

# Détection robuste de signaux acoustiques de mammifères marins

Florian Dadouchi

► **To cite this version:**

Florian Dadouchi. Détection robuste de signaux acoustiques de mammifères marins. Sciences de la Terre. Université de Grenoble, 2014. Français. NNT : 2014GRENU059 . tel-01099566v2

**HAL Id: tel-01099566**

**<https://tel.archives-ouvertes.fr/tel-01099566v2>**

Submitted on 30 Jun 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE

Pour obtenir le grade de

**DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE**

Spécialité : **Terre, Univers, Environnement**

Arrêté ministériel : 7 Août 2006

Présentée par

**Florian Dadouchi**

Thèse dirigée par **Jérôme I. Mars**

et codirigée par **Cédric Gervaise**

et par **Cornel Ioana**

préparée au sein du laboratoire **Grenoble Images Parole Signal Automatique (GIPSA-Lab)**

et de l'école doctorale **Terre Univers Environnement**

**Détection robuste de signaux  
acoustiques de mammifères marins**

Thèse soutenue publiquement le **8 Octobre 2014**,  
devant le jury composé de :

**M. Patrick Flandrin**

Directeur de Recherche, ENS de Lyon, CNRS, Président

**Mme. Marie A. Roch**

Professeur, San Diego State University , Rapporteur

**M. Ljubiša Stanković**

Professeur, University of Montenegro, Rapporteur

**M. Julien Huillery**

Maître de conférence, École Centrale de Lyon, Examinateur

**M. Dominique Fattaccioli**

Ingénieur Ph.D., DGA Techniques Navales, Examinateur

**M. Jérôme I. Mars**

Professeur, Grenoble INP, GIPSA-Lab, Directeur de thèse

**M. Cédric Gervaise**

Chercheur, Chaire CHORUS, GIPSA-Lab, , Co-Directeur de thèse

**M. Cornel Ioana**

Maître de conférence, Grenoble-INP, GIPSA-Lab, Co-Directeur de thèse

**Mme. Odile Gérard**

Ingénieur Ph.D., DGA Techniques Navales, Invité





UNIVERSITÉ DE GRENOBLE  
ECOLE DOCTORALE TUE

Terre - Univers - Environnement

THÈSE

pour obtenir le titre de

Docteur en Sciences

de l'université de Grenoble

Mention : Océan, Atmosphère, Hydrologie

présentée et soutenue par :

Florian DADOUCHI

Détection robuste de signaux acoustiques  
de mammifères marins

Thèse dirigée par Jérôme I. Mars

co-dirigée par Cédric Gervaise

et par Cornel Ioana

préparée au laboratoire Grenoble, Images, Parole, Signal,  
Automatique (GIPSA-Lab)

soutenue le 8 Octobre 2014

**Jury :**

Président :	M. Patrick Flandrin	-	ENS de Lyon, CNRS
Rapporteurs :	Mme. Marie A. Roch	-	San Diego State University
	M. Ljubiša Stanković	-	University of Montenegro
Examineurs :	M. Julien Huillery	-	Laboratoire Ampère, École Centrale de Lyon
	M. Dominique Fattaccioli	-	DGA Techniques Navales
Directeur :	Jérôme I. Mars	-	GIPSA-Lab, Grenoble INP
Co-directeurs :	M. Cédric Gervaise	-	Chaire CHORUS
	M. Cornel Ioana	-	GIPSA-Lab, Grenoble INP
Invitée :	Mme. Odile Gérard	-	DGA Techniques Navales



# Remerciements

À l'aube des vacances, je dois remercier. Avant tout, je remercie le mauvais temps de l'été 2014, mauvais temps qui a considérablement réduit toute tentation de fuite, m'obligeant par la force des choses à écrire ce manuscrit. Nimbostratus, je ne vous remercierai jamais assez.

Ces remerciements nuageux je les adresse aussi bien évidemment à mon jury de thèse pour avoir accepté de lire mes travaux, de s'être levé aussi tôt (beaucoup trop pour certains) ou déplacé d'aussi loin pour la soutenance : Londres, Toulon, et même Lyon pour certains !

Je remercie. Mes directeurs de thèse, Jérôme, Cédric et Cornel. Jérôme, je t'ai donné bien du mal, et fais faire quelques cauchemars, je te délivre à présent de ce fardeau, vogue en paix ! Merci Cédric à toi aussi. Cornel, on se refait un séjour au pays du soleil levant quand tu veux !

Je remercie. Mr. Joseph Fourier, sans qui rien de tout ceci n'aurait été possible.

Je remercie. Cindy pour nous avoir toujours tenu au courant de la météo en temps-réel, pour garder le chauffage au chaud, et pour les multiples coinches. Manuel, que l'on nomme MfCC dans le milieu, un toulousain aux multiples features. Quentin pour ses blagues plus élaborées que les miennes. Céline pour les nombreuses discussions théoriques autour du très prometteur hyperspectre de Meillier. Romain, mon psychiatre personnel du lundi matin, co-bureau de toujours, partenaire de brainstorming continu (l'attaque de la mélasse a de l'avenir). Florian Dadouchi d'avoir toujours été là pour moi. Florian (un autre), mon psychanalyste quand j'avais besoin d'aide. Lucas, pour m'avoir autant laissé empiéter sur ton bureau et de n'avoir jamais rien dit. La team des coincheurs assidus du midi. Nikola, pour qui mon français est resté plus que obscur pendant longtemps. Antoine pour m'avoir permis de faire un grand nombre de siestes lors de ses expériences oculaires. Tim, mon service info personnel (ne le dites à personne). L'autre Tim, même s'il ne comprends rien en info. L'ensemble des baleines, dauphins et orques qui ont accepté de me parler pendant trois ans. Alexis, pour sa maîtrise parfaite

du bleu, mais aussi du rouge. Rémy pour ses fonctions Vertes. Taia mon compagnon d'exil tahitien. Buchoud, pour son (bref mais intense) costard de soutenance. Aude, pour son hypothétique Buchoud de soutenance. Robin, pour aider Aude à trouver un Buchoud de soutenance. Guillaume et Pascal, va aussi falloir les aider. Olivier Le Bot pour avoir patiemment accepté de démêler les clics. Francesca de m'avoir appris que le latin n'était pas complètement une langue morte. Fakhri pour ta gentillesse. Lucia BT pour m'avoir toujours donné la possibilité de rentrer dans mon bureau, mes tentatives de m'en auto-exclure seront donc restées vaines jusqu'au bout. Je remercie Olivier Michel, mon encyclopédie personnelle, maintes fois ouverte lors de mes errements scientifiques. Mais aussi Gabriel, François et Jocelyn pour toutes ces interactions diverses et variées.

Je remercie. Gières, SMLV, Humbert II, Foch, Fontaine, Lesduiguière et l'île verte dans leur ensemble. Vous êtes trop nombreux pour être cités, l'encre coûte cher, pourtant vous êtes les premiers d'entre tous à me supporter (au sens propre). Aziza, Selma et Ryad, des voisins pas comme les autres. All the people from Victoria and Seattle, who will probably never read this. Le G2E-Lab de m'avoir hébergé quand je fuyais le GIPSA.

Je remercie. Ma famille. Mon père et ma mère sans qui je ne serais pas là (encore une fois au sens propre). Mes frères d'être venus par surprise à Grenoble pour endurer, pieds nus, ma soutenance.

Je remercie. Les gens que je n'ai pas encore eu l'occasion de rencontrer. Ça ne saurait tarder.

A tout ceux que j'ai oublié dans ces remerciements, dites-vous que c'était volontaire.







# Introduction

Notre connaissance des océans est extrêmement limitée, ils couvrent 71% de la surface de la planète mais on estime cependant que 95% restent inexplorés. Nous sommes pourtant redevables envers les océans pour de nombreuses raisons. Pour des raisons climatiques d'abord, les océans participent en effet à la régulation du climat en étant les principaux vecteurs de transfert de la chaleur de l'équateur vers les pôles, ILS jouent également le rôle de puit de carbone naturel majeur et produisent la moitié de l'oxygène mondial. Pour des raisons biologiques également, car abritant une très large biodiversité, importante en elle-même, mais aussi source nutritionnelle pour un très grand nombre d'habitants de la planète.

Ces océans, loin d'être protégés - 99% des océans ne le sont pas - sont soumis à de très fortes pressions anthropiques : pollution physico-chimique, pêche industrielle non raisonnée, forte augmentation du niveau sonore d'origine anthropique et développement des zones côtières.

La pêche intensive draine les océans de toute vie marine, menaçant l'équilibre des écosystèmes et mettant en danger les populations, et les millions de pêcheurs à travers le monde vivant principalement des ressources offertes par les océans. Au sujet de la perte de la biodiversité dans les écosystèmes engendrée par la surpêche, un article de Worm *et. al* [WBB<sup>+</sup>06] est conclu de manière très alarmiste :

*"This trend is of serious concern because it projects the global collapse of all taxa currently fished by the mid 21st century [...]. Our findings further suggest that the elimination of locally adapted populations and species not only impairs the ability of marine ecosystems to feed a growing human population but also sabotages their stability and recovery potential in a rapidly changing marine environment"*

Avant de poursuivre :

*"By restoring marine biodiversity through sustainable fisheries management, pollution control, maintenance of essential habitats, and the*

*creation of marine reserves, we can invest in the productivity and reliability of the goods and services that the ocean provides to humanity. Our analyses suggest that business as usual would foreshadow serious threats to global food security, coastal water quality, and ecosystem stability, affecting current and future generations"*

Les conséquences de la surpêche sont sérieuses, mais ne sont pourtant pas la seule menace qui pèse sur les écosystèmes marins. L'eutrophisation des milieux côtiers, c'est à dire la réponse de l'écosystème à l'apport massif de substances telles que les nitrates et phosphates, par les fertilisants et eaux usées en particulier, crée des "zones mortes". Cet apport excessif de nutriments favorise en effet la croissance d'algues et favorise l'activité microbienne, consommant ainsi l'oxygène dissout dans l'eau. Plus de 245,000 kilomètres carrés seraient affectés [DR08].

Au delà de la perte de biodiversité et de toutes ses implications, le rôle de puit de carbone joué par les océans pourrait être altéré de manière dramatique par le réchauffement climatique [SHSM98].

Une autre modification essentielle de l'environnement est liée à l'augmentation continue du niveau de bruit dans les océans. Ainsi, en raison de l'augmentation importante du trafic maritime côtier, l'impact de l'activité humaine sur la faune sous-marine est devenu un sujet d'intérêt majeur afin d'éviter d'affecter trop fortement les écosystèmes (intérêt écologique) et/ou les destructions matérielles (intérêt économique). D'autre part, le nombre important d'opérations militaires en environnement côtier implique une prise en compte judicieuse de l'impact d'émissions SONAR sur la vie des mammifères marins car les signaux émis sont souvent caractérisés par des puissances élevées et des formes d'onde définies dans les mêmes bandes de fréquence que celles utilisées par les mammifères marins. Bruits de trafic maritime et émissions impulsives sont les deux pressions qui ont été identifiées dans le descripteur 11 : 'Bruit sous marin' de la Directive Cadre sur la Stratégie du Milieu Marin (DCSMM, UE2009). En conséquence, l'analyse des signaux de mammifères marins permettrait d'obtenir des informations sur l'impact des activités militaires sur le comportement des animaux marins en vue de préparer des actions de mitigation. Ainsi, une première motivation de cette étude repose sur la protection de l'environnement sous-marin. En effet, plusieurs fois, lors d'exercices SONAR, des mammifères marins (baleine à bec), ont été retrouvés échoués. Suite à ce phénomène, la DGA a ouvert un PEA Impact Sonar mammifères Marins où des études sur la détection et la classification sont nécessaires. Finalement, la simple augmentation du niveau de bruit moyen peut avoir des impacts insidieux et plus difficilement mesurables sur l'ensemble de l'écosystème : accroissement des niveaux de stress, abandon d'un habitat important, masquage ou réduction des

sons naturels, changement dans le comportement vocal, diminution de l'aptitude à chasser ou à éviter les prédateurs, et réduction des occasions d'accouplement [Wei07]. Cette thèse pourrait servir à des études faites par des organismes de recherche civiles pour l'écologie marine et la cétologie par exemple.

## Contexte et objectif scientifique de la thèse :

Ce travail de thèse, financé par une bourse DGA, s'inscrit dans un contexte d'étude des océans par acoustique passive (PAM)<sup>1</sup>. L'acoustique passive se place désormais comme une pierre angulaire pour la compréhension de la dynamique des écosystèmes marins. Les mammifères marins, en plus d'être très emblématiques de l'environnement marin auprès du grand public, se placent haut dans la chaîne alimentaire, et sont à ce titre à la fois des maillons très importants du réseau trophique marin et des éléments fragiles, car en compétition directe avec l'industrie de la pêche.

L'utilisation abondante qu'ils font des sons, que ce soit pour communiquer ou chasser fait de l'acoustique passive un allié de grande valeur pour les détecter, les localiser et estimer leur nombre.

Un des verrous techniques actuels limitant le traitement automatisé de grandes bases de données sur les mammifères marins concerne la difficulté à détecter efficacement les sons qu'ils émettent. Ce travail de thèse se concentre précisément sur cet aspect du problème.

La non-stationnarité des signaux émis rend naturelle l'utilisation de représentations temps-fréquence pour l'interprétation et le traitement automatique de ces sons. Parmi toutes les représentations nous choisissons le spectrogramme.

La partie technique de ce manuscrit sera donc dédié à la détection statistique d'évènements acoustiques au sein du spectrogramme. Le schéma directeur de ce travail de thèse concerne donc l'automatisation autant que se peut d'une segmentation efficace des signatures acoustiques émis par les mammifères marins à partir de tests statistiques. Derrière ce schéma, l'idée principale véhiculée dans ce manuscrit est la suivante : le nombre de fausses alarmes engendrées par la résolution d'un test d'hypothèse binaire sur l'énergie du spectrogramme est prévisible si le test est résolu à probabilité de fausse alarme constante. Le nombre de détections dans des régions du plan binaire peut alors être utilisée pour discriminer les régions abritant le support temps-fréquence du signal de celles n'abritant que du bruit.

---

1. Passive Acoustic Monitoring

## Organisation du document :

Dans le **premier chapitre** nous décrivons les enjeux de l'acoustique passive ainsi que les difficultés inhérentes à ce problème. Nous y décrivons les différents contributeurs acoustiques dans les océans : la géophonie, la biophonie et l'anthropophonie. Cette polyphonie océanique est composée de sources par nature intermittentes et les paysages acoustiques marins seront donc très divers. Une implication à cette description du paysage acoustique concerne la nécessité d'estimer localement toutes les grandeurs spectrales, localement au niveau temporel, fréquentiel et spatial. Ce constat rend l'utilisation de représentation temps-fréquence naturelle. Nous en choisissons une - le spectrogramme - et nous discutons la pertinence de ce choix.

Le **chapitre 2** est une prolongation des travaux de Julien Huillery sur la détection sur spectrogramme (travaux réalisés au GIPSA-Lab). Nous discutons de la description des statistiques du spectrogramme. Les coefficients du spectrogramme issus d'un bruit Gaussien sont modélisés par une loi du  $\chi^2$  à deux degrés de liberté. Nous montrons que l'estimation de cette loi est équivalent à estimer la densité spectrale de puissance du bruit. La coloration et la non-stationnarité du bruit Gaussien rendent nécessaire l'estimation de cette grandeur en chaque point temps-fréquence du spectrogramme. La connaissance de la densité de probabilité des coefficients du bruit dans le spectrogramme permet de résoudre un test d'hypothèse binaire en utilisant un critère de Neyman-Pearson.

**Les principales contributions de ce travail de thèse** sont présentées à partir du **chapitre 3**. Dans ce chapitre nous présentons les limitations du détecteur quadratique (i.e. le détecteur d'énergie sur le spectrogramme) en terme de performance à bas rapport-signal-à-bruit (RSB) ( $\leq 8dB$ ). Nous proposons alors une méthode, basée sur l'analyse de la densité des fausses alarmes, que nous appelons FADA (pour False Alarm Density Analysis) et qui permet de surmonter les limitations du détecteur classique. Nous cherchons à discriminer les régions temps-fréquence abritant le support temps-fréquence du signal de celles n'abritant que du bruit. Nous basons cette analyse sur le nombre de détections dans des régions du plan binaire. Les fausses alarmes sont générées par un test d'hypothèse résolu à probabilité de fausse alarme constante, nous interprétons alors le spectrogramme binaire comme un ensemble de réalisations de variables de Bernoulli. Le spectrogramme binaire est découpé en régions temps-fréquence sur lesquelles les détections sont comptées et la distribution de ce nombre aléatoire dans le cas "bruit seul" est modélisée par une loi binomiale. Un second test d'hypothèse portant sur ce nombre de détections est résolu, à nouveau avec un critère de Neyman-Pearson.

Le **chapitre 4** est un double raffinement du chapitre 3. Nous redéfinissons d'abord le découpage et le comptage des détections dans le spectrogramme binaire comme une convolution bidimensionnelle du spectrogramme binaire et d'une

matrice uniforme. Cette approche rend le problème invariant temporellement et modifie légèrement la philosophie du problème. En particulier le test d'hypothèse du chapitre 3 prend une décision sur une région du plan temps-fréquence, alors que le test d'hypothèse découlant de la nouvelle définition prend une décision sur un point temps-fréquence en tenant compte de son voisinage. Le second raffinement concerne la description de la densité de probabilité utilisée. La loi binomiale du chapitre 3 suppose l'indépendance des variables de Bernoulli. La corrélation intrinsèque du spectrogramme rend cette approximation non tenable dans la plupart des cas. La vraie loi est binomiale corrélée et ne possède pas d'expression analytique explicite. Nous proposerons une estimation numérique de cette loi.

Le **chapitre 5** discute du choix des paramètres de la méthode présentée et en évalue numériquement les performances. Finalement la méthode globale de détection des sifflements de mammifères marins à bas RSB et en milieu fortement impulsif est présenté.

Notre objectif principal dans cet thèse est de séparer et de détecter de manière automatique les sons harmoniques et impulsifs rencontrés dans les enregistrements acoustiques. La méthode sera validée sur de nombreux exemples réels de signaux de mammifères marins. Nous exploiterons en particulier trois bases de données. La première est issue de la campagne de mesure **ERATO-09** dans le golfe de Gascogne en Septembre 2009 lors de laquelle la présence de dauphins communs *delphinus delphis* a été constatée. La seconde base de données est issue de **MobySound** (<http://www.mobysound.org/>). Cette base de données, fournie par le Cooperative Institute for Marine Resources Studies (Oregon, USA), est librement accessible. Nous utiliserons tout particulièrement les données d'odontocètes rendues accessibles à l'occasion du 5ème Workshop on Detection, Classification, Localization and Density Estimation of Marine Mammals using Passive Acoustics (2011, Oregon). Ces enregistrements contiennent, entre autres des sons de *delphinus delphis* et *tursiops truncatus*. La troisième base de données est issue d'enregistrement réalisés au large de l'archipel de **Molène**. Ces enregistrements contiennent des sons de *tursiops truncatus* ainsi qu'une forte activité d'origine benthique.

# Table des matières

<b>1</b>	<b>Du paysage acoustique sous-marin au monitoring par acoustique passive</b>	<b>18</b>
1.1	Paysage acoustique sous-marin . . . . .	19
1.2	Le monitoring par acoustique passive . . . . .	26
1.3	L'acoustique passive en pratique . . . . .	34
1.4	Représentations temps-fréquence . . . . .	39
1.5	Conclusion . . . . .	42
<b>2</b>	<b>Détection adaptative de signaux non-stationnaires</b>	<b>45</b>
2.1	Introduction . . . . .	46
2.2	Modèle du signal en temps . . . . .	46
2.3	Modèle du signal en temps-fréquence . . . . .	49
2.4	Densité de probabilité du spectrogramme sous l'hypothèse bruit seul	51
2.5	Estimation locale de la densité spectrale de puissance (DSP) du bruit	53
2.6	Résolution du test d'hypothèse binaire . . . . .	58
2.7	Exemples sur données réelles . . . . .	60
2.8	Conclusions et problématique . . . . .	62
<b>3</b>	<b>Étude statistique de l'organisation des maxima temps-fréquence</b>	<b>67</b>
3.1	Introduction . . . . .	68
3.2	Segmentation de composantes temps-fréquence . . . . .	70
3.3	Formulation du problème . . . . .	73
3.4	Degrés de liberté de la méthode . . . . .	79
3.5	Conclusion . . . . .	83
<b>4</b>	<b>Analyse de la densité des fausses alarmes (FADA)</b>	<b>89</b>
4.1	Introduction . . . . .	90
4.2	Approche convolutive du comptage . . . . .	91
4.3	Complément sur les statistiques sous $\mathcal{H}_0^R$ . . . . .	93
4.4	Conclusion . . . . .	103

<b>5</b>	<b>Analyse des performances en simulation, et validation en contexte réel</b>	<b>107</b>
5.1	Introduction . . . . .	108
5.2	Choix de la probabilité de fausse alarme $p_{fa}$ . . . . .	109
5.3	Forme de la région de comptage $\mathcal{R}$ . . . . .	118
5.4	Évaluation de la méthode . . . . .	125
5.5	Illustration sur données réelles . . . . .	131
5.6	Conclusion . . . . .	136
	<b>Conclusions et perspectives</b>	<b>149</b>



## Notations et acronymes/ Notations and acronyms

Français	English	Notation	Acronym
Transformée de Fourier à Court Terme	Short Time Fourier Transform		TFCT
Rapport signal à bruit	Signal to Noise Ratio		RSB
Spectrogramme	Spectrogram	$S_x[n, k]$	
Longueur de la fenêtre d'apodisation	Length of the window function	M	
Densité spectrale de puissance	Power Spectral Density	$\gamma[n, k]$	DSP
Voisinage d'estimation de la DSP du bruit	Neighborhood of estimation of the noise PSD	V	
Test d'hypothèse binaire	Binary hypothesis test		BHT
Hypothèses BHT1	Hypotheses BHT1	$\mathcal{H}_0, \mathcal{H}_1$	
Hypothèses BHT2	Hypotheses BHT2	$\mathcal{H}_0^{\mathcal{R}}, \mathcal{H}_1^{\mathcal{R}}$	
Probabilité de fausse alarme pour BHT1	False alarm probability BHT1	$p_{fa}$	
Seuil sur le spectrogramme	Threshold on the spectrogram	$S_{p_{fa}}[n, k]$	
Région de comptage	Counting region	$\mathcal{R}$	
Probabilité de fausse alarme pour BHT2	False alarm probability for BHT2	$p_{fa}^{\mathcal{R}}$	
Support temps-fréquence du signal	Time-frequency support of the signal	$\supset_s$	
Spectrogramme binaire	Counting region	$\Phi_{p_{fa}} = \{\phi_{p_{fa}}\}$	
Carte de comptage	Counting map	$\Psi_{p_{fa}} = \{\psi_{p_{fa}}\}$	
Seuil sur la carte de comptage	Counting map threshold	$\Psi_T$	



# Chapitre 1

## Du paysage acoustique sous-marin au monitoring par acoustique passive

### Sommaire

---

<b>1.1</b>	<b>Paysage acoustique sous-marin . . . . .</b>	<b>19</b>
1.1.1	Géophonie . . . . .	19
1.1.2	Anthropophonie . . . . .	19
1.1.3	Biophonie . . . . .	22
<b>1.2</b>	<b>Le monitoring par acoustique passive . . . . .</b>	<b>26</b>
1.2.1	Bénéfices de l'acoustique passive . . . . .	30
1.2.2	Objectifs de l'acoustique passive . . . . .	33
<b>1.3</b>	<b>L'acoustique passive en pratique . . . . .</b>	<b>34</b>
1.3.1	Chaîne de traitement . . . . .	34
1.3.2	Quels enseignements tirer de la description du paysage acoustique ? . . . . .	36
1.3.3	Difficultés inhérentes à l'acoustique passive sous-marine	38
<b>1.4</b>	<b>Représentations temps-fréquence . . . . .</b>	<b>39</b>
<b>1.5</b>	<b>Conclusion . . . . .</b>	<b>42</b>

---

Dans ce chapitre nous décrivons le paysage acoustique sous marin et la diversité des signaux émis par les cétacés. Nous présentons les enjeux et les promesses liées à l'acoustique passive. Finalement, nous introduisons les outils d'analyses privilégiés qui seront utilisés tout au long de ce manuscrit.

## 1.1 Paysage acoustique sous-marin

En 1962, Wenz propose une description du paysage acoustique sous-marin [Wen62], cette description est reprise et affinée dans [Hil09]. Les influences respectives, en termes d'amplitudes, des différents contributeurs acoustiques y sont décrites. L'influence respective des différentes sources est donné en dB ref  $1\mu\text{Pa}$  (décibels référencés à  $1\mu\text{Pa}$ ). Trois classes contribuant au paysage acoustique sous-marin sont identifiées : la géophonie, la biophonie et l'anthropophonie. Un spectre générique tel que construit par Wenz est proposé en figure 1.1 (Source de la figure : [B<sup>+</sup>03])

**Remarque :** La référence de pression pour mesurer l'intensité d'un son sous la mer est  $1\mu\text{Pa}$ . Cette référence n'est pas la même dans le contexte d'acoustique aérienne où la référence est de  $20\mu\text{Pa}$ . À titre de comparaison les clics émis par les Bélugas ont une amplitude oscillant entre 206 et 225 dB (ref  $1\mu\text{Pa}$  @ 1m), dans l'air une amplitude de 202 dB (ref  $20\mu\text{Pa}$  @ 1m) provoque la mort d'un être humain. L'erreur d'unité est donc fatale.

### 1.1.1 Géophonie

La géophonie recouvre l'ensemble des sources acoustiques issues de processus physiques et géophysiques [Hil09, KSG<sup>+</sup>13]. Nous pouvons citer en exemple l'agitation de surface due au vent, aux vagues et à la pluie (100Hz-100kHz), les sources intermittentes comme les tremblements de terre ( $\leq 100\text{Hz}$ ) et les volcans sous-marins ( $\leq 100\text{Hz}$ ), les craquements et déformations de la glace (10Hz-20kHz) (voir figure 1.2), les éclairs (1kHz) et le bruit thermique d'agitation moléculaire ( $> 60\text{kHz}$ )

L'influence de ces sources se fait ainsi sentir sur l'ensemble de la bande de fréquence mesurable : du plus bas de l'échelle avec les tremblements de terre ( $\leq \text{Hz}$ ) au plus haut de l'échelle avec le bruit thermique ( $> 60\text{kHz}$ ).

### 1.1.2 Anthropophonie

L'anthropophonie regroupe l'ensemble des émissions d'origine humaine. Il existe un grand nombre de sources différentes [Hil09, SW07].

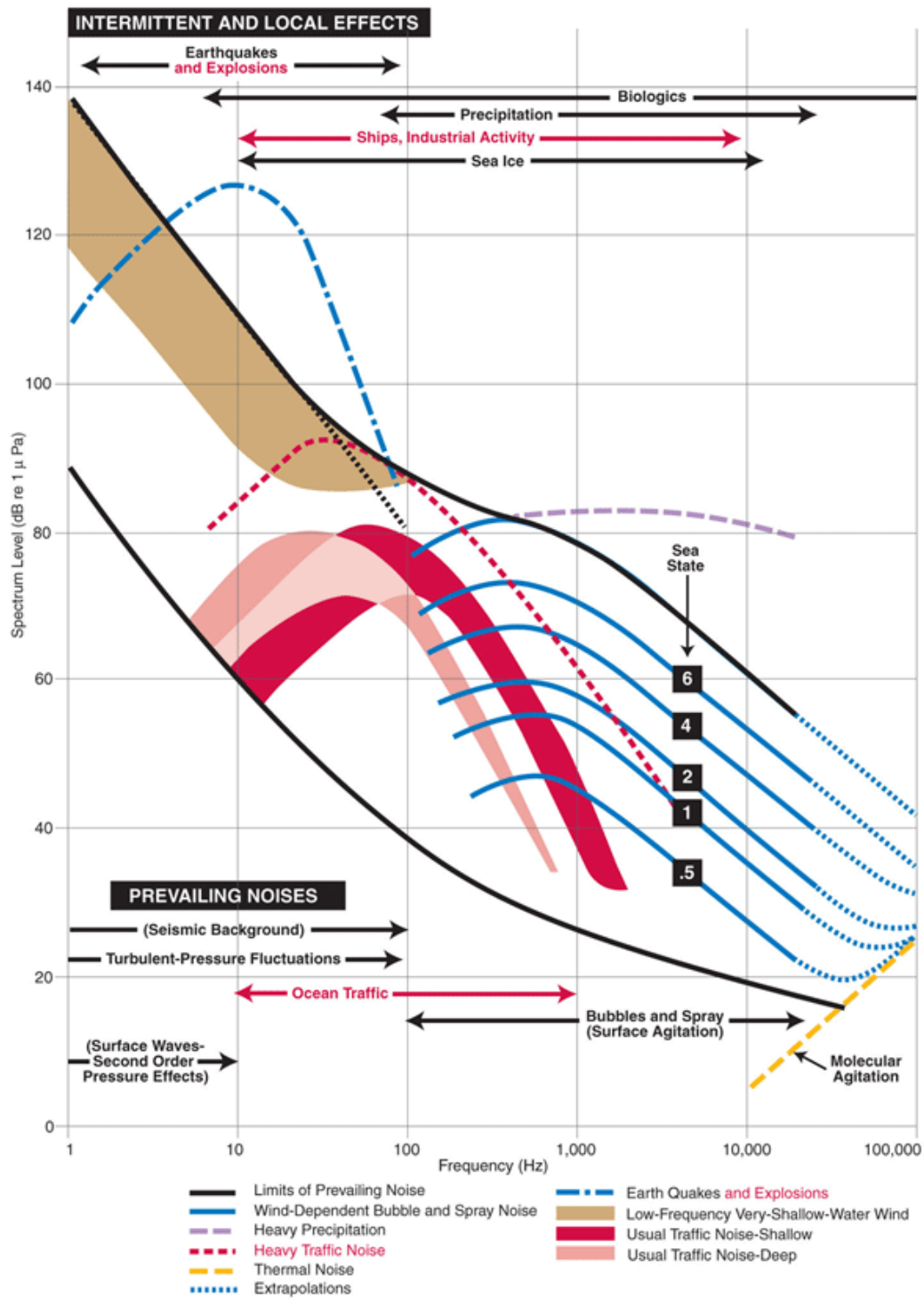


FIGURE 1.1 – Modélisation du bruit océanique et des différents contributeurs au spectre, tel que donné par Wenz [Wen62] (Source : [B<sup>+</sup>03]).

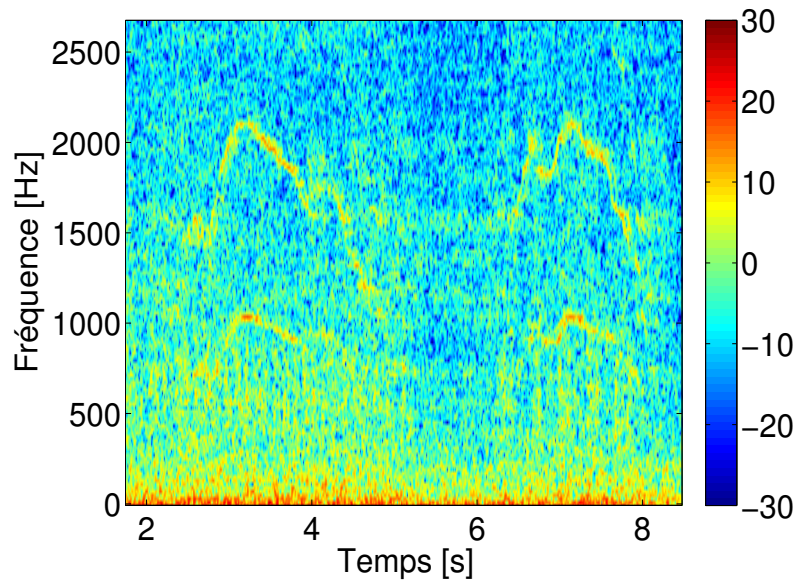


FIGURE 1.2 – Spectrogramme d’un son émis par la friction de blocs de glace dans l’antarctique.

Le **trafic maritime** est la principale source de bruit d’origine humaine. Le bruit rayonné par les navires dépend de leur taille et de leur poids (voir figure 1.3 montrant l’exemple du spectrogramme d’une embarcation légère). Il est en général plutôt bas en fréquence ( 10Hz-1kHz) et se propage donc bien dans le milieu marin. Les **communications sous-marines** ainsi que les **sonars** et sondeurs (la figure 1.4 montre une représentation schématique d’un sondeur en action) participent fortement à l’augmentation du niveau de bruit. Les sonars (sound navigation and ranging) sont utilisés dans un contexte militaire pour localiser les bâtiments présent dans l’eau et ont des puissances d’émission de l’ordre de 235 dB ref  $1\mu\text{Pa}$  @ 1m [Car06]. Les sonars sont aussi utilisés dans le domaine civil pour la navigation maritime ou pour cartographier le fond des océans, ils sont en général de forte amplitude.

Une autre source de bruit provient des activités de **prospection sismique** et d’analyse des couches sédimentaires. Des airguns sont classiquement utilisés pour créer des sources impulsionnelles de forte amplitude ( $\sim 260\text{dB}$  re  $1\mu\text{Pa}$  @ 1m). Les airguns relâchent des bulles d’air sous pression dans l’eau qui implosent. Les sons produits sont principalement dans la bande de fréquence 5Hz-300Hz et se propagent donc à longue distance [Hil09].

Une des plus fortes sources de bruit anthropique dans l’océan provient du battage (pile-driving). Cette méthode est entre autre utilisée pour installer les fondations de bâtiments au large des côtes comme les plateformes pétrolières ou les éoliennes offshore. Le contenu spectral du bruit de battage se situe dans une bande de fréquence inférieure à 10kHz avec un pic d’énergie se situant entre 100Hz et 2kHz. Le niveau de bruit reçu peut aller jusqu’à 237dB ref  $1\mu\text{Pa}$  @ 1m, un

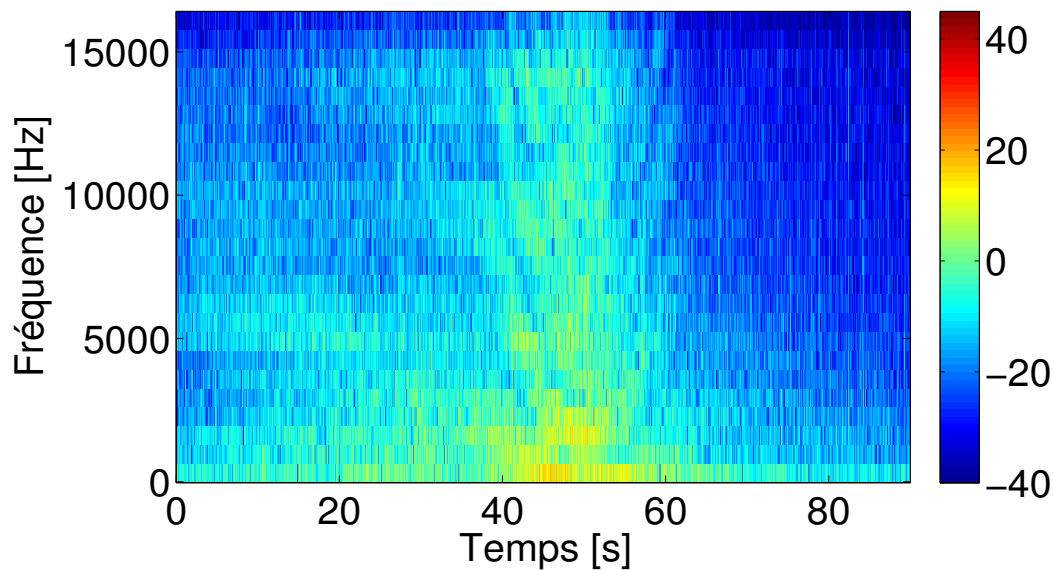


FIGURE 1.3 – Spectrogramme du bruit rayonné par une embarcation légère. Ce bruit est non-stationnaire et large bande.

niveau de bruit suffisant pour créer de sérieux dommages au système auditif des mammifères marins, et se propageant suffisamment loin pour altérer le comportement des mammifères marins jusqu'à 50km de distance [BSS<sup>+</sup>10]. Des efforts de recherches sont menés afin de mettre au point des barrières acoustiques limitant la propagation acoustique dans le milieu océanique. Des rideaux de bulles<sup>1</sup> sont ainsi utilisés pour absorber les ondes acoustiques émises par les airguns, les explosions et les pile-drivers, réduisant les intensités sonores de 3 à 5dB [WGJJ00]. De nouvelles recherches (théoriques) laissent entrevoir la possibilité d'un abaissement des niveaux sonores transmis de 10 à 20dB [SCW<sup>+</sup>10].

Une table recensant un certain nombre de sources d'origine anthropique et leur niveaux acoustiques respectifs est donné en figure 1.5.

### 1.1.3 Biophonie

La biophonie regroupe l'ensemble des sons émis par les organismes biologiques. Les sons d'origine biologique se font entendre sur l'ensemble de la bande de fréquence, de quelques Hertz à plusieurs centaines de kilo-Hertz et sont produits par les mammifères marins, les poissons et les invertébrés. Tous participent activement au chorus de sons enregistrés sous l'eau [Cat78].

Les différentes espèces ne sont ni présentes partout, ni en quantité égale, ni à tout moment. Il existe donc une forte diversité spatiale et temporelle des paysages acoustiques. Nous allons lister un certains nombre d'espèces qui sont particulièrement présentes acoustiquement.

---

1. bubble curtains

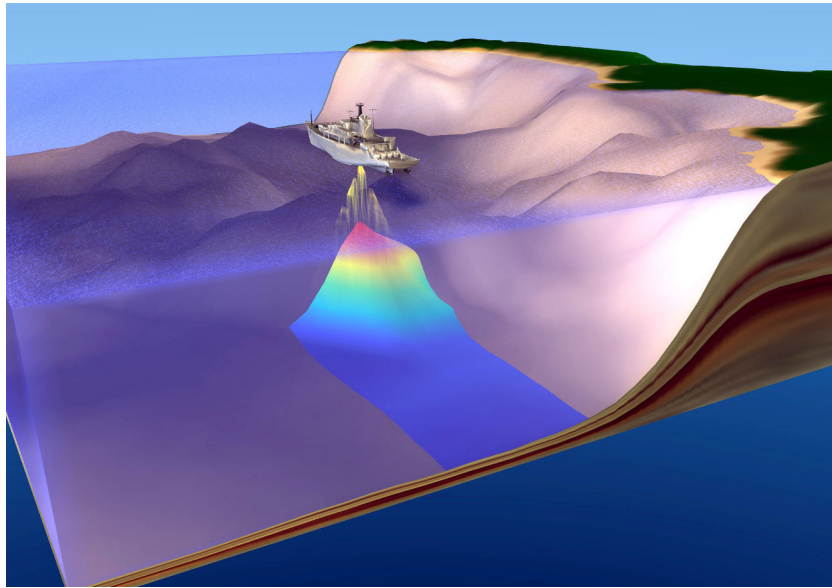


FIGURE 1.4 – Sondeur acoustique utilisé pour cartographier les fonds marins (Source : U.S. Navy, [http://www.navy.mil/view\\_image.asp?id=2767](http://www.navy.mil/view_image.asp?id=2767))

Sound source	Source level (dB re 1 $\mu$ Pa @ 1 m)	Power (W)	Total energy per pulse (J)	Bandwidth $\Delta = 10$ dB (Hz)	Source direction	Pulse duration (s)
Ship shock trial (10000 lb explosive)	304	$0.021 \times 10^{15}$	$0.042 \times 10^{15}$	0.5–50	Omni	2
Torpedo MK-46 (98 lb explosive)	289	$0.66 \times 10^{12}$	$0.066 \times 10^{12}$	10–200	Omni	0.1
Air-gun array	260	$0.21 \times 10^9$	$6.2 \times 10^6$	5–300	$60 \times 180^\circ$ V	0.03
US Navy 53C ASW sonar	235	$0.77 \times 10^6$	$1.5 \times 10^6$	2000–8000	$40 \times 360^\circ$ H	2
SURTASS LFA sonar	235	$0.59 \times 10^6$	$0.029 \times 10^9$	100–500	$30 \times 360^\circ$ H	6–100
Pile-driving 1000 kJ hammer	237	$0.46 \times 10^6$	$0.023 \times 10^6$	100–1000	$15 \times 360^\circ$ H	0.05
Multibeam sonar deep-water EM 122	245	$0.077 \times 10^6$	760	11 500–12 500	$1.0 \times 120^\circ$ V	0.01
Seal bombs (2.3 g charge)	205	$2.6 \times 10^3$	79	15–100	Omni	0.03
Multibeam sonar shallow EM 710	232	$2.2 \times 10^3$	4.5	70 000–100 000	$0.5 \times 140^\circ$ V	0.002
Sub-bottom profiler SBP 120	230	$2.1 \times 10^3$	210	3000–7000	$3 \times 35^\circ$ V	0.1
Acoustic harassment device	205	$1.3 \times 10^3$	330	8000–30 000	$90 \times 360^\circ$	0.15–0.5
Cargo vessel (173 m length, 16 knots)	192	66	–	40–100	$80 \times 180^\circ$	CW
Acoustic telemetry SIMRAD HTL 300	190	42	–	25 000–26 500	$90 \times 360^\circ$	CW
Small boat outboard engine (20 knots)	160	$42 \times 10^{-3}$	–	1000–5000	$80 \times 180^\circ$	CW
Acoustic deterrent device	150	$4.2 \times 10^{-3}$	$1.4 \times 10^{-3}$	5000–160 000	$90 \times 360^\circ$	0.2–0.3
Operating windmill turbine	151	$2.6 \times 10^{-3}$	–	60–300	$15 \times 360^\circ$ H	CW

FIGURE 1.5 – Caractéristiques de sources acoustiques d'origine anthropique (V : vertical, H : horizontal, CW : Continuous Wave) (Source : [Hil09])



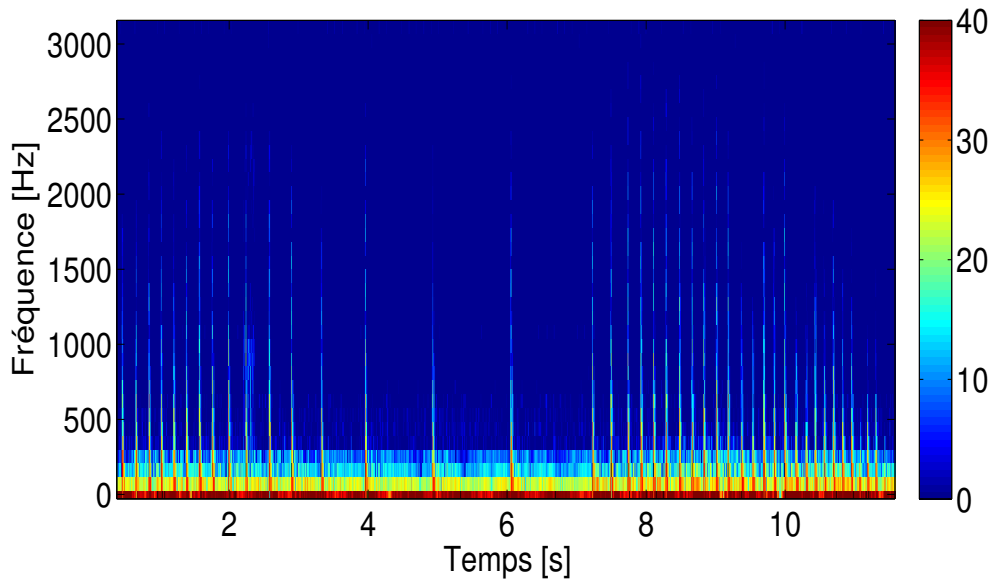


FIGURE 1.6 – Spectrogramme d’un son émis par un aiglefin (*Melanogrammus aeglefinus*) (Source : Chaire Chorus)

## Les poissons

Les **poissons** produisent une large variété de sons [LP46, ZMP99, MF02] qu’ils utilisent pour la sélection de leur habitat, la reproduction et la communication [Bas90]. Certaines larves de poissons utilisent par exemple la spécificité des paysages acoustiques pour trouver l’habitat convenant à leurs besoins [SMM<sup>+</sup>05]. Certains sons produits sont de nature impulsionnelle, leur vessie natatoire se contracte de manière rythmée pour émettre des sons proches des clics (voir le spectrogramme de sons émis par un aiglefin en figure 1.6). D’autres sont de nature harmonique, dont les fréquences sont comprises entre 10Hz et 300Hz [MB05].

## Les invertébrés

Les **crustacés** émettent eux ainsi, volontairement ou non, des sons perceptibles au niveau de l’hydrophone.

Les crevettes claqueuses peuvent par exemple être la source de bruit dominante entre 2kHz et 5kHz [EYJ48]. La fermeture rapide d’une de leurs pinces crée une bulle de cavitation [VSvdHL00], en implosant celle-ci génère des niveaux acoustiques compris entre 183 et 189 dB re 1  $\mu$ Pa @ 1 m, l’énergie est présente jusqu’à 200kHz et peut augmenter le niveau de bruit moyen de 20dB [AB98] (voir figure 1.7).

Plus discrets, les bivalves comme par exemple les coquilles Saint-Jacques émettent aussi des sons lors de la respiration. Ces sons sont mesurables à une distance de 1 mètre, ils sont plutôt large bande et sont dus à l’ouverture et fermeture de la coquille (voir figure 1.8 pour un exemple) [KGC<sup>+</sup>10].

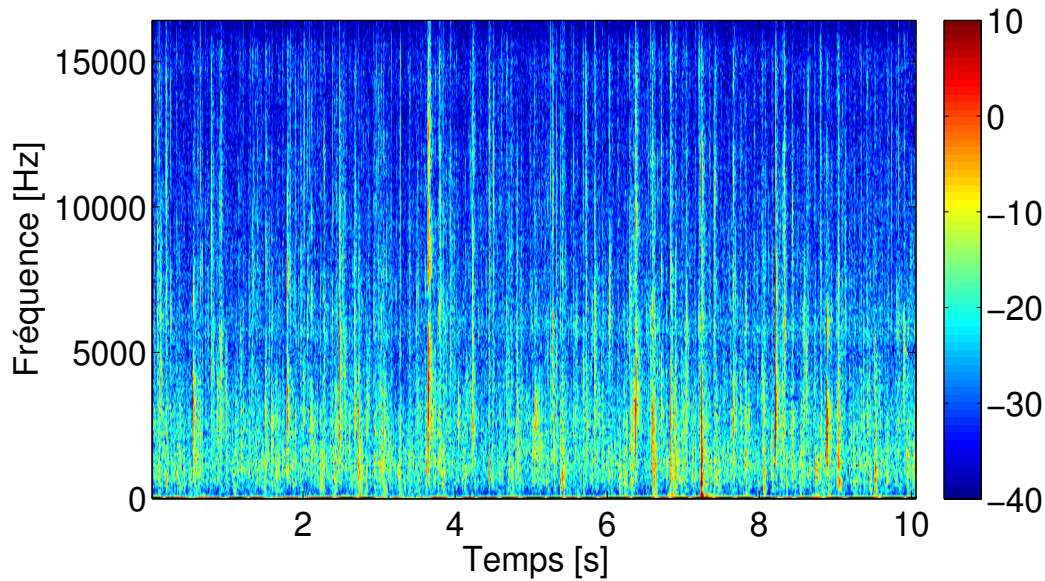


FIGURE 1.7 – Spectrogramme du son émis par l'activité benthique, et en particulier par des crevettes claqueuses, au large de l'île de Molène (Source : Chaire Chorus)

### Les mammifères marins

Parmi les **mammifères marins** nous trouvons les cétacés, les siréniens et les pinnipèdes. Les cétacés comprennent les baleines à fanons (ou mysticètes) parmi lesquels nous trouvons les rorquals, baleines franches et baleines à bosse, et les baleines à dents (ou odontocètes) qui comprennent entre autres les cachalots, bélugas, dauphins et marsouins. Les siréniens sont représentés par le lamantin et le dugong. Et enfin les pinnipèdes comprennent les phoques, otaries et morses (un exemple de son émis par un phoque barbu est donné en figure 1.9).

Les cétacés utilisent une grande diversité de signaux. Nous pouvons globalement distinguer trois classes de signaux : les sifflements (pour les odontocètes) ou chants (pour les mysticètes), les clics et les "buzzs" (séries de clics très rapprochés). Toutes les espèces n'émettent pas tous les types de sons.

Les clics sont des sons de nature impulsionnelle. Ils sont de très courte durée et couvrent une large gamme de fréquence (voir figure 1.10). Les clics sont utilisés à des fins d'écholocation c'est à dire pour se repérer dans l'espace et chasser mais pas seulement, le sonar des odontocètes possède aussi d'excellentes capacités de discrimination d'objets, leur permettant ainsi déterminer la taille, la forme et la composition d'un matériau [Au93]. Une hypothèse évoquée dans [NM83] concerne la possibilité de l'utilisation des clics pour désorienter ou assommer leurs proies. Cette hypothèse a été testée expérimentalement et révoquée dans [BBAK06] mais leurs conclusions n'excluent pas la possibilité d'une capacité à étourdir les proies. Celle-ci ne serait pas uniquement due à l'utilisation de clics.

Les sifflements sont des modulations de fréquence qui peuvent être harmoniques

(voir figures 1.11, 1.12 et 1.13), et couvrent une large gamme de fréquences (voir figure 1.14). Ces sifflements sont utilisés dans un contexte plutôt social même si leur utilisation spécifique reste mal connue et discutée [JSW06, MR01]. Il a par exemple été montré que les dauphins *tursiops* se répondent de manière cohérente à plusieurs centaines de mètres les uns des autres [Jan00]. Il existe une grande diversité de sifflements, à titre informatif nous montrons le répertoire tel que classifié par McCowan en figure 1.15 [McC95]. Ce répertoire n'a été obtenu qu'en utilisant 5 dauphins en captivité, sur un nombre limité de sifflements. Bien qu'il ne puisse donc pas être utilisé pour la détection de *tursiops* dans la nature, ce catalogue donne un aperçu de la diversité à laquelle nous pouvons nous attendre.

Les mysticètes utilisent de manière générale des sons plutôt bas en fréquence ( $\leq 5kHz$ ) et de très forte amplitude (188 dB re  $1\mu Pa$  @ 1m) pour le rorqual bleu [CT71]). Les mysticètes peuvent ainsi communiquer à grande distance (les sons portent jusqu'à plusieurs centaines de kilomètres). La raison en est probablement la propension à se déplacer seuls ou en groupes réduits (mère et enfant). Les odontocètes quant à eux utilisent en général des sons plus haut en fréquence ( $\sim 5kHz - 30kHz$ ) qui ont donc une portée beaucoup moins grande (quelques kilomètres [ORBL07]). Des signatures acoustiques sont utilisées par certains odontocètes. Ces signatures sont soit des signatures de groupe, soit des signatures individuelles par lesquelles les odontocètes se reconnaissent [JS98]. Il est donc théoriquement possible de reconnaître des groupes en écoutant les sons qu'ils émettent. Les odontocètes émettent aussi des buzzs, qui sont des séries de clics très rapprochés, dont nous montrons un exemple tiré de [TZM<sup>+</sup>11] en figure 1.16.

La communication chez les odontocètes semble être complexe, pour illustrer cette complexité nous prenons l'exemple des orques au large de l'île de Vancouver, en Colombie britannique. John Ford discute dans [For91] des différentes traditions vocales des orques résidents. L'étude de 16 groupes stables d'individus apparentés pendant de nombreuses années a permis de différencier 4 clans ayant des traditions vocales très distinctes, c'est à dire n'ayant aucun sons en commun dans leur répertoires. Au sein de chaque clan, des variations structurales des sons ont été observées, c'est ce qu'il qualifie de dialectes spécifiques à chaque bande au sein du clan. De manière surprenante, des bandes de différents clans, donc n'ayant pas de sons en commun, se déplacent souvent ensemble, et les relations sociales qu'ils entretiennent sont indépendantes des relations acoustiques.

## 1.2 Le monitoring par acoustique passive

Notons tout d'abord que nous nous intéressons au cas particulier de l'acoustique passive, par opposition à l'acoustique active pour son caractère non intrusif. L'acoustique active cherche à obtenir de l'information sur un milieu par émission

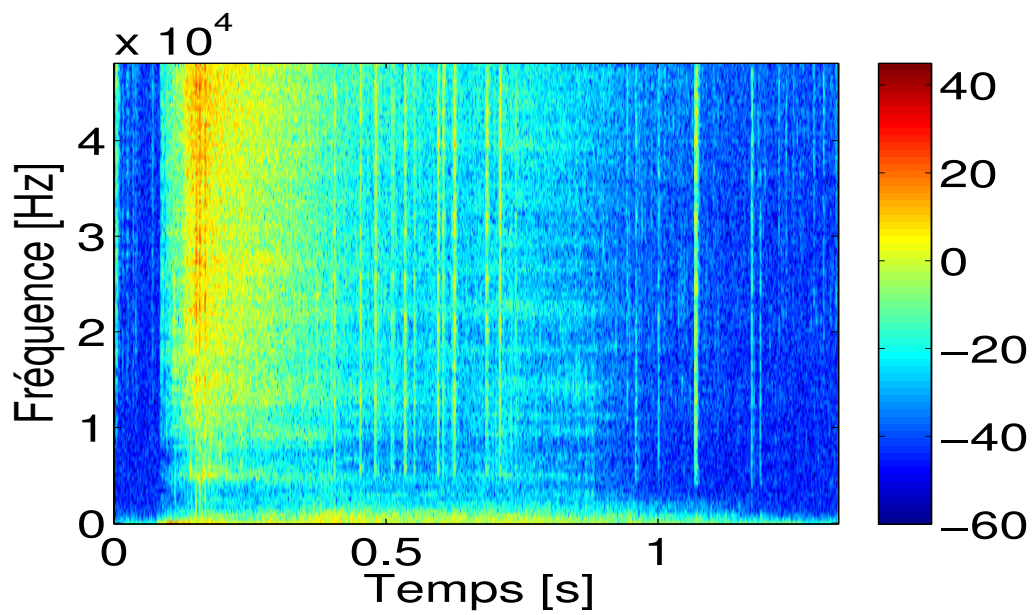


FIGURE 1.8 – Spectrogramme du son émis par une coquille Saint Jacques lors de l'activité de respiration (Source : Chaire Chorus).

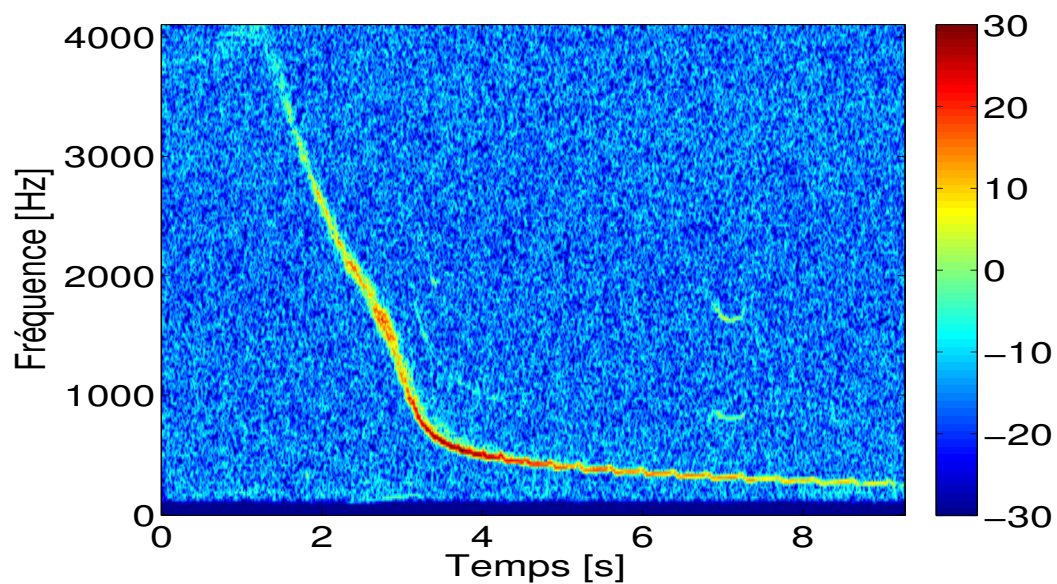


FIGURE 1.9 – Spectrogramme du son émis par un phoque barbu.

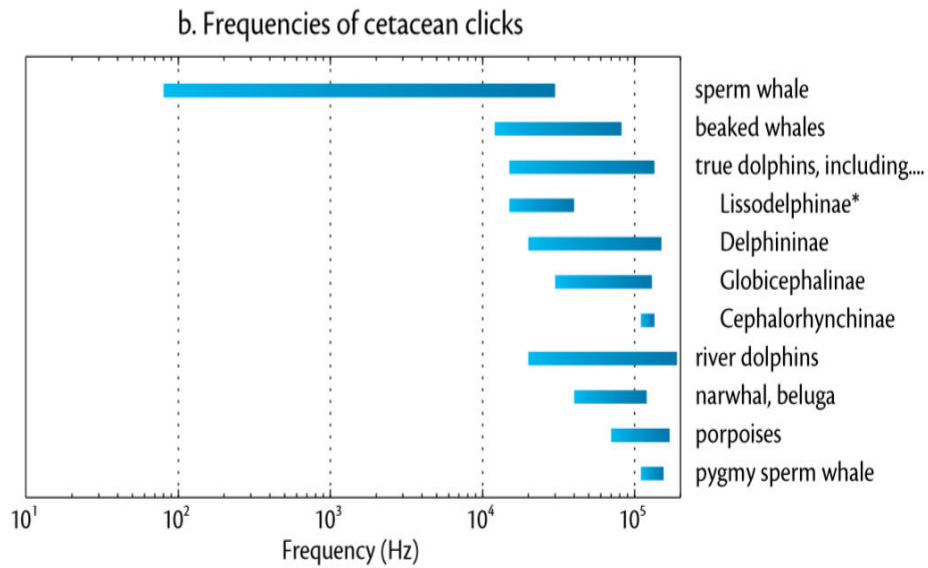


FIGURE 1.10 – Illustration des gammes de fréquences occupées par les clics de certains mammifères marins (Source : [MSM<sup>+</sup>07])

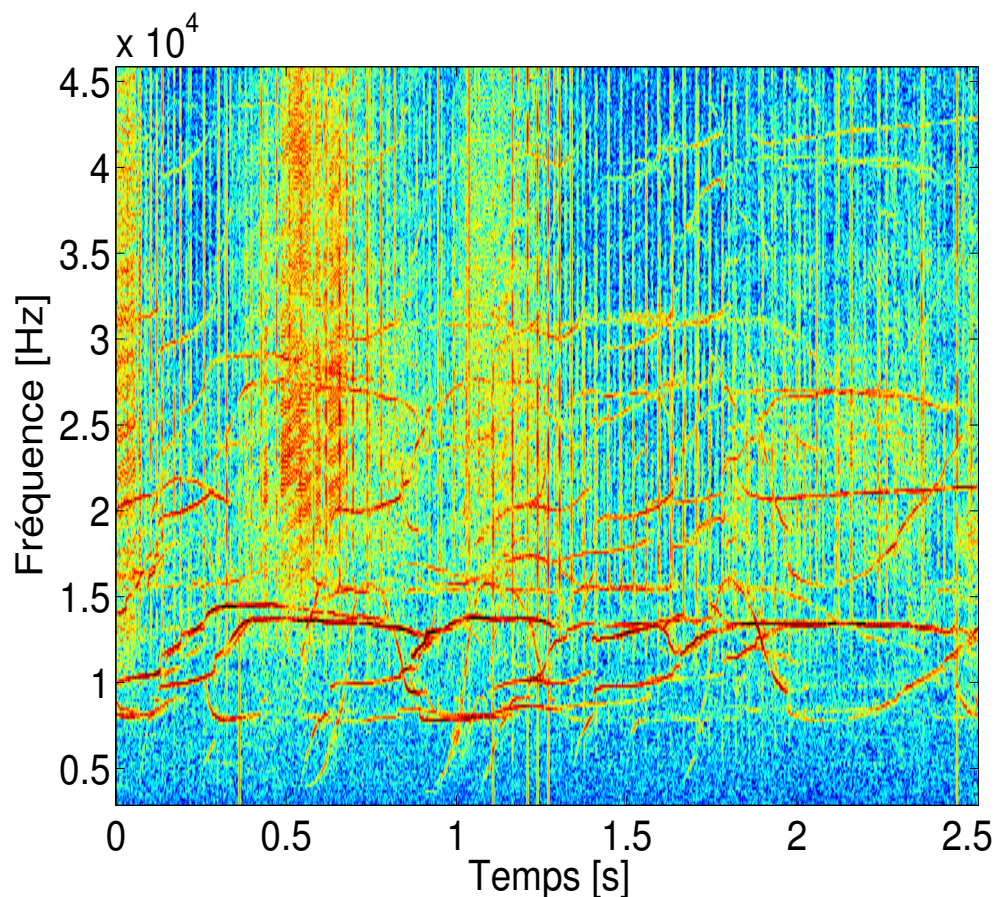


FIGURE 1.11 – Spectrogramme tracé à partir d'un extrait d'enregistrement de *delphinus delphis*. Nous y observons un grand nombre de sifflements se croisant dans le plan temps-fréquence ainsi qu'un grand nombre de clics d'écholocation. (Mobysound : 'Qx-Dc-FLIP0610-VLA-061015-163500.wav')

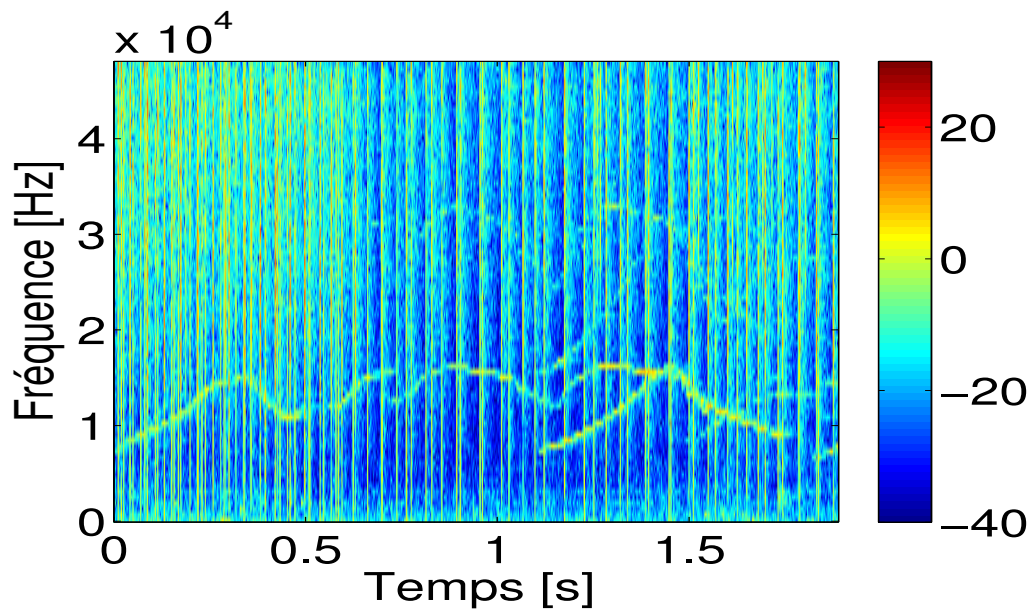


FIGURE 1.12 – Spectrogramme d’un enregistrement de grands dauphins en présence de clics et d’activité benthique). Enregistré au large de l’île de Molène (Source : Chaire Chorus).

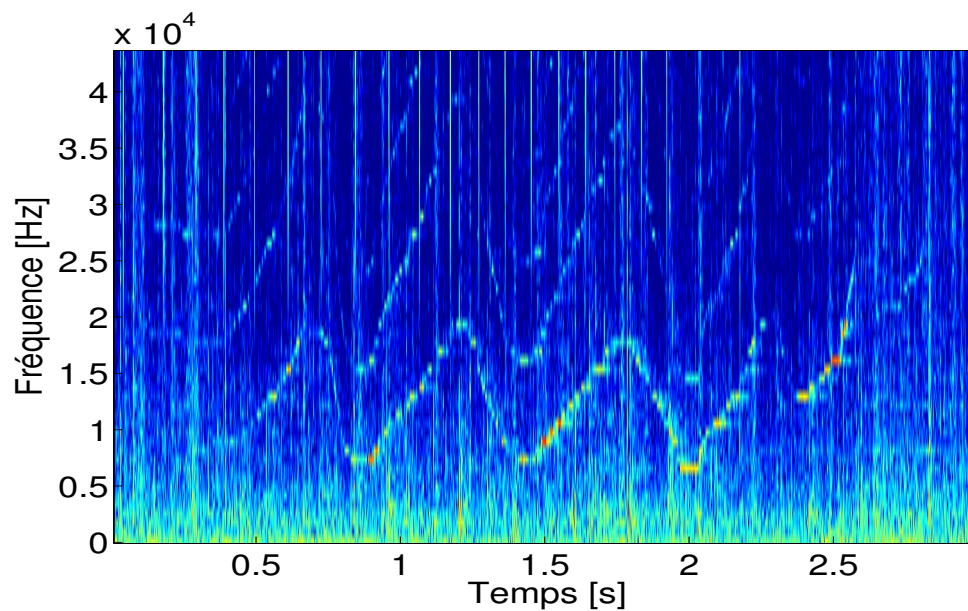


FIGURE 1.13 – Exemple de son enregistré près de l’île de Molène. Signal harmonique dans un environnement impulsionnel principalement dû à des crevettes claqueuses.

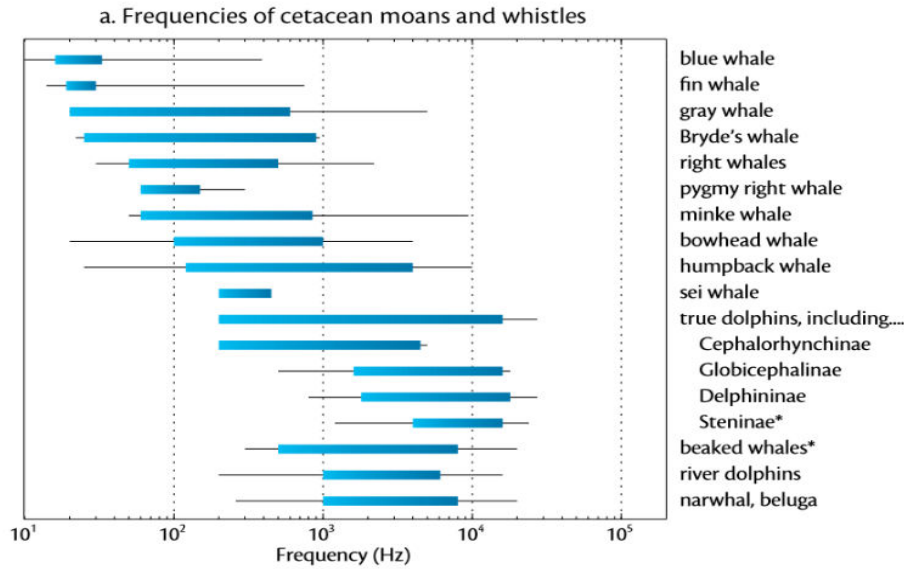


FIGURE 1.14 – Illustration des gammes de fréquences occupées par les sifflements de certains mammifères marins (Source [MSM<sup>+</sup>07])

d'une onde sur, ou au sein, de celui-ci. L'acoustique passive, au contraire, n'utilise que le bruit rayonné par le milieu pour extraire de l'information sur celui-ci. Le but du monitoring par acoustique passive (PAM<sup>2</sup>) est donc de se servir du paysage acoustique sous-marin décrit précédemment pour obtenir de l'information. Cette information possède deux composantes différentes.

La première composante concerne l'analyse de la nature des sources émettrices : géophysiques [Alp99], biologiques [MSM<sup>+</sup>07, LOB02] ou industrielles [TLKP06]. Les signaux sont détectés, extraits et classifiés. La seconde composante concerne cette fois non pas la source émettrice en elle-même mais le milieu traversé par l'onde acoustique (tomographie passive). Les signaux sont alors détectés, extraits, et la réponse impulsionnelle du canal de propagation estimée [Jos10, BNMW10].

Dans les deux cas, une étape critique dans l'extraction de l'information consiste en la détection du signal d'intérêt et de sa séparation du bruit de fond. Cette étape est difficile, et nous en détaillerons les raisons dans la section suivante.

### 1.2.1 Bénéfices de l'acoustique passive

Malgré la difficulté rencontrée pour extraire ces deux composantes informatives, il existe plusieurs raisons expliquant le fort développement de l'acoustique passive sous-marine, et en particulier pour l'étude des mammifères marins [MSM<sup>+</sup>07], que nous résumons ici.

Les méthodes d'observation visuelles classiques des cétacés souffrent de plusieurs difficultés. L'observation est dépendante des conditions météorologiques, et

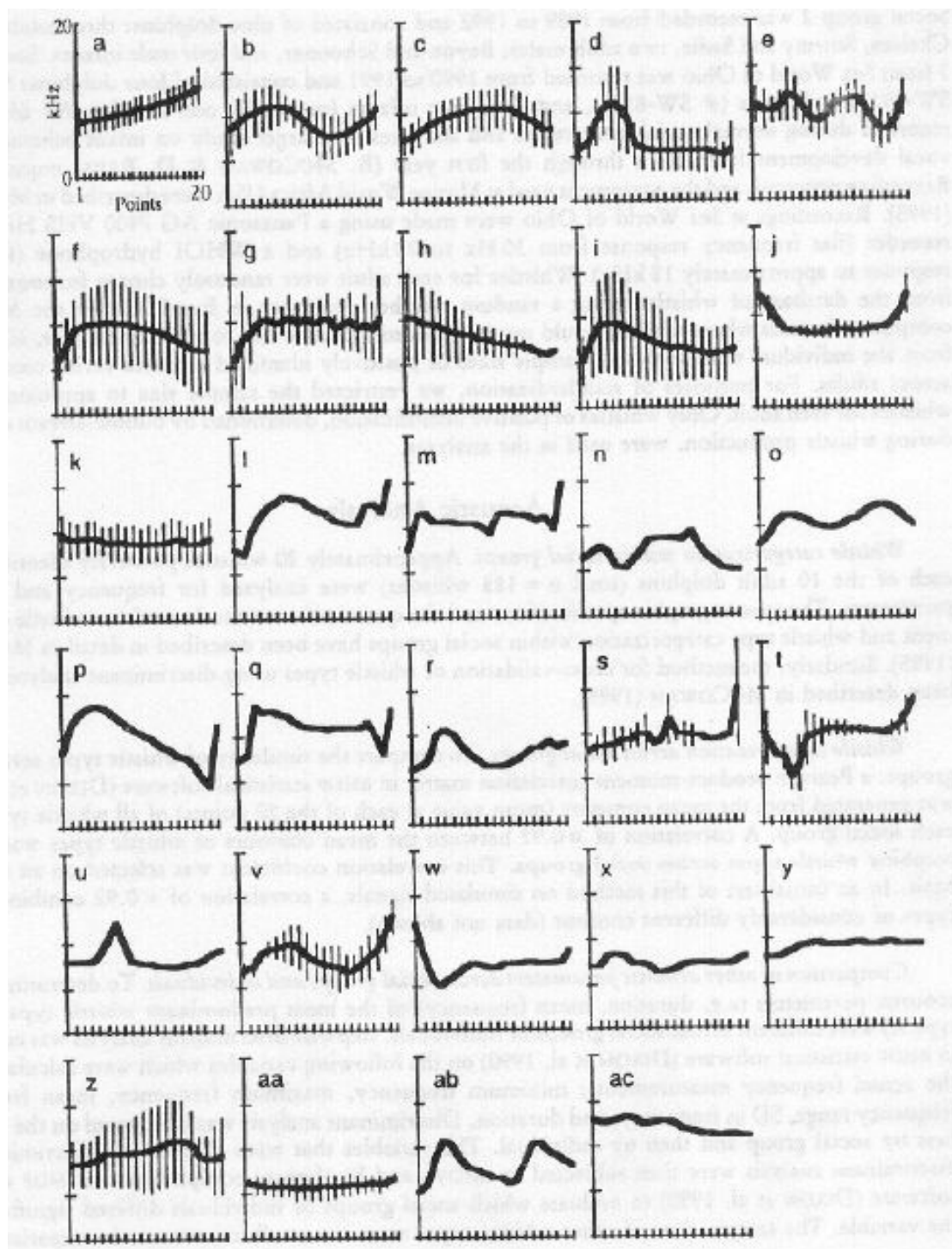


FIGURE 1.15 – Exemple d'un répertoire vocal d'un groupe de *tursiops* tel que estimé par [MR95]



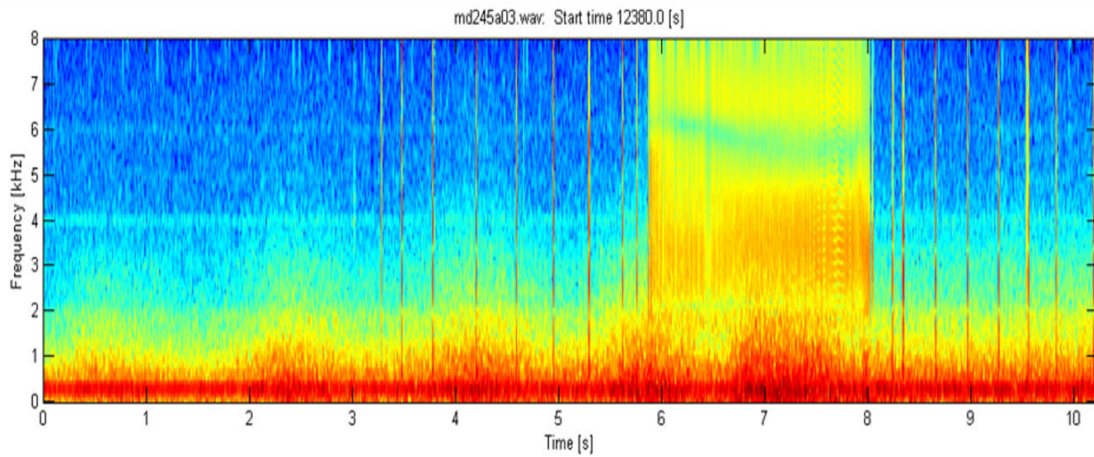


FIGURE 1.16 – Clics et buzz de baleine à bec de Blainville *Mesoplodon densirostris* (Source : [TZM<sup>+</sup>11])

est très difficile lorsque l'on cherche à étudier des régions reculées, difficiles d'accès ou dangereuses. L'information recueillie présente une faible densité spatiale car il est extrêmement difficile d'observer de vastes étendues d'océans. L'observation présente aussi une faible résolution temporelle, l'observation n'est par exemple possible que de jour, et les sorties en mer n'ont lieu que quelques jours par année. De plus l'observation ne peut avoir lieu que lorsque les cétacés remontent à la surface.

L'acoustique passive au contraire est particulièrement **adaptée à l'environnement marin** et aux espèces marines. Alors que les ondes électromagnétiques subissent une très forte atténuation de leur amplitude sur de courtes distances, les ondes acoustiques quant à elles peuvent potentiellement se propager sur de très grandes distances (dépendant des espèces en question et des conditions de propagation). Par exemple, en 1985, il est noté que l'acoustique passive rend possible la détection de baleines boréales à une distance supérieure à celle permise par l'observation visuelle [CH85]. L'information sur certaines sources peut donc être captée sur de très longues distances (plusieurs centaines de kilomètres pour les mysticètes). Il est donc en théorie possible de surveiller de vastes étendues d'océans avec peu de capteurs (au moins pour certaines espèces). L'acoustique passive est relativement indépendante des conditions météorologiques, et peut avoir lieu dans des régions reculées et peu accessibles comme l'océan arctique [KSG<sup>+</sup>13].

Ainsi, en plaçant un ou plusieurs capteurs à des endroits stratégiques pendant de longues périodes, il sera possible d'obtenir une information ayant une très **haute résolution temporelle**, et même spatiale dans certaines conditions (la résolution spatiale est dépendante du nombre de capteurs mais aussi de l'espèce étudiée). Une fois les hydrophones placés, l'information récoltée est enregistrée en continu sur des périodes pouvant aller de quelques jours à plusieurs années. Il est même possible d'obtenir une information continue en **temps-réel** pour des endroits proches des côtes.

La résolution est haute spatialement car il est facile de multiplier le nombre de capteurs, et que la propagation du son est très bonne dans l'environnement sous-marin.

Les campagnes PAM sont souvent relativement **peu chères comparées aux coûts d'expéditions classiques**. Le déploiement des hydrophones a lieu une seule fois, et fonctionne en continu sur de longues périodes pendant desquelles l'hydrophone est complètement autonome. Les campagnes d'observation visuelle habituelles nécessitent une sortie en mer à chaque observation.

À noter que les méthodes PAM ne cherchent pas à remplacer les méthodes d'observation visuelles, mais à en être complémentaire.

### 1.2.2 Objectifs de l'acoustique passive

Une fois les signaux de mammifères marins détectés, il existe de nombreuses manières possibles de les utiliser.

Une première application immédiate concerne le recensement des espèces de mammifères marins (pour les espèces dont une classification des sons est connue), la compréhension des migrations et de leur occupation de l'habitat. En effet, nous avons vu que les campagnes de recensement et de suivi de ces animaux sont difficiles à mettre en oeuvre et sont majoritairement basées sur le comptage visuel. Certaines espèces sont cependant très difficiles à observer. La spécificité des sons émis par ces animaux rend cependant possible le comptage acoustique, et augure même d'excellents résultats dans le futur, la détection acoustique pouvant avoir lieu à une distance bien plus grande que la détection visuelle [MTW<sup>+</sup>09, MTM<sup>+</sup>13]. La difficulté majeure (non résolue) que rencontrent ces méthodes d'estimation concerne le manque de connaissance du comportement acoustique des espèces étudiées. La fréquence d'utilisation des sons et ses liens avec la taille du groupe, le comportement ou l'activité dans laquelle s'engage l'animal étudié, ou encore la saison d'étude, sont inconnues.

Une autre application intéressante concerne la compréhension de l'interaction entre les activités d'origine anthropique et la biocénose<sup>3</sup>.

Les activités industrielles et maritimes connaissent un essor très important : augmentation très importante du trafic maritime, installation d'éoliennes offshore et d'hydroliennes, prospection sismique [GGP<sup>+</sup>03] etc. Il en résulte une considérable augmentation du niveau de bruit ambiant moyen dans les écosystèmes marins. L'historique de mesure montre une augmentation moyenne de 3dB par décen-

---

3. La biocénose est définie comme l'ensemble des êtres vivants cohabitants dans un biotope. L'étude la biocénose part du principe que le cadre d'étude adéquat ne se situe pas au niveau de l'individu spécifique mais de l'ensemble des individus en interaction dans un espace.

nie dans le nord-est du pacifique [Hil09]. Cette augmentation impacte les cétacés et la biocénose dans son ensemble. Cette augmentation du bruit peut provoquer des effets de masquage des communications sociales et une diminution des capacités à détecter certains sons comme ceux de prédateurs ou de proies. Elle peut aussi provoquer l'abandon de certaines zones de manière temporaire ou définitive. Une forte augmentation du bruit peut aussi induire stress, désorientation, dommage de l'audition potentiellement irréversible, échouages etc. [Tho95].

La pollution des environnements marins par l'Homme n'est pas seulement acoustique mais aussi physico-chimique. Les polluants ont des effets microscopiques mais aussi macroscopiques, certains organismes benthiques sont très sensibles à la pollution de l'eau. L'état de la bio-diversité benthique est alors un indicateur de la bonne santé de celui-ci. Certains organismes sensibles comme les bivalves, crevettes et autres invertébrés sont producteurs de sons. L'acoustique passive pourrait alors être utilisée comme moyen d'estimation rapide du taux de pollution en étudiant ses effets sur la faune [DIGJ<sup>+</sup>12].

Un autre aspect lié à l'acoustique passive concerne non pas la recherche d'information sur la source émettrice mais sur la manière dont le milieu de propagation déforme l'onde acoustique. Nous nous servons alors des sons captés comme des sources d'opportunité pour faire de la tomographie passive. Des travaux ont déjà été réalisés afin d'utiliser les signaux émis par les cétacés pour faire de l'inversion géo-acoustique [BGS<sup>+</sup>13, Jos10]

## 1.3 L'acoustique passive en pratique

Nous avons exposé les bénéfices et objectifs de l'acoustique passive. Nous présentons maintenant la chaîne de traitement et les difficultés techniques inhérentes au choix de l'acoustique passive comme outil d'analyse.

### 1.3.1 Chaîne de traitement

Dans ce manuscrit nous nous intéressons particulièrement au monitoring des mammifères marins. Ces activités de monitoring sont classiquement dénommées par l'acronyme DCLDE pour "Detection, Classification, Localization and Density Estimation". Ces activités peuvent être menées en utilisant tous les sons émis par les mammifères marins. Aujourd'hui, les traitements se servant des clics émis par les mammifères marins sont plus avancés que ceux utilisant les sifflements. Il existe une littérature conséquente de méthodes traitant de la localisation, de la classification et du dénombrement de mammifères marins par utilisation des clics reçus à l'hydrophone [MWD<sup>+</sup>06, GG06, HZVDSA10, ZVDSH<sup>+</sup>10, MTS<sup>+</sup>13, SHC<sup>+</sup>08, RSG10].

Celles-ci sont quasiment opérationnelles. À l’opposé les méthodes utilisant les sifflements sont encore à l’état embryonnaire. En particulier, le lien entre la détection et l’estimation des sifflements [EK08, RBP<sup>+</sup>11, GCGW13], et leur utilisation effective dans une étape de classification reste encore très problématique [BD13]. L’étape de classification est alors souvent réalisée sur des sifflements extraits manuellement [Ste81, RSB<sup>+</sup>07, ORBL07, BOC<sup>+</sup>11], bien que certaines méthodes automatiques liant la détection et la classification commencent à voir le jour [GCGW13].

L’explication de ce retard est simple, il reste très difficile de détecter et d’estimer correctement les sifflements. Les potentiels liés à l’utilisation des sifflements sont cependant très attractifs, ces signaux possèdent une très grande diversité, et sont donc potentiellement très discriminants pour la reconnaissance d’espèces et de groupes. **Nous nous intéresserons donc tout particulièrement au cas des sifflements.**

La résolution des tâches DCLDE par utilisation des sifflements peut être grossièrement découpée en six étapes génériques :

1. Une étape de **détection d’évènements acoustiques** : le but est de détecter les pics d’énergie, en général sur une représentation temps-fréquence.
2. Une seconde étape d’**estimation de la loi de fréquence instantanée**. Le but est de relier les pics d’énergie détectés en une seule composante temps-fréquence.
3. Une troisième étape d’**extraction de caractéristiques** ("features"). Ces caractéristiques sont centrales lors de l’étape de classification. Ces critères de classification sont souvent relatifs à la forme du signal : fréquence centrale, nombre de points d’inflexion, durée, etc. Ils doivent être les plus discriminants possibles.
4. L’étape de **classification** proprement dite.
5. **Dénombrement** des mammifères marins
6. **Localisation** des mammifères marins.

La qualité de l’étape de détection de l’énergie dans le plan temps-fréquence apparaît comme **critique pour l’ensemble de la chaîne de traitement**. Si cette étape est mal réalisée, ses résultats impacteront négativement l’ensemble de la chaîne de traitement. Ainsi, afin de limiter le nombre de fausses alarmes, et pour limiter la complexité algorithmique des méthodes cherchant à estimer des lois de fréquence instantanées à partir des pics temps-fréquence détectés, le seuil de détection est habituellement placé haut. Dans la communauté PAM, il est en général placé autour de 8-10dB (voir par exemple [SRG<sup>+</sup>10, RBP<sup>+</sup>11, GCGW13]). Ce faisant nous ignorons volontairement un grand nombre de signaux utiles.

Dans ce travail de thèse nous proposons de nous attaquer à ce problème de détection. Nous nous imposons **deux contraintes** fortes tout au long de ce manuscrit .

La première contrainte est de proposer **un détecteur ne se basant que sur les statistiques du bruit**, aucune hypothèse ne devra être faite sur les signaux d'intérêt afin de garantir une large applicabilité de la méthode proposée. Le détecteur sera donc basé sur la résolution d'un test d'hypothèse binaire, le signal étant supposé inconnu, l'outil central de ce manuscrit sera le critère de Neyman-Pearson.

La seconde contrainte est de proposer une méthode ne mettant en jeu que **peu de paramètres**. Nous pensons qu'une méthode présentant peu de paramètres aura plus de chances d'être applicable telle quelle sur une large gamme de signaux.

**Remarque :** Les six étapes distinctes que nous venons de décrire ne sont pas nécessairement distinctes. Par exemple certaines méthodes sont spécialement conçues pour la détection de certains mysticètes produisant des sifflements très stéréotypés. Des méthodes dites de corrélation de spectrogramme sont alors utilisées [MC00]. Le spectrogramme est corrélé avec un template, c'est à dire un modèle temps-fréquence approximatif du signal recherché. Ces méthodes concatènent alors la détection, l'extraction et la classification en une seule et même étape, une partie du travail ayant déjà été faite en amont (i.e. lors de la définition du template). Ces méthodes offrent de bons résultats pour ces espèces mais ne sont pas transposables aux cas des odontocètes pour qui la définition de templates n'est pas possible car les différents sifflements ne sont pas, ou peu, connus. De même les méthodes utilisant des réseaux de neurones détectent, extraient les features et classifient simultanément les signaux [Erb00, Mel04] mais nécessitent cependant de grandes bases d'entraînement. La création de celles-ci requièrent un travail titanesque, qui doit être recommencé pour chaque nouvelle espèce d'étude.

### 1.3.2 Quels enseignements tirer de la description du paysage acoustique ?

Nous effectuons ici plusieurs observations découlant du modèle de Wenz du paysage acoustique.

#### Définition de la notion de bruit

Tout d'abord, il nous faudra définir précisément la notion de bruit. Nous cherchons à discriminer le signal du bruit. Tout au long de ce manuscrit, notre méthodologie de détection sera basée sur un critère de Neyman-Pearson. Nous allons donc décrire de manière très précise les statistiques du bruit et détecter le signal

en conséquence. La définition de ce que nous considérons comme étant du bruit dans cette étude, sera fondamentalement liée à notre objectif final. En particulier le bruit pour l'un est potentiellement un signal pour l'autre.

### **Non-stationnarité temporelle**

Les facteurs influençant le spectre observé sont par nature intermittents (vent, pluie, tremblements de terre, émissions biologiques, trafic maritime, ...), quelle que soit la définition du bruit adoptée, celui-ci sera donc par nature non-stationnaire. Ainsi, si nos signaux d'intérêt sont biologiques, alors tous les sons d'origine géophysique sont considérés comme du bruit qui sera alors non-stationnaire. Inversement, si nos signaux d'intérêt sont les signaux géophysiques, alors les signaux biologiques et anthropiques seront considérés comme du bruit, là encore notre bruit de fond sera non-stationnaire.

### **Non-stationnarité spatiale**

Le troisième enseignement concerne la stationnarité spatiale. La nature et l'amplitude des sources est dépendante de la position de l'hydrophone. Un hydrophone placé sur une voie très pratiquée par les navires, dans un environnement grand fond ou petit fond, en présence de benthos ou non, recevra un signal différent.

### **Non-blancheur du bruit**

Un dernier enseignement concerne la non-blancheur du bruit. L'amplitude du bruit décroît fortement avec la fréquence. Toute méthode faisant l'hypothèse du bruit blanc nécessitera donc un blanchiment préalable du spectre afin de se rendre applicable.

### **Conséquence**

L'enseignement principal que nous devons tirer de ces observations est le suivant : afin de pouvoir détecter les signaux d'intérêt, il est nécessaire de **travailler localement**. L'estimation du bruit se doit d'être faite localement, à la fois **temporellement, fréquemment et spatialement**. Toute méthode ayant vocation à être applicable en pratique, c'est à dire en dehors des conditions très spécifiques dans lesquelles elle a été développée se doit donc de respecter cet impératif.

L'exemple type est le cas du navire équipé d'hydrophones et implémentant des méthodes de détection et localisation des mammifères marins pour éviter les collisions. Ce navire se déplace et subira des conditions de bruit très changeantes. Un deuxième exemple est celui du réseau de capteurs. L'espacement spatial entre capteurs nécessitera d'estimer le bruit au niveau du capteur et non à l'échelle du réseau de capteurs.

### 1.3.3 Difficultés inhérentes à l'acoustique passive sous-marine

Dans cette section nous analysons les différentes contraintes auxquelles nous faisons face lorsque nous nous intéressons à l'acoustique passive pour le monitoring des mammifères marins. Nous listons trois classes de contraintes : 1) les contraintes liées au contexte passif de l'étude, 2) les contraintes liées à la nature des signaux d'intérêt et 3) les contraintes liées au milieu océanique.

#### Contraintes liées au contexte passif

1. La **distance** par rapport à l'hydrophone est inconnue.
2. Les **dates** d'émissions et de réceptions sont inconnues.
3. Le **nombre** de signaux est inconnu.
4. La gamme des **rapports signaux à bruit** (RSB) rencontrée est très large, elle est dépendante de la localisation de l'animal, de son orientation et de sa puissance d'émission. De plus, le RSB d'un même signal évolue très rapidement au cours du temps.
5. Les bases de données générées ont une **grande dimension**. Pour donner une idée de l'ordre de grandeur : la base de données Orhive représente un enregistrement continu de 20000 heures échantillonnées à 44,1kHz. Soit environ trois mille milliards d'échantillons à traiter, soit environ 10TB de données [NSST13]. De même les bases de données générées par les High-Frequency Acoustic Recording Package (HARP) [WH07] ont vocation à devenir des bases de données immenses : 2-12TB par an et par instrument, pour 25 instruments déployés, environ 300TB de données sont générées par an.

#### Contraintes liées à l'étude des mammifères marins

1. Les sifflements émis par les mammifères marins sont des modulations de fréquence non linéaires, et fortement **non-stationnaires**. De même les clics, sons impulsionnels, sont loin d'être stationnaires.
2. La **forme d'onde** des signaux est inconnue.
3. La **diversité** des signaux observés est grande (voir section biophonie).
4. Non seulement le nombre de signaux est inconnu mais les cétacés, et en particulier les odontocètes sont des animaux très sociaux qui communiquent simultanément, et sur les mêmes bandes de fréquences. Nous attendons donc potentiellement un **grand nombre de signaux se superposant dans le plan temps-fréquence** (voir la figure 1.11 pour s'en convaincre).
5. Alors que les sifflements sont quasi omnidirectionnels, les clics émis par les cétacés sont très directifs [ABMF12, LA03]. La puissance reçue est donc dépendante de l'orientation de l'animal par rapport à l'hydrophone (voir

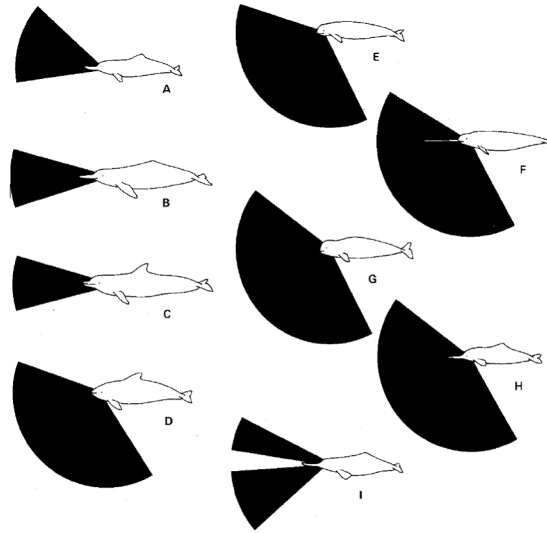


FIGURE 1.17 – Représentation schématique de la directivité des émissions de clics et de leur diversité chez les odontocètes. A : *Lipotes vexillifer* (dauphin de Chine), B : *Inia geoffrensis* (dauphin rose de l’Amazonie), C : *Tursiops truncatus* (grand dauphin), D : *Cephalorhynchus commersonii* (dauphin de Commerson), E : *Neophocaena phocaenoides* (marsouin aptère), F : *Monodon monoceros* (narval), G : *Delphinapterus leucas* (baleine blanche), H : *Pontoporia blainvillei* (dauphin de la Plata) et I : *Platanista indi* (dauphin de l’Indus). Figure prise dans [Pil83]

l’exemple du diagramme de directivité de plusieurs odontocètes en figure 1.17).

6. Les cétacés se déplacent lors de l’émission des signaux. Le signal reçu subit donc un **effet Doppler**. Ce Doppler peut être utilisé pour déduire la direction et la vitesse de l’animal [JGI<sup>+</sup>12].

### Contraintes liées au milieu océanique

1. Nous observons des **multi-trajets et évanouissements de canal** lors de la propagation dans le milieu sous-marin [BGS<sup>+</sup>13]. (voir figures 1.18 et 1.19).
2. La **propagation** est complexe dans le milieu marin (voir [Bon10])
3. Le bruit océanique est **non stationnaire et non blanc** (voir section précédente)

## 1.4 Représentations temps-fréquence

Au vu de la nature non-stationnaire des signaux d’intérêts et du bruit les représentations temps-fréquence s’imposent naturellement comme outils privilégiés d’analyse. Il existe une très grande diversité de représentations. Le choix que nous devons faire doit être guidé à la fois par les objectifs et les contraintes de l’étude.



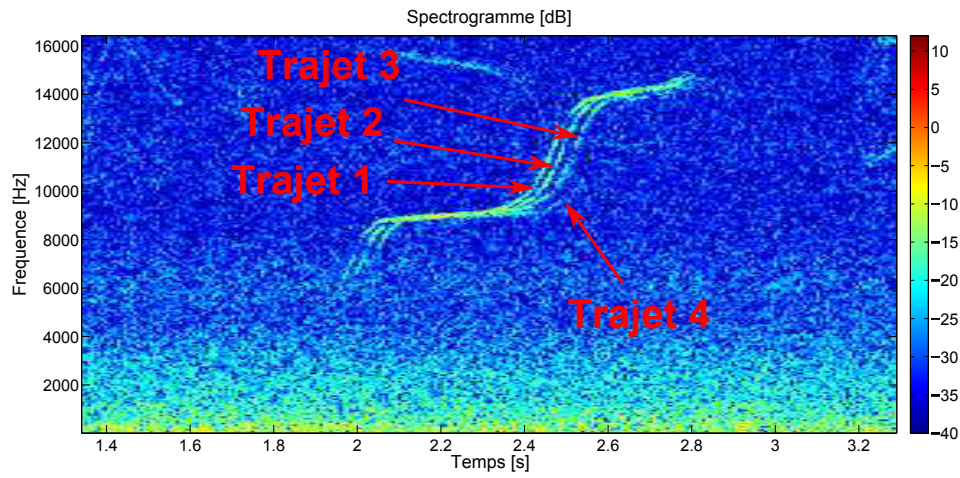


FIGURE 1.18 – Représentation d’un signal de dauphin avec multi-trajet (source : SHOM).

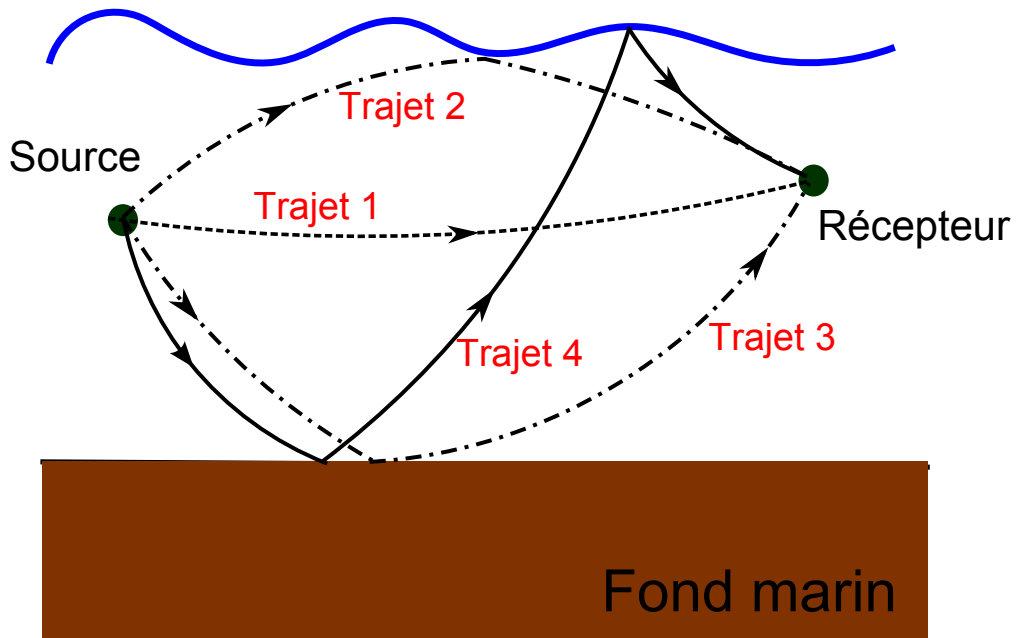


FIGURE 1.19 – Représentation schématique des multi-trajets dans un environnement océanique.

Nous cherchons à localiser l'énergie d'un signal dans le plan temps-fréquence. Les représentations quadratiques apparaissent donc comme un choix naturel. Parmi toutes les classes de représentations disponibles, celles appartenant à la classe de Cohen [Coh89] sont un choix pertinent. Cette classe a reçu une attention considérable ces 25 dernières années [HBB92, LPA93, BR92, JB95, SS95, MF85, Fla98, DS01], les propriétés de cette classe sont ainsi bien connues.

La caractéristique majeure de la classe de Cohen est le compromis qui doit être fait entre la précision de la localisation du signal conjointement en temps et en fréquence, et la présence de termes d'interférence. Aux deux extrêmes nous trouvons le **spectrogramme** et la distribution de **Wigner-Ville**. Alors que la distribution de Wigner-Ville permet une localisation parfaite des signaux à modulation de fréquence linéaire, le spectrogramme est soumis au principe d'incertitude temps-fréquence [Bal81]. Il ne sera pas possible de représenter un signal non-stationnaire conjointement en temps et en fréquence avec une précision parfaite en utilisant le spectrogramme. Il semblerait donc logique à première vue de choisir la représentation de Wigner-Ville. Nous choisirons cependant dans cette étude le spectrogramme pour plusieurs raisons.

La raison principale motivant ce choix est l'absence de terme d'interférence. En particulier, il est montré dans [IDS03] que les performances des estimateurs de lois de fréquences instantanées dépendent fortement du nombre de composantes présentes lors de l'analyse. Les performances décroissent très rapidement avec le nombre de composantes (en pratique pour un nombre de composantes supérieur à 3 ou 4). Nous nous intéressons dans ce travail à des signaux qui présentent un très grand nombre de composantes (potentiellement plusieurs dizaines en quelques secondes). La contribution des termes d'interférences, même réduite, est donc absolument rédhibitoire. Le spectrogramme en revanche conserve des performances égales qui ne dépendent ni du nombre de composantes présentes dans l'enregistrement, ni de leur espacement respectifs (aussi longtemps que leurs supports temps-fréquence ne se chevauchent pas).

La seconde raison est la connaissance précise des statistiques du spectrogramme en présence de bruit Gaussien [HMM08, HMC02]. Le spectrogramme est issu du module carré d'une transformation linéaire, ce qui facilite grandement l'expression, la manipulation, et l'interprétation de ses statistiques. Les statistiques d'ordre 1 et 2 d'un signal aléatoire sont connues pour tous les membres de la classe de Cohen dans trois cas : 1. bruit blanc stationnaire, 2. bruit coloré stationnaire et 3. bruit blanc non-stationnaire [Sta02]. Le cas coloré non-stationnaire en revanche n'est pas connu. C'est pourtant dans ce cadre que se place ce travail. Même dans les cas connus, la manipulation des statistiques engendrées par des transformations

bilinéaires de signaux Gaussiens sera difficile.

Le spectrogramme est déjà largement utilisé par les communautés scientifiques n'appartenant pas au monde du traitement du signal [HBB92]. Les raisons de ce succès sont la facilité d'interprétation du spectrogramme (il s'apparente à une représentation spectrale de type densité spectrale de puissance, très classique), sa robustesse au bruit et sa faible complexité algorithmique. Il apparaît donc intéressant de fournir un outil de détection robuste et d'interprétation aisée pour des non spécialistes.

Enfin, le spectrogramme est une transformation inversible du signal sous condition de normalisation de la fenêtre d'analyse.

Pour toutes ces raisons nous choisissons donc le spectrogramme comme outil d'analyse tout au long de ce manuscrit.

### Définition du spectrogramme

Celui-ci est défini comme le module carré de la transformée de Fourier à court terme :

$$\mathbb{S}_x[n, k] = \left| \sum_{p=0}^{M-1} x[nD + p]w[p]e^{-j2\pi k \frac{p}{K}} \right|^2 \quad (1.1)$$

avec  $w[p]$  une fenêtre de pondération de longueur  $M$ ,  $D$  est l'incrément temporel entre deux fenêtres d'analyse successives.  $K$  est la longueur de la transformée de Fourier discrète (TFD). La grandeur  $K/M$  est appelée facteur de zero-padding. La grandeur  $D/M$  quant à elle est appelée facteur de recouvrement.

La fenêtre de pondération  $w$  [Har78] remplit la condition de normalisation énoncée plus tôt et qui assurent l'inversibilité de la TFCT [All77] :

$$\sum_{m=0}^{M-1} w[m]^2 = 1 \quad (1.2)$$

C'est donc sur cette représentation  $\mathbb{S}_x[n, k]$  du signal  $x[n]$  que nous allons travailler tout au long de ce manuscrit.

## 1.5 Conclusion

Dans ce chapitre nous avons décrit le paysage acoustique rencontré en mer. Les différents contributeurs au bruit ambiant sont d'origine géologique, anthropique et biologique ; le bruit ambiant est non-stationnaire et coloré. Ce paysage acoustique

contient de l'information sur les différentes sources émettrices, notre intérêt dans ce manuscrit se porte spécialement sur les sons émis par les mammifères marins. Nous avons décrit les différentes étapes de la chaîne de traitement du signal en acoustique passive, et nous avons identifié l'étape de détection des événements acoustiques d'intérêt comme critique. **L'étude de cette étape de détection fait ainsi l'objet de ce travail de thèse.** Au vu de la nature non stationnaire des signaux d'intérêt, et de la nature fortement multi-composante des signaux d'intérêt, nous avons choisi le spectrogramme comme outil privilégié d'analyse. Nous étudierons donc dans le chapitre suivant les méthodes permettant de détecter les pics d'énergie dans le spectrogramme de manière adaptée. La méthode doit ainsi prendre en compte l'ensemble des contraintes formulées dans ce chapitre. Nous choisirons une méthode statistique, ne se basant que sur les statistiques du bruit, et ne faisant donc pas intervenir de connaissance *a priori* sur le signal.



# Chapitre 2

## Détection adaptative de signaux non-stationnaires

### Sommaire

---

<b>2.1</b>	<b>Introduction</b>	<b>46</b>
<b>2.2</b>	<b>Modèle du signal en temps</b>	<b>46</b>
2.2.1	Définition du bruit	47
2.2.2	Définition du signal :	47
2.2.3	Conséquence pour le processus de détection :	47
<b>2.3</b>	<b>Modèle du signal en temps-fréquence</b>	<b>49</b>
2.3.1	Définition du support temps-fréquence du signal	49
2.3.2	Formulation temps-fréquence du test d'hypothèse binaire	49
<b>2.4</b>	<b>Densité de probabilité du spectrogramme sous l'hypothèse bruit seul</b>	<b>51</b>
<b>2.5</b>	<b>Estimation locale de la densité spectrale de puissance (DSP) du bruit</b>	<b>53</b>
2.5.1	Méthodes de lissage	55
2.5.2	Méthodes conjointes d'estimation du bruit et de détection	55
2.5.3	Méthodes à statistiques minimales	56
2.5.4	Méthode choisie dans ce travail : estimation par les Z-minima	56
<b>2.6</b>	<b>Résolution du test d'hypothèse binaire</b>	<b>58</b>
<b>2.7</b>	<b>Exemples sur données réelles</b>	<b>60</b>
<b>2.8</b>	<b>Conclusions et problématique</b>	<b>62</b>

---

---

**Résumé du chapitre précédent :** Les enjeux du monitoring par acoustique passive ont été exposés, et une description du paysage acoustique sous-marin a été proposé.

**Objectif de ce chapitre :** Ce chapitre est dédié à la formulation temps-fréquence d'un test d'hypothèse binaire. La densité de probabilité du bruit dans le spectrogramme est décrite. Le problème d'estimation de la densité spectrale de puissance du bruit dans le cas coloré et non-stationnaire est abordé. Finalement un critère de Neyman-Pearson est utilisé pour résoudre ce test d'hypothèse binaire et détecter le signal en chaque point temps-fréquence.

---

## 2.1 Introduction

Le but de ce second chapitre est de présenter une méthode automatique de détection des signaux d'intérêt. Ce chapitre sert de base pour la suite de ce manuscrit. La question centrale que nous nous posons dans ce chapitre est celui de l'existence d'un évènement acoustique à une date et une fréquence donnée. Par évènement acoustique nous désignons une réalisation d'une variable non assimilable au bruit. En ce sens, la méthode présentée repose sur une analyse fine des statistiques du bruit.

## 2.2 Modèle du signal en temps

Tout au long du manuscrit nous considérerons un signal temporel réel  $x[n]$ . Ce signal est composé d'un signal d'intérêt inconnu  $s[n]$  corrompu par un bruit additif Gaussien  $b[n]$  :

$$x[n] = s[n] + b[n], \quad (2.1)$$

Le contexte dans lequel nous nous plaçons induit plusieurs contraintes fortes sur le signal  $s[n]$  et sur le bruit  $b[n]$  que nous avons détaillé dans le chapitre précédent. Nous souhaitons développer une méthode applicable à la plus large gamme de signaux possibles. Pour ce faire nous devons donc faire le moins d'hypothèse possible sur le signal d'intérêt.

Au chapitre précédent nous avons montré qu'il est nécessaire de définir et d'estimer les statistiques du bruit localement. Ces statistiques doivent être estimées en chaque point temporel, fréquentiel et spatial. Nous nous plaçons dans cette étude dans le cas mono-capteur, donc en un seul point spatial. La problématique se réduit donc à l'estimation locale des statistiques du bruit au niveau temporel et fréquentiel.

### 2.2.1 Définition du bruit

Nous montrerons dans la section 2.4 que l'hypothèse de non-blancher et de non-stationnarité totale conduit à un problème de détection insoluble, nous relâchons donc ces hypothèses et considérons un sous-problème de ce cas (trop) général. Nous définissons donc le bruit de fond comme la réalisation d'une variable aléatoire Gaussienne, localement blanche et stationnaire. La notion de localité est définie par rapport à un voisinage temps-fréquence  $V$ .

### 2.2.2 Définition du signal :

Au sein de ce voisinage temps-fréquence  $V$ , le bruit est blanc et stationnaire. Le signal est donc défini comme ayant des caractéristiques différentes de celles du bruit sur  $V$ . La formulation du problème permet donc de discriminer le bruit du signal si :

1. Le signal est une variable aléatoire non stationnaire et blanche sur  $V$ ,
2. Le signal est une variable aléatoire stationnaire et non blanche sur  $V$ ,
3. Le signal est une variable aléatoire non-stationnaire et non blanche sur  $V$ ,

Nous excluons donc le cas d'un signal blanc stationnaire sur  $V$ , car ses statistiques seront assimilées à celles du bruit. Le signal est ainsi essentiellement défini comme "non-bruit" sur  $V$ .

### 2.2.3 Conséquence pour le processus de détection :

La définition du voisinage temps-fréquence  $V$  a ainsi une importance primordiale pour le résultat final. Le détecteur que nous décrivons dans ce chapitre est basé sur l'analyse des coefficients du spectrogramme sur ce voisinage temps-fréquence. L'information accessible au détecteur pour décider de la présence éventuelle de signal est celle qui est incluse dans  $V$ . En clair, le détecteur ne "voit" qu'une partie du plan temps-fréquence et décide en conséquence.

Afin de pouvoir détecter un signal d'intérêt, le voisinage  $V$  ne doit pas être inclus dans le support temps-fréquence du signal d'intérêt. Ce voisinage doit contenir au moins quelques réalisations du "bruit". Nous illustrons les implications du choix du voisinage temps-fréquence  $V$  avec l'exemple du signal acoustique rayonné par



un navire (voir figure 2.1). Nous définissons deux voisinages temps-fréquence différents :  $V_1 = [1min \times 500Hz]$  et  $V_2 = [1s \times 500Hz]$ . La fenêtre  $V_1$  n'est pas strictement incluse dans le support temps-fréquence émis par le navire, il existera donc un contraste statistique entre l'énergie émise par le navire et le bruit de fond, le signal sera donc détectable. À l'opposé, lorsque le voisinage  $V_2$  est choisi, la fenêtre est incluse dans le support temps-fréquence du signal. La fenêtre "voit" alors un signal stationnaire et blanc, le son émis par le navire ne sera alors pas détectable (sauf aux frontières).

Un enseignement à tirer de cette petite analyse est que la forme de la fenêtre a un impact considérable sur les signaux que nous considérons comme détectables. Un choix judicieux de ce voisinage d'estimation du bruit pourra être utilisé pour discriminer certaines classes de signaux. Par exemple les bruits de navire au profit des sifflements de dauphins.

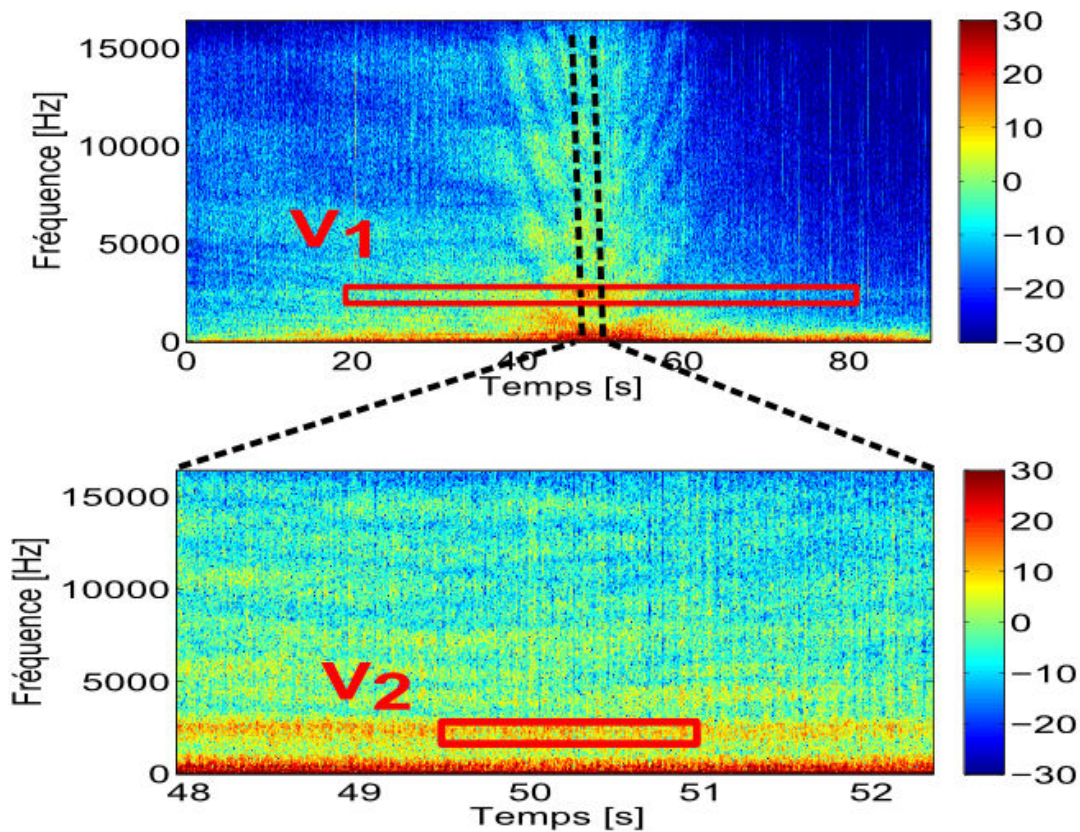


FIGURE 2.1 – La figure du haut montre le signal rayonné par une embarcation légère. Nous illustrons les conséquences du choix du voisinage  $V$  d'estimation du bruit. Le voisinage  $V_1$  n'est pas inclus dans le support temps-fréquence du signal, le signal émis par le navire est donc détectable. La figure du bas est un zoom sur quelques secondes de ce signal. Le voisinage  $V_2$  est inclus dans le support temps-fréquence du signal. Le signal émis par le navire n'est alors pas détectable.

## 2.3 Modèle du signal en temps-fréquence

### 2.3.1 Définition du support temps-fréquence du signal

Nous définissons  $\mathfrak{Z}(s, \mathbb{S}_s)$  le support temps-fréquence du signal  $s[n]$  pour la représentation temps-fréquence  $\mathbb{S}_s$ . La manière intuitive de définir  $\mathfrak{Z}(s, \mathbb{S}_s)$  est la suivante :

$$\mathfrak{Z}(s, \mathbb{S}_s) = \{[n, k] / \mathbb{S}_x[n, k] = \mathbb{S}_{s+b}[n, k]\} \quad (2.2)$$

C'est à dire que le support  $\mathfrak{Z}(s, \mathbb{S}_s)$  correspond à l'ensemble des points  $[n, k]$  dont l'énergie provient en partie du signal  $s[n]$ .

### 2.3.2 Formulation temps-fréquence du test d'hypothèse binaire

Nous cherchons donc à estimer le support temps-fréquence  $\mathfrak{Z}(s, \mathbb{S}_s)$  du signal  $s[n]$  à partir du spectrogramme  $\mathbb{S}_x[n, k]$  du signal  $x[n]$ . Cette estimation revient à faire un choix en chaque point  $[n, k]$  du spectrogramme. Ce choix est fait par comparaison de l'énergie observée en  $[n, k]$  et de l'énergie "attendue" dans le cas bruit seul.

Nous reformulons ceci en terme de test d'hypothèse binaire. En chaque point temps-fréquence  $[n, k]$  l'une ou l'autre de ces deux hypothèses doit être choisie : " $\mathcal{H}_0$  : bruit seul" ou " $\mathcal{H}_1$  : signal + bruit".

$$\begin{cases} \mathcal{H}_0 : \mathbb{S}_x[n, k] = \mathbb{S}_b[n, k] \\ \mathcal{H}_1 : \mathbb{S}_x[n, k] = \mathbb{S}_{s+b}[n, k] \end{cases} \quad (2.3)$$

Résoudre ce test d'hypothèse binaire en chaque point temps-fréquence revient donc à attribuer 1 lorsque l'hypothèse  $\mathcal{H}_1$  est retenue et 0 lorsque l'hypothèse  $\mathcal{H}_0$  est retenue.

Ce test d'hypothèse est extrêmement général au sens où aucune information sur la nature du signal rencontré au point  $[n, k]$  n'est générée lors de sa résolution. Nous ne concluons pas sur l'origine physique du signal. Ce test indique uniquement la présence d'énergie en  $[n, k]$  ne trouvant pas son origine dans notre définition du bruit.

En pratique la résolution de ce test d'hypothèse revient à comparer une grandeur d'étude, ici l'énergie observée  $\mathbb{S}_x[n, k]$ , à un seuil  $S^T[n, k]$ . Ce seuil n'est donc

pas nécessairement constant et peut varier en fonction du point d'étude  $[n, k]$ . La résolution du test correspond donc à la prise de décision suivante :

$$\begin{cases} \text{décider } \mathcal{H}_0 \text{ si } \mathbb{S}_x[n, k] \leq S^T[n, k] \\ \text{décider } \mathcal{H}_1 \text{ si } \mathbb{S}_x[n, k] > S^T[n, k] \end{cases} \quad (2.4)$$

Tout l'art de la résolution d'un test binaire réside donc dans le choix de la grandeur d'étude et le choix du seuil  $S^T[n, k]$ .

Nous rappelons encore une fois que le signal d'intérêt  $s[n]$  est aléatoire, et que la densité de probabilité  $p_{\mathcal{H}_1}(\mathbb{S}_x)$  de ses coefficients est inconnue. Les stratégies optimales classiques, au sens de minimisation du risque bayésien [Kay98], nous sont donc inaccessibles. La stratégie de seuillage que nous adoptons ici suit une approche dite de Neyman-Pearson [NP92]. Celle-ci consiste en la maximisation de la probabilité de détection sous contrainte d'une certaine probabilité de fausse alarme. Pour cela nous devons connaître les densités de probabilité  $p_{\mathcal{H}_0}(\mathbb{S}_x)$  des coefficients du spectrogramme sous l'hypothèse bruit seul.

Nous avons trois remarques à formuler pour conclure cette section :

**Remarque 1 :** La définition de  $\beth(s, \mathbb{S}_s)$  donnée à l'équation (2.2) n'est pas satisfaisante. L'effet d'étalement spectral dû au fenêtrage du signal disperse l'énergie du signal sur l'ensemble de la bande de fréquence correspondant à une date  $n$  à laquelle le signal  $s[n]$  est défini [Har78]. L'application stricte de cette définition reviendrait donc à détecter l'ensemble des dates temporelles pour lesquelles le signal est défini. Le support temps fréquence auquel nous nous intéressons en pratique est donc plus restrictif que cela. Nous cherchons à déterminer uniquement le support temps-fréquence du premier lobe du signal. Par facilité nous le dénommerons quand même par le terme de "support temps-fréquence du signal".

**Remarque 2 :** Le support  $\beth(s, \mathbb{S}_s)$  est dépendant de la représentation temps-fréquence choisie et des paramètres de construction de celle-ci. Un changement de représentation ou un changement des paramètres de construction changera le support temps-fréquence. Ainsi, le changement de la fenêtre de pondération change la taille du lobe principal et donc la forme de  $\beth(s, \mathbb{S}_s)$  (voir figure 2.2), de même la longueur de la fenêtre de pondération, adaptée aux non-stationnarités du signal, modifie là encore la définition de  $\beth(s, \mathbb{S}_s)$  (voir figure 2.3).

**Remarque 3 :** Pour simplifier les notations, nous abandonnons dans la suite de ce manuscrit le rappel de la dépendance au choix de la représentation temps-fréquence. Le support  $\beth(s, \mathbb{S}_s)$  devient donc  $\beth(s)$ . Nous devons cependant garder à

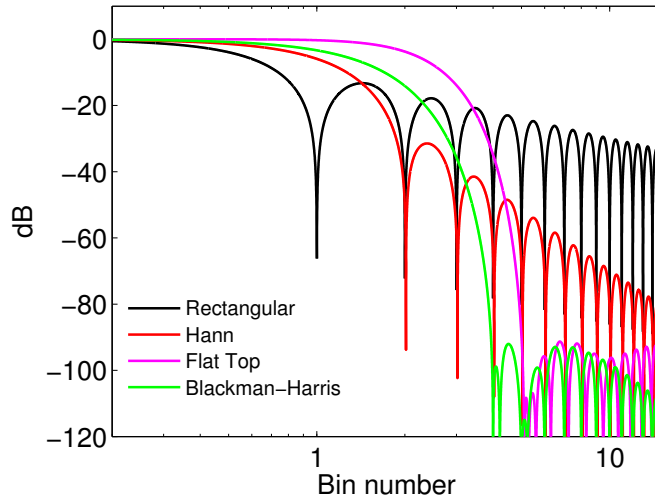


FIGURE 2.2 – Représentation spectrale de plusieurs fenêtres de pondération. La forme du lobe principal est dépendante de la forme de la fenêtre [Har78].

l'esprit que cette dépendance existe, et donc que le support temps-fréquence estimé n'est forcément pas unique.

## 2.4 Densité de probabilité du spectrogramme sous l'hypothèse bruit seul

Nous exprimons dans cette section les statistiques du spectrogramme dans le cas "bruit seul". Le bruit  $b[n]$  est une variable aléatoire Gaussienne. La fonction d'autocorrélation du bruit, notée  $\Gamma_b[n, \tau]$  à l'instant  $n$  et pour le décalage temporel  $\tau$  est définie de la manière suivante :

$$\Gamma_b[n, \tau] = E\left(b[n - \tau/2] b[n + \tau/2]\right) \quad (2.5)$$

La non-stationnarité du bruit se traduit par une dépendance à l'instant  $n$  auquel est calculée la fonction d'autocorrélation. De même la non blancher du bruit se traduit par une dépendance à  $\tau$ . Dans le cas stationnaire nous observons  $\Gamma_b[n, \tau] = \Gamma_b[\tau]$ . Le module carré de la TFD de  $\Gamma_b[\tau]$  est alors un estimateur asymptotiquement non-biaisé<sup>1</sup> de la densité spectrale de puissance (DSP)  $\gamma_b[k]$  du bruit (relation de Wiener-Khintchine).

Dans notre cas, la non-stationnarité des signaux d'intérêt nous oblige à prendre en compte la variation temporelle de la densité spectrale de puissance que nous notons alors  $\gamma_b[n, k]$ . Cette quantité est plus difficile à définir. Une adaptation di-

1. Un estimateur est asymptotiquement non-biaisé s'il le biais d'estimation tend vers 0 lorsque la taille de l'échantillon d'estimation tend vers l'infini

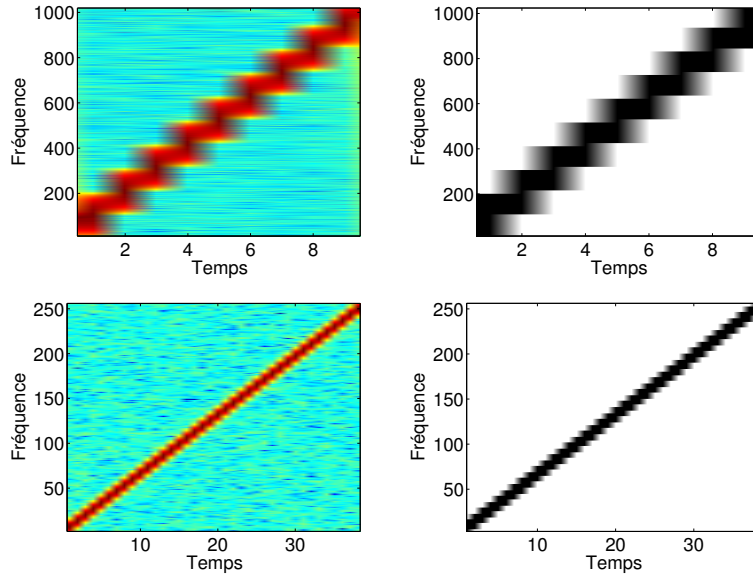


FIGURE 2.3 – Représentations temps-fréquence d’une même modulation linéaire de fréquence. La première ligne montre le spectrogramme du signal calculé avec une fenêtre de pondération de 2048 points et son support temps-fréquence. La seconde ligne montre le spectrogramme du même signal et son support temps-fréquence correspondant, cette fois pour une fenêtre de pondération de longueur 512 points. Le support temps-fréquence du signal tel que nous le définissons est dépendant du choix de la représentation.

recte de l’estimateur précédent dans le cas non-stationnaire consiste donc à estimer  $\gamma_b[n, k]$  en considérant le module carré de la TFCT du signal  $s[n]$  sur un segment de longueur  $M$ . La condition asymptotique n’est cependant plus remplie, nous avons affaire à un estimateur biaisé .

Nous utilisons ici une définition très intuitive de la densité spectrale de puissance du bruit introduite par [Mar70] :

$$\gamma_b[n, k] = E(\mathbb{S}_b[n, k]) \quad (2.6)$$

Ainsi au lieu de considérer directement le module carré de la TFCT discrète, nous nous intéressons à son espérance mathématique. Cette grandeur a été introduite sous le nom de spectre physique. Nous la confondrons avec la notion de densité spectrale de puissance dans la suite de ce travail.

Utilisant cette définition de la DSP, il est montré dans [HMM08] que la densité de probabilité des coefficients du spectrogramme d’un signal centré Gaussien est très bien approximée par une loi du  $\chi^2$  à deux degrés de liberté prenant la forme suivante :

$$p_{\mathcal{H}_0}(\mathbb{S}_x[n, k]) = \frac{1}{\gamma_b[n, k]} \exp\left(-\frac{\mathbb{S}_x[n, k]}{\gamma_b[n, k]}\right), \quad (2.7)$$

Cette expression est intéressante car elle montre que l'expression de  $p_{\mathcal{H}_0}(\mathbb{S}_x[n, k])$  est uniquement dépendante de  $\gamma_b[n, k]$ .

**Remarque 1 :** Dans le cas d'un bruit présentant de la corrélation ou de fortes non-stationnarités, la modélisation de  $p_{\mathcal{H}_0}(\mathbb{S}_x)$  par une loi du  $\chi^2$  (voir équation 2.7) n'est valable que sous réserve de l'utilisation d'une fenêtre de pondération atteignant 0 à ses bornes [HMM08]. Le calcul du spectrogramme avec fenêtre d'analyse rectangulaire est donc prohibé sous nos hypothèses.

**Remarque 2 :** L'émergence de la loi du  $\chi^2$  à deux degrés de liberté est naturelle. La cause en est la nature linéaire de la transformation appliquée sur le signal. La transformée de Fourier d'une variable Gaussienne reste Gaussienne. Son module carré est donc la somme du carré de deux variables Gaussiennes. La loi résultante est donc une loi du  $\chi^2$  à deux degrés de liberté.

**Remarque 3 :** En suivant le même raisonnement nous pouvons donc espérer obtenir une loi simple pour quantité d'autres transformations : ondelettes, chirplettes, transformée de Fourier fractionnaire, etc. Il reste cependant à en étudier les conditions d'application (conditions de circularité de ces transformations [HMM08, MM11]) .

## 2.5 Estimation locale de la densité spectrale de puissance (DSP) du bruit

La connaissance de  $\gamma_b$  en chaque point  $[n, k]$  apparaît donc comme cruciale dans le processus d'estimation de  $p_{\mathcal{H}_0}(\mathbb{S}_x)$ . La DSP du bruit  $\gamma_b[n, k]$  est définie comme l'espérance de  $\mathbb{S}_x[n, k]$  sous  $\mathcal{H}_0$ . Nous sommes alors confrontés à un problème de taille : nous n'avons accès qu'à une seule réalisation de  $\mathbb{S}_x[n, k]$  et nous ne savons même pas si elle provient de  $\mathcal{H}_0$  ou de  $\mathcal{H}_1$ . Nous avons donc, au mieux, une seule réalisation pour calculer l'espérance des coefficients du spectrogramme provenant du bruit, ce qui est assez peu... Afin de rendre cette tâche d'estimation réalisable nous relâchons légèrement l'hypothèse de non-blancité et de non-stationnarité du bruit. Nous supposons le bruit localement blanc et stationnaire, c'est à dire qu'il existe un voisinage temps-fréquence  $V$ , centré en  $[n, k]$  sur lequel la DSP du bruit est approximativement constante. L'estimation de la DSP aura lieu sur ce voisinage.

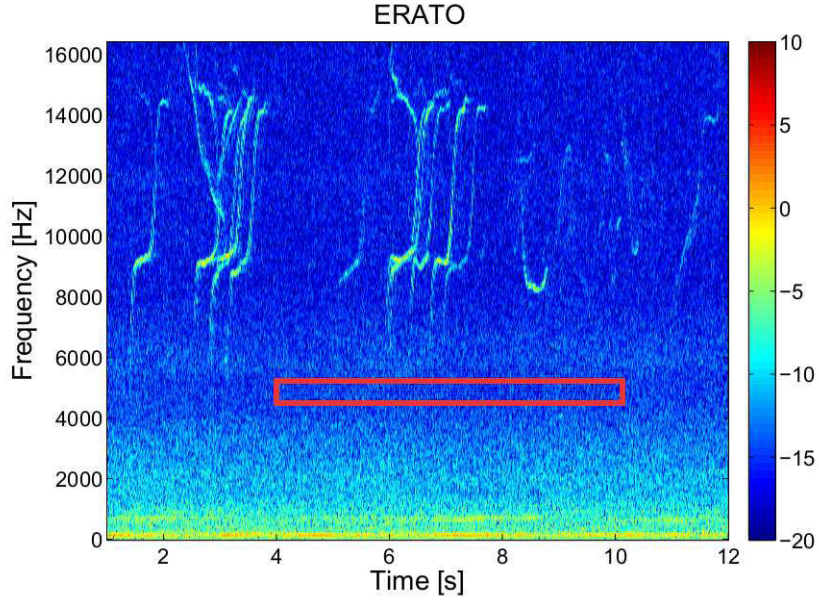


FIGURE 2.4 – Exemple de forme de voisinage temps-fréquence  $V$  d'estimation de la densité spectrale de puissance  $\gamma_b[n, k]$ . (Source : ERATO-09 SHOM)

Prenons l'exemple des données provenant de la campagne ERATO-09 (voir figure 2.4). Le bruit n'est ni blanc ni stationnaire mais il semble raisonnable de considérer le bruit comme localement stationnaire et blanc sur un voisinage temps-fréquence de longueur 6 secondes et sur une bande de fréquence de largeur 500Hz centré en  $[n, k]$ . Le résultat est donc le choix d'une région temps-fréquence de taille  $[6s \times 500Hz]$  centré sur le point  $[n, k]$  et sur laquelle la DSP du bruit  $\gamma_b[n, k]$  est considérée constante. L'exemple d'une des régions d'estimation est tracée en rouge sur la figure 2.4.

Nous cherchons donc à estimer l'espérance des coefficients  $\mathbb{S}_b[n, k]$  sur ce voisinage  $V$ . Grâce à la relaxe des hypothèses de non-blanchéur et non-stationnarité du signal nous avons maintenant accès à plusieurs réalisations de  $\mathbb{S}_x[n, k]$ . A partir de ces réalisations de  $\mathbb{S}_x$  nous pouvons maintenant fournir une estimation de  $E\{\mathbb{S}_b[n, k]\}$ .

Le problème d'estimation se formule de la manière suivante. Nous avons accès à  $\mathbb{S}_x^V[n, k]$ , l'ensemble des coefficients du spectrogramme dans le voisinage  $V$  d'estimation, centré autour du point  $[n, k]$  :

$$\mathbb{S}_x^V[n, k] = \{\mathbb{S}_x(1), \dots, \mathbb{S}_x(T)\} \quad (2.8)$$

Parmi ces  $T$  réalisations  $\mathbb{S}_x(i)$ , un certain nombre (que nous appelons  $N_b$ ) sont des coefficients issus du bruit et un certain nombre  $N_s = T - N_b$  sont issus du signal.

Dans le cas  $N_s = 0$ , c'est à dire que le voisinage  $V$  ne contient que des coefficients issus du bruit, le problème d'estimation est simple. Il suffit de prendre la moyenne arithmétique des  $\mathbb{S}_x(i)$ , nous avons alors un estimateur non-biaisé de l'espérance et donc de  $\gamma_b[n, k]$ .

Les problèmes deviennent plus sérieux lorsque  $N_s > 0$ , c'est à dire que le voisinage  $V$  contient des coefficients issus du bruit  $\mathcal{H}_0$  et du signal  $\mathcal{H}_1$ . La moyenne arithmétique est alors un estimateur biaisé de l'espérance. Si nous étions capables de différencier les coefficients provenant du bruit de ceux provenant du signal alors nous pourrions calculer  $\gamma_b[n, k]$ , mais nous n'en aurions plus besoin puisque nous cherchons précisément à résoudre ce problème !

Nous ne considérons ici que les méthodes permettant d'estimer le bruit ne nécessitant pas d'avoir accès à des trames temporelles ne comprenant que du bruit seul. Trois types de stratégies sont alors mises en place pour résoudre ce problème : les méthodes itératives, les méthodes par lissage, et les méthodes utilisant les statistiques minimales [RD01, YR06, SFB00].

### 2.5.1 Méthodes de lissage

Ces méthodes cherchent à utiliser des opérateurs de lissage pour estimer la DSP du bruit. Ces opérations de lissage sont pondérées par une probabilité de présence de signal, afin de fournir une estimation la moins biaisée possible de la DSP du bruit. La probabilité de présence peut être établie en mesurant un RSB *a priori* [Coh04], ou en mesurant le ratio entre le spectre et sa valeur minimale [RL06] ou encore par l'établissement d'un modèle de présence du signal [RD01, Coh03]. Elles présentent l'avantage de s'adapter très rapidement aux non-stationnarités du bruit. Ces méthodes présentent en général plusieurs paramètres affectant fortement les performances de l'algorithme. La question de la probabilité de présence du signal reste ouverte. De plus, rappelons que dans ce manuscrit notre philosophie est de faire le moins d'hypothèses possibles sur le signal ou sa présence, nous ne choisirons donc pas cette classe de méthodes.

### 2.5.2 Méthodes conjointes d'estimation du bruit et de détection

Ces méthodes cherchent à estimer itérativement la DSP du bruit. Pour cela les points représentatifs du signal au sein du voisinage  $V$  sont éliminés à chaque itération. Une estimation de la DSP est alors donnée, ce qui permet d'éliminer à nouveau certains des coefficients. Ces méthodes résolvent donc simultanément le problème d'estimation et de détection du signal. Ces méthodes présentent les



défauts habituels des méthodes itératives telles que les temps de calcul, spécialement lorsque le critère est un moment d'ordre supérieur [MM11, YR06], et le problème de convergence du critère d'arrêt. Un avantage majeur de la méthode présentée dans [MM11] est qu'aucune hypothèse sur la présence de signal ou ses statistiques n'est utilisée. La méthode d'estimation est seulement basée sur la valeur du Kurtosis, qui est connue dans le cas Gaussien. Les grandes valeurs sont alors supprimées jusqu'à atteindre la valeur théorique du Kurtosis d'une variable Gaussienne (i.e. une valeur nulle). Le principale inconvénient vient du nombre d'échantillons nécessaires à l'estimation du Kurtosis.

### 2.5.3 Méthodes à statistiques minimales

L'idée retenue par cette classe de méthode est la suivante : plutôt que de faire des hypothèses sur le signal, la DSP du bruit est estimée à partir des plus petits coefficients au sein du voisinage. Ceux-ci sont très probablement des réalisations du bruit seul :

Un estimateur utilisant la valeur minimale observée sur un voisinage est utilisée dans [Mar94], cette valeur minimale doit être corrigée (par un gain multiplicatif) afin d'estimer la DSP. La détermination du gain multiplicatif est ici le problème central. Additionnellement, le minimum possède une forte variance, des stratégies de filtrage ont donc été mises en place pour réduire cette variance [Mar01].

Pour réduire la variance de l'observation du minimum sans filtrage, des estimateurs de la DSP à partir du  $q^{ime}$  quantile ont été proposés [SFB00]. Le choix du gain multiplicatif n'est cependant pas résolu.

Un cas particulier est celui du quantile médian pour lequel aucun gain multiplicatif n'est nécessaire, la médiane étant directement un estimateur de l'espérance. Cet estimateur est robuste à la présence de signal tant que le signal n'occupe pas plus de 50% de la fenêtre d'estimation. Sans aller jusqu'à formaliser la méthode en termes d'estimation de densité spectrale de puissance du bruit, la méthode d'estimation par la médiane est déjà utilisée dans la communauté PAM [GCGW13, BHTM14, Mel13]. Cette méthode surestimera la densité spectrale de puissance du bruit lorsque un grand nombre de signaux est observé simultanément (en particulier en présence de clics, buzzs et sifflements émis simultanément tel qu'observé en figure 2.9a).

### 2.5.4 Méthode choisie dans ce travail : estimation par les Z-minima

Nous choisissons l'approche du problème par les statistiques minimales car elle semble très intuitive et ne fait pas intervenir d'hypothèses sur le signal. Le problème d'estimation de la DSP du bruit à partir du  $q^{eme}$  quantile (en utilisant un

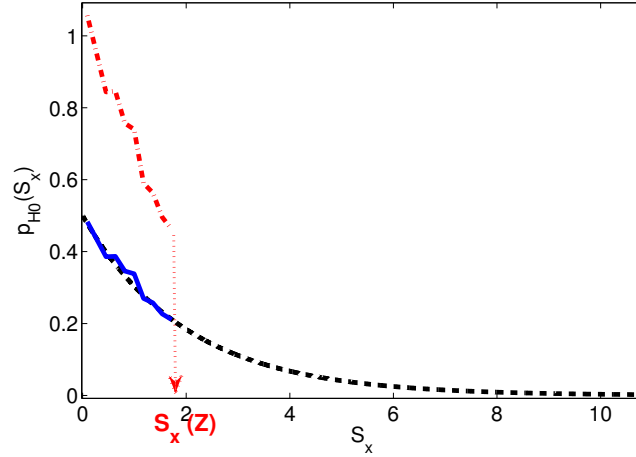


FIGURE 2.5 – Représentation de la densité de probabilité  $p_{\mathcal{H}_0}(\mathbb{S}_x)$  théorique (en pointillé), de  $p_{\mathcal{H}_0}(\mathbb{S}_x | \mathbb{S}_x \leq \mathbb{S}_x(Z)) \times p_{\mathcal{H}_0}(\mathbb{S}_x \leq \mathbb{S}_x(Z))$  (en ligne pleine) et de  $p_{\mathcal{H}_0}(\mathbb{S}_x | \mathbb{S}_x \leq \mathbb{S}_x(Z))$ . En pratique nous observons uniquement la densité de probabilité  $p_{\mathcal{H}_0}(\mathbb{S}_x | \mathbb{S}_x \leq \mathbb{S}_x(Z))$  à partir de laquelle nous estimons  $\gamma_b[n, k]$ .

quantile inférieur à la médiane) est adressé dans [Hui08]. Nous résumons ce travail succinctement, pour une analyse détaillée, se reporter directement à l’original.

Nous supposons que les  $Z$  valeurs les plus faibles sur le voisinage  $V$  sont des réalisations du bruit seul :

$$\mathbb{S}_x^V[n, k] = \underbrace{\{\mathbb{S}_x(1), \dots, \mathbb{S}_x(Z), \dots, \mathbb{S}_x(N_b), \dots, \mathbb{S}_x(T)\}}_{\text{bruit}} \quad (2.9)$$

En pratique il suffit de choisir un quantile suffisamment faible pour être sûr de n’observer que du bruit, nous choisissons  $Z = T/10$  (c’est à dire le 10<sup>ème</sup> quantile). Comment estimer la DSP du bruit à partir de ces  $Z$  valeurs les plus faibles observées sur le voisinage  $V$  ? En d’autres termes, nous cherchons à estimer  $\gamma_b[n, k]$  à partir de :

$$p_{\mathcal{H}_0}(\mathbb{S}_x | \mathbb{S}_x \leq \mathbb{S}_x(Z)) \quad (2.10)$$

La figure 2.5 est une représentation visuelle du problème auquel nous sommes confrontés. Nous observons la densité de probabilité tracée en rouge à partir de laquelle nous devons estimer la loi du  $\chi^2$  tracée en noir. La courbe tracée en bleu correspond à  $p_{\mathcal{H}_0}(\mathbb{S}_x | \mathbb{S}_x \leq \mathbb{S}_x(Z)) \times p_{\mathcal{H}_0}(\mathbb{S}_x \leq \mathbb{S}_x(Z))$ .

La méthodologie proposée par [Hui08] propose de résoudre ce problème d’estimation. Elle repose sur une estimation conjointe du nombre de coefficients bruit seul et de la DSP du bruit. Ce problème est un sous-problème de ceux présentés précédemment qui cherchent à réaliser conjointement l’estimation de la DSP du

bruit et la recherche des coefficients "signal". Un voisinage donné contient  $T$  coefficients, parmi lesquels  $N$  sont du bruit et  $T-N$  du signal. En ajustant un seuil équivalent à un RSB de 0dB, le nombre de coefficients  $N_{cand}$  supérieurs à ce seuil sera composé des  $N - T$  coefficients "signal" et du nombre de fausses alarmes  $N \times p_{fa,0dB}$  :

$$N_{cand} = (T - N) + Np_{fa,0dB} \quad (2.11)$$

Il est montré à partir de l'équation (2.10) qu'un estimateur de  $\gamma_b[n, k]$  est le suivant :

$$\gamma_b[n, k] = m_Z + \mathbb{S}_x(Z) \left( \frac{N}{Z} - 1 \right) \quad (2.12)$$

avec  $m_Z$  la moyenne des  $Z$  plus petits coefficients sur le voisinage. L'estimation itérative conjointe de  $N$  et  $\gamma_b$  est effectuée de la manière suivante :

$$\begin{aligned} N_0 &= T \\ \text{for } i &= 0 : 10 \\ &\hat{\gamma}_{b,i} = m_Z + \mathbb{S}_x^{(Z)} \left( \frac{N}{Z} - 1 \right) \\ &seuil_i = 1.26\hat{\gamma}_{b,i} \\ &N_{cand,i} = \text{Card}\{\mathbb{S}_x^{(V)} > seuil_i\} \\ &\hat{N}_i = \frac{T - N_{cand,i}}{1 - p_{fa,0dB}} \\ \text{end} \end{aligned} \quad (2.13)$$

Trois exemples d'estimations du bruit sur données réelles sont données en figure 2.7b, 2.8b et 2.9b. Une fois  $\gamma_b[n, k]$  estimée en chaque point temps-fréquence nous pouvons passer à l'étape de détection.

## 2.6 Résolution du test d'hypothèse binaire

Nous cherchons à résoudre le test d'hypothèse binaire par une approche de Neyman-Pearson, c'est à dire que nous imposons une probabilité de fausse alarme constante lors de la résolution. Cette approche ne nécessite que la connaissance des densités de probabilité sous  $\mathcal{H}_0$ . La probabilité de fausse alarme est définie de la manière suivante :

$$p_{fa} = \int_{ST}^{+\infty} p_{\mathcal{H}_0}(\mathbb{S}_x) d\mathbb{S}_x \quad (2.14)$$

La figure 2.6 représente visuellement la notion de probabilité de fausse alarme,

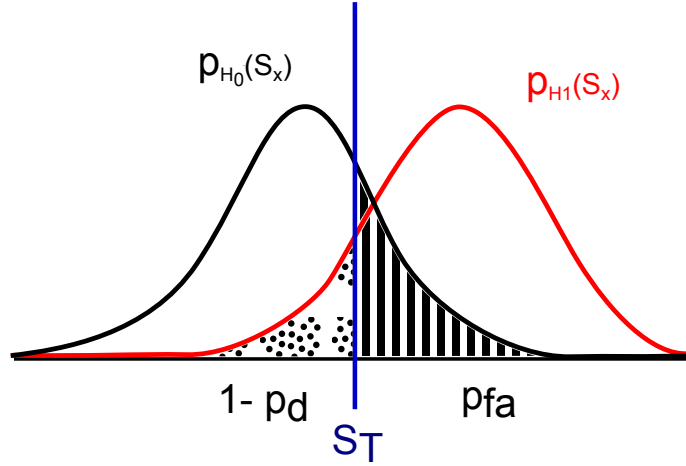


FIGURE 2.6 – Représentation schématique de deux lois de probabilité. La densité de probabilité noire correspond à la loi observée sous  $\mathcal{H}_0$ , celle en rouge correspond à la loi sous  $\mathcal{H}_1$ . L'aire hachurée représente la probabilité de fausse alarme correspondant au seuil  $S_T$ . L'aire avec des points correspond à 1 moins la probabilité de détection.

c'est à dire l'aire sous la courbe de la densité de probabilité sous l'hypothèse  $\mathcal{H}_0$ , qui est supérieure à un seuil. Dans notre cas, en introduisant la densité de probabilité décrivant les statistiques du spectrogramme et présentée dans l'équation (2.7) et estimée grâce au processus itératif (2.13) nous obtenons :

$$p_{fa} = \exp\left(-\frac{S^T[n, k]}{\gamma_b[n, k]}\right) \quad (2.15)$$

Le choix de  $S^T[n, k]$  est donc localement adapté à  $\gamma_b[n, k]$  de manière à obtenir  $p_{fa}$  constante. Le seuil  $S^T[n, k]$  s'exprime de la manière suivante :

$$S_T[n, k] = -\hat{\gamma}_b[n, k] \ln(p_{fa}) \quad (2.16)$$

Connaissant le seuil nous pouvons maintenant résoudre le test d'hypothèse binaire. Nous appelons  $\phi_{p_{fa}}[n, k]$  la fonction indicatrice de la décision prise au point  $[n, k]$  telle que :

$$\phi_{p_{fa}}[n, k] = \begin{cases} 0, & \text{if } \mathbb{S}_x[n, k] < S_T[n, k] \\ 1, & \text{if } \mathbb{S}_x[n, k] \geq S_T[n, k], \end{cases} \quad (2.17)$$

Ce test binaire résolu nous obtenons une estimation  $\hat{\Xi}(s, p_{fa})$  de  $\Xi(s)$  (dépendante de  $p_{fa}$ ).

**Remarque :** Une grandeur équivalente à la probabilité de fausse alarme est introduite dans [HM06]. Celle-ci prend le nom de RSB de focus et est définie comme

$\xi_0$ (dB)	0	1	2	3	5	7	9	10	12
$p_{fa}$	0.28	0.26	0.24	0.22	0.16	0.1	0.05	0.03	0.006

TABLE 2.1 – Répartition des  $p_{fa}$  et leur équivalence en terme de RSB de focus

le RSB pour lequel le détecteur est réglé de manière optimale au sens du maximum de vraisemblance. Sans donner les détails de calcul le lien entre  $p_{fa}$  et le RSB de focus  $\xi_0$  est le suivant :

$$p_{fa} = \exp(-(0,26\xi_0 + 1)) \quad (2.18)$$

Ce RSB de focus quantifie l'émergence minimale (en terme de décibels) que doit avoir le signal pour être détecté. Cette formulation est très intuitive, dans la suite nous donnerons donc très souvent l'équivalence entre  $p_{fa}$  et RSB de focus. Le tableau 2.1 montre quelques exemples d'équivalence entre RSB de focus et  $p_{fa}$ .

## 2.7 Exemples sur données réelles

Nous proposons quelques exemples d'application de cette méthodologie d'estimation et de détection sur données réelles pour des cas assez divers.

Le premier exemple est tiré de la base de données ERATO-09 fournie par le SHOM<sup>2</sup>. L'enregistrement contient notamment des sifflements de dauphins *delphinus delphis*. Nous traçons le spectrogramme de cet enregistrement (figure 2.7a) dans le quel nous observons deux sifflements superposés. La densité spectrale de puissance du bruit est estimée à partir du spectrogramme (2.7b). Le voisinage temps-fréquence d'estimation  $V$  est  $[1s \times 500Hz]$ . L'estimation n'est pas affectée par la présence de signal. Le spectrogramme est seuillé pour quatre probabilités de fausse alarme différentes :  $p_{fa} = 0.21, 0.1, 0.027, 0.006$ . Ce qui correspond à des RSB de focus de 3, 7, 10 et 12 dB. Bien entendu, plus la probabilité de fausse alarme est élevée plus la probabilité de détection du signal l'est aussi, c'est ce que nous vérifions sur les différentes figures.

Le second exemple est tiré d'un enregistrement de grands dauphins *tursiops truncatus* réalisés près de l'île de Molène réalisés dans le cadre des activités de la Chaire Chorus<sup>3</sup>. Le spectrogramme est tracé en figure 2.8 Cette fois les sifflements émis par les dauphins sont mélangés à des signaux impulsionnels émis par l'activité benthique, et en particulier les crevettes claqueuses. Le voisinage temps-fréquence

2. Service Hydro-Océanographique de la Marine

3. Chaire industrielle d'excellence Chorus issue de la collaboration entre GIPSA-Lab et l'IUEM (Institut Universitaire Européen de la Mer) à Brest. L'activité principale de Chorus concerne le monitoring des écosystèmes marins par acoustique passive. Les titulaires de la chaire sont Cédric Gervaise et Lucia Di Iorio.

d'estimation  $V$  est  $[1s \times 500Hz]$ . L'estimation de la DSP du bruit semble cohérente malgré la présence de bruit impulsionnel. Le spectrogramme est seuillé pour quatre probabilités de fausse alarme différentes :  $p_{fa} = 0.21, 0.1, 0.027, 0.006$ . Ce qui correspond à des RSB de focus de 3, 7, 10 et 12 dB. Dans ce cas nous détectons à la fois les sifflements et les impulsions, ce qui est logique puisque aucun des signaux en question n'a un support temps-fréquence plus grand que le voisinage d'estimation  $V$ . Là encore nous observons le compromis entre probabilité de fausse alarme et probabilité de détection.

Le troisième exemple est le plus complexe des trois (voir figure 2.9a). Il provient de la base de données MobySound. Le spectrogramme est généré à partir de cet enregistrement (calculé sur 1024 points avec une fenêtre de Hann de 1024 points et 512 points de recouvrement). Nous observons dans le spectrogramme un grand nombre de sifflements et de clics. La DSP du bruit est estimée encore une fois sur un voisinage temps-fréquence  $V$  de  $[1s \times 500Hz]$ . L'estimation obtenue semble cohérente avec ce que nous observons, même si certaines zones semblent avoir été biaisées par la présence de signal (voir par exemple la région à 0.3s et autour de 25kHz). Le spectrogramme est à nouveau seuillé pour les quatre mêmes probabilités de fausse alarme. Nous détectons à la fois les sifflements et les impulsions, ce qui est encore une fois logique, l'explication étant la même que pour l'exemple précédent.

## Analyse des résultats

Nous sommes capables, à partir du spectrogramme, d'estimer la densité spectrale de puissance du bruit. Cette connaissance est utilisée pour seuiller le spectrogramme sans connaissance *a priori* du signal d'intérêt. Nous avons trois remarques à formuler sur les résultats obtenus.

Nous constatons d'abord que dans le cas d'un bruit impulsionnel, les détections associées aux sifflements et celles associées aux impulsions sont mélangées, que ces impulsions proviennent de clics de dauphins ou de l'activité benthique. À ce stade, la méthode de détection quadratique que nous avons présenté ne permet pas de séparer les sifflements des impulsions.

La seconde remarque est liée à la méthode de détection elle-même. Le critère de Neyman-Pearson nous assure de maximiser la probabilité de détection sous contrainte de probabilité de fausse alarme donnée. Nous ne pouvons cependant garantir que la probabilité de fausse alarme choisie permette de détecter quoi que ce soit. Si nous choisissons une probabilité de fausse alarme trop faible, le signal ne sera ainsi pas détecté (voir figure 2.7f par exemple). Nous ne possédons pas de critère nous permettant de choisir cette probabilité de fausse alarme puisque le

RSB du signal est inconnu.

La troisième remarque est liée à la seconde. Ne possédant pas de critère concernant le choix de la probabilité de fausse alarme, celle-ci est donc généralement choisie de manière *ad hoc*. Deux manières de procéder sont alors possibles. La première consiste à seuiller le spectrogramme avec une probabilité de fausse alarme très faible, une partie du signal est manquée mais le spectrogramme binaire est composée majoritairement de détections auxquelles nous pouvons "faire confiance", c'est à dire qu'elle sont très probablement issues du signal d'intérêt. L'alternative consiste alors à choisir une grande probabilité de fausse alarme. Ce faisant, nous espérons pouvoir détecter une grande majorité du signal, la contrepartie étant le très grand nombre de fausses alarmes générées, le tri entre "vraie détection" et "fausse alarme" est alors plus difficile à faire.

Nous nous plaçons dans ce travail dans le cadre de la détection de signaux possédant une énergie faible. Nous nous attacherons donc à développer une méthode permettant d'être capable de traiter un grand nombre de fausses détections.

## 2.8 Conclusions et problématique

Dans ce chapitre nous avons formulé le contexte de la détection d'un signal dans le plan temps-fréquence. La nature non-stationnaire et non-blanche des signaux étudiés nous force à appliquer le processus de détection du signal en chaque point temps-fréquence. L'absence d'*a priori* sur le signal d'intérêt nous incite à adopter une approche de Neyman-Pearson. Cette approche nécessite la connaissance de la densité de probabilité du spectrogramme sous l'hypothèse bruit seul. La densité de probabilité, définie en chaque point  $[n, k]$  prend la forme du loi du  $\chi^2$  à deux degrés de liberté, et sa forme ne dépend que de  $\gamma_b[n, k]$ , le spectre physique du bruit que nous confondons avec la notion de densité spectrale de puissance du bruit. Nous utilisons une méthode d'estimation de cette densité spectrale de puissance basée sur les statistiques minimales. La loi de probabilité dans le cas bruit seul, un critère de Neyman-Pearson permet de binariser le spectrogramme de manière adaptée au bruit en chaque point  $[n, k]$ .

Nous nous intéressons dans ce travail aux signaux à bas RSB en milieu impulsif (clics et activité benthique). Nous avons identifié trois limitations principales à la méthode de détection quadratique présentée dans ce chapitre.

1. L'approche de Neyman-Pearson nous permet de garantir la maximisation de la probabilité de détection sous contrainte d'une certaine probabilité de fausse alarme, mais ne garantit aucunement la probabilité de détection du signal. Le RSB du signal étant inconnu, le choix de la probabilité de fausse alarme est difficile.

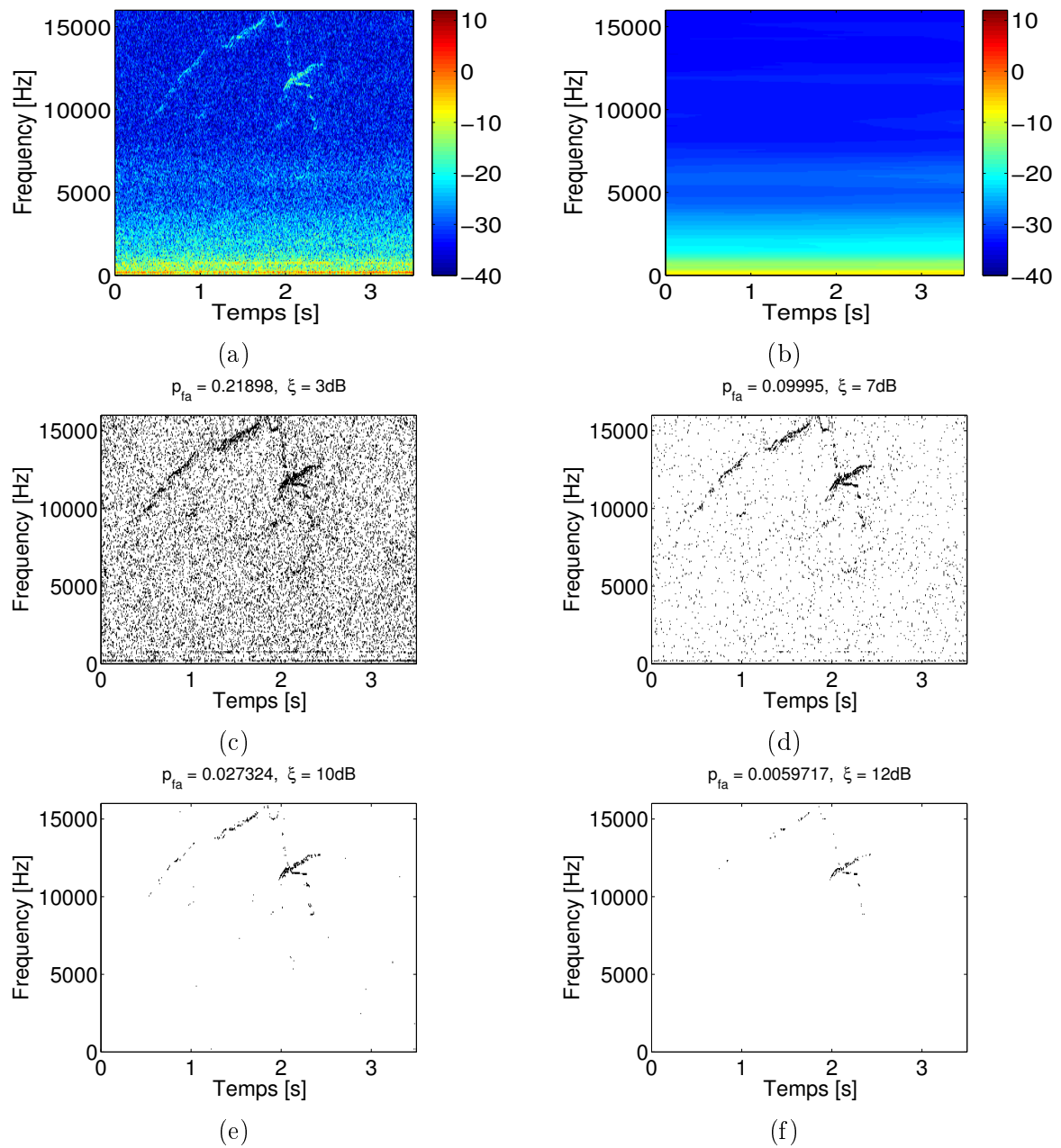


FIGURE 2.7 – Exemple d’un signal de dauphin commun *delphinus delphis* enregistré lors de la campagne ERATO-09 (SHOM) a) Spectrogramme du signal (calculé sur 512 points avec une fenêtre de Hann de 512 points et 256 points de recouvrement), b) Densité spectrale de puissance estimée, c) Seuillage à  $p_{fa} = 0.21$ , d) Seuillage à  $p_{fa} = 0.1$ , e) Seuillage à  $p_{fa} = 0.027$  et f) Seuillage à  $p_{fa} = 0.006$



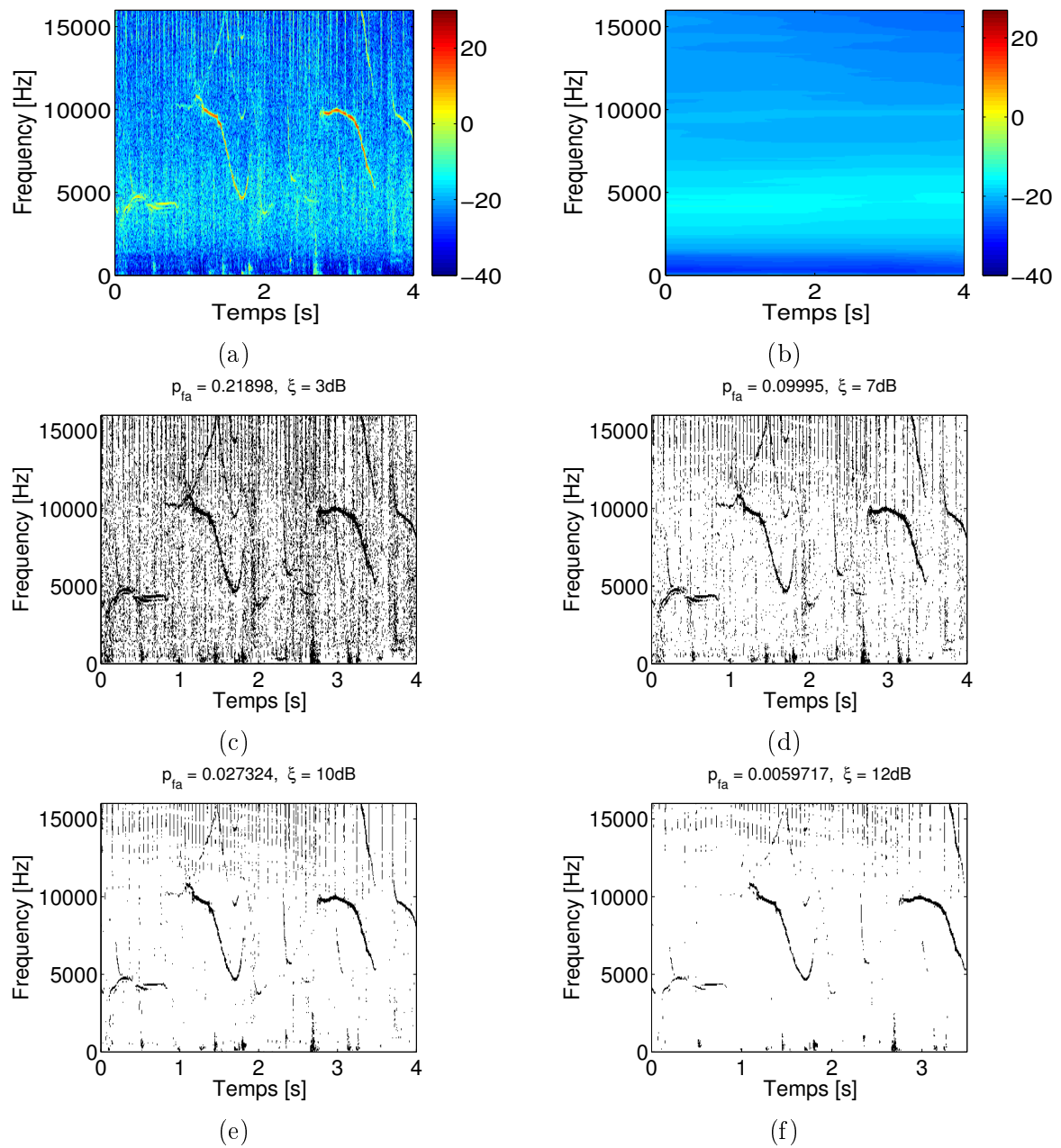


FIGURE 2.8 – Exemple d'un signal de grands dauphins *tursiops truncatus* enregistré au large de l'île de Molène (Chaire Chorus) a) Spectrogramme du signal (calculé sur 512 points avec une fenêtre de Hann de 512 points et 256 points de recouvrement), b) Densité spectrale de puissance estimée, c) Seuillage à  $p_{fa} = 0.21$ , d) Seuillage à  $p_{fa} = 0.1$ , e) Seuillage à  $p_{fa} = 0.027$  et f) Seuillage à  $p_{fa} = 0.006$

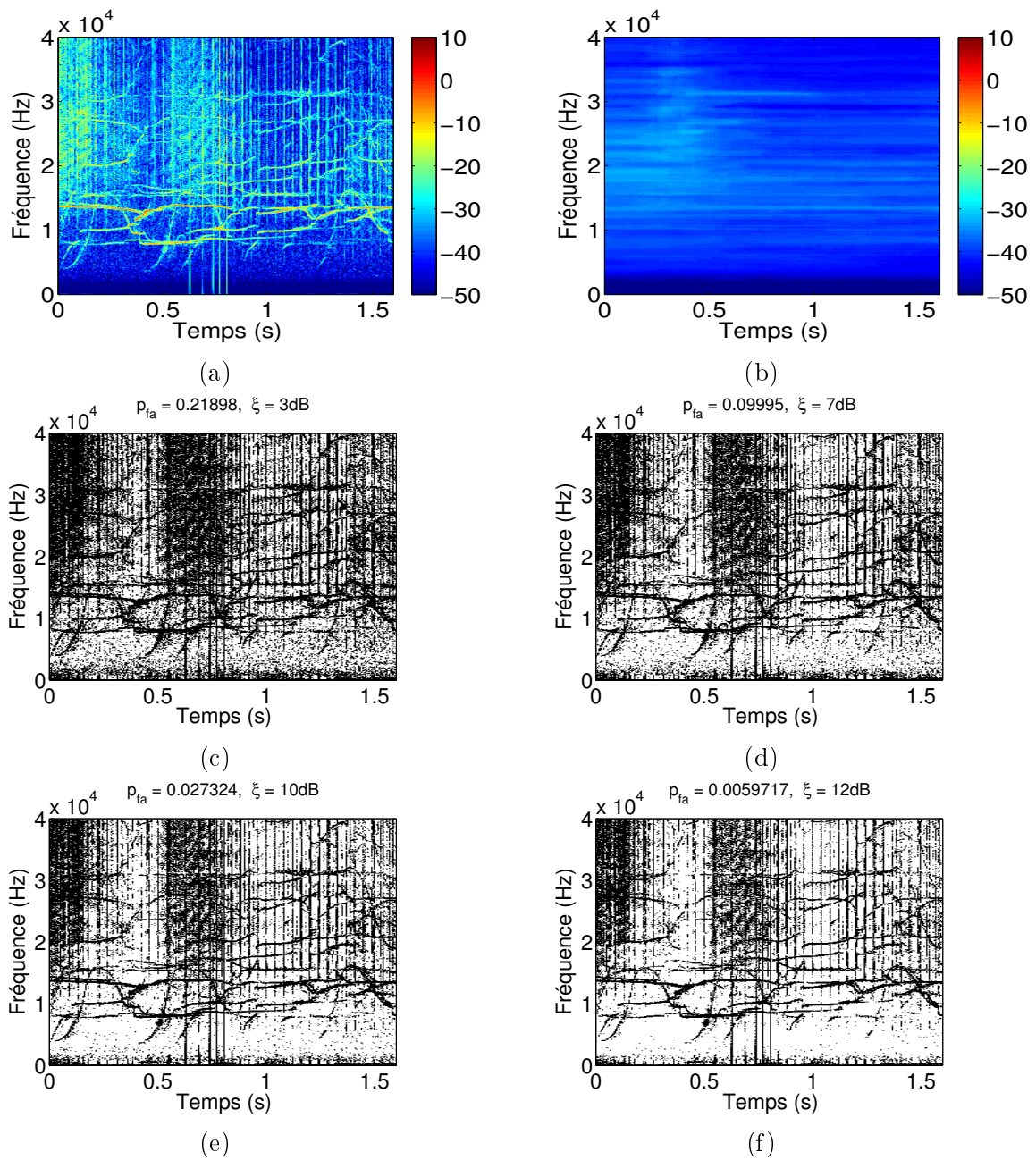


FIGURE 2.9 – Exemple d’un signal de dauphins communs *delphinus delphis* faisant partie de la base de données Mobysound a) Spectrogramme du signal (calculé sur 1024 points avec une fenêtre de Hann de 1024 points et 512 points de recouvrement), b) Densité spectrale de puissance estimée, c) Seuillage à  $p_{fa} = 0.21$ , d) Seuillage à  $p_{fa} = 0.1$ , e) Seuillage à  $p_{fa} = 0.027$  et f) Seuillage à  $p_{fa} = 0.006$  (MobySound : ‘Qx-Dc-FLIP0610-VLA-061015-163500.wav’)

2. Nous nous intéressons à des signaux à faible RSB. Afin de pouvoir les détecter nous serons donc naturellement amenés à choisir de grandes probabilités de fausse alarme. Un enjeu concerne donc la discrimination entre vraies détections et fausses alarmes au sein du spectrogramme binaire. Cette discrimination est difficile lorsque le nombre de fausses alarmes est très grand.
3. Le détecteur quadratique ne permet pas de différencier à lui seul les détections issues de signaux à bande étroite (modulation non linéaires de fréquences) et les signaux impulsionnels large bande. Nous observons en pratique que ces deux catégories de signaux se superposent temporellement chez les mammifères marins.

Ces trois limitations posent le cadre du travail développé dans cette thèse. Le chapitre suivant est dédié à l'analyse de la densité des fausses alarmes dans le spectrogramme binaire. Nous développons un cadre statistique permettant de discriminer les régions temps-fréquence abritant du signal de celles abritant uniquement du bruit (i.e. des fausses alarmes). Ce cadre statistique permet de gérer efficacement des spectrogrammes binaires fortement corrompus par des fausses alarmes.

# Chapitre 3

## Étude statistique de l'organisation des maxima temps-fréquence

### Sommaire

---

<b>3.1</b>	<b>Introduction</b>	<b>68</b>
<b>3.2</b>	<b>Segmentation de composantes temps-fréquence</b>	<b>70</b>
<b>3.3</b>	<b>Formulation du problème</b>	<b>73</b>
3.3.1	Définition du support temps-fréquence étendu $\mathfrak{Q}_{\mathcal{R}}$	75
3.3.2	Formulation du test d'hypothèse binaire BHT2	76
3.3.3	Densité de probabilité de la carte de comptage $\Psi$ sous $\mathcal{H}_0^R$	77
3.3.4	Résumé de la méthode	78
<b>3.4</b>	<b>Degrés de liberté de la méthode</b>	<b>79</b>
3.4.1	Probabilité de fausse alarme $p_{fa}$	79
3.4.2	Partitionnement du spectrogramme binaire : choix de $\mathcal{R}$	82
3.4.3	Choix de la probabilité de fausse alarme $p_{fa}^{\mathcal{R}}$	83
<b>3.5</b>	<b>Conclusion</b>	<b>83</b>

---

---

**Résumé du chapitre précédent :** Le chapitre précédent apporte une réponse au problème de la détection sur le spectrogramme d'un signal inconnu en présence de bruit Gaussien. L'approche retenue dans ce manuscrit est celle de Neyman-Pearson car elle ne fait intervenir aucune connaissance *a priori* sur le signal à détecter. Les trois grandes étapes de la détection sont donc : **1.** les coefficients du spectrogramme sont modélisés par une loi du  $\chi^2$  à deux degrés de liberté, **2.** cette loi est estimée et **3.** celle-ci est utilisée pour seuiller le spectrogramme à probabilité de fausse alarme constante.

Ce chapitre a aussi identifié trois limitations que nous traitons dans ce manuscrit : **1.** Comment discriminer les vraies détections des fausses alarmes? **2.** Comment choisir la probabilité de fausse alarme? **3.** Comment séparer les clics des sifflements?

**Objectif de ce chapitre :** Nous nous intéressons dans ce chapitre à la discrimination entre vraies détections et fausses alarmes. Nous définissons un cadre statistique de détection basé sur des régions du plan temps-fréquence. En particulier nous analysons l'organisation des détections au sein du spectrogramme binaire.

---

## 3.1 Introduction

Le chapitre 2 décrit un détecteur statistique chargé de prendre une décision en chaque point temps fréquence  $[n, k]$  du spectrogramme  $\mathbb{S}_x[n, k]$  : " $\mathcal{H}_0$  : coefficient bruit seul" ou " $\mathcal{H}_1$  : coefficient signal + bruit". Le signal étant de nature inconnue, le détecteur ne se base que sur l'analyse de la densité de probabilité des coefficients du spectrogramme dans le cas bruit seul. Il ne fait aucune hypothèse *a priori* sur le signal à détecter. Le processus de décision est basé sur un critère de Neyman-Pearson, maximisant ainsi la probabilité de détection sous contrainte d'une certaine probabilité de fausse alarme  $p_{fa}$ , choisie par l'utilisateur. Ce processus de décision génère quatre types de sorties :

- Les vrais positifs (ou vraies détections) : la décision  $\hat{\mathcal{H}}_1$  est prise alors que  $\mathcal{H}_1$  est vraie.
- Les vrais négatifs (ou vraies non détections) : la décision  $\hat{\mathcal{H}}_0$  est prise, et  $\mathcal{H}_0$  est vraie.
- Les fausses alarmes (ou faux positifs) : la décision  $\hat{\mathcal{H}}_1$  est prise alors que  $\mathcal{H}_0$  est vraie.

	$\hat{H}_0$	$\hat{H}_1$
$H_0$	Vrai négatif	Fausse alarme
$H_1$	Faux négatif	Vraie détection

FIGURE 3.1 – Décisions possibles d’un test d’hypothèse binaire. L’intersection entre  $H_i$  et  $\hat{H}_j$  correspond à prendre la décision  $\hat{H}_j$  lorsque l’hypothèse vraie est  $H_i$

- Les faux négatifs (ou détections manquées) : la décision  $\hat{H}_0$  est prise alors que  $H_1$  est vraie.

Le tableau 3.1 résume visuellement les quatre décisions que l’on observe en sortie d’un test d’hypothèse binaire. Les deux erreurs en sortie d’un test d’hypothèse binaire étant donc les fausses alarmes et les détections manquées. Ces deux grandeurs sont inversement reliées : pour faire baisser le nombre de détections manquées, il est nécessaire d’augmenter le nombre de fausses alarmes, et inversement.

Rappelons encore une fois que nous nous intéressons tout particulièrement aux signaux de faibles RSB. Nous avons vu dans le chapitre précédent que pour pouvoir les détecter, il est nécessaire de choisir une probabilité de fausse alarme  $p_{fa}$  qui soit élevée. Ce faisant il devient difficile de discriminer les fausses alarmes des vraies détections. En ce sens, nous considérons que les compromis probabilité de fausse alarme/ probabilité de détection accessibles au détecteur quadratique ne sont pas satisfaisants. Nous nous retrouvons alors devant une impasse : le simple critère énergétique sur le spectrogramme ne suffit pas à résoudre ce problème de manière satisfaisante pour les signaux de faible RSB.

Nous décidons arbitrairement de séparer en deux catégories de méthodes permettant de faire émerger un compromis plus satisfaisant : les **changements de représentation** et les approches par **segmentation**.

### Changement de représentation :

Cette catégorie de méthodes part du principe que la raison pour laquelle le signal est difficile à détecter est que l’espace de représentation choisi n’est pas adapté. Ainsi, un signal sinusoïdal noyé corrompu par un bruit blanc est très facilement détectable dans le domaine de Fourier car cette représentation lui est très bien adaptée. Le bruit additif est ainsi dispersé sur l’ensemble de la bande de fréquence alors que l’énergie associée au signal est concentrée sur un faible nombre de coefficients. Le spectrogramme possède de mauvaises propriétés de concentration de l’énergie pour des signaux non-stationnaires [Fla98]. Les représentations

bi-linéaires de la classe de Cohen sont plus adaptées aux signaux non stationnaires mais souffrent de termes d'interférences et d'une faible robustesse au bruit [Coh89]. Des méthodes telles que la S-transform [SIP96], les L-spectrogrammes moyennés [SOS12] ou une combinaison de multitapers et de réallocation des représentations [XF07] semblent être robustes à la fois au bruit à la nature multi-composante des signaux observés. Nous avons passé en revue plusieurs de ces représentations dans le Chapitre 1 et expliqué nos motivations pour le choix du spectrogramme. En particulier nous nous intéressons tout particulièrement à la détection statistique (et non à un problème d'estimation) des signaux. Le spectrogramme possède des statistiques très simples qui sont mises à profit dans cette thèse. Nous ne nous intéresserons donc pas plus avant à ce problème de représentation mais plutôt à la seconde catégorie d'approches.

### **Segmentation :**

Cette catégorie de méthodes part du principe que l'information d'amplitude seule n'est pas suffisamment discriminante pour faire émerger un compromis probabilité de fausse alarme/ probabilité de détection satisfaisant pour de faibles RSB. Ce que nous appelons ici les "approches par segmentation" doivent être comprises comme des approches régionales de la détection. Il s'agit ici de rajouter de l'information "spatiale" dans le processus de décision. L'idée est de penser la détection à l'échelle de la composante temps-fréquence, et de ne pas baser la décision uniquement sur une analyse point par point. Les signaux que nous essayons de détecter ont un support temps-fréquence qui n'est pas discret, celui-ci est au moins partiellement connexe (modulations de fréquence et impulsions large bande). Nous pensons donc que la stratégie de décision ne doit pas ignorer cette proximité des pics temps-fréquences lors du processus de décision.

Il est important de souligner que cette classe de méthode est indépendante de la problématique de changement de représentation. Elle intervient en effet une fois la représentation du signal choisie. Ainsi, quelle que soit la représentation choisie, une partie des signaux possèdera un faible RSB, nous retrouverons donc le même problème que précédemment.

La section suivante développe plus avant les différentes stratégies d'intégration d'information régionale dans le processus de décision.

## **3.2 Segmentation de composantes temps-fréquence**

Nous détaillons ici plusieurs des stratégies mises en place afin d'intégrer de l'information régionale dans le processus de décision : ajout de considération physiques *a priori*, considération du noyau reproduisant, utilisation d'outils de morphologie

mathématique, méthodes à contours actifs, méthodes de croissance de régions et méthodes de tracking temps-fréquence.

### **Considérations physiques *a priori* :**

La segmentation temps-fréquence n'est qu'un outil au service d'une tâche applicative précise. Dans un contexte applicatif, les caractéristiques du signal d'intérêt sont souvent connues. Par exemple l'apparition du signal peut être attendue dans une certaine bande de fréquence ce qui est le cas pour certains mammifères marins [RBP<sup>+</sup>11] et des ondes gravitationnelles [Syl02]. Ou encore une durée minimale du signal peut être connue approximativement [OBN03]. Il existe trois inconvénients majeurs à cette méthode :

1. Elle manque de généralité, les caractéristiques du signal doivent être (partiellement) connus
2. Une sur-segmentation des composantes est souvent observée pour les bas RSB
3. Une baisse de la probabilité de détection est également observée

### **Noyau reproduisant :**

D'abord développé pour l'élimination de détections non significatives au sein du scalogramme [MKH07], puis adaptée à la détection sur spectrogramme [Hui08], cette méthode exploite la redondance intrinsèque de la représentation utilisée. En effet, chaque coefficient du spectrogramme partage une certaine quantité d'énergie avec ses voisins. Les détections sont rassemblées en composantes connexes et leur taille (en nombre de pixels) est comparée avec la taille minimale que doit avoir une composante pour avoir un sens physique. Cette taille minimale est dépendante des paramètres de construction du spectrogramme et du seuil utilisé pour binariser le spectrogramme. Deux inconvénients à cette méthode :

1. Baisse de la probabilité de détection
2. Sur-segmentation des composantes pour les bas RSB

### **Morphologie mathématique :**

Introduits par Jean Serra [Ser86], les opérateurs de morphologie mathématique ont été largement utilisés pour la segmentation d'images [DL03]. Dans le domaine de l'acoustique passive, les opérateurs de fermeture mathématique ont été utilisés pour la réduction du nombre de fausses alarmes issues des spectrogrammes seuillés [UCKP09, TKB<sup>+</sup>12]. Le choix de l'élément structurant est cependant crucial pour le résultat final.



## Contours actifs :

Développé par Kass *et al.* [KWT88], la méthode des contours actifs (ou snakes), consiste à laisser une courbe bidimensionnelle se déformer de manière itérative pour épouser la forme du contour à extraire. L'algorithme converge lorsque des contraintes énergétiques sont minimisées. Ces contraintes énergétiques sont l'énergie interne, liée à la forme de la courbe bidimensionnelle, et l'énergie externe, liée aux caractéristiques de l'image (contraste, bruit, ...). Ces modèles ont déjà été appliqués à des représentations temps-fréquence [TMG08]. Plusieurs inconvénients limitent l'applicabilité de cette méthode : **1)** la formulation et la pondération des critères énergétiques sont cruciales, et doivent être adaptées aux signaux d'intérêt, **2)** l'initialisation de l'algorithme est déterminante, la courbe doit être soit à l'intérieur de la forme à segmenter, soit à l'extérieur, en aucun cas entre les deux, et **3)** le temps de calcul lié à cette méthode est souvent important sur de grandes images.

## Croissance de régions :

L'algorithme de croissance de région initialement proposé par Adams [AB94] a été adapté au cas des images temps-fréquence [MM11, HMC02]. A partir d'un germe initial, l'idée est de chercher dans les voisins proches ceux qui possèdent les mêmes caractéristiques.

## Tracking temps-fréquence

Des méthodes très classiques en traitement du signal visant à créer des "pistes temps-fréquence" peuvent être utilisées. Nous pouvons citer deux grands représentants de ces méthodes que sont le filtrage de Kalman [Kal60] et le filtrage particulaire [AMGC02]. Ces deux méthodes ont déjà été utilisées, sur le spectrogramme binaire, dans le cadre de la détection de sifflements d'odontocètes [MOCT08, RBP<sup>+</sup>11]. Ces méthodes de suivi de pistes temps-fréquence bénéficient de solides fondations théoriques, ce qui est appréciable dans notre contexte. L'inconvénient majeur de ce type de méthode réside dans la nécessité de **formuler un modèle du signal**. Dès lors que l'on s'écarte du modèle prédéfini, il n'est plus possible de garantir la pertinence des résultats obtenus.

## Pourquoi développer une nouvelle méthode ?

Alors que pour la détection de pixels temps-fréquence des méthodes statistiques ont été développées, la détection de motifs temps-fréquence connexes, ou partiellement connexes, n'a pour le moment été réalisée que de manière *ad hoc*, algorithmique ou en imposant des modèles sur le signal d'intérêt.

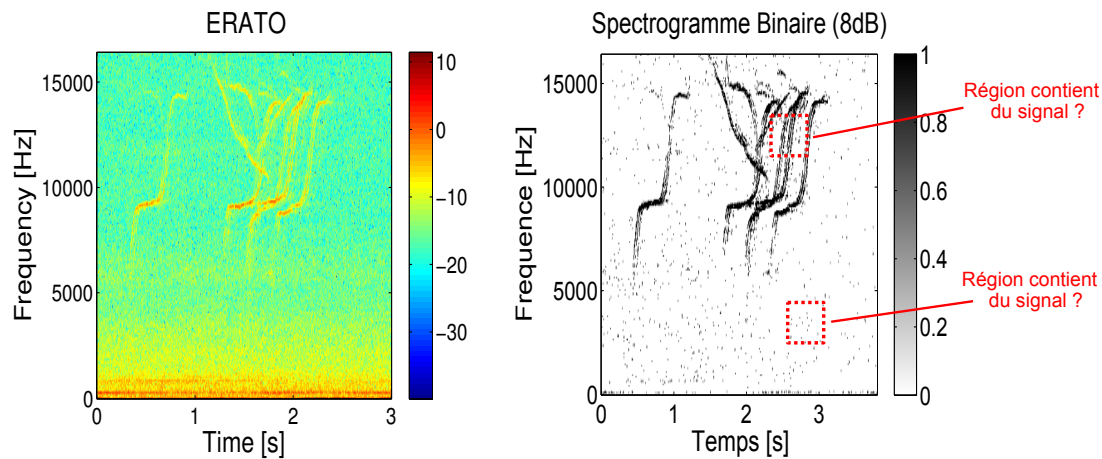


FIGURE 3.2 – Illustration schématique du test d'hypothèse binaire. La question posée concerne la présence de signal dans une région du plan temps-fréquence binaire (Source : ERATO-09, SHOM).

Nous proposons dans ce chapitre une méthode statistique, basée sur un test d'hypothèse binaire, permettant de discriminer les régions contenant du signal des régions ne contenant que du bruit. Ce test d'hypothèse binaire ne pose pas de modèle sur le signal d'intérêt, et est résolu par une approche de Neyman-Pearson.

Ce chapitre reprend les résultats présentés dans [DGI<sup>+</sup>13a].

**Pour résumer :** le chapitre précédent répondait à la question : Comment détecter les pixels temps-fréquence dont l'énergie provient en partie du signal ? La question que nous nous posons dans ce chapitre est la suivante : Comment détecter une région abritant une partie du support temps-fréquence du signal (voir figure 3.2) ?

### 3.3 Formulation du problème

Avant de formuler plus rigoureusement le problème de détection en question, nous expliquons brièvement le principe "philosophique" de la méthode. Nous donnons aussi une très brève description technique de ce qui sera développé plus avant dans ce chapitre.

**Principe "philosophique" :** La résolution du test d'hypothèse BHT1 est nécessairement génératrice de fausses alarmes. Cependant puisque elle se fait à probabilité de fausse alarme constante nous avons une idée grossière de la quantité que nous allons générer. Lorsque l'on seuille le spectrogramme à  $p_{fa} = 10^{-6}$  nous n'attendons qu'un nombre très faible de fausses détections. Au contraire le seuillage du spectrogramme à  $p_{fa} = 0.3$  laisse augurer un grand nombre de fausses alarmes. Ces

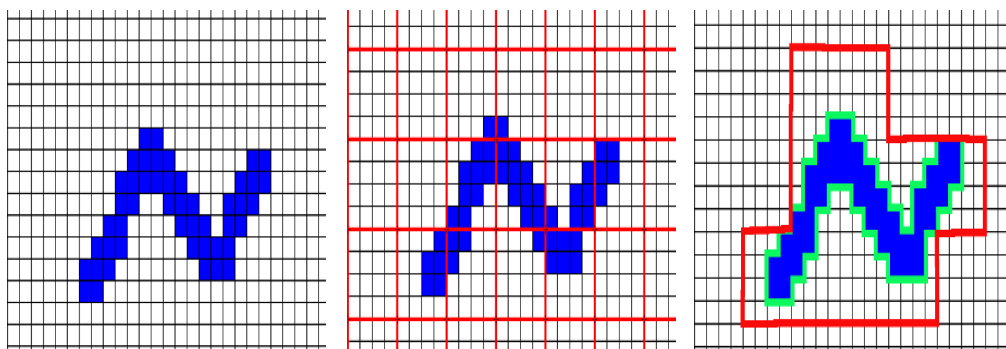


FIGURE 3.3 – Exemple schématique montrant la différence entre le support temps-fréquence  $\mathfrak{S}$  et le support temps-fréquence étendu  $\mathfrak{S}_{\mathcal{R}}$ . À gauche le spectrogramme binaire du signal, au milieu le spectrogramme binaire partitionné (grille rouge) ; à droite le support temps-fréquence  $\mathfrak{S}$  en vert, et le support temps-fréquence étendu  $\mathfrak{S}_{\mathcal{R}}$  en rouge.

fausses détections se répartissent uniformément sur tout le plan temps-fréquence binaire. Par conséquent, si l'on choisit une sous-région d'aire  $A$  du spectrogramme binaire, nous y attendons en moyenne un nombre  $A \times p_{fa}$  de détections. Si le nombre de détections que nous observons effectivement dans cette région est proche de  $A \times p_{fa}$  alors nous pouvons raisonnablement en déduire que la région en question n'abrite que des fausses détections. De même, si le nombre de détections observé est grand devant  $A \times p_{fa}$  la probabilité que ces détections ne provienne que du bruit semble plus faible : la région abrite très probablement une partie du support temps-fréquence du signal d'intérêt.

**Principe technique :** Nous proposons donc une méthode statistique de détection des régions abritant du signal basée sur la résolution séquentielle de deux tests d'hypothèse :

1. Le premier est basé sur résolution par une approche de Neyman-Pearson d'un test d'hypothèse binaire. Nous l'appelons BHT1, la grandeur sur laquelle le test est résolu est l'énergie des pixels temps-fréquence (Cf. Chapitre 2).
2. Le second est la résolution par une approche de Neyman-Pearson d'un second test d'hypothèse binaire, que nous appelons BHT2. Le spectrogramme binaire est découpé en sous-régions, les détections sont comptées dans chacune de ces sous-régions, et leur nombre est comparé à une fonction densité de probabilité décrivant le nombre de fausses alarmes attendu. Le nombre de fausses alarmes est modélisé par une loi binomiale.

### 3.3.1 Définition du support temps-fréquence étendu $\beth_{\mathcal{R}}$

Pour rappel, le support temps-fréquence  $\beth$  (prononcer Beth, ou Bet<sup>1</sup>) du signal  $s(t)$  est défini comme l'ensemble des pixels temps-fréquence dont l'énergie provient en partie du signal :

$$\beth(s) = \{[n, k] / \mathbb{S}_x[n, k] = \mathbb{S}_s[n, k] + \mathbb{S}_b[n, k]\} \quad (3.1)$$

avec  $\mathbb{S}_s[n, k]$  l'énergie provenant du signal, et  $\mathbb{S}_b[n, k] = \gamma_b[n, k]$  la densité spectrale de puissance du bruit. Le chapitre 1 propose une estimation  $\hat{\beth}(p_{fa})$  de ce support défini comme l'ensemble des points temps-fréquence dépassant un seuil  $S^{p_{fa}}$  :

$$\hat{\beth}(p_{fa}) = \{[n, k] / \mathbb{S}_x[n, k] \geq S^{p_{fa}}[n, k]\} \quad (3.2)$$

Cet estimateur est donc dépendant du choix de la probabilité de fausse alarme  $p_{fa}$ . C'est un estimateur non-biaisé du support temps-fréquence, c'est à dire que pour  $RSB \rightarrow \infty$  nous obtenons exactement le support temps-fréquence désiré pour  $p_{fa} \rightarrow 0$ . A noter que ce support dépend lui aussi de la représentation temps-fréquence choisie ainsi que de ses paramètres de construction (voir chapitre 2).

Nous partitionnons maintenant le spectrogramme binaire en une grille régulière sans recouvrement (voir figure 3.4). Nous définissons  $\mathcal{R}[n, k]$  la région temps-fréquence rectangulaire d'aire  $A$  centrée en  $[n, k]$ . Le nouveau test d'hypothèse binaire BHT2 que nous introduisons ici est réalisé sur ces régions  $\mathcal{R}[n, k]$ .

Nous définissons aussi  $\beth_{\mathcal{R}}$ , que nous appelons support temps-fréquence étendu du signal (voir figure 3.3). Il est défini comme l'ensemble des régions temps-fréquence  $\mathcal{R}[n, k]$  centrées en  $[n, k]$  qui intersectent  $\beth$  :

$$\beth_{\mathcal{R}} = \{\mathcal{R}[n, k] / \mathcal{R}[n, k] \cap \beth \neq \{\emptyset\}\} \quad (3.3)$$

Le but de ce chapitre est d'estimer  $\beth_{\mathcal{R}}$ . Il convient de noter la propriété suivante :

$$\beth \subset \beth_{\mathcal{R}} \quad (3.4)$$

Le support étendu  $\beth_{\mathcal{R}}$  est donc un estimateur biaisé du support  $\beth$ . La méthode présentée aura donc tendance à épaissir les bordures du support  $\beth$  du signal d'intérêt. Cet épaississement des bordures est le prix à payer pour obtenir une plus grande robustesse au bruit. La figure 3.3 montre une représentation visuelle des deux supports  $\beth_{\mathcal{R}}$  et  $\beth$  que nous cherchons à déterminer. Nous pouvons aussi remarquer que dans le cas où  $\mathcal{R}[n, k]$  est une région composée d'un seul pixel, nous

1. Ici écrit dans l'alphabet hébreu,  $\beth$  est une lettre commune aux alphabets phénicien, araméen, syriaque, hébreu et arabe et signifie "maison". Symbole adéquat pour la recherche du support temps-fréquence d'un signal. Source : Wikipedia

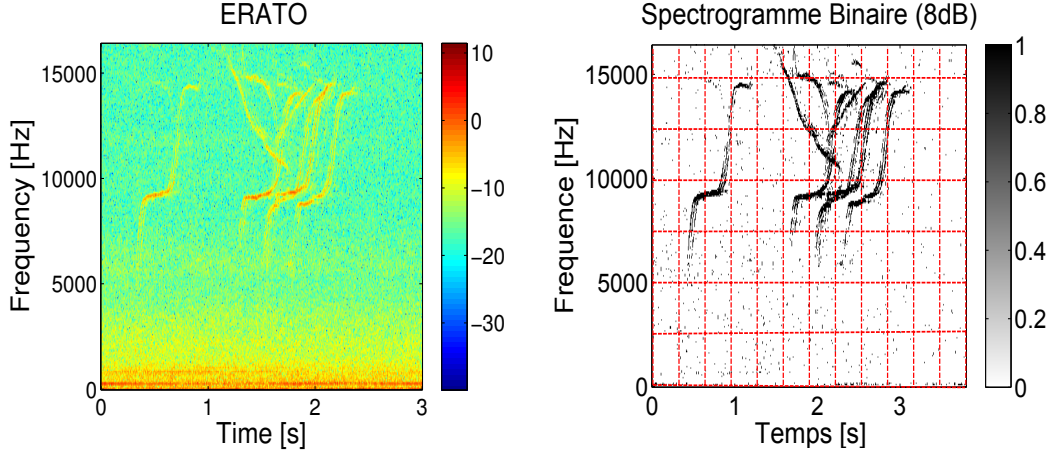


FIGURE 3.4 – Exemple schématique du partitionnement d'un spectrogramme binaire généré à partir du spectrogramme original. Le partitionnement est réalisé avec une grille fixe, sans recouvrement. (Source : ERATO-09, SHOM)

obtenons logiquement  $\sqsupset = \sqsupset_{\mathcal{R}}$ .

### 3.3.2 Formulation du test d'hypothèse binaire BHT2

Ce test d'hypothèse binaire, appelé BHT2 est relatif à la question suivante : la région  $\mathcal{R}[n, k]$  est elle une "région  $\mathcal{H}_1^{\mathcal{R}}$  : abrite une partie du support temps-fréquence du signal" ou une "région  $\mathcal{H}_0^{\mathcal{R}}$  : n'abrite que du bruit" ? Nous définissons alors BHT2 de la manière suivante pour toutes les régions  $\mathcal{R}[n, k]$  du spectrogramme binaire :

$$BHT2 : \begin{cases} \mathcal{H}_0^{\mathcal{R}} : \lambda = 0 \\ \mathcal{H}_1^{\mathcal{R}} : \lambda > 0, \end{cases} \quad (3.5)$$

Avec  $\lambda$  la proportion de signal présente dans la région temps-fréquence  $\mathcal{R}[n, k]$ . Notons que  $\lambda$  est une variable discrète définie entre 0 et 1 :  $\lambda \in \{0, 1/A, 2/A, \dots, 1\}$ . Afin de résoudre BHT2 nous introduisons une variable aléatoire  $\Psi = \{\psi[n, k]\}$ , comptant le nombre de détections au sein des  $\mathcal{R}[n, k]$  :

$$\psi[n, k] = \sum_{[t_0, f_0] \in \mathcal{R}[n, k]} \phi_{p_{f_a}}[t_0, f_0] \quad (3.6)$$

La carte  $\psi$  ainsi générée est appelée **carte de comptage**. Les valeurs qu'elle peut prendre sont entières et appartiennent à  $\{0, 1, \dots, A\}$ . Le test d'hypothèse BHT2 est résolu grâce à la variable  $\Psi$  :

$$BHT2 : \begin{cases} \mathcal{H}_0^{\mathcal{R}} : \psi[n, k] \leq \psi_T \\ \mathcal{H}_1^{\mathcal{R}} : \psi[n, k] > \psi_T, \end{cases} \quad (3.7)$$

Avec  $\psi_T$  un seuil sur le nombre de détections observées au sein des régions. Nous souhaitons résoudre ce test par une approche de Neyman-Pearson. Quelle est la fonction densité de probabilité de  $\Psi$  sous  $\mathcal{H}_0^R$  ? C'est l'objet de la section suivante.

### 3.3.3 Densité de probabilité de la carte de comptage $\Psi$ sous $\mathcal{H}_0^R$

Le test BHT1 est résolu par une approche de Neyman-Pearson, cela signifie que le seuil adaptatif est déterminé en chaque point temps-fréquence de manière à maximiser la probabilité de détection sous contrainte d'une certaine probabilité de fausse alarme  $p_{fa}$ . Cette probabilité de fausse alarme est un degré de liberté de la méthode. La variable aléatoire  $\phi_{p_{fa}}[n, k]$  est binaire et la probabilité de prendre chacune des deux valeurs sous  $\mathcal{H}_0$  est connue :

$$\begin{cases} p_{\mathcal{H}_0}(\phi_{p_{fa}} = 0) = 1 - p_{fa} \\ p_{\mathcal{H}_0}(\phi_{p_{fa}} = 1) = p_{fa} \end{cases} \quad (3.8)$$

Sous l'hypothèse  $\mathcal{H}_0$ , la variable aléatoire  $\phi_{p_{fa}}[n, k]$  suit donc une loi de Bernoulli de paramètre  $p_{fa}$  :

$$\phi_{p_{fa}}^{\mathcal{H}_0} \sim \text{Bernoulli}(p_{fa}) \quad (3.9)$$

Sous l'hypothèse  $\mathcal{H}_0^R$ , la région  $\mathcal{R}[n, k]$  n'abrite que des fausses alarmes,  $\psi[n, k]$  est donc la somme de variables aléatoires de Bernoulli de paramètres  $p_{fa}$ . Sous hypothèse d'indépendance des variables de Bernoulli, la variable  $\Psi$  suit alors une loi binomiale de paramètres  $\mathcal{A}$  et  $p_{fa}$  sous  $\mathcal{H}_0^R$  :

$$\Psi \sim \text{Binom}(\mathcal{A}, p_{fa}) \quad (3.10)$$

Cette loi de distribution binomiale est définie comme :

$$p_{\mathcal{H}_0^R}(\psi[n, k] = k) = \binom{\mathcal{A}}{k} p_{fa}^k (1 - p_{fa})^{\mathcal{A}-k} \quad (3.11)$$

avec

$$\binom{\mathcal{A}}{k} = \frac{\mathcal{A}!}{k!(\mathcal{A} - k)!} \quad (3.12)$$

le nombre de combinaisons sans répétitions de  $k$  éléments parmi  $\mathcal{A}$  éléments. La fonction de distribution cumulative associée à  $p_{\mathcal{H}_0^R}(\psi)$  est la suivante :

$$p_{\mathcal{H}_0^R}(\psi[n, k] \leq k) = I_{1-p_{fa}}(\mathcal{A} - k, 1 + k) \quad (3.13)$$

avec  $I_x(a, b)$  la fonction beta incomplète régularisée [AS12] définie comme :

$$I_x(a, b) = \frac{B(x; a, b)}{B(a, b)} \quad (3.14)$$

et

$$B(x; a, b) = \int_0^x t^{a-1}(1-t)^{b-1} dt, \quad (3.15)$$

la fonction Beta incomplète, et  $B(a, b) = B(1; a, b)$  la fonction Beta. La probabilité de fausse alarme est alors définie comme :

$$p_{fa}^{\mathcal{R}} = p_{\mathcal{H}_0^{\mathcal{R}}}(\psi[n, k] > \psi_T) \quad (3.16)$$

avec  $\psi_T$  le seuil induit par la probabilité de fausse alarme  $p_{fa}^{\mathcal{R}}$ . Nous pouvons alors résoudre BHT2 à probabilité de fausse alarme constante  $p_{fa}^{\mathcal{R}}$ , et ce quelle que soit la  $p_{fa}$  utilisée pour binariser le spectrogramme :

$$p_{fa}^{\mathcal{R}} = 1 - I_{1-p_{fa}}(\mathcal{A} - \psi_T, 1 + \psi_T) \quad (3.17)$$

Il convient de noter que la variable  $p_{fa}^{\mathcal{R}}$  est discrète et ne peut prendre que  $A+1$  valeurs puisque  $\psi_T \in \{0, 1, \dots, A\}$ . Celles-ci sont dépendante du choix de  $p_{fa}$ .

**Remarque 1 :** Sous l'hypothèse  $\mathcal{H}_1$  la variable  $\phi_{p_{fa}}[n, k]$  suit aussi une loi de Bernoulli. Le paramètre de cette loi est la probabilité de détection  $p_d$  qui est fonction de la  $p_{fa}$  mais aussi du RSB - inconnu - du signal. Afin d'établir une méthodologie applicable à la plus grande gamme de signaux possibles (et en particulier aux signaux dont les RSB varient rapidement) nous choisissons de ne caractériser que les statistiques du bruit. Nous ne nous intéresserons donc pas aux réalisations des variables liées à l'hypothèse  $\mathcal{H}_1$  dans la suite de ce manuscrit.

**Remarque 2 :** L'hypothèse d'indépendance des variables de Bernoulli est une approximation forte. Considérer les coefficients comme indépendants revient à considérer qu'il est possible de générer un pavage non redondant du plan temps-fréquence, ce qui est en contradiction avec le théorème de Balian-Low [Bal81]. Bien que grossière, cette hypothèse nous permet de formuler analytiquement le mécanisme d'estimation de  $\mathfrak{I}_{\mathcal{R}}$ . Nous discuterons la (non-)pertinence de ce choix dans le chapitre suivant. L'hypothèse sera alors relâchée afin de prendre en compte la redondance intrinsèque du spectrogramme.

### 3.3.4 Résumé de la méthode

À ce stade nous sommes capables de discriminer les régions abritant du signal de celles n'abritant que du bruit au sein d'un spectrogramme binaire à probabilité de fausse alarme constante. Cette discrimination est adaptée au bruit et à la manière

dont le spectrogramme est seuillé. Pour plus de clarté, nous résumons ici les six étapes de la méthode présentée :

1. Le spectrogramme est calculé (à partir de la forme d'onde temporelle du signal bruité)
2. La densité spectrale de puissance du bruit  $\gamma_b[n, k]$  est estimée en chaque point temps-fréquence
3. La loi du  $\chi^2$  décrivant les coefficients du bruit dans le spectrogramme est ainsi estimée en chaque point temps-fréquence, et le spectrogramme seuillé à probabilité de fausse alarme constante  $p_{fa}$
4. Le spectrogramme binaire est partitionné en une grille régulière de taille  $\mathcal{A}$  et les détections comptées,
5. Le modèle binomial  $Binom(\mathcal{A}, p_{fa})$  correspondant est défini
6. Une décision est prise sur la présence ou non de signal dans les régions temps-fréquence

Une représentation schématique des différentes étapes de la méthode proposée est donnée figure 3.5.

### 3.4 Degrés de liberté de la méthode

La méthode proposée possède plusieurs degrés de liberté que nous résumons dans la table 3.1 : ceux liés à la représentation du signal et à sa binarisation (i.e. les paramètres de construction du spectrogramme, d'estimation du bruit et la probabilité de fausse alarme choisie  $p_{fa}$ ) ainsi que les deux degrés de liberté supplémentaires introduits dans ce chapitre : la taille des régions  $\mathcal{R}$  sur lesquelles les détections sont comptées et la seconde probabilité de fausse alarme  $p_{fa}^{\mathcal{R}}$ .

Les paramètres de construction du spectrogramme et le voisinage d'estimation de la DSP du bruit sont choisis en fonction de considérations physiques liées à la fois au signal d'intérêt et à l'environnement au sein duquel il est enregistré. Le choix de ces paramètres est discuté brièvement dans le Chapitre 2 et pour plus de détail se référer à l'excellent travail de thèse de J. Huillery [Hui08].

Nous illustrons ci-après l'influence de trois paramètres liés à l'analyse de la densité des fausses alarmes : la probabilité de fausse alarme  $p_{fa}$ , la forme de la région  $\mathcal{R}$  et la seconde probabilité de fausse alarme  $p_{fa}^{\mathcal{R}}$ .

#### 3.4.1 Probabilité de fausse alarme $p_{fa}$

La probabilité de fausse alarme  $p_{fa}$  est un paramètre déterminant dans le processus de détection. Si la probabilité de fausse alarme est fixée trop basse, le signal n'est pas détecté et quelle que soit la probabilité  $p_{fa}^{\mathcal{R}}$  choisie, le signal ne sera pas



## Détection des régions d'intérêt dans le plan temps-fréquence

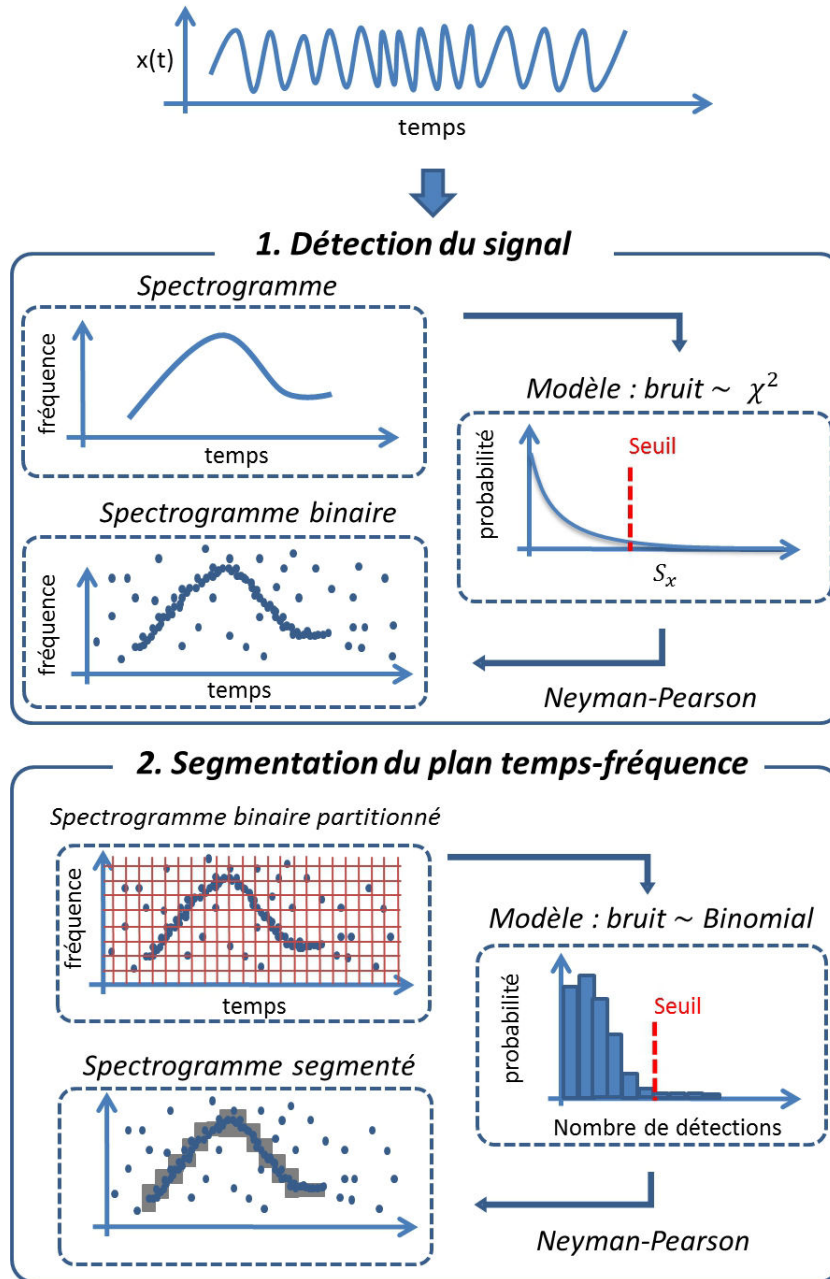


FIGURE 3.5 – Schéma bloc montrant les différentes étapes du processus de segmentation du spectrogramme. **1.** Le bruit est estimé localement, puis le spectrogramme seuillé à  $p_{fa}$  constante. **2.** Le spectrogramme binaire est partitionné, la carte de comptage modélisée par une loi binomiale, et le seuil sur le nombre de détections correspondant à une probabilité de fausse alarme  $p_{fa}^{\mathcal{R}}$  est calculé. Nous obtenons alors la carte segmentée.

Étapes du processus	Degrés de liberté
Représentation du signal	Longueur de la fenêtre d'apodisation Forme de la fenêtre d'apodisation Recouvrement Zéro-padding
Estimation du support $\sqsupset$	Voisinage d'estimation de $\gamma_b[n, k]$ Choix du percentile pour l'estimation de $\gamma_b[n, k]$ Probabilité de fausse alarme $p_{fa}$
Estimation du support étendu $\sqsupset_{\mathcal{R}}$	Partitionnement du spectrogramme binaire $\mathcal{R}$ Probabilité de fausse alarme $p_{fa}^{\mathcal{R}}$

TABLE 3.1 – Récapitulatif des degrés de liberté de la méthode à chaque étape du processus de détection.

déteçté. De même si  $p_{fa} \approx 1$ , le signal d'intérêt est noyé dans les fausses déteçtions. Il est alors impossible de séparer les deux distributions  $p_{\mathcal{H}_0^{\mathcal{R}}}(\psi)$  et  $p_{\mathcal{H}_1^{\mathcal{R}}}(\psi)$ . De manière générale,  $p_{fa}$  va fixer la gamme des points de fonctionnement probabilité de fausse alarme/ probabilité de déteçtion accessibles à BHT2.

La figure 3.6 illustre l'importance du choix de  $p_{fa}$ . Un spectrogramme est généré et binarisé en utilisant trois probabilités de fausse alarme différentes :  $p_{fa} = 0.35$ ,  $0.07$  et  $10^{-4}$  (respectivement équivalentes à des RSB de focus de -10dB, 8dB et 15dB). Le spectrogramme est ensuite partitionné en régions de taille  $A = 8 \times 8$  pixels temps-fréquence sur lesquelles les déteçtions sont comptées, nous obtenons alors les cartes de comptage  $\Psi$ . Nous formulons trois remarques

**Remarque 1 :** Les valeurs accessibles à  $\psi[n, k]$  dépendent de la taille de la région de comptage. Ces valeurs seront comprises entre 0 et A. Les valeurs prises par  $\psi[n, k]$  dépendent du choix de  $p_{fa}$ . Une  $p_{fa}$  très faible induira, dans le cas bruit seul, des valeurs de  $\psi[n, k]$  proches de  $A \times p_{fa} \approx 0$  et une variance  $A \times p_{fa} \times (1 - p_{fa}) \approx 0$  faible (voir figure 3.6d). Au contraire, une probabilité de fausse alarme élevée déplacera  $\psi[n, k]$  vers les hautes valeurs, tout en augmentant aussi la variance (voir figure 3.6b).

**Remarque 2 :** La probabilité de déteçtion du signal, et par conséquent la densité de probabilité de  $\Psi$  sous l'hypothèse signal, dépend de  $p_{fa}$ . Le seuillage à  $p_{fa} = 10^{-4}$  (équivalent à un RSB de focus de 15dB) ne permet pas de déteçter le signal dans sa totalité (voir figure 3.6d), il paraît donc très difficile de "corriger" ces

détections manquées grâce à BHT2. À l’opposé, le seuillage à  $p_{fa} = 0.35$  (équivalent à un RSB de focus de  $-10\text{dB}$ ) détecte la quasi-totalité du signal mais noie alors les détections liées au signal dans les fausses alarmes, baissant drastiquement le contraste de la carte de comptage. Il est intéressant de noter que malgré un nombre très important de fausses alarmes, l’œil humain est encore approximativement capable de discerner des régions d’intérêt, à la fois sur le spectrogramme binaire, et sur la carte de comptage.

**Remarque 3 :** La capacité d’un détecteur statistique à différencier deux hypothèses dépend du contraste entre deux distributions. Idéalement nous souhaiterions avoir que les deux distributions de probabilités de  $\Psi$ , i.e. dans les cas  $\mathcal{H}_0^{\mathcal{R}}$  et  $\mathcal{H}_1^{\mathcal{R}}$ , soient complètement séparées. Le choix de  $p_{fa}$  est donc crucial. Les cas  $p_{fa} = 0$  et  $p_{fa} = 1$  représentent les deux cas extrêmes pour lesquels il ne sera plus possible de séparer les deux distributions. Le choix de la  $p_{fa}$  apparaît donc un point crucial et est dépendant du RSB du signal. Pour un RSB très élevé, il n’est pas nécessaire de choisir une grande  $p_{fa}$ . Pour un signal de faible RSB la situation est bien moins évidente... Il est sans doute nécessaire d’augmenter  $p_{fa}$ , mais jusqu’à quel point ? Comment choisir ce seuil ? Ce problème sera discuté plus en détail dans le chapitre 5.

### 3.4.2 Partitionnement du spectrogramme binaire : choix de $\mathcal{R}$

La manière de partitionner le spectrogramme binaire, et en particulier le choix de la taille des régions  $\mathcal{R}$  est aussi un choix pour le processus de décision global.

Nous illustrons l’influence de la taille des régions en générant des cartes de comptage à partir de deux spectrogrammes binaires. Le premier est généré avec une  $p_{fa} = 0.07$  (équivalent à  $8\text{dB}$ ) (figure 3.7), le second correspond à  $p_{fa} = 0.21$  (équivalent à  $3\text{dB}$ ) (figure 3.8). Pour chacun des ces deux spectrogrammes binaires nous générons les cartes de comptage correspondante pour trois régions différentes :  $3 \times 3$ ,  $8 \times 8$  et  $15 \times 15$ . La gamme de valeurs accessibles à  $\psi[n, k]$  est donc différente dans les trois cas. De 0 à 9 pour le découpage en  $3 \times 3$ , de 0 à 64 pour le découpage en  $8 \times 8$ , de 0 à 225 pour le découpage en  $15 \times 15$ .

Le choix du découpage impacte le contraste au sein de la carte de comptage entre les distributions  $p_{\mathcal{H}_0^{\mathcal{R}}}(\psi)$  et  $p_{\mathcal{H}_1^{\mathcal{R}}}(\psi)$ . Ce contraste va être différent en fonction de la valeur de la  $p_{fa}$  décidée à BHT1 et de la taille des régions de comptage considérées. Ainsi, les grandes régions auront tendance à "noyer" le signal au milieu des fausses détections (voir les fenêtres de taille  $15 \times 15$ ). À l’opposé les petites régions ne tirent pas suffisamment de la proximité des pics temps-fréquence, le contraste entre régions "bruit seul" et régions "signal" n’est alors pas très grand.

Le choix du partitionnement, apparaît donc lui aussi comme un point crucial pour deux raisons : **1.** il définit la forme de  $\mathcal{R}$ , et **2.** il influence le contraste de la carte de comptage et donc la capacité à discriminer les régions "bruit seul" des régions "signal". Le choix de  $\mathcal{R}$  n'est pas résolu ici, ce problème sera discuté en détail dans le chapitre 5.

### 3.4.3 Choix de la probabilité de fausse alarme $p_{fa}^{\mathcal{R}}$

La probabilité de fausse alarme  $p_{fa}^{\mathcal{R}}$  est un degré de liberté laissé à l'utilisateur qui détermine le point de fonctionnement sur la courbe COR. La figure 3.9 montre un exemple de segmentation du spectrogramme pour trois différentes probabilités de fausse alarme  $p_{fa}^{\mathcal{R}}$ . Le spectrogramme a été binarisé à 5dB, ce qui est un seuil très bas. Malgré cela nous sommes capables d'estimer le support temps-fréquence étendu  $\mathcal{R}$  de manière satisfaisante.

## 3.5 Conclusion

Nous avons formulé dans ce chapitre un test d'hypothèse binaire BHT2. Ce test cherche à déterminer la présence éventuelle de signal dans une région du spectrogramme. Ce test s'applique sur une variable  $\Psi$ , appelée variable de comptage. Cette variable compte le nombre de détections dans une région temps-fréquence  $\mathcal{R}$  du spectrogramme binaire centrée en  $[n, k]$ . Le nombre de détections observé est comparé avec le nombre de détections attendu dans l'hypothèse où la région abrite du bruit seul. Nous avons approximé la loi de probabilité donnant le nombre de fausses alarmes attendu dans une région bruit seul par une loi binomiale. Nous avons finalement montré que le contraste statistique entre les régions signal  $\mathcal{H}_1^{\mathcal{R}}$  et les régions bruit seul  $\mathcal{H}_0^{\mathcal{R}}$  est suffisamment discriminant pour établir la présence ou non de signal dans une région temps-fréquence même lorsque le spectrogramme a été binarisé avec un seuil inhabituellement bas (seuillé à 5dB dans l'exemple donné en figure 3.9). La méthode proposée possède donc des potentialités en terme de robustesse de détection et de "critique" des détections générées par un premier test d'hypothèse binaire qui méritent d'être développées.

À ce stade, la méthode développée possède trois paramètres cruciaux qui conditionnent les résultats de détection.

Le choix de la probabilité de fausse alarme  $p_{fa}$  d'abord. Cette probabilité est utilisée pour binariser le spectrogramme, et nous n'avons aucun critère de choix pour le moment. Ce choix fixe la représentation (le spectrogramme binaire) sur laquelle BHT2 va être appliqué. Ce choix doit être adapté au RSB qui est inconnu.

Le choix de la forme des régions de comptage est lui aussi  $\mathcal{R}$  crucial. Alors qu'une région de petite taille réduit le contraste statistique entre les régions signal

et les régions bruit, des régions trop grandes "noient" les détections provenant du signal dans un grand nombre de fausses alarmes, réduisant là aussi le contraste. Il est donc nécessaire de trouver un critère de sélection adapté au signal d'intérêt, ce choix sera discuté au chapitre 5.

Nous pointons deux faiblesses sur ce chapitre. La première concerne l'approximation d'indépendance des variables de Bernoulli que nous avons fait au sein de ce chapitre, celle-ci mérite d'être critiquée. La redondance intrinsèque du spectrogramme, due aux paramètres de construction du spectrogramme, corréle temporellement et fréquentiellement les pixels temps-fréquence. Quel est l'impact de cette approximation en terme d'écart entre la loi effectivement observée et le modèle binomial ? La seconde faiblesse de ce chapitre concerne le positionnement de la grille de comptage des détections. La taille des régions est cruciale mais leur positionnement l'est aussi. Ces deux point font l'objet du chapitre suivant.

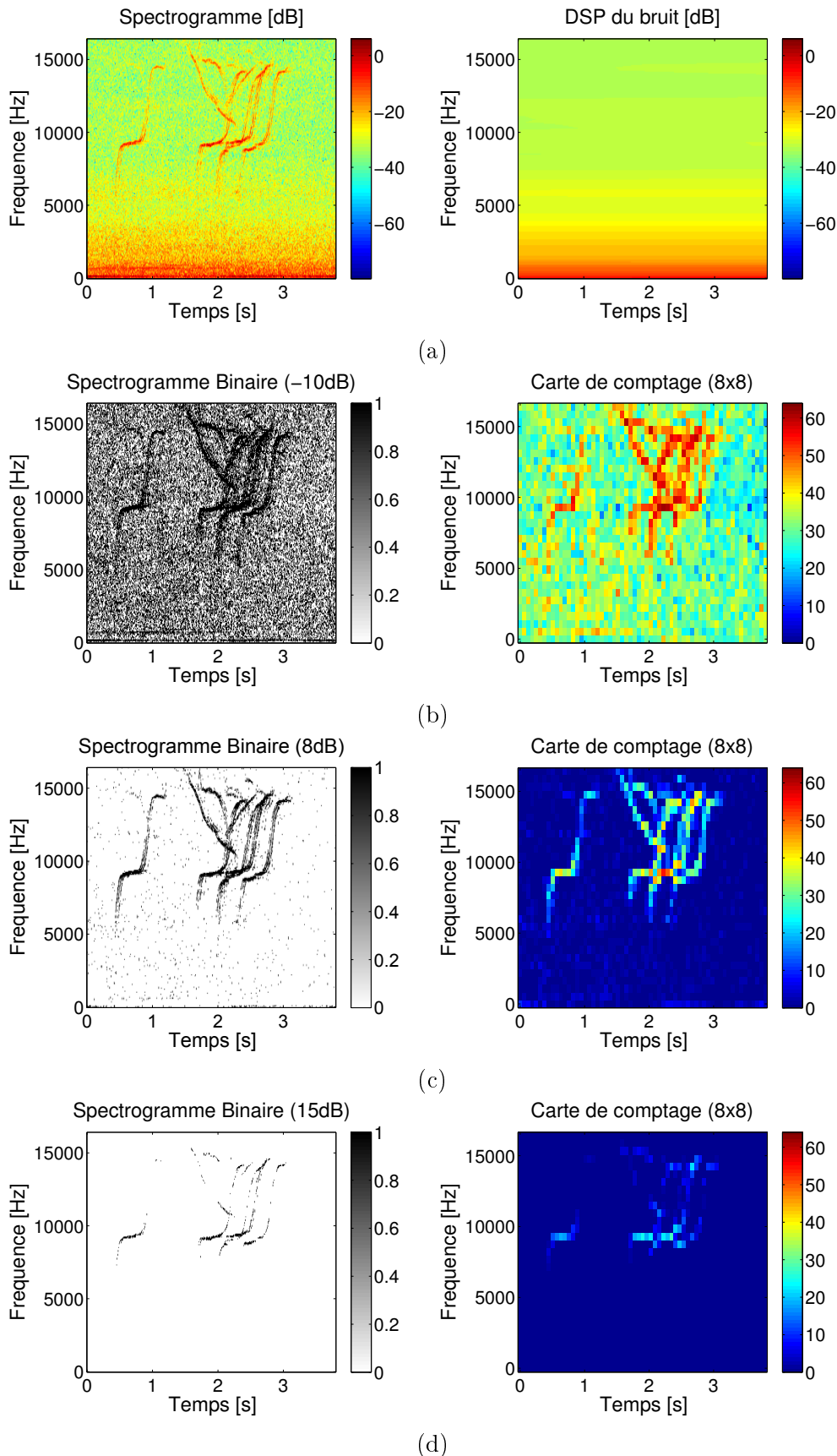


FIGURE 3.6 – **Illustration de l'influence de la probabilité de fausse alarme.** a) Le spectrogramme calculé et la DSP du bruit estimée. Trois spectrogrammes binaires sont obtenus par seuillage : b)  $p_{fa} = 0.35$  (équivalent -10dB), c)  $p_{fa} = 0.07$  (équivalent 8dB), c)  $p_{fa} = 10^{-4}$  (équivalent 15dB), et d)  $p_{fa} = 10^{-4}$  (équivalent 15dB). Les spectrogrammes binaires sont partitionnés et les détections comptées de manière identique : régions de  $8 \times 8$  pixels temps-fréquence.

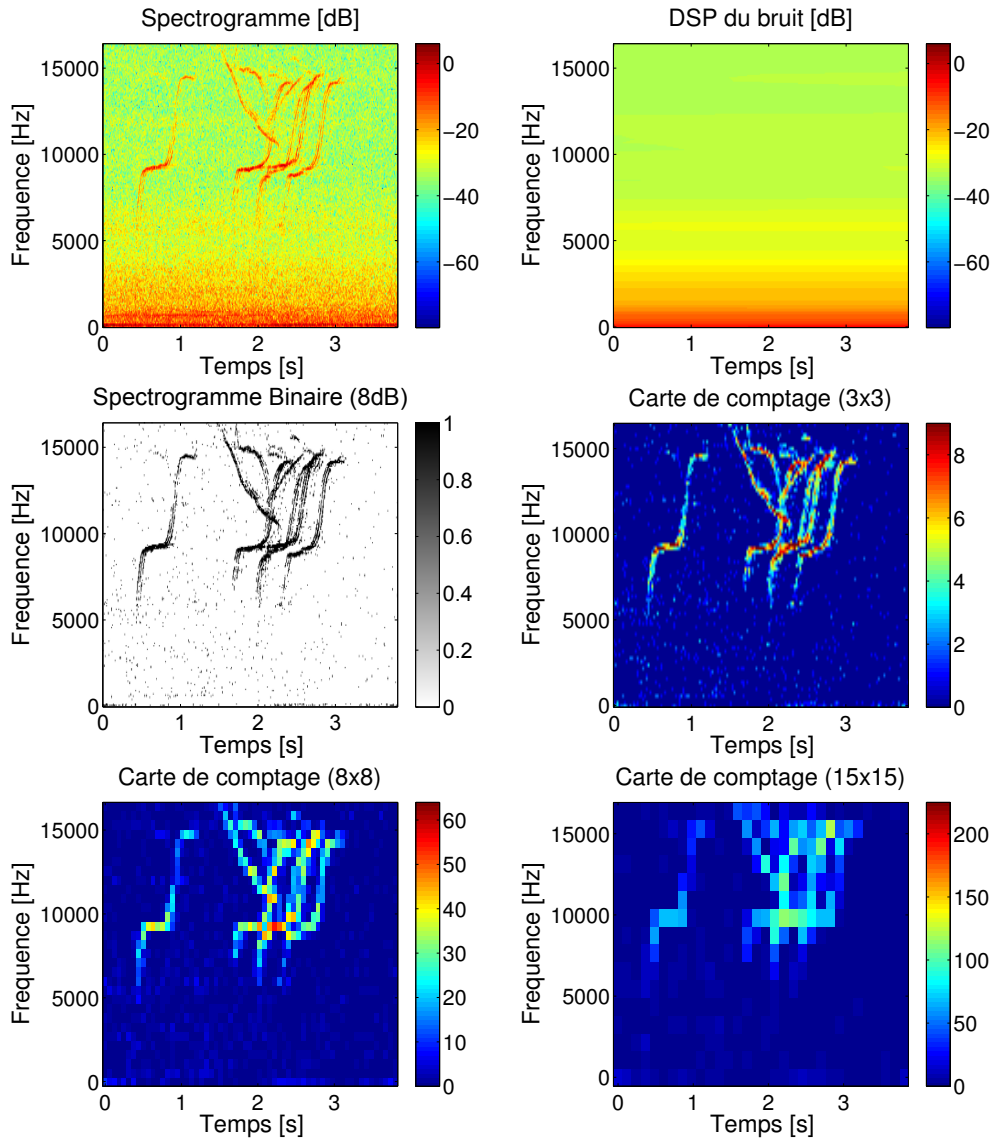


FIGURE 3.7 – **Illustration de l’influence de la taille de la région de comptage.** Spectrogramme du signal calculé avec 50% de recouvrement et une fenêtre de Hann de 512 points, sans zero-padding et sa DSP estimée. Le spectrogramme est généré avec une probabilité de fausse alarme  $p_{fa} = 0.07$  (équivalent 8dB). Trois cartes de comptage sont générées à partir du spectrogramme seuillé binaire en utilisant des régions de taille  $3 \times 3$ ,  $8 \times 8$  et  $15 \times 15$  pixels temps-fréquence.

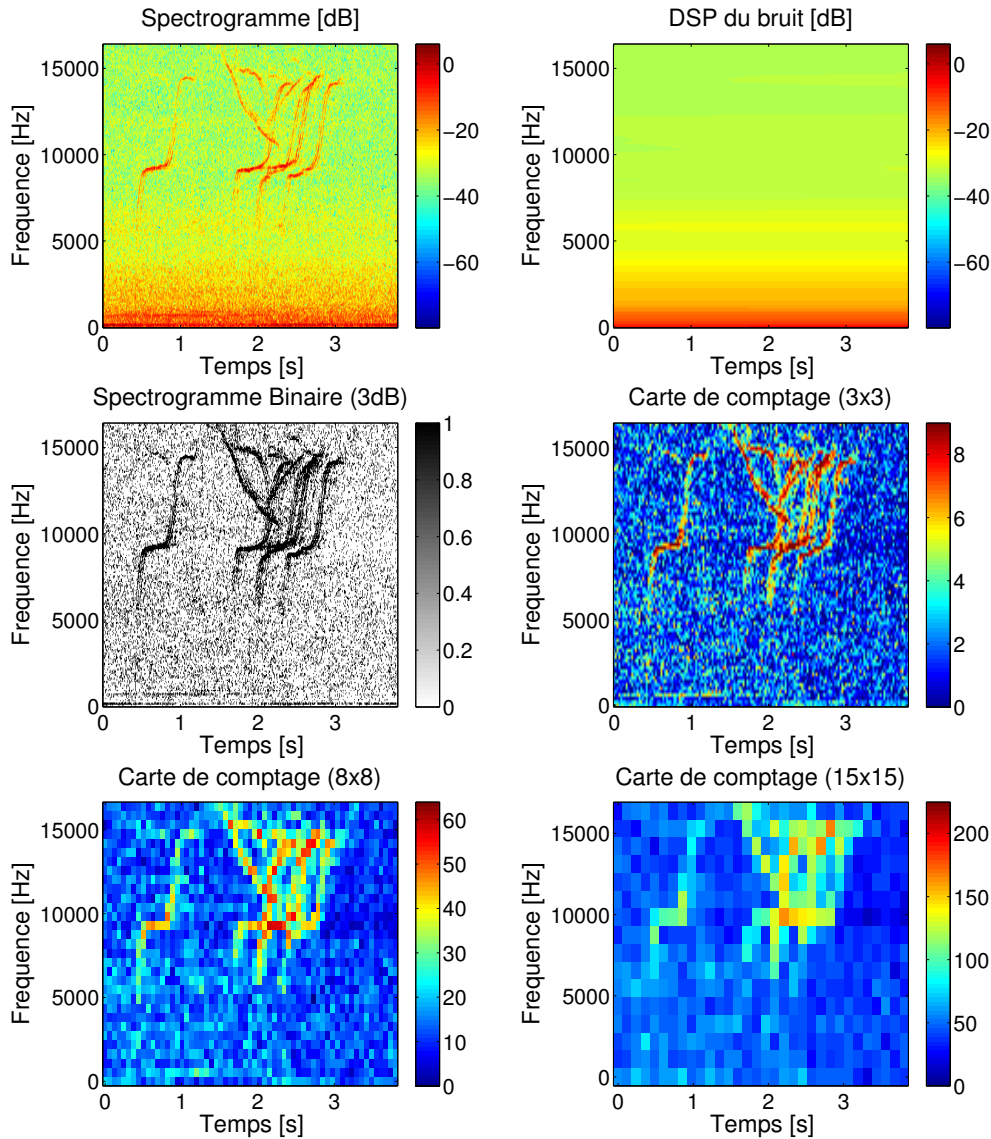


FIGURE 3.8 – **Illustration de l'influence de la taille de la région de comptage.** Spectrogramme du signal calculé avec 50% de recouvrement et une fenêtre de Hann de 512 points, sans zero-padding et sa DSP estimée. Le spectrogramme est généré avec une probabilité de fausse alarme  $p_{fa} = 0.07$  (équivalent 8dB). Trois cartes de comptage sont générées à partir du spectrogramme seuillé binaire en utilisant des régions de taille  $3 \times 3$ ,  $8 \times 8$  et  $15 \times 15$  pixels temps-fréquence.



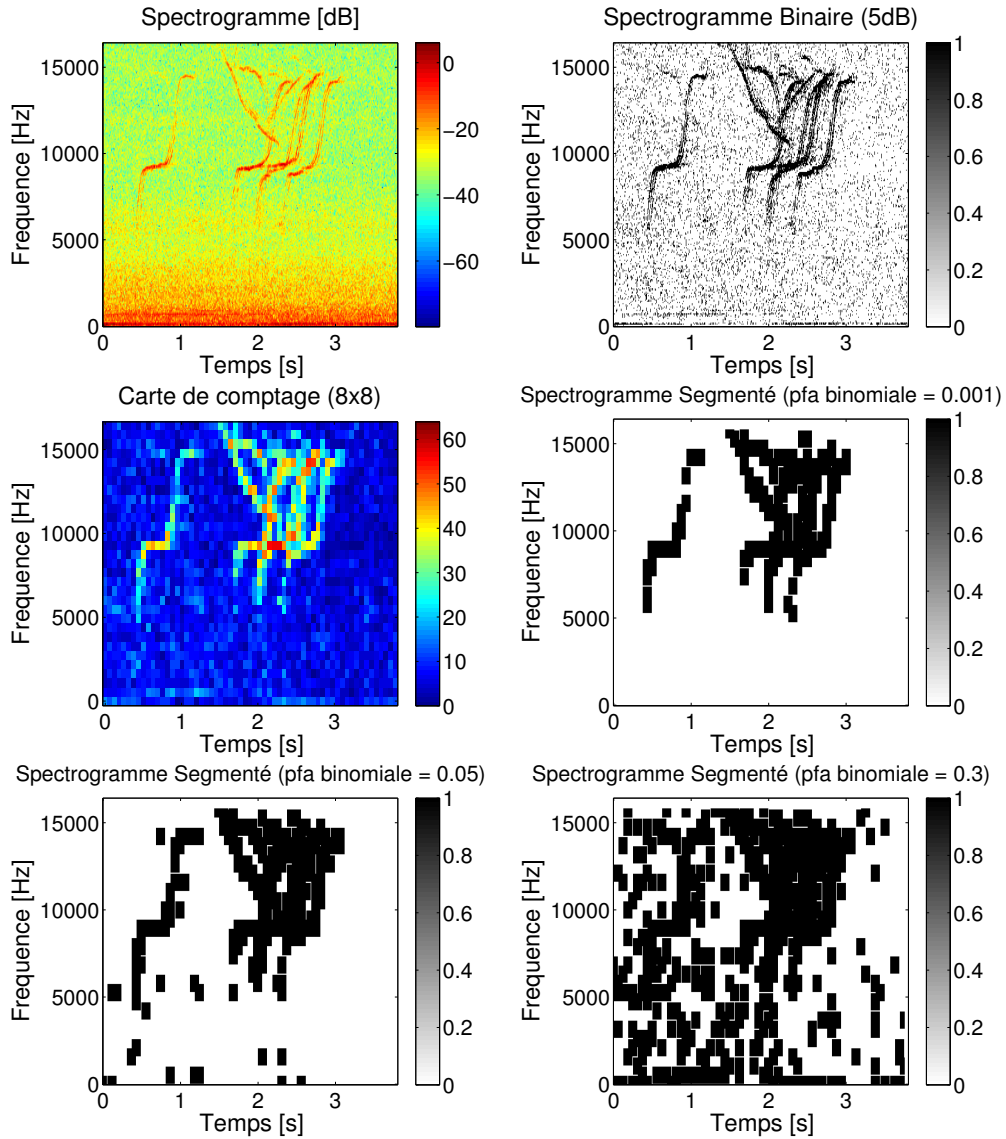


FIGURE 3.9 – Illustration de l’influence du choix de la probabilité de fausse alarme  $p_{fa}^{\mathcal{R}}$ . Spectrogramme du signal calculé avec 50% de recouvrement et une fenêtre de Hann de 512 points, sans zero-padding et le spectrogramme binaire seuillé à 5dB. La carte de comptage est générée sur des régions de taille  $8 \times 8$  pixels que nous seuillons ensuite à  $p_{fa}^{\mathcal{R}} = 10^{-3}$ , 0.05 et 0.3.

# Chapitre 4

## Analyse de la densité des fausses alarmes (FADA)

### Sommaire

---

<b>4.1</b>	<b>Introduction . . . . .</b>	<b>90</b>
<b>4.2</b>	<b>Approche convolutive du comptage . . . . .</b>	<b>91</b>
<b>4.3</b>	<b>Complément sur les statistiques sous <math>\mathcal{H}_0^{\mathcal{R}}</math> . . . . .</b>	<b>93</b>
4.3.1	Redondance intrinsèque du spectrogramme . . . . .	94
4.3.2	Loi binomiale corrélée . . . . .	98
4.3.3	Estimation numérique par les spectrogrammes assistants.	99
<b>4.4</b>	<b>Conclusion . . . . .</b>	<b>103</b>

---

---

**Résumé du chapitre précédent :** L'information énergétique pour la détection de signaux sur le spectrogramme n'est pas suffisamment discriminante pour engendrer des compromis probabilité de fausse alarme/ probabilité de détection satisfaisants à bas RSB. Partant de ce constat, le chapitre précédent propose de réaliser un second test d'hypothèse binaire BHT2 sur des régions  $\mathcal{R}$  du spectrogramme binaire. Ce test cherche à déterminer s'il y a présence de signal dans ces régions. Pour ce faire nous comparons le nombre de pics temps-fréquence détectés dans  $\mathcal{R}$  (pour une  $p_{fa}$  donnée) à un nombre de pics attendu sous  $\mathcal{H}_0^{\mathcal{R}}$  (donné par un modèle binomial). BHT2 est résolu grâce à un critère de Neyman-Pearson.

**Objectif de ce chapitre :** Ce chapitre raffine la description du spectrogramme proposée dans le chapitre précédent : **1.** Nous constaterons d'abord que la méthode n'est pas invariante par translation temps-fréquence, et nous résoudrons ce problème par une approche convolutive. **2.** Nous constaterons que la prise en compte de la redondance intrinsèque du spectrogramme est nécessaire pour une bonne estimation de la loi  $p_{\mathcal{H}_0^{\mathcal{R}}}(\psi)$ .

---

## 4.1 Introduction

Le chapitre précédent propose une méthode de détection du support temps-fréquence étendu  $\mathcal{Q}_{\mathcal{R}}$  d'un signal. Cette méthode est basée sur quatre étapes successives. La première étape est la binarisation du spectrogramme du signal en utilisant un critère de Neyman-Pearson. La seconde étape est le partitionnement du spectrogramme binaire en sous-régions  $\mathcal{R}$  sur lesquelles le nombre de détections est compté. La troisième étape est la modélisation de la loi décrivant le nombre de détections attendues au sein de ces régions sous  $\mathcal{H}_0^{\mathcal{R}}$ . La quatrième et dernière étape est le calcul d'un seuil (par une approche de Neyman-Pearson) sur le nombre de détections au delà duquel une sous région est décidée  $\hat{\mathcal{H}}_1^{\mathcal{R}}$ . Se référer à la figure 3.5 du chapitre précédent pour un rappel schématique de la méthode.

Le principe ayant été exposé, nous raffinons maintenant les différentes étapes du processus. Ce chapitre a deux objectifs. Le premier objectif est traité en section 1. Il montre que le partitionnement du spectrogramme binaire en régions sans recouvrement n'assure pas l'invariance temporelle de la méthode. **Une approche convolutive, plus flexible, et possédant la propriété d'invariance temporelle est proposée.**

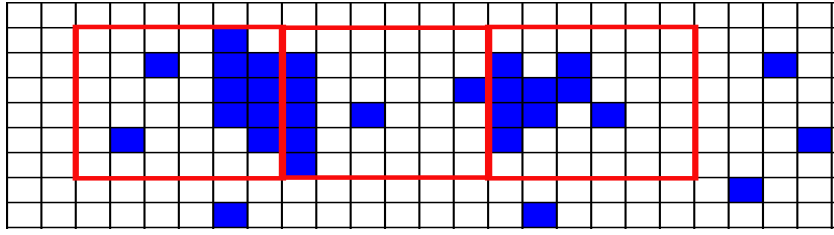
La section 2 montre que l'approximation binomiale proposée au premier chapitre est trop imprécise. La redondance intrinsèque du spectrogramme, induite par le recouvrement et l'étalement spectral dû au fenêtrage, corrèle les variables de Bernoulli. La loi théorique décrivant le comportement de la variable de comptage  $\Psi$  est la loi binomiale corrélée. Cette loi n'est exprimée de manière explicite dans la littérature que dans le cas d'une dépendance à l'ordre 1, ce qui n'est pas le cas du spectrogramme au sein duquel les variables sont corrélées aux ordres supérieurs. **Une méthode numérique d'estimation de cette loi est proposée.**

## 4.2 Approche convolutive du comptage

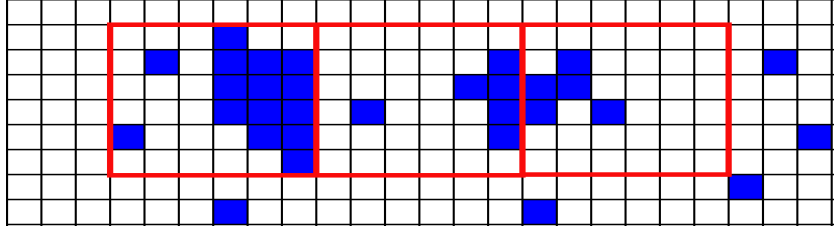
Le chapitre précédent propose de partitionner le spectrogramme binaire en un pavage régulier sans recouvrement sur lequel les détections sont comptées. Deux limitations fondamentales à cette méthode de pavage sont observées : l'ensemble des formes de régions possibles est limité et les résultats de segmentation ne sont pas invariants par translation temps-fréquence.

**Pavages possibles** : Les régions  $\mathcal{R}$  accessibles ne sont pas de forme arbitraire. Tel que défini au chapitre précédent, un partitionnement du spectrogramme binaire en régions  $\mathcal{R}$  n'est possible que si ces régions peuvent paver la totalité du plan sans se recouvrir. Un exemple de région simple mais pourtant non accessible est une région de forme circulaire. Il est impossible de paver l'ensemble du plan sans recouvrement avec des régions de forme circulaire.

**Non invariance par translation temps-fréquence** : le principal problème de la méthode de segmentation présentée précédemment est son instabilité. La segmentation obtenue est dépendante du positionnement de la grille. Une translation temps-fréquence d'un seul pixel peut altérer fortement les résultats de segmentation. La figure 4.1 en est un exemple schématique éloquent. Cette figure représente une partie d'un spectrogramme binaire fictif. Celui-ci est découpé en régions de taille  $6 \times 6$  sans recouvrement. Le seuil au delà duquel une région est décidée  $\hat{\mathcal{H}}_1^{\mathcal{R}}$  (appartenant à  $\mathcal{R}$ ) est arbitrairement fixé à 7. Deux exemples sont donnés. Dans le premier (figure 4.1.a), le nombre de détections est compté au sein de chaque région et les trois régions sont décidées  $\hat{\mathcal{H}}_1^{\mathcal{R}}$ . Dans le second exemple (figure 4.1.b) la grille est décalée d'un pixel vers la droite. Les détections sont à nouveau comptées, mais cette fois seule la région de gauche remplit la condition suffisante pour être détectée comme appartenant à  $\mathcal{R}$ . Pour le même spectrogramme binaire, une translation de la grille d'un seul pixel temps-fréquence altère dramatiquement le résultat de segmentation. Dans une logique de stabilité des résultats, ce cas de figure est extrêmement problématique.



(a) Nombre de détections dans les sous régions. A gauche : 10 détections, au milieu : 7 détections, à droite : 9 détections



(b) Nombre de détections dans les sous régions. A gauche : 15 détections, au milieu : 6 détections, à droite : 5 détections

FIGURE 4.1 – Illustration de l’instabilité dans l’estimation de  $\sqsupset_{\mathcal{R}}$  dans le cas sans recouvrement. Cas théorique : le seuil calculé  $\psi_T$  pour détecter le support est  $\psi_T = 7$ . Nous voyons ici qu’un décalage de 1 pixel vers la droite change complètement l’estimation du support. Cas (a) les trois régions sont décidées  $\hat{\mathcal{H}}_1^{\mathcal{R}}$ , cas (b) seule la région de gauche est décidée  $\hat{\mathcal{H}}_1^{\mathcal{R}}$ .

Nous décidons pour ces raisons de réinterpréter la manière de générer la carte de comptage, et donc la manière de détecter le support temps-fréquence du signal d’intérêt :

- Dans le chapitre 2 nous détectons l’ensemble des points appartenant à  $\sqsupset$ , le support temps-fréquence du signal.
- Dans le chapitre 3 nous estimons l’ensemble des régions temps-fréquence qui abritent une partie du support  $\sqsupset$ .
- Dans ce chapitre nous détectons l’ensemble des points temps-fréquence dont le voisinage contient au moins un pixel appartenant à  $\sqsupset$ .

La principale différence provient du fait que dans ce cas-ci, l’affectation  $\mathcal{H}_1^{\mathcal{R}}$  vaut pour le pixel  $[n, k]$  uniquement et non pas pour toute la région. Pour chaque pixel temps-fréquence du spectrogramme binaire, le nombre de ses voisins, au sein de  $\mathcal{R}$ , est compté et comparé à un seuil. Si ce nombre est supérieur au seuil alors le pixel est affecté à  $\mathcal{H}_1^{\mathcal{R}}$ . Cette manière de compter revient en pratique à convoluer le spectrogramme binaire  $\Phi$  avec une région  $\mathcal{R}$  contenant des 1 et des 0. La région est de forme arbitraire et ne nécessite pas la connexité du voisinage en question (voir figure 4.2). La carte de comptage  $\Psi$  est alors redéfinie de la manière suivante :

$$\Psi = \Phi_{pfa} * \mathcal{R} \quad (4.1)$$

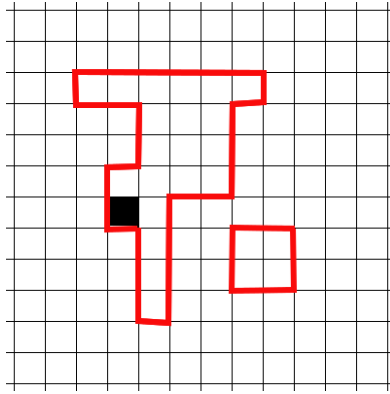


FIGURE 4.2 – Exemple d’une région  $\mathcal{R}$ . Le pixel temps-fréquence auquel sera affecté la valeur du nombre de détection dans son voisinage est le pixel noir. Son voisinage est défini comme l’intérieur de la région en rouge. Cette région possède une forme arbitraire et n’est pas connexe.

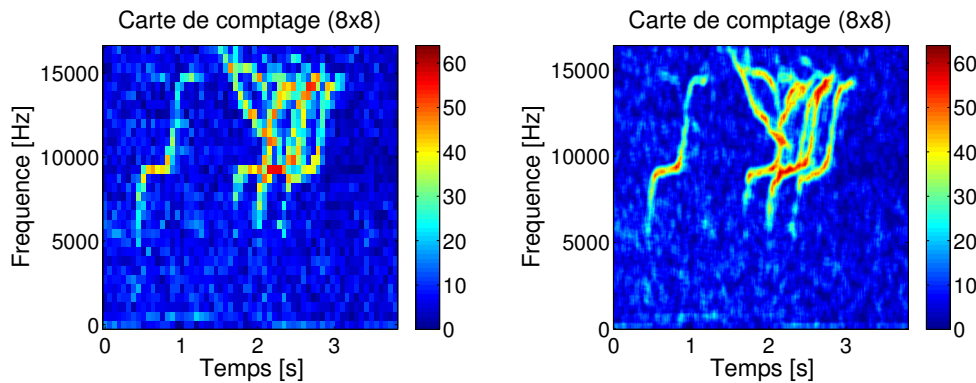


FIGURE 4.3 – Cartes de comptage.  $\mathcal{R} = 8 \times 8$  pixels temps-fréquence. À gauche : découpage sans recouvrement. À droite : convolution avec  $\mathcal{R}$

avec  $*$  l’opérateur de convolution bidimensionnelle. De manière explicite :

$$\psi[n, k] = \sum_{t_0=-\infty}^{+\infty} \sum_{f_0=-\infty}^{+\infty} \phi_{p_{fa}}[t - t_0, f - f_0] \mathcal{R}[t_0, f_0] \quad (4.2)$$

La figure 4.3 montre une comparaison de cartes de comptage. À gauche la carte de comptage générée par partitionnement et comptage des détections au sein des régions, à droite la carte générée par convolution du spectrogramme binaire avec  $\mathcal{R}$ . Nous constatons qu’en plus de posséder cette bonne propriété d’invariance par translation temps-fréquence, la carte générée par convolution localise mieux les groupes de pics temps-fréquence. La carte possède une meilleure résolution.

### 4.3 Complément sur les statistiques sous $\mathcal{H}_0^{\mathcal{R}}$

Cette section montre que l’approximation d’indépendance des variables de Bernoulli proposée au chapitre précédent n’est pas tenable. En effet, ces variables sont

fortement corrélées et la densité de probabilité de leur somme n'est plus binomiale. Nous étudions la forme de la densité de probabilité dans le cas corrélé. La formulation théorique ne possédant d'expression analytique explicite, nous introduisons une méthode numérique d'estimation de cette loi.

### 4.3.1 Redondance intrinsèque du spectrogramme

Le Chapitre 3 proposait de négliger la corrélation entre les variables de Bernoulli afin d'obtenir une expression analytique explicite de  $p_{\mathcal{H}_0^R}(\psi)$ . L'hypothèse d'indépendance est fautive de manière générale. Il existe deux causes de dépendance qui sont liées à la construction du spectrogramme : **1.** la première cause est le recouvrement entre fenêtres d'analyse qui induit une redondance temporelle de la répartition d'énergie, **2.** la seconde cause est l'étalement spectral liée au fenêtrage de la transformée de Fourier à court terme qui induit une redondance fréquentielle de la répartition d'énergie.

Cette redondance induit une sur-dispersion des lois de comptage observées. La figure 4.4 montre l'effet de cette redondance. Nous calculons d'abord le spectrogramme calculé sans recouvrement et nous comptons les détections sur une seule bande de fréquence de longueur 25 pixels (figure 4.4a). Nous évitons ainsi toute dépendance entre les coefficients, les lois binomiales se superposent parfaitement. Nous calculons ensuite le spectrogramme avec 50% de recouvrement et nous comptons les détections sur une fenêtre de taille  $5 \times 5$  pixels (figure 4.5b). Si l'effet de la corrélation était négligeable alors la loi devrait aussi se superposer parfaitement aux observations. Ce n'est pas le cas, les lois observées sont plus dispersées que celles prédites.

Quels sont les facteurs influençant la corrélation entre les coefficients ? Nous en étudions quatre : le recouvrement entre fenêtres, la forme de la fenêtre, sa longueur et le zéro-padding. Nous étudions l'impact de la corrélation sous l'angle de la divergence de Kullback-Leibler entre la PDF observée et celle prédite :

**Recouvrement entre fenêtres :** Le recouvrement entre fenêtres lors de la construction du spectrogramme est bien entendu un facteur influant sur la corrélation. Le recouvrement est la quantité d'énergie partagée entre les pixels adjacents temporellement. Plus il y a de recouvrement, plus grande sera la corrélation entre variables. C'est bien entendu ce que nous observons sur la figure 4.7c pour toutes les formes de fenêtres de pondération présentées.

**Forme de la fenêtre de pondération :** Le choix de la fenêtre de pondération reflète le compromis fait par l'utilisateur entre la capacité à localiser précisément un pic temps-fréquence et la capacité à séparer deux pics proches dont les amplitudes

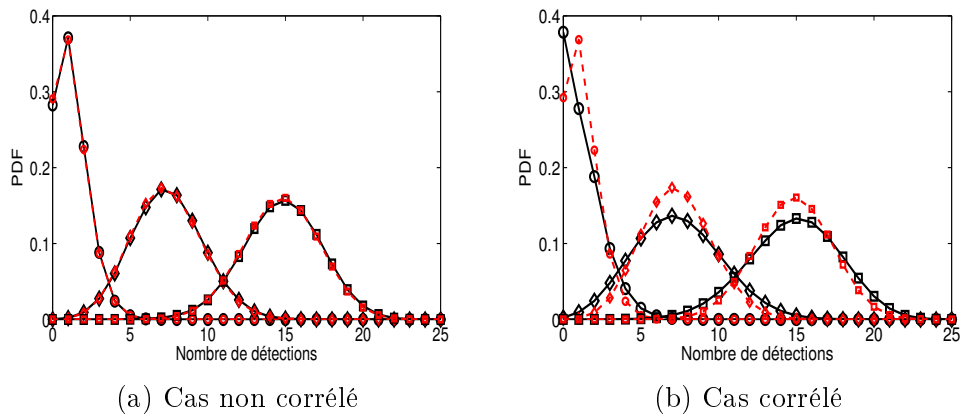


FIGURE 4.4 – Effet de la corrélation sur la PDF de  $\Psi$ . En trait plein : PDF observées, en tirets : lois binomiales prédites. Les courbes avec différents marqueurs correspondent à différentes  $p_{fa}$  : "O" correspond à  $p_{fa} = 0.05$ , "◇" à  $p_{fa} = 0.3$ , "□" à  $p_{fa} = 0.6$ . A gauche le comptage est effectué sur une seule bande de fréquence, et le spectrogramme est calculé sans recouvrement. Les lois observées et prédites se superposent parfaitement. A droite le comptage est effectué sur un carré temps-fréquence de taille  $5 \times 5$  avec 50% de recouvrement. Les PDF observées sont sur-dispersées par rapport aux lois binomiales.

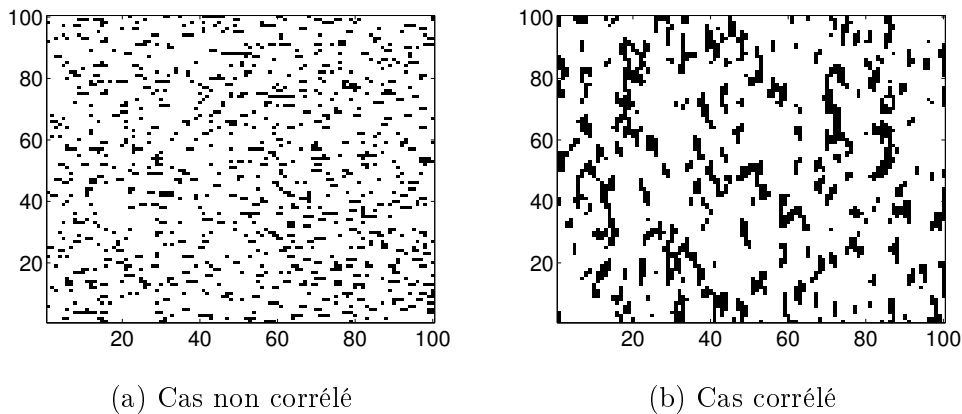


FIGURE 4.5 – Effet de la corrélation sur la répartition des fausses alarmes sur le spectrogramme binaire. Un signal Gaussien synthétique de variance unité est généré. : (a) Le spectrogramme est construit avec une fenêtre rectangulaire, sans recouvrement entre fenêtres, et seuillé avec une probabilité de fausse alarme  $p_{fa} = 0.1$ . Les fausses alarmes se répartissent uniformément sur le plan ; (b) Le spectrogramme est construit avec une fenêtre de Hanning, 90% de recouvrement entre fenêtres et seuillé avec une probabilité de fausse alarme  $p_{fa} = 0.1$ . Les fausses alarmes forment des agrégats. La répartition n'est plus uniforme.



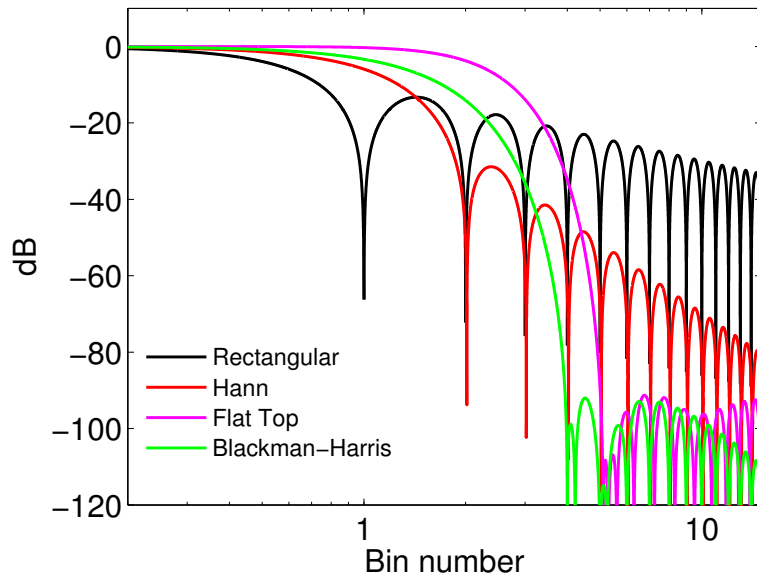
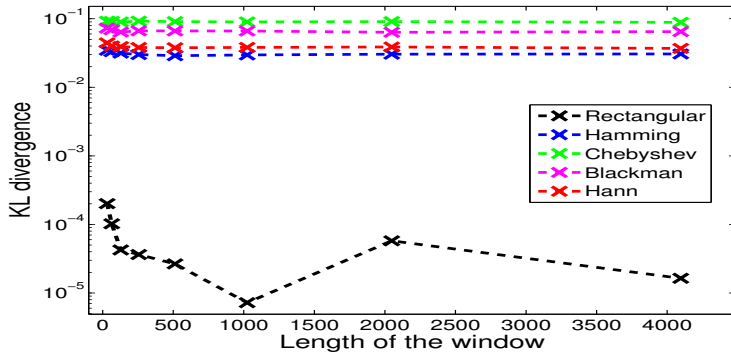


FIGURE 4.6 – Transformée de Fourier de plusieurs fenêtres d’apodisation [Har78].

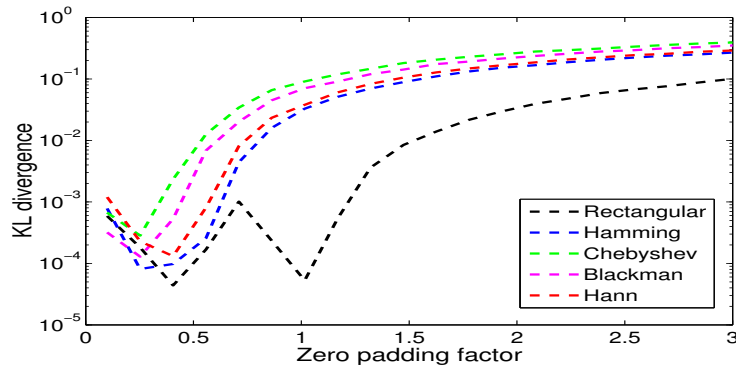
sont très différentes. Plus explicitement, nous choisissons l’importance relative des lobes secondaires par rapport au lobe principal (voir figure 4.6). Le compromis classique lorsque nous sommes confrontés à des signaux ayant des dynamiques d’amplitude très grandes consiste alors à choisir une fenêtre de pondération non rectangulaire. Ainsi plus le lobe principal est large, plus forte est la corrélation entre pixels adjacents fréquentiellement. Nous pouvons le constater pour toutes les configurations possibles, plus étroit est le lobe principal et plus grande est la divergence (figures 4.7a, 4.7b et 4.7c).

**Longueur de la fenêtre de pondération :** L’importance relative des pixels adjacents par rapport au lobe principal ne dépend pas de la longueur de la fenêtre de pondération. C’est pourquoi la longueur de la fenêtre n’a pas d’impact sur la corrélation des coefficients (voir figure 4.7a).

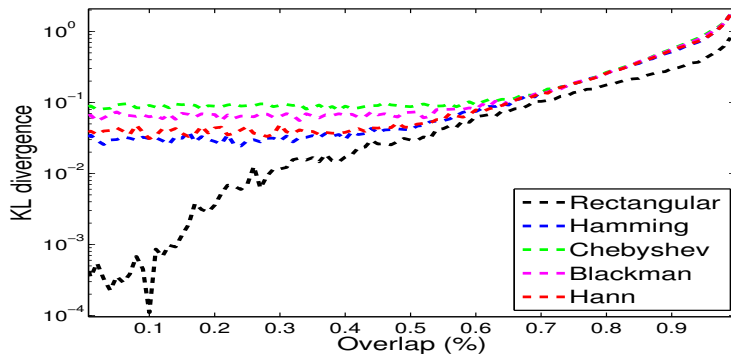
**Zéro padding :** Le zéro-padding consiste à compléter le signal temporel par des zéros. Le nombre de points de calcul de la transformée de Fourier discrète est alors augmenté mais il n’y a pas de rajout d’information (puisque l’on ne fait que rajouter des zéros). Sans zéro-padding le contenu spectral d’un signal de  $M$  points est décrit par  $M$  points espacés uniformément entre 0 et  $f_s/2$ . Avec zéro-padding le contenu spectral du même signal est décrit à l’aide de plus de points. Dans le cas de l’analyse d’une fréquence pure, il y aura donc plus de points pour décrire le lobe principal de cette fréquence dans le spectre, les valeurs seront donc toutes corrélées. Plus les variables sont corrélées plus loin nous serons du modèle binomial et donc plus grande la divergence (voir figure 4.7b).



(a) Évolution de la divergence de Kullback-Leibler entre la PDF observée et la loi binomiale prédite en fonction de la longueur de la fenêtre de pondération. Pas d'impact de cette longueur, quelle que soit la fenêtre utilisée.



(b) Évolution de la divergence de Kullback-Leibler entre la PDF observée et la loi binomiale prédite en fonction du facteur de zéro-padding. Plus grand est ce facteur, plus grande est la corrélation et donc la divergence.



(c) Évolution de la divergence de Kullback-Leibler entre la PDF observée et la loi binomiale prédite en fonction du recouvrement entre fenêtres lors de la construction du spectrogramme. Plus grand est le recouvrement, plus grande est la divergence.

FIGURE 4.7 – Divergence de Kullback-Leibler entre la PDF observée et la loi binomial prédite pour plusieurs paramètres.

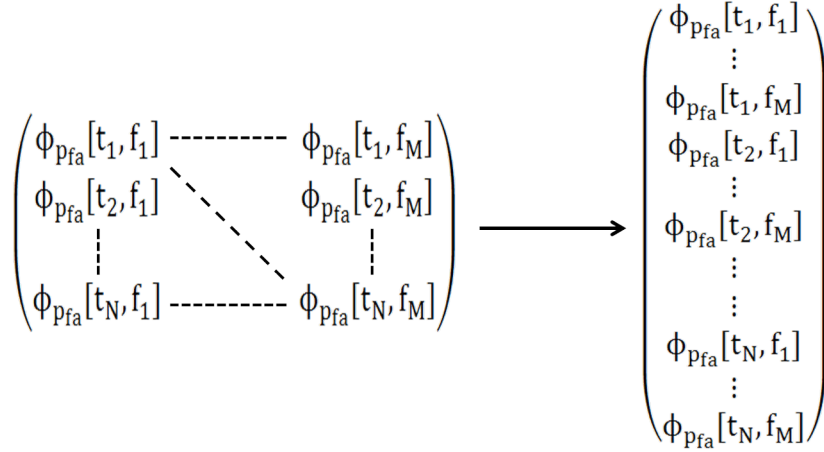


FIGURE 4.8 – Notation linéaire

Tous ces paramètres ont donc un impact sur la forme de la PDF observée. La question principale étant la suivante : comment prédire la forme que va prendre la PDF sous  $\mathcal{H}_0^R$  en fonction des paramètres de construction du spectrogramme, de la  $p_{fa}$  utilisée et de  $\mathcal{R}$  ?

### 4.3.2 Loi binomiale corrélée

La loi binomiale corrélée est une loi de probabilité discrète à support compact décrivant les valeurs prises par la somme de variables de Bernoulli corrélées. Nous cherchons à la définir dans le cas du spectrogramme.

Afin de simplifier les notations nous définissons  $\phi_{lin} = \{\phi_1, \phi_2, \dots, \phi_A\}$  la notation linéaire correspondant à l'ensemble des points  $\phi_{pfa}[n, k]$  dans la région  $\mathcal{R}$  (voir figure 4.8). Une expression analytique exacte de la distribution jointe d'une variable binaire en tenant compte des coefficients de corrélation est donnée en [Bah61]. Afin d'obtenir une expression simple à visualiser, la variable  $z_i$  est introduite, elle est la version normalisée de  $\phi_i$  :

$$z_i = (\phi_i - p_{fa}) / \sqrt{p_{fa}(1 - p_{fa})} \quad (4.3)$$

Il est alors montré dans [Bah61] que la loi de probabilité jointe de  $\phi_{lin}$  tenant compte de la corrélation entre tous les coefficients s'écrit de manière générale de la façon suivante :

$$\check{p}(\phi_{lin}) = p(\phi_{lin}) \left( 1 + \sum_{1 \leq i < j \leq A} E(Z_i Z_j) z_i z_j + \sum_{1 \leq i < j < k \leq A} E(Z_i Z_j Z_k) z_i z_j z_k + \dots + E(Z_i Z_j \dots Z_A) z_i z_j \dots z_A \right) \quad (4.4)$$

Cette loi de distribution corrigée est alors non seulement une fonction de la corrélation entre variables mais aussi de toutes les corrélations d'ordres supérieurs. La loi décrivant la valeur de la somme  $\psi$  de ce vecteur  $\phi_{lin}$  est la suivante :

$$p(\psi = k) = \sum_{\phi:t(\phi)=k} \check{p}(\phi_{lin}) \quad (4.5)$$

Avec  $t(\phi_{lin}) = \sum_{i=1}^A \phi_i$ . L'expression générale décrivant cette probabilité n'est pas utilisable en pratique, c'est pourquoi une approximation à l'ordre 2 de cette loi est proposée [Bah61, ZK13]. Les corrélations d'ordres supérieurs à 3 sont alors considérés nuls, i.e.  $E(\phi_i\phi_j\phi_k) = \dots = E(\phi_1\phi_2\dots\phi_A) = 0$ .

**Remarque :** Si  $E(\phi_i\phi_j)$  est lui aussi choisi nul pour tous les couples  $(i, j)$  alors on se retrouve dans le cas non corrélé, et la somme s'exprime comme une loi binomiale.

Dans le cas du spectrogramme il n'est cependant pas possible de négliger ces corrélations d'ordres supérieurs. La figure 4.9 montre l'exemple de  $E(\phi_i\phi_j)$ ,  $E(\phi_i\phi_j\phi_k)$  et  $E(\phi_i\phi_j\phi_k\phi_l)$ . Si ces corrélations étaient négligeables alors les seuls éléments non nuls seraient les  $E(\phi_i\phi_i\phi_i)$ ,  $E(\phi_i\phi_i\phi_i\phi_i)$ , etc. Ce qui n'est pas le cas.

Nous faisons alors face à deux problèmes avec la formulation analytique implicite de  $p_{\mathcal{H}_0^{\mathcal{R}}}(\Psi)$  donnée dans l'équation (4.5) :

1. Nous n'avons pas de formulation de  $E[\phi_{i_1}\dots\phi_{i_k}]$  en fonction des paramètres de construction du spectrogramme, de la probabilité de fausse alarme  $p_{fa}$  et de  $\mathcal{R}$ .
2. Nous n'avons pas de formulation explicite de  $p_{\mathcal{H}_0^{\mathcal{R}}}(\Psi)$  en fonction des  $E[\phi_{i_1}\dots\phi_{i_k}]$  lorsque ceux-ci ne s'annulent pas pour  $k \geq 2$ .

Une formulation théorique explicite de  $p_{\mathcal{H}_0^{\mathcal{R}}}(\psi)$  semble difficile à obtenir, et la loi implicite n'est pas utilisable en pratique. C'est pourquoi nous proposons une méthode d'estimation numérique de cette loi.

### 4.3.3 Estimation numérique par les spectrogrammes assistants.

L'expression analytique explicite de  $p_{\mathcal{H}_0^{\mathcal{R}}}(\psi)$  en fonction de tous les paramètres est difficile à obtenir, **nous introduisons par conséquent la notion de spectrogramme assistant.**

Un spectrogramme, dit assistant, est construit de manière à imiter la structure de corrélation du spectrogramme d'étude. Il n'a pas de sens physique et ne sert que comme intermédiaire de calcul pour obtenir très simplement une estimation

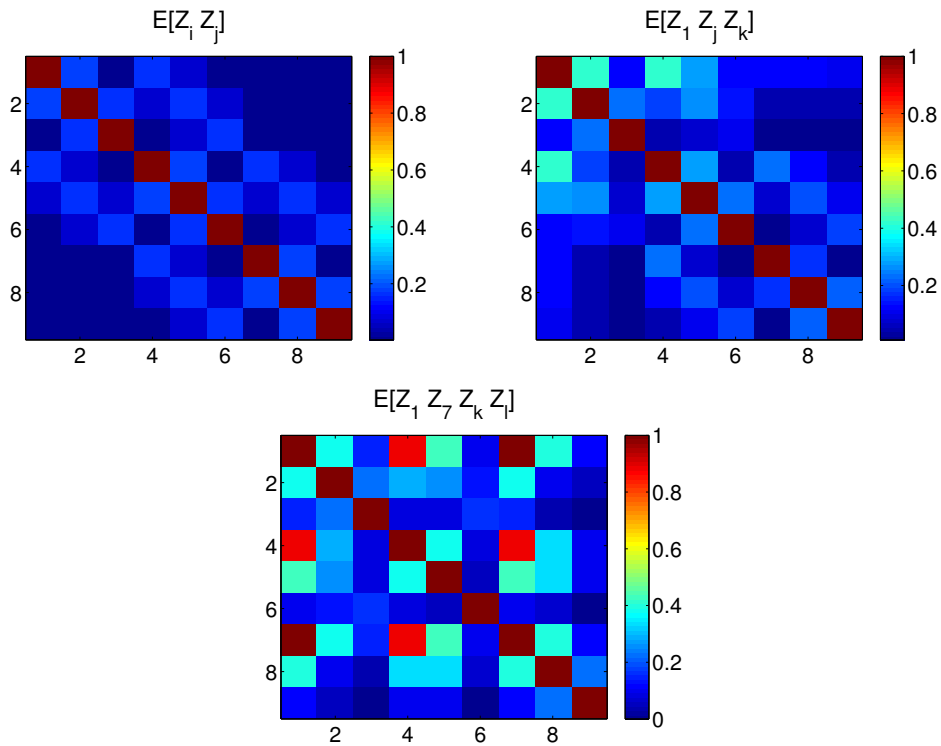


FIGURE 4.9 – Calcul des moments d’ordres supérieurs  $E[Z_i Z_j]$ ,  $E[Z_1 Z_i Z_j]$  et  $E[Z_1 Z_7 Z_i Z_j]$ . Le spectrogramme a été calculé avec 75% de recouvrement, une fenêtre de Hann de 512 points. Le spectrogramme a été binarisé avec  $p_{fa} = 0.01$ . Le spectrogramme binaire a été découpé en régions de taille  $3 \times 3$  sans recouvrement. Un total de  $2.5 \times 10^6$  réalisations de  $\Phi$  sont utilisées pour l’estimation. Les matrices ne sont pas diagonales, y compris pour les ordres supérieurs.

numérique de  $p_{\mathcal{H}_0^{\mathcal{R}}}(\psi)$ .

L'idée est de synthétiser un signal simple, et de lui appliquer les mêmes transformations que celles que nous appliquons au signal d'intérêt. **Le but de ces spectrogrammes est d'imiter la structure de la corrélation des variables au sein du spectrogramme.** Nous estimons alors les propriétés du signal réel à partir du signal synthétique.

Cette méthode est très flexible dans le sens où elle permet d'estimer  $p_{\mathcal{H}_0^{\mathcal{R}}}(\psi)$  quelle que soit la structure de corrélation des variables. Finalement, la connaissance de la structure de corrélation des variables n'est même pas requise, il est seulement nécessaire de savoir comment générer une corrélation du même type qu'effectivement observée dans le cas qui nous intéresse.

Bien que différent en pratique, le principe que nous proposons ici, est similaire aux "knockoffs filters" développés dans le cadre du contrôle du "False Discovery Rate" pour des variables corrélées [BC14].

### **En pratique :**

Un signal synthétique, aléatoire, blanc, Gaussien, stationnaire et de variance unité est généré. Le spectrogramme de ce signal est construit en utilisant les mêmes paramètres que ceux utilisés pour le signal d'intérêt. **Il est fondamental de noter** qu'il n'est pas nécessaire de générer un signal dont la DSP imite celle de la DSP du bruit, seule la structure de corrélation nous intéresse ici. La raison en est la suivante : dans le cas réel nous estimons la DSP du bruit et seuillons le spectrogramme avec un critère de Neyman-Pearson. En supposant que nous réalisons correctement (i.e. une estimation non biaisée et de faible variance) l'étape d'estimation de la DSP, le résultat obtenu par seuillage est indépendant de la forme de la DSP du bruit. C'est pour cela que le signal synthétisé est blanc et stationnaire.

Le spectrogramme assistant est ensuite seuillé avec la même probabilité de fausse alarme  $p_{fa}$  que celle utilisée pour le spectrogramme du signal d'intérêt. Ce spectrogramme assistant binaire possède alors les mêmes statistiques que le spectrogramme binaire d'intérêt sous l'hypothèse  $\mathcal{H}_0^{\mathcal{R}}$ . Pour finir nous le convoluons avec la région de comptage  $\mathcal{R}$ .

Nous estimons la densité de probabilité de cette carte de comptage assistante par l'histogramme, qui est donc par définition une estimation de  $p_{\mathcal{H}_0^{\mathcal{R}}}(\psi)$ . La précision de cette estimation est uniquement dépendante de la longueur du signal synthétique utilisé.

La grande flexibilité des spectrogrammes assistants offre un avantage non négligeable. Des lois plus complexes que la loi binomiale corrélée peuvent être estimées très simplement. Nous pouvons par exemple choisir des régions de comptage  $\mathcal{R}$  non uniformes. Ainsi, au lieu de n'être composée que de 0 et de 1, la fenêtre  $\mathcal{R}$  peut par exemple être choisie Gaussienne et assigner un poids différent aux pixels en fonction de la distance au point d'étude. La méthode d'estimation numérique est identique quand bien même le (très difficile) problème analytique associé est le suivant : "Quelle est la densité de probabilité de la somme pondérée de variables binaires corrélées fréquemment et temporellement aux ordres supérieurs ?"

Dans le chapitre suivant nous utiliserons cette simplicité pour étudier les possibilités d'éliminations de la probabilité de fausse alarme  $p_{fa}$  en tant que degré de liberté de FADA.

Nous résumons en cinq étapes la **méthode d'estimation de  $p_{\mathcal{H}_0}^{\mathcal{R}}(\Psi)$  par les spectrogrammes assistants** :

- **Étape 1** : Génération d'un signal Gaussien centré blanc stationnaire de variance unité
- **Étape 2** : Calcul de son spectrogramme en utilisant les mêmes paramètres que ceux désirés pour l'analyse du signal (imitation de la structure de corrélation)
- **Étape 3** : Seuillage du spectrogramme avec la même probabilité de fausse alarme  $p_{fa}$
- **Étape 4** : Génération de la carte de comptage  $\Psi$
- **Étape 5** : Estimation de la densité de probabilité  $p_{\mathcal{H}_0}^{\mathcal{R}}(\Psi)$  par l'histogramme.

La figure 4.10 illustre la pertinence de cette approche sur données réelles. Nous avons choisi un extrait d'une minute d'un enregistrement de bruit sous-marin tiré de la base de données ERATO-09 du SHOM. Nous avons vérifié manuellement qu'aucun signal non-stationnaire n'était présent dans cette partie de l'enregistrement. Nous avons alors seuillé ce spectrogramme avec une probabilité de fausse alarme égale à 0.5. Les détections sont ensuite comptées sur des régions de taille  $7 \times 7$ . La densité de probabilité de  $\Psi$  est estimée à partir de l'histogramme des valeurs observées. Nous comparons cette densité de probabilité avec celle prédite par le modèle binomial. Nous constatons que la PDF observée est sur-dispersée par rapport à ce modèle. Nous traçons aussi la densité de probabilité estimée grâce aux spectrogrammes assistants. Nous constatons que le modèle prédit numériquement

colle bien à l'observation.

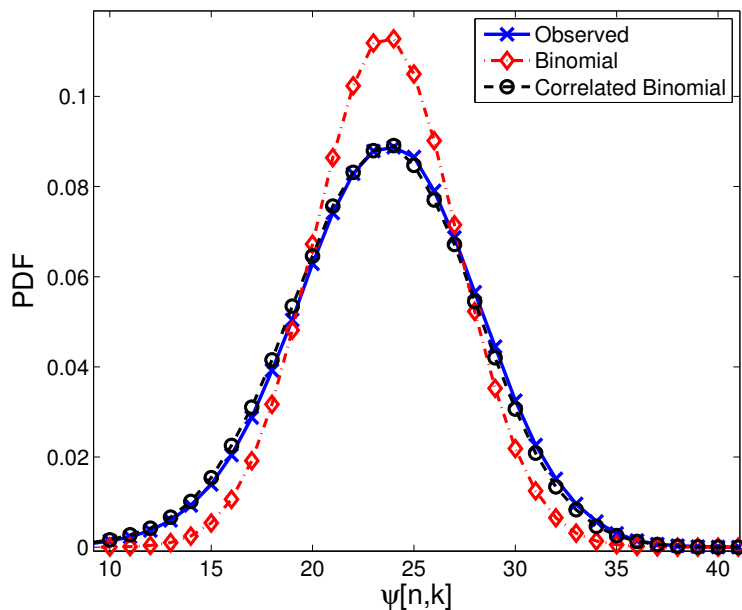


FIGURE 4.10 – Illustration de la méthode d'estimation de la densité de probabilité de  $\Psi$  sous  $\mathcal{H}_0^R$  par les spectrogrammes assistants sur données réelles. Trois courbes sont représentées. Le marqueur 'croix' montre la densité de probabilité effectivement observée sur données réelles. Le marqueur 'diamant' montre le modèle binomial associé (Cf chapitre 3). Le marqueur 'rond' montre la densité de probabilité prédite par les spectrogrammes assistants.

Nous montrons sur un exemple réel tiré du même enregistrement (ERATO-09) la capacité de la méthode proposée à traiter un très grand nombre de fausses alarmes (voir figure 4.11). Le spectrogramme du signal est seuillé avec une probabilité égale à 0.5. Nous estimons la densité de probabilité des fausses alarmes en utilisant la méthode des spectrogrammes assistants et résolvons BHT2 avec une probabilité de fausse alarme égale à 0.01. Malgré le très grand nombre de fausses alarmes générées par BHT1, nous sommes capables de discriminer de manière très efficace les vraies détections des fausses alarmes. Il est à noter cependant que le résultat obtenu n'est pas nécessairement le meilleur atteignable. Nous souhaitons ici illustrer la capacité à traiter un grand nombre de fausses alarmes tout en fournissant des résultats qui restent cohérents.

## 4.4 Conclusion

Ce chapitre est une extension directe du chapitre 3. Il apporte et discute deux points fondamentaux.

Le premier est lié à la définition des régions de comptage, alors que le partition-



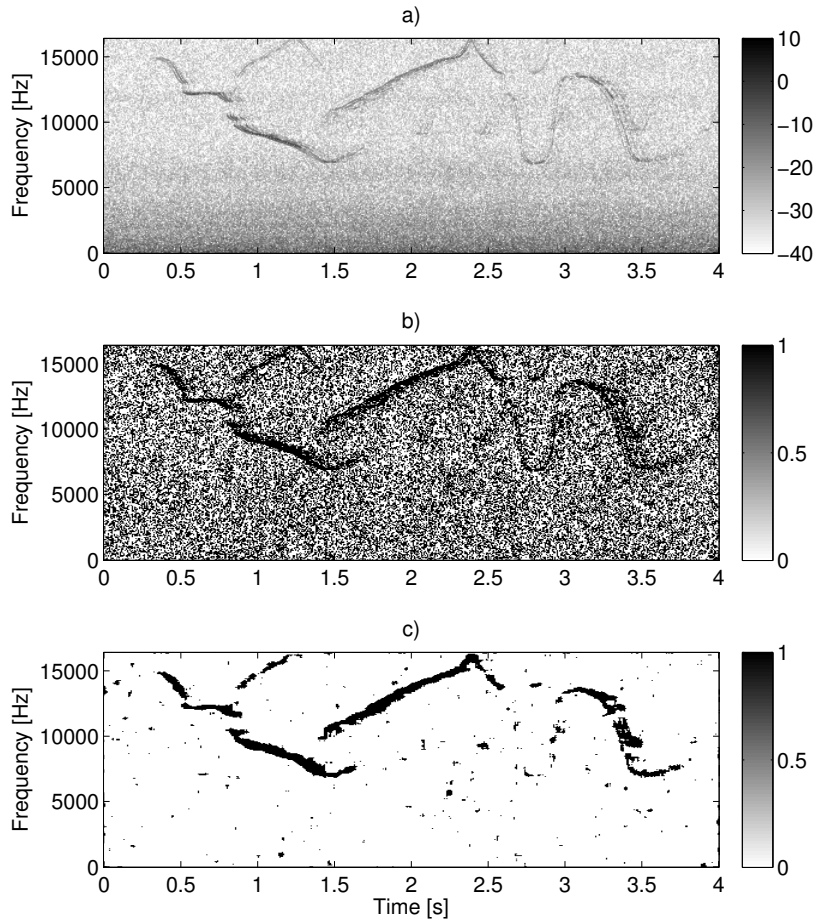


FIGURE 4.11 – Illustration de la capacité de FADA à gérer un grand nombre de fausses alarmes. Exemple de la détection de sifflements de dauphins sur un enregistrement sous-marin (ERATO-09, SHOM) : a) Spectrogramme du signal (calculé sur 512 points avec une fenêtre de Hann de 512 points et 256 points de recouvrement), b) Sortie de BHT1 : spectrogramme binaire obtenu avec  $p_{fa} = 0.5$  et c) sortie de BHT2 : spectrogramme segmenté avec  $p_{fa} = 0.01$

nement sans recouvrement formulé dans le Chapitre 2 permettait de comprendre le principe de la méthode, cette formulation n'est pas efficace dans un contexte réel. En particulier, la méthode n'est pas invariante par translation temps-fréquence. Nous reformulons la génération de la carte de comptage en terme de convolution bi-dimensionnelle avec une matrice uniforme, ce qui résout le problème.

Le second point est une critique de l'approximation binomiale. Nous montrons que la corrélation entre les coefficients du spectrogramme n'est pas négligeable dans notre contexte. La corrélation disperse fortement la densité de probabilité  $p_{\mathcal{H}_0^{\mathcal{R}}}(\psi)$  effectivement observée par rapport au modèle binomial formulé au chapitre 3. Nous formulons alors la loi théorique, qui est binomiale corrélée. Aucune formulation explicite de cette loi n'est disponible lorsque les corrélations d'ordres supérieurs ne s'annulent pas. Par conséquent nous introduisons les spectrogrammes assistants qui ont pour but d'imiter la structure de corrélation du spectrogramme, ce qui nous permet d'estimer numériquement la loi binomiale corrélée. Cette méthode se montre très efficace pour la gestion d'un grand nombre de fausses alarmes.



# Chapitre 5

## Analyse des performances en simulation, et validation en contexte réel

### Sommaire

---

<b>5.1</b>	<b>Introduction</b>	<b>108</b>
<b>5.2</b>	<b>Choix de la probabilité de fausse alarme <math>p_{fa}</math></b>	<b>109</b>
5.2.1	Maximisation du nombre de détections sous contrainte	109
5.2.2	Vraisemblance intégrale	111
5.2.3	Norme du vecteur de vraisemblance	116
<b>5.3</b>	<b>Forme de la région de comptage <math>\mathcal{R}</math></b>	<b>118</b>
5.3.1	Pour la détection de sifflements	118
5.3.2	Pour la détection de clics	123
<b>5.4</b>	<b>Évaluation de la méthode</b>	<b>125</b>
<b>5.5</b>	<b>Illustration sur données réelles</b>	<b>131</b>
5.5.1	Segmentation de sifflements	131
5.5.2	Segmentation de sifflements en milieu impulsionnel	132
<b>5.6</b>	<b>Conclusion</b>	<b>136</b>

---

---

**Résumé du chapitre précédent :** Le chapitre précédent a affiné la méthode présentée au Chapitre 3. La carte de comptage  $\Psi$  y a été réinterprétée comme une convolution bidimensionnelle du spectrogramme binaire avec une région  $\mathcal{R}$  de forme arbitraire. Ensuite, nous avons affiné la description des statistiques de  $\Psi$  pour prendre en compte la redondance du spectrogramme. La formulation analytique de la PDF de  $\Psi$  étant implicite et donc non exploitable, nous avons proposé une méthode permettant de l'estimer numériquement.

**Objectif de ce chapitre :** Ce chapitre a trois objectifs principaux. D'abord nous discutons du choix de la probabilité de fausse alarme  $p_{fa}$ . Nous proposons quatre manières de la choisir et étudions les performances de chacune. Nous nous intéressons ensuite au problème du choix de la région de comptage  $\mathcal{R}$ . Nous proposons des critères de choix pour la détection de sifflements et pour la détection de clics puis nous évaluons numériquement ces méthodes. Finalement nous donnons une méthode de segmentation des sifflements en milieu impulsif. Nous illustrons la méthode sur données réelles.

---

## 5.1 Introduction

Jusqu'à présent nous avons décrit un cadre statistique permettant de détecter des groupes de pics temps-fréquence dans le spectrogramme. Cette méthode qui tire profit de la densité des fausses alarmes dans le spectrogramme binaire pour prendre une décision régionale possède 5 degrés de liberté que nous listons ci-dessous :

1. Le premier degré de liberté concerne les paramètres de construction du spectrogramme.
2. Le second degré de liberté concerne la région  $V$  sur laquelle la densité spectrale de puissance du bruit  $\gamma_b[n, k]$  est estimée. Ce choix a déjà été discuté dans le chapitre 2. Il dépend de la stationnarité et de la blancheur du bruit ambiant. Nous ne reviendrons pas dessus dans ce chapitre.
3. Le troisième degré de liberté concerne le choix de la probabilité de fausse alarme  $p_{fa}$  utilisée pour résoudre le premier test d'hypothèse binaire.
4. Le quatrième degré de liberté concerne le choix de la région de comptage  $\mathcal{R}$ .
5. Le cinquième degré de liberté concerne le choix de la probabilité de fausse alarme du second test d'hypothèse  $p_{fa}^{\mathcal{R}}$ .

Pour être applicable de manière efficace en contexte opérationnel, c'est à dire pour diverses conditions de bruit et/ou de signaux et/ou d'enregistrement (hydrophone embarqué ou fixe par exemple), nous pensons que la méthode construite doit avoir le moins de paramètres possibles à régler au cas par cas. Ce chapitre s'intéresse ainsi dans un premier temps aux choix de la probabilité de fausse alarme  $p_{fa}$  d'une part, et à la forme de la région de comptage  $\mathcal{R}$  d'autre part. Nous donnerons des critères de choix pour la détection de sifflements et pour la détection de clics.

Un second objectif de ce chapitre concerne l'évaluation des performances de FADA sur signaux synthétiques. Nous comparerons les performances du détecteur FADA à celles du détecteur quadratique.

Enfin, nous illustrerons la méthode FADA sur données réelles d'odontocètes. Nous montrerons que la méthode proposée est applicable pour une large variété de signaux et de conditions de bruit. Nous montrerons en particulier des résultats de segmentation de sifflements en milieu fortement impulsif.

## 5.2 Choix de la probabilité de fausse alarme $p_{fa}$

Dans cette section nous nous posons la question du choix de  $p_{fa}$ . Nous avons vu dans le chapitre 3 que les résultats finaux de détection en sortie du second test d'hypothèse binaire sont dépendants du choix de la probabilité de fausse alarme. Il suffit de considérer les deux cas extrêmes pour s'en rendre compte, c'est à dire les cas  $p_{fa} = 0$  et  $p_{fa} = 1$ . Dans le cas  $p_{fa} = 0$  le signal n'est jamais détecté. Il n'y a pas non plus de fausses alarmes. En sortie du second test d'hypothèse binaire rien n'est détecté. Dans le cas  $p_{fa} = 1$  le signal est effectivement détecté, comme d'ailleurs l'ensemble des pixels du spectrogramme. Il n'est alors plus possible de faire la différence entre régions abritant du signal et régions n'abritant que des fausses alarmes puisque tout est détecté. Nous ne pouvons alors plus faire confiance aux détections obtenues, et ne détectons rien en sortie du second test d'hypothèse binaire.

Nous nous posons ici la question suivante : quel choix devons-nous faire lors du seuillage du spectrogramme ? Le choix de la  $p_{fa}$  doit, d'une certaine manière être adapté au signal d'intérêt. Si une  $p_{fa}$  égale à 0 ne laisse rien à détecter à BHT2, une  $p_{fa}$  trop proche de 1 noie les vraies détections parmi les fausses alarmes et il devient alors impossible de séparer  $p_{\mathcal{H}_0^{\mathcal{R}}}(\psi)$  de  $p_{\mathcal{H}_1^{\mathcal{R}}}(\psi)$ .

### 5.2.1 Maximisation du nombre de détections sous contrainte

Une première solution que nous avons présenté dans [DGI<sup>+</sup>13b] repose sur l'idée suivante. Puisque nous ne sommes pas capable de choisir une probabilité de fausse alarme  $p_{fa}$  *a priori* n'ayant pas d'information sur le signal, nous en essayons plusieurs.

Nous créons le vecteur  $\mathcal{P}_{\mathcal{FA}} = \{p_{fa}^{(1)}, p_{fa}^{(2)}, \dots, p_{fa}^{(N)}\}$ . Pour chaque  $p_{fa}^{(i)}$  un spectrogramme binaire  $\Phi_{p_{fa}^{(i)}}$  est généré.

Nous nous retrouvons donc avec un ensemble de spectrogrammes binaires  $\Phi = \{\Phi_{p_{fa}^{(1)}}; \Phi_{p_{fa}^{(2)}}; \dots; \Phi_{p_{fa}^{(N)}}\}$ .

Des cartes de comptage associées à chaque spectrogramme binaire sont générées par convolution avec la même matrice uniforme  $\mathcal{R} : \Psi = \{\Psi_{p_{fa}^{(1)}}; \Psi_{p_{fa}^{(2)}}; \dots; \Psi_{p_{fa}^{(N)}}\}$ .

Nous résolvons ensuite le second test d'hypothèse binaire pour chaque  $\Psi_{p_{fa}^{(i)}}$  en utilisant un critère de Neyman-Pearson.

La probabilité de fausse alarme  $p_{fa}^{\mathcal{R}}$  utilisée pour résoudre le second test d'hypothèse binaire est unique et ne dépend pas de  $p_{fa}^{(i)}$ . Nous cherchons donc à établir pour chaque  $\Psi_{p_{fa}^{(i)}}$  le seuil  $\psi_T^{(i)}$  tel que

$$p_{\mathcal{H}_0^{\mathcal{R}}}(\psi_{p_{fa}^{(i)}} > \psi_T^{(i)}) \leq p_{fa}^{\mathcal{R}} \quad (5.1)$$

Notons que dans notre cas il n'y a pas nécessairement égalité entre la probabilité de fausse alarme objectif  $p_{fa}^{\mathcal{R}, Obj}$  et la probabilité de fausse alarme  $p_{fa}^{\mathcal{R}, Reel}$ . Nous ne faisons que garantir que la probabilité de fausse alarme observée sera inférieure à la probabilité de fausse alarme objectif.

Pour chaque  $p_{fa}^{(i)}$  nous pouvons donc résoudre  $BHT2^{(i)}$  avec un critère de Neyman-Pearson :

$$BHT2^{(i)} : \quad \mathbb{F}^{(i)}[n, k] = \begin{cases} 0 & \text{if } \psi_{p_{fa}^{(i)}}[n, k] \leq \psi_T^{(i)} \\ 1 & \text{if } \psi_{p_{fa}^{(i)}}[n, k] > \psi_T^{(i)}, \end{cases} \quad (5.2)$$

$\mathbb{F}^{(i)}$  correspond au spectrogramme binaire en sortie de  $BHT2^{(i)}$ , c'est à dire le spectrogramme binaire obtenu après analyse de la densité des fausses alarmes dans  $\Phi_{p_{fa}^{(i)}}$ . Nous nous retrouvons donc avec un ensemble de spectrogrammes binaires  $\mathbb{F} = \{\mathbb{F}^{(1)}; \mathbb{F}^{(2)}; \dots; \mathbb{F}^{(N)}\}$ . Et nous choisissons finalement le spectrogramme binaire  $\mathbb{F}^*$  maximisant le nombre de détections en sortie de  $BHT2^{(2)}$ , sous contrainte de probabilité de fausse alarme constante :

$$\mathbb{F}^*[n, k] = \mathbb{F}^{(i_{\max})}[n, k] \quad (5.3)$$

avec  $i_{\max} = \underset{i}{\operatorname{argmax}} \sum_{n, k} \mathbb{F}^{(i)}[n, k]$

Nous résumons la méthode FADA avec maximisation sous contrainte avec un schéma en figure 5.1. Le spectrogramme du signal est tout d'abord calculé. Ce spectrogramme est ensuite seuillé en utilisant N probabilités de fausse alarme différentes. Nous obtenons alors N spectrogrammes binaires, qui représentent N compromis probabilité de fausse alarme/ probabilité de détection différents. N cartes de comptage sont générées par convolution avec une région  $\mathcal{R}$ . Le second

test BHT2 est alors résolu  $N$  fois, pour chacune des cartes de comptage en utilisant un critère de Neyman-Pearson identiques pour toutes. La probabilité de fausse alarme est fixée à  $p_{fa}^{\mathcal{R}}$ . Le spectrogramme segmenté que nous conservons est celui qui maximise le nombre de détections en sortie de BHT2.

## 5.2.2 Vraisemblance intégrale

Dans [Dad14] nous avons proposé une seconde manière d'aborder le problème. Plutôt que de choisir une probabilité de fausse alarme  $p_{fa}$  unique pour tous les pixels nous proposons de nous adapter localement au signal. Nous définissons alors la variable de vraisemblance  $L_{p_{fa}}$  :

$$L_{p_{fa}}[n, k] = p_{\mathcal{H}_0^{\mathcal{R}}}(\psi_{p_{fa}}[n, k]) \quad (5.4)$$

Elle traduit la probabilité d'observer  $\psi_{p_{fa}}$  détections au pixel  $[n, k]$  sous l'hypothèse  $\mathcal{H}_0^{\mathcal{R}}$  sachant que la probabilité de fausse alarme utilisée pour générer le spectrogramme binaire  $\Phi_{p_{fa}}$  est  $p_{fa}$ . Afin d'obtenir une grandeur qui ne dépende pas de  $p_{fa}$ . Nous introduisons alors la grandeur  $L^*$ . Celle-ci est la version intégrée de la vraisemblance par rapport à  $p_{fa}$  en chaque point  $[n, k]$  :

$$L^*[n, k] = \int_0^1 L_{p_{fa}}[n, k] dp_{fa}. \quad (5.5)$$

Nous proposons de résoudre le test d'hypothèse BHT2 sur cette grandeur. C'est à dire que nous prenons une décision sur la présence d'un groupe de pics temps-fréquence autour de  $[n, k]$  en se basant sur la vraisemblance intégrale. BHT2 est alors résolu de la manière suivante :

$$BHT2 : \begin{cases} \mathcal{H}_0^{\mathcal{R}} : L^*[n, k] < L_T^* \\ \mathcal{H}_1^{\mathcal{R}} : L^*[n, k] \geq L_T^* \end{cases} \quad (5.6)$$

avec  $L_T^*$  un seuil sur cette vraisemblance intégrale que nous devons définir. En particulier, comme tout au long de ce manuscrit nous souhaitons choisir le seuil avec un critère de Neyman-Pearson. Il est nécessaire d'estimer la densité de probabilité de  $L^*$  sous  $\mathcal{H}_0^{\mathcal{R}}$ .

Une expression analytique explicite de cette loi semble hors de notre portée. En particulier, n'ayant pas été en mesure de fournir une expression de la loi binomiale corrélée, une version intégrée de celle-ci semble encore plus difficile à obtenir. Nous nous servons donc une fois de plus des spectrogrammes assistants qui fournissent une solution numérique simple à ce problème. Le principe de la méthode est détaillé ci-dessous.



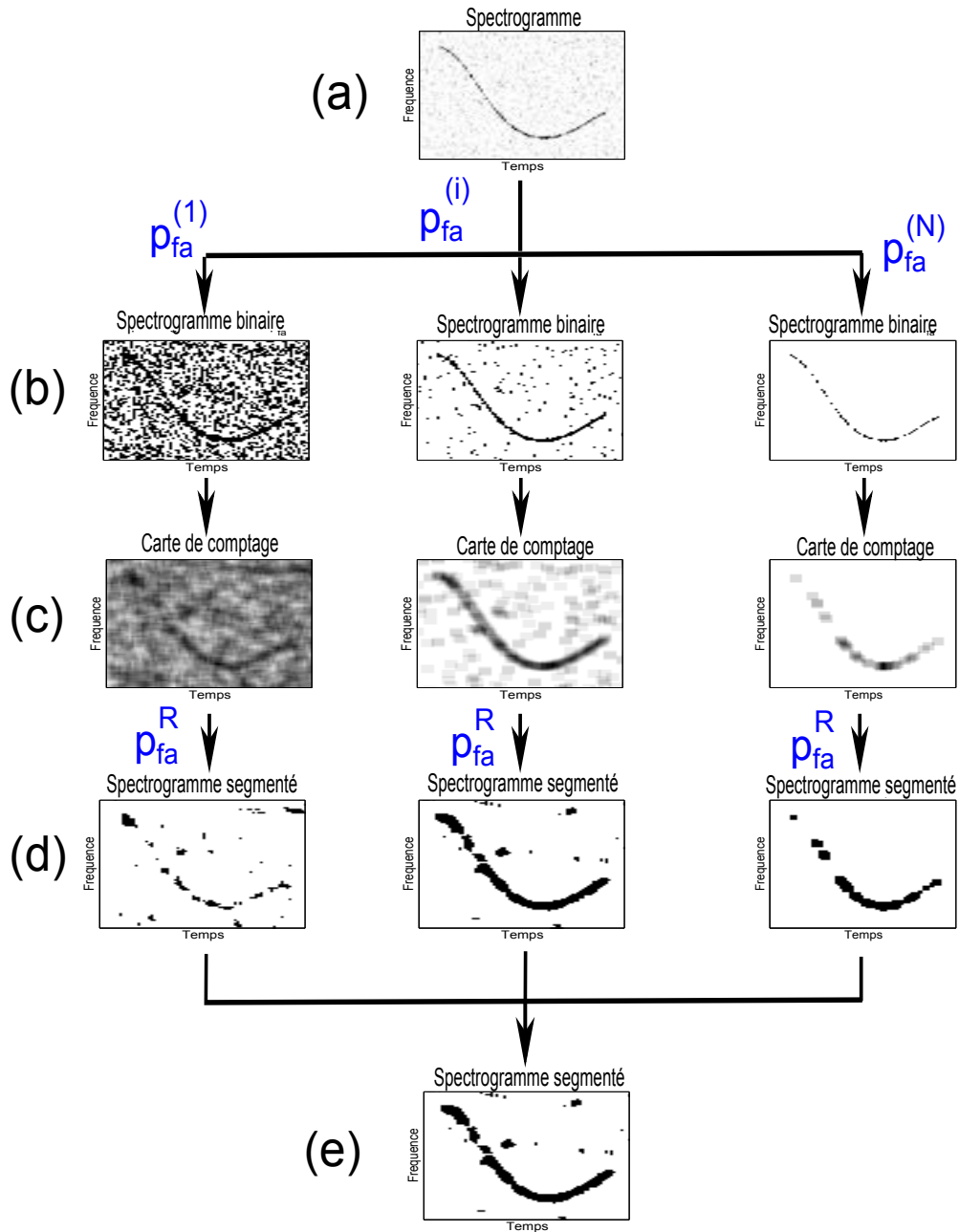


FIGURE 5.1 – Illustration schématique de la méthode FADA avec maximisation du nombre de détections sous contrainte de fausse alarme constante. a) le spectrogramme est généré et la DSP du bruit estimée en chaque point temps-fréquence  $[n,k]$ , b)  $N$  probabilités de fausses alarmes différentes sont utilisées pour générer  $N$  spectrogrammes binaires, c) les cartes de comptage sont calculées à partir de ces spectrogrammes binaires, d) les cartes de comptage sont seuillées à probabilité de fausse alarme constante  $p_{fa}^R$  et f) le spectrogramme segmenté final est celui qui maximise le nombre de détections.

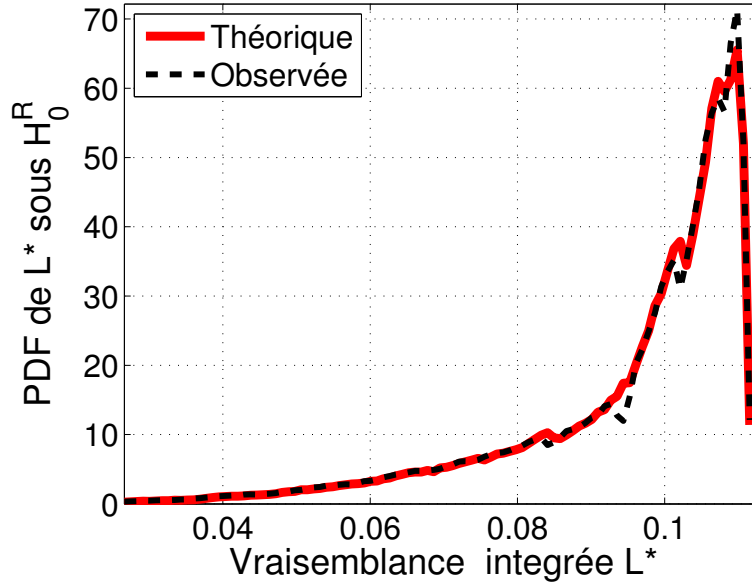


FIGURE 5.2 – Densité de probabilité de  $L^*$  sous  $\mathcal{H}_0^R$ . En rouge est tracée la version prédite par les spectrogrammes assistants. En pointillés noirs est représentée la densité de probabilité mesurée sur un enregistrement réel de bruit sous-marin tiré sur les données ERATO-09.

Méthode d'estimation de  $p_{\mathcal{H}_0^R}(L^*)$  par les spectrogrammes assistants :

- **Étape 1** : Génération d'un signal Gaussien centré blanc stationnaire de variance unité
- **Étape 2** : Calcul de son spectrogramme en utilisant les mêmes paramètres que ceux désirés pour l'analyse du signal (imitation de la structure de corrélation)
- **Étape 3** : Seuillage du spectrogramme pour chaque probabilité de fausse alarme dans  $\mathcal{P}_{FA} = \{p_{fa}^{(1)}, p_{fa}^{(2)}, \dots, p_{fa}^{(N)}\}$ .
- **Étape 4** : Génération des cartes de comptage  $\Psi = \{\psi_{p_{fa}^{(1)}}, \dots, \psi_{p_{fa}^{(N)}}\}$
- **Étape 5** : Mesure de la vraisemblance  $L_{p_{fa}^{(i)}}[n, k] = p_{\mathcal{H}_0}(\psi_{p_{fa}^{(i)}}[n, k])$  en chaque point  $[n, k]$  et pour chaque  $i \in \{1, \dots, N\}$ . Cette vraisemblance est connue (chapitre précédent).
- **Étape 6** : Estimation de l'intégrale  $L^*[n, k] = \int_0^1 L_{p_{fa}}[n, k] dp_{fa}$  en chaque point  $[n, k]$
- **Étape 7** : Estimation de la densité de probabilité  $p_{\mathcal{H}_0^R}(L^*)$  par l'histogramme.

Nous souhaitons analyser la pertinence de cette méthode d'estimation de  $p_{\mathcal{H}_0^{\mathcal{R}}}(L^*)$ . Pour cela, de manière identique au chapitre 4, nous avons choisi une minute d'enregistrement de bruit sous-marin (tiré de la campagne ERATO-09, SHOM) sur lequel nous avons visuellement vérifié qu'aucun son d'origine biologique, anthropique ou géologique majeur n'était présent. L'enregistrement est effectué avec une fréquence d'échantillonnage de 32 768Hz. Le spectrogramme a été calculé avec une fenêtre de Hann de 512 points, et un décalage de 256 points entre fenêtres d'analyse. Le voisinage d'estimation de la densité spectrale de puissance du bruit est  $V = 5s \times 192Hz$  (soit un voisinage  $V = 641 \times 3$  pixels) Ce jeu de paramètre est le même pour le signal synthétique et pour le signal réel. La région utilisée pour compter les détections est  $\mathcal{R} = 6 \times 6$  pixels. Nous avons alors tracé en figure 5.2 la densité de probabilité prédite par la méthode décrite ci-avant, en trait plein rouge. Sur cette courbe nous avons superposé, en pointillés noirs, la densité de probabilité de  $L^*$  sous  $\mathcal{H}_0^{\mathcal{R}}$  estimée à partir du bruit sous-marin réel. Les deux courbes se superposent convenablement.

Maintenant que nous connaissons la densité de probabilité de  $L^*$  sous  $\mathcal{H}_0^{\mathcal{R}}$  nous pouvons résoudre le test d'hypothèse binaire BHT2 en suivant une approche de Neyman-Pearson. Le seuil  $L_T^*$  est ainsi défini de la manière suivante :

$$L_T^* = \operatorname{argmax}_{L_T^*} \left( p_{\mathcal{H}_0^{\mathcal{R}}}(L^* \geq L_T^*) \right) \quad (5.7)$$

sous contrainte  $p_{\mathcal{H}_0^{\mathcal{R}}}(L^* \geq L_T^*) \leq p_{fa}^{\mathcal{R}}$

À noter que la condition d'égalité entre la probabilité de fausse alarme obtenue et celle désirée n'est pas nécessairement remplie.

La figure 5.3 illustre schématiquement le fonctionnement du détecteur que nous appelons FADA-intégrale. Tout d'abord le spectrogramme est seuillé pour N probabilités de fausse alarme différentes (figure 5.3.b). Les N cartes de comptage sont générées à partir de ces spectrogrammes binaires (figure 5.3.c). Nous calculons la vraisemblance  $L_{p_{fa}}[n, k]$  en chaque point. Nous avons donc N cartes de vraisemblance (figure 5.3.d) qu'il faut combiner en une seule. La solution retenue consiste à calculer l'intégrale de la vraisemblance par rapport à la  $p_{fa}$ . Une carte de vraisemblance intégrale est calculée. L'utilisation des spectrogrammes assistants permet d'avoir accès à la densité de probabilité de  $L^*$  sous  $\mathcal{H}_0^{\mathcal{R}}$ . La carte de vraisemblance intégrale est seuillé en utilisant un critère de Neyman-Pearson (figure 5.3.f)

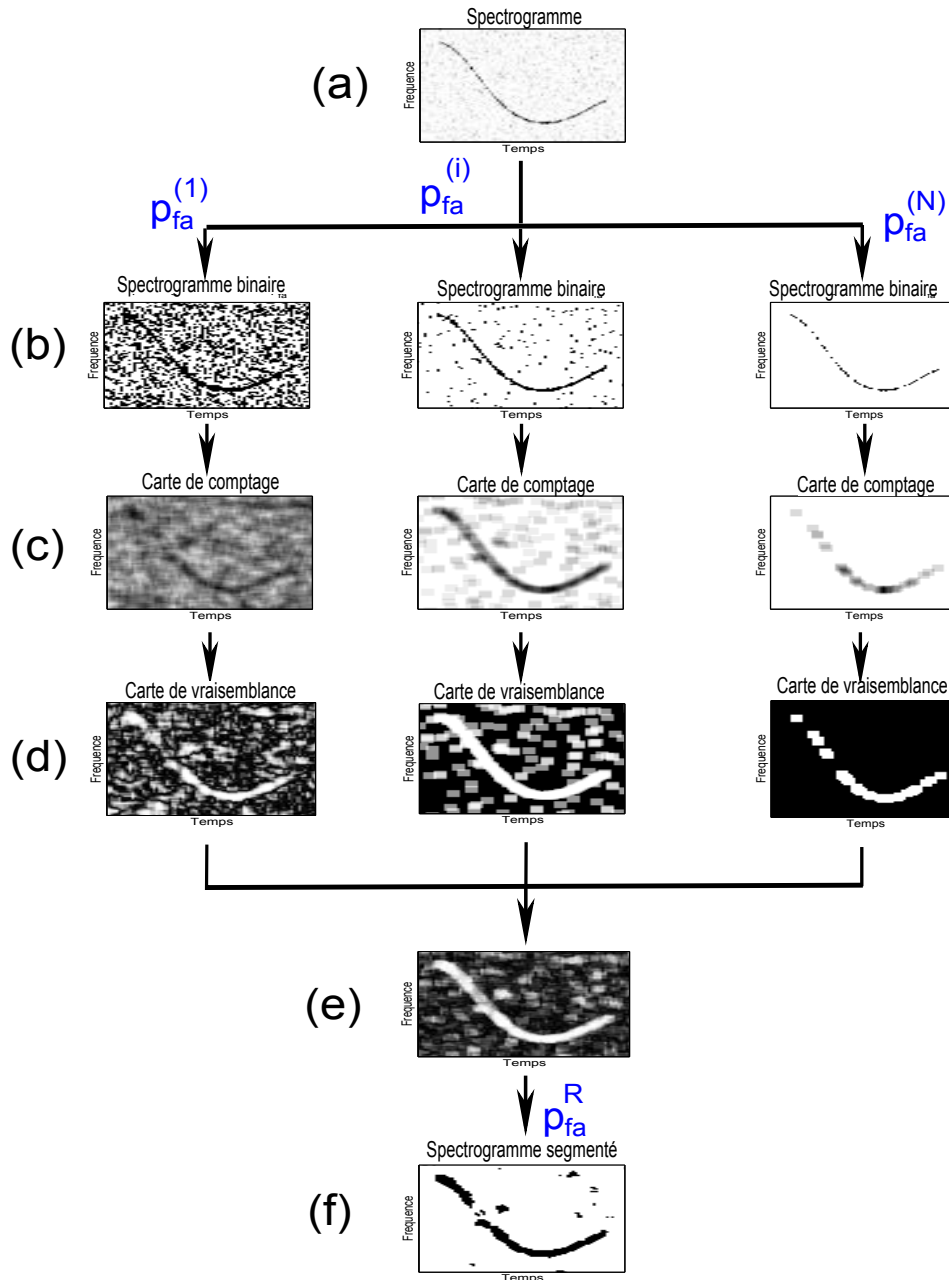


FIGURE 5.3 – Illustration schématique de la méthode FADA-intégrale. a) le spectrogramme est généré et la DSP du bruit estimée en chaque point temps-fréquence  $[n,k]$ , b)  $N$  probabilités de fausses alarmes différentes sont utilisées pour générer  $N$  spectrogrammes binaires, c) les cartes de comptage sont calculées à partir de ces spectrogrammes binaires, d) les cartes de vraisemblance sont calculées, i.e. la probabilité d'observer la valeur  $\psi[n, k]$  sous  $\mathcal{H}_0^R$  connaissant  $p_{fa}$  et  $\mathcal{R}$ , e) calcul de la carte de vraisemblance intégrale (ou alternativement de la Norme 1 ou 2) et f) Seuillage de la carte de vraisemblance intégrale par Neyman-Pearson.

$\xi_0$ (dB)	0	1	2	3	5	7	9	10	12
$p_{fa}$	0.28	0.26	0.24	0.22	0.16	0.1	0.05	0.03	0.006

TABLE 5.1 – Répartition des  $p_{fa}$  et leur équivalence en terme de RSB de focus

### 5.2.3 Norme du vecteur de vraisemblance

Nous définissons dans ce manuscrit une autre manière de résoudre ce test d'hypothèse, à l'aide de normes.

#### Norme 1

Nous définissons  $\delta_1$ , une norme équivalente à la norme 1, que nous définissons ainsi :

$$\delta_1 = \frac{1}{N} \|L_{\mathcal{P}_{FA}}[n, k]\|_1 = \frac{1}{N} \sum_{i=1}^N L_{p_{fa}^{(i)}}[n, k] \quad (5.8)$$

(Sans valeur absolue car toutes les réalisations de  $L_{p_{fa}^{(i)}}[n, k]$  sont comprises dans l'intervalle  $[0, 1]$ ). L'idée derrière le passage de la vraisemblance intégrale au calcul de la norme  $\delta_1$  provient du désir d'attribuer un "poids" différent aux différentes probabilités de fausse alarme. Nous voudrions par exemple être capable de répartir logarithmiquement les probabilités de fausse alarme que nous testons, et de leur attribuer le même poids, ce qui n'est pas le cas lors de l'intégration si les points de mesure de sont pas uniformément répartis. Pour rappel, le RSB de "focus" est défini de la manière suivante [HM06] :

$$p_{fa} = \exp(-(0,26\xi_0 + 1)) \quad (5.9)$$

Il correspond au RSB pour lequel le détecteur est réglé de manière optimale au sens du maximum de vraisemblance. Il peut paraître intéressant de mesurer la vraisemblance de l'hypothèse  $\mathcal{H}_0^R$  pour des RSB de focus  $\xi_0$  (en dB) répartis linéairement, la correspondance entre les deux grandeurs est donc la suivante :

Calculer la norme 1 revient donc à accorder la même importance à des  $p_{fa}$  réparties non-uniformément.

La méthode d'estimation de la densité de probabilité  $p_{\mathcal{H}_0^R}(\delta_1)$  est exactement la même que pour  $p_{\mathcal{H}_0^R}(L^*)$ . Nous utilisons les spectrogrammes assistants. La seule différence se situe à l'étape 6 : au lieu d'estimer l'intégrale de  $L_{p_{fa}}$  par rapport à  $p_{fa}$  il suffit de calculer  $\delta_1$

De même que pour la vraisemblance intégrale, nous comparons la densité de probabilité prédite par les spectrogrammes assistants avec la densité de probabilité observée sur données réelles. Les paramètres de construction du spectrogramme,

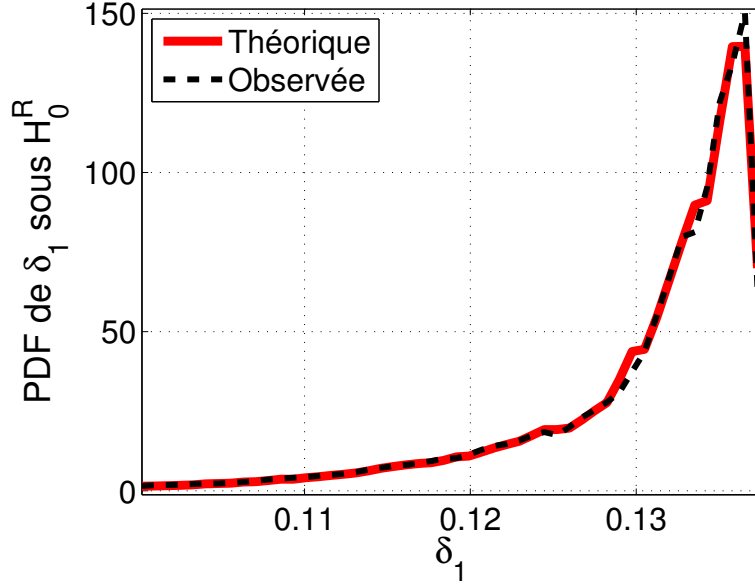


FIGURE 5.4 – Densité de probabilité de  $\delta_1$  sous  $\mathcal{H}_0^{\mathcal{R}}$ . En trait plein est tracée la version prédite par les spectrogrammes assistants. En pointillé est représentée la densité de probabilité mesurée sur un enregistrement réel de bruit sous-marin tiré sur les données ERATO-09.

d'estimation de la densité spectrale de puissance du bruit et de comptage des détections sont les mêmes.

### Norme euclidienne

De même nous définissons  $\delta_2$ , une norme équivalente à la norme euclidienne (ou norme 2), que nous définissons ainsi :

$$\delta_2 = \frac{1}{N} \|L_{\mathcal{P}_{FA}}[n, k]\|_2 = \frac{1}{N} \sqrt{\sum_{i=1}^N L_{p_{fa}}^{(i)}[n, k]^2} \quad (5.10)$$

Comme pour la norme 1 nous accordons un poids égal à toutes les probabilités de fausse alarmes, et ce même si elles ne sont pas uniformément réparties. La norme euclidienne est potentiellement un choix viable pour la raison suivante : nous cherchons à trouver une grandeur qui sépare le mieux possible les distributions sous  $\mathcal{H}_0^{\mathcal{R}}$  et sous  $\mathcal{H}_1^{\mathcal{R}}$ . Prendre la valeur de la vraisemblance au carré va ainsi pénaliser fortement les valeurs proches de 0 par rapport à celles proches de 1.

De la même manière que pour la norme 1. Nous comparons la densité de probabilité  $p_{\mathcal{H}_0^{\mathcal{R}}}(\delta_2)$  prédite par les spectrogrammes assistants et celle obtenue par les données réelles.

**Remarque :** Les valeurs prises par  $L_{p_{fa}}^{(i)}[n, k]$  sont toutes comprises entre 0 et 1. Les valeurs prises par les normes  $\| \cdot \|_1$  et  $\| \cdot \|_2$  sont donc toujours inférieures

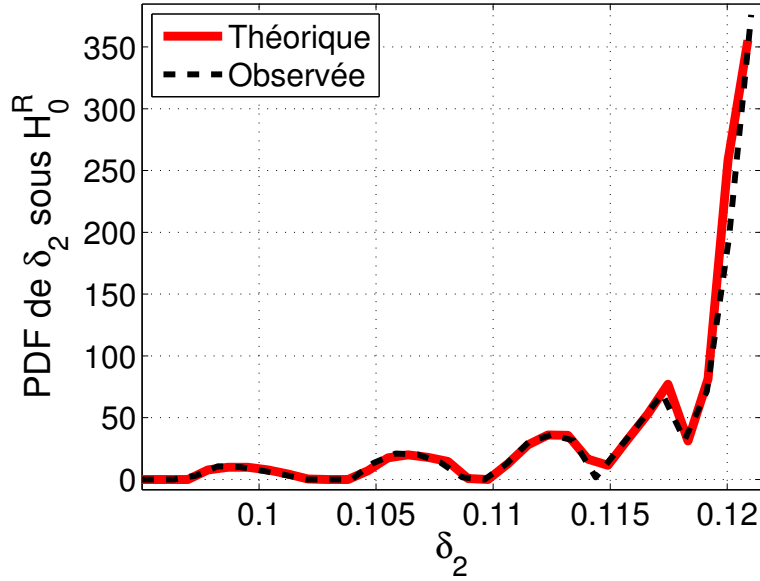


FIGURE 5.5 – Densité de probabilité de  $\delta_2$  sous  $\mathcal{H}_0^{\mathcal{R}}$ . En trait plein est tracée la version prédite par les spectrogrammes assistants. En pointillé est représentée la densité de probabilité mesurée sur un enregistrement réel de bruit sous-marin tiré sur les données ERATO-09.

à 1. Nous sommes en fait en train de mesurer la distance entre la matrice de vraisemblance et une hypersphère de rayon 1.

### 5.3 Forme de la région de comptage $\mathcal{R}$

Le choix de  $\mathcal{R}$  (quatrième degré de liberté dans la liste) peut, et doit être adapté aux signaux que l'on cherche à détecter et aux objectifs de l'étude. Nous donnons ici des critères de choix de cette région de comptage pour la détection de sifflements, et pour la détection de clics.

#### 5.3.1 Pour la détection de sifflements

Nous nous posons ici la question du choix de la forme de la région de comptage  $\mathcal{R}$  dans le cas où nous cherchons à détecter un sifflement inconnu. Nous définissons ici le sifflement comme une modulation de fréquence continue dans le temps. Cette définition a des implications pour la forme du support temps-fréquence attendu, et en particulier sur la longueur temporelle et fréquentielle que doit avoir la région de comptage :

**Longueur fréquentielle :** La représentation du spectrogramme d'un signal composé d'une fréquence pure n'est pas idéalement concentrée au sens du principe d'incertitude, c'est à dire que le pavage du spectrogramme possède une certaine

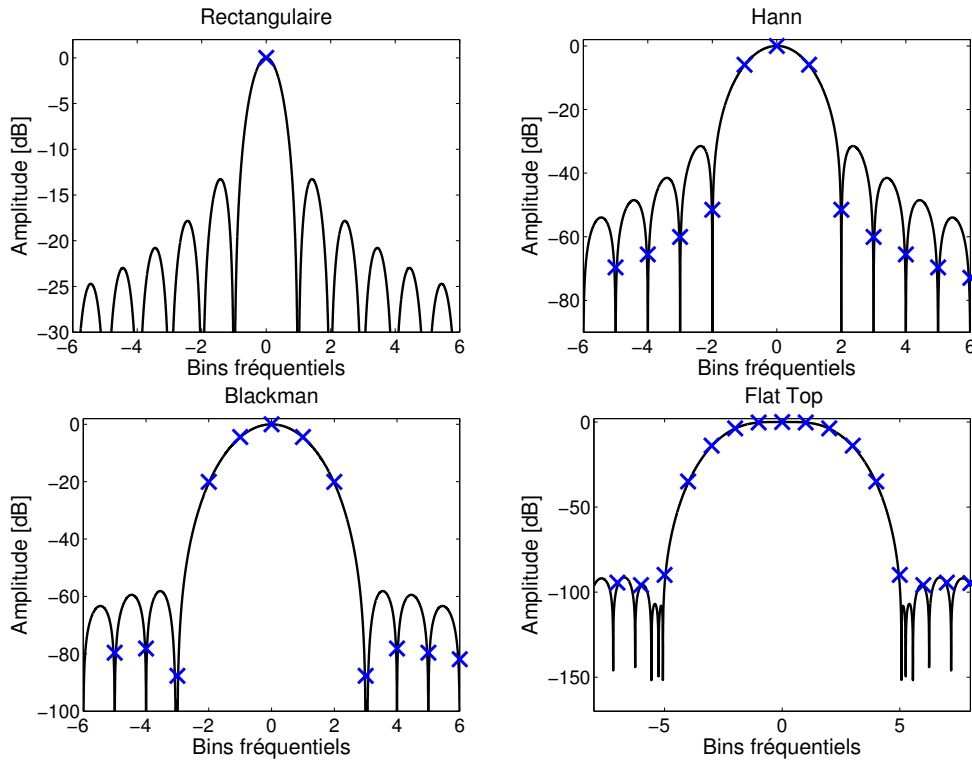
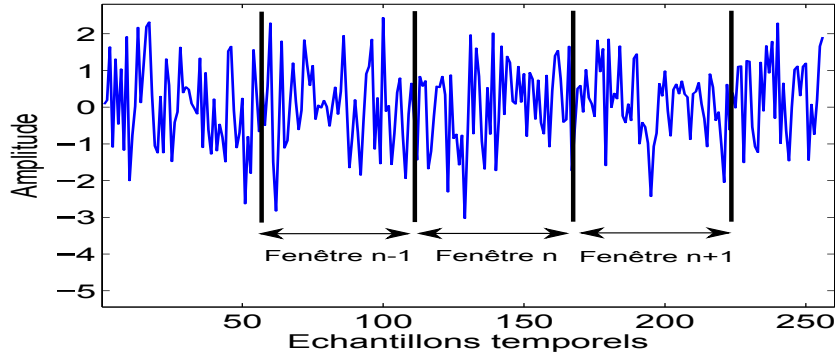


FIGURE 5.6 – Transformée de Fourier discrètes de plusieurs fenêtres de pondération : Rectangulaire, Hann, Blackman-harris et Flat Top. Les croix marquent les valeurs calculées à chaque bin fréquentiel, sans zéro-padding.

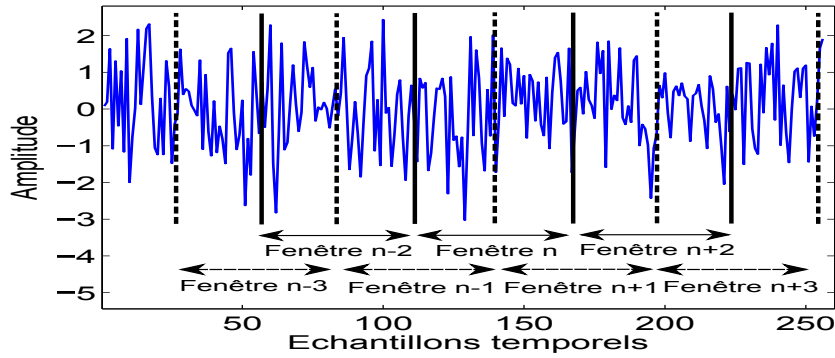
"taille" dans le plan temps-fréquence [Fla98, SDT13]. Mais elle n'est pas idéalement concentrée non plus dans les différentes "cases" du pavage. Son énergie sera dispersée sur l'ensemble de la bande de fréquence, et en particulier sur le premier lobe, qui contiendra la plus grande partie de l'énergie du signal. Le nombre de bins fréquentiels dont l'énergie provient du premier lobe est dépendant de la forme de la fenêtre de pondération (voir figure 5.6). Si nous cherchons à détecter les bins contenant de l'énergie due au premier lobe de la fréquence pure, cela revient ainsi à détecter 1 seul pixel dans le cas d'une fenêtre rectangulaire, 3 pixels dans le cas d'une fenêtre de Hann, 5 pixels dans le cas d'une fenêtre de Blackman-harris et 9 pixels dans le cas d'une fenêtre "Flat top" (en omettant l'énergie des deux pixels de bordure qui recevront 100dB de moins que le pixel central). Pour notre problème, cela signifie que la région de comptage  $\mathcal{R}$  **doit être fréquentiellement au moins aussi longue que le nombre de pixels appartenant au premier lobe.**

**Longueur temporelle :** calculer la TFCT revient à découper le signal en tronçons (voir figure 5.7), puis à pondérer (ou non) le signal et enfin à calculer la transformée de Fourier discrète, puis à en prendre le module carré. L'hypothèse de continuité de la fréquence que nous avons fait revient à dire que si la fréquence existe sur une fenêtre  $n$ , alors elle existe aussi sur la fenêtre  $n - 1$  et  $n + 1$ . Sans





(a) Sans recouvrement



(b) Avec 50% de recouvrement

FIGURE 5.7 – Fenêtrage du signal avec et sans recouvrement.

recouvrement (voir figure 5.7a), les détections peuvent alors être comptées sur les pixels temps fréquence indexés par  $n - 1$ ,  $n$  et  $n + 1$ . S'il y a recouvrement, par exemple de 50% (voir figure 5.7b), alors les fenêtres indexées dans la gamme  $\{n - 3, n - 2, \dots, n + 3\}$  reçoivent de l'énergie de la part du signal. Nous faisons le choix de ne pas considérer les fenêtres  $n - 3$  et  $n + 3$  qui n'abritent, au mieux que 50% de présence de signal dans leur bornes. Nous proposons alors **la longueur temporelle par défaut suivante** :

$$T_R = 1 + \frac{2}{1 - \frac{D}{M}}, \quad (5.11)$$

avec  $M$  la longueur de la fenêtre et  $D$  la pas de glissement entre deux fenêtres.  $D/M$  et donc le facteur de recouvrement du spectrogramme. Le facteur  $\frac{1}{1 - D/M}$  compte le nombre de fenêtres qu'il y a avant de recouvrir complètement la première fenêtre complètement disjointe. Le facteur 2 indique juste que l'on parcourt la fenêtre dans les indices de avant  $n$  et après  $n$ . L'ajout de 1 est là pour prendre en compte le pixel central. Voici un tableau récapitulant la longueur temporel  $T$  de la région de comptage en fonction du facteur de recouvrement :

D/M	0	0.5	0.75	0.8	0.9	0.95
$T_R$	1	5	9	11	21	41

(5.12)

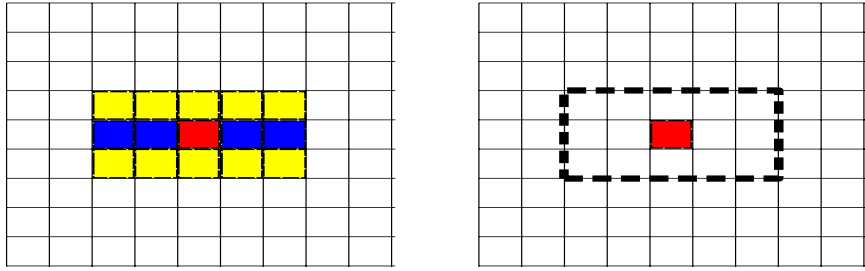
Dans le cas stationnaire la région de comptage recommandée est donc de forme rectangulaire et de taille  $T_R \times F_R$ , avec  $T_R$  la longueur temporelle de la région  $\mathcal{R}$  et  $F_R$  la longueur fréquentielle.

**(Non-)stationnarité du signal :** la forme de la région que nous venons de décrire est adaptée à un signal stationnaire. Nous montrons en figure 5.8a la forme d'une partie de  $\sqsupset$  pour ce signal stationnaire. En couleur nous montrons la raison de l'appartenance à  $\sqsupset$ . Est colorié en rouge le pixel d'intérêt, celui pour lequel nous sommes amenés à une prendre une décision. Sont coloriés en bleu les pixels faisant partie de  $\sqsupset$  à cause du recouvrement et de l'hypothèse liée à la modulation de fréquence continue. Sont coloriés en jaune les pixels qui appartiennent  $\sqsupset$  car ils font partie du lobe principal d'un pixel appartenant déjà à  $\sqsupset$ . Dans la colonne de droite de cette même figure est tracée la région  $\mathcal{R}$  qui découle de cette analyse.

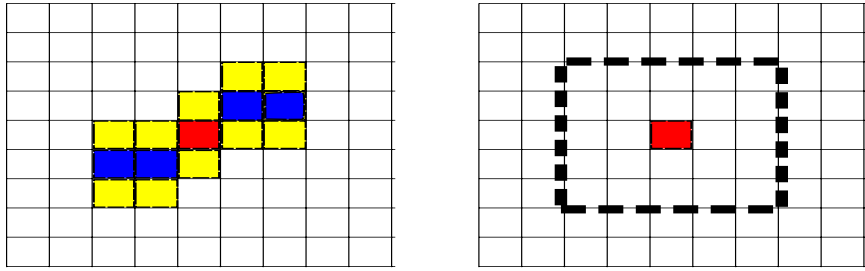
L'intérêt du spectrogramme réside cependant dans sa capacité à traiter des signaux non-stationnaires. Nous montrons l'exemple schématique de trois signaux présentant des degrés de non-stationnarité divers : peu non-stationnaire (figure 5.8b), moyennement non-stationnaire (figure 5.8c) et très non-stationnaire (figure 5.8d). Cette non-stationnarité se traduit par une montée très rapide de la fréquence, la région de comptage doit donc s'adapter à ces non-stationnarités en s'étirant sur l'axe vertical. Dans le cas de l'analyse du spectrogramme d'un signal stationnaire construit avec une fenêtre de Hann et un recouvrement de 50% la région de comptage  $\mathcal{R}$  est rectangulaire de taille  $T_R \times F_R = 5 \times 3$ . Elle sera de taille  $5 \times 5$  dans le cas d'un signal peu non-stationnaire, de taille  $5 \times 7$  dans le cas d'un signal moyennement non-stationnaire, de taille  $5 \times 9$  pour un signal fortement non-stationnaire.

**Récapitulatif :** nous récapitulons sous forme de tableau (Table 5.2) la forme des régions recommandées pour plusieurs fenêtres de pondération, plusieurs facteurs de recouvrement et plusieurs degrés de non-stationnarité. Les tailles données sont évidemment sujettes à caution pour plusieurs raisons :

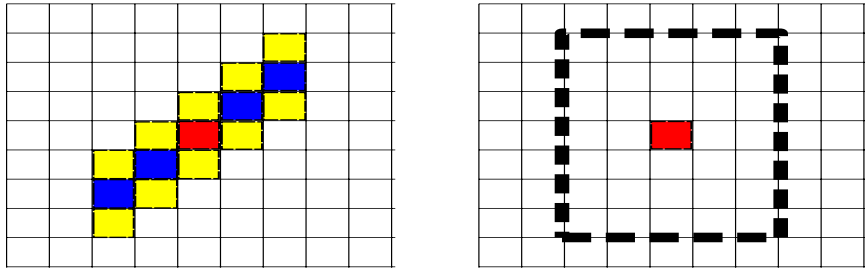
1. La longueur temporelle  $T_R$  de la région peut être adaptée au degré de non-stationnarité et à la durée d'existence du signal. Par exemple, pour un signal très stationnaire ayant une longue durée d'existence, la région de comptage peut être allongée temporellement.
2. Les notions "peu non-stationnaire", "moyennement non-stationnaire" et "très non-stationnaire" sont définies de manière très qualitative. Nous n'avons pas



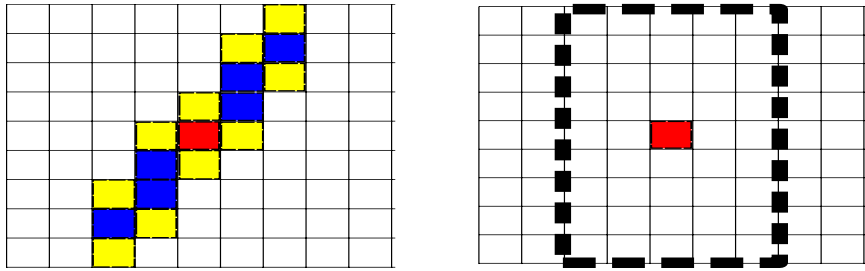
(a) Signal stationnaire



(b) Signal peu non stationnaire



(c) Signal moyennement non-stationnaire



(d) Signal très non-stationnaire

FIGURE 5.8 – Représentation du support temps-fréquence  $\mathcal{R}$  du signal et de la région de comptage  $\mathcal{R}$  déduite. La colonne de gauche montre l'exemple de support temps-fréquence de signaux, en fonction du degré de non-stationnarité observé. Les pixels colorés font partie de  $\mathcal{R}$ , et leur couleur indique la raison de leur appartenance. Le pixel en rouge est le pixel d'intérêt pour lequel la décision est prise. Les pixels en bleu sont les pixels adjacents temporellement qui par hypothèse font aussi partie de  $\mathcal{R}$ . Les pixels jaunes font partie du lobe principal du pixel bleu (ou rouge) présent dans leur fenêtre temporelle.

défini de manière rigoureuse ces notions de degré de (non-)stationnarité, et nous ne le ferons pas dans ce manuscrit (des pistes de réflexion sur ces questions pourront par exemple être trouvées dans [BFH<sup>+</sup>10])

3. Les tailles des régions  $\mathcal{R}$  données sont des ordres de grandeur informatifs. En pratique, lorsque nous sommes confrontés à une grande dynamique de stationnarité des signaux, nous conseillons de choisir les tailles données par les catégories "peu non-stationnaires" et "moyennement non-stationnaire".
4. Des  $\mathcal{R}$  plus grandes peuvent être choisies, mais le résultat de segmentation final sera évidemment plus grossier.

	Hann			Blackman-harris		
Recouvrement	0.5	0.75	.9	0.5	0.75	.9
Stationnaire	5 × 3	9 × 3	21 × 3	5 × 5	9 × 5	21 × 5
Peu non stat.	5 × 5	9 × 5	21 × 5	5 × 7	9 × 7	21 × 7
Moy. non stat.	5 × 7	9 × 7	21 × 7	5 × 9	9 × 9	21 × 9
Très non stat.	5 × 9	9 × 9	21 × 9	5 × 11	9 × 11	21 × 11

TABLE 5.2 – Tailles des régions de comptage  $\mathcal{R}$  en fonction de la fenêtre de pondération et du recouvrement. Les tailles sont données dans l'ordre  $T_R \times F_R$ . (abréviations : stat. = stationnarité, moy. = moyennement)

### 5.3.2 Pour la détection de clics

Avant de décrire la forme de la région de comptage pour la détection de clics, nous avons une remarque fondamentale à formuler ici : **le spectrogramme ne peut pas être adapté simultanément aux clics et aux sifflements.**

Il suffit d'observer les figures 5.9a et 5.9b pour s'en convaincre. La figure 5.9a est un zoom sur 25ms d'un spectrogramme calculé à partir d'un enregistrement d'odontocètes. Cet enregistrement est accessible sur la base de données MobySound<sup>1</sup>. Le signal est échantillonné à une fréquence de 192kHz. Le spectrogramme est construit avec une fenêtre de Hann de 2048 points, c'est à dire une fenêtre d'une durée de 10ms, ce qui semble être un compromis assez bien adapté aux sifflements

1. Ces enregistrements font partie de la base de données utilisée lors de la 5ème "Workshop on Detection, Classification, Localization and Density Estimation of Marine Mammals using Passive Acoustics". Tous ces enregistrements sont librement téléchargeables sur le site de MobySound (<http://www.mobysound.org/>)

observés dans cet enregistrement. Les sifflements sont en effet assez bien localisés en fréquence.

Cette longueur de fenêtre n'est cependant pas adaptée aux clics. Ils ont une forte amplitude mais leur durée d'existence est très brève, inférieure à la milliseconde. Cependant dans ce cas de figure, leur énergie est intégrée sur une dizaine de milli-secondes. La conséquence est l'étalement de cette énergie, et une très mauvaise localisation temporelle du clic en question. A l'opposé, la figure 5.9b montre un spectrogramme construit avec une fenêtre de Hann de 128 points, soit une durée de 0.6ms. Dans ce cas-ci, il n'est plus possible de distinguer les trois sifflements présents dans le spectrogramme, leur énergie étant trop mal localisée fréquentiellement. Le clic au contraire est très bien localisé.

Nous ne faisons ici qu'observer le compromis très classique sur la forme du pavage du plan temps-fréquence. La leçon à tirer de cette analyse est que le spectrogramme ne pourra jamais être simultanément adapté aux deux grands types de signaux émis par les mammifères marins : harmoniques et impulsionnels. Nous fermons ici cette parenthèse, avant de la réouvrir dans une prochaine section pour parler du problème de la séparation des sifflements et des clics.

Le spectrogramme est donc construit avec une fenêtre de pondération très courte. Le choix de la forme de la région  $\mathcal{R}$  est bien plus simple à faire que dans le cas des sifflements. La raison étant que nous avons un *a priori* très fort sur la forme du signal recherché. Nous recherchons un signal de très courte durée couvrant l'ensemble (ou une grande partie) de la bande fréquentielle. La région de comptage  $\mathcal{R}$  la mieux adaptée dans ce cas est donc naturellement verticale et couvre l'ensemble de la bande de fréquence.

La figure 5.21 montre un exemple de détection de clics (pour un enregistrement issu de MobySound). Y est représenté le spectrogramme du signal, la forme d'onde temporelle des clics après détection et un zoom sur la forme d'onde détectée.

Nous résumons l'ensemble des paramètres de la méthode pour la détection de sifflements et de clics dans la table 5.3.

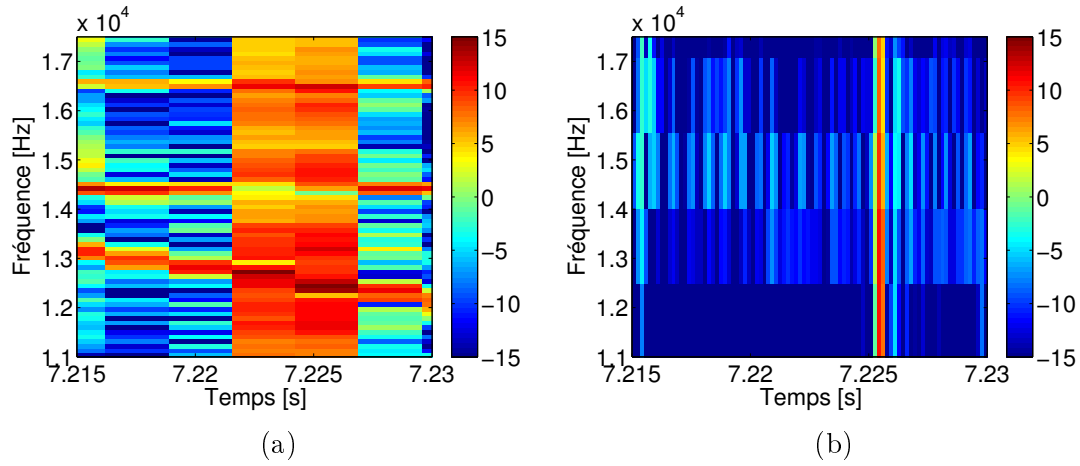


FIGURE 5.9 – (a) Spectrogramme adapté aux sifflements : fenêtre de 2048 points, soit 10ms ; (b) Spectrogramme adapté aux clics : fenêtre de 128 points, soit 0.6ms.

	Paramètres	Commentaires
Longueur de la fenêtre de pondération	$L$	À adapter au signal
Incrément temporel entre fenêtres d'analyse	$D$	Au choix (défaut : recouvrement $D/L = 0.5$ )
Longueur FFT	$M$	Au choix (défaut : facteur de zéro-padding $M/L = 1$ )
Voisinage d'estimation DSP du bruit	$V$	À adapter à la stationnarité et à la blancheur du bruit
Forme de $\mathcal{R}_A$ (sifflements)	$T_R = 1 + \frac{2}{1 - \frac{D}{M}}$ , $F_R$	Le choix de $F_R$ est discuté précédemment.
Forme de $\mathcal{R}_A$ (clics)	$T_R = 1, F_R = M$	
Probabilité de fausse alarme	$p_{fa}^{\mathcal{R}}$	Au choix (défaut : 0.01)

TABLE 5.3 – Récapitulatif sur le choix des paramètres.

## 5.4 Évaluation de la méthode

Dans cette section nous analysons numériquement les performances des différents détecteurs proposés :

- Le détecteur FADA avec maximisation du nombre de détections sous contrainte
- Le détecteur FADA avec vraisemblance intégrale

- Le détecteur FADA avec norme 1
- Le détecteur FADA avec norme euclidienne

Tout au long de cette analyse, nous nous servons du détecteur quadratique, basé sur la loi du  $\chi^2$ , comme référence.

**Une remarque préliminaire s'impose :** les simulations numériques ne concerneront que le cas de la détection de sifflements inconnus. Nous faisons le choix de ne pas analyser les performances du détecteur FADA pour la détection de sons impulsionsnels. Nous pensons en effet que la détection de clics gagnerait à être faite dans un domaine plus adapté : par exemple le domaine temporel, en utilisant l'énergie du signal et/ou ses statistiques d'ordres supérieures du signal [GBB<sup>+</sup>10], par des méthodes vectorielles [BCI<sup>+</sup>12], par des méthodes temps-échelle ou par des méthodes dédiées à l'analyse d'impulsions rythmées [LBGBM13]. Le cadre théorique FADA que nous avons mis en place est cependant suffisamment flexible pour permettre de traiter les signaux impulsionsnels - nous le montrerons sur de nombreux exemples sur signaux réels dans la prochaine section - mais nous éviterons de revendiquer une quelconque optimalité.

Nous utilisons les Caractéristiques Opérationnelles de Réception<sup>2</sup> (courbe COR) comme méthode de comparaison des différents détecteurs [Faw06]. Pour chaque RSB nous traçons l'ensemble des points de fonctionnement  $(p_{fa}, p_d)$  accessibles au détecteur.

Nous mesurons les performances pour des sifflements synthétiques. Ces sifflements sont modélisés par des signaux de la forme suivante :

$$s[n] = \sin(2\pi n f[n]) + b[n] \quad (5.13)$$

avec

$$f[n] = \sum_{i=1}^3 \frac{1}{2} A_i \left( 1 + \sin(2\pi f_i n + phase_i) \right) \quad (5.14)$$

La fréquence d'échantillonnage est supposée égale à 32kHz. Les amplitudes  $A_i$  sont choisies aléatoirement entre 0 et  $f_s/6$ . Les phases sont tirées aléatoirement entre 0 et  $2\pi$ , les fréquences  $f_i$  sont tirées aléatoirement entre 0 et 4 Hz, ce qui donne une pente absolue maximale atteignable de  $8 \times f_s/2 = 131.000$  Hz/s. Ce qui semble largement supérieur en terme de pente à ce que les mammifères marins peuvent émettre dans le cas de sifflements. Nous montrons en figure 5.10 quatre tirages aléatoires de fréquence instantanées.

---

2. Anglais : Receiving Operating Curves (ROC)

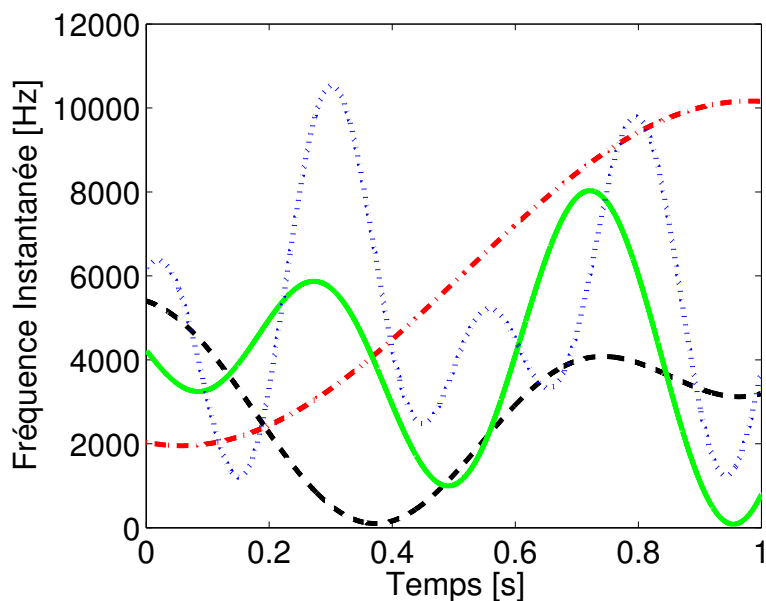


FIGURE 5.10 – Exemple de quatre lois de fréquences instantanées utilisées pour la mesure des performances de FADA

La détection FADA nécessite le choix d'un vecteur de probabilité de fausse alarme, que ce soit pour maximiser le nombre de détection sous contrainte, ou pour calculer la vraisemblance intégrale, la norme 1 ou la norme 2. Nous choisissons un vecteur de 26 probabilités de fausses alarmes. Ces probabilités sont choisis de manière à ce que les RSB de focus correspondants soient linéairement répartis entre 0 et 12dB (voir l'éq. 5.9 et la table 5.1).

Les RSB donnés par la suite pour le tracé des courbes COR correspond à un RSB moyen du signal. Les signaux synthétiques générés sont très non-stationnaires. Le RSB local du signal est donc en général très différent si l'on est dans une zone presque stationnaire, ou dans un zone très non-stationnaire. Nous avons généré 250000 signaux aléatoires pour chaque RSB.

La figure 5.11 montre sur un même graphique les courbes COR des quatre détecteurs FADA (maximisation sous contrainte, vraisemblance intégrale, Norme 1 et Norme 2) ainsi que la courbe COR du détecteur quadratique, pour un signal de RSB moyen égal à 3dB.

Tout d'abord nous remarquons que toutes les courbes COR données par la méthode FADA sont très nettement au dessus de celle du détecteur quadratique.

La seconde remarque que nous formulons concerne les performances obtenues pour les détecteurs FADA-intégrale, FADA-Norme1 et FADA-Norme2. Nous le voyons sur cette figure, mais nous l'avons également vérifié pour de multiples RSB, les performances de ces trois méthodes sont très similaires. Nous ne le montrons pas



ici pour ne pas surcharger ce manuscrit. Ce résultat est surprenant et nous pensons qu'une étude théorique plus poussée permettrait d'en comprendre la raison, et peut être de mieux choisir le vecteur de  $p_{fa}$ .

La troisième remarque concerne le comportement erratique de courbe COR obtenue par la méthode FADA de maximisation du nombre de détections sous contrainte. Dans le cas de la maximisation sous contrainte, nous garantissons que la probabilité de fausse alarme observée est plus faible que celle désirée, mais en aucun cas nous ne garantissons qu'elle soit proche. Nous montrons en exemple les  $p_{fa}^{\mathcal{R}}$  accessibles en fonction du choix de  $p_{fa}$  dans le tableau suivant :

	$p_{fa}^{\mathcal{R}}$ accessibles						
$p_{fa} = 0.1$	0.86	0.66	0.44	0.26	0.13	0.06	0.03
	0.01	0.004	0.001	0.0004	0.0001	0	
$p_{fa} = 0.05$	0.60	0.33	0.14	0.053	0.018	0.0059	0.0014
	0.0002	0.0001	0				
$p_{fa} = 0.01$	0.25	0.068	0.013	0.0023	0.0004	0.0001	0

Le choix de  $p_{fa}$  optimal (au sens de la maximisation sous contrainte) est inconnu à priori. Le choix de  $p_{fa}^{\mathcal{R}}$  que nous faisons *a priori* n'est ainsi par nécessairement adapté. Il suffit de voir le cas où l'on choisit  $p_{fa}^{\mathcal{R}} = 0.01$  et où il se trouve que la  $p_{fa}$  optimale est  $p_{fa} = 0.01$ . Notre  $p_{fa}^{\mathcal{R}}$  objectif est 0.01, mais celle que nous obtenons effectivement est 0.0023, soit environ cinq fois moins.

Dans la suite de ce manuscrit, tous les exemples, et résultats que nous donnerons seront obtenus à partir de FADA-intégrale uniquement (y compris les courbes COR puisqu'elles sont quasiment identiques pour la norme 1 et 2).

La figure 5.12 montre l'exemple de 4 courbes COR obtenues pour quatre RSB différents (1dB, 3dB, 7dB et 12dB). La méthode FADA-intégrale montre de meilleures performances en terme de compromis probabilité de fausse alarme/probabilité de détection que le détecteur quadratique pour ces quatre RSB.

Nous souhaitons pouvoir comparer les détecteurs sur une large gamme de RSB. Cette comparaison est difficile en utilisant uniquement les courbes COR. Il faudrait soit multiplier le nombre de figures, soit tracer toutes les courbes (détecteur quadratique et FADA-intégrale) sur le même graphique, le rendant alors illisible.

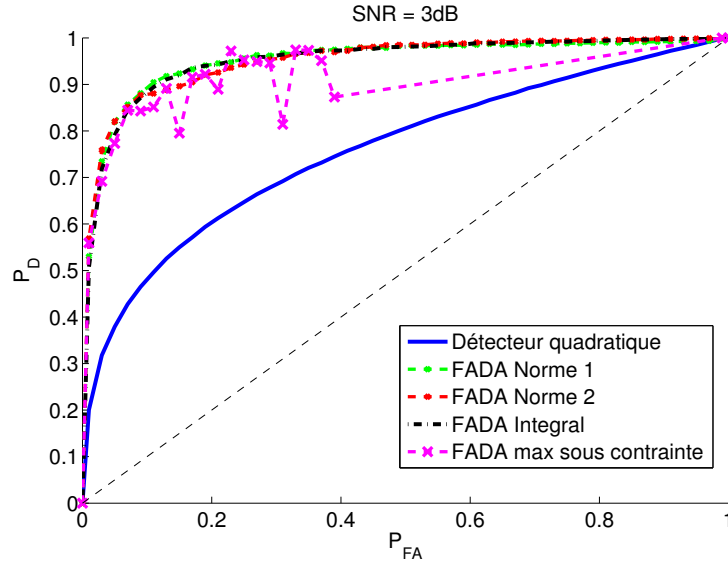


FIGURE 5.11 – Courbe COR pour les quatre méthodes FADA (maximisation sous contrainte, vraisemblance intégrale, Norme 1 et Norme 2), ainsi que pour le détecteur quadratique.

Nous proposons d'utiliser l'aire sous la courbe COR, notée AUC (pour Area Under the Curve) [Bra97] :

$$AUC(RSB) = \int_0^1 p_d(p_{fa}, RSB) dp_{fa} \quad (5.15)$$

Un détecteur performant doit ainsi avoir une AUC proche de 1, c'est à dire  $p_d = 1$  quelle que soit la  $p_{fa}$ . Un détecteur aléatoire qui se situe au niveau de la ligne de chance a une AUC proche de 0.5. Cette grandeur a l'avantage de pouvoir représenter simplement le comportement global du détecteur. Nous perdons cependant une partie de l'information, nous n'avons ainsi plus accès aux points de fonctionnement  $(p_{fa}, p_d)$  des détecteurs. Deux détecteurs peuvent ainsi avoir une même AUC mais deux courbes COR différentes [Faw06].

Nous représentons ainsi sur la figure 5.14 les courbes AUC obtenues pour les deux détecteurs. La courbe AUC correspondant à FADA-intégrale est très nettement au dessus de celle du détecteur quadratique et ce pour tous les RSB.

Il existe deux raisons à ces très bonnes performances. Tout d'abord, la nature régionale du détecteur nous permet de discriminer très efficacement les fausses alarmes des vraies détections, et donc de mettre à 0 beaucoup de fausses alarmes. La seconde raison est la capacité de la méthode à "rattraper" une partie des détections manquées, c'est à dire attribuer la valeur 1 à un point temps-fréquence auquel on avait affecté par erreur la valeur 0. Prenons l'exemple suivant : un point temps-fréquence  $[n, k]$  possède un RSB local très faible, afin d'être détecté

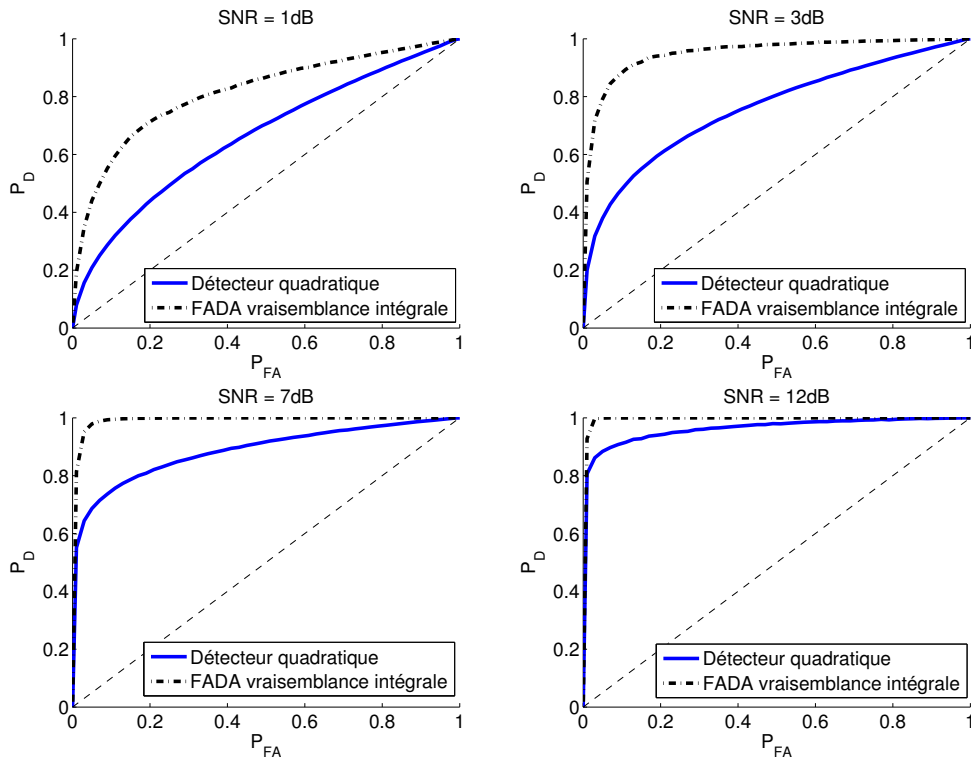


FIGURE 5.12 – Comparaison des courbes COR du détecteur quadratique (trait plein) et de FADA-intégrale (tirets) pour quatre RSB différents : 1dB, 3dB, 7dB et 12dB

par le détecteur quadratique il faudra considérablement augmenter la probabilité de fausse alarme. Dans le cas FADA, il suffit que les voisins de ce point  $[n, k]$  aient un RSB plus élevé pour pouvoir le détecter sans augmenter démesurément la probabilité de fausse alarme. Un schéma explicite ce point en figure 5.13. Nous supposons que ces spectrogrammes binaires schématiques sont issus d'un seuillage avec une probabilité de fausse alarme identique. Il semble évident que le point d'étude sur le spectrogramme de gauche sera mis à 1, "rattrapant" donc la détection manquée, sans avoir besoin d'augmenter  $p_{fa}$  comme cela est nécessaire avec le détecteur quadratique. À l'inverse, le point d'étude dans le spectrogramme binaire de droite sera mis à 0, corrigeant donc la fausse alarme.

La prise en compte d'une information de voisinage temps-fréquence dans le processus de détection permet au détecteur FADA de fournir des performances supérieures au détecteur énergétique simple en terme de probabilité de fausse alarme/probabilité de détection. L'information de proximité des pics temps-fréquence nous permet à la fois de discriminer efficacement les régions "signal" des régions "bruit seul" (et donc de mettre à zéro une partie des fausses alarmes), mais aussi de corriger une partie des détections manquées. Par rapport à l'état de l'art (présenté au chapitre 3) sur les différentes manières de prendre en compte l'information de voisinage temps-fréquence nous conservons avec FADA la formulation en test d'hy-

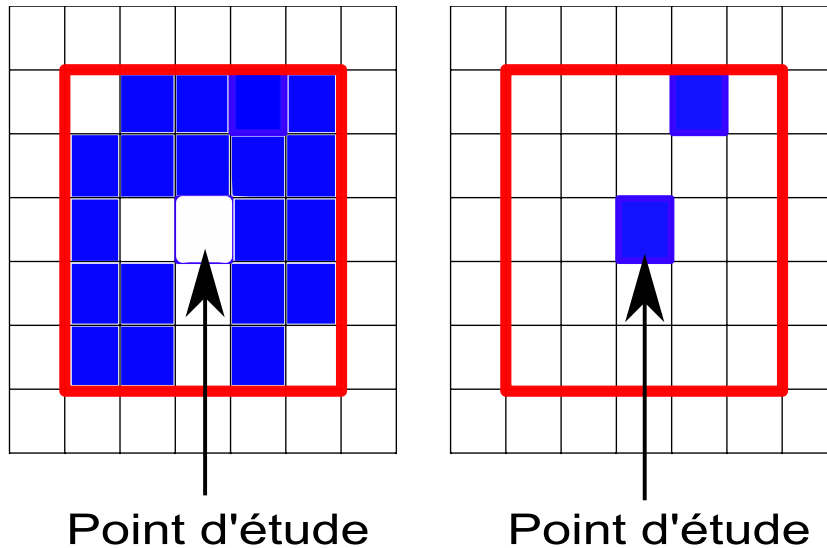


FIGURE 5.13 – Deux spectrogrammes binaires hypothétiques seuillés avec une même probabilité de fausse alarme. À gauche le point d'étude est faussement mis à 0, mais tous un grand nombre de voisins sont à 1. Il sera donc très probablement mis à 1, on corrige les non détections. À droite nous observons la situation opposée. Le point d'étude est faussement mis à 1, mais un grand nombre de ses voisins est à 0. Il sera probablement mis à 0, on corrige les fausses alarmes.

pothèse binaire et la possibilité d'être résolu avec un critère de Neyman-Pearson. Ce test d'hypothèse binaire (FADA-intégrale, voir eq. 5.6) revêt un caractère intuitif puisque qu'il cherche à détecter des groupes de pics temps-fréquence qui émergent du bruit. Finalement, les performances de FADA-intégrale, FADA-Norme 1 et FADA-Norme 2 sont très similaires. Pour cette raison nous illustrerons dans les prochaines sections la méthode sur données réelles à l'aide du détecteur FADA-intégrale.

## 5.5 Illustration sur données réelles

Dans cette section nous illustrons maintenant la méthode FADA-intégrale sur données réelles d'odontocètes.

### 5.5.1 Segmentation de sifflements

Nous présentons tout d'abord les résultats de la méthode sur des signaux relativement simples. Comme indiqué dans la section précédente, afin de ne pas multiplier inutilement le nombre de figures, nous ne montrons ici que les résultats obtenus avec FADA-intégrale.

Les figures 5.15 et 5.16 montrent le cas de sifflements avec des RSB élevés. Les différentes composantes harmoniques sont bien segmentées. Nous remarquons que les sifflements proches sont réunis dans une seule et même région temps-fréquence,

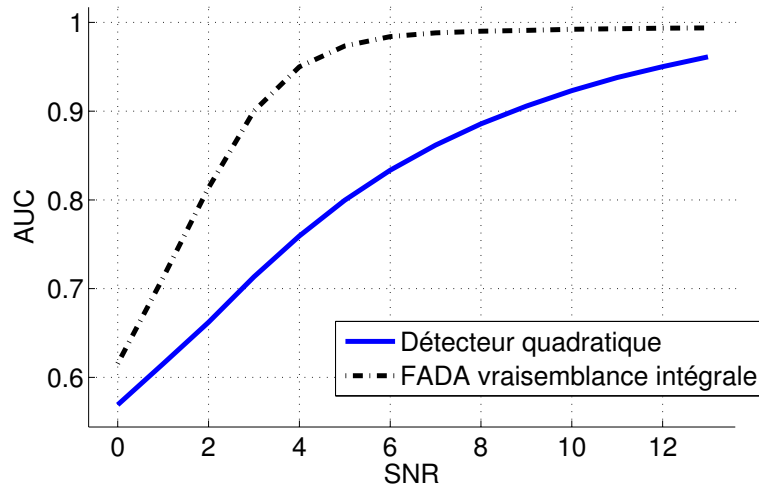


FIGURE 5.14 – Courbes AUC pour le détecteur quadratique (trait plein) et pour le détecteur FADA-intégrale. Le détecteur proposé fournit de meilleures performances pour tous les RSB.

il en va de même pour les multi-trajets. Ce comportement n'est pas une surprise, la méthode que nous proposons n'a pas la capacité de séparer les sifflements proches. Cette étape de séparation devra être faite en post-traitement (voir par exemple [JW11]).

### 5.5.2 Segmentation de sifflements en milieu impulsionnel

Nous nous intéressons maintenant au cas, bien plus difficile, de la segmentation de sifflements en milieu impulsionnel. Comme nous l'avons expliqué précédemment, il est très difficile de trouver une représentation qui soit simultanément adaptée aux sifflements et aux clics. **Par conséquent, il sera difficile de détecter simultanément les clics et les sifflements.**

Nous proposons de résoudre ce problème de manière **séquentielle**. L'idée est de d'abord détecter les clics, car ceux-ci sont très problématiques lors de la détection des sifflements [MOCT08, RBP<sup>+</sup>11], puis d'utiliser ces détections pour obtenir un spectrogramme débruité, sur lequel les sifflements seront détectés.

#### Suppression des impulsions dans le spectrogramme

Nous nous intéressons tout d'abord à la suppression des impulsions dans le spectrogramme. Nous montrons en figure 5.17 le processus original de suppression des impulsions dans le spectrogramme.

À partir de la série temporelle (figure 5.17.a) un **spectrogramme adapté aux clics** est construit (figure 5.17.c), c'est à dire que nous utilisons une fenêtre de pondération très courte. La méthode de détection FADA est alors appliquée pour

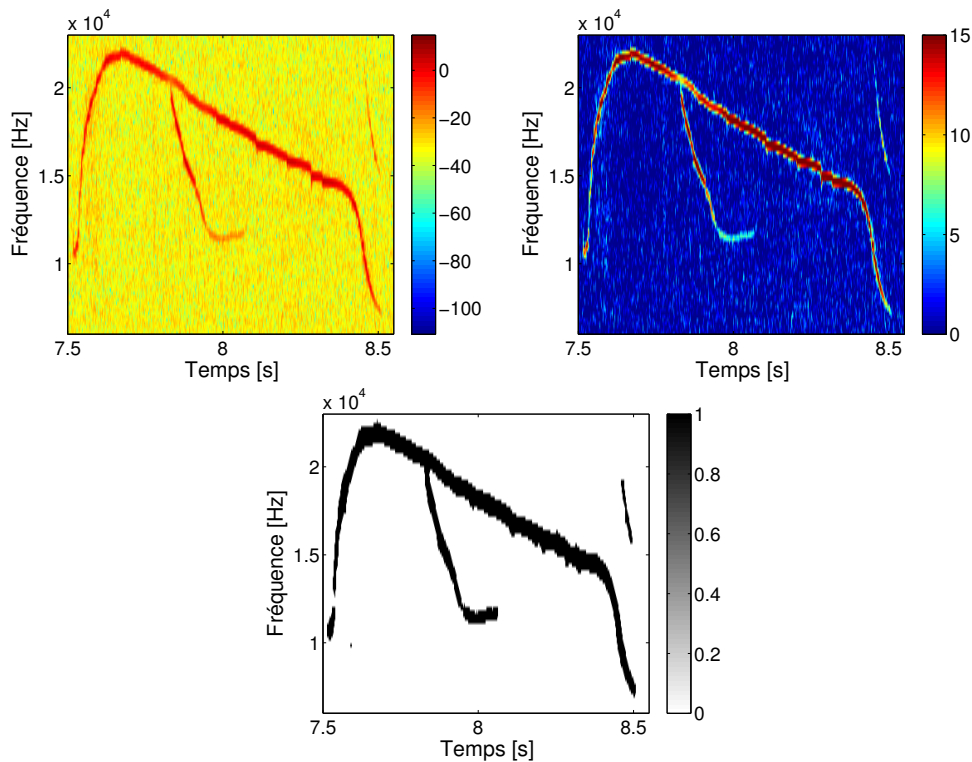


FIGURE 5.15 – (a) Spectrogramme original du signal, puis (b) Carte représentant le RSB du signal (affichée entre 0 et 15dB), (c) Spectrogramme segmenté. Les parties de fort RSB sont correctement segmentées, tout comme les parties avec un RSB plus faible ( $\sim 5$ dB)

la **détection de clics**, avec une région verticale pour le comptage des détections. Les **échantillons temporels** correspondants aux clics ainsi détectés sont **mis à zéro** (figure 5.17.d). Une série temporelle ne contenant que les clics est alors générée (figure 5.17.e). Un **spectrogramme adapté aux sifflements**, c'est à dire avec une fenêtre de pondération plus longue, est construit à partir de la série temporelle sans clics. Nous obtenons alors un **spectrogramme sans bruit impulsionnel** (figure 5.17.f).

La mise à zéro des échantillons temporels et non des coefficients du spectrogramme permet ainsi d'éviter les "trous" dans le spectrogramme adapté aux sifflements. Une partie des échantillons temporels mis à zéro trouvaient leur origine à la fois dans les clics mais aussi dans les sifflements. La conséquence de cette méthode de suppression des impulsions est donc la perte de puissance des sifflements.

Nous illustrons le potentiel de cette méthode sur trois exemples (figures 5.19, 5.20 et 5.22) de signaux réels issus de la base de données MobySound. Nous appliquons la méthodologie de suppression des impulsions telle que nous venons de la décrire. Le spectrogramme adapté à la détection de clics est calculé sur 64 points avec une fenêtre de Hann de 64 points et 32 points de recouvrement entre fenêtres. La DSP

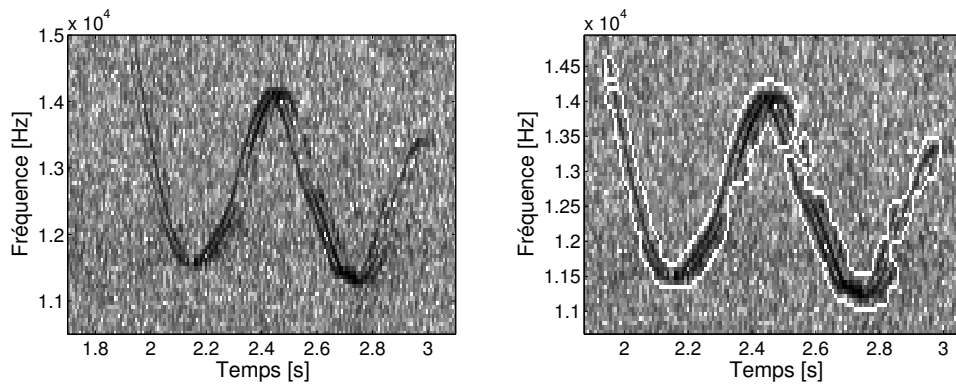


FIGURE 5.16 – Spectrogramme original du signal et segmentation. Source : ERATO - SHOM.

du bruit est estimé sur un voisinage  $V = [1s \times 500Hz]$ . La probabilité de fausse alarme est fixé à  $p_{fa}^{\mathcal{R}} = 0.001$ . Les deux spectrogrammes montrés sont adaptés aux sifflements, ils sont calculés sur 1024 points en utilisant une fenêtre de Hann de 1024 points et 512 points de recouvrement. Le premier signal (figure 5.19) montre un sifflement au milieu d'un train de clics. Les impulsions sont enlevées de manière très efficace. À noter que le sifflement n'est pas "morcelé" en plusieurs composantes temps-fréquence. Le second signal (figure 5.20) montre plusieurs sifflements au milieu d'un train de clics très rapproché. De même nous sommes capable de supprimer les impulsions de manière extrêmement efficace. Le troisième exemple est un signal plus long, observé sur une gamme de fréquence de 0 à 80kHz. Ce signal est très difficile dans le sens où nous observons des sifflements au milieu de "buzz" qui forment des pavés temps-fréquence. Le débruitage proposé n'est pas parfait, il reste encore des résidus d'énergie des sons impulsionnels et nous avons aussi perdu une partie de l'énergie des sifflements (voir par exemple entre 20 et 20.5s autour de 10kHz).

### Détection des sifflements sur spectrogramme débruité

Après avoir détecté les clics par FADA, nous sommes en mesure de détecter les sifflements par FADA sur le spectrogramme libéré des impulsions.

Il faut cependant noter une subtilité lors de l'utilisation de FADA pour la détection de sifflement sur spectrogramme débruité. Lorsque nous mettons à zéro les échantillons temporels correspondants aux clics **nous modifions les statistiques du spectrogramme** que nous observons. L'ensemble de la méthode FADA repose sur une hypothèse de gaussianité du signal étudié. Or en mettant les échantillons à zéro ces statistiques sont modifiées. **L'estimation de la DSP du bruit**, au sens où nous l'avons décrite tout au long de ce manuscrit doit donc être **effectuée sur le spectrogramme original**, non modifié.

Cette estimation est cependant difficile dans le cas où les clics sont très rapprochés. Le voisinage d'estimation  $V$  doit donc être choisi suffisamment grand (quelques secondes suffisent) pour que le nombre de coefficients "bruit seul" dans le voisinage permette une estimation fiable. La conséquence étant que le détecteur ne pourra pas s'adapter rapidement à une évolution des statistiques du bruit lorsque nous sommes en présence de nombreux clics.

La méthode de suppression des impulsions et de segmentation du spectrogramme en figure 5.23. À partir de la forme d'onde temporelle nous calculons deux spectrogrammes : un adapté aux clics, l'autre adapté aux sifflements. Nous estimons alors la DSP du bruit dans le spectrogramme adapté aux impulsions, ce qui nous permet de détecter les clics et de mettre à zéro les échantillons temporels correspondants. Afin de segmenter les sifflements nous appliquons alors FADA sur le spectrogramme sans impulsions

Nous montrons un exemple de segmentation de sifflements dans un train de clics en figure 5.24. Nous appliquons la méthodologie de segmentation de sifflements en milieu impulsif décrite dans la figure 5.23. Les paramètres pour la détection d'impulsions sont les mêmes que précédemment ( $V = [1s \times 500Hz]$ , et probabilité de fausse alarme 0.01). Les paramètres de FADA pour la détection des sifflements sont la forme de la région (choisie pour s'adapter à un signal "moyennement non stationnaire", soit  $5 \times 7$  pixels temps-fréquence) et la probabilité de fausse alarme  $p_{fa}^{\mathcal{R}}$  choisie égale à 0.01. Nous sommes capable de segmenter les sifflements au milieu des trains de clics de manière très efficace.

### **Quel usage de ces résultats de segmentation ?**

L'approche que nous avons proposé dans ce manuscrit n'est pas une alternative au méthode traditionnellement utilisées en acoustique passive. Au contraire, elle a été pensée et construite de manière à être complémentaire des méthodes classiques de suivi des composantes temps-fréquence utilisées en acoustique passive (Silbido [RBP<sup>+</sup>11], filtres de Kalman [MOCT08], filtres notch [JW11], inférence Bayésienne [HE06] et trackers utilisant la cohérence de phase par exemple [IGSM10]). L'exemple que nous prenons ici est celui de la réallocation de spectrogramme [AF95]. En particulier nous utilisons les résultats de segmentation obtenus comme masque binaire sur le spectrogramme réalloué.

Les figures 5.25, 5.26, 5.27 et 5.28 sont ainsi des exemples d'application de FADA sur données réelles issues de la base de données MobySound.

La figure 5.25.a) montre l'exemple d'un sifflement au milieu de clics. Le spectrogramme est calculé sur 1024 points avec une fenêtre de Hann de 1024 points et



512 points de recouvrement. Les impulsions au sein de ce spectrogramme sont ensuite supprimées en utilisant FADA adapté aux clics (figure 5.25.b). Pour cela un spectrogramme adapté aux clics est calculé (longueur de 64 points, fenêtre de Hann de 64 points et 32 points de recouvrement) et la probabilité de fausse alarme est fixée à 0.01 pour la détection. Le spectrogramme débruité est alors segmenté par utilisation de FADA adapté aux sifflements (région de comptage  $5 \times 7$  pixels et probabilité de fausse alarme fixée à 0.01). Les bordures des régions détectées sont tracées en noir (figure 5.25.c). Finalement nous calculons le spectrogramme réalloué et nous masquons celui-ci à l'aide des résultats de segmentation (voir figure 5.25.d). Le résultat obtenu est une représentation très fine du sifflement et presque entièrement libre de tout bruit.

La même méthode, avec les mêmes paramètres est appliquée pour trois autres exemples en figures 5.26, 5.27 et 5.28, avec un degré de difficulté croissant. Dans les différents cas, nous parvenons à segmenter les sifflements de manière très efficace.

## 5.6 Conclusion

Ce chapitre avait trois objectifs. Le premier concerne le choix des paramètres de la méthode. Ces deux paramètres sont la probabilité de fausse alarme  $p_{fa}$  utilisée pour binariser le spectrogramme. Nous avons étudié quatre manières de gérer le choix de  $p_{fa}$  : **1.** la maximisation du nombre de détections sous contrainte de probabilité de fausse alarme  $p_{fa}^{\mathcal{R}}$  constante au second test d'hypothèse binaire, **2.** l'intégration de la vraisemblance par rapport à  $p_{fa}$ , **3.** le calcul de la norme 1 sur un vecteur de vraisemblance et finalement **4.** le calcul de la norme euclidienne sur un vecteur de vraisemblance.

Le choix de la forme de la région de comptage est abordé. Nous donnons des critères de choix et différencions le cas de la détection de sifflements de celui de la détection de clics. Dans le cas des sifflements nous relient le choix de la région de comptage aux paramètres de construction du spectrogramme : forme de la fenêtre de pondération et recouvrement entre fenêtres. Le choix final de la région de comptage est finalement dépendant de la dynamique fréquentielle attendue de la part du signal.

Le second objectif concerne l'analyse numérique des performances de FADA pour les quatre manières de gérer  $p_{fa}$ . L'analyse des performances a été étudié sous l'angle de courbes COR, qui nous permettent d'avoir une idée très précise des comportements des détecteurs à SNR constant, mais aussi sous l'angle des courbes AUC, qui permettent d'étudier un comportement global pour tous les SNR. Enfin nous avons introduit la courbe iso-AUC qui mesure la variation de RSB nécessaire pour que deux détecteurs aient les mêmes performances globales. Ces différentes

mesures de performances montrent que le détecteur FADA montre des performances nettement supérieures au détecteur quadratique simple.

Finalement nous traitons le cas de la détection de sifflements en milieu fortement impulsionnel pour des signaux réels. Nous montrons que la méthode proposée est parfaitement applicable en conditions réelles.



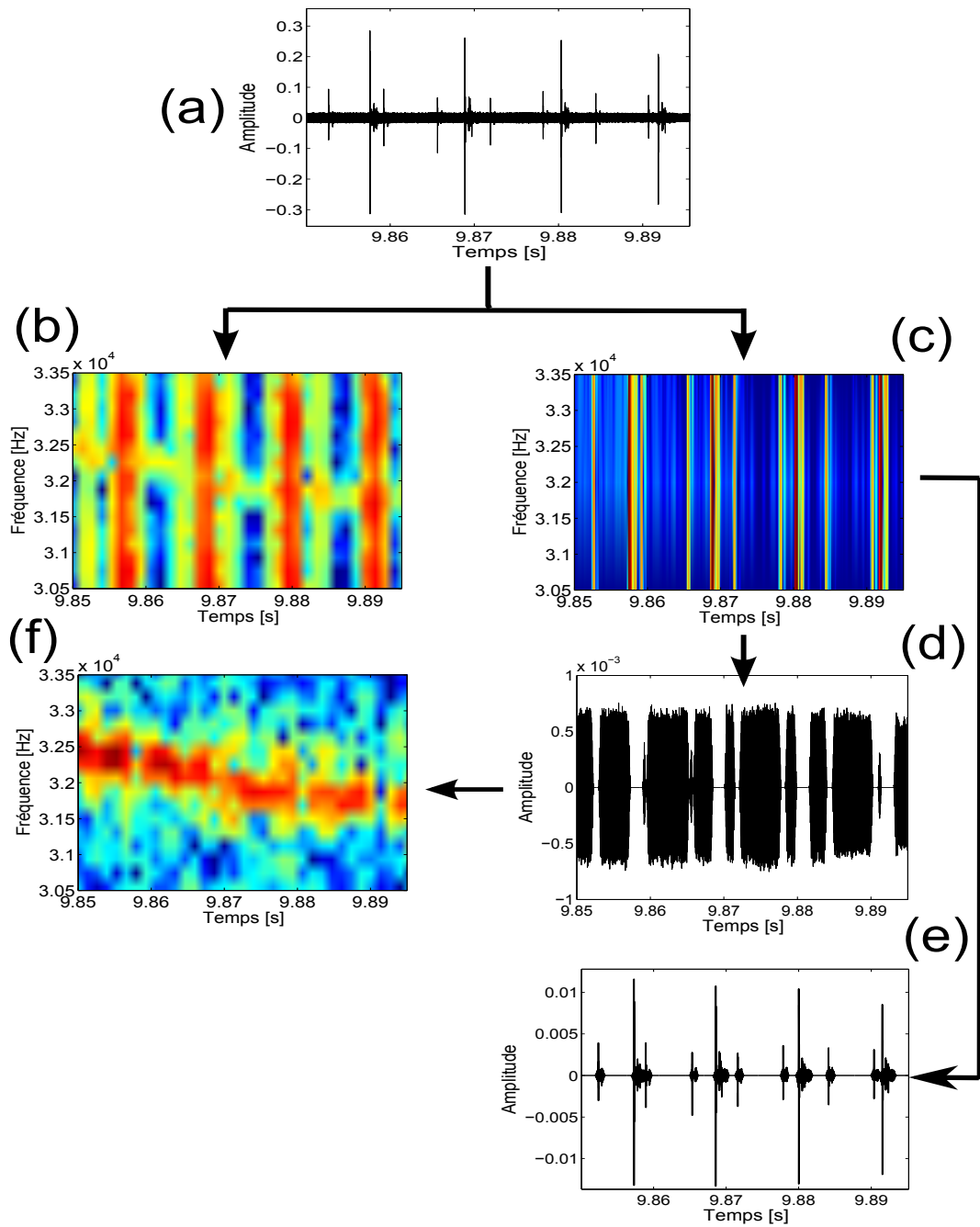


FIGURE 5.17

FIGURE 5.18 – **Débruitage de sifflements en milieu impulsionnel** : (a) Forme d'onde du signal, (b) Spectrogramme adapté aux sifflements : fenêtre de 2048 points (équivalent à 10ms); (c) Spectrogramme adapté aux clics : fenêtre de 128 points (équivalent à 0.6ms). La détection est appliquée aux clics; (c) Spectrogramme adapté aux sifflements avec clics filtrés; (d) Forme d'onde temporelle sans les clics détectés par FADA, (e) Forme d'onde temporelle des clics détectés et (f) Spectrogramme débruité et adapté aux sifflements.

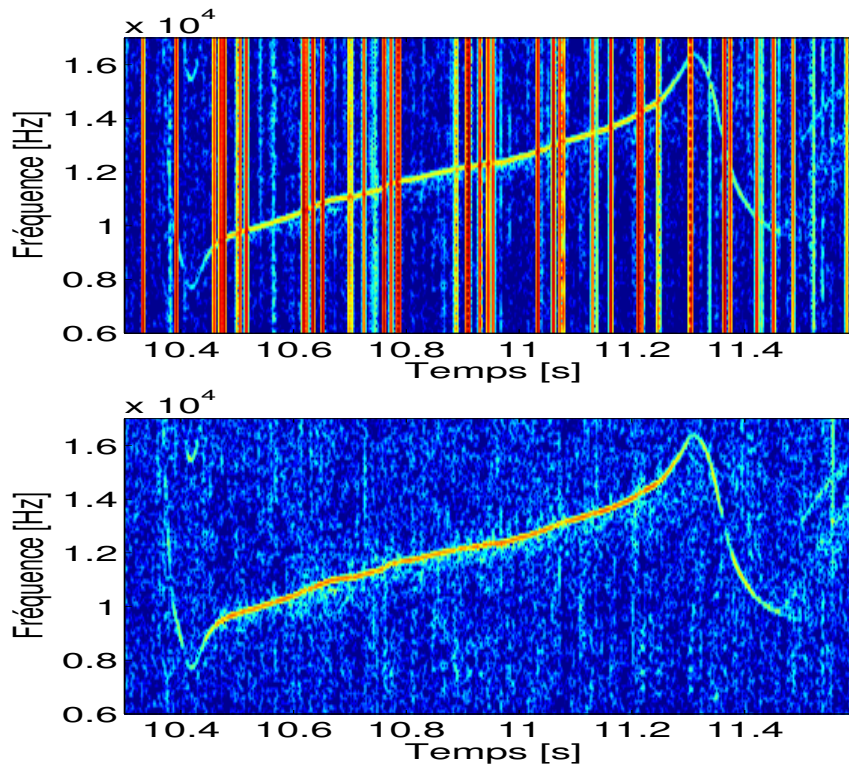


FIGURE 5.19 – Débruitage du spectrogramme par suppression des clics dans le domaine temporel (Source : 'Qx-Tt-SCI0608-N1-060814-121518.wav', MobySound).

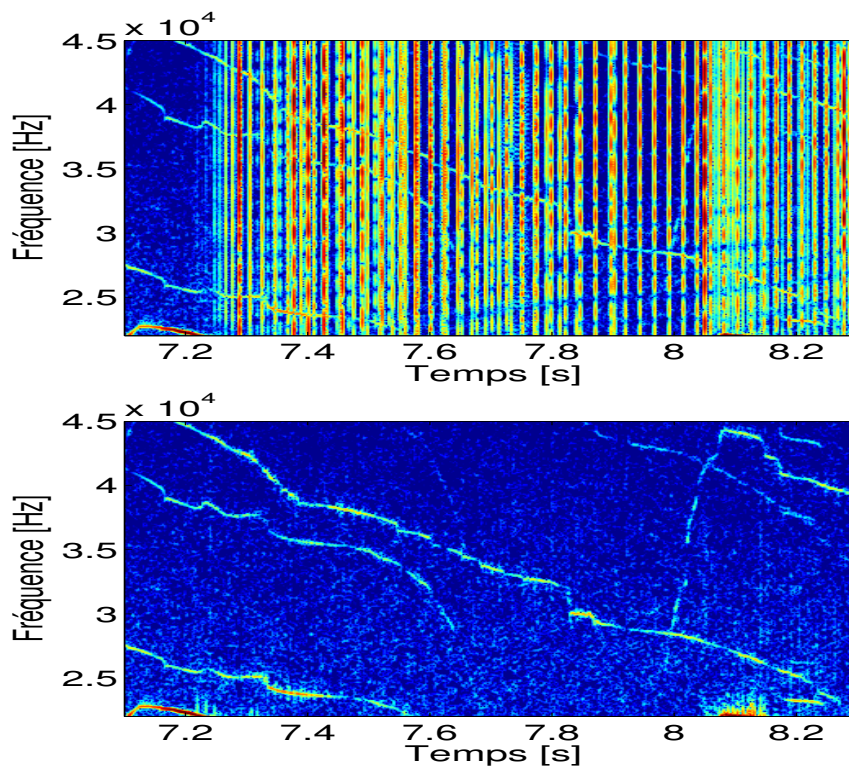


FIGURE 5.20 – Débruitage du spectrogramme par suppression des clics dans le domaine temporel (Source : 'Qx-Tt-SCI0608-N1-060814-121518.wav', MobySound).

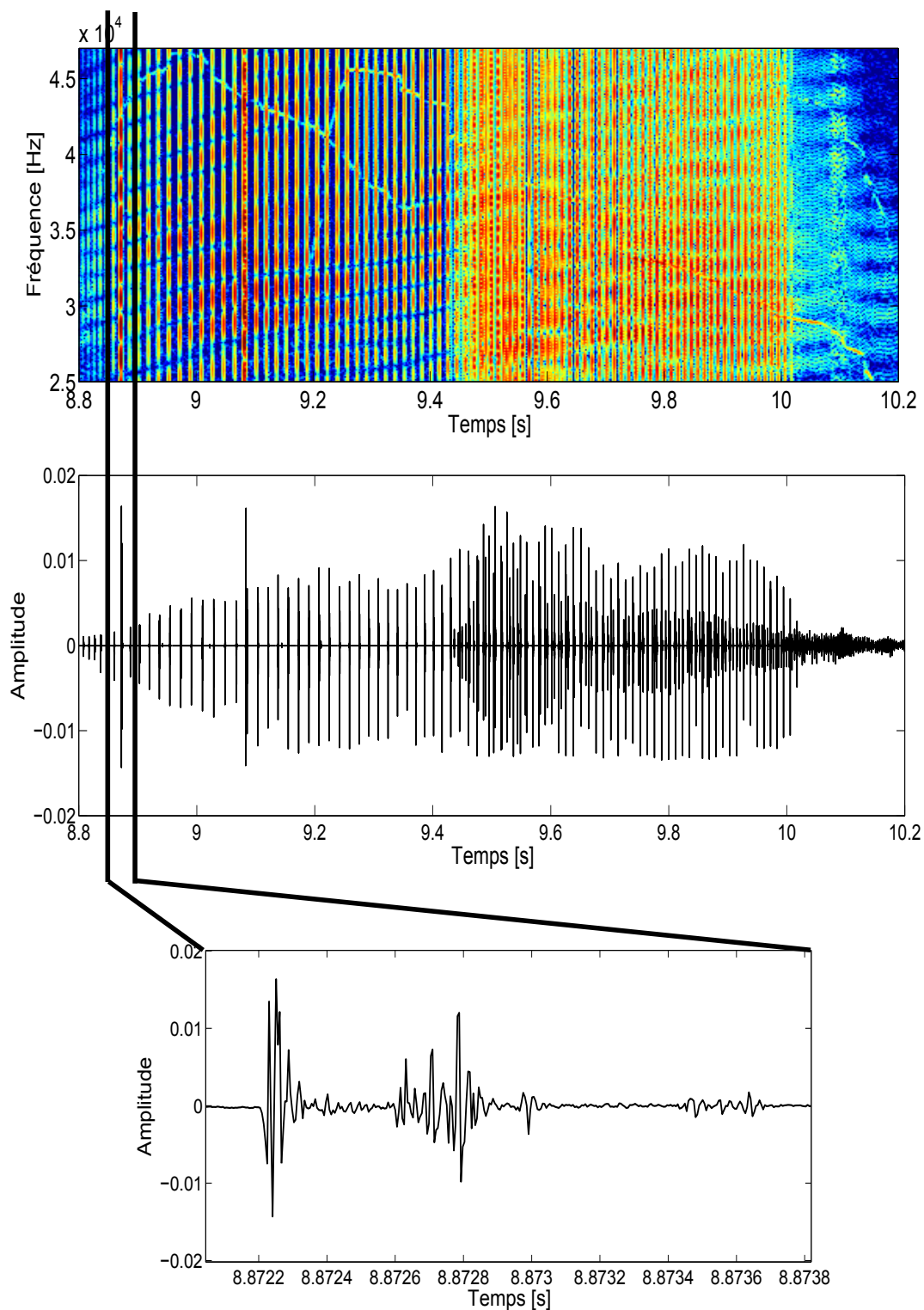
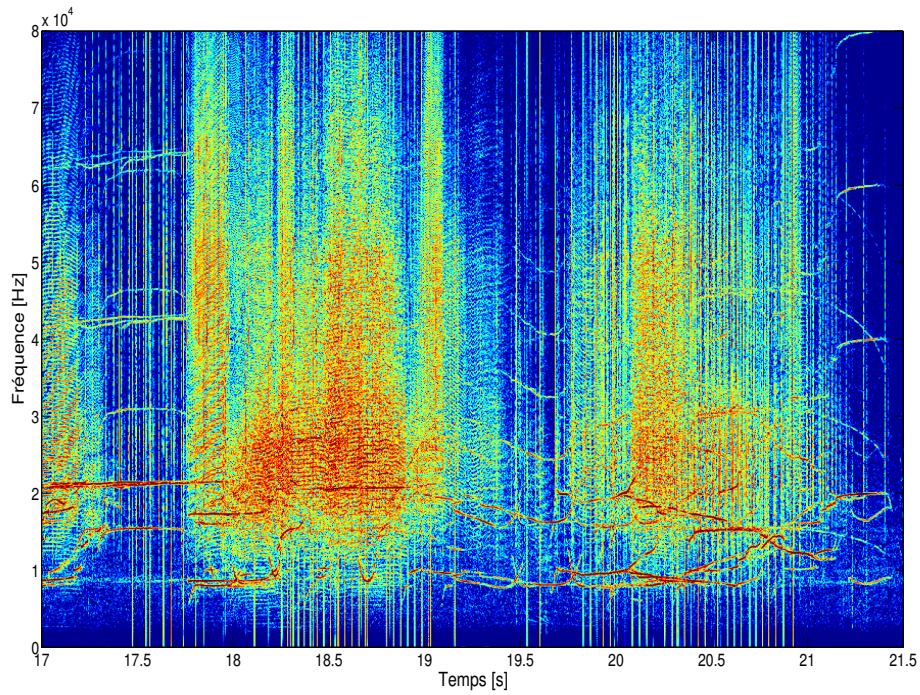
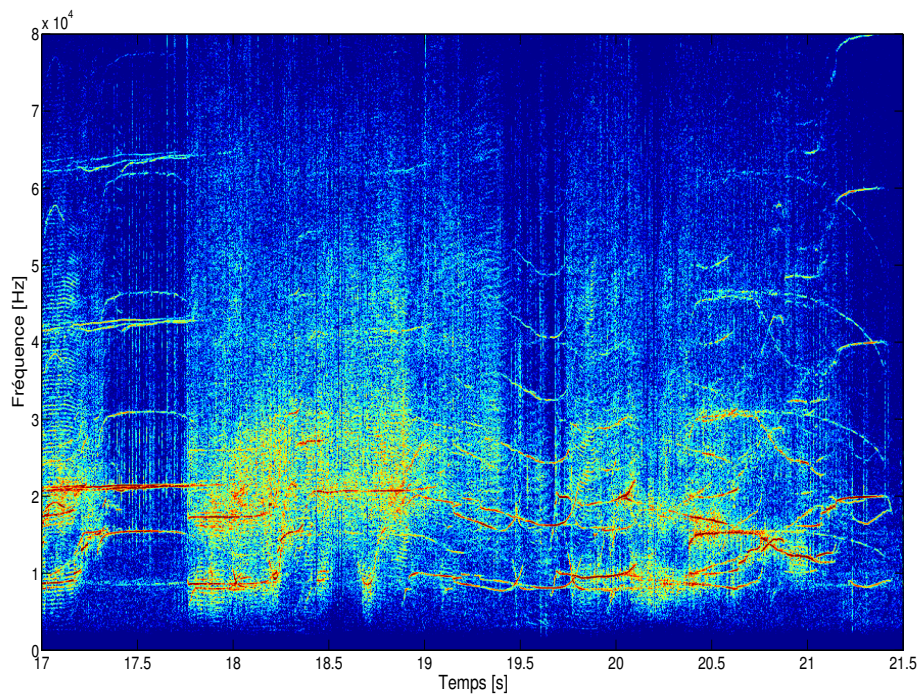


FIGURE 5.21 – Présentation des résultats de détection de clics. La méthode FADA est appliquée à la détection de clics. Les fenêtres temporelles au sein desquelles des clics ont été détectés sont extraites (a) Spectrogramme du signal (Source : 'Qx-Tt-SCI0608-N1-060814-121518.wav', Mobysound) ; (b) Forme d'onde temporelle des clics détectés ; (c) Zoom sur une des formes d'onde.



(a)



(b)

FIGURE 5.22 – Illustration de FADA appliqué à la détection et à la suppression de clics (a) Spectrogramme du signal issu de la base de données Moby-sound, (b) Spectrogramme débruité ( $p_{fa} = 0.01$ ) (MobySound : 'Qx-Dc-FLIP0610-VLA-061015-163500.wav')

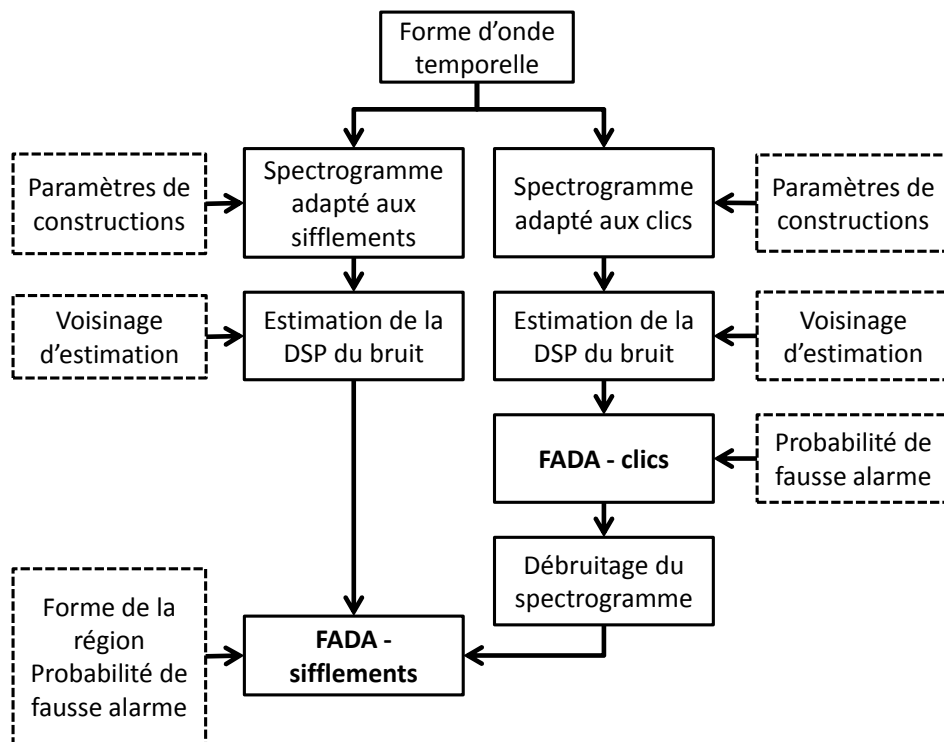
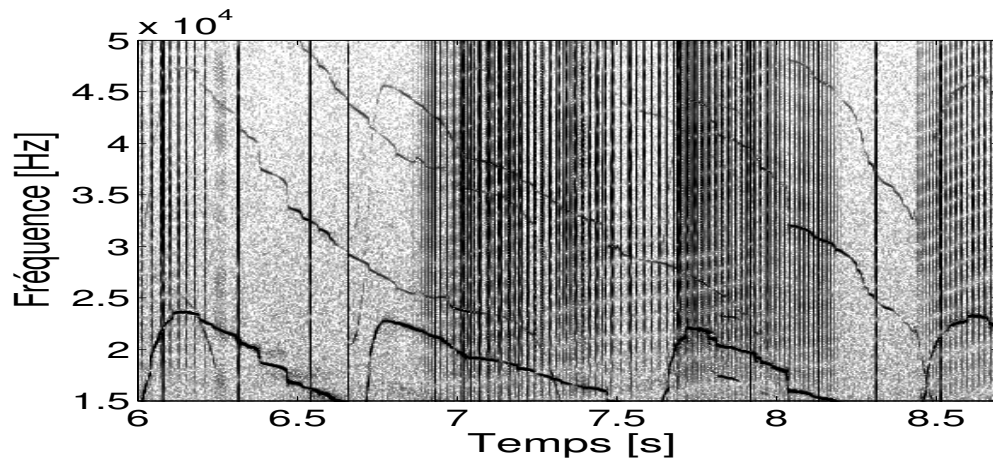
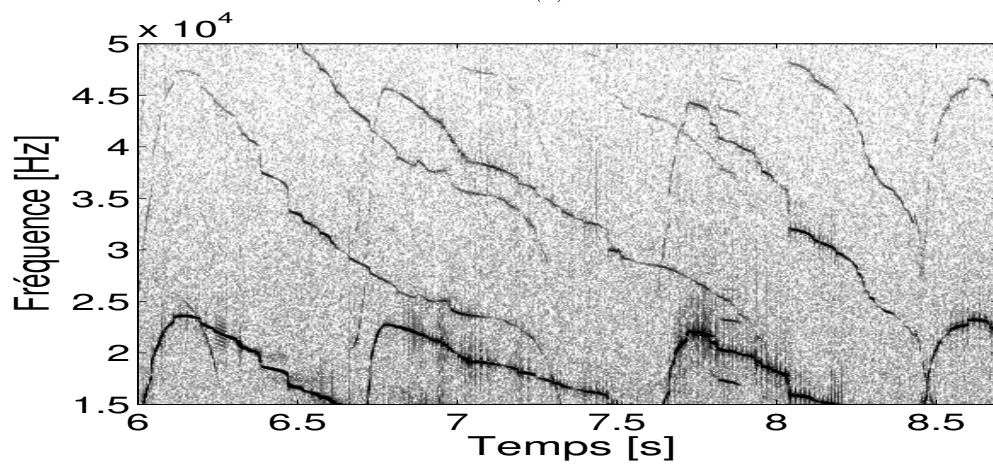


FIGURE 5.23 – **Schéma global de segmentation des sifflements en milieu impulsionnel.** Les boîtes en trait plein représentent les différentes étapes du processus. Les boîtes en pointillés indiquent pour chaque étape du processus les paramètres à régler. Le processus de segmentation est composé de deux étapes principales. À partir de la forme d'onde temporelle sont calculés deux spectrogrammes. Le premier adapté aux sifflements, le second aux clics. Les DSP du bruit sont estimées pour chacun des deux spectrogrammes. Les clics sont détectés par FADA et supprimés dans le domaine temporel : un spectrogramme sans impulsions est obtenu. Nous appliquons ensuite FADA sur ce spectrogramme sans impulsions (en utilisant la DSP estimée avant débruitage) pour détecter les sifflements.

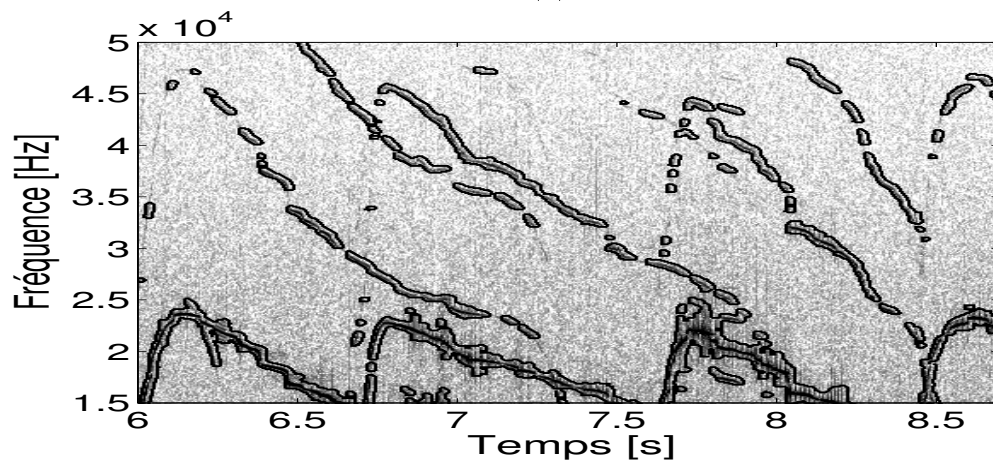




(a)

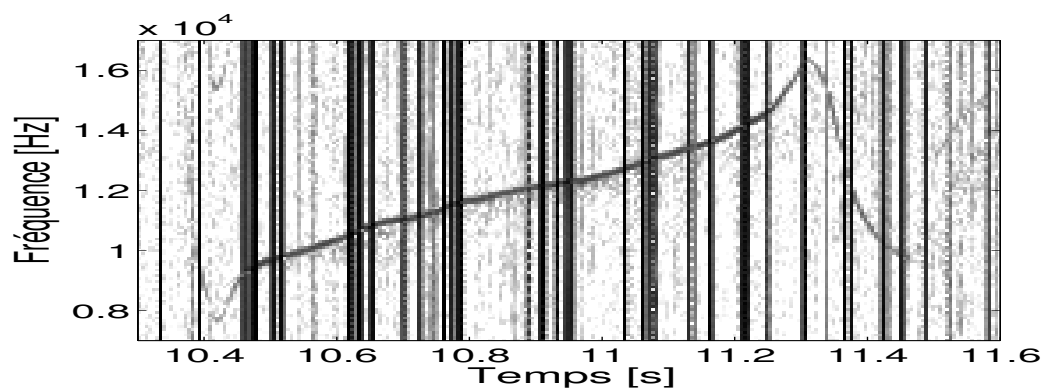


(b)

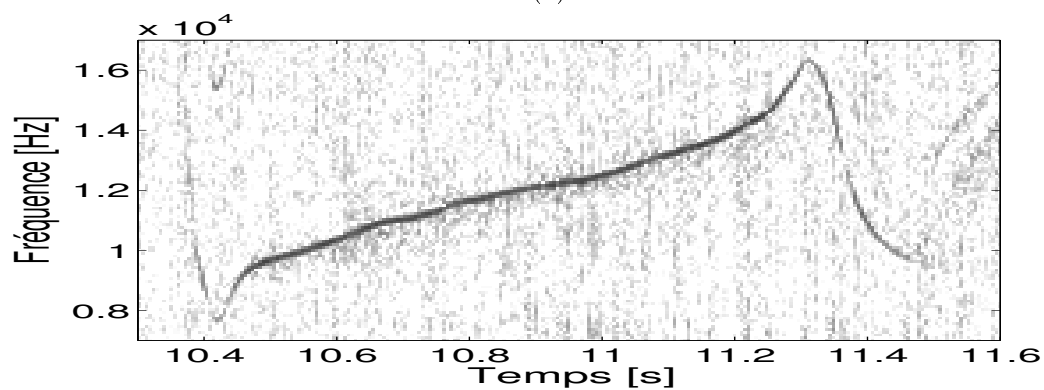


(c)

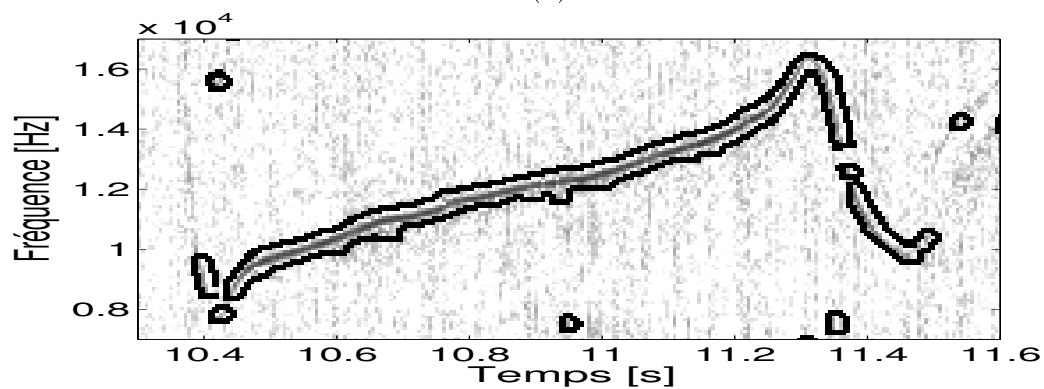
FIGURE 5.24 – Illustration de FADA appliqué à la détection de sifflements au milieu d'un train de clics (a) Spectrogramme d'un signal de la base de données Mobysound, (b) Spectrogramme débruité ( $p_{fa} = 0.01$ ), (c) Spectrogramme segmenté ( $p_{fa} = 0.05$ ) (Source : 'Qx-Tt-SCI0608-N1-060814-121518.wav', MobySound)



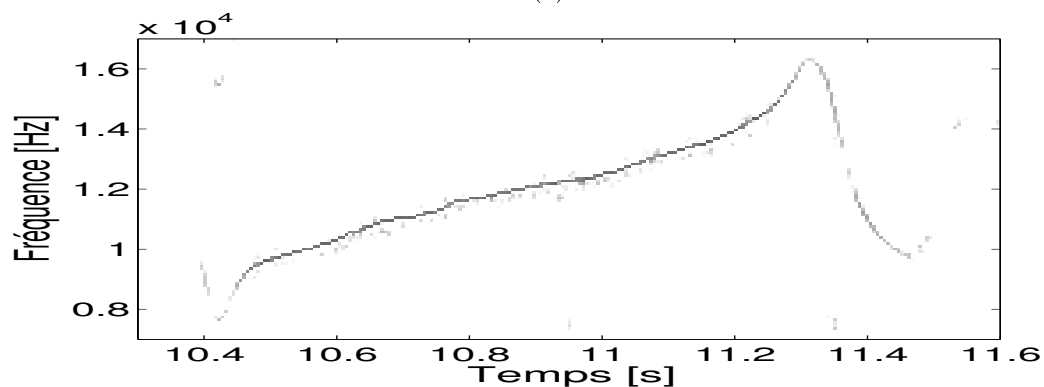
(a)



(b)

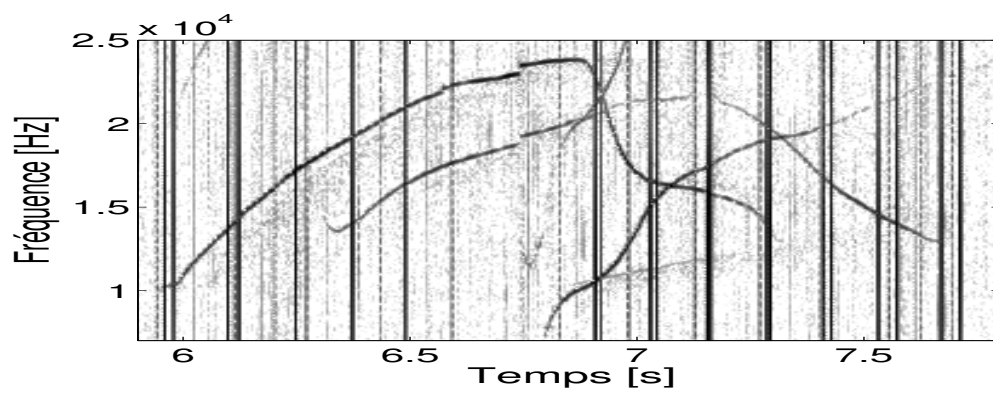


(c)

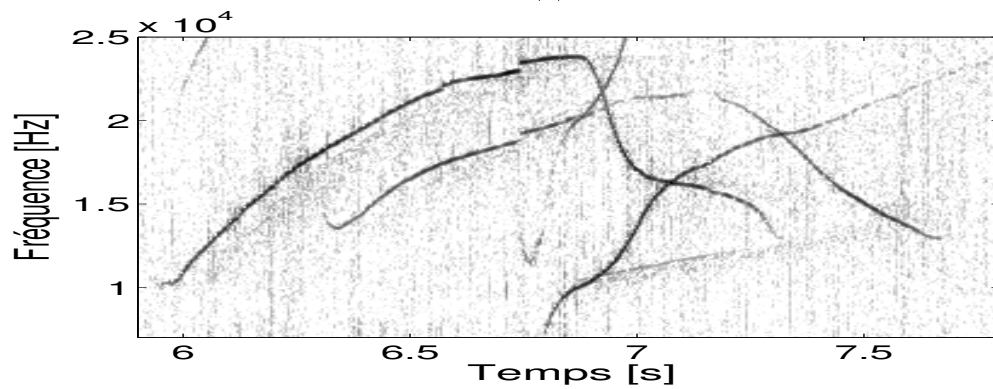


(d)

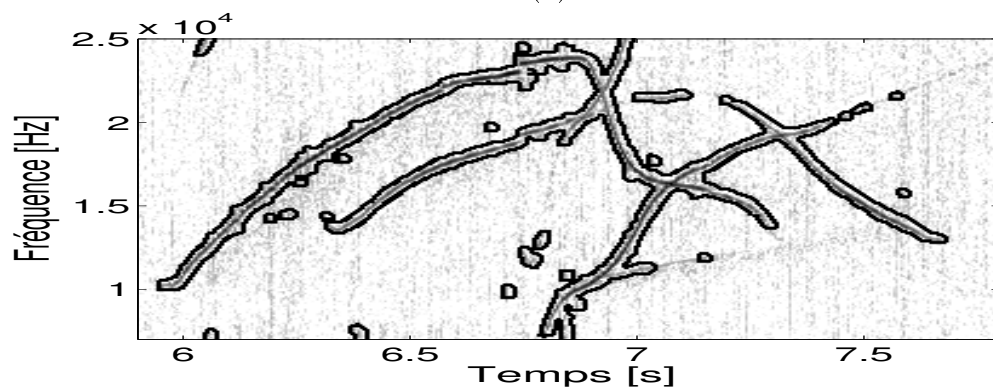
FIGURE 5.25 – Illustration de FADA appliqué à la détection et à la réallocation de sifflements au milieu d'un train de clics (a) Spectrogramme d'un signal de la base de données Mobysound, (b) Spectrogramme débruité ( $p_{fa} = 0.01$ ), (c) Spectrogramme segmenté ( $p_{fa} = 0.05$ ), (d) Spectrogramme réalloué (Source : 'Palmyra102006-061103-213127\_4.wav', MobySound)



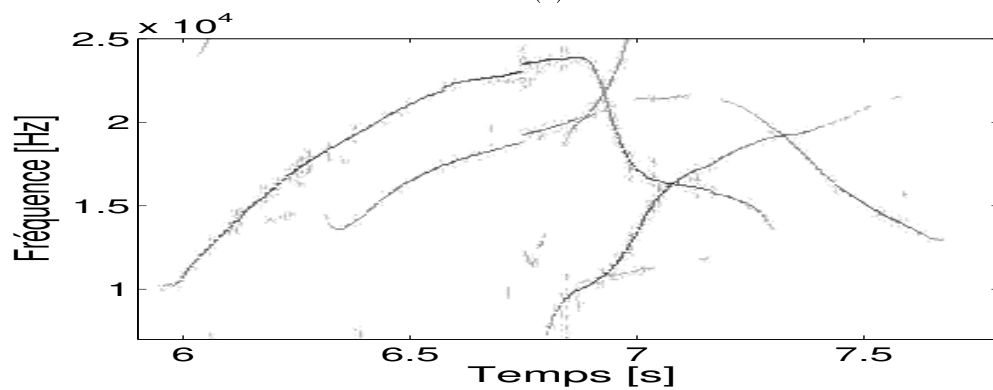
(a)



(b)

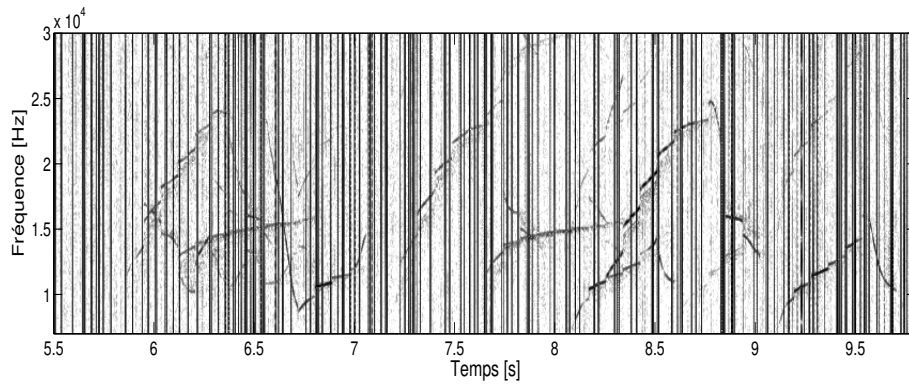


(c)

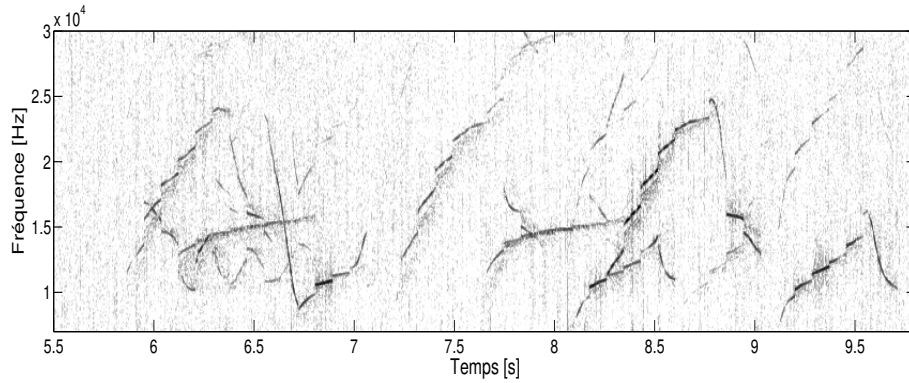


(d)

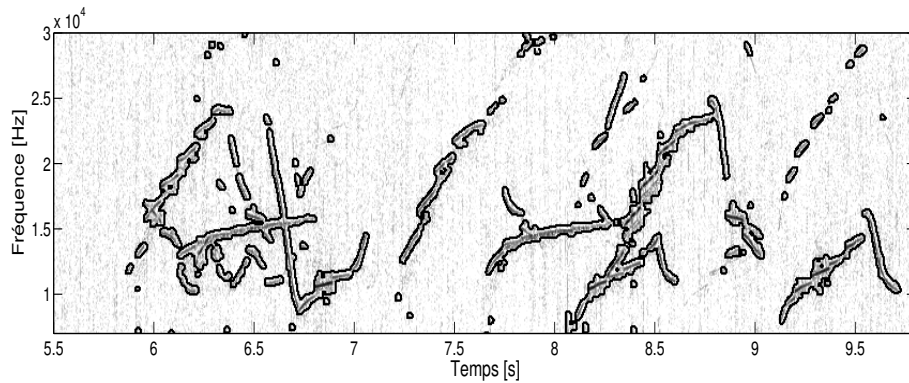
FIGURE 5.26 – Illustration de FADA appliqué à la détection et à la réallocation de sifflements au milieu d'un train de clics (a) Spectrogramme d'un signal de la base de données Mobysound, (b) Spectrogramme débruité ( $p_{fa} = 0.01$ ), (c) Spectrogramme segmenté ( $p_{fa} = 0.05$ ), (d) Spectrogramme réalloué (Source : 'Palmyra102006-061103-213127\_4.wav', MobySound)



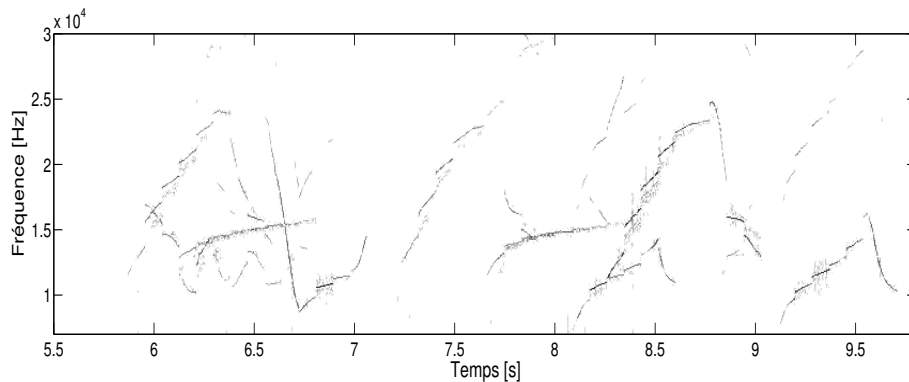
(a)



(b)

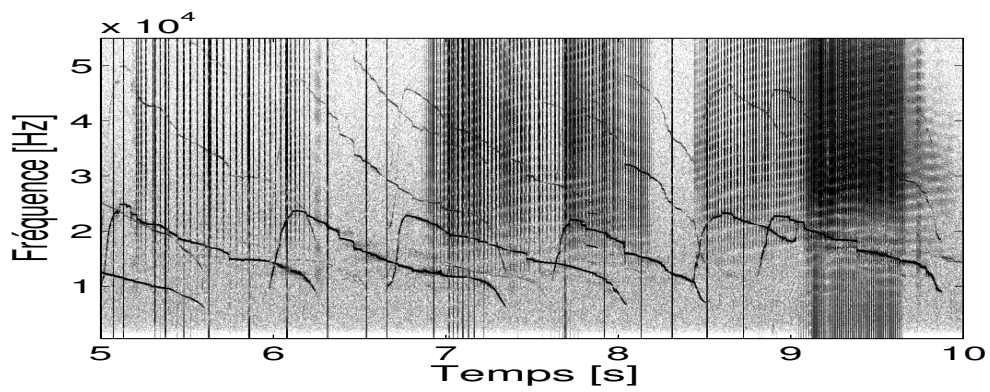


(c)

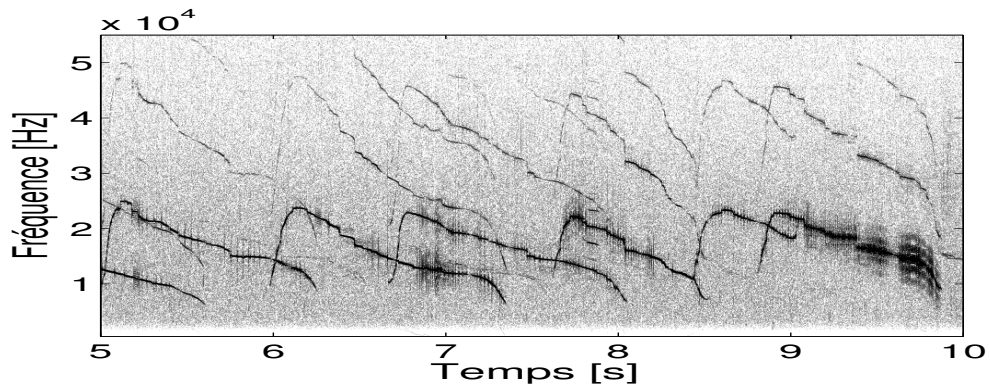


(d)

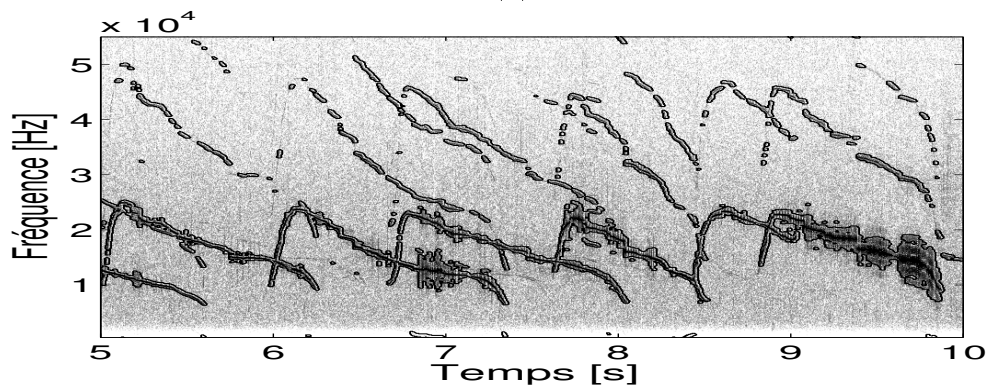
FIGURE 5.27 – Illustration de FADA appliqué à la détection et à la réallocation de sifflements au milieu d'un train de clics (a) Spectrogramme d'un signal de la base de données Mobysound, (b) Spectrogramme débruité ( $p_{fa} = 0.01$ ), (c) Spectrogramme segmenté ( $p_{fa} = 0.05$ ), (d) Spectrogramme réalloué (Source : 'Palmyra102006-061103-213127\_4.wav', MobySound)



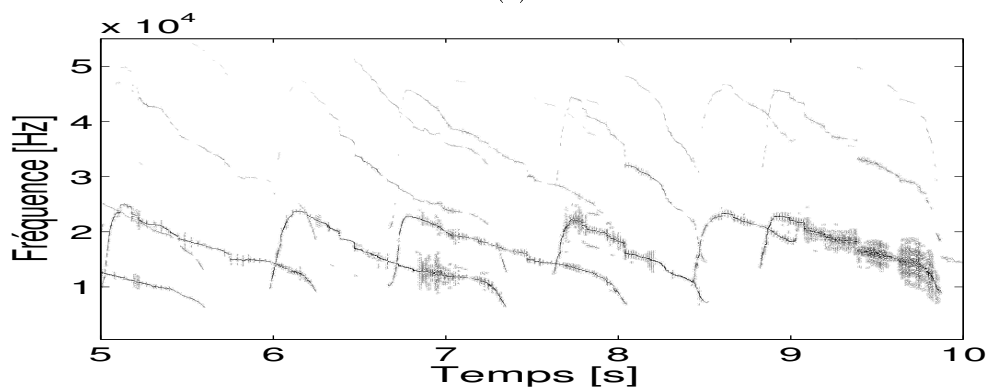
(a)



(b)



(c)



(d)

FIGURE 5.28 – Illustration de FADA appliqué à la détection et à la réallocation de sifflements au milieu d'un train de clics (a) Spectrogramme d'un signal de la base de données Mobysound, (b) Spectrogramme débruité ( $p_{fa} = 0.01$ ), (c) Spectrogramme segmenté ( $p_{fa} = 0.05$ ), (d) Spectrogramme réalloué (Source : 'Qx-Tt-SCI0608-N1-060814-121518.wav', MobySound)

# Conclusions et perspectives

Cette thèse s'insère dans un contexte de **monitoring du milieu marin par acoustique passive** (PAM). Les océans subissent des pressions d'origine anthropique fortes, liées en particulier à la constante augmentation du bruit rayonné par les activités industrielles, et militaires. L'impact de cette augmentation sur les écosystèmes marins, et en particulier sur les mammifères marins, est largement inconnu. Les ondes acoustiques sont particulièrement bien adaptées au milieu marin, et sont donc abondamment utilisées par la faune marine. L'écoute de ce paysage acoustique sous-marin permet donc d'avoir accès de manière non intrusive à une information unique sur cet écosystème. Un enjeu tout particulier du PAM concerne donc les activités dites de **détection, classification, localisation et dénombrement des mammifères marins**.

La qualité de la détection des signaux est critique quel que soit l'objectif final. C'est pourquoi, l'**objectif principal** de cette thèse concerne la **détection robuste de signaux acoustiques** émis par les mammifères marins en utilisant le **spectrogramme**. Les mammifères marins produisent deux types de signaux : des sifflements (signaux à bande étroite) et des clics (signaux impulsionnels large bande). Nous nous intéressons tout particulièrement à la détection de sifflements, à **bas rapport signal à bruit** ( $RSB \leq 8dB$ ) et en **milieu impulsionnel** (i.e. en présence de clics et de l'activité benthique).

Les statistiques du bruit océanique sont **non stationnaires et colorées**, le processus de détection mis en place dans ce travail est par conséquent pensé de manière **locale**. En chaque point temps-fréquence du spectrogramme nous formulons un **test d'hypothèse binaire** qui vise à discriminer l'hypothèse "signal" de l'hypothèse "bruit seul". Le caractère **passif** de l'étude impose une contrainte forte : le signal d'intérêt que nous cherchons à détecter est **inconnu a priori**. En ce sens, l'ensemble de ce travail de thèse repose sur une description des statistiques du bruit, faisant du **critère de Neyman-Pearson** un outil central de ce travail.

Une présentation du paysage acoustique sous-marin et des enjeux (et difficultés)

liés aux activités PAM sont données dans le Chapitre 1. Nous présentons dans le Chapitre 2 le test d'hypothèse binaire utilisé pour la détection statistique du signal sur le spectrogramme. Ce test d'hypothèse, résolu par une approche de Neyman-Pearson, se base sur la comparaison entre l'énergie d'un point temps-fréquence, et la densité spectrale de puissance du bruit.

Ce test souffre de deux maux : **1.** les performances en terme de compromis entre probabilité de fausse alarme/ probabilité de détection ne sont pas satisfaisantes pour de faibles rapports signal à bruit et **2.** la séparation entre sifflements et clics n'est pas aisée. La littérature propose alors un certain nombre de solutions visant à résoudre ces deux problèmes de manière empirique, ou itératives, ou encore en imposant un modèle sur le signal observé. **L'originalité de ce travail réside dans la formulation statistique du problème et dans l'absence de modèle imposé sur le signal.**

**L'idée principale** véhiculée par ce manuscrit est la suivante : l'analyse de l' **organisation spatiale des détections sur le spectrogramme binaire** permet de résoudre les deux problèmes évoqués précédemment. Nous construisons alors une méthode statistique, que nous appelons FADA (pour False Alarm Density Analysis), qui intègre explicitement l'information spatiale dans le processus de détection.

Pour ce faire, le nombre de détections observé dans des régions du spectrogramme binaire est utilisé pour discriminer les régions abritant une partie du support temps-fréquence du signal de celles n'abritant que du bruit. Les Chapitres 3 et 4 sont donc dédiés à la description statistique de la densité des fausses alarmes dans le spectrogramme binaire. La densité de probabilité du nombre de fausses alarmes dans les régions du spectrogramme binaire est d'abord modélisée par une loi **binomiale** (Chapitre 3) puis par une loi **binomiale corrélée**, qui permet la prise en compte de la **redondance du spectrogramme** (Chapitre 4). Un apport significatif du Chapitre 4 concerne la définition de la notion de **spectrogramme assistant** qui permet d'estimer simplement les densités de probabilité du spectrogramme binaire.

Le chapitre 5 s'intéresse tout d'abord aux degrés de liberté de FADA. Nous pensons qu'une condition d'applicabilité de FADA en conditions réelles passe par la réduction du nombre de degrés de liberté. Nous proposons ainsi des stratégies de seuillage multiple du spectrogramme et des critères adaptés à la détection de sifflements et de clics. Nous évaluons ensuite la pertinence de la méthode proposée numériquement et la validons sur données réelles réputées difficiles.

## Perspectives

Nous identifions quatre perspectives à ce travail. Les deux premières perspectives sont directement issues de ce travail et visent à combler deux lacunes que nous identifions dans ce travail. Les deux dernières sont des perspectives à plus long terme.

La première perspective concerne le développement théorique des performances de la méthode présentée. Les performances de la méthode ont été évaluées numériquement mais il serait bon de les évaluer théoriquement, au moins dans le cas d'un signal déterministe perturbé par un bruit aléatoire.

La seconde perspective concerne la quantification des performances sur de grandes bases de données d'odontocètes. La méthode présentée semble prometteuse au vu des exemples présentés sur données réelles mais il serait bon de la tester, seule ou en combinaison avec des méthodes de suivi temps-fréquence classique, et d'en évaluer les gains.

La troisième perspective concerne l'extension de la méthode proposée à d'autres classes de représentation des signaux. Une première adaptation semble immédiate à toutes les représentations temps-fréquence quadratiques issues de transformations linéaires, dont en première ligne les représentations temps-échelle : scalogramme et chirplettes par exemple. Le cas des représentations temps-échelles nécessitera de redéfinir la notion de voisinage temps-fréquence pour compter les détections. Nous avons bon espoir de pouvoir appliquer cette méthode à toute représentation dont les statistiques sont connues. En particulier, la densité de probabilité de la variable de comptage sera obtenue très facilement à partir du moment où nous sommes capables de seuiller la représentation à probabilité de fausse alarme constante. Il nous suffira alors ensuite de calculer la représentation assistante correspondante à partir d'un bruit blanc Gaussien pour obtenir la densité de probabilité de la variable de comptage dans le cas "bruit seul".

La quatrième perspective, plus lointaine, concerne cette fois la classification non supervisée. La détection telle que nous l'avons présentée dans ce travail de thèse n'est pas un objectif en soi. Un objectif intéressant serait de parvenir à détecter de manière automatique des signaux inconnus, et à les classer de manière non supervisée. Ce processus pourrait permettre d'évoluer dans le sens de la découverte de connaissance automatique [FPSS96] dans les très grandes bases de données sous-marines. En particulier nous pensons que le processus de classification doit explicitement prendre en compte la (mauvaise) qualité de la détection. Par qualité de la détection nous désignons la probabilité de fausse alarme, mais aussi la frag-



mentation des composantes temps-fréquences détectées et les erreurs d'estimation de la loi de fréquence instantanée. Certaines méthodes présentent dans la littérature semblent fournir de bonnes pistes de réflexion en ce sens, voir par exemple [EK SX96]. Le Saint-Graal étant, de mon opinion personnelle, la propagation de l'incertitude sur l'ensemble de la chaîne de traitement : de la détection d'évènement acoustique jusqu'à la classification, la localisation et le dénombrement des mammifères marins.

# Bibliographie

- [AB94] Rolf Adams and Leanne Bischof. Seeded region growing. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 16(6) :641–647, 1994.
- [AB98] Whitlow WL Au and Kiara Banks. The acoustics of the snapping shrimp *synalpheus parneomeris* in kaneohe bay. *The Journal of the Acoustical Society of America*, 103(1) :41–47, 1998.
- [ABMF12] Whitlow WL Au, Brian Branstetter, Patrick W Moore, and James J Finneran. The biosonar field around an atlantic bottlenose dolphin (*tursiops truncatus*). *The Journal of the Acoustical Society of America*, 131(1) :569–576, 2012.
- [AF95] François Auger and Patrick Flandrin. Improving the readability of time-frequency and time-scale representations by the reassignment method. *Signal Processing, IEEE Transactions on*, 43(5) :1068–1089, 1995.
- [All77] J Allen. Short-term spectral analysis, and modification by discrete fourier transform. *IEEE Transactions on Acoustics Speech and Signal Processing*, 25 :235–238, 1977.
- [Alp99] Bedri Alpar. Underwater signatures of the kocaeli earthquake (august 17th 1999). *Turk. J. Mar. Sci*, 5(3) :111–130, 1999.
- [AMGC02] M Sanjeev Arulampalam, Simon Maskell, Neil Gordon, and Tim Clapp. A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking. *Signal Processing, IEEE Transactions on*, 50(2) :174–188, 2002.
- [AS12] Milton Abramowitz and Irene A Stegun. *Handbook of mathematical functions : with formulas, graphs, and mathematical tables*. Courier Dover Publications, 2012.
- [Au93] Whitlow WL Au. *The sonar of dolphins*. Springer, 1993.
- [B<sup>+</sup>03] Ocean Studies Board et al. *Ocean noise and marine mammals*. National Academies Press, 2003.

- [Bah61] RR Bahadur. A representation of the joint distribution of responses to  $n$  dichotomous items. in studies in item analysis and prediction, (ed. h. solomon), 158-168, 1961.
- [Bal81] Roger Balian. Un principe d'incertitude fort en theorie du signal ou en mecanique quantique. *CR Acad. Sci. Paris*, 292(2) :1357–1361, 1981.
- [Bas90] Andrew H Bass. Sounds from the intertidal zone : vocalizing fish. *Bioscience*, pages 249–258, 1990.
- [BBAK06] Kelly J Benoit-Bird, Whitlow WL Au, and Ronald Kastelein. Testing the odontocete acoustic prey debilitation hypothesis : No stunning results. *The Journal of the Acoustical Society of America*, 120(2) :1118–1123, 2006.
- [BC14] Rina Foygel Barber and Emmanuel Candes. Controlling the false discovery rate via knockoffs. *arXiv preprint arXiv :1404.5609*, 2014.
- [BCI<sup>+</sup>12] Florin-Marian Birleanu, Ion Candel, Cornel Ioana, Cedric Gervaise, Alexandru Serbanescu, and Gheorghe Serban. A vector approach to transient signal processing. In *Information Science, Signal Processing and their Applications (ISSPA), 2012 11th International Conference on*, pages 1141–1146. IEEE, 2012.
- [BD13] Michael Bittle and Alec Duncan. A review of current marine mammal detection and classification algorithms for use in automated passive acoustic monitoring. *Proceedings of Acoustics 2013*, 2013.
- [BFH<sup>+</sup>10] Pierre Borgnat, Patrick Flandrin, Paul Honeine, Cédric Richard, and Jun Xiao. Testing stationarity with surrogates : A time-frequency approach. *Signal Processing, IEEE Transactions on*, 58(7) :3459–3470, 2010.
- [BGS<sup>+</sup>13] Amélie Barazzutti, Cedric Gervaise, Yann Stéphan, Florian Dadouchi, Jean-Pierre Sessarego, et al. Inversion géoacoustique passive en milieux petits fonds à partir de signaux représentatifs des émissions de cétacés. *Traitement du signal*, 30(3-4-5) :169–194, 2013.
- [BHTM14] Benjamin B Bougher, Joey Hood, James Theriault, and Hilary Moors. Generalized marine mammal detection based on improved band-limited processing. In *Proceedings of Meetings on Acoustics*, volume 17, page 070067. Acoustical Society of America, 2014.
- [BNMW10] Julien Bonnel, Barbara Nicolas, Jérôme I Mars, and Shane C Walker. Estimation of modal group velocities with a single receiver for geoacoustic inversion in shallow water. *The Journal of the Acoustical Society of America*, 128(2) :719–727, 2010.

- [BOC<sup>+</sup>11] Yvonne Barkley, Julie Nicola Oswald, James V Carretta, Shannon Rankin, Alexis Rudd, and Marc O Lammers. Comparison of real-time and post-cruise acoustic species identification of dolphin whistles using rocca (real-time odontocete call classification algorithm). Technical report, 2011.
- [Bon10] Julien Bonnel. *Analyse de la dispersion acoustique UBF (0-150 Hz) pour la surveillance et la caractérisation du milieu marin*. PhD thesis, Institut National Polytechnique de Grenoble-INPG, 2010.
- [BR92] Boualem Boashash and AP Reilly. *Algorithms for time-frequency signal analysis*. Longman Cheshire, 1992.
- [Bra97] Andrew P Bradley. The use of the area under the roc curve in the evaluation of machine learning algorithms. *Pattern recognition*, 30(7) :1145–1159, 1997.
- [BSS<sup>+</sup>10] Helen Bailey, Bridget Senior, Dave Simmons, Jan Rusin, Gordon Picken, and Paul M Thompson. Assessing underwater noise levels during pile-driving at an offshore windfarm and its potential effects on marine mammals. *Marine Pollution Bulletin*, 60(6) :888–897, 2010.
- [Car06] William M Carey. Sound sources and levels in the ocean. *Oceanic Engineering, IEEE Journal of*, 31(1) :61–75, 2006.
- [Cat78] Douglas H Cato. Marine biological choruses observed in tropical waters near australia. *The Journal of the Acoustical Society of America*, 64(3) :736–743, 1978.
- [CH85] William C Cummings and DV Holliday. Passive acoustic location of bowhead whales in a population census off point barrow, alaska. *The Journal of the Acoustical Society of America*, 78(4) :1163–1169, 1985.
- [Coh89] Leon Cohen. Time-frequency distributions-a review. *Proceedings of the IEEE*, 77(7) :941–981, 1989.
- [Coh03] Israel Cohen. Noise spectrum estimation in adverse environments : Improved minima controlled recursive averaging. *Speech and Audio Processing, IEEE Transactions on*, 11(5) :466–475, 2003.
- [Coh04] Israel Cohen. Speech enhancement using a noncausal a priori snr estimator. *Signal Processing Letters, IEEE*, 11(9) :725–728, 2004.
- [CT71] William C Cummings and Paul O Thompson. Underwater sounds from the blue whale, balaenoptera musculus. *The journal of the Acoustical Society of America*, 50(4B) :1193–1198, 1971.
- [Dad14] Florian Dadouchi. False alarm density analysis (fada) : A tool for statistical time-frequency segmentation. 2014.

- [DGI<sup>+</sup>13a] Florian Dadouchi, Cedric Gervaise, Cornel Ioana, Julien Huillery, and Jérôme I Mars. Automated segmentation of linear time-frequency representations of marine-mammal sounds. *The Journal of the Acoustical Society of America*, 134(3) :2546–2555, 2013.
- [DGI<sup>+</sup>13b] Florian Dadouchi, Cedric Gervaise, Cornel Ioana, Julien Huillery, Jerome I Mars, et al. Segmentation temps-fréquence statistique : Analyse de la répartition spatiale des détections sur spectrogramme n fois seuillé. In *24ème édition du colloque Grets*, 2013.
- [DIGJ<sup>+</sup>12] Lucia Di Iorio, Cédric Gervaise, Virginie Jaud, Anthony A Robson, and Laurent Chauvaud. Hydrophone detects cracking sounds : Non-intrusive monitoring of bivalve movement. *Journal of Experimental Marine Biology and Ecology*, 432 :9–16, 2012.
- [DL03] Edward R Dougherty and Roberto A Lotufo. Hands-on morphological image processing. Spie Bellingham, WA, 2003.
- [DR08] Robert J Diaz and Rutger Rosenberg. Spreading dead zones and consequences for marine ecosystems. *science*, 321(5891) :926–929, 2008.
- [DS01] Igor Djurovic and LJubisa Stankovic. Robust wigner distribution with application to the instantaneous frequency estimation. *Signal Processing, IEEE Transactions on*, 49(12) :2985–2993, 2001.
- [EK08] Christine Erbe and Andrew R King. Automatic detection of marine mammals using information entropy. *The Journal of the Acoustical Society of America*, 124(5) :2833–2840, 2008.
- [EKSX96] Martin Ester, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Kdd*, volume 96, pages 226–231, 1996.
- [Erb00] Christine Erbe. Detection of whale calls in noise : Performance comparison between a beluga whale, human listeners, and a neural network. *The Journal of the Acoustical Society of America*, 108(1) :297–303, 2000.
- [EYJ48] F Alton Everest, Robert W Young, and Martin W Johnson. Acoustical characteristics of noise produced by snapping shrimp. *The Journal of the Acoustical Society of America*, 20(2) :137–142, 1948.
- [Faw06] Tom Fawcett. An introduction to roc analysis. *Pattern recognition letters*, 27(8) :861–874, 2006.
- [Fla98] Patrick Flandrin. *Time-frequency/time-scale analysis*, volume 10. Academic Press, 1998.

- [For91] John KB Ford. Vocal traditions among resident killer whales (*orcinus orca*) in coastal waters of british columbia. *Canadian journal of zoology*, 69(6) :1454–1483, 1991.
- [FPSS96] Usama Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth. From data mining to knowledge discovery in databases. *AI magazine*, 17(3) :37, 1996.
- [GBB<sup>+</sup>10] Cedric Gervaise, A Barazzutti, Sylvain Busson, Y Simard, and N Roy. Automatic detection of bioacoustics impulses based on kurtosis under weak signal to noise ratio. *Applied Acoustics*, 71(11) :1020–1026, 2010.
- [GCGW13] Douglas Gillespie, Marjolaine Caillat, Jonathan Gordon, and Paul White. Automatic detection and classification of odontocete whistles. *The Journal of the Acoustical Society of America*, 134(3) :2427–2437, 2013.
- [GG06] Pascale Giraudet and Hervé Glotin. Real-time 3d tracking of whales by echo-robust precise tdoa estimates with a widely-spaced hydrophone array. *Applied Acoustics*, 67(11) :1106–1117, 2006.
- [GGP<sup>+</sup>03] Jonathan Gordon, Douglas Gillespie, John Potter, Alexandros Frantzis, Mark P Simmonds, René Swift, and David Thompson. A review of the effects of seismic surveys on marine mammals. *Marine Technology Society Journal*, 37(4) :16–34, 2003.
- [Har78] Fredric J Harris. On the use of windows for harmonic analysis with the discrete fourier transform. *Proceedings of the IEEE*, 66(1) :51–83, 1978.
- [HBB92] Franz Hlawatsch and G Faye Boudreaux-Bartels. Linear and quadratic time-frequency signal representations. *IEEE Signal Processing Magazine*, 9(2) :21–67, 1992.
- [HE06] Xanadu C Halkias and Daniel PW Ellis. Call detection and extraction using bayesian inference. *Applied Acoustics*, 67(11) :1164–1174, 2006.
- [Hil09] John A Hildebrand. Anthropogenic and natural sources of ambient noise in the ocean. *Marine Ecology Progress Series*, 395(5), 2009.
- [HM06] Julien Huillery and Nadine Martin. Detection of time-frequency components of signals : Bayes and focus snr. In *Seventh International Conference on Mathematics in Signal Processing. IMA 2006*, 2006.
- [HMC02] Cyril Hory, Nadine Martin, and Alain Chehikian. Spectrogram segmentation by means of statistical features for non-stationary signal interpretation. *Signal Processing, IEEE Transactions on*, 50(12) :2915–2925, 2002.

- [HMM08] Julien Huillery, Fabien Millioz, and Nadine Martin. On the description of spectrogram probabilities with a chi-squared law. *Signal Processing, IEEE Transactions on*, 56(6) :2249–2258, 2008.
- [Hui08] Julien Huillery. *Support temps-fréquence d’un signal inconnu en présence de bruit additif gaussien*. These, Institut National Polytechnique de Grenoble - INPG, July 2008. Département Images et Signal Département Images et Signal.
- [HZVDSA10] Ludwig Houéganigan, S Zaugg, M Van Der Schaar, and Michel André. Space–time and hybrid algorithms for the passive acoustic localisation of sperm whales and vessels. *Applied Acoustics*, 71(11) :1000–1010, 2010.
- [IDS03] Veselin N Ivanovic, Milos Dakovic, and Ljubisa Stankovic. Performance of quadratic time–frequency distributions as instantaneous frequency estimators. *Signal Processing, IEEE Transactions on*, 51(1) :77–89, 2003.
- [IGSM10] Cornel Ioana, Cédric Gervaise, Yann Stéphan, and Jérôme I Mars. Analysis of underwater mammal vocalisations using time–frequency–phase tracker. *Applied Acoustics*, 71(11) :1070–1080, 2010.
- [Jan00] Vincent M Janik. Whistle matching in wild bottlenose dolphins (*tursiops truncatus*). *Science*, 289(5483) :1355–1357, 2000.
- [JB95] Douglas L Jones and Richard G Baraniuk. An adaptive optimal–kernel time–frequency representation. *Signal Processing, IEEE Transactions on*, 43(10) :2361–2371, 1995.
- [JGI<sup>+</sup>12] Nicolas Josso, Cedric Gervaise, Cornel Ioana, Jerome I Mars, Yann Stéphan, et al. Dolphin localization and motion estimation with a single hydrophone multipath propagation. In *Proceedings of the 5th International Workshop on Detection, Classification, Localization, and Density Estimation of Marine Mammals using Passive Acoustics*, 2012.
- [Jos10] Nicolas Josso. *Caractérisation des milieux sous marins en utilisant des sources mobiles d’opportunité*. PhD thesis, Institut National Polytechnique de Grenoble-INPG, 2010.
- [JS98] Vincent M Janik and Peter JB Slater. Context-specific use suggests that bottlenose dolphin signature whistles are cohesion calls. *Animal behaviour*, 56(4) :829–838, 1998.
- [JSW06] Vincent M Janik, Laela S Sayigh, and RS Wells. Signature whistle shape conveys identity information to bottlenose dolphins. *Proceedings of the National Academy of Sciences*, 103(21) :8293–8297, 2006.

- [JW11] A Torbjörn Johansson and Paul R White. An adaptive filter-based method for robust, automatic detection and frequency estimation of whistles. *The Journal of the Acoustical Society of America*, 130(2) :893–903, 2011.
- [Kal60] Rudolph Emil Kalman. A new approach to linear filtering and prediction problems. *Journal of Fluids Engineering*, 82(1) :35–45, 1960.
- [Kay98] Steven M Kay. Fundamentals of statistical signal processing, vol. ii : Detection theory. *Signal Processing. Upper Saddle River, NJ : Prentice Hall*, 1998.
- [KGC<sup>+</sup>10] Bazile Kinda, Cedric Gervaise, L Chauvaud, Virginie Jaud, Sylvain Busson, T Robson, et al. " sono-proxy" : Un capteur non intrusif du comportement de la coquille saint-jacques. In *Annales hydrographiques*, volume 7. Service hydrographique et océanographique de la marine, 2010.
- [KSG<sup>+</sup>13] G Bazile Kinda, Yvan Simard, Cédric Gervaise, Jérôme I Mars, and Louis Fortier. Under-ice ambient noise in eastern beaufort sea, canadian arctic, and its relation to environmental forcing. *The Journal of the Acoustical Society of America*, 134(1) :77–87, 2013.
- [KWT88] Michael Kass, Andrew Witkin, and Demetri Terzopoulos. Snakes : Active contour models. *International journal of computer vision*, 1(4) :321–331, 1988.
- [LA03] Marc O Lammers and Whitlow WL Au. Directionality in the whistles of hawaiian spinner dolphins (*stenella longirostris*) : A signal feature to cue direction of movement ? *Marine Mammal Science*, 19(2) :249–264, 2003.
- [LBGBM13] Olivier Le Bot, Cedric Gervaise, Julien Bonnel, and Jerome I Mars. Click trains deinterleaving by rhythm analysis. *TRAITEMENT DU SIGNAL*, 30(3-5) :195–218, 2013.
- [LOB02] PHILLIP S LOBEL. Diversity of fish spawning sounds and the application of passive acoustic monitoring. *Bioacoustics*, 12(2-3) :286–289, 2002.
- [LP46] Donald P Loye and Don A Proudfoot. Underwater noise due to marine life. *The Journal of the Acoustical Society of America*, 18(2) :446–449, 1946.
- [LPA93] Patrick J Loughlin, James W Pitton, and Les E Atlas. Bilinear time-frequency representations : New insights and properties. *Signal Processing, IEEE Transactions on*, 41(2) :750–767, 1993.



- [Mar70] William D Mark. Spectral analysis of the convolution and filtering of non-stationary stochastic processes. *Journal of Sound and Vibration*, 11(1) :19–63, 1970.
- [Mar94] Rainer Martin. Spectral subtraction based on minimum statistics. *power*, 6 :8, 1994.
- [Mar01] Rainer Martin. Noise power spectral density estimation based on optimal smoothing and minimum statistics. *Speech and Audio Processing, IEEE Transactions on*, 9(5) :504–512, 2001.
- [MB05] Herman Medwin and Joseph E Blue. *Sounds in the sea : From ocean acoustics to acoustical oceanography*. Cambridge University Press, 2005.
- [MC00] David K Mellinger and Christopher W Clark. Recognizing transient low-frequency whale sounds by spectrogram correlation. *The Journal of the Acoustical Society of America*, 107(6) :3518–3529, 2000.
- [McC95] Brenda McCowan. A new quantitative technique for categorizing whistles using simulated signals and whistles from captive bottlenose dolphins (delphinidae, tursiops truncatus). *Ethology*, 100(3) :177–193, 1995.
- [Mel04] David K Mellinger. A comparison of methods for detecting right whale calls. *Canadian Acoustics*, 32(2) :55–65, 2004.
- [Mel13] David Mellinger. Conditioning for marine bioacoustic signal detection and classification. In *Proceedings of Meetings on Acoustics*, volume 19, page 010017. Acoustical Society of America, 2013.
- [MF85] Wolfgang Martin and Patrick Flandrin. Wigner-ville spectral analysis of nonstationary processes. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 33(6) :1461–1470, 1985.
- [MF02] AA Myrberg and Lee A Fuiman. The sensory world of coral reef fishes. *Coral reef fishes : dynamics and diversity in a complex ecosystem*. Academic, San Diego, pages 123–148, 2002.
- [MKH07] Douglas Maraun, J Kurths, and M Holschneider. Nonstationary gaussian processes in wavelet domain : Synthesis, estimation, and significance testing. *Physical Review E*, 75(1) :016707, 2007.
- [MM11] Fabien Millioz and Nadine Martin. Circularity of the stft and spectral kurtosis for time-frequency segmentation in gaussian environment. *Signal Processing, IEEE Transactions on*, 59(2) :515–524, 2011.
- [MOCT08] Asitha Mallawaarachchi, SH Ong, Mandar Chitre, and Elizabeth Taylor. Spectrogram denoising and automated extraction of the fun-

- damental frequency variation of dolphin whistles. *The Journal of the Acoustical Society of America*, 124(2) :1159–1170, 2008.
- [MR95] Brenda McCowan and Diana Reiss. Quantitative comparison of whistle repertoires from captive adult bottlenose dolphins (delphinidae, *tursiops truncatus*) : a re-evaluation of the signature whistle hypothesis. *Ethology*, 100(3) :194–209, 1995.
- [MR01] Brenda McCowan and Diana Reiss. The fallacy of ‘signature whistles’ in bottlenose dolphins : a comparative perspective of ‘signature information’ in animal vocalizations. *Animal Behaviour*, 62(6) :1151–1162, 2001.
- [MSM<sup>+</sup>07] David K Mellinger, Kathleen M Stafford, S Moore, Robert P Dziak, and Haru Matsumoto. Fixed passive acoustic observation methods for cetaceans. *Oceanography*, 20(4) :36, 2007.
- [MTM<sup>+</sup>13] Tiago A Marques, Len Thomas, Stephen W Martin, David K Mellinger, Jessica A Ward, David J Moretti, Danielle Harris, and Peter L Tyack. Estimating animal population density using passive acoustics. *Biological Reviews*, 88(2) :287–309, 2013.
- [MTS<sup>+</sup>13] Delphine Mathias, Aaron Thode, Jan Straley, Russ Andrews, Olivier Le Bot, Cédric Gervaise, Jerome I Mars, et al. Range-depths tracking of multiple sperm whales over large distances using a two-element vertical array and rhythmic properties of clicks-trains. In *Workshop : Neural Information Processing Scaled for Bioacoustics : NIPS4B*, 2013.
- [MTW<sup>+</sup>09] Tiago A Marques, Len Thomas, Jessica Ward, Nancy DiMarzio, and Peter L Tyack. Estimating cetacean population density using fixed passive acoustic sensors : an example with blainville’s beaked whales. *The Journal of the Acoustical Society of America*, 125(4) :1982–1994, 2009.
- [MWD<sup>+</sup>06] RP Morrissey, J Ward, N DiMarzio, S Jarvis, and DJ Moretti. Passive acoustic detection and localization of sperm whales (*physeter macrocephalus*) in the tongue of the ocean. *Applied Acoustics*, 67(11) :1091–1105, 2006.
- [NM83] Kenneth S Norris and Bertel Mohl. Can odontocetes debilitate prey with sound? *American Naturalist*, pages 85–104, 1983.
- [NP92] Jerzy Neyman and Egon S Pearson. *On the problem of the most efficient tests of statistical hypotheses*. Springer, 1992.
- [NSST13] Steven Ness, Helena Symonds, Paul Spong, and George Tzanetakis. The orchive : Data mining a massive bioacoustic archive. *arXiv preprint arXiv :1307.0589*, 2013.

- [OBN03] Julie N Oswald, Jay Barlow, and Thomas F Norris. Acoustic identification of nine delphinid species in the eastern tropical pacific ocean. *Marine Mammal Science*, 19(1) :20–037, 2003.
- [ORBL07] Julie N Oswald, Shannon Rankin, Jay Barlow, and Marc O Lamers. A tool for real-time acoustic species identification of delphinid whistles. *The Journal of the Acoustical Society of America*, 122(1) :587–595, 2007.
- [Pil83] Georg Pilleri. The sonar system of the dolphins. *Endeavour*, 7(2) :59–64, 1983.
- [RBP<sup>+</sup>11] Marie A Roch, T Scott Brandes, Bhavesh Patel, Yvonne Barkley, Simone Baumann-Pickering, and Melissa S Soldevilla. Automated extraction of odontocete whistle contours. *The Journal of the Acoustical Society of America*, 130(4) :2212–2223, 2011.
- [RD01] Christophe Ris and Stéphane Dupont. Assessing local noise level estimation methods : Application to noise robust asr. *Speech Communication*, 34(1) :141–158, 2001.
- [RL06] Sundarrajan Rangachari and Philipos C Loizou. A noise-estimation algorithm for highly non-stationary environments. *Speech communication*, 48(2) :220–231, 2006.
- [RSB<sup>+</sup>07] Marie A Roch, Melissa S Soldevilla, Jessica C Burtenshaw, E Elizabeth Henderson, and John A Hildebrand. Gaussian mixture model classification of odontocetes in the southern california bight and the gulf of california. *The Journal of the Acoustical Society of America*, 121(3) :1737–1748, 2007.
- [RSG10] Nathalie Roy, Yvan Simard, and Cédric Gervaise. 3d tracking of foraging belugas from their clicks : experiment from a coastal hydrophone array. *Applied Acoustics*, 71(11) :1050–1056, 2010.
- [SCW<sup>+</sup>10] Ann Stokes, Kevin Cockrell, Joshua Wilson, Dwight Davis, and David Warwick. Mitigation of underwater pile driving noise during offshore construction : Final report. *Applied Physical Sciences Corp*, 2010.
- [SDT13] Ljubiša Stanković, Miloš Daković, and Thayannathan Thayaparan. *Time-frequency signal analysis with applications*. Artech House, 2013.
- [Ser86] Jean Serra. Introduction to mathematical morphology. *Computer vision, graphics, and image processing*, 35(3) :283–305, 1986.
- [SFB00] Volker Stahl, Alexander Fischer, and Rolf Bippus. Quantile based noise estimation for spectral subtraction and wiener filtering. In

- Acoustics, Speech, and Signal Processing, 2000. ICASSP'00. Proceedings. 2000 IEEE International Conference on*, volume 3, pages 1875–1878. IEEE, 2000.
- [SHC+08] Melissa S Soldevilla, E Elizabeth Henderson, Gregory S Campbell, Sean M Wiggins, John A Hildebrand, and Marie A Roch. Classification of risso's and pacific white-sided dolphins using spectral properties of echolocation clicks. *The Journal of the Acoustical Society of America*, 124(1) :609–624, 2008.
- [SHSM98] Jorge L Sarmiento, Tertia MC Hughes, Ronald J Stouffer, and Syukuro Manabe. Simulated response of the ocean carbon cycle to anthropogenic climate warming. *Nature*, 393(6682) :245–249, 1998.
- [SIP96] LJubisa Stanković, Veselin Ivanović, and Zoran Petrović. Unified approach to noise analysis in the wigner distribution and spectrogram. In *Annales des télécommunications*, volume 51, pages 585–594. Springer, 1996.
- [SMM+05] Stephen D Simpson, Mark Meekan, John Montgomery, Rob McCauley, and Andrew Jeffs. Homeward sound. *Science*, 308(5719) :221–221, 2005.
- [SOS12] Srdjan Stanković, Irena Orović, and Victor Sucic. Averaged multiple l-spectrogram for analysis of noisy nonstationary signals. *Signal Processing*, 92(12) :3068–3074, 2012.
- [SRG+10] Yvan Simard, Nathalie Roy, Samuël Giard, Cedric Gervaise, Manuela Conversano, and Nadia Ménard. Estimating whale density from their whistling activity : Example with st. lawrence beluga. *Applied Acoustics*, 71(11) :1081–1086, 2010.
- [SS95] Ljubisa Stankovic and Srdjan Stankovic. An analysis of instantaneous frequency representation using time-frequency distributions-generalized wigner distribution. *Signal Processing, IEEE Transactions on*, 43(2) :549–552, 1995.
- [Sta02] L Stankovic. Analysis of noise in time-frequency distributions. *IEEE Signal Processing Letters*, 9(9) :286–289, 2002.
- [Ste81] William W Steiner. Species-specific differences in pure tonal whistle vocalizations of five western north atlantic dolphin species. *Behavioral Ecology and Sociobiology*, 9(4) :241–246, 1981.
- [SW07] Gerry E Studds and Andrew J Wright. A brief review of anthropogenic sound in the oceans. *International Journal of Comparative Psychology*, 20(2), 2007.
- [Syl02] Julien Sylvestre. Time-frequency detection algorithm for gravitational wave bursts. *Physical Review D*, 66(10) :102004, 2002.

- [Tho95] Denis H Thomson. *Marine mammals and noise*. Gulf Professional Publishing, 1995.
- [TKB<sup>+</sup>12] Aaron M Thode, Katherine H Kim, Susanna B Blackwell, Charles R Greene Jr, Christopher S Nations, Trent L McDonald, and A Michael Macrander. Automated detection and localization of bowhead whale sounds in the presence of seismic airgun surveys. *The Journal of the Acoustical Society of America*, 131(5) :3726–3747, 2012.
- [TLKP06] Frank Thomsen, Karin Lüdemann, Rudolf Kafemann, and Werner Piper. Effects of offshore wind farm noise on marine mammals and fish. *Biola, Hamburg, Germany on behalf of COWRIE Ltd*, page 62, 2006.
- [TMG08] Jérémy Terrien, Catherine Marque, and Guy Germain. Ridge extraction from the time–frequency representation (tfr) of signals based on an image processing approach : application to the analysis of uterine electromyogram ar tfr. *Biomedical Engineering, IEEE Transactions on*, 55(5) :1496–1503, 2008.
- [TZM<sup>+</sup>11] Peter L Tyack, Walter MX Zimmer, David Moretti, Brandon L Southall, Diane E Claridge, John W Durban, Christopher W Clark, Angela D’Amico, Nancy DiMarzio, Susan Jarvis, et al. Beaked whales respond to simulated and actual navy sonar. *PLoS One*, 6(3) :e17009, 2011.
- [UCKP09] Ildar R Urazghildiiev, Christopher W Clark, Timothy P Krein, and Susan E Parks. Detection and recognition of north atlantic right whale contact calls in the presence of ambient noise. *Oceanic Engineering, IEEE Journal of*, 34(3) :358–368, 2009.
- [VSvdHL00] Michel Versluis, Barbara Schmitz, Anna von der Heydt, and Detlef Lohse. How snapping shrimp snap : through cavitating bubbles. *Science*, 289(5487) :2114–2117, 2000.
- [WBB<sup>+</sup>06] Boris Worm, Edward B Barbier, Nicola Beaumont, J Emmett Duffy, Carl Folke, Benjamin S Halpern, Jeremy BC Jackson, Heike K Lotze, Fiorenza Micheli, Stephen R Palumbi, et al. Impacts of biodiversity loss on ocean ecosystem services. *science*, 314(5800) :787–790, 2006.
- [Wei07] Lindy S Weilgart. The impacts of anthropogenic ocean noise on cetaceans and implications for management. *Canadian Journal of Zoology*, 85(11) :1091–1116, 2007.
- [Wen62] Gordon M Wenz. Acoustic ambient noise in the ocean : spectra and sources. *The Journal of the Acoustical Society of America*, 34(12) :1936–1956, 1962.

- [WGJJ00] B Würsig, CR Greene Jr, and TA Jefferson. Development of an air bubble curtain to reduce underwater noise of percussive piling. *Marine Environmental Research*, 49(1) :79–93, 2000.
- [WH07] Sean M Wiggins and John A Hildebrand. High-frequency acoustic recording package (harp) for broad-band, long-term marine mammal monitoring. In *Underwater Technology and Workshop on Scientific Use of Submarine Cables and Related Technologies, 2007. Symposium on*, pages 551–557. IEEE, 2007.
- [XF07] Jun Xiao and Patrick Flandrin. Multitaper time-frequency reassignment for nonstationary spectrum estimation and chirp enhancement. *Signal Processing, IEEE Transactions on*, 55(6) :2851–2860, 2007.
- [YR06] Chunghsin Yeh and Axel Röbel. Adaptive noise level estimation. In *Proc. of the 9th Int. Conf. on Digital Audio Effects*, pages 145–148, 2006.
- [ZK13] Alexander Zaigraev and Serguei Kaniovski. A note on the probability of at least successes in correlated binary trials. *Operations Research Letters*, 41(1) :116 – 120, 2013.
- [ZMP99] Randy Zelick, David A Mann, and Arthur N Popper. Acoustic communication in fishes and frogs. In *Comparative hearing : fish and amphibians*, pages 363–411. Springer, 1999.
- [ZVDSH<sup>+</sup>10] Serge Zaugg, Mike Van Der Schaar, Ludwig Houégnigan, Cédric Gervaise, and Michel André. Real-time acoustic classification of sperm whale clicks and shipping impulses from deep-sea observatories. *Applied Acoustics*, 71(11) :1011–1019, 2010.







---

**RÉSUMÉ** - Les océans subissent des pressions d'origine anthropique particulièrement fortes comme la surpêche, la pollution physico-chimique, et le bruit rayonné par les activités industrielles et militaires. Cette thèse se place dans un contexte de compréhension de l'impact du bruit rayonné dans les océans sur les mammifères marins. L'acoustique passive joue donc un rôle fondamental dans ce problème. Ce travail aborde la tâche de détection de signatures acoustiques de mammifères marins dans le spectrogramme. Cette tâche est difficile pour deux raisons : **1.** le bruit océanique a une structure complexe (non-stationnaire, coloré), **2.** les signaux de mammifères marins sont inconnus et possèdent eux aussi une structure complexe (non-stationnaires bande étroite et/ou impulsions). Le problème doit donc être résolu de manière locale en temps-fréquence, et ne pas faire d'hypothèse *a priori* sur le signal.

Des détecteurs statistiques basés uniquement sur la connaissance des statistiques du bruit dans le spectrogramme existent, mais souffrent de deux lacunes : **1.** leurs performances en terme de probabilité de fausse alarme/ probabilité de détection se dégradent fortement à faible rapport signal à bruit, et **2.** ils ne sont pas capables de séparer les signaux à bande étroite des signaux impulsifs. Ce travail apporte des pistes de réflexion sur ces problèmes.

L'originalité de ce travail de thèse repose dans la formulation d'un test d'hypothèse binaire prenant explicitement en compte l'organisation spatiale des pics temps-fréquence. Nous introduisons une méthode d'Analyse de la Densité des Fausses Alarmes (FADA) qui permet de discriminer les régions temps-fréquence abritant le signal de celles n'abritant que du bruit. Plus précisément, le nombre de fausses alarmes dans une région du plan est d'abord modélisé par une loi binomiale, puis par une loi binomiale corrélée, afin de prendre en considération la redondance du spectrogramme. Le test d'hypothèse binaire est résolu par une approche de Neyman-Pearson.

Nous démontrons numériquement la pertinence de cette approche et nous la validons sur données réelles de mammifères marins disposant d'une grande variété de signaux et de conditions de bruit. En particulier, nous illustrons la capacité de FADA à discriminer efficacement le signal du bruit en milieu fortement impulsif.

**Mots-clés :** Acoustique passive, représentation temps-fréquence, segmentation statistique, loi du  $\chi^2$ , loi binomiale, loi binomiale corrélée, analyse de la densité des fausses alarmes.

---

**ABSTRACT** - The oceans experience heavy anthropogenic pressure due to overfishing, physico-chemical pollution, and noise radiated by industrial and military activities. This work focuses on the use of passive acoustic monitoring of the oceans, as a tool to understand the impact of radiated noise on marine ecosystems, and particularly on marine mammals. This work tackles the task of detection of acoustical signals of marine mammals using the spectrogram. This task is uneasy for two reasons : **1.** the ocean noise structure is complex (non-stationary and colored) and **2.** the signals of interest are unknown and also shows a complex structure (non-stationary narrow band and/or impulsive). The problem therefore must be solved locally without making *a priori* hypothesis on the signal.

Statistical detectors only based on the local analysis of the noise spectrogram coefficients are available, making them suitable for this problem. However, these detectors suffer two disadvantages : **1.** the trade-offs false alarm probability/ detection probability that are available for low signal to noise ratio are not satisfactory and **2.** the separation between narrow-band and impulsive signals is not possible. This work brings some answers to these problems.

The main contribution of this work is to formulate a binary hypothesis test taking explicitly in account the spatial organization of time-frequency peaks. We introduce the False Alarm Density Analysis (FADA) framework that is efficiently discriminates time-frequency regions hosting signal from the ones hosting noise only. In particular the number of false alarms in regions of the binary spectrogram is first modeled by a binomial distribution, and then by a correlated binomial distribution to take in account the spectrogram redundancy. The binary hypothesis test is solved using a Neyman-Pearson criterion.

We demonstrate the relevance of this approach on simulated data and validate the FADA detector on a wide variety of real signals. In particular we show the capability of the proposed method to efficiently detect signals in highly impulsive environment.

**Key words :** Passive acoustic monitoring, time-frequency representation, statistical segmentation,  $\chi^2$  distribution, binomial and correlated binomial distribution, false alarm density analysis.

---