale calls: scuba-divers inhalations. These two types of signals are different in terms of structure and frequency bands: whale calls are ultra-low frequency deterministic signals (§ 2.3) whereas scuba-divers inhalations are characterized as high-frequency modulated noise. Due to that aspect, divers inhalations can not be simulated from varying instantaneous phase and, clean signals can not be used as templates. Their stochastic properties do not satisfy classical MFs assumptions (§ 3.3). However, second-order statistics can be estimated based on the modulation in the signature: the SMF can, therefore, be applied.

Additional experiments should be performed to assess the detection range reached by the SMF and compared to the DEMON (as well as both combined). However, the SMF remains limited by the *a priori* knowledge on the modulated frequency bands and, scuba divers signatures have been proven to vary with the equipment (regulator) (Donskoy et al., 2008) and might be sensitive to variations in the volume of liquid air available in the bottles.

3.7 Conclusion

As a conclusion, Figure 3.9 summarizes the extended version of the SMF for SNR maximization and detection in a passive context. The SMF, originally based on the MF, is based on the second-order statistics of the signal and the noise. It can be seen as a time-varying linear filter adjusted in accordance with the estimation of the input SNR. The processing can be separated into two steps. First, the offline processing only requires the reduced covariance matrix of the signal and simulated sea-colored noise to compute the filter bank. Then, during the online application of the SMF to the observation, the background noise reduced covariance matrix, and the time-varying SNR are estimated. They are based on the observation's median-filtered TF analysis and, bypass the selection by an operator.

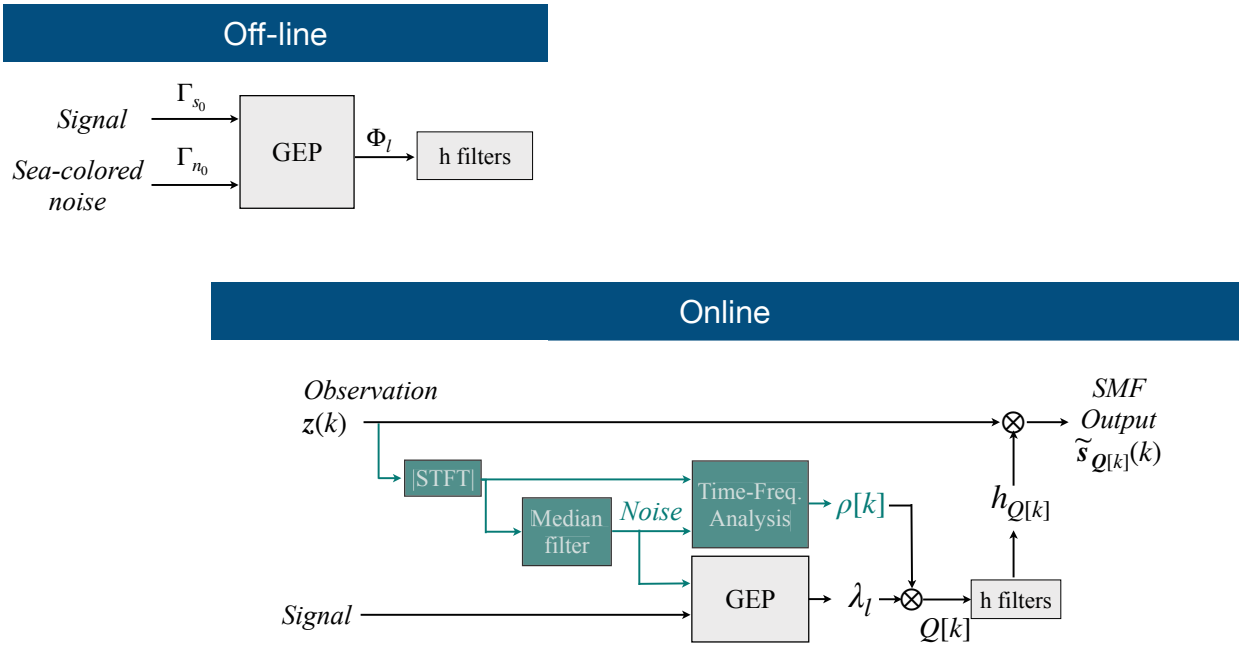

Figure 3.9: Scheme of the SMF improved for a passive application.

The SMF has been applied to two different types of signals. Results obtained demonstrated its ability to detect low-SNR ABW calls and, showed greater robustness to impulsive noise and false alarm reduction than the MF. In addition to call detection and counting, the output of the filter can be used for subsidiary analysis, such as measurement of similarities between calls or, determine call-time of arrivals, providing new opportunities for automatic source localization and whale tracking. The application to scuba divers breathing detection highlighted the ability of the SMF of detecting other types of signals (other than whale calls), in a different frequency band. Results also demonstrated how the SMF could be used to extend detection ranges.

This first and instinctive quantification of the quality of detectors with respect to, e.g., detection thresholds or false detections, introduces the need of a more robust and systematic analysis of methods' range of action and limitations. These types of analysis are grouped under the label performance analysis and are the subject of Chapter 4.

Chapter 4

Performance analysis of stereotyped sounds detectors

Contents

4.1 The performance analysis dilemma

Passive detection algorithm performance assessment is a tricky subject (Leroy et al., 2017b). The formal expression of a detector's performances are derived from information on the probability of false alarm and the probability of *true* detection, when they are available (Van Trees, 2002). However, in the absence of this knowledge, the theoretical analysis of the performances can not be easily carried out. A solution to this problem is to resort to ad-hoc approaches such as Monte-Carlo simulations. They rely on an extensive computation over an artificial dataset. Monte-Carlo simulations have been conducted for the classic SMF in numerous publications (Courmontagne, 2010; Bénard et al., 2011; Mori and Gounon, 2000). The problem is that performance simulations might not be representative of the method's robustness to noise in real conditions. The third option is to confront the algorithm to a ground truth dataset (Mellinger and Clark, 2000; Socheleau et al., 2015; Gavrilov and McCauley, 2013), keeping in mind that the annotation process has been shown to be subjective and quite variable, especially in low SNR conditions (Leroy et al., 2017b) and, that in a passive context, there are no *a priori* knowledge of the contents of the recordings. The choice was made to use additional information from multi-sensor observation and whale localization to reduce the subjectivity of the annotation process and evaluate the SMF's performances on a robust ground-truth dataset. The SMF's performances are assessed on real marine signals by the comparison of the method output with human inspected and annotated data (Section 4.2). Results are compared to the commonly used temporal MF (§ 3.3) and to the Z-detector presented in Socheleau et al. (2015).

4.2 Groundtruth dataset context

The localization method (§ 2.4.2) is applied to track a calling ABW swimming through the SWIR array on May 31st, 2013 (Fig. 4.1). This individual's song is recorded continuously on the array, for more than 21 hours (from 01:20 to 22:40 on OBS RR48). OBS RR48 is the farthest from the trajectory. It is therefore chosen for annotations to provide the widest SNR range, with the use additional information from multi-sensor observation and whale localization to reduce the subjectivity of the annotation process (Leroy et al., 2017b; Bouffaut et al., 2018)[1].

This dataset, is corrupted with different noise sources (Figure 4.2 (a)):

- tonal noise radiated from a close ship between 01:20 and 08:00,
- P-calls from 20:10 to the end,
- D-calls from 10:40 to the end,
- a continuous chorus all-day, strong in the FW frequency range before 08:00,
- more than 50 seismic events.

[1]The annotated dataset, SMF + MF detections and corresponding whale locations is available online, DOI: 10.5281/zenodo.3624145

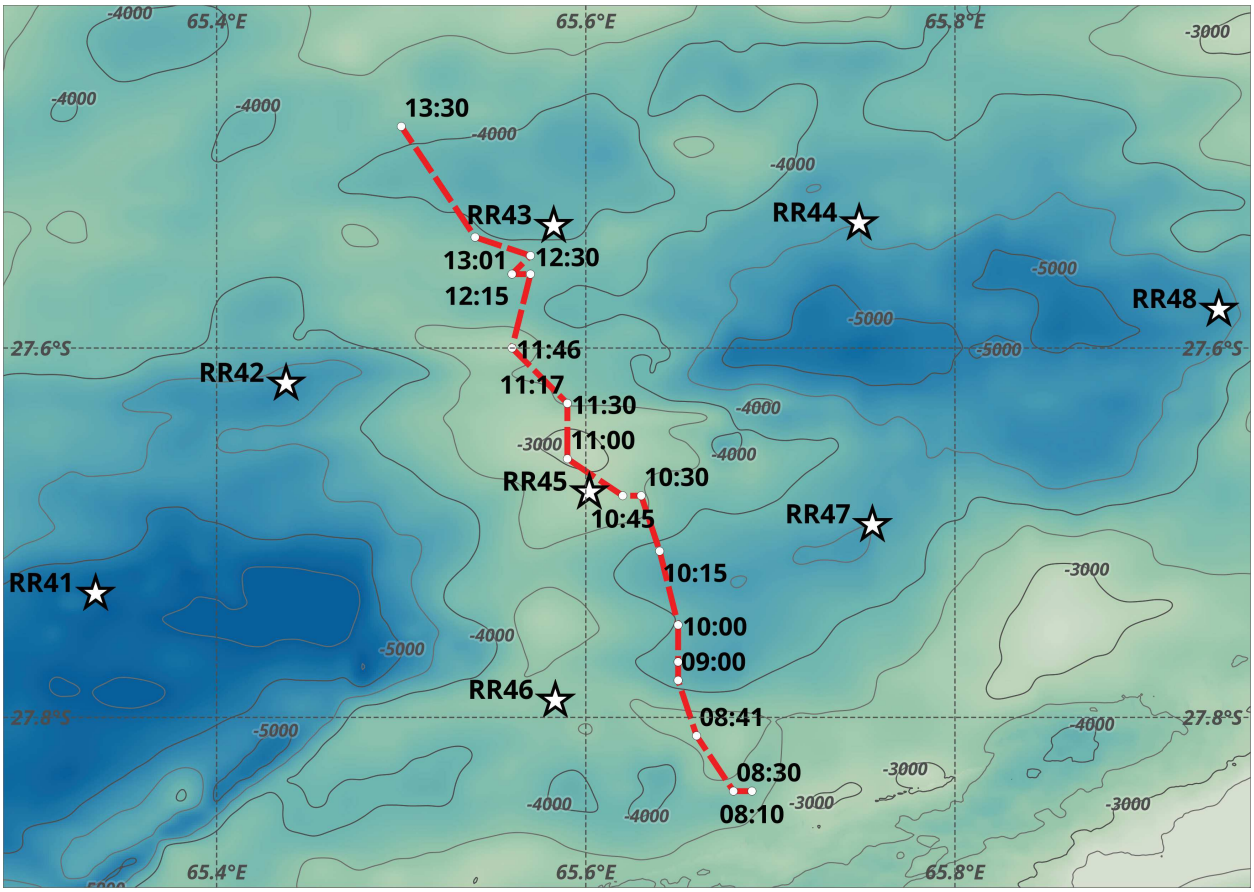

Figure 4.1: ABW passive acoustic monitoring tracking through the SWIR array. Star denotes OBSs locations (Dréo et al., 2019).

Considering an average swim speed of 10 km/h, the whale trajectory is extrapolated (linear South-East to North-West trajectory relative to the SWIR array) as range indicator (Figure 4.2 (b)). Form the extrapolation, at 1:30, the ABW is estimated about 100 km away in the South direction from OBS RR48 ($\overline{\rho[k]} \simeq 2$ dB) and 130 km North West at 22:30 ($\overline{\rho[k]} \simeq 3$ dB). The closest point of approach at 10:30 is 35 km away from the sensor ($\overline{\rho[k]} \simeq 13.4$ dB).

Along this trajectory and recording period, 845 ABW calls are annotated. Calls SNRs are estimated with the method described in § 3.5.2.2. The dataset shows significant SNR variations from the first contact to the last (Figure 4.2 (c)), that mainly follow a $-20\log_{10}(r)$ tendency. It appears that the $-20\log_{10}(r)$ fitting is slightly shifted in time and should have a maximum near 12:00. One of the possible explanations is that this estimation of the whale's location ±5 km (the difference between the distances estimated at 10:30 and 12:00) is within the localization margin of error. The trajectory is assimilated to a straight line, the best resolution of the localization method is 900 m (grid definition § 2.4.2) and uncertainties on the OBS location is ±500 m (§ 2.1.1). The other possible explanation could be, changes in the background noise delaying the maximum measured SNR: in the first hours of recording, FW and ship noise increases ambient noise levels in the frequency band of the call impacting the SNR. Variability in SNRs measurement methods is discussed in § 4.6.1.

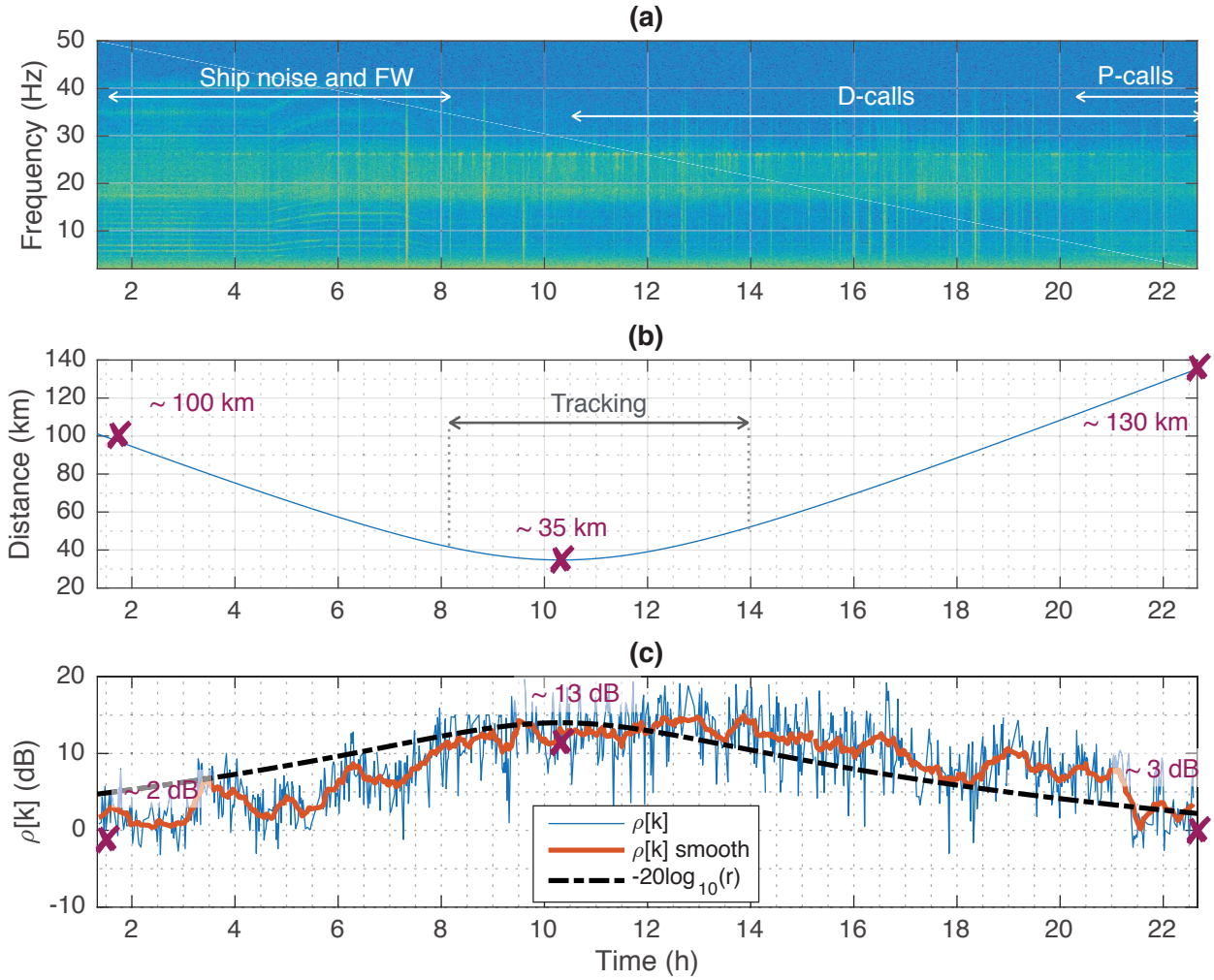

Figure 4.2: Temporal evolution of (a) the soundscape (nfft = 2048, overlap = 95%), (b) estimated and extrapolated ABW-OBS RR48 distance, (c) SNR variations (ρ_k estimated according to § 3.5.2.2) relative to the ABW of May 31st, 2013 dataset.

4.3 Method

4.3.1 General experiment setup

This day-long record is divided by portions of 80 min for the TF analysis. As described in section 3.6.1, the signal reference for both the SMF and the Z-detector is similar. This signal is also used for inter-correlation with $z_{[15-30]}(k)$ for the MF and on the SMF-noise-reduced cleaned data (SMF + MF) for similar comparison. It is a 20 s-long signal of 2001 samples.

For the SMF, $|\boldsymbol{\gamma}_z(k', f)|$ is evaluated using Hann window of 2048 samples, with 98% overlapping. The median filter applied to each frequency canal for the background noise estimation has 201 samples (Section 3.5.2.1). Its equivalent duration is 83 s, which is four times longer than an ABW call. After retrieving real-time k (Section 3.5.2.2), the size of the sliding window, used to assume the stationarity of the noise during the online application of the SMF, is set to match the signal with L = 2001 samples.

Before applying the Z-detector, the observation is down-sampled to a base-band signal for better computation time. The Ts of the Z-detector automatically adapts to deal with noise variations in the environment (e.g. transient signal) and authorizes ABW call variations in frequency, amplitude or duration (Socheleau et al., 2015; Leroy et al., 2016). The Z-detector also has an additional threshold (denoted τ in Socheleau et al. (2015)), used to separate the signal and the noise. It is fixed to 0.15% and, the worst-case user-defined probability of false alarm is fixed to 3%, as in Socheleau et al. (2015).

For each annotated call, the maximum of the SMF $\rho[k]$ estimation (Section 3.5.2.2) is measured and, an arbitrary confidence index is assigned by the operator. The index reveals to be following SNR variations (the lower the SNR, the harder it is to classify an event). Outcomes of the detection methods are automatically checked. The performance analysis is carried out with the help of different standard scoring metrics presented in § 4.3.2.

4.3.2 Scoring metrics

In order to evaluate the detection performances of the MF, SMF + MF and the Z-detector, outputs are compared with the groundtruth annotations and, according to the detection theory (Van Trees, 2002), they are sorted into four classes (Table 4.1).

		Groundtruth	
		signal present	signal absent
Detection	signal present	**true positive**	false positive
	signal absent	false negative	**true negative**

Table 4.1: Detection theory (Van Trees, 2002).

The number of detections corresponding to each class is counted. In this context true negative is a silent class. Performances are then evaluated as rates:

- the true positive rate (TPR) (also called recall, sensitivity or detection probability depending on the field) indicates the rate of detected true calls overall annotations, i.e., how many relevant items are detected

$$\text{TPR} = \frac{\text{nb. of true positives}}{\text{nb. of true positives} + \text{nb. of false negatives}}; \tag{4.1}$$

- the missed detection rate (MDR) (also called miss rate or false negative rate) indicates the rate of true calls that are not detected

$$\text{MDR} = \frac{\text{nb. of false negatives}}{\text{nb. of true positives} + \text{nb. of false negatives}}; \tag{4.2}$$

- the positive predictive values (PPV) (also called precision) indicates the rate of detected true calls overall detections, i.e., how many detected items are relevant

$$\text{PPV} = \frac{\text{nb. of true positives}}{\text{nb. of true positives} + \text{nb. of false positives}}; \tag{4.3}$$

- the false discovery rate (FDR) indicates the rate of detections that are not calls

$$\text{FDR} = \frac{\text{nb. of false positives}}{\text{nb. of true positives} + \text{nb. of false positives}}. \tag{4.4}$$

Note that TPR + MDR = 1 and, PPV + FDR = 1.

4.4 Performances versus threshold

The first performance analysis is completed by characterizing the detection outputs for different Ts varying from 0 to 0.1 of correlation amplitude with an increment of 0.001. The choice of the threshold is indeed essential to adjust the method for the intended application. If set too low, it might not be able to differentiate true positives from false positives and, set too high, it might miss detections (false negatives). Figure 4.3 illustrates the threshold limitations for low SNR data recorded simultaneously as data presented in section 3.6.1 but by OBS RR47.

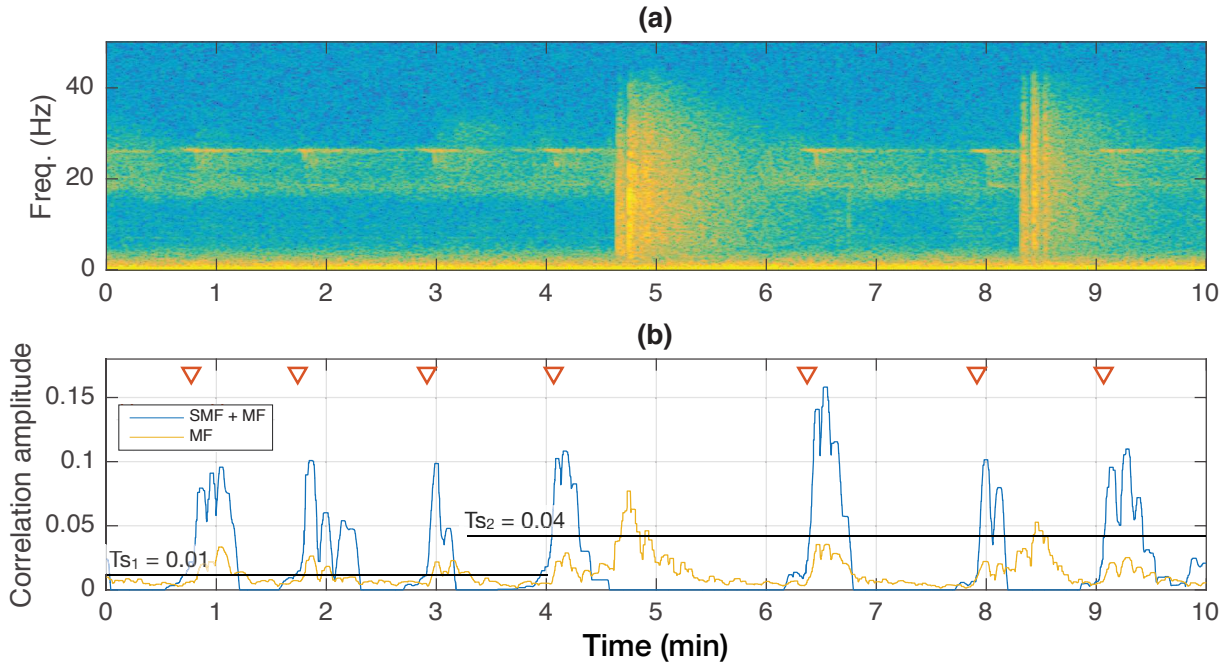

Figure 4.3: Illustration of detection threshold limits on an observation from May 31st, 2013 at 12.30, recorded by OBS RR47.

The threshold $Ts_1 = 0.01$ seems adequate for both detectors, it differentiates peaks from background line. Because SMF + MF peaks are higher, the second threshold $Ts_2 = 0.04$ is also satisfactory. However, Ts_2 is not appropriate for the MF: only the seismic events have amplitude above this value. This threshold is too high to provide true positives and only generate false positives.

4.4.1 Scores depending on the threshold

In order to compare the SMF + MF and the MF, Figure 4.4 presents the evolution of the supplementary TPR and MDR along with the FDR against different threshold values. Overall observation of the performances curves highlights that increasing the threshold value reduces the FDR faster than it decreases the TPR. For a given threshold value, the SMF+MF generally performs better than the MF. Below Ts = 0.005, the MF reaches higher TPR than the SMF + MF but with 2 to 4 times its FDR. This higher MF TPR might be due to the concurrence of some ABW calls and seismic events, impacting $\boldsymbol{\rho}(k)$ estimation (§ 3.5.2.2) and hence leading to a SMF false negative. However, for all other threshold values, the SMF + MF TPR is always greater than the MF's.

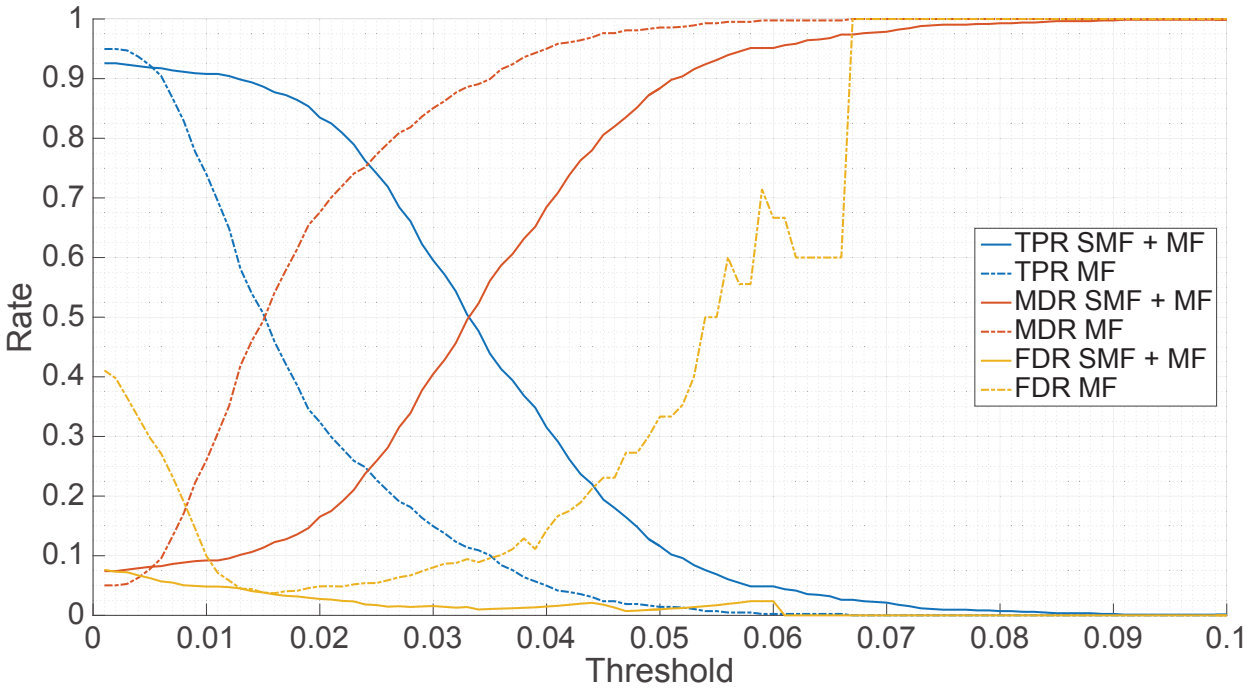

Figure 4.4: TPR, MDR and FDR of the MF and the SMF + MF as a function of the detection threshold.

The SMF + MF FDR is always lower than the MF's , except for Ts = [0.013 - 0.015] where they perform similarly. For this threshold range, SMF + MF and MF TPRs are respectively [90 - 89]% and [58 - 50]%, therefore the first method still performs better. The maximum threshold delimiting this equality (Ts = 0.015) marks the beginning of the invalidity of the MF: the FDR increases again, corresponding to the limit presented in Figure 4.3. Therefore, in § 4.4.2, MF performances are only considered for thresholds between 0 and 0.015.

4.4.2 Receiver operating characteristics

To compare directly TPRs and FDRs (or PPVs and TPRs) independently from the threshold, receiver operating characteristics (ROC) and Precision-Recall curves are displayed for the MF and the SMF + MF respectively in Figures 4.5 and 4.6. The Z-detector performances on the ABW

dataset are also indicated on both figures for comparison.

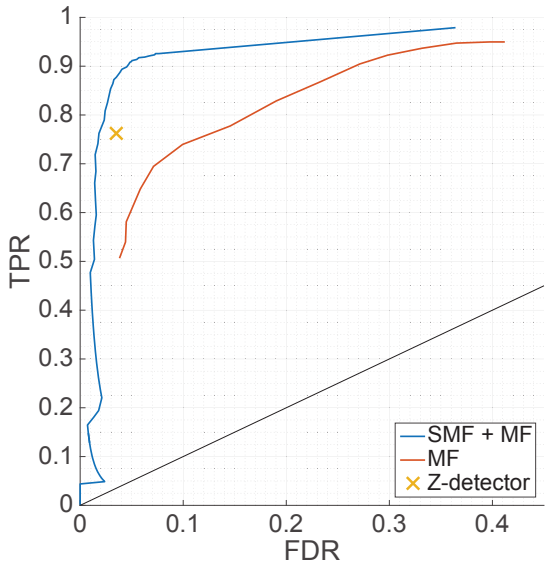

Figure 4.5: ROC comparison between the MF, the SMF + MF and the Z-detector.

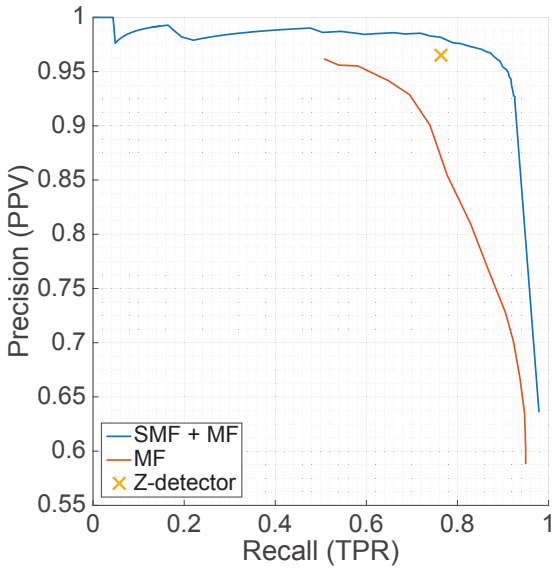

Figure 4.6: Precision-Recall comparison between the MF, the SMF + MF and the Z-detector.

The interpretation of a ROC curve is simple; the ideal detector should have a TPR as high as possible for, at the same time a low FDR (ideally as close as possible to the upper left corner). The ROC curve of a detector should be greater than the TPR = FDR line, that marks a random outcome of the detector. Same types of observations are applied to Precision-Recall graphs, PPV > 0.5, and the curve should approach the upper right corner. On both representations, the SMF + MF shows the best performances. The Z-detector point, shows a FDR of approximately 3.5%, which is consistent with its settings (§ 4.3.1 (Socheleau et al., 2015)). The MF curve for Ts = [0 - 0.015], shows lower performances, but they are still over the random line.

4.4.3 Discussion

Results presented in this section (§ 4.4) underline the performances of each detector against a varying threshold. This threshold is crucial to provide reliable detection performances. However, to perform the analysis on a significant number of calls and provide a robust performance analysis, this study uses the entire dataset, regardless of the SNR. On the one hand, it means that the results are valid for any SNRs in the dataset. On the other hand, it does not indicate the performances of each detector for a specific SNR range. Evaluation of the response of methods to SNR changes is the purpose of the second part of the analysis of the performances and, is investigated in section 4.5.

4.5 Performances against the SNR

The evaluation of the performances of the MF, SMF + MF and the Z-detector against SNR variations is also carried out on the dataset presented in § 4.2. Detection thresholds are then

fixed, according to the results of section 4.4, to Ts = 0.016 Ts = 0.01 and Ts = 0.005 for the SMF + MF and, to Ts= 0.01 for the MF. According to Figure 4.4, same FDR than the Z-detector (of 3.5 %) is reached by the SMF + MF and the MF at a Ts= 0.016.

4.5.1 True positive rate against the SNR

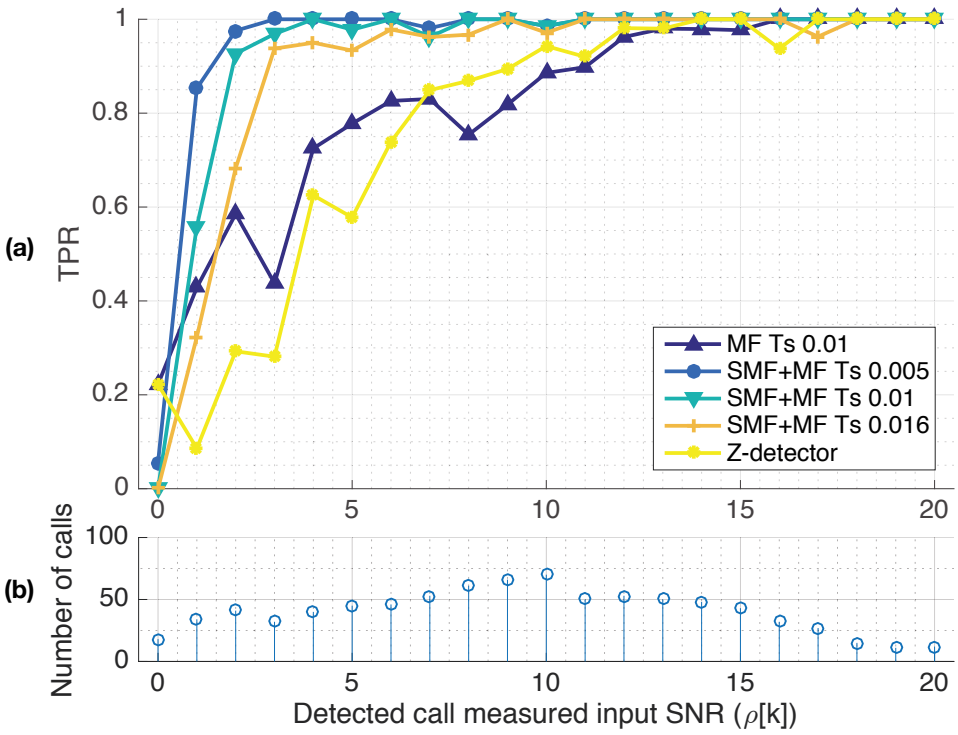

Figure 4.7: Comparative performance analysis between the MF Ts = 0.01, the SMF +MF Ts = {0.005; 0.01; 0.016} and, the Z-detector on real data. (a) TPR against measured SNR applied on the correlation output. (b) Stacked bar representation of the number of annotated calls per SNR for TPR estimation.

Figure 4.7 (a) displays detection results, as TPR against the SNR for the MF Ts = 0.01, the SMF + MF Ts = 0.005, 0.01 and 0.016 and, the Z-detector. Figure 4.7 (b) bar representation, highlights the number of annotated calls on which the TPR is estimated. The higher the number of annotated calls the more reliable the estimate of the TPR is.

TPR reaches one at a lower SNR for the SMF +MF ($\simeq$ 3 dB, $\simeq$ 4 dB, and $\simeq$ 9 dB respectively for Ts = 0.005, Ts = 0.01 and Ts = 0.016) than for the Z-detector (14 dB) or the MF ($\simeq$ 16 dB).

4.5.2 Performances evolution during the day

Figure 4.8 (a) highlights the correspondence between May 31st, 2013 timing and SNR variations introduced by the whale movement and background noise evolution. Figure 4.8 (b) presents the estimated TPR, (c) FDR and (d) MDR per 80 min portions of observation for the three compared methods, the MF Ts = 0.01, the SMF + MF Ts = 0.005, Ts = 0.01 and Ts = 0.016 and, the Z-detector.

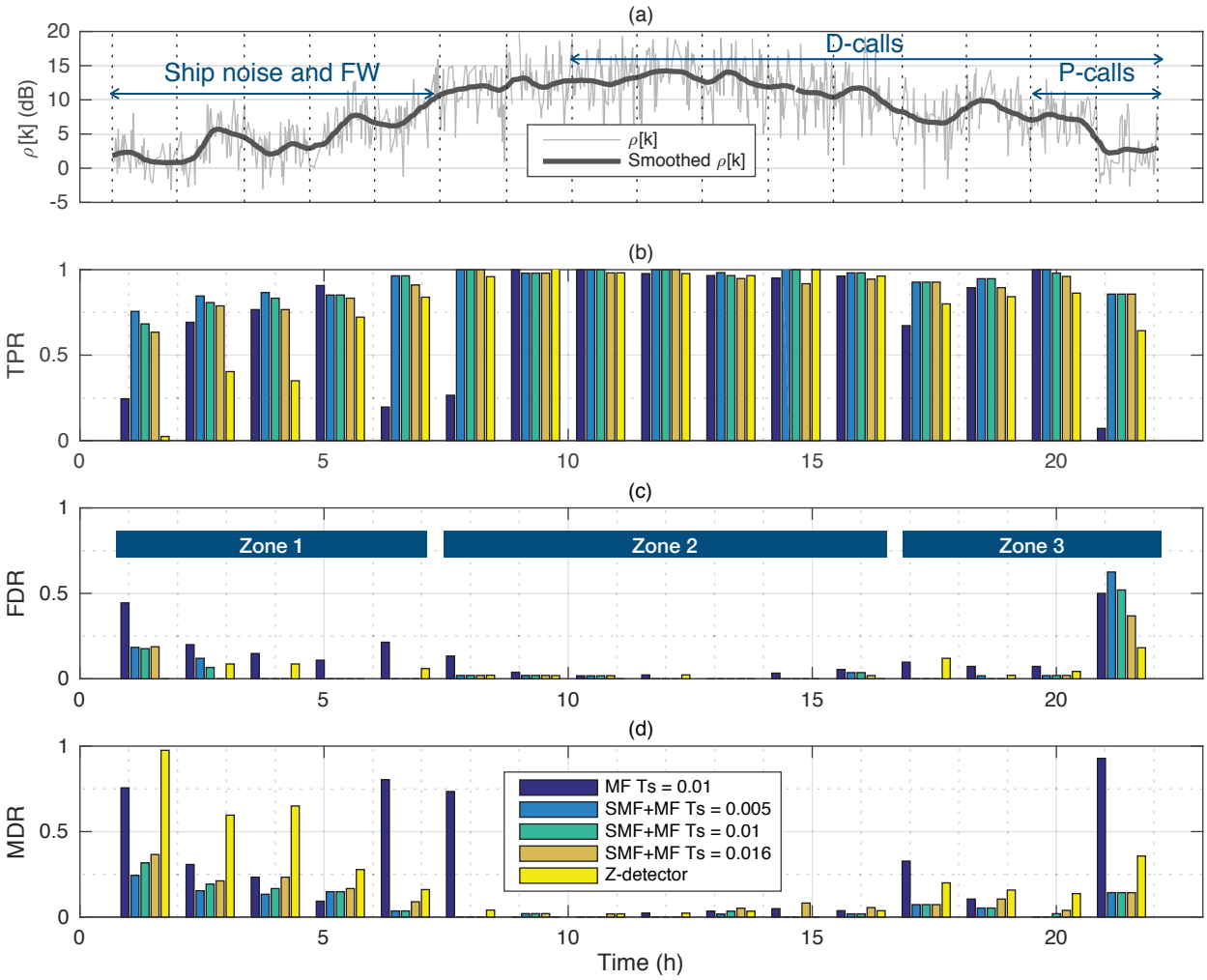

Figure 4.8: Time-dependent compared performance analysis of MF, SMF + MF and the Z-detector on a ground truth dataset of 845 annotated calls, on May 31st, 2013. (a) Detected call measured input SNR $\rho[k]$ (Section 3.5.2.2). Performance criterion are measured on 80 min portions of observation containing on average 53 calls (min. 43-max. 62). (b) TPR (c) FDR (d) MDR.

First, the analysis of Figure 4.8 shows that, for all methods, the TPR increases with the measured input SNR $\rho[k]$, i.e., when the whale-sensor distance is shorter, while the FDR and the MDR decrease. This first observation agrees with the results presented in Figure 4.7 and with general understanding of the detection theory (if the SNR increases, so should the TPR). The MF has higher FDR than the other two methods, independently of the other method thresholds, except for the last portion of observation (addressed in the following paragraphs). The Z-detector has higher MDR (and therefore lower TPR) than the SMF + MF, particularly when ship noise and fin whale pulses are present in the recordings, before 08:00. The MF also presents high MDR which is non-consistent with the SNR variations. More specific analysis of the data is conducted in the following paragraphs, separating the day in three zones: before 08:00, between 08:00 and 16:00 and after 16:00.

In zone 1, before 08:00, $\rho[k]$ is < 10 dB and ship noise is present. For the first three portions of observation, the best TPR is reached for the SMF + MF for all thresholds, then for the MF. The

smallest TPR is obtained by the Z-detector. On the 4th portion starting at 05:20, TPR of the MF is greater than the other two methods: the SMF + MF misses a detection occurring at the same time as a seismic event. In this portion, the Z-detector reaches TPR > 0.5. For the observation starting at 06:40 (the last of zone 1), TPR of the MF is low. It might be due to changes in the background noise when the ship noise fades away while the whale calls are still quite low: the detection threshold is not adapted.

The SMF +MF false alarms, as well as some of the MF's in the first two windows of analysis, might be triggered by a continuous remaining signal in the unit A band. This signal probably comes from distant calls (§ 2.3.5) but, could not be strictly assigned to one. The other MF false alarms were mostly due to seismic events.

In the second zone, between 08:00 and 16:00, the $\rho[k]$ is > 10 dB and the three methods perform similarly. However, in the first portion, the MF behaves similarly to the last portion of zone 1, likely for the same reasons. There are only a few missed detections probably due to the simultaneous occurrence of ABW calls and short larger band noise (fin whale pulse or seismic event).

After 16:00 in the third zone, $\rho[k]$ decreases to values lower than 10 dB. For the first three portions, MDRs raise as TPRs decrease but, the methods FDRs are still lower than before 08:00 probably because there are less continuous sound-sources. However, TPR and MDR results are similar to the ones before 08:00. For the last observed portion, FDRs are high for all methods, but still with TPR > 0.5. At that moment, the signal is faint and, as for the firsts sections of zone 1, it is hard to differentiate a call from the reverberated unit A in the annotation process. A closer look at the results shows that the MF is only triggered by a call and a seismic event, hence the FDR = 0.5. However, the SMF still shows patterns of $\simeq 60-70$ s spaced detections, that could be attributed to ABW but are classified as false detections.

Zones 1 and 3 are comparable in terms of SNR. However, the whale is estimated to be 65 km away from OBS RR48 at 05:00 and 130 km away at 22:00. Due to the ship noise present at the beginning of the record, $\rho[k]$ is of the same order (Section 3.5.2.2). Those too low or negative estimations of $\rho[k]$ impact not only the SMF +MF but also the other methods by increasing the number of missed detections.

4.6 Discussion

4.6.1 SNR estimation

It is important to give some perspectives to the analysis of the comparison between the SMF + MF and the Z-detector. Although the SNR definition remains the power ratio between the signal and the noise in the data, its estimation differs between the SMF and the Z-detector.

As presented in Section 3.5.2.2, the SMF's input SNR $\rho[k]$ is measured continuously in the TF domain, with a parameter to prevent positive values on non-Z-call events (and therefore avoid reconstruction of noise events). This definition is set to allow positive evaluation of $\boldsymbol{\rho}(k)$ (3.33) even in the presence of uncompleted calls. For the Z-detector, a base-band representation of the observation is used (15 Hz bandwidth centered around 22.5 Hz). The input SNR is then estimated from energy variations and the diagonal values of the noise covariance matrix (Socheleau et al., 2015): it can be negative.

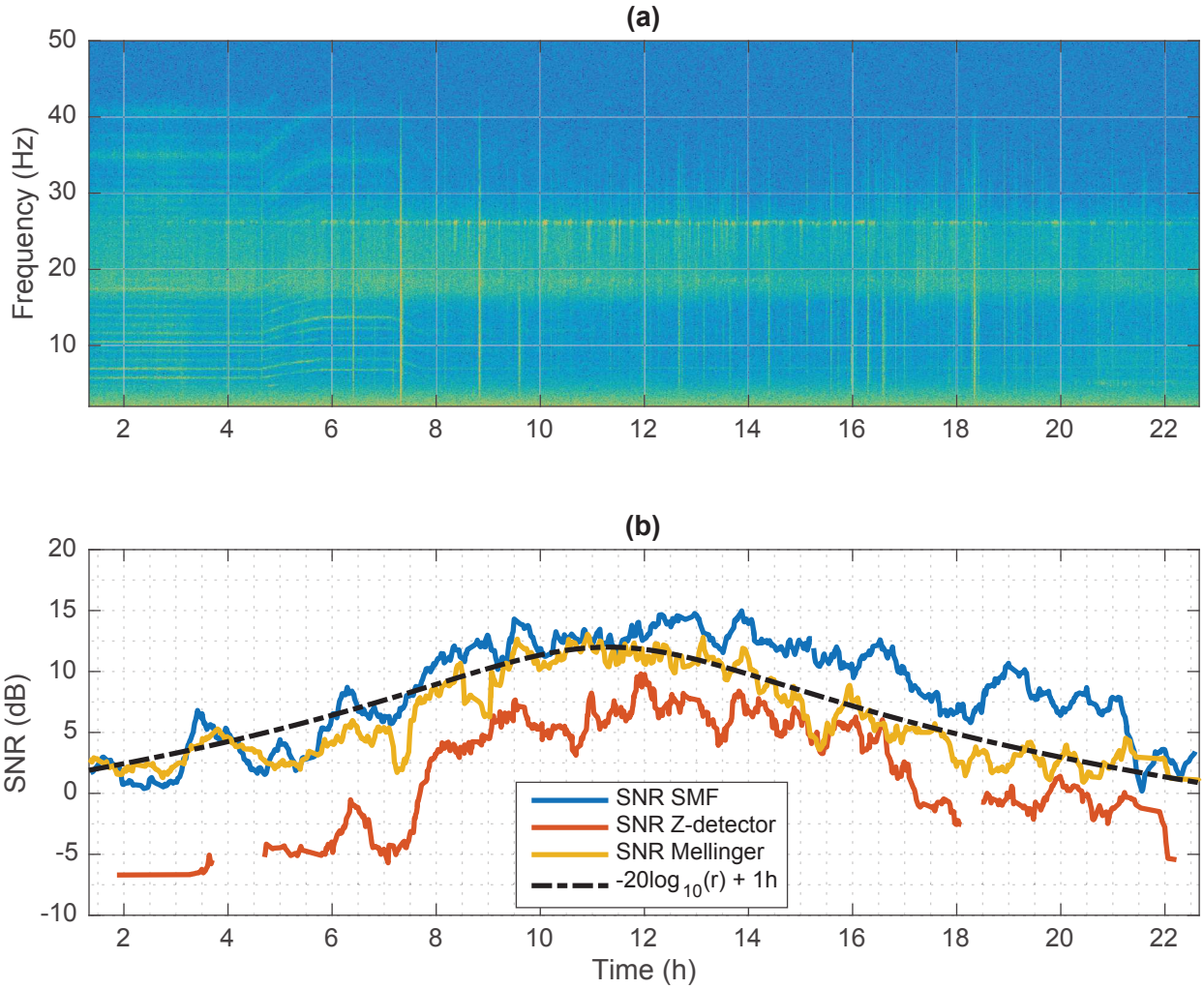

Figure 4.9: Comparison of different SNR estimation methods: SMF (Section 3.5.2.2), Z-detector (Socheleau et al., 2015) and, Mellinger's (Mellinger and Clark, 2006)) on the ABW of May 31^{st}, 2013 dataset.

To better understand differences between the two SNR estimation methods, a comparison is carried out with the "standard" SNR used in the PAM community (Mellinger and Clark, 2006). Results are illustrated in Figure 4.9, on the ABW of May 31^{st} dataset. The method presented in Mellinger and Clark (2006) is based on the annotations of the dataset, where a call power is measured within the indicated time and frequency bounds and, the noise power is estimated in the times between adjacent calls. Here, powers are strictly measured between 26 and 27 Hz. SNRs are smoothed using a Savitzky-Golay finite impulse response smoothing filter.

Estimated SNRs evolution is consistent with the whale-sensor distance and follow a $-20\log_{10}(r)$ (Figure 4.2, shifted by + 1 hour to correspond to the ground-truth, the Mellinger and Clark (2006) method. It appears that the SMF estimation of $\boldsymbol{\rho}(k)$ is similar to the ground-truth, especially in the presence of ship noise, before 08:00. After 12:00 $\boldsymbol{\rho}(k)$ is between 5 and 7 dB higher than the ground-truth, and returns to similar values after 21:00. On the contrary, the Z-detector method mostly underestimates the SNR, with value consistently 5 to 7 dB below the ground-truth estimation. They are closer when the ground-truth shows a "drop": at 09:00 due to a seismic event and between 15:00 and 16:00. This analysis highlights the variability of SNR estimation in the passive context but also validates $\boldsymbol{\rho}(k)$ as a robust SNR estimator that does not require any *a priori* information.

4.6.2 Detection range

Results presented in Figure 4.8 and knowledge of the whale-OBS distance from the tracking (paragraph 4.2) are used to estimate the detection range at a fixed TPR threshold. With a TPR> 75%, Table 4.2 presents the estimated detection ranges for each method. Ranges are measured at the extreme time of each 80 min analysis window and are rounded down (to the tens). In the absence of ship noise, detection ranges increase by at lesr 30 km. They are even

	Detection range (km)	
Method	with boat noise	without boat noise
MF Ts = 0.005	85 *	120 *
SMF + MF Ts = 0.005	100	130
SMF + MF Ts = 0.01	85	120
SMF + MF Ts = 0.016	85	120
Z-detector	60	120

Table 4.2: Detection range of the MF, SMF + MF and the Z-detector reached for TPR> 75% and estimated from the whale tracking paragraph 4.2 and the results of Figure 4.8. The symbol * indicates that there are gaps in the TPR values.

doubled for the Z-detector. The SMF + MF Ts = 0.005 expands the detection range of the sensor up to 100 km in the presence of ship noise, which is 40 km more than the Z-detector.

4.6.3 SMF limitations

The method reaches its limitations when the estimated value of $\rho[k]$ is equal to zero or negative (3.33). In that case, the SMF applies the first filter, and therefore, the observation is not considered as a signal. Due to the definition of the estimation of $\rho[k]$ used in § 3.5.2.2, it might occur in different contexts. If the call is completely embedded in short duration noises, but in the same bandwidth, e.g., seismic event or high-intensity fin whale pulses, the noise estimation might overcome the estimated presence of the call. This combination of events is quite unusual and

might be hard to detect even by an experimented analyst. The second source of error occurs on remote calls. As discussed in § 2.3.5, only unit A remains from long-distance propagation. The emitted calls follow multiple-path spreading denoted on recorded data by multiple echoes. In some circumstances, it leads to an almost uninterrupted signal in the unit A frequency band, where the dissociation of singular calls is complicated, even for an automatic detection algorithm. It might lead to missed detections.

4.7 Conclusion

To conclude, Figure 4.10 compares the performances of the studied methods over the entire dataset in the form of a bubble chart where the axis represents the TPR and the FDR and circles size represents the shorter detection radius, in the presence of ship noise. The best detector should be in the lower right corner, with the biggest circle. This is where the SMF + MF circles

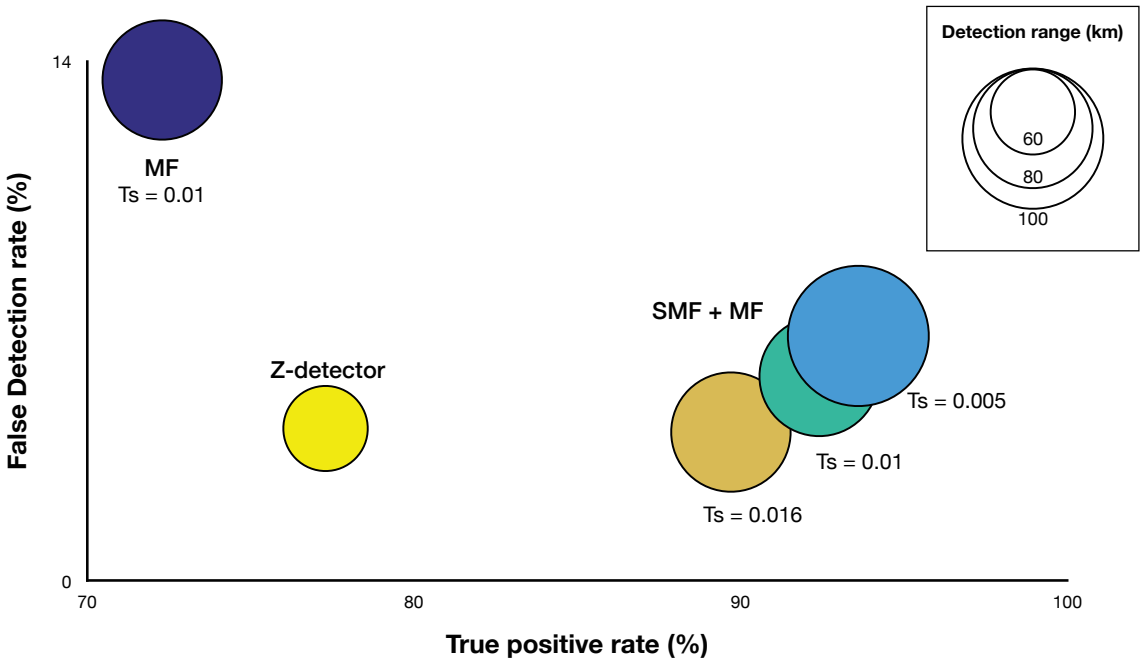

Figure 4.10: Recap of the experimental performances of MF, SMF +MF and, Z-detector on a 22h long noise-corrupted database with 845 annotated ABW calls. Circles represent the most limited detection radius, in presence of ship noise.

are located, whatever th threshold. With Ts= 0.005, the method performs the best TPR and best detection range. The Z-detector and SMF + MF with Ts= 0.016 share the same FDR however, the SMF + MF exceed the Z-detector performances on the two other criteria.

Overall, in passive contexts, it is challenging to control the content of recorded data. Therefore, the evaluation of detection algorithm performances has to be either assessed using simulations or confronting the detector to a ground truth dataset. The first one, related to the detection theory, provides a probabilistic approach of the method's performances and satisfies classic

signal processing techniques. It is often not representative of the method's robustness against noise, to the detection of degraded signals, etc. Ground truth datasets provide a large variety of observations and set of events that would be hard to recreate. However, data annotation is subjective and highly variable between data analysts, and even one can not always be consistent (Leroy et al., 2017b). Creation of an open-source dataset would be beneficial for algorithm training, performance analysis, and detector comparison.

Until this point, this thesis focused on binary detection, the detection of one call type at a time. However, there are often multiple calling species in the recordings and, detectors such as the MF, SMF and, Z-detector are not designed to detect different types of signals at once. Multi-species detection, therefore, requires another set of signal processing strategies. The detection of multiple types of signals often resorts to more generalized detection and classification systems: pattern recognition algorithms (§ 1.5.1). A generalized method, based on the classification of BW tonal signals and their reconstruction is presented in Chapter 5.

Chapter 5

Automatic transcription

Contents

5.1 Introduction

As discussed in section 1.5, binary detection is rapidly limited when there is more than one target signal. Figure 5.1 displays an example of concurrently calling species with simultaneous calls from ABWs, MPBWs and remote P-calls. In this context, where signals of interest are overlapping in time and frequency and, recordings are subject to various noise conditions, the detection of multi-class signals often relies on a pattern recognition system. The proposed method is an automatic transcription algorithm which can identify multiple concurrently calling species in sound recordings. This algorithm is based on pattern recognition of tonal calls in the TF domain, and follow a classical sequence (Figure 5.2): (1) detection of signals of interest, (2) features extraction and, (3) classification. The classified tonal signals are then used to reconstruct, separately, the underlying songs.

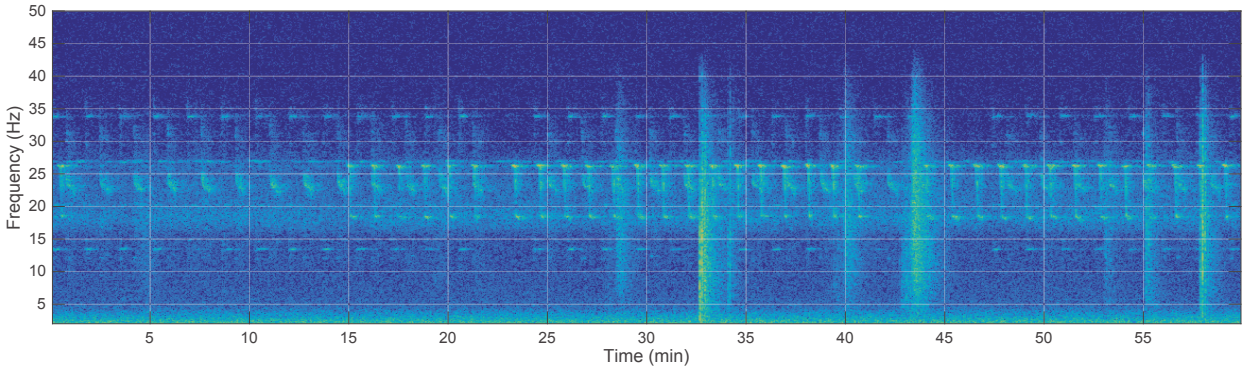

Figure 5.1: Illustration of concurrently calling species identified as TF overlapping calls (ABW, MPBW and, P-calls).

The success of the transcription algorithm depends on the quality of the automatic extraction of signals of interest. This crucial step relies on what makes BW calls distinct from each other and, from the noise or other transient look-alike signals. Based on the knowledge of their calls, portions of calls that favor the transmission of the information with less degradation through propagation (§ 2.3.6) are selected: the tonal parts. The determination of which call parts identify as tonal is discussed in § 5.2.1 before the presentation of the different steps and results.

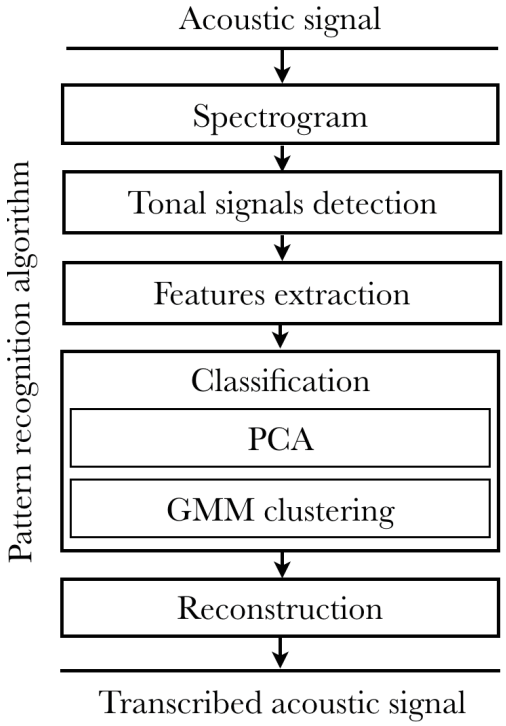

Figure 5.2: Automatic transcription algorithm flow chart.

5.2 Segmentation of blue whale calls

5.2.1 Are baleen whale calls tonal signals?

Mathematically, a pure tone is determined as a simple sine function with a constant frequency (and linear phase). However, a chirp is determined as a pseudo-periodic signal, sine function with non-linear phase and varying frequency. For the observed baleen whale song units (or in the case of MPBW calls, partials), most are not perfect pure tones. To analyze which units (or partial), can be considered tonal for subsidiary detection and transcription, two parameters are estimated (Figure 5.3):

- the rate of frequency change (or chirpyness) $\frac{f_{max} - f_{min}}{\Delta T}$ and,
- frequency variations $f_{\rightsquigarrow}$, representing how much a signal frequency evolves through time compared to its maximum frequency (in %/s) such as

$$f_{\rightsquigarrow} = \frac{f_{max} - f_{min}}{f_{max}} \times \frac{100}{\Delta T}. \tag{5.1}$$

For the characterization, measurements were done using Raven Pro[1] on recordings centered on the high SNR observations from section 2.3 and, 20 to 30 units were analyzed for each category. Results are presented in Table 5.1. The evaluation of chirpyness and $f_{\rightsquigarrow}$ stresses faster and steeper frequency variations for FW pulses and ABW unit B, with values > 1 (when other units are < 1), due to important frequency variations over short periods.

[1] http://ravensoundsoftware.com

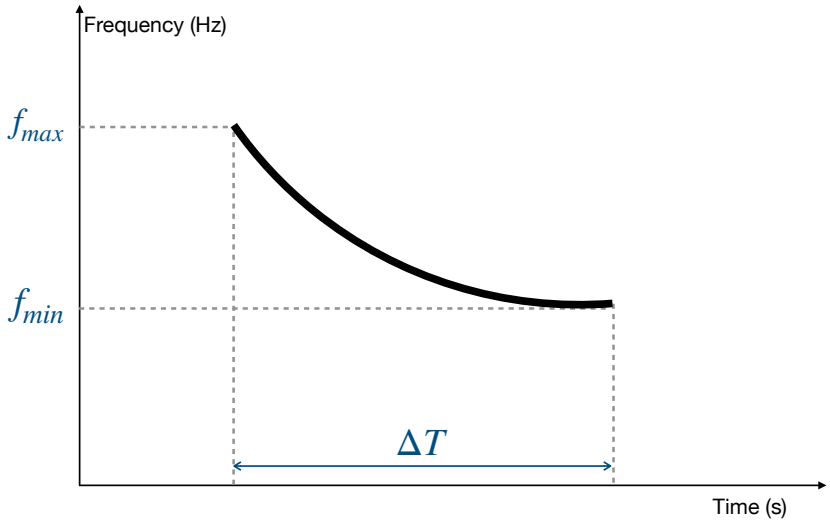

Figure 5.3: Illustration of the time and frequency parameters for chirpyness and frequency variation analysis.

	f_{max} (Hz)	f_{min} (Hz)	ΔT (s)	Chirpyness (Hz/s)	$f_{\rightsquigarrow}$ (%/s)
FW	29.9	16.2	1.5	**9.13**	**30.5**
MPBW unit 1 high	34.9	33.9	18.6	0.05	0.2
MPBW unit 1 low	14.1	13.4	20.6	0.03	0.2
MPBW unit 2 DS	24.0	22.3	23.6	0.07	0.3
ABW unit A	26.6	25.9	12.7	0.06	0.2
ABW unit B	25.9	18.8	2	**3.55**	**13.7**
ABW unit C	18.8	18.3	11.1	0.05	0.2
P-call	27.3	26.6	19.8	0.04	0.1

Table 5.1: Analysis of baleen whale call units and partials for chirpyness (Hz/s) and frequency variations $f_{\rightsquigarrow}$ (%/s). Analysis performed on 20-30 high SNR units per type.

A decrease in ABW call frequencies, among other large whales, have been observed over the past decade (Leroy et al., 2018). Annual decay is estimated to be −0.14 Hz/year. At these very low frequencies, it represents a 0.6% change of unit A peak frequency. Hence, small frequency variations are significant. This is why, in accordance with results of Table 5.1, FW pulses and ABW unit B are considered as chirps and the other analyzed units (or partial) are considered tonal. Automatic tonal signal detection methods are investigated in § 5.2.2 for further use in the transcription algorithm.

5.2.2 Tonal signal detection

Multiple tonal signal detection algorithms are available in the literature. They can be classified as time-domain, frequency-domain, or spectro-temporal domain methods. Some of these are from the field of speech and musical signal processing where they are often referred to as *pitch detection* or *pitch tracking* algorithms (Babacan et al., 2013). Methods operating in the time

domain, such as the instantaneous frequency estimator (Boashash, 1992) (§ 5.2.2.1), are straightforward to implement but are known to not being robust in case of polychromatic signals. For speech and musical sounds, fundamental frequency (f_0) estimators such as auto-correlation function (ACF) and its extension, the YIN estimator are well known and are commonly used (De Cheveigné and Kawahara, 2002) (§ 5.2.2.2).

Methods operating in the frequency-domain exploit the harmonicity of voiced or instrumental signals to estimate f_0 and are extensively used in speech processing. Examples include cepstrum-based methods, the summation of residual harmonics, and harmonic product spectrum (HPS)(Noll, 1969). HPS measures the maximum coincidence of harmonics for each spectral frame (§ 5.2.2.3). Another method, the statistical maximum likelihood pitch detection algorithm (Noll, 1969; De La Cuadra et al., 2001), also uses this characteristic of voiced sound by searching through a set of possible ideal spectra (impulse train convoluted with the signal window's spectrum) and choosing the one which best matches the shape of the input spectrum. Since BW calls do not always exhibit harmonicity in their songs (or at least not in the investigated frequency band), HPS was the only frequency-domain method that provided results worth presenting here.

Methods operating in the spectro-temporal domain often employ image-processing operations such as image-thresholding, edge-detection (Baumgartner and Mussoline, 2011; Gillespie, 2004) and ridge (intensity) detection (Kershenbaum and Roch, 2013; Madhusudhana, 2015). Two such methods are included to our comparative study (Cost-function based detector § 5.2.2.4 and ridge detector § 5.2.2.5).

Theoretical elements of the five compared methods (instantaneous frequency estimator, Yin estimator, HPS, Cost-function-based detector, and ridge detector) are briefly described in the following paragraphs [2].

5.2.2.1 Instantaneous Frequency estimator

The instantaneous frequency is a time-dependent characteristic of a signal. It searches for the location of its "peak frequency" (frequency with maximum amplitude) over time. For the discrete and complex observation, denoted $z(n) = \mathrm{A}(n)e^{j\varphi(n)}$, where $\mathrm{A}(n)$ and $\varphi(n)$, respectively, stand for the time-varying amplitude and phase that can be obtained using the discrete Hilbert transform operation (Boashash, 1992), the instantaneous frequency can be estimated as

$$\hat{f}(n) = \frac{f_s}{2\pi}(\varphi(n+1) - \varphi(n)). \tag{5.2}$$

For increased accuracy, this estimation is followed by a third-order one-dimensional median filter of 21 samples ($\simeq 0.2$ s). Relation (5.2) indicates that the instantaneous frequency estimator can detect only one frequency at a time and therefore is only meaningful for monochromatic and narrow-band signals.

[2]The Matlab code developed for each method is available online: `https://leabouffaut.github.io/tonal_detectors/`

5.2.2.2 YIN estimator

The YIN algorithm (De Cheveigné and Kawahara, 2002), originally developed for speech and musical applications, is a fundamental frequency (f_0) estimator derived from the ACF. The ACF is the inverse Fourier transform of the power spectrum. In response to a periodic signal, it shows peaks at multiples of the signal's period. Frequency f_0 can then be estimated by choosing the highest non-zero-lag peak (zero is always the highest peak). Using this approach, an error might occur if the lowest peak is too close to zero or by choosing higher-order peaks (if their amplitudes are higher). In order to prevent sensitivity to amplitude changes, erroneous peak selection, global error, and, to improve precision, the YIN estimator presents multiple improvements to the ACF.

5.2.2.3 Harmonic Product Spectrum

The HPS has been used for detecting fundamental frequency in periodic signals and for estimating pitch in human speech (Schroeder, 1968; Noll, 1969). It relies on the simple idea that pitch peaks in the log spectrum are harmonic multiples of the fundamental frequency. Therefore the product of r-integer compressed versions of the log spectrogram should enhance the peak of the fundamental frequency. Estimation of the fundamental frequency is given by

$$\hat{f}_0 = \arg\max_f \prod_{r=1}^{\mathrm{R}} |\mathrm{Z}(rf)|, \tag{5.3}$$

where R is the number of harmonics to be considered (and also is the number of frequency-compressed copies of the original spectrum), r is the decimation factor, $|\mathrm{Z}(f)|$ is the amplitude spectrum of the observation and $|\mathrm{Z}(rf)|$ is the r^{th} compressed version of the original spectrum. It is applied to each time frame of the spectrogram. However, the observed baleen whale calls do not present harmonic characteristics.

5.2.2.4 Cost-function-based detector

The detection method described in Baumgartner and Mussoline (2011), referred to as the cost-function-based detector, is applied to the spectrogram. First, the spectrogram is amplitude thresholded. Tonal detection is performed by applying a cost function for each successive time bin of the remaining spectrogram pixels. The cost function penalizes large frequency, and amplitude jumps over short time steps. The path with the minimum cost is reconstructed. During tonal tracking, amplitude weighting is applied, so that louder parts of the call are weighted higher compared to quieter parts.

5.2.2.5 Ridge detector

This method relies on ridge detection, a widely used image-processing technique for automatic feature selection and image segmentation. This detector treats spectrogram regions corresponding to tonal signals as intensity ridges, and determines their TF contours (Kershenbaum and Roch, 2013; Madhusudhana, 2015).

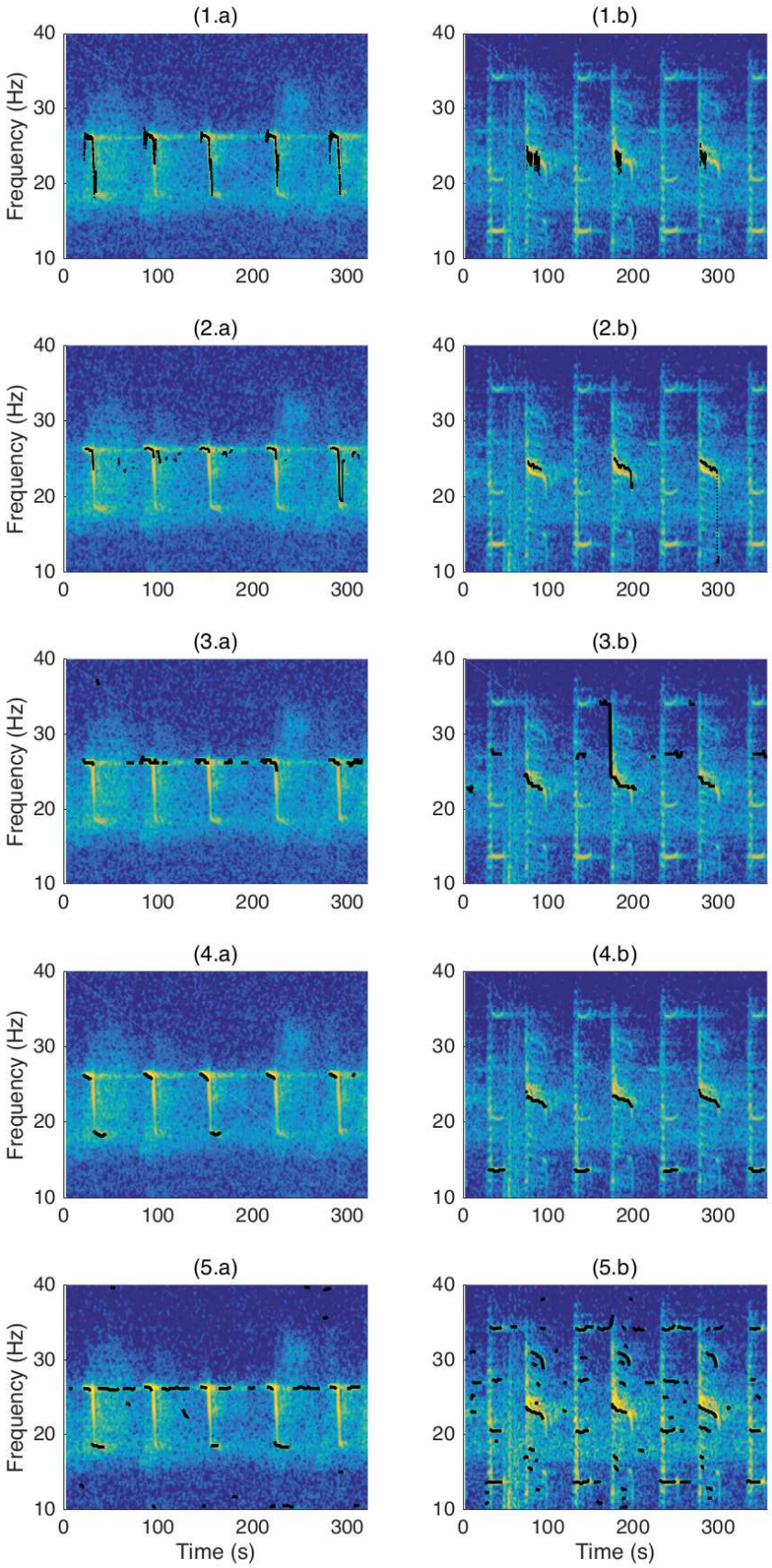

Figure 5.4: Illustrations of ABW (column a) and MPBW (column b) calls obtained with a spectrogram window of 5.12 s (512 samples, 80% overlap) and 0.2 Hz frequency resolution. Dark lines show output from: the instantaneous frequency estimator* (row 1), YIN estimator* (row 2), HPS (row 3), cost-function-based detector (row 4; Ts = 40%), and ridge detector (row 5).

*Outputs derived from the time-domain processing results, the spectrogram is computed for display only.

First, TF points along intensity ridges in the spectrogram are identified. Then, using additional information from the immediate spectro-temporal neighborhood of the previously identified points, TF points corresponding to tonal signals are connected using a Bayesian filtering strategy. The source code and method are available online (Madhusudhana, 2018).

5.2.2.6 Application

The different methods are illustrated on series of five ABW calls (FIG 5.4 column (a)) and three calls and one unit produced by MPBW (FIG 5.4 column (b)), respectively recorded by OBS RR43 (on May 31^{st}, 2013 at 12:33 UTC - § 2.3.1) and RR47 (on May 28^{th}, 2013 at 18:18 UTC - § 2.3.2). This observation set can be used for qualitative evaluation of the detectors. It can be used to analyze the response of each tonal detector to (1) a single tonal (ABW unit A), (2) sharp frequency variations (ABW unit B), (3) varying frequency content (MPBW unit 2 Down-sweep) and (4) to polychromatic signals (on ABW when unit A echoes and C are concurrent or on MPBW unit 1).

Spectrograms are computed for cost-function and ridge-detector methods using a Hann window of 5.12 s (512 samples) with 80% overlap, resulting in a $\simeq 0.2$ Hz frequency resolution and 1.02 s time resolution. The YIN estimator minimum detectable frequency is set to 4 Hz. HPS number of harmonics is set to 2, due to the low f_s of our data. A spectral threshold is empirically set to -20 dB for both the cost-function and ridge detector. Further, to increase computational efficiency, the frequency range was limited to the range of 10 Hz to 40 Hz. For a 'fair' comparison, detection results of the time-domain methods (instantaneous frequency and YIN estimator) were limited to $[10-40]$ Hz also.

In order to compare the methods similarly, detection outputs of the instantaneous frequency estimator, YIN estimator, and HPS are sorted into detected tonals TF vectors. A few rules are applied using the knowledge on the content of each recording: tonal minimum duration is set to 5 s and, minimum TF distance between two successive tonals is set to 3 s or 0.8 Hz. When these requirements are not fulfilled, detected points are discarded. Unlike the first three methods, the cost-function and ridge detector perform in the TF domain and, fill detection vectors only in the presence of salient tonal signals.

With the ABW Z-calls (Figure 5.4 column a), unit A and the beginning of unit B, where most energy is concentrated, are well detected by all the methods. However, this is not the case for unit C with the first three methods (Instantaneous frequency (1.a), YIN (2.a) and HPS (3.a)). This is likely due to the simultaneous reverberation of unit A. HPS (3.a) generates upper harmonics artifacts, but thanks to the cleaning process, they do not affect the estimation of the Z-calls (except at $\simeq 40$ s where few points remain at $\simeq 38$ Hz, which is $2 \times$ unit C's frequency). The cost-function-based approach (4.a) detects Z-call unit As and some of the unit Cs while the ridge detector (5.a) detects all unit As (including the reverberation) and most unit Cs.

For the MPBW calls (Figure 5.4 column b), the first units are not detected by either the in-

stantaneous frequency (1.b) nor the YIN estimator (2.b), probably because of multiple strong co-existing tonal signals. The second unit down-sweep is not precisely determined by the instantaneous frequency (large frequency span surrounding the "true" tonal signal) and estimated a few hertz off by the YIN estimator. When the HPS (3.b) is used, harmonic artifacts emphasize a $\simeq$ 27 Hz tonal signal instead of the expected $\simeq$ 13.5 Hz MPBW unit 1 fundamental. It is one of the known limitations of the HPS (octave error) (De La Cuadra et al., 2001). The 34 Hz tonal of unit 1 appears on the estimate, only when reverberated (otherwise, $\simeq$ 27 Hz is predominant). Since unit 2 down-sweep does not present harmonics, it is correctly estimated. The cost-function detector (4.b) detects the first unit lowest tonal (13.5 Hz) as well as the second unit down-sweep. The ridge detector (5.b) shows that one of the main advantages of this method is that it can detect more than the f_0. Indeed, all main tonal components of both the first and second unit are detected. However, some unexpected tonals are detected, e.g., at $\simeq$ 27 Hz, which are attributable to non-target P-calls.

5.2.2.7 Discussion

The application of five tonal signal detectors (instantaneous frequency estimator, YIN estimator, HPS, cost-function detector, and ridge detector) to high SNR recordings of ABW and MPBW calls showed that, two methods stand out:

- the cost-function detector due to its ability to retrieve the highest intensity frequency (peak frequency) and,

- the ridge detector for its ability to detect MPBW concurrent tonal signals.

Ground truth information of tonal contour is required to evaluate the performances of these algorithms. In order to characterize methods robustness against SNR variations, crucial in passive acoustic monitoring applications, annotated data should cover various SNR ranges. Such publicly available annotated databases do not exist for low-frequency baleen whale calls. It is yet possible to add simulated signals to various recorded background noises with a controlled SNR. ABW calls were chosen for the simulations because of the simplicity of their TF contours. Simulation for detectors comparison and performance analysis are discussed in section 5.2.3.

5.2.3 Tonal detector comparison and performances analysis

5.2.3.1 Evaluation data

A ground-truth dataset is generated by embedding synthetic Z-calls with controlled SNR values in real recordings obtained with OBSs. First, a 79-hour noise dataset is selected from the RHUM-RUM data. Records are picked from November 2012 (7 hours on RR42) and from January 2013 (72 hours on RR47) and do not contain any biological sources (Figure 2.20)(Dréo et al., 2019). However, the noise dataset features various ship noises and seismic events.

Synthetic ABW Z-calls are generated according to the parametric model presented in section 3.5.1.1, from which it is possible to retrieve the phase information of the call and simulate a Z-call waveform (Socheleau et al., 2015). Z-calls present two tonal units (A and C), each one lasting ≃ 8 s and, units are linked by a 4 s linear down-chirp, meaning that 75% of the call is fully tonal. Twenty 20 s-long Z-calls, with a non-overlapping random time of occurrence, are added to 1800 s long noise tracks randomly extracted from the 79-hour noise dataset. Each SNR is controlled according to the description proposed in Mellinger and Clark (2006) and varies between 0 and 15 dB. The accurate knowledge of the signal instantaneous frequency, SNR, and offset of superposition constitute the ground-truth information.

5.2.3.2 Evaluation method

Silbido[3] scoring metrics (Roch et al., 2011) are employed to describe the effectiveness of these detectors to retrieve tonal signals and to describe the quality of the detections. *Silbido* scoring examines the set of detected tonal signals for possible correspondence to tonal signals in the ground-truth dataset (Roch et al., 2011). The detection outputs are therefore processed as in § 5.2.2.6.

Silbido and its scoring tool were originally developed (in Matlab) for automated extraction of higher frequency odontocete whistles. Therefore, the scoring method need some alterations to be able to work with the low-frequency calls of baleen whales. The following parameters are changed in the `dtPerformances.m` file: the framing duration is set to `Length_ms = 512` ms and the advance to `Advance_ms = 256` ms. The search range for detection around the ground-truth frequency is changed to ± `PeakTolerance_Hz = 1.5` Hz (± 4 frequency bins). For each point retrieved in this search range, the absolute frequency difference between the detection and ground truth is computed. The tolerance for a match (a valid detection) is changed to `MatchTolerance_Hz = 1.5` Hz. Detections outside of this range are considered false positives. The higher and lower cutoff frequencies are set to `thr.high_cutoff_Hz = 40` Hz and `thr.low_cutoff_Hz = 10` Hz.

Silbido offers multiple scoring metrics when using the function `dtAnalyzeResults.m`: Precision, Recall, Frequency deviation, Coverage and Excess. **Precision** (or TPR equation (4.1)) measures the percentage of detections that are correct, and, **Recall** (or PPV, equation (4.3))) evaluates the percentage of the expected detections that are retrieved. The **Frequency deviation** quantifies the average distance (in Hz) between points on the TF contour of the ground-truth tonal signal to those of a matched detection. **Coverage** is the percentage of duration of each ground-truth signal that is detected. **Excess** indicates 'extra' duration (in seconds) compared to the ground-truth signal. **Fragmentation** counts the number of fragmented detections per ground-truth signal (Figure 5.5).

[3] `https://roch.sdsu.edu/index.php/software/`

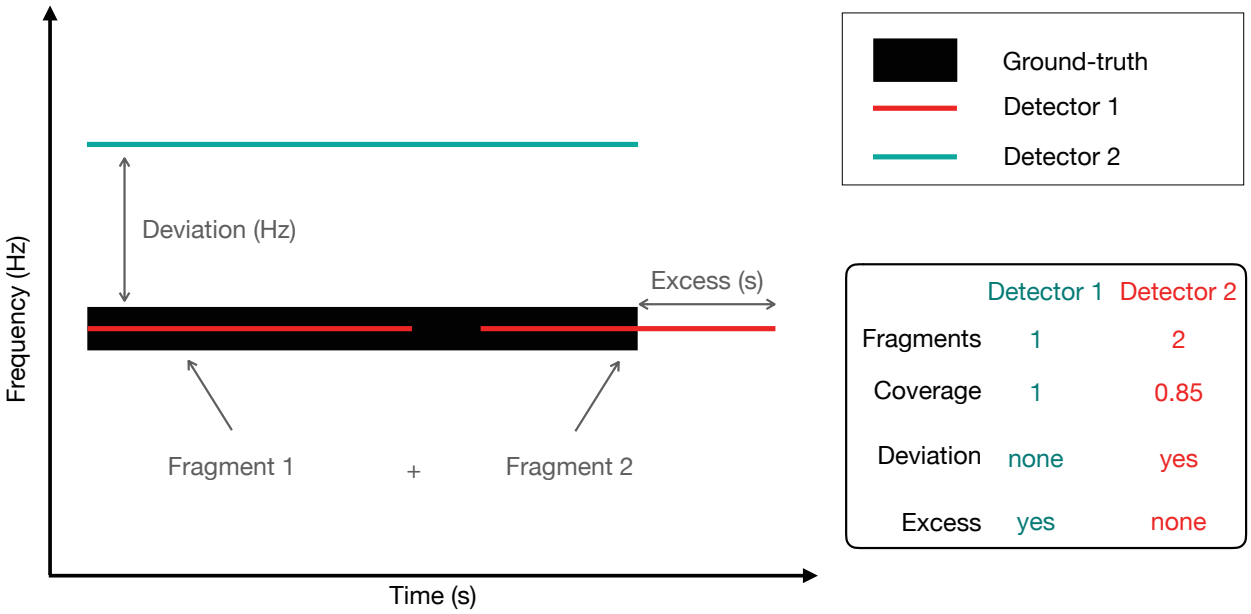

Figure 5.5: Illustration of *Silbido* scoring metrics (Fragmentation, Coverage, Deviation and Excess).

For statically robust evaluation of the performance scores, Monte-Carlo experiments are conducted with 20 iterations per SNR in the range 0 to 15 dB. Results produced by Silbido's `dtAnalyzeResults.m` function are discussed in § 5.2.3.3.

5.2.3.3 Results

Mean scores over the 20 Monte-Carlo simulations for different SNR ranges are presented in Figure 5.6. For illustration purposes, values are displayed as ratios between 0 and 1 except for the frequency deviation (absolute value in Hz). The displayed **Fragmentation** is calculated as 1/N, where N is the number of fragments describing a ground-truth tonal, e.g., if a tonal is detected as one single track, Fragmentation = 1, however, if it is detected as 4 fragments, Fragmentation = 0.25. The **Excess** is plotted as a percentage of the duration of the ground-truth call, e.g., if the detected tonal exceed the signal by 5 s, Excess = 5/20 = 0.25. Ideal values are also displayed on the spider plots for reference: 1 for Recall, Precision, Fragmentation and Coverage, and 0 for Deviation and Excess.

As a general trend, all methods perform better with increasing SNR. Note that the instantaneous frequency does not detect tonals below SNR = 3 dB and thus is not shown on the corresponding plot.

At SNRs below 3 dB, **Precision** is $\simeq 0.4$ for the HPS and about 0.8 for the other methods. With increasing SNR, the ridge detector and cost function detector reach a Precision of 1 in the $[4-6]$ dB SNR range, followed by the instantaneous frequency in the $[7-9]$ dB SNR range.

Recall spans between $\simeq 0.7$ and 0.85 for SNRs below 3 dB, the highest Recall is obtained by

the ridge detector. At higher SNR values, a Recall close to 1 is obtained by the YIN estimator, closely followed by the two TF methods in the SNR [7 – 9] dB range.

Best results in the **Deviation** category are obtained by the HPS and ridge detector (≃ 0.1 Hz), about half of the spectrogram frequency resolution. Results suggest that half of the detected frequencies are in the correct frequency bin while the other half is in an adjacent one. Other methods are closer to 0.25 Hz.

About 65 – 75 % of the call is covered by the cost-function and ridge detectors, 75% corresponding to 100% of the Z-call fully tonal signals (unit A and C § 5.2.3.3). The YIN estimator **Coverage** evolves from 50% at SNRs lower 3 dB to 75 – 80% for higher SNR values. The instantaneous frequency estimator Coverage raises from 50% for the [4 – 6] dB SNR range to 90% for higher SNR ranges. HPS Coverage is lower than the other methods, around 50%.

For most methods, **Excess** is close to 0, except for the HPS with an **Excess** of ≃ 0.2 s.

Fragmentation also improves with the SNR, getting closer to one. No values are lower than 0.5, i.e., calls are detected in a maximum of two parts. Using the cost-function and the ridge detector, it is close to 0.6: ≃ 65% of the calls are detected in 2 parts, which is consistent with Coverage values (no detections of unit B). For the Yin estimator, HPS and instantaneous frequency estimator (from the [7 – 9] dB SNR range and higher), Fragmentation is between 0.8 – 0.9. Only 10 – 20% of the calls are fragmented in 2 parts (so for 80 – 90% of the calls, units A, B and C are detected at once, consistently with Coverage values).

5.2.3.4 Discussion

Comparison against common scoring metrics is performed on those simulations (§ 5.2.3.3). Methods that achieved best overall scores and are the least sensitive to SNR variations are the YIN estimator (temporal method), the cost-function and ridge detectors (both operating in the TF plane). As for the other methods, robustness could be improved by reducing noise in a pre-processing step.

The choice of the detector highly relies on the intended application. The YIN estimator is straightforward to implement. It provides results that are robust to noise, where almost all tonals are retrieved, and with few false positives. For higher SNRs, this estimator achieves satisfying Coverage, but, its Frequency deviation is higher than other methods. It can be used in cases where no further classification is required, e.g., for source-level measurement. Both the cost-function and the ridge detector are reliable from low to high SNR values. They provide similar tonal-retrieval scores and generate only a few false positives. At SNRs lower than 3 dB, the ridge detector generates a lower number of false negatives that manifests through higher recall. The tendency is inverted at higher SNRs, but both methods keep similar recall values. Independently of the SNR, ridge detection has lower Frequency deviation. Low deviation is of importance if the

detected tonals are being classified at a second stage.

For the subsequent use of the detected tonal signals for classification and reconstruction, the chosen tonal detector should (1) be reliable with high Precision, and Recall scores, (2) provide satisfying Coverage of the tonal units, and equivalent Fragmentation and, (3) detect with low-Frequency deviation for robust features extraction. Besides, even if not taken into account in the analysis of the performances, the method should be able to detect multiple concurrent tonal signals, e.g., in the presence of MPBW partials, in recordings of multiple concurrently calling species or, when strong reverberation of ABW unit A leaking into unit C occurs. In light of the above, the chosen method is the ridge detector (Kershenbaum and Roch, 2013; Madhusudhana, 2015).

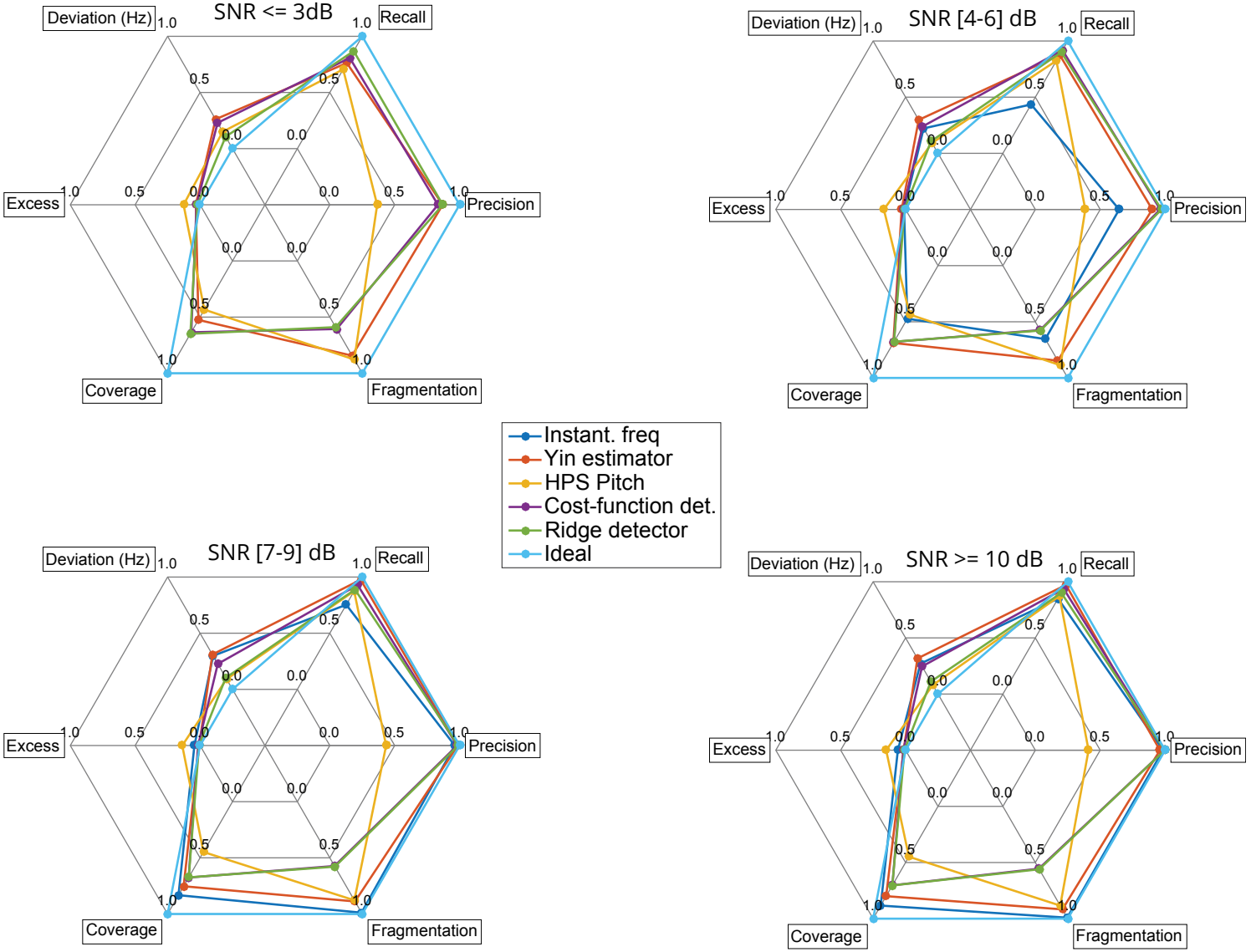

Figure 5.6: Spider plots of the tonal-detection algorithms mean scores for different SNR ranges ($\leqslant 3$ dB, $[4-6]$ dB, $[7-9]$ dB, $\geqslant 10$ dB.). They are assessed on simulations run over 20 Monte-Carlo iterations on 20 synthetic Z-calls randomly drawn in a 30-min long real marine noise records with a controlled SNR. The compared algorithms are the instantaneous frequency estimator, the YIN estimator, the HPS, the Cost function detector and the Ridge detector. Precision, Recall, Coverage, Excess and Fragmentation are expressed as percentages between 0 and 1. Deviation is given in Hz (Bouffaut et al., 2020).

5.3 Classification and reconstruction

5.3.1 Training dataset

The training dataset comprises data from OBS RR48 recorded on May 18^{th}, 2013 (day 138) and from OBS RR41 recorded on May 15^{th}, 2013 (day 135), respectively. All BW vocalizations within the data were manually annotated by drawing a box (describing begin and end times and, minimum and maximum frequencies) around the units using the software *Raven Pro 1.5*. Out of the 48 hours of recordings, more than 4000 individual units were selected in various SNR ranges [4]. Tonal units distribution over the dataset is represented in Figure 5.7. The dataset is representative of the different types of signals present in the recordings, the classes are evenly distributed and contain between 555 and 1133 tonals.

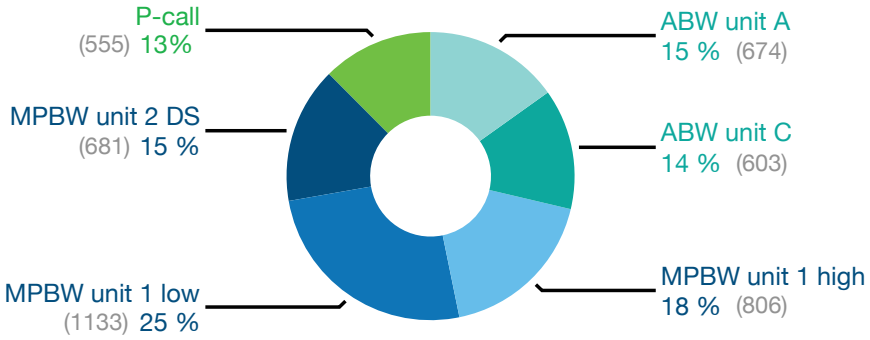

Figure 5.7: Tonal units number of instances and distribution in the training dataset, out of 4000 annotated signals.

5.3.2 Features extraction

Tonal signals detected by the ridge detector are characterized using different attributes *or features*. Features have to facilitate the sorting of the detected tonal signals into different categories. They should also be simple to measure and robust to noise (Duda et al., 2012). They are generally closely related to the application. Based on the work presented in Urazghildiiev et al. (2009) and Baumgartner and Mussoline (2011) for baleen whale signal characterization, the selected set of features measures temporal, spectral, and amplitude variations and includes:

- average frequency $\overline{f}$ (Hz - Figure 5.8(a)),
- center frequency (frequency reached at half of the cumulative signal amplitude; Hz),
- bandwidth (Hz),
- average amplitude and amplitude standard deviation (dB),
- minimal, maximal, average and instantaneous slopes (Hz/s) and,

[4] The annotated dataset is available online: DOI: 10.5281/zenodo.3624145

- least concurrent frequency ratio Ω (dimensionless; described below - Figure 5.8(b)).

When multiple tonal signals occur concurrently (M, number of tonals), ratios of the average frequencies ($\overline{f}$) are computed for each pair of concurrent tonal signals, with the higher $\overline{f}$ of a pair as the numerator so that the ratios are ≥ 1. For a tonal signal m_i ($i \in \{1,2,...,M\}$), the associated least concurrent frequency ratio Ω_{m_i} is taken as the least of such ratios among all pairs m_i and m_j ($j \in \{1,2,...,M\}$ and $i \neq j$). In the absence of concurrent signals, Ω = 1. Ω provides a way for quantifying the polychromatic nature of tonal signals (Figure 5.8):

· for ABW calls when units A and C are concurrent, $\Omega = \frac{\overline{f}_{\text{unit A}}}{\overline{f}_{\text{unit C}}} = \frac{26.2}{18.7} = 1.45$;

· for MPBW unit 1, $\Omega = \frac{\overline{f}_{\text{unit 1 high}}}{\overline{f}_{\text{unit 1 low}}} = \frac{34}{13.5} = 2.5$ and;

· for MPBW unit 2 DS and P-calls (no expected concurrent tonals), ω = 1.

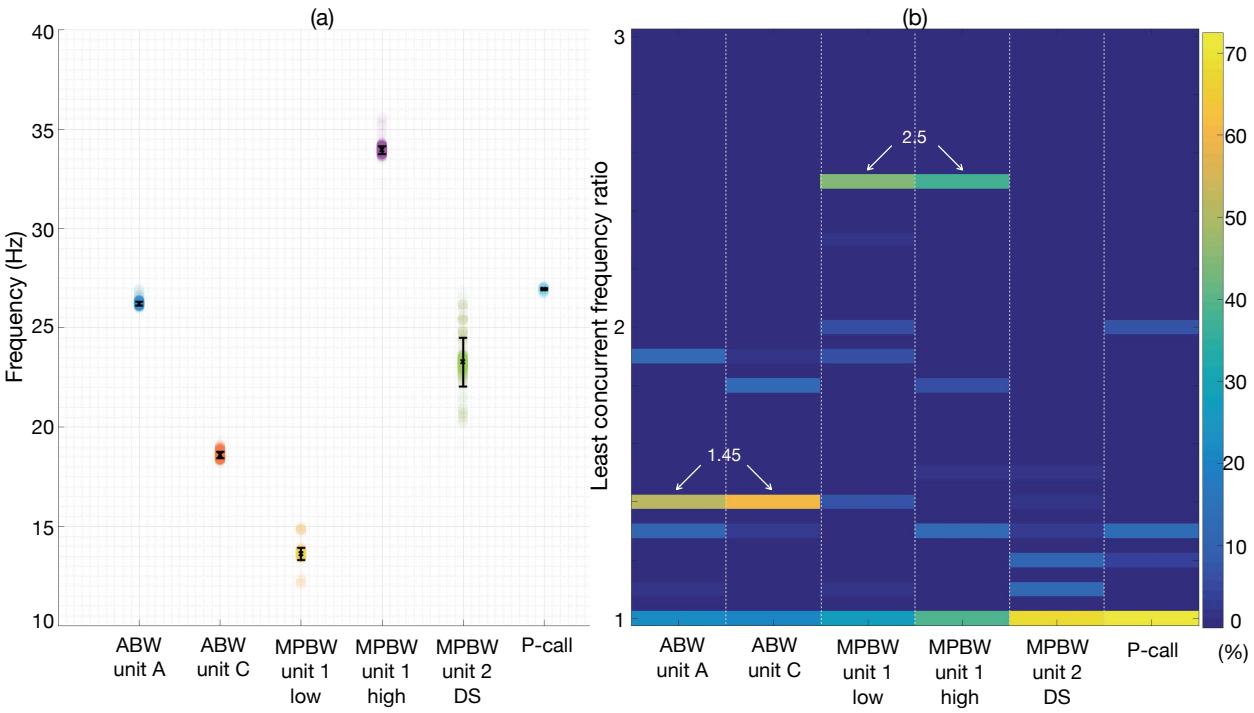

Figure 5.8: Illustration of features of (a) average frequency and (b) least concurrent frequency ratio Ω on the training dataset. (a) Each tonal is represented as a round marker with transparency: accumulation of markers at a specific location shows as more intense color. Each color corresponds to a specific unit and, mean values and standard deviations are indicated. (b) Ω values are indicated and, amplitudes are displayed as the percentage of concurrent tonal units found with each ratio among all annotated units of the same type.

However, Ω could be adversely impacted by the simultaneous occurrence of multiple whale species tonal signals or shipping noise. As an example, an examination of ABW measured ratios for unit A on the training dataset (Figure 5.8), indicates that 50 % of the annotated units are recorded simultaneously with unit C (Ω = 1.45) and 20 % are recorded alone (Ω = 1). Yet, a

smaller portion of the set is recorded with MPBW Unit 1 high ($\Omega = \frac{34}{26.2} = 1.3$; 11 %) and low ($\Omega = \frac{26.2}{13.5} = 1.9$; 13 %). Shared ratio between classes indicate tonals that have been recorded simultaneously.

Multiple-path arrivals (reverberation) caused by the deep-sea bathymetry makes it difficult to isolate and extract the direct signal. Measuring the duration of tonal signals in these conditions can produce inaccurate estimates that may not be representative of the actual signal. For example, estimated duration of ABW unit A calls from the training dataset is 20 ± 12 s, whereas the known duration is 12 s (cf. § 2.3.1). Hence, the signal duration is not considered as a feature for classification purposes.

5.3.3 Dimension reduction and classification

PCA is a tool for features transformation that aims at finding mutually orthogonal global direction in data that maximize variance (Comon and Jutten, 2010). PCA is an application of the KLE (§ 3.4.2), often used for dimension reduction. In the present case, features measured on each detected tonal signal are put together as a table, or input matrix D of size [N tonal signals ×10 features]. PCA transformation of the input matrix into L dimensions (denoted T_L is expressed as the product between the original data matrix D and a weight matrix W_L such as

$$T_L = DW_L. \tag{5.4}$$

The weight matrix is composed of the eigenvectors of the covariance matrix of D (D^TD), sorted by descending eigenvalues order. The eigenvalue problem is solved by singular value decomposition. Explicitly, W_L expresses the influence of each one of the original features on the principal component (PC) dimensions. For dimension reduction, L is chosen smaller than the original number of dimensions ($L < 10$). Here the number of PCs is reduced to 2, conveying 98.5% of the data total variance. Compared to similar algorithms such as independent component analysis, PCA gives the best reconstruction: each tonal signal is represented in the new L dimensions. T_L is of size [N tonal signals ×L]. Training is used to estimate W_L and, is saved to reproduce the same transformation on other datasets.

In the reduced 2-dimensional space, points corresponding to call units are grouped into clusters using Gaussian mixture models (GMMs). GMMs are commonly used to estimate of the probability density functions in statistical classification systems (Duda et al., 2012). GMMs, unlike K-means, account for data variance (hence it makes sense to use it after variance maximization by PCA). For the training data, six distinct clusters were observed, and they corresponded well with the six annotated tonal types (see Figure 5.7).

5.3.4 Reconstruction

Detected tonal signals associated with a particular class are used in the reconstruction of a putative independent song. First, the STFT $\mathrm{X}(t,f) \in \mathbb{C}$ of the input signal $x(t)$ is calculated. Complex STFT enables the subsequent reconstruction of a signal without phase information losses. A binary mask $\mathrm{Y}_i(t,f) \in \{0,1\}$ for the i^{th} class (prepared by setting points along all detected TF contours in the i^{th} class to 1, and 0 elsewhere) is applied to the STFT as

$$\mathrm{Z}_i(t,f) = \mathrm{Y}_i(t,f) \odot \mathrm{X}(t,f), \tag{5.5}$$

where $\odot$ is the Hadamard product. Finally, the time-series data representing an independent song is obtained by computing the inverse STFT of $\mathrm{Z}_i(t,f)$, i.e., $z_i(t) = \mathrm{iSTFT}\{\mathrm{Z}_i(t,f)\}$.

5.4 Results

5.4.1 Training performances

Training data projected on the first and second PC are presented in Figure 5.9 and color-coded according to the clustering. PC1 conveys 55.5% of the total variance, 94% of PC1's weight is attributed to frequency features ($\overline{f}$ and center frequency). PC2 conveys 43% of the total variance, 88% of PC2's weight is attributed to the average amplitude.

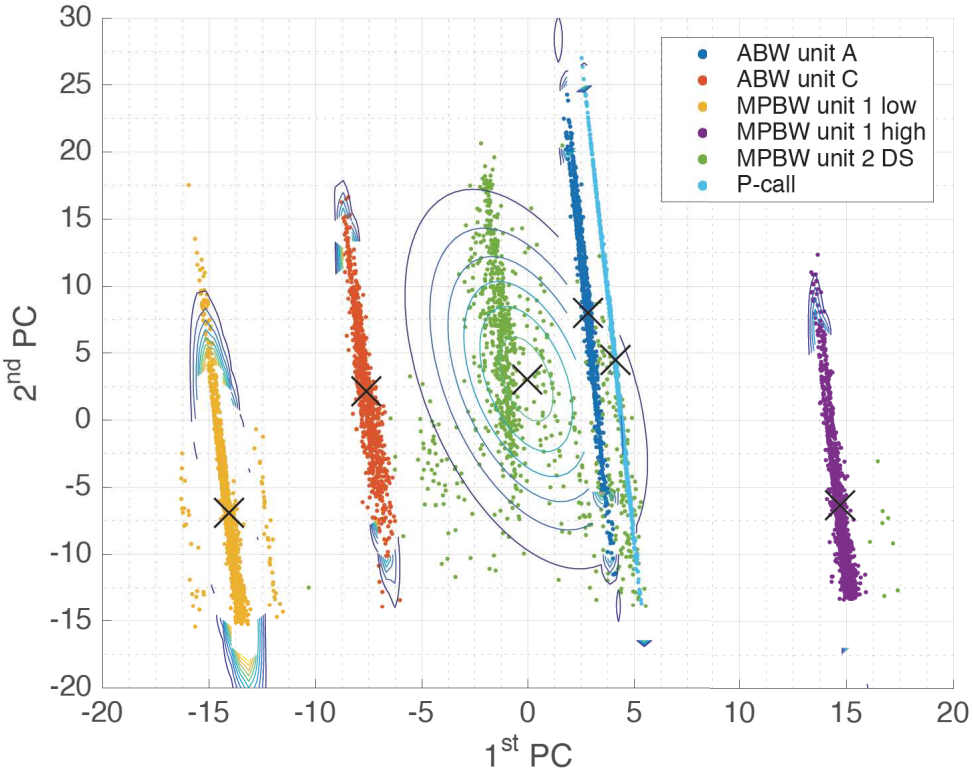

Figure 5.9: Training data projected on the first and second PCs. Colors represent the results of the GMM clustering with precision and recall of Table 5.2.

Three classes are completely distinct due to their distinct frequency differences: MPBW unit 1 low, ABW unit C and, MPBW unit 1 high. Most of GMMs are narrow ellipses. However, the MPBW unit 2 DS ellipse is wider than the others and, its orientation is different: the cluster overlaps with ABW unit A and P-call clusters. Therefore, the clustering performances are quantified using the metrics Precision and Recall (§ 4.3.2). A confusion matrix of the classification outcomes, along with the per-class Precision and Recall values, are presented in Table 5.2.

		ABW		MPBW			?	Precision (%)	Recall (%)
		unit A	unit C	unit 1 low	unit 1 high	unit 2 DS	P-call		
Cluster No.	1	93.28	-	-	-	2.09	0.12	97.69	93.28
	2	-	99.67	-	-	0.25	-	99.75	99.67
	3	-	-	99.92	-	0.25	-	99.75	99.92
	4	-	-	-	99.37	-	-	100.00	99.37
	5	6.24	0.33	0.08	0.63	97.29	7.87	86.52	97.29
	6	0.47	-	-	-	0.12	92.01	99.36	92.01

Table 5.2: Classification results presented as a confusion matrix, along with per-cluster Recall and Precision values.

The higher values in the confusion matrix occur along its primary diagonal, indicating high Recall rates. ABW unit A and P-calls that exhibit strong similarities in frequency and, for this reason they are easily mistaken (§ 2.3.3) (Leroy et al., 2017a; Ward et al., 2017). However, PCA and clustering approaches employed here readily separate the two signal types (see Figure 5.9). As can be seen from Table 5.2, only 0.12% of P-call occurrences were incorrectly classified as ABW unit A and, reversely 0.47% of ABW unit A were assimilated to P-calls.

The spreading of the MPBW unit 2 DS ellipse affects the classification within the classes it overlaps. P-calls (7.87%) and ABW unit A (6.24%) are wrongly attributed to cluster 5, lowering the cluster's precision and, decreasing their recall scores. Yet, average precision and recall of respectively 97.18% and 96.92%, indicate satisfying classification performances for the different units of the annotated training dataset.

5.4.2 Testing: unsupervised application

The proposed transcription process is applied and illustrated on a recording containing multiple MPBW calls and P-calls as well as FW chorus and seismic noise (Figure 5.10(a-b)). The output of the ridge detector is shown in Figure 5.10(c) where colors represent the associated data clusters. Detected tonals occurring outside of the $[10-40]$ Hz frequency range and with power (on the normalized spectrogram) below -60 dB were discarded. Transcribed and reconstructed waveforms are displayed in Figure 5.10(d). A MPBW song, constructed by associating clusters 3 (MPBW unit 1 low) and 4 (MPBW unit 1 high), is displayed in pink. Another MPBW song

consisting of the unit 2 DS only (cluster 5) is displayed in purple. A song consisting of tonals from the P-call cluster (6) is plotted in orange.

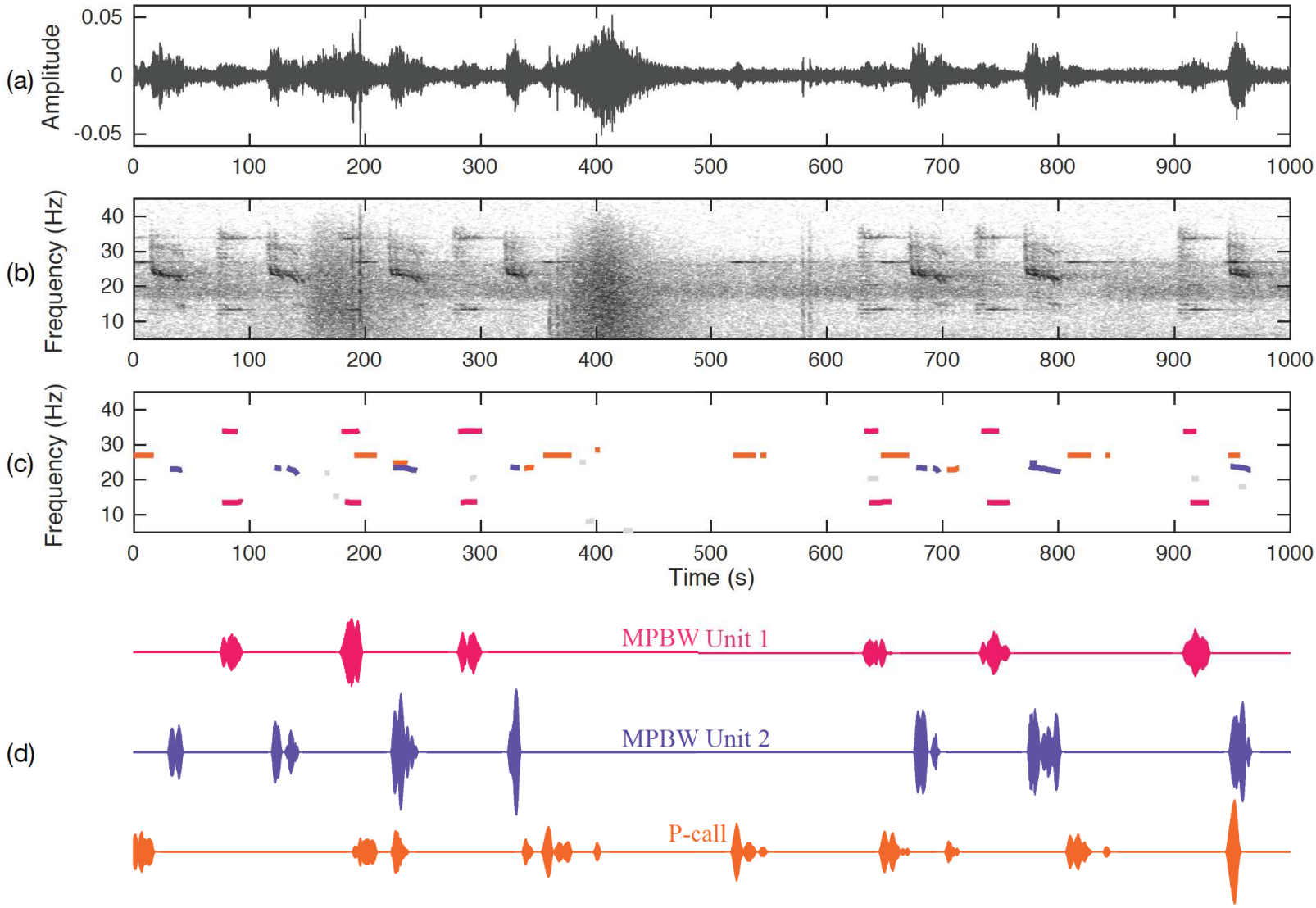

Figure 5.10: Illustration of the performance of the developed method using a recording (waveform (a) and spectrogram (b)) containing MPBW calls and P-calls, FW chorus and two strong seismic events (at 180 s and 420 s, respectively). Tonal detector outputs are shown in (c) and color-coded by the classification results. Waveforms of reconstructed songs are shown in (d) (Bouffaut et al., 2019).

5.5 Discussion

The results presented in Figure 5.10 highlight the effectiveness of the proposed automatic transcription algorithm to retrieve and regroup tonal signals for the reconstruction of independent song tracks. Interfering noises, such as seismic events, have been successfully removed from the resulting tracks. However, the obtained results on tonal extraction using the ridge detector are impacted by the differences between a unit and its echoes: such spectro-temporally disjoint components are detected as independent units. For example, in Figure 5.10, at 110-120 s the MPBW unit 2 DS is detected in at least two parts (fragmentation). This can have many consequences.

First, in the annotated dataset, a unit and its echoes are considered as one annotated tonal signal with a unique label. Therefore, multiple fragments are affected with a label corresponding to a single annotation. As a consequence, in the training process, the percentages reported in Table 5.2 are reflective of the number of fragmented signals associated with each class and are not representative of the actual number of annotated signals.

Fragmentation also impacts classification performances. As discussed in § 5.4.1, MPBW unit 2 DS, is a relatively complex signal in comparison to the other units and, its frequency range is close to that of P-calls and ABW unit A. When the extracted TF contours are fragmented, subsequent estimation of attributes (especially $\overline{f}$ and center frequency) is less accurate. Given that frequency attributes convey most information on the PC1 axis (§ 5.4.1), Fragmentation significantly influences the location of the data point on the PCs axes. The spread of the MPBW cluster and the resulting misclassifications can be attributed to Fragmentation of the extracted TF contours. Furthermore, echoes of the following segments of MPBW unit 2 DS might also yield into incorrect classifications as the corresponding detections present non-typical attributes. As an example, the echoes, at $\simeq$ 340 s and $\simeq$ 710 s in Figure 5.10, were wrongly classified as P-calls.

5.6 Conclusion

To address the issue of the automatic analysis of PAM recording of BW, the proposed strategy is to perform song transcription based on the TF representation of acoustic signals and pattern recognition. The different steps are summarized in Figure 5.11. First, the observation is represented in the TF domain. Then, signals of interest are detected: the tonal parts of BW calls are chosen because they are less degraded by the propagation channel. This step can be assimilated to segmentation and, is carried out using the ridge detector (that was chosen after performance comparison with other tonal detectors). Then, features describing tonal signal TF-amplitude information are extracted. For classification, data are represented on the first two principal components, describing 98.5% of the total variance and, GMM clustering is then applied, performing training performances of 97.2% precision and 96.9% recall. Reconstruction of the hence classified signals is finally performed, providing one separate waveform for each

class of signals. In a preliminary application, non-supervised transcription of a recording of MPBW and P-calls, polluted with seismic noise, provides supportive results of the interest of such method.

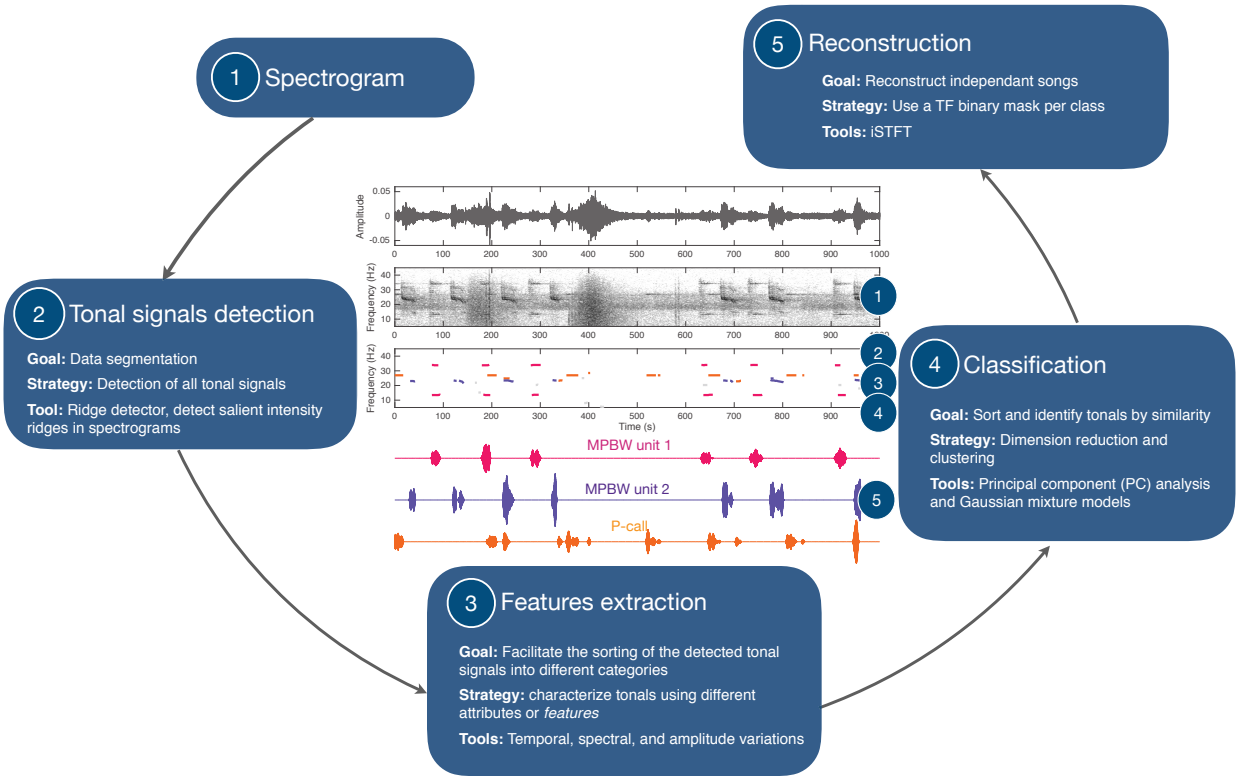

Figure 5.11: Flowchart of the pattern recognition strategy for automatic transcription of BW songs.

Application of the pattern recognition method can be used for a multi-species detection and classification in different PAM applications. Because of the low complexity of each task, it is possible to have a solid understanding of all inputs and outputs and, adapt the method to other types of signals. Besides, PCA transformation and resulting GMMs (after training) are computationally efficient and, could be implemented without much adaptation to real-time on-board processing for event detection, e.g., on gliders or communicating buoys. The additional reconstruction step exhibited possible application for aural and visual analysis. For example, it could intervene in annotation processes, to improve efficiency and reduce variability in order to generate larger ground-truth datasets.

Future work discussed with more details in the general conclusion should focus on:

- improving the algorithm to lower false alarms on broader units such as MPBW unit 2 DS,
- separating songs from individuals of the same species and,
- finding scoring metrics, able to evaluate the complete transcription process performances and conduct comparisons to similar algorithms.

Conclusion and future work

This thesis was devoted to the challenging problem of the detection and classification of stereotyped signals modified by the propagation channel (echoes, low and varying SNRs...) in a passive acoustic context. Solutions to this problem were proposed in a framework that can be adapted to diverse types of stereotyped signals. Strategies were demonstrated and validated on BW low-frequency signals. In the following, an overview of the contributions of this thesis is presented and directions for future work are outlined.

The specificity of the application to BW signals was discussed in **Chapter 1**. First, an overview of BW classical methods for visual surveys was presented, and their limitations were pointed out. It led to the conclusion that information collected visually is too sparse for continuous monitoring in remote areas where most BW populations are. To overcome this limitation, PAM was then introduced as a supplementary approach to conduct long-term surveys, resulting in the combination of different fields such as ecology, acoustics, and signal processing. A focus on propagation showed that BW low-frequency calls are good candidates for long-distance propagation and, a literature review on detection methods revealed that the MF is the most common detection strategy.

The development of appropriate detection and classification methods requires prior knowledge on signals of interest, the detection context, and the possible issues (e.g. overlapping transient noise sources). To that extent, data used in this work, recorded from the bottom of the Indian Ocean by the RHUM-RUM OBSs was presented in **Chapter 2**. The different components of the soundscapes were then analyzed and, signals of interest were described (ABW, MPBW, and P-calls). The calls were shown to be overlapping in frequency and share the same bandwidth than FW pulses, all contributing to the whale chorus. Besides, this already busy soundscape was also affected by ship noise and seismic events.

Chapter 3 first confirmed that stereotyped signals propagating over long-distances are received with echoes and varying SNRs and, should be considered as random signals. Therefore, MFs are no longer optimal. To improve signal detection in such changing conditions, the proposed strategy was based on the SMF which is an extension of the MF, in the sense of optimal filtering and SNR maximization, for stochastic signals embedded in additive colored noise. The filtering strategy uses a new formulation of the observation input SNR, expressed in the form of a Rayleigh quotient. The optimization of this ratio leads to the construction of a set of sub-band

filters designed to maximize the output SNR. The adaptation of the SMF to the passive context is one of the principal contributions of the work. The proposed method overcomes the noise selection step and a new method is introduced for the automatic estimation of the time-varying SNR. In this Chapter, the SMF + MF was successfully applied and compared to the MF for the detection of ABW calls at high and low SNRs. It presented better detection amplitudes and was less inclined to false detection. The SMF was also illustrated on the detection of scuba-divers breathing signals. The two detected signals are different in terms of structure and frequency bands: whale calls are ultra-low frequency (initially) deterministic signals, whereas scuba-divers inhalations are characterized as high-frequency modulated noise.

Performance analysis of the SMF + MF compared to the MF and, the Z-detector was investigated in **Chapter 4**. It was performed on recordings from an ABW singing continuously for more than 21 hours with various SNRs and available tracking information. Received calls were annotated on the farthest OBS from the trajectory, chosen for important SNR variability and, compared to closer sensor for cross-reference. Numerous other sound sources were also denoted during the day (ship noise, P-calls, D-calls, chorus, seismic events), providing subsequent SNR variations. Classic supplementary TPR/MDR and PPV/FDR were estimated after the automatic comparison between detector outputs and the ground-truth data. Performances were displayed: depending on the threshold, as ROC curves; and, with fixed thresholds: against the SNR and during the day. These analyses showed that all methods performed similarly on high SNRs (estimated > 13 dB). However, for lower SNRs, the MF performances rapidly decreased with higher MDR and FDR. The analysis also underlined the critical role of the MF detection threshold. Comparison between the Z-detector and the SMF + MF (with threshold set so both method achieve similar FDR of 3.5 %) showed that the SMF + MF reached TPR = 1 at a lower SNR, therefore increasing the detection range in the day-long analysis. In presence of ship noise, the detection range was extended by 25 km (with a minimum TPR of 75 %) and, for a threshold corresponding to an FDR of 6 %, the SMF + MF was proven to detect signals up to 100 km under the same conditions.

The proposed method for the automatic evaluation of the SNR was compared to the Z-detector's and to the reference method in the DCLDE community (that requires annotation information). Our method matched the reference in the presence of ship noise but overestimated the SNR ($+[5-7]$ dB) on less noisy parts. The Z-detector mostly underestimated the SNR ($-[5-7]$ dB).

Performance analysis also helped to demonstrate the SMF limitations. Detections were missed when the estimated value of the input SNR $\rho[k]$ was equal to zero or negative. This situation occurred when a call was embedded entirely in short duration noises, but with the same bandwidth, e.g., seismic event or high-intensity fin whale pulses. However, this combination of events was not common in the dataset and might be hard to detect even for a trained analyst. Source of error might also be related to remaining energy in the ABW unit A frequency band

due to echoes making the dissociation of singular calls complicated. On the tested data, ship noise did not cover this frequency band. However, if that had happened, it would have prevented detections or trigger a false alarm.

Finally, **Chapter 5** dealt with the detection and separation of concurrent calls from different species. The proposed strategy was based on a pattern recognition scheme. A first analysis was proposed to determine how to detect BW tonal signals, responsible for long-range propagation and, often used for calls characterization. Different tonal detection methods (instantaneous frequency estimator, Yin estimator, HPS, Cost-function-based detector, and ridge detector) were investigated and tested on a ground-truth dataset generated by embedding synthetic ABW calls in noise recordings for an OBS, with controlled SNR values. Performances of these different methods were evaluated using *Silbido* scoring metrics: Precision, Recall, Frequency deviation, Coverage, Excess, Precision, and Recall. Results showed that the ridge detector satisfied all criteria for subsequent classification (high Precision and Recall, satisfying Coverage and low-frequency deviation) in addition to being able to detect concurrent tonal signals. The elaboration of performances evaluation and method comparison strategies for passive acoustic is another contribution of this thesis.

Once a tonal detection method was selected, the following steps of pattern recognition consisted in the extraction of TF-amplitude features to characterize the tonal signals and, their classification using GMM clustering on their projection onto a reduced number of PCs (2). The method was trained on more than 4000 annotated units with overall precision and recall of respectively 97.2% and 96.9%, showing reliable classification. Notably, despite being close in frequency, P-calls and ABW unit A were correctly identified, with less than 0.5% error between classes. However, the MPBW unit 2 DS class was problematic due to the fragmentation of the detected signal. It led to the misclassification of other close units into its cluster.

A first unsupervised application to a recording containing MPBW calls and P-calls showed promising results for the proposed method. The last step of the process, the reconstruction of the classified signals was then applied, providing a waveform for each class of signals and therefore, reconstruction of each underlying song. Even if preliminary examples of song reconstruction require further work, they constitute the last contribution of this thesis. Our analysis of the developed methods and obtained results led to interesting observations; some of which could be lines for future work and, are detailed in the following paragraphs.

Perspectives for the Stochastic Matched Filter

The SMF was extended to the passive context in Section 3.5.1.1 under considerations of a time-varying linear filter for SNR maximization. This filter, was composed of sub-band filters that can be designed *offline* and, the adaptation to the passive context relied on the *online* estimation of noise covariance matrix and, the estimate the time-varying SNR.

To design a filter bank that can be applied to a large class of observations with no *a priori* knowledge on the instantaneous variations of the noise, the chosen approach was to resort to the simulation of low-frequency synthetic sea-colored noise (white noise with -6 dB/decade) to estimate the noise covariance matrix $\mathbf{\Gamma}_{\boldsymbol{n}_0}$ and solve the GEP. Even though this simulated noise was sufficient for OBS recorded data, it might be of interest to consider "a statistical" approach, for example, when SOFAR-recorded data are considered. The chorus has more impacts on this type of data and consequently, SMF performances could be improved with specific SNR maximizing sub-band filters.

Another topic for future work is the signal's covariance matrix $\mathbf{\Gamma}_{\boldsymbol{s}_0}$ in the design of the filters. In the developed method, it was estimated based on a parametric model of the signal of interest. However, recent studies have shown a global decay and seasonal variations in the frequency of BW calls, ABW included (−0.14 Hz/year) (McDonald et al., 2009; Leroy et al., 2018). As illustrated in Figure 3.3, the bandwidth of the maximum filter (applied when $Q[k] = Qmax$) is wide around ABW peak frequencies. However, if the frequency decay persists, the re-evaluation of the signal parameters would be necessary. For an improved use of the method, an additional "learning" step could be added, learning the parametric model values from a set of selected signals. Another option would be the automatic analysis of the detected signals, to learn directly from the incoming signal and provide auto-adaptative parametric values (this would, hence, require a regular actualization of the filter bank).

Based on the work from Xerri and Borloz (2004); Borloz and Xerri (2011); Chagmani et al. (2017), an additional study could be conducted to evaluate the optimal number of filters for SNR maximization in an automatic way. In addition to spectrogram set up, the last user-defined parameter is the size of the median filter for the noise estimation (median absolute deviation). The size of such filter could be automatically estimated from knowledge on the signal's duration and the spectrogram parameters.

Finally, regarding the application of the SMF, an additional study could be conducted to precisely evaluate the detection range (and gain from using the SMF). Such information is crucial for both of the presented applications: for ABW it intervenes in density estimation and, for scuba-divers, it is essential to perform threat detection at the earliest.

Perspectives for the automatic transcription

Results of the automatic transcription method were presented in Section 5.4.2. This experiment focused on a short (< 17 min) recording containing MPBW and P-calls, used for unsupervised testing. Next step of this work would be to test the method on a new, extended annotated dataset and, the evaluation of the classification performances (confusion matrix) on this test dataset.

Regarding the classification performances on the training data, lower miss-classification on broader units such as MPBW unit 2 DS, could be achieved with less fragmentation. One

of the options would be the improvement of the detection method to increase robustness to reverberated sound. However, this could be fulfilled only by imposing a detection threshold or, by introducing less strict rules in amplitude variation separation to connect units and reverberated signals. These solutions are not ideal: the first would reject lower intensity signals such as echoes but also remote calls and, the second would impact the classification because tonal and reverberated signals do not necessarily share the same features. The last option to improve classification precision and, to increase its robustness to reverberations is to apply a probability threshold to the GMM clusters. It would delimit the confidence region for the assignment of a tonal signal to the distribution. In this case, tonals out of the predefined boundaries could automatically be discarded. Note that other set of features (e.g., cepstral coefficients) or different classifiers (e.g., machine learning) could be considered. However, the interest of the presented approach is the low complexity of each step, where all parameters are understandable and can be checked.

The separation of songs of individuals from the same species could improve the utility of the transcription algorithm in visual and aural analyses. It could subsequently be used to study individual behavior (acoustic behavior, communication, tracking), count, and, on a further extent, improve density estimations. Consideration of ICI (Stafford et al., 2011) in the signal reconstruction step of the algorithm could help achieve individuals song separation. Besides, other options might be taken into account, i.e., received level comparison and echoes analysis.

Perspectives for performance analysis

Systematic performance analysis of detection and classification methods developed for the passive context is crucial. For example, estimating the detection probability within a specific detection range is essential for call density estimation. In addition, performance scores based on a statistical evaluation of methods outcome, can be used for reliable and unbiased comparison between methods. Comparison is essential to make a knowledgeable choice between the available detection techniques to address a specific application. However, performance evaluation should be conducted preferably on ground-truth data, resulting from the annotations of experienced analysts. Indeed, recordings are more representative of soundscapes diversity and the effects of the propagation canal on received signals.

The problem is that performance evaluation suffers from the few ground-truth datasets available (especially for BW low-frequency signals) and the lack of standardized procedures for assessment and method comparison. Besides, it is often impaired by the difficulty and variability of the annotation process. This is why initiatives to provide freely available annotations such as offered by the DCLDE community should be encouraged as well as collaborative annotation actions such as the currently developed Aplose, the online multi-analyst annotation platform of the Ocean Data Explorer project [5].

[5] https://oceandataexplorer.org

Traditional performance evaluation metrics (e.g., precision and recall) do not suffice in quantifying the performance of transcription methods. The evaluation of the performances requires to determine new scoring metrics to assess the accuracy of the transcription or, the percentage of transcribed songs, compare its efficiency on different types of signals or, evaluate units associations in a multi-individuals context. Perception experiences could be conducted to evaluate the quality of transcribed signals. However, for human auditory analysis, they require the transposition of the signals to a higher frequency range. Another approach consists in relying on simulated signals. This procedure could be completed with a method already used for odontocetes whistle detection and, part of the *Silbido*[6] software: it consists in comparing point by point the original and reconstructed signals in the TF domain (Roch et al., 2011). However, such dataset does not yet exist for BW signals.

Aware of the difficulties of cross-fields collaborative efforts required by PAM and, to generalize the use of automatic detection methods, the community should support (or continue to support) the development of comprehensive methods with thorough explanations, examples of the parameters to set and, shared code; the setting of standardized procedures for performance evaluation and, the development of comparison datasets to favor exchanges on the existing methods.

[6] https://roch.sdsu.edu/index.php/software/

Bibliography

Adam, O., Cazau, D., Gandilhon, N., Fabre, B., Laitman, J. T., and Reidenberg, J. S. (2013). New acoustic model for humpback whale sound production. *Applied Acoustics*, 74(10):1182–1190. 21, 26

Adam, O., Cazau, D., Samaran, F., Laitman, J. T., and Reidenberg, J. S. (2018). Anatomical evidence of blue whale tonal and pulsed sounds. *Personal communication*. 21, 26

Arveson, P. T. and Vendittis, D. J. (2000). Radiated noise characteristics of a modern cargo ship. *J. Acoust. Soc. Am.*, 107(1):118–129. 42

Babacan, O., Drugman, T., d'Alessandro, N., Henrich, N., and Dutoit, T. (2013). A comparative study of pitch extraction algorithms on a large variety of singing sounds. In *Proc. ICASSP*, pages 7815–7819. IEEE. 107

Balcazar, N. E., Klinck, H., Nieukirk, S. L., Mellinger, D. K., Klinck, K., Dziak, R. P., and Rogers, T. L. (2017). Using calls as an indicator for Antarctic blue whale occurrence and distribution across the Southwest Pacific and Southeast Indian Oceans. *Marine Mammal Science*, 33(1):172–186. 46

Balcazar, N. E., Tripovich, J. S., Klinck, H., Nieukirk, S. L., Mellinger, D. K., Dziak, R. P., and Rogers, T. L. (2015). Calls reveal population structure of blue whales across the Southeast Indian Ocean and Southwest Pacific Ocean. *Journal of Mammalogy*, 96(6):1184. 21, 29, 66, 73

Barlow, D. R., Torres, L. G., Hodge, K. B., Steel, D., Baker, C. S., Chandler, T. E., Bott, N., Constantine, R., Double, M. C., Gill, P., et al. (2018). Documentation of a New Zealand blue whale population based on multiple lines of evidence. *Endangered Species Research*, 36:27–40. 12

Barruol, G. and Sigloch, K. (2013). Investigating La Réunion hot spot from crust to core. *Eos, Transactions American Geophysical Union*, 94(23):205–207. 37

Barruol, G., Sigloch, K., and group, R.-R. (2017). RHUM-RUM experiment, 2011-2015, code YV (Réunion Hotspot and Upper Mantle - Réunion's Unterer Mantel) funded by ANR, DFG, CNRS-INSU, IPEV, TAAF, instrumented by DEPAS, INSU-OBS, AWI and the Universities of Muenster, Bonn, La Réunion. 37

Baumgartner, M. F., Bonnell, J., Van Parijs, S. M., Corkeron, P. J., Hotchkin, C., Ball, K., Pelletier, L.-P., Partan, J., Peters, D., Kemp, J., et al. (2019). Persistent near real-time passive acoustic monitoring for baleen whales from a moored buoy: system description and evaluation. *Methods in Ecology and Evolution*. 16

Baumgartner, M. F., Fratantoni, D. M., Hurst, T. P., Brown, M. W., Cole, T. V., Van Parijs, S. M., and Johnson, M. (2013). Real-time reporting of baleen whale passive acoustic detections from ocean gliders. *J. Acoust. Soc. Am.*, 134(3):1814–1823. 18

Baumgartner, M. F. and Mussoline, S. E. (2011). A generalized baleen whale call detection and classification system. *J. Acoust. Soc. Am.*, 129(5):2889–2902. 30, 31, 108, 109, 118

Baumgartner, M. F., Stafford, K. M., and Latha, G. (2018). Near real-time underwater passive acoustic monitoring of natural and anthropogenic sounds. In Venkatesan, R., Tandon, A., D'Asaro, E., and Atmanand, M. A., editors, *Observing the Oceans in Real Time*, pages 203–226. Springer International Publishing, Cham. 16, 18

Bénard, F., Glotin, H., and Giraudet, P. (2011). Highly defined whale group tracking by passive acoustic stochastic matched filter. *Advances in Sound Localization*. 64, 75, 88

Benesty, J., Chen, J., Huang, Y., and Cohen, I. (2009). *Noise reduction in speech processing*, volume 2. Springer Science & Business Media. 65, 67

Binder, C. M. and Hines, P. (2012). Applying automatic aural classification to cetacean vocalizations. *Proceedings of Meetings on Acoustics ECUA2012*, 17(1):070029. 31

Binder, C. M. and Hines, P. C. (2019). Range-dependent impacts of ocean acoustic propagation on automated classification of transmitted bowhead and humpback whale vocalizations. *J. Acoust. Soc. Am.*, 145(4):2480–2497. 32, 61

Bittle, M. and Duncan, A. (2013). A review of current marine mammal detection and classification algorithms for use in automated passive acoustic monitoring. In *Proceedings of Acoustics*, volume 2013. 29

Boashash, B. (1992). Estimating and interpreting the instantaneous frequency of a signal. II. Algorithms and applications. *Proc. of the IEEE*, 80(4):540–568. 108

Bonnal, J., Danes, P., and Renaud, M. (2010). Detection of acoustic patterns by stochastic matched filtering. In *Int. Conf. Intelligent Robots and Systems*, pages 1970–1975. 67

Borloz, B. and Xerri, B. (2011). Subspace SNR maximization: The constrained stochastic matched filter. *IEEE Transactions on Signal Processing*, 59(4):1346–1355. 67, 130

Bouffaut, L. and Boudraa, A.-O. (2017). Passive stochastic matched filter: Application to scuba divers detection. In *ICASSP 2017*, USA. 81

Bouffaut, L., Dréo, R., Labat, V., and Boudraa, A.-O. (2017a). Filtrage adapté stochastique passif pour la détection de plongeurs. In *GRETSI 2017*, France. 81

Bouffaut, L., Dréo, R., Labat, V., Boudraa, A.-O., and Barruol, G. (2017b). Antarctic blue whale calls detection based on an improved version of the stochastic matched filter. In *EUSIPCO 2017*, Greece. 75

Bouffaut, L., Dréo, R., Labat, V., Boudraa, A.-O., and Barruol, G. (2018). Passive stochastic matched filter for Antarctic blue whale call detection. *J. Acoust. Soc. Am.*, 144(2):955–965. 31, 66, 73, 88

Bouffaut, L., Madhusudhana, S., Labat, V., Boudraa, A.-O., and Klinck, H. (2019). Automated blue whale song transcription across variable acoustic contexts. In *OCEANS 2019*, Marseille, France. 124

Bouffaut, L., Madhusudhana, S., Labat, V., Boudraa, A.-O., and Klinck, H. (2020). A performance comparison of tonal detectors for low frequency vocalizations of Antarctic blue whales. *(Accepted) J. Acoust. Soc. Am.* 117

Brodie, D. C. and Dunn, R. A. (2015). Low frequency baleen whale calls detected on ocean-bottom seismometers in the Lau basin, Southwest Pacific Ocean. *J. Acoust. Soc. Am.*, 137(1):53–62. 17, 21, 28

Cannat, M., Sauter, D., Mendel, V., Ruellan, E., Okino, K., Escartin, J., Combier, V., and Baala, M. (2006). Modes of seafloor generation at a melt-poor ultraslow-spreading ridge. *Geology*, 34(7):605–608. 56

Cato, D. H. and McCauley, R. D. (2002). Australian research in ambient sea noise. *Acoustics Australia*, 30(1):13–20. 45, 51

Caudal, F. and Glotin, H. (2008). Stochastic Matched Filter outperforms Teager-Kaiser-Mallat for tracking a plurality of sperm whales. *IEEE PASSIVE 2008*, pages 1–9. 64, 67

Cavassilas, J.-F., Xerri, B., and Borloz, B. (1997). Filtre adapté stochastique. Contribution à la détection de textures bidimensionnelles. In *GRETSI 1997*, pages 491–494. 71

Cavassillas, J.-F. (1991). Stochastic matched filter. *Int. Conf. Sonar Signal Processing*, 13, part. 9:194–199. 67

Cavassillas, J.-F. and Xerri, B. (1993). Extension de la notion de filtre adapté. Contribution à la détection de signaux courts en présence de termes perturbateurs. *Traitement du Signal*, 10(3):215–221. 67, 70

Cazau, D., Adam, O., Aubin, T., Laitman, J. T., and Reidenberg, J. S. (2016). A study of vocal nonlinearities in humpback whale songs: from production mechanisms to acoustic analysis. *Scientific reports*, 6:31660. 21, 26

Cazau, D., Adam, O., Laitman, J. T., and Reidenberg, J. S. (2013). Understanding the intentional acoustic behavior of humpback whales: a production-based approach. *J. Acoust. Soc. Am.*, 134(3):2268–2273. 26

Cerchio, S., Andrianantenaina, B., Lindsay, A., Rekdahl, M., Andrianarivelo, N., and Rasoloarijao, T. (2015). Omura's whales (balaenoptera omurai) off Northwest Madagascar: ecology, behaviour and conservation needs. *Royal Society open science*, 2(10):150301. 21, 28

Chagmani, M., Xerri, B., Borloz, B., and Jauffret, C. (2017). The constrained stochastic matched filter subspace tracking. In *ISPA 2017*, pages 205–209. 130

Chapp, E., Bohnenstiehl, D. R., and Tolstoy, M. (2005). Sound-channel observations of ice-generated tremor in the Indian Ocean. *Geochemistry, Geophysics, Geosystems*, 6(6). 44

Clark, C., Ellison, W., and Beeman, K. (1986). Acoustic tracking of migrating bowhead whales. In *OCEANS'86*, pages 341–346. IEEE. 18

Cole, T. and, C. K. (2019). *NEFSC Right Whale Aerial Survey*. NOOA. 12

Comon, P. and Jutten, C., editors (2010). *Handbook of Blind Source Separation: Independent component analysis and applications*. Academic press. 120

Courmontagne, P. (1999). Le filtrage adapté stochastique adaptatif appliqué aux images bruitées. *Traitement du Signal*, 16(4):303–318. 67

Courmontagne, P. (2010). The stochastic matched filter and its applications to detection and de-noising. In *Stochastic Control*. InTech. 67, 68, 70, 71, 75, 88

Courmontagne, P. and Cavassilas, J.-F. (1997). Une nouvelle écriture du développement de Karhunen-Loève - Application à l'interpolation d'images. In *GRETSI 1997*, pages 455–458. 67

Courmontagne, P. and Chaillan, F. (2006). The adaptative stochatic matched filter for SAS images denoising. In *OCEANS 2006*. IEEE. 67

Courmontagne, P., Julien, G., and Bouhier, M. (2010). An improvement to the pulse compression scheme. In *OCEANS 2010 IEEE - Sydney*, pages 1–5. 67, 69, 76

Cubaynes, H. C., Fretwell, P. T., Bamford, C., Gerrish, L., and Jackson, J. A. (2019). Whales from space: Four mysticete species described using new VHR satellite imagery. *Marine Mammal Science*, 35(2):466–491. 13

Cummings, W. C. and Thompson, P. O. (1971). Underwater sounds from the blue whale, balaenoptera musculus. *J. Acoust. Soc. Am.*, 50(4B):1193–1198. 14, 21, 27

Curtis, K. R., Howe, B. M., and Mercer, J. A. (1999). Low-frequency ambient sound in the North Pacific: Long time series observations. *J. Acoust. Soc. Am.*, 106(6):3189–3200. 56

De Cheveigné, A. and Kawahara, H. (2002). Yin, a fundamental frequency estimator for speech and music. *J. Acoust. Soc. Am.*, 111(4):1917–1930. 108, 109

De La Cuadra, P., Master, A. S., and Sapp, C. (2001). Efficient pitch detection techniques for interactive music. In *International Computer Music Conference - ICMC 2001*. 108, 112

Delarue, J., Todd, S. K., Van Parijs, S. M., and Di Iorio, L. (2009). Geographic variation in Northwest Atlantic fin whale (balaenoptera physalus) song: Implications for stock structure assessment. *J. Acoust. Soc. Am.*, 125(3):1774–1782. 21, 50

Donskoy, D. M., Sedunov, N. A., Sedunov, A. N., and Tsionskiy, M. A. (2008). Variability of scuba diver's acoustic emission. In *Optics and Photonics in Global Homeland Security IV*, volume 6945, page 694515. International Society for Optics and Photonics. 81, 84

Dréo, R., Bouffaut, L., Guillon, L., Labat, V., Boudraa, A.-O., and Barruol, G. (2017). Antarctic blue whale localization with ocean bottom seismometer in Southern Indian Ocean. In *UACE 2017*, Greece. 56

Dréo, R., Bouffaut, L., Leroy, E., Barruol, G., and Samaran, F. (2019). Baleen whale distribution and seasonal occurrence revealed by an ocean bottom seismometer network in the Western Indian Ocean. *Deep Sea Research Part II: Topical Studies in Oceanography*, 161:132–144. 17, 33, 37, 45, 47, 48, 54, 55, 56, 89, 112

Duda, R. O., Hart, P. E., and Stork, D. G. (2012). *Pattern classification*. John Wiley & Sons, second edition. 118, 120

Duennebier, F. K., Blackinton, G., and Sutton, G. H. (1981). Current-generated noise recorded on ocean bottom seismometers. *Marine Geophysical Researches*, 5(1):109–115. 44

Dugan, P. J., Rice, A. N., Urazghildiiev, I. R., and Clark, C. W. (2010). North Atlantic right whale acoustic signal processing: Part I. comparison of machine learning recognition algorithms. In *2010 IEEE Long Island Systems, Applications and Technology Conference*, pages 1–6. IEEE. 31

Dunn, R. A. and Hernandez, O. (2009). Tracking blue whales in the Eastern Tropical Pacific with an ocean-bottom seismometer and hydrophone array. *J. Acoust. Soc. Am.*, 126(3):1084–1094. 16, 55

Fillinger, L., Hunter, A., Zampolli, M., and Clarijs, M. (2012). Passive acoustic detection of closed-circuit underwater breathing apparatus in an operational port environment. *J. Acoust. Soc. Am.*, 132(4):EL310–EL316. 81

Frank, S. D. and Ferris, A. N. (2011). Analysis and localization of blue whale vocalizations in the Solomon Sea using waveform amplitude data. *J. Acoust. Soc. Am.*, 130(2):731–736. 15, 16, 55

Fretwell, P. T., Staniland, I. J., and Forcada, J. (2014). Whales from space: counting southern right whales by satellite. *PLoS One*, 9(2):e88655. 13

Garcia-Rojas, M. I., Jenner, K. C. S., Gill, P. C., Jenner, M.-N. M., Sutton, A. L., and McCauley, R. D. (2018). Environmental evidence for a pygmy blue whale aggregation area in the subtropical convergence zone south of Australia. *Marine Mammal Science*, 34(4):901–923. 18, 21

Gavrilov, A. N. and McCauley, R. D. (2013). Acoustic detection and long-term monitoring of pygmy blue whales over the continental slope in Southwest Australia. *J. Acoust. Soc. Am.*, 134(3):2505–2513. 17, 29, 48, 76, 88

Gavrilov, A. N., McCauley, R. D., Salgado-Kent, C., Tripovich, J., and Burton, C. (2011). Vocal characteristics of pygmy blue whales and their change over time. *J. Acoust. Soc. Am.*, 130(6):3651–3660. 17, 27, 30, 31

Gedamke, J. (2009). Geographic variation in Southern Ocean fin whale song. *International Whaling Commission report SC/61/SH16*, pages 1–8. 50

Gedamke, J., Gales, N., Hildebrand, J., and Wiggins, S. (2007). Seasonal occurrence of low frequency whale vocalisations across Eastern Antarctic and Southern Australian waters, february 2004 to february 2007. *White paper presented to the Scientific Committee of the International Whaling Commission*. 21, 30

Gill, P. C. (2002). A blue whale (balaenoptera musculus) feeding ground in a southern Australian coastal upwelling zone. *Journal of Cetacean Research and Management*, 4(2):179–184. 12

Gillespie, D. (2004). Detection and classification of right whale calls using an 'edge' detector operating on a smoothed spectrogram. *Canadian Acoustics*, 32(2):39–47. 30, 31, 108

Goldbogen, J. A., Southall, B. L., DeRuiter, S. L., Calambokidis, J., Friedlaender, A. S., Hazen, E. L., Falcone, E. A., Schorr, G. S., Douglas, A., Moretti, D. J., et al. (2013). Blue whales respond to simulated mid-frequency military sonar. *Proceedings of the Royal Society B: Biological Sciences*, 280(1765):20130657. 22

Gomez, C., Lawson, J., Wright, A. J., Buren, A., Tollit, D., and Lesage, V. (2016). A systematic review on the behavioural responses of wild marine mammals to noise: the disparity between science and policy. *Canadian journal of zoology*, 94(12):801–819. 22

Guilment, T., Socheleau, F.-X., Pastor, D., and Vallez, S. (2018). Sparse representation-based classification of mysticete calls. *J. Acoust. Soc. Am.*, 144(3):1550–1563. 31

Hable, S., Sigloch, K., Barruol, G., Stähler, S. C., and Hadziioannou, C. (2018). Clock errors in land and ocean bottom seismograms: high-accuracy estimates from multiple-component noise cross-correlations. *Geophysical Journal International*, 214(3):2014–2034. 38

Halkias, X. C., Paris, S., and Glotin, H. (2013). Classification of mysticete sounds using machine learning techniques. *J. Acoust. Soc. Am.*, 134(5):3496–3505. 31

Hammond, P., Sears, R., and Berube, M. (1990). A note on problems in estimating the number of blue whales in the Gulf of St Lawrence from photo-identification data. *Rep. int. Whal. Commn*, pages 141–142. 12

Hanson, J. A. and Bowman, J. R. (2005). Dispersive and reflected tsunami signals from the 2004 Indian Ocean tsunami observed on hydrophones and seismic stations. *Geophysical Research Letters*, 32(17). 44

Harcourt, R., van der Hoop, J., Kraus, S., and Carroll, E. L. (2019). Future directions in eubalaena spp.: comparative research to inform conservation. *Frontiers in Marine Science*, 5:530. 21

Harland, E. J. and Armstrong, M. S. (2004). The real-time detection of the calls of cetacean species. *Canadian Acoustics*, 32(2):76–82. 30

Harris, C. M., Thomas, L., Falcone, E. A., Hildebrand, J., Houser, D., Kvadsheim, P. H., Lam, F.-P. A., Miller, P. J., Moretti, D. J., Read, A. J., et al. (2018). Marine mammals and sonar: Dose-response studies, the risk-disturbance hypothesis and the role of exposure context. *Journal of applied ecology*, 55(1):396–404. 22

Harris, D., Matias, L., Thomas, L., Harwood, J., and Geissler, W. H. (2013). Applying distance sampling to fin whale calls recorded by single seismic instruments in the Northeast Atlantic. *J. Acoust. Soc. Am.*, 134(5):3522–3535. 17

Haver, S. M., Gedamke, J., Hatch, L. T., Dziak, R. P., Van Parijs, S., McKenna, M. F., Barlow, J., Berchok, C., DiDonato, E., Hanson, B., et al. (2018). Monitoring long-term soundscape trends in US waters: The NOAA/NPS ocean noise reference station network. *Marine Policy*, 90:6–13. 51

Haver, S. M., Klinck, H., Nieukirk, S. L., Matsumoto, H., Dziak, R. P., and Miksis-Olds, J. L. (2017). The not-so-silent world: measuring Arctic, Equatorial, and Antarctic soundscapes in the Atlantic Ocean. *Deep Sea Research Part I: Oceanographic Research Papers*, 122:95–104. 45, 56

Hendricks, B., Keen, E. M., Wray, J. L., Alidina, H. M., McWhinnie, L., Meuter, H., Picard, C. R., and Gulliver, T. A. (2018). Automated monitoring and analysis of marine mammal vocalizations in coastal habitats. In *2018 OCEANS-MTS/IEEE Kobe Techno-Oceans (OTO)*, pages 1–10. IEEE. 16

Hildebrand, J. A. (2009). Anthropogenic and natural sources of ambient noise in the ocean. *Marine Ecology Progress Series*, 395:5–20. 40, 43

Irvine, L. M., Mate, B. R., Winsor, M. H., Palacios, D. M., Bograd, S. J., Costa, D. P., and Bailey, H. (2014). Spatial and temporal occurrence of blue whales off the US West coast, with implications for management. *PLoS One*, 9(7):e102959. 21

IWC (1980). Chairman's report of the thirty-first annual meeting. *Report of the International Whaling Commission*, page 30:25–41. 2

Jolliffe, C. D., McCauley, R. D., Gavrilov, A. N., Jenner, K. C. S., Jenner, M.-N. M., and Duncan, A. J. (2019). Song variation of the South Eastern Indian ocean pygmy blue whale population in the Perth Canyon, Western Australia. *PloS one*, 14(1):e0208619. 21

Josso, N. (2010). *Caractérisation des milieux sous marins en utilisant des sources mobiles d'opportunité*. PhD thesis, Institut National Polytechnique de Grenoble - INPG. 23, 61

Juennard, N. (2007). *Detection and localisation of very high energy particles in underwater acoustic*. PhD thesis, Université du Sud Toulon Var. 67, 71

Julien, G. (2012). *Filtrage adapté stochastique et amélioration des performances des systèmes de positionnement d'engins sous-marins en milieu bruyant*. PhD thesis, Université du Sud Toulon Var. 67, 72

Kershenbaum, A. and Roch, M. A. (2013). An image processing based paradigm for the extraction of tonal sounds in cetacean communications. *J. Acoust. Soc. Am.*, 134(6):4435–4445. 108, 109, 116

Kewley, D., Browning, D., and Carey, W. (1990). Low-frequency wind-generated ambient noise source levels. *J. Acoust. Soc. Am.*, 88(4):1894–1902. 45

Khalil, H. H., Rahmat, R. O., and Mahmoud, W. A. (2008). Estimation of noise in gray-scale and colored images using median absolute deviation (mad). In *2008 3rd International Conference on Geometric Modeling and Imaging*, pages 92–97. IEEE. 75

Krause, B. (2008). Anatomy of the soundscape: evolving perspectives. *Journal of the Audio Engineering Society*, 56(1/2):73–80. 40

Labat, V. and Daré, D. (2014). Analyse de signaux acoustiques marins: identification de fréquences caractéristiques via la méthode DEMON. *Congrès Français d'Acoustique (12; 2014; Poitiers)*, pages 345–350. 81

Lennartsson, R., Dalberg, E., Persson, L., and Petrović, S. (2009). Passive acoustic detection and classification of divers in harbor environments. *OCEANS 2009*, pages 1–7. 81

Leroy, E. (2017). *Signaux acoustiques de baleines bleues dans l'océan indien austral : traitement, analyse et interprétation des signaux*. PhD thesis, Université Bretagne Occidentale. 2, 5, 21, 55

Leroy, E. C., Royer, J.-Y., Bonnel, J., and Samaran, F. (2018). Long-term and seasonal changes of large whale call frequency in the Southern Indian Ocean. *Journal of Geophysical Research: Oceans*, 123(11):8568–8580. 17, 21, 107, 130

Leroy, E. C., Samaran, F., Bonnel, J., and Royer, J.-Y. (2016). Seasonal and diel vocalization patterns of Antarctic blue whale (balaenoptera musculus intermedia) in the Southern Indian Ocean: A multi-year and multi-site study. *PLOS ONE*, 11(11):1–20. 17, 45, 46, 52, 73, 91

Leroy, E. C., Samaran, F., Bonnel, J., and Royer, J.-Y. (2017a). Identification of two potential whale calls in the Southern Indian Ocean, and their geographic and seasonal occurrence. *J. Acoust. Soc. Am.*, 142(3):1413–1427. 21, 45, 49, 122

Leroy, E. C., Thomisch, K., and Van Opzeeland, I. (2017b). Variability in ground-truth data sets and the performance of two automated detectors for Antarctic blue whale calls in different soundscape conditions. *J. Acoust. Soc. Am.*, 141(5):3604–3605. 28, 32, 49, 51, 88, 101

Lewis, L. A., Calambokidis, J., Stimpert, A. K., Fahlbusch, J., Friedlaender, A. S., McKenna, M. F., Mesnick, S. L., Oleson, E. M., Southall, B. L., Szesciorka, A. R., et al. (2018). Context-dependent variability in blue whale acoustic behaviour. *Royal Society open science*, 5(8):180241. 19, 21, 53

Lin, T.-H., Chou, L.-S., Akamatsu, T., Chan, H.-C., and Chen, C.-F. (2013). An automatic detection algorithm for extracting the representative frequency of cetacean tonal sounds. *J. Acoust. Soc. Am.*, 134(3):2477–2485. 30

Madhusudhana, S. (commit 387c420 on 21/11/2018). Github ridge detector repository. 111

Madhusudhana, S., Gavrilov, A., and Erbe, C. (2016). A generic system for the automatic extraction of narrowband signals in underwater audio. *J. Acoust. Soc. Am.*, 140(4):3182–3182. 30

Madhusudhana, S. K. (2015). *Automatic Detectors for Underwater Soundscape Measurements*. PhD thesis, Curtin University. 108, 109, 116

Marques, T. A., Thomas, L., Martin, S. W., Mellinger, D. K., Ward, J. A., Moretti, D. J., Harris, D., and Tyack, P. L. (2013). Estimating animal population density using passive acoustics. *Biological Reviews*, 88(2):287–309. 18

Matsumoto, H., Bohnenstiehl, D. R., Tournadre, J., Dziak, R. P., Haxel, J. H., Lau, T.-K., Fowler, M., and Salo, S. A. (2014). Antarctic icebergs: A significant natural ocean sound source in the Southern hemisphere. *Geochemistry, Geophysics, Geosystems*, 15(8):3448–3458. 44, 45

Max, J. (1981). *Méthodes et techniques de traitement du signal et applications aux mesures physiques. 3eme édition*. Masson. 66, 76

McCauley, R. D., Gavrilov, A. N., Jolliffe, C. D., Ward, R., and Gill, P. C. (2018). Pygmy blue and Antarctic blue whale presence, distribution and population parameters in Southern Australia based on passive acoustics. *Deep Sea Research Part II: Topical Studies in Oceanography*, 157:154–168. 52, 66

McDonald, M. A., Hildebrand, J. A., and Mesnick, S. (2009). Worldwide decline in tonal frequencies of blue whale songs. *Endangered Species Research*, 9(1):13–21. 21, 130

McDonald, M. A., Hildebrand, J. A., and Webb, S. C. (1995). Blue and fin whales observed on a seafloor array in the Northeast Pacific. *J. Acoust. Soc. Am.*, 98(2):712–721. 16

McDonald, M. A., Mesnick, S. L., and Hildebrand, J. A. (2006). Biogeographic characterization of blue whale song worldwide: Using song to identify populations. *Journal of Cetacean Research and Management*, 8(1):55–65. 14, 21, 45, 48

Melcon, M. L., Cummins, A. J., Kerosky, S. M., Roche, L. K., Wiggins, S. M., and Hildebrand, J. A. (2012). Blue whales respond to anthropogenic noise. *PLoS One*, 7(2):e32681. 22

Mellinger, D. K. and Clark, C. W. (1997). Methods for automatic detection of mysticete sounds. *Marine & Freshwater Behaviour & Phy*, 29(1-4):163–181. 29, 31

Mellinger, D. K. and Clark, C. W. (2000). Recognizing transient low-frequency whale sounds by spectrogram correlation. *J. Acoust. Soc. Am.*, 107(6):3518–3529. 29, 88

Mellinger, D. K. and Clark, C. W. (2006). Mobysound: A reference archive for studying automatic recognition of marine mammal sounds. *Applied Acoustics*, 67(11-12):1226–1242. 72, 98, 99, 113

Miksis-Olds, J. L., Bradley, D. L., and Maggie Niu, X. (2013). Decadal trends in Indian Ocean ambient sound. *J. Acoust. Soc. Am.*, 134(5):3464–3475. 41, 42, 56

Miller, B. S., Calderan, S., Collins, K., Leaper, R., Kelly, N., Peel, D., Olson, P., Ensor, P., and Double, M. C. (2013). Long-range acoustic tracking of Antarctic blue whales. *NOAA*. 18

Miller, B. S., Calderan, S., Gillespie, D., Weatherup, G., Leaper, R., Collins, K., and Double, M. C. (2016). Software for real-time localization of baleen whale calls using directional sonobuoys: A case study on Antarctic blue whales. *J. Acoust. Soc. Am.*, 139(3):EL83–EL89. 18

Miller, B. S., Wotherspoon, S., Rankin, S., Calderan, S., Leaper, R., and Keating, J. L. (2018). Estimating drift of directional sonobuoys from acoustic bearings. *J. Acoust. Soc. Am.*, 143(1):EL25–EL30. 18

Moore, S. E., Howe, B. M., Stafford, K. M., and Boyd, M. L. (2007). Including whale call detection in standard ocean measurements: Application of acoustic seagliders. *Marine Technology Society Journal*, 41(4):53–57. 18

Mori, J.-L. and Gounon, P. (2000). The use of stochastic matched filter in active sonar. In *Signal Processing Conference, 2000 10th European*, pages 1–4. IEEE. 66, 67, 88

Mouy, X., Leary, D., Martin, B., and Laurinolli, M. (2008). A comparison of methods for the automatic classification of marine mammal vocalizations in the Arctic. In *2008 New Trends for Environmental Monitoring Using Passive Systems*, pages 1–6. IEEE. 30, 31

Munk, W. H., Spindel, R. C., Baggeroer, A., and Birdsall, T. G. (1994). The heard island feasibility test. *J. Acoust. Soc. Am.*, 96(4):2330–2342. 17

Naguib, M. and Wiley, R. H. (2001). Estimating the distance to a source of sound: mechanisms and adaptations for long-range communication. *Animal behaviour*, 62(5):825–837. 53

Nieukirk, S. L., Fregosi, S., Mellinger, D. K., and Klinck, H. (2016). A complex baleen whale call recorded in the mariana trench marine national monument. *J. Acoust. Soc. Am.*, 140(3):EL274–EL279. 18, 19

Noll, A. M. (1969). Pitch determination of human speech by the harmonic product spectrum, the harmonic sum spectrum, and a maximum likelihood estimate. In *Symposium on Computer Processing in Communication, ed.*, volume 19, pages 779–797. University of Broodlyn Press, New York. 108, 109

Norris, T., Oswald, J., and Sousa-Lima, R. (2010). A review and inventory of fixed installation passive acoustic monitoring methods and technologies. *Final report to the International Association of Oil and Gas Producers Joint Industry Programme on E&P Sound and Marine Life*. 16

Nowacek, D. P., Christiansen, F., Bejder, L., Goldbogen, J. A., and Friedlaender, A. S. (2016). Studying cetacean behaviour: new technological approaches and conservation applications. *Animal behaviour*, 120:235–244. 12, 20

Pace, F., White, P., and Adam, O. (2012). Hidden markov modeling for humpback whale (megaptera novaeanglie) call classification. *Proceedings of Meetings on Acoustics ECUA2012*, 17(1):070046. 31

Payne, R. and Webb, D. (1971). Orientation by means of long range acoustic signaling in baleen whales. *Annals of the New York Academy of Sciences*, 188(1):110–141. 15, 27

Payne, R. S. and McVay, S. (1971). Songs of humpback whales. *Science*, 173(3997):585–597. 13, 14

Pijanowski, B. C., Villanueva-Rivera, L. J., Dumyahn, S. L., Farina, A., Krause, B. L., Napoletano, B. M., Gage, S. H., and Pieretti, N. (2011). Soundscape ecology: the science of sound in the landscape. *BioScience*, 61(3):203–216. 40, 53

Porter, M. B. and Bucker, H. P. (1987). Gaussian beam tracing for computing ocean acoustic fields. *J. Acoust. Soc. Am.*, 82(4):1349–1359. 25

Putland, R. L., Merchant, N. D., Farcas, A., and Radford, C. A. (2018). Vessel noise cuts down communication space for vocalizing fish and marine mammals. *Global change biology*, 24(4):1708–1721. 22

Rankin, S., Miller, B., Crance, J. L., Sakai, T., and Keating, J. L. (2019). Sonobuoy acoustic data collection during cetacean surveys. *NOAA Technical Memorandum NMFS*. 18

Redfern, J. V., Hatch, L. T., Caldow, C., DeAngelis, M. L., Gedamke, J., Hastings, S., Henderson, L., McKenna, M. F., Moore, T. J., and Porter, M. B. (2017). Assessing the risk of chronic shipping noise to baleen whales off Southern California, USA. *Endangered Species Research*, 32:153–167. 22

Reidenberg, J. S. and Laitman, J. T. (2007). Discovery of a low frequency sound source in mysticeti (baleen whales): anatomical establishment of a vocal fold homolog. *The Anatomical Record*, 290(6):745–759. 26

Reidenberg, J. S. and Laitman, J. T. (2008). Sisters of the sinuses: cetacean air sacs. *The Anatomical Record*, 291(11):1389–1396. 26

Roch, M. A., Scott Brandes, T., Patel, B., Barkley, Y., Baumann-Pickering, S., and Soldevilla, M. S. (2011). Automated extraction of odontocete whistle contours. *J. Acoust. Soc. Am.*, 130(4):2212–2223. 30, 113, 132

Ross, D. (1976). *Mechanics of Underwater Noise*. Pergamon. 42

Rousseeuw, P. J. and Croux, C. (1993). Alternatives to the median absolute deviation. *Journal of the American Statistical association*, 88(424):1273–1283. 75

Samaran, F., Adam, O., and Guinet, C. (2010a). Detection range modeling of blue whale calls in Southwestern Indian Ocean. *Applied Acoustics*, 71(11):1099–1106. 45, 66

Samaran, F., Adam, O., and Guinet, C. (2010b). Discovery of a mid-latitude sympatric area for two Southern hemisphere blue whale subspecies. *Endangered Species Research*, 12(2):157–165. 21, 47

Samaran, F., Adam, O., Motsch, J.-F., and C., G. (2008). Definition of the Antarctic and pygmy blue whale call templates. Application to fast automatic detection. *Canadian Acoustics*, 36(1):98–103. 29, 73

Samaran, F., Guinet, C., Adam, O., Motsch, J.-F., and Cansi, Y. (2010c). Source level estimation of two blue whale subspecies in Southwestern Indian Ocean. *J. Acoust. Soc. Am.*, 127(6):3800–3808. 21, 27, 55, 73

Samaran, F., Stafford, K. M., Branch, T. A., Gedamke, J., Royer, J.-Y., Dziak, R., and Guinet, C. (2013). Seasonal and geographic variation of Southern blue whale subspecies in the Indian Ocean. *PLoS ONE, Public Library of Science*, 8(8):e71561. 5, 17, 21, 29, 45, 54, 55

Schafer, R. M. (1980). *The tuning of the world: Toward a theory of soundscape design*. University of Pennsylvania Press Philadelphia, PA, USA. 39

Schevill, W. E. (1964). Underwater sounds of cetaceans. *Marine bio-acoustics*, 1:307–316. 14

Schmid, F. and Schlindwein, V. (2016). Microearthquake activity, lithospheric structure, and deformation modes at an amagmatic ultraslow spreading Southwest Indian ridge segment. *Geochemistry, Geophysics, Geosystems*, 17(7):2905–2921. 39, 44

Schreiber, O. (1952). Some sounds from marine life in the Hawaiian area. *J. Acoust. Soc. Am.*, 24(1):116–116. 14

Schroeder, M. R. (1968). Period histogram and product spectrum: New methods for fundamental-frequency measurement. *J. Acoust. Soc. Am.*, 43(4):829–834. 109

Shabangu, F. W., Yemane, D., Stafford, K. M., Ensor, P., and Findlay, K. P. (2017). Modelling the effects of environmental conditions on the acoustic occurrence and behaviour of Antarctic blue whales. *PloS one*, 12(2):e0172705. 29

Širović, A., Hildebrand, J. A., and Wiggins, S. M. (2007). Blue and fin whale call source levels and propagation range in the Southern Ocean. *J. Acoust. Soc. Am.*, 122(2):1208–1215. 16, 21, 26

Širović, A., Hildebrand, J. A., Wiggins, S. M., McDonald, M. A., Moore, S. E., and Thiele, D. (2004). Seasonality of blue and fin whale calls and the influence of sea ice in the Western Antarctic Peninsula. *Deep Sea Research Part II: Topical Studies in Oceanography*, 51(17):2327–2344. 16, 21, 30, 46, 51, 66

Širović, A., Oleson, E. M., Buccowich, J., Rice, A., and Bayless, A. R. (2017). Fin whale song variability in Southern California and the Gulf of California. *Scientific reports*, 7(1):10126. 21

Slamnoiu, G., Radu, O., Rosca, V., Pascu, C., Damian, R., Surdu, G., Curca, E., and Radulescu, A. (2016). DEMON-type algorithms for determination of hydro-acoustic signatures of surface ships and of divers. *IOP Conference Series: Materials Science and Engineering*, 145(8):082013. 81

Slămnoiu, G., Radu, O., Roşca, V., Pascu, C., Surdu, G., Curcă, E., Damian, R., and Rădulescu, A. (2017). Risk factors detection for strategic importance objectives in littoral areas. *IOP Conference Series: Materials Science and Engineering*, 209(1):012099. 81

Socheleau, F.-X., Leroy, E., Carvallo Pecci, A., Samaran, F., Bonnel, J., and Royer, J.-Y. (2015). Automated detection of Antarctic blue whale calls. *J. Acoust. Soc. Am.*, 138(5):3105–3117. 5, 30, 73, 75, 88, 91, 94, 98, 113

Socheleau, F.-X. and Samaran, F. (2017). Detection of mysticete calls: A sparse representation-based approach. *Unpublished*. 31

Stafford, K. M., Bohnenstiehl, D. R., Tolstoy, M., Chapp, M., Mellinger, D. K., and Moore, S. E. (2004). Antarctic-type blue whale calls recorded at low latitudes in the Indian and Eastern Pacific Oceans. *Deep Sea Research Part I: Oceanographic Research Papers*, 51(10):1337 – 1346. 73

Stafford, K. M., Chapp, E., Bohnenstiel, D. R., and Tolstoy, M. (2011). Seasonal detection of three types of "pygmy" blue whale calls in the Indian Ocean. *Marine Mammal Science*, 27(4):828–840. 21, 29, 48, 55, 131

Stähler, S. C., Sigloch, K., Hosseini, K., Crawford, W. C., Barruol, G., Schmidt-Aursch, M. C., Tsekhmistrenko, M., Scholz, J.-R., Mazzullo, A., and Deen, M. (2016). Performance report of the RHUM-RUM ocean bottom seismometer network around La Réunion, western Indian Ocean. *Advances in Geosciences*, 41:43–63. 17, 37, 38, 39

Stimpert, A. K., DeRuiter, S. L., Falcone, E. A., Joseph, J., Douglas, A. B., Moretti, D. J., Friedlaender, A. S., Calambokidis, J., Gailey, G., Tyack, P. L., et al. (2015). Sound production and associated behavior of tagged fin whales (balaenoptera physalus) in the Southern California Bight. *Animal Biotelemetry*, 3(1):23. 19, 20, 53

Stolkin, R., Sutin, A., Radhakrishnan, S., Bruno, M., Fullerton, B., Ekimov, A., and Raftery, M. (2006). Feature based passive acoustic detection of underwater threats. *Proc. SPIE*, 6204:620408–620410. 81

Sutton, G. H. and Barstow, N. (1990). Ocean-bottom ultralow-frequency (ULF) seismo-acoustic ambient noise: 0.002 to 0.4 Hz. *J. Acoust. Soc. Am.*, 87(5):2005–2012. 44

Talandier, J., Hyvernaud, O., Reymond, D., and Okal, E. A. (2006). Hydroacoustic signals generated by parked and drifting icebergs in the Southern Indian and Pacific Oceans. *Geophysical Journal International*, 165(3):817–834. 44, 45

Thode, A. M., Kim, K. H., Blackwell, S. B., Greene Jr, C. R., Nations, C. S., McDonald, T. L., and Macrander, A. M. (2012). Automated detection and localization of bowhead whale sounds in the presence of seismic airgun surveys. *J. Acoust. Soc. Am.*, 131(5):3726–3747. 30, 31

Thomisch, K. (2017). Distribution patterns and migratory behavior of Antarctic blue whales. *Berichte zur Polar-und Meeresforschung= Reports on polar and marine research*, 707. 21, 46

Trabattoni, A., Barruol, G., Dreo, R., Boudraa, A., and Fontaine, F. (2020). Orienting and locating ocean-bottom seismometers from ship noise analysis. *Geophysical Journal International*, 220(3):1774–1790. 38

Tripovich, J. S., Klinck, H., Nieukirk, S. L., Adams, T., Mellinger, D. K., Balcazar, N. E., Klinck, K., Hall, E. J., and Rogers, T. L. (2015). Temporal segregation of the Australian and Antarctic blue whale call types (balaenoptera musculus spp.). *Journal of mammalogy*, 96(3):603–610. 21, 30

Trygonis, V., Gerstein, E., Moir, J., and McCulloch, S. (2013). Vocalization characteristics of North Atlantic right whale surface active groups in the calving habitat, Southeastern United States. *J. Acoust. Soc. Am.*, 134(6):4518–4531. 31

Urazghildiiev, I. R., Clark, C. W., Krein, T. P., and Parks, S. E. (2009). Detection and recognition of North Atlantic right whale contact calls in the presence of ambient noise. *IEEE Journal of Oceanic Engineering*, 34(3):358–368. 30, 31, 118

Urick, R. J. (1983). Principles of underwater sound 3rd edition. *Peninsula Publising Los Atlos, California*. 22, 23, 24, 27, 28

Van Trees, H. L. (2002). *Optimum Array Processing: Part IV of Detection, Estimation, and Modulation Theory*. John Wiley and Sons, Inc. 29, 66, 88, 91

Walker, R. A. (1963). Some intense, low-frequency, underwater sounds of wide geographic distribution, apparently of biological origin. *J. Acoust. Soc. Am.*, 35(11):1816–1824. 14

Ward, R., Gavrilov, A. N., and McCauley, R. D. (2017). "Spot" call: A common sound from an unidentified great whale in Australian temperate waters. *J. Acoust. Soc. Am.*, 142(2):EL231–EL236. 16, 17, 21, 45, 49, 122

Weirathmueller, M. J., Stafford, K. M., Wilcock, W. S., Hilmo, R. S., Dziak, R. P., and Tréhu, A. M. (2017). Spatial and temporal trends in fin whale vocalizations recorded in the NE Pacific Ocean between 2003-2013. *PloS one*, 12(10):e0186127. 29

Weirathmueller, M. J., Wilcock, W. S., and Soule, D. C. (2013). Source levels of fin whale 20 Hz pulses measured in the Northeast Pacific Ocean. *J. Acoust. Soc. Am.*, 133(2):741–749. 21

Wenz, G. M. (1962). Acoustic ambient noise in the ocean: Spectra and sources. *J. Acoust. Soc. Am.*, 34(12):1936–1956. 40, 41

Wilcock, W. S. D., Stafford, K. M., Andrew, R. K., and Odom, R. I. (2014). Sounds in the ocean at 1–100 Hz. *Annual review of marine science*, 6:117–140. 45

Williams, R., Clark, C., Ponirakis, D., and Ashe, E. (2014). Acoustic quality of critical habitats for three threatened whale populations. *Animal Conservation*, 17(2):174–185. 16, 22

Worzel, J. L., Ewing, M., and Pekeris, C. L. (1948). *Propagation of Sound in the Ocean*. Geological Society of America. 17

Xerri, B. and Borloz, B. (2004). Detection by SNR maximization: Application to the blind source separation problem. In *International Conference on Independent Component Analysis and Signal Separation*, pages 602–609. Springer. 67, 130

Zimmer, W. M. (2011). *Passive acoustic monitoring of cetaceans*. Cambridge University Press. 11, 15, 21

Širović, A., McDonald, M., Balcazar-Cabrera, N., Buchan, S., Cerchio, S., Clark, C., Davis, G., Findlay, K., Gagnon, G., Kyo, N., Leroy, E., Miksis-Olds, J., Miller, B., Oleson, E., Pangerc, T., TH, R., Samaran, F., Simard, Y., Stafford, K., Stevenson, D., Sugioka, H., Tomish, K., Tripovich, J., Truong, G., Van Opzeeland, I., Van Parijs, S., Yoshida, R., and Brownell Jr, R. (2017). Blue whale songs worldwide: an update. In *22nd Biennial Conference of the Biology of Marine Mammals*. Society for Marine Mammalogy. 14

UNIVERSITE BRETAGNE LOIRE / SCIENCES DE LA MER ET DU LITTORAL

UBO Université de Bretagne Occidentale

Titre : Détection et classification dans un contexte acoustique passive : application à la détection des signaux basse-fréquences des baleines bleues

Mots clés : Détection automatique, filtrage adapté stochastique, performances de détection, transcription sonore, monitoring par acoustique passive, baleine bleue.

Résumé : L'analyse des grands volumes de données générés par la surveillance par acoustique passive long-terme et continue des baleines bleues (BW) est améliorée par la détection automatisée des signaux d'intérêt. Le travail présenté dans cette thèse s'attaque au problème de la détection et classification de signaux stéréotypés dans un contexte passif basse fréquence où les signaux sont modifiés par le canal de propagation, bruités et où le SNR varie continuellement. Les méthodes développées sont appliquées à des enregistrements issus d'OBS déployés dans l'océan Indien occidental.

Premièrement, le filtrage adapté stochastique (SMF) est étendu au contexte passif en adaptant l'estimation du bruit et du SNR. Ce filtre est appliqué avec succès pour la détection des calls de baleine bleue antarctique et est comparé aux MF et Z-detector sur données annotées présentant de nombreux bruits et d'importantes variations du SNR. Les excellentes performances du SMF passif permettent d'augmenter la portée de détection jusqu'à 100 km en présence de bruit de bateau.

La détection simultanée de différentes espèces s'appuie sur un schéma de reconnaissance de formes où les signaux tonaux de BW sont extraits, caractérisés et classifiés pour la transcription automatique des chants. Les signaux ainsi identifiés sont ensuite reconstruits avec des formes d'onde distinctes reproduisant les chants sous-jacents. Le succès de la reconstruction repose sur la qualité de la détection de tonales: le détecteur de crêtes est choisi pour son efficacité. Les résultats d'apprentissage et la première application non supervisée de la transcription ont révélé des résultats prometteurs et son utilité pour l'analyse multi-espèces.

Title: Detection and classification in passive acoustic context: Application to blue whale low-frequency signals

Keywords: Automatic detection, stochastic matched filter, detection performances, sound transcription, passive acoustic monitoring, blue whale.

Abstract: The analysis of the large volumes of data resulting from continuous and long-term monitoring efforts of blue whales (BWs) benefits from the automated detection of target signals. This thesis investigates the challenging problem of the detection and classification of stereotyped signals in a low-frequency passive acoustic context where (1) signals traveling long distances are deteriorated by the propagation channel, (2) overlapping noises interfere and, (3) SNRs vary continuously. Developed methods are applied to recordings from ocean bottom seismometers deployed in the western Indian Ocean.

First, the stochastic matched filter (SMF) is adapted to the passive context by overcoming noise estimation and estimating the SNR automatically. This filter is successfully applied to the detection of Antarctic blue whales calls and is compared to the MF and the Z-detector on an annotated ground-truth dataset exhibiting various SNRs and noises. The passive SMF showed better performances, increasing the detection range up to 100 km in the presence of ship noise.

The problematic of the detection of concurrently calling species is addressed based on a pattern recognition development for the automatic transcription of BW songs where, tonal signals are extracted, characterized, and classified. The hence identified signals are then reconstructed as separate waveforms reconstructing of the underlying songs. The success of the reconstruction relies on the quality of the tonal detector: the ridge detector was chosen for its efficiency. Training and unsupervised application revealed promising results of the proposed transcription method and its utility for multi-species analysis.