UNIVERSITÉ DU QUÉBEC

DÉTECTION ET IDENTIFICATION AUTOMATIQUE EN TEMPS-RÉEL DES VOCALISES DE RORQUAL BLEU (*BALAENOPTERA MUSCULUS*) ET DE RORQUAL COMMUN (*BALAENOPTERA PHYSALUS*) DANS L'ESTUAIRE DU SAINT-LAURENT

MÉMOIRE
PRÉSENTÉ À
L'UNIVERSITÉ DU QUÉBEC À RIMOUSKI
COMME EXIGENCE PARTIELLE
DU PROGRAMME DE MAÎTRISE EN OCÉANOGRAPHIE

PAR
XAVIER MOUY

OCTOBRE 2007

Remerciements

Au terme de ce travail, je tiens à remercier mon directeur de recherche, le Dr Yvan Simard (ISMER, UQAR) et mon co-directeur le Dr Mohammed Bahoura (DMIG, UQAR) pour m'avoir donner l'opportunité d'effectuer ce projet de recherche. Avec sa disponibilité, sa pédagogie et son sens de l'humour, Yvan a su me transmettre sa passion pour l'acoustique sous-marine. Il m'a permis de mieux connaître le monde de la recherche en me donnant l'opportunité de participer à des congrès scientifiques. Je lui suis très reconnaissant de m'avoir offert plusieurs contrats de recherche, qui m'ont permis d'élargir mes connaissances dans différents domaines de l'acoustique. Mohammed a su m'aider avec une grande grande patience pour l'apprentissage des méthodes de traitement du signal qui m'ont servi au cours de ma maîtrise. Je ne saurais assez le remercier pour ses nombreux encouragements et pour m'avoir remonté le moral dans les périodes difficiles.

Je tiens à remercier mon président du Jury, le Dr Jean-François Dumais (ISMER, UQAR) et mon correcteur externe, le Dr Jean Rouat (GEGI, Université de Sherbrooke) pour avoir aimablement accepté de lire et de corriger ce travail.

Merci à Catherine Bédard, Marc Sourisseau, Richard Lepage, Nathalie Roy et Pierre Saint-Laurent pour toutes les discussions fructueuses et stimulantes au cours desquelles ils ont su me faire partager leurs compétences.

Je désire exprimer ma reconnaissance à l'Association Canadienne d'Acoustique qui m'a gratifié du Prix Fessenden 2006 pour mon projet de maîtrise.

Je remercie ma famille qui m'a épaulé et conseillé au cours de ces années passées loin

d'eux. Merci à tous mes amis qui m'ont soutenu de près ou de loin. Un merci spécial à Pierre, Marc, Emma, Julien, Catherine, Aurélie et Karine pour ces bières au Baro, ces *raclées* au badminton et ces soirées inoubliables.

Et enfin je ne remercierai jamais assez Héloïse pour m'avoir écouté, encouragé, et soutenu.

Cette recherche à été rendue possible grâce au support financier de la chaire de recherche de Pêches et Océans Canada en acoustique marine appliquée aux ressources et à l'écosystème à l'Institut des Sciences de la Mer de Rimouski (ISMER). Une aide financière a également été accordée par l'ISMER, Québec-Océan et l'Association Canadienne d'Acoustique.

Résumé

La détection et l'identification automatique des vocalises d'animaux est un outil utile pour documenter leur distribution saisonnière, leur abondance relative ainsi que leur comportement dans leur habitat naturel. La performance des méthodes de traitement du signal utilisées est cependant dépendante du type de vocalisations (bande de fréquence, variabilité du patron temps-fréquence) et des caractéristiques environnementales (bruit, effets de propagation sonore). Ce projet de recherche compare plusieurs méthodes de détection et d'identification dans le domaine temps-fréquence, des vocalises de rorquals bleus (*Balaenoptera musculus*) et de rorquals communs (*Balaenoptera physalus*) dans le Saint-Laurent. Trois des vocalises de ces balaenoptéridés sont des patrons réguliers d'infrasons ($<$ 30 Hz) stéréotypés et une autre est de fréquence plus élevée (30–110 Hz), irrégulière et variable à la fois en fréquence et en durée (1-5 s). À cause du trafic maritime important, des caractéristiques bathymétriques et physiques de la Voie maritime du Saint-Laurent, la majorité des vocalises sont polluées par du bruit intense dans les basses fréquences et étirées en temps par les trajets multiples. Toutes les méthodes commencent par le calcul du spectrogramme puis d'une étape de réduction du bruit basée sur des techniques de traitement d'image. Ensuite la première approche consiste à binariser le spectrogramme et à calculer la coïncidence avec un modèle temps-fréquence binarisé, via une opération logique *AND*. La seconde approche consiste à sélectionner les maxima locaux à chaque pas de temps du spectrogramme et à extraire les contours temps-fréquence des vocalises en utilisant un algorithme de suivi. Ensuite deux méthodes de reconnaissance sont testées pour classifier ces contours, la déformation temporelle dynamique (DTW) et la quantifi-

cation vectorielle (VQ). Les taux de faux négatifs et de faux positifs sont évalués sur une série de plus de 2000 vocalises extraites d'enregistrements continus collectés dans l'aire d'étude. La méthode de coïncidence des spectrogrammes se trouve être plus performante pour les vocalises stéréotypées (vocalises A, B et 20 Hz), tandis que l'approche par extraction de contours s'avère être plus performante pour la vocalise variable. L'interprétation des indices de performance selon le contexte rythmique des vocalises montre que toutes ces méthodes sont utilisables pour un suivi d'animaux (*Monitoring*). Finalement, l'approche par extraction de contours montre un potentiel intéressant pour la détection et l'identification de vocalises de mammifères marins et est adaptable à différents types de vocalises.

Table des matières

Liste des tableaux

Liste des figures

Liste des abréviations

ESA	: *Endangered Species Act*	: Loi sur les espèces en péril (USA)
MFCC	: *Mel Frequency Cepstral Coefficients*	: Coefficients Cepstraux à échelle de Mel
DTW	: *Dynamic Time Warping*	: Déformation temporelle dynamique
VQ	: *Vector quantization*	: Quantification vectorielle
FFT	: *Fast Fourier Transform*	: Transformée de Fourier rapide
RSB	: Rapport Signal sur Bruit	
COSEPAC	: Comité sur la Situation des Espèces en Péril Au Canada	
LBG	: *Linde-Buzo-Gray*	

Liste des notations

Calcul du spectrogramme

F_s	: Fréquence d'échantillonnage
s	: Signal acoustique échantillonné
m	: Pas de temps de l'échantillonnage
w	: Fenêtre de pondération
K	: Nombre d'échantillons de la fenêtre w
S	: Spectrogramme original
n	: Indice des segments en temps du spectrogramme
N	: Nombre de segments en temps du spectrogramme
k	: Indice de fréquence du spectrogramme
L	: Pas d'avancement des segments du spectrogramme

Réduction du bruit

S	: Spectrogramme original
$\overline{S}$	: Spectrogramme S moyenné en temps
Δt	: Longueur de la fenêtre mobile d'égalisation
S_{eq}	: Spectrogramme égalisé
G	: Masque bidimensionnel gaussien
l	: Nombre de lignes du masque gaussien G

c : Nombre de colonnes du masque gaussien G

S_{lis} : Spectrogramme lissé

$\overline{e}$: Énergie moyenne du spectrogramme S_{lis} dans le temps

$\overline{e}_{lis}$: Énergie moyenne $\overline{e}$ lissée par une moyenne mobile

d_1 : Longueur de la fenêtre mobile pour le lissage de $\overline{e}$

T_1 : Seuil adaptatif 1 (seuil haut)

δ_1 : Constante pour la définition du seuil T_1

T_2 : Seuil adaptatif 2 (seuil bas)

δ_2 : Constante pour la définition du seuil T_2

E : Fonction intermédiaire pour la définition des seuils T_1 et T_2

λ : Facteur d'adaptivité de la fonction E

T_3 : Seuil adaptatif 3

σ : Écart-type des valeurs du spectrogramme S_{lis} à chaque pas de temps n

δ_3 : Constante pour la définition du seuil T_3

$\widehat{S}$: Spectrogramme débruité

Détection par coïncidence des spectrogrammes

f_1 : Fréquence de début du modèle

f_2 : Fréquence de fin du modèle

D_{voc} : Durée du modèle

Δf : Étalement fréquentiel du modèle

D_{ini} : Durée qui précède et succède l'image du modèle

T_{cs} : Seuil de détection

Extraction des contours

d_2 : Longueur de la fenêtre mobile pour le lissage des fragments

$\boldsymbol{x}_i$: Paire de fragments connus suffisamment proches en temps

T_{seg} : Durée maximale entre les fragments des paires $\boldsymbol{x_i}$

$\boldsymbol{\alpha_i}$: Vecteur contenant les pentes des bouts des fragments de $\boldsymbol{x_i}$

β_i : Distance de cassure des fragments de $\boldsymbol{x_i}$

L_α : Série d'observations du paramètre α

L_β : Série d'observations du paramètre β

Θ_α : Distribution normale associée à la série d'observations L_α

Σ_α : Matrice de covariance de la série d'observations L_α

$\boldsymbol{\mu_\alpha}$: Vecteur des moyennes de la série d'observations L_α

Θ_β : Distribution normale associée à la série d'observations L_β

σ_β : Écart-type de la série d'observations L_β

μ_β : Moyenne de la série d'observations L_β

x : Paire de fragments inconnus suffisamment proches en temps

P : Vraisemblance d'une connexion de fragments inconnus

T_c : Valeur minimale de P pour connecter deux fragments

T_{min} : Durée minimale tolérée pour les contours extraits

T_{max} : Durée maximale tolérée pour les contours extraits

Reconnaissance par DTW

T : Vocalise inconnue

R_k : Modèles de référence

k : Indice du modèle de référence

f_{inst} : Fréquence instantanée d'une vocalise

v : Vitesse de la fréquence instantanée

a : Accélération de la fréquence instantanée

I : Nombre de trames de temps de la vocalise T

J_k : Nombre de trames de temps de la vocalise de référence R_k

X : Matrice des distances séparant T de R_k

w : Parcourt minimal

c : Point du parcourt w

C : Nombre de points du parcourt w

g : Fonction de pondération

N : Facteur de normalisation

$\tilde{D}$: Distance accumulée

D_w : Distance du parcourt w

RA : Paramètre délimitant la zone de recherche de w dans la matrice X

T_{DTW} : Seuil d'identification

Reconnaissance par VQ

Q : Processus d'apprentissage

E : Espace des descripteurs

k : Indice de la classe de vocalise

N_k : Nombre d'exemplaires de vocalises pour une classe k

X_k : Ensemble de points représentant les vocalises d'une classe k dans l'espace E

M : Nombre de régions dans l'espace E

C_k : Ensemble des centroïdes d'une classe k

T : Vocalise inconnue

D : Distance séparant une vocalise T de C_k

T_{VQ} : Seuil d'identification

Évaluation de la performance

TFN : Taux de faux négatifs

N_d : Nombre de vocalises oubliées

N_{voc} : Nombre total de vocalises de la base de données

P_{voc} : Puissance d'une vocalise

P_{bruit} : Puissance du bruit au voisinage de la vocalise

T_{voc} : Durée d'une vocalise

I_r : Indice de temps-réel

T_p : Durée du processus de reconnaissance d'un enregistrement

T_r : Durée d'un enregistrement

Chapitre 1

Introduction

Le rorqual bleu, *Balaenoptera musculus* (aussi appelé baleine bleue) et le rorqual commun, *Balaenoptera physalus*, font partie de la famille des Balaenopetridés du sous-ordre des *mysticetes* (baleines à fanons) (Linnæus, 1758 ; tableau 1.1). Trois sous-espèces de rorquals bleus sont reconnues : *B. m. musculus* (Linnæus, 1758) dans l'hémisphère nord, *B. m. intermedia* de l'Antarctique (*e.g.* Rankin *et al.*, 2005) et *B. m. brevicauda*, aussi connue sous le nom de baleine bleue pygmée (Ichihara, 1966), dans la zone subantarctique du sud de l'Océan Indien et au sud-ouest de l'Océan Pacifique (*e.g.* Stafford *et al.*, 2004). Pour le rorqual commun deux sous-espèces ont été identifiées : *B. p. physalus* (Linnæus, 1758) dans l'Atlantique Nord et *B. p. quoyi* (Fisher, 1829), dans les océans du sud. Le rorqual bleu et le rorqual commun sont cosmopolites et se retrouvent dans la quasi-totalité des océans (McDonald *et al.*, 2006a, mais non dans l'Océan Arctique). Ils sont tous deux classés *espèces menacées* par l'ESA (*Endangered Species Act*, U.S.A.). Les populations Nord-Atlantique sont classées par le COSEPAC (Comité sur la Situation des Espèces en Péril Au Canada) comme *en voie de disparition*, pour le rorqual bleu et *espèce préoccupante* pour le rorqual commun (Sears et Calambokidis, 2002; COSEPAC, 2004). La principale raison de ces statuts est la baisse considérable des stocks mondiaux engendrée par la chasse à la baleine intensive menée au cours des siècles précédents. Roman et Palumbi (2003) estiment que la population de rorquals communs de l'Atlantique Nord avant la chasse à

Tableau 1.1 – Taxonomie du rorqual bleu et du rorqual commun. Les sous-espèces qui fréquentent le Saint-Laurent sont en caractères gras (tiré de NOAA : http ://www.afsc.noaa.gov/nmml/education/)

	Rorqual bleu	Rorqual commun
Règne	Animal	Animal
Embranchement	Chordés	Chordés
Sous-embranchement	Vertébrés	Vertébrés
Classe	Mammifères	Mammifères
Sous-classe	Thériens	Thériens
Infra-classe	Placentaires	Placentaires
Ordre	Cétacés	Cétacés
Sous-ordre	Mysticètes	Mysticètes
Famille	Balaenopteridés	Balaenopteridés
Genre	*Balaenoptera*	*Balaenoptera*
Espèce	*musculus*	*physalus*
Sous-espèces	***musculus***	***physalus***
	intermedia	*quoyi*
	brevicauda	

la baleine était six fois supérieure à aujourd'hui. Les stocks actuels de l'Atlantique Nord-Ouest sont évalués selon Perry *et al.* (1999) entre 100 et 600 individus pour le rorqual bleu et entre 3500 et 6300 pour le rorqual commun. Si l'occurrence des rorquals dans certaines régions du monde semble augmenter (Nord-Est Pacifique), on ignore si cela est dû à un accroissement du stock ou à une modification dans la répartition des espèces (Calambokidis *et al.*, 1990; Barlow, 1995).

Actuellement, on peut citer deux menaces potentielles pour la survie de ces espèces : l'interaction directe avec les navires et les divers bruits sous-marins d'origine anthropique (Reeves *et al.*, 1998; Anonyme, 2006). Laist *et al.* (2001) ont mis en avant l'importance des collisions sur les populations de rorquals communs. Cette étude montre que les collisions avec les rorquals communs sont les plus fréquentes et dans certaines régions sont la cause d'environ un tiers des échouages. Jensen et Silber (2003) montrent également que parmi toutes les régions du monde recensées, la côte Nord-Atlantique du Canada et des États-Unis est celle qui compte le plus grand nombre de collisions. Ils recensent également

que les navires de marchandise (cargos, bateaux conteneurs, etc.) et les embarcations de tourisme pour l'observation des baleines (*Whale watching*) sont impliqués dans la majorité de ces collisions. Ces embarcations d'observation des baleines sont également soupçonnées de créer un stress sur les mammifères marins, amenant une perturbation du comportement (*e.g.* changements dans le rythme des plongées et dans la durée des respirations en surface, etc.) (Jahoda *et al.*, 2003). D'autre part, la contribution humaine au bruit ambiant dans les océans a augmenté depuis les 50 dernières années et est dominée par les sons de basses fréquences des navires, le développement pétrolier et les activités de recherche menées par l'armée (Andrew *et al.*, 2002; McDonald *et al.*, 2006b). Il est possible que les mysticètes soient affectés par cette pollution sonore car eux-même produisent et probablement perçoivent des sons dans les basses fréquences (Richardson *et al.*, 1995; National Research Council, 2003). Les effets du bruit sur l'audition, la communication et le comportement des baleines bleues et rorquals communs ne sont pas bien estimés. Bien que l'exposition ponctuelle à des bruits intenses en basses fréquences (sonars militaires de basses fréquences) ne semble pas avoir d'impact majeur, les effets d'une exposition à long terme sont inconnus (Croll *et al.*, 2001).

L'estuaire et le golfe du Saint-Laurent forment un habitat critique pour l'alimentation et la survie des baleines du Nord-Ouest Atlantique qui est exposé à un bon nombre des activités anthropiques préoccupantes pour la survie des baleines citées ci-dessus. Ce garde-manger traditionnel des baleines à la tête du chenal principal du continent (Simard et Lavoie, 1999) est aussi un très important site d'observation des baleines (Hoyt, 2001). Le Saint-Laurent est également une voie de trafic maritime très intense permettant la liaison entre les ports de l'Atlantique et l'intérieur du continent. Le site d'alimentation intensive est situé à un carrefour de la voie maritime (à la tête du chenal Laurentien). L'impact que ces activités anthropiques pourraient avoir sur la survie des rorquals bleus et des rorquals communs qui fréquentent la zone est fort mal connu en raison du manque de connaissances précises sur la fréquentation dans le temps et dans l'espace de ces animaux. Afin de mieux les comprendre et les protéger, il est donc nécessaire de mettre au point des méthodes efficaces permettant d'observer ces mammifères en continu 24 heures sur

24, pendant de longues périodes, sur tout le bassin et de façon non intrusive.

La méthode la plus ancienne pour le monitorage de mammifères marins est l'observation visuelle. Cette approche consiste à identifier et compter les mammifères marins lors de leur remontée en surface à partir d'un bateau, d'un avion ou d'une station côtière. Cette méthode, encore très utilisée de nos jours, comporte certains inconvénients. Tout d'abord, le biais lors du comptage et de l'identification n'est pas constant, il est dépendant de l'observateur et de son état de fatigue. Ensuite, les observations sont dépendantes des périodes d'apparition des baleines en surface (10% du temps pour les rorquals bleus et communs ; Lagerquist *et al.* (2000), Goldbogen *et al.* (2006)), des conditions météorologiques (brume, vent, vagues) et des conditions de luminosité (jour/nuit) (Costa, 1996). Enfin, la présence d'un bateau ou d'un avion peut perturber le comportement des animaux observés et ainsi biaiser les résultats d'observation (Salvado *et al.*, 1992).

L'avancée dans la miniaturisation des composants électroniques a permis le développement de balises géoréférencées (Johnson et Tyack, 2003). Ces balises, attachées sur l'animal par une ventouse, contiennent différents senseurs permettant par exemple de mesurer la profondeur, l'orientation et les vocalises de l'animal durant la plongée. Les données peuvent être stockées sur un disque dur embarqué ou transférées via ondes radios ou par satellite à une station côtière. La pose de balises permet notamment de suivre avec précision la plongée d'une baleine et d'en étudier le comportement vocal (Lagerquist *et al.*, 2000; Goldbogen *et al.*, 2006; Oleson *et al.*, 2007; Calambokidis, 2002, 2003). Elle impose une forte proximité de la baleine et de l'embarcation de recherche, est difficile et demande souvent plusieurs essais. Ainsi les approches rapides et erratiques du bateau de recherche peuvent engendrer un stress important et provoquer un changement de comportement sur la population étudiée (Edds et MacFarlane, 1987). Cette méthode est donc problématique pour l'étude d'espèces menacées ou en voie de disparition. Il est à noter que le suivi ne peut être effectué que sur quelques individus d'une population et pendant une courte durée (à cause du détachement de la balise ou de l'autonomie limitée des batteries). Ce procédé reste intrusif et les changements sur le comportement de l'animal ne

sont pas connus.

Le milieu sous-marin est propice à l'utilisation des sons. La propagation des ondes sonores y est rapide (env. 1500 m/s) et l'atténuation beaucoup plus faible que dans l'air (Urick, 1983). La présence d'un couloir de son (« *SOFAR Channel* ») permet aux ondes sonores de basse fréquence de parcourir de très grandes distances dans l'océan profond (jusqu'a 10000 km, *e.g.* Bannister *et al.*, 1993). Les mammifères marins, qui émettent des sons dans l'eau, utilisent ces propriétés afin de communiquer entre eux et/ou s'orienter dans leur environnement (Richardson *et al.*, 1995). Grace à l'acoustique passive sous-marine, il est possible de détecter, identifier et localiser les animaux qui émettent régulièrement des sons spécifiques afin d'évaluer leur distribution saisonnière, leur abondance relative, leur comportement et leur habitat (Costa, 1996). L'opération consiste à enregistrer l'environnement sonore sous-marin en installant des mouillages équipés d'enregistreurs autonomes (*e.g.* Wiggins, 2003), ou d'autres systèmes d'acquisition connectés à des hydrophones, puis de les analyser. Contrairement aux observations visuelles faites en surface, l'acoustique passive présente l'avantage de pouvoir observer l'animal sans être dépendant des conditions météorologiques, des périodes d'apparition des baleines en surface et de la luminosité (Costa, 1996). Cette approche permet également de détecter les baleines sur plusieurs dizaines de kilomètres par opposition aux 1 ou 2 kilomètres par observation visuelle (Clark *et al.*, 1996; Cummings et Holliday, 1985; Clark et Fristrup, 1997; Ko *et al.*, 1986). L'acoustique passive permet également de suivre de façon non intrusive et à coût raisonnable un nombre élevé d'individus sur de longues périodes (*e.g.* Watkins *et al.*, 2004) et dans des environnements inhospitaliers où l'homme n'a pas accès facilement (les régions polaires par exemple).

L'analyse des enregistrements acoustiques sous-marins peut se faire « manuellement » avec un casque d'écoute (pour les vocalises audibles) et la représentation temps-fréquence des enregistrements. Un opérateur expérimenté analyse alors visuellement un spectrogramme (*cf.* section 2.3.1) puis identifie et localise en temps les vocalises visibles et/ou audibles (*e.g.* figure 1.1). Un tel processus est très long et la fatigue ou le changement

d'opérateur engendre un biais dans l'analyse qui est variable, donc difficile à estimer. Grâce aux méthodes de traitement de signal il est possible d'analyser les enregistrements de façon automatique en utilisant des algorithmes de détection et d'identification. La performance des méthodes de traitement de signal utilisées est très dépendante de la complexité de la vocalise de l'animal à reconnaître et de l'environnement sonore (propagation dans le milieu, caractéristiques du bruit). Une méthode peut donc difficilement être universelle. La méthode du filtre adapté (*Matched filter*) a été l'une des premières méthodes à être utilisée pour la détection de signaux acoustiques d'animaux (Mellinger et Clark, 1996). Cette méthode consiste à corréler un signal modèle avec le signal inconnu. Les *pics* de corrélation déterminent alors l'emplacement temporel des vocalises détectées. Cette méthode s'est avérée adaptée pour des signaux très stéréotypés et pour des conditions de bruits de type gaussien peu intenses (Mellinger et Clark, 2000). Bien qu'elle ait été utilisée dans plusieurs études (*e.g.* Stafford *et al.*, 1998, pour le rorqual bleu pacifique), cette méthode n'est pas efficace en environnements bruités. La méthode de corrélation des spectrogrammes suit le même principe que la méthode précédente excepté que l'étape de corrélation est effectuée sur les images temps-fréquence (spectrogramme) du modèle et de l'enregistrement inconnu. Cette méthode est très répandue et a mené à des travaux sur plusieurs vocalises de baleines (Mellinger, 2004 et Munger *et al.*, 2005 pour la baleine franche, Wiggins *et al.*, 2005 et Mellinger et Clark, 2000 pour le rorqual bleu). Cette méthode, comme également la méthode du filtre adapté, est bien adaptée pour des vocalises stéréotypées sans trop de variations temporelles ni fréquentielles. D'autres techniques sont issues du domaine de la reconnaissance automatique de la parole. On retrouve notamment les méthodes de reconnaissance telles que les réseaux neuronaux et les modèles de Markov cachés. Ces deux méthodes sont plus complexes mais ont montré de bons résultats pour l'identification des vocalises d'éléphants (Clemins *et al.*, 2005), d'oiseaux (Kogan et Margoliash, 1998) et de baleines (Mellinger, 2004; Potter *et al.*, 1994; Mellinger et Clark, 2000). Pour ces deux méthodes, l'extraction des descripteurs des vocalises est une étape très importante. Plusieurs ont été testés : les coefficients de Fourier (Mellinger, 2004), les coefficients d'ondelettes (Potter *et al.*, 1994; Selin *et al.*, 2007), les coefficients cepstraux à échelle de Mel, *MFCC*

(Clemins *et al.*, 2005) et les coefficients issus de la prédiction linéaire (Clemins et Johnson, 2006). Le choix de ces paramètres est très dépendant des caractéristiques de la vocalise à reconnaître (la bande fréquentielle par exemple). Les réseaux neuronaux et les modèles de Markov cachées permettent une bonne adaptivité aux variations des patrons des vocalises mais demandent un grand nombre de vocalises modèles pour la phase d'entraînement. Une autre approche utilisée pour la détection/identification des vocalises consiste à extraire le patron temps-fréquence des vocalises à partir du spectrogramme, de les représenter par une série de paramètres caractéristiques puis de les identifier en utilisant des algorithmes de classification. Les patrons de vocalises peuvent être extraits par des algorithmes de détection de bord (Gillespie, 2004) ou grâce à des algorithmes de suivi de trajet (« *tracking* » de la fréquence fondamentale) (Halkias et Ellis, 2006; Brown *et al.*, 2006). L'extraction automatique des contours de vocalises par algorithmes de *tracking* peut être très coûteuse en calcul (Brown et Zhang, 1991). Plusieurs algorithmes de classification ont été testés comme par exemple l'analyse discriminante (Gillespie, 2004) et la déformation temporelle dynamique, *DTW* (Buck et Tyack, 1993; Brown *et al.*, 2006).

La découverte des vocalises de rorqual bleu et de rorqual commun est assez récente. Tout d'abord ce sont de longs sons forts et de très basse fréquence (en dessous de 30 Hz) qui ont été reportés dans différents endroits du monde par les chercheurs et attribués à une source biologique (Walker, 1963; Weston et Black, 1965; Kibblewhite *et al.*, 1967). Ce n'est que plus tard, à partir d'observations visuelles et d'enregistrements, que les rorquals bleus et communs ont été identifiés comme étant les sources de certains de ces sons (Cummings et Thompson, 1971; Thompson *et al.*, 1979; Schevill *et al.*, 1964). Par la suite, d'autres études ont révélé qu'il existait différents répertoires de vocalises selon les populations de rorquals bleu (Thompson et Friedl, 1982; McDonald *et al.*, 2006a).

Le rorqual bleu produit des sons en basses fréquences (< à 200 Hz) de forte puissance ($\backsim$150-190 dB re 1 μPa @ 1 m, Thode *et al.*, 2000; Berchok *et al.*, 2006). Le répertoire du rorqual bleu dans le Saint-Laurent est le même que celui du rorqual bleu Nord-Atlantique. Il est composé de trois types de vocalises, les vocalises de type A, B et D. (Edds, 1982;

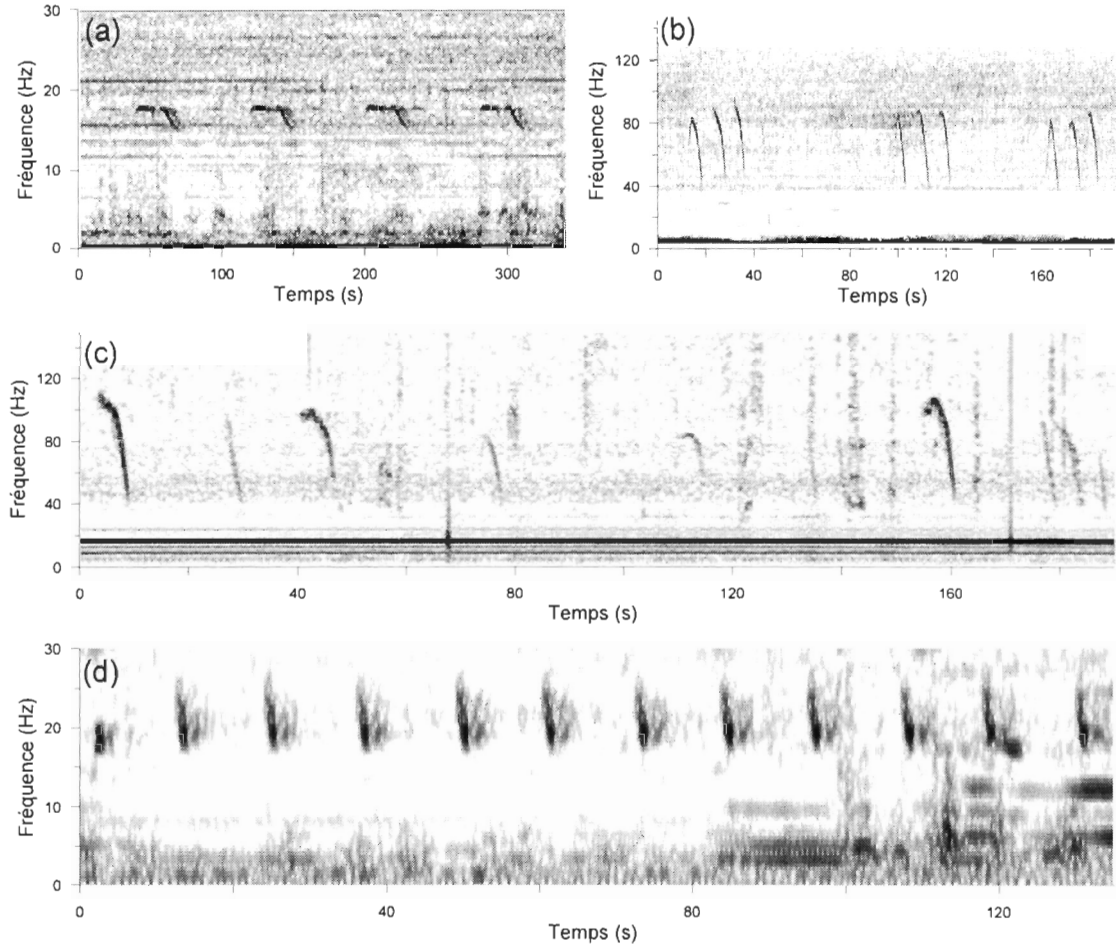

Figure 1.1 – Vocalises du rorqual bleu et du rorqual commun. Spectrogrammes représentant (a) 4 phrases AB, (b) 9 vocalises D groupées en trio répétés et (c) 11 vocalises D très variables, du rorqual bleu. (d) Spectrogramme représentant 12 vocalises de 20 Hz du rorqual commun avec présence d'échos. Les paramètres des spectrogrammes sont :(a) taille de FFT 4,1 s, taille de fenêtre 4,1 s, fenêtre de Hamming ; (b) et (c) taille de FFT 1 s, taille de fenêtre 0,5 s, fenêtre de Hamming ; (d) taille de FFT 4,1 s, taille de fenêtre 2 s, fenêtre de Hamming. La fréquence d'échantillonnage est de 2000 Hz.

Mellinger et Clark, 2003; Berchok *et al.*, 2006). La vocalise A est un son tonal de 18 Hz qui dure environ 8 secondes (figure 1.1a). La vocalise B qui est un son modulé en fréquence allant de 18 à 15 Hz en approximativement 11 secondes (figure 1.1a). Ces deux vocalises se succèdent souvent pour former des *phrases* répétées environ toutes les 74 secondes (Berchok *et al.*, 2006; Mellinger et Clark, 2003). Il est à noter qu'il peut ne pas y avoir de silence entre les deux vocalises ; Berchok *et al.* (2006) nomment ces vocalises les *vocalises hybrides* (figure 1.2a). Il est possible également de rencontrer des séries de vocalises A seulement ainsi que des séries de B seulement (Berchok *et al.*, 2006). La vocalise D (ou *Arch sound*) est moins stéréotypée [1]. Elle est caractérisée par une descente en fréquence dans la bande de fréquence 120-40 Hz (figure 1.1b,c). La vocalise peut débuter par une montée en fréquence plus ou moins prononcée formant ainsi une arche (*e.g.* 1^{ere} et 7^{eme} vocalises de la figure 1.1b). Les durées et les fréquences de début et de fin sont variables (figure 1.1c). Les vocalises D ne sont pas répétées en séquences régulières comme les vocalises A et B, cependant il est possible de voir sur les enregistrements quelques successions de vocalises régulièrement espacées en temps (figure 1.1b). Berchok *et al.* (2006) recensent trois types de *vocalises audibles* : les *Downsweeps*, les *Blurp* et les *Grunts*. Les vocalises D auxquelles nous nous intéressons dans cette étude correspondent aux *Downsweeps*. Mellinger et Clark (2003) reportent un autre type de vocalise qui consiste en un son de 9 Hz. Dans le Saint-Laurent, le bruit étant très intense aux alentours de 9 Hz, aucune de ces vocalises n'a pu être reportée.

L'utilisation des vocalises chez le rorqual bleu n'est pas bien comprise. Oleson *et al.* (2007) observent que les séries de *phrases* AB sont reliées principalement à une activité de déplacement d'un seul individu alors que les successions de vocalises A seulement ou B seulement sont plutôt utilisées lors de l'alimentation ou le déplacement de plusieurs individus ensembles. Ces vocalises semblent produites par les mâles uniquement et auraient une fonction pour la reproduction (Oleson *et al.*, 2007; McDonald *et al.*, 2001). Les vocalises D se trouvent être produites par des baleines bleues de plusieurs régions du monde (Rankin *et al.*, 2005; McDonald *et al.*, 2001; Mellinger et Clark, 2003). Oleson *et al.*

[1] *i.e.* ne suit pas rigoureusement un patron temps-fréquence

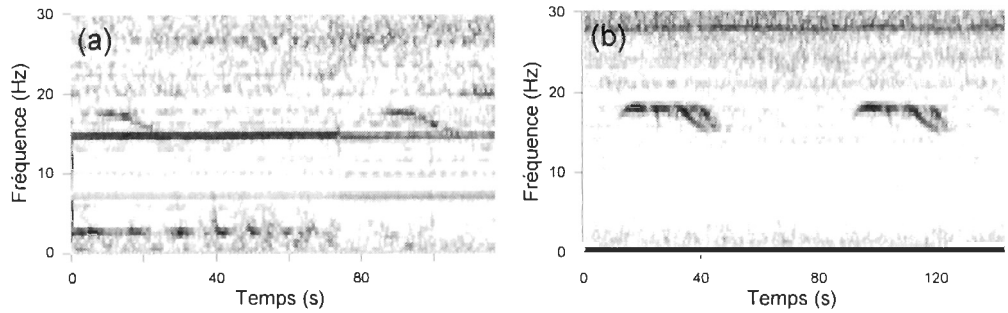

Figure 1.2 – Vocalises A et B du rorqual bleu. (a) Spectrogramme de 2 vocalises *hybrides*, (b) 2 phrases AB avec des forts échos. Les paramètres des spectrogrammes sont : taille de FFT 4,1 s, taille de fenêtre 4,1 s, fenêtre de Hamming. La fréquence d'échantillonnage est 2000 Hz.

(2007) observent que ces vocalises sont produites par les mâles et les femelles et émettent l'hypothèse qu'elles seraient reliées à un comportement social possiblement associé à la nourriture. Pour plus d'information sur le comportement vocal des balaenoptéridés se reporter aux travaux de Richardson *et al.* (1995), Payne et Webb (1971), Clark (1998) et Clark et Ellison (2003).

Le rorqual commun émet également des sons en basses fréquences à forte intensité (159-184 dB re 1 μPa @ 1 m, Charif *et al.*, 2002). Les vocalises du rorqual commun les plus présentes sur les enregistrements du Saint-Laurent sont les mêmes que celles reportées dans les autres régions du monde (Edds, 1988; Schevill *et al.*, 1964; Thompson et Friedl, 1982; Thompson *et al.*, 1992; Samaran, 2004). La plus fréquente est la vocalise de 20 Hz (ou impulsion de 20 Hz) qui est un son modulé en fréquence passant d'une fréquence de 23 à 18 Hz en environ 1 seconde (figure 1.1d). La fréquence haute peut être parfois supérieure. Ces impulsions peuvent être répétées sous forme de longues séquences régulières ou sous forme de séries de courtes durées avec un intervalle irrégulier entre chaque impulsion (Watkins *et al.*, 1987; Thompson et Friedl, 1982; McDonald *et al.*, 1995; Samaran, 2004). Edds (1988) et Samaran (2004) reportent également des vocalises tonales de 140 Hz. Un type de vocalise du rorqual commun à été reporté uniquement dans le Saint-Laurent. Il correspond à des impulsions ressemblant aux impulsions de 20 Hz typiques mais dont la fréquence est plus faible (nommées « *backbeat* », *e.g.* première vocalise de la figure 1.1d).

Ces deux dernières vocalises sont beaucoup moins fréquentes que les impulsions typiques de 20 Hz. Au cours de cette étude, elles ne sont pas considérées.
Les vocalises du rorqual commun sont produites uniquement par les mâles ce qui encourage à penser qu'elles jouent un rôle dans la reproduction (Croll *et al.*, 2002). Les longues périodes de silence observées entre les séquences de vocalises seraient corrélées avec l'activité respiratoire de l'animal (remontée à la surface) (Watkins *et al.*, 1987; Cummings *et al.*, 1986). Samaran (2004) pose l'hypothèse que le changement de rythme et de forme (*e.g. backbeats*) des vocalisations pourrait être attribuable à un changement dans le mécanisme de production de son causé par exemple par un comportement alimentaire comme la déglutition.

Bien que plusieurs méthodes de détection et de reconnaissance aient déjà été mises en place pour ces vocalises, aucune n'a été testée pour les conditions acoustiques et les vocalises du Saint-Laurent. A cause de l'importance du trafic, le bruit dans la Voie maritime du Saint-Laurent est très intense dans les basses fréquences (Simard *et al.*, 2006b). Cette plage de fréquence du bruit correspond exactement à celle utilisée par les rorquals bleus et les rorquals communs. La majorité des vocalises enregistrées se trouvent donc avec un rapport signal sur bruit très faible (figure 1.3). La présence d'une couche intermédiaire froide dans la moitié supérieure de la colonne d'eau et les murs abrupts à la tête du chenal Laurentien favorisent les échos et la propagation sonore par trajets multiples (figures 1.1d et 1.2b). Les vocalises captées peuvent donc se retrouver étirées en temps ou déformées. Une autre difficulté peut résider dans le fait que les vocalises de rorqual bleu et de rorqual commun se trouvent dans la même bande étroite de fréquence et peuvent être présentes simultanément (figure 1.3).
L'objectif de ce travail est donc de développer des méthodes de traitement de signal pour la détection et la reconnaissance automatique des vocalises du rorqual bleu et du rorqual commun adaptées au contexte acoustique spécifique du Saint-Laurent dans le but d'une application de monitorage en temps réel. Les algorithmes développés doivent donc avoir la capacité de détecter les signaux acoustiques de baleines dans un bruit intense, d'être adaptés à la variabilité en temps et en fréquence des vocalises et de pouvoir être implantés

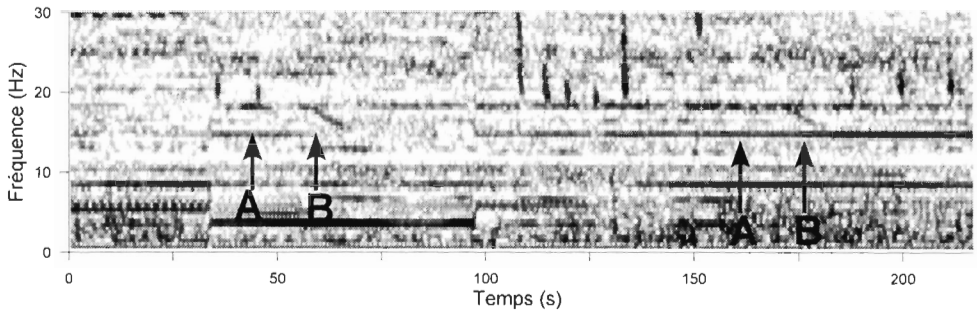

Figure 1.3 – Exemple d'enregistrement contenant des vocalises de rorquals bleus et communs avec du bruit intense. Les paramètres du spectrogramme sont : taille de FFT 4,1 s, taille de fenêtre 4,1 s, fenêtre de Hamming. La fréquence d'échantillonnage est 2000 Hz.

en temps réel. Dans un premier temps, une approche classique déjà utilisée dans plusieurs études pour les vocalises du rorqual bleu et du rorqual commun est adaptée aux conditions acoustiques du Saint-Laurent. Ensuite, une approche plus nouvelle est élaborée afin d'essayer de palier aux faiblesses de l'approche classique.

Chapitre 2

Matériel et méthodes

2.1 Collecte des données

Les données ont été récoltées grâce à un réseau côtier d'hydrophones posés sur le fond, situé à Cap-de-Bon-Désir en 2003 (sites A et B de la figure 2.1) et, à des enregistreurs acoustiques autonomes de type AURAL M-1 (Multi-Electronique Inc., Rimouski, QC, Canada) situés dans la colonne d'eau, déployés en 2003 et 2004 (sites C, D, E, F et G sur la figure 2.1) pendant la période estivale dans l'estuaire maritime du Saint-Laurent (tableau 2.1). Les enregistrements ont été numérisés sur 16 bits à une fréquence d'échantillonnage de 10 ou 20 kHz aux sites A et B, et 2 kHz pour les autres sites, puis décimés pour obtenir une fréquence F_s voulue (*cf.* tableaux A.1 et A.2 en annexe).

2.2 Bases de données

Afin de tester la performance des méthodes de reconnaissance utilisées, une base de données de test a été crée. La première partie est composée d'approximativement 500 vocalises de chaque type (tableau 2.2), identifiées puis localisées en temps (début et fin de la vocalise) manuellement, grâce au logiciel *Cool Edit Pro* (Syntrillium Software Corpora-

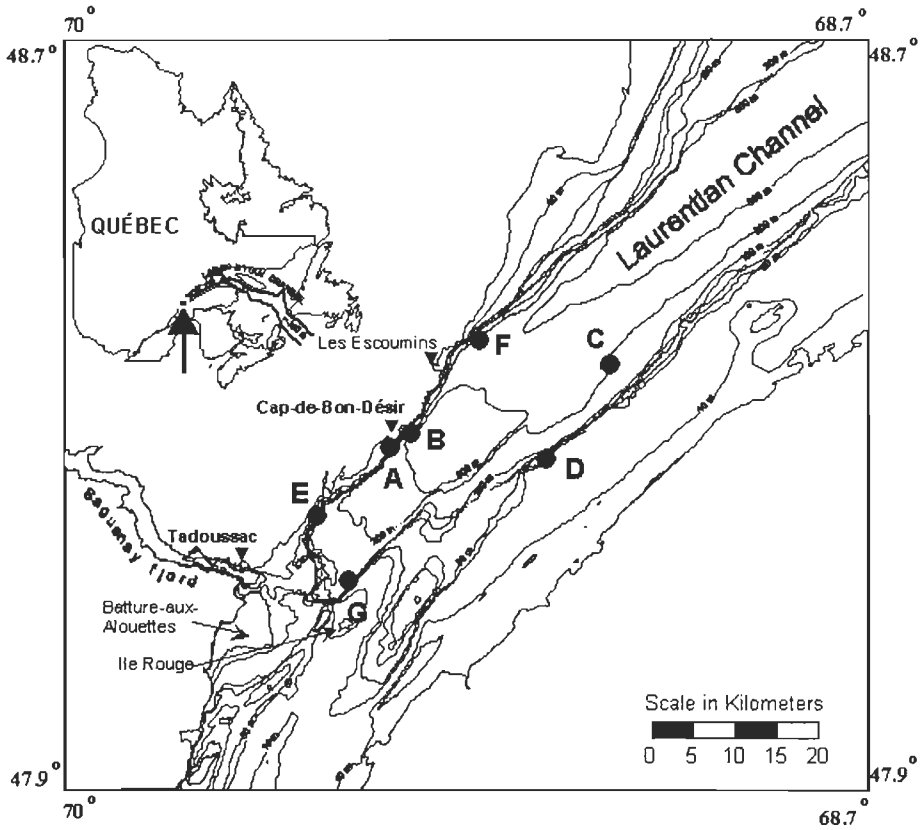

Figure 2.1 – Emplacements des enregistreurs acoustiques.

Tableau 2.1 – Années de collecte des données, coordonnées et profondeurs des enregistreurs acoustiques.

Site	Année	Période d'échan- -tillonnage	Position		Profondeur de l'hydrophone	Profondeur du fond
			Lat.	Long.		
A	2003	Août - Sept.	48,2683° N	69,4663° W	125 m	130 m
B	2003	Août - Sept.	48,2687° N	69,4636° W	188 m	193 m
C	2003	Sept. - Oct.	48,3544° N	69,1167° W	54 m	88 m
D	2003	Sept. - Oct.	48,2504° N	69,2225° W	53 m	55 m
E	2003	Sept. - Oct.	48,1900° N	69,5925° W	43 m	60 m
F	2004	Août - Oct.	48,3783° N	69,3300° W	155 m	239 m
G	2004	Août - Oct.	48,1233° N	69,5433° W	126 m	135 m

Tableau 2.2 – Nombre de vocalises par catégorie dans les bases de données

Base de données	# Voc. A	# Voc. B	# Voc. D	# Voc. 20Hz
Test	568	490	510	495
Apprentissage	117	91	97	104

tion, AZ, U.S.A.), par un opérateur entraîné (l'auteur) en utilisant le spectrogramme des enregistrements et un casque d'écoute si nécessaire (valable uniquement pour la vocalise de type D du rorqual bleu). Les vocalises de la base de données ont été choisies de façon à représenter au mieux les conditions acoustiques du Saint-Laurent (*i.e.* bruit, échos, etc.). La seconde partie de la base de données de test est composée de 14 heures d'enregistrements ne contenant aucune vocalise (échantillonnage de bruits sous-marins divers). Les périodes sélectionnées sont celles qui ont été jugées les plus problématiques pour les méthodes de détection (*e.g.* bruit de navigation intense dans la bande de fréquence d'intérêt). Plusieurs algorithmes utilisés dans cette étude nécessitent une phase d'apprentissage. Une base de données d'entraînement a ainsi été créée. Celle-ci est composée d'une centaine de vocalises de chaque type, sélectionnées parmi l'ensemble des enregistrements qui ne sont pas utilisés dans la base de données de test (tableau 2.2).

2.3 Détection et identification des vocalises

Deux approches de détection/identification des vocalises basées sur une représentation temps-fréquence du signal sont expérimentées. Les deux premières étapes communes aux deux approches sont le calcul du spectrogramme puis la réduction du bruit. Ensuite, la première approche (figure 2.2a), consiste à détecter les vocalises en calculant la coïncidence du spectrogramme avec un modèle temps-fréquence de vocalise. Pour la seconde approche (figure 2.2b), les vocalises sont détectées en extrayant leur contour temps-fréquence sur le spectrogramme. Les contours extraits sont caractérisés par des paramètres puis identifiés par des algorithmes de classification. Deux type d'algorithmes sont testés (DTW et VQ).

Ces méthodes ont été programmées avec le logiciel *Matlab* ® (The Mathworks Inc., MA, U.S.A.).

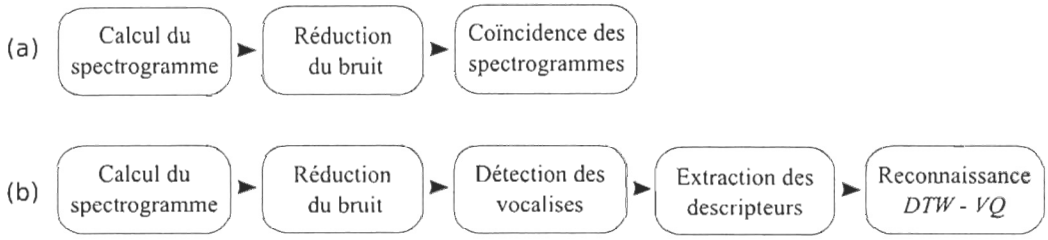

Figure 2.2 – Description des deux approches utilisées pour la détection/identification des vocalises. (a) Étapes pour l'approche par coïncidence des spectrogrammes (*cf.* section 2.3.3),(b) étapes pour l'approche par reconnaissance de contours (*cf.* section 2.3.4).

2.3.1 Calcul du spectrogramme

L'ensemble des méthodes utilisées dans cette étude se basent sur la représentation temps-fréquence du signal (spectrogramme). La technique choisie ici est la transformée de Fourier à fenêtre glissante. Le signal acoustique échantillonné, $s[m]$, est multiplié par une fenêtre glissante $w[m]$ de K échantillons. La transformée de Fourier de chaque segment, n, obtenu est calculée. La fenêtre glissante se déplace avec un pas de L échantillons (figure 2.3). On peut définir le spectrogramme, $S[n,k]$, par le module des transformées de Fourier rapide (FFT) des différentes fenêtres pondérées du signal. On a alors

$$S[n,k] = \left| \sum_{m=0}^{K-1} s[nL+m]w[m]e^{\frac{-j2\pi mk}{K}} \right|, \tag{2.1}$$

où $n = 0, 1, \cdots, N-1$ et $k = 0, 1, \cdots, M-1$, représentent respectivement le temps et la fréquence. Le nombre de fréquences discrètes M (taille de la FFT) doit être supérieur ou égal au nombre d'échantillons K de la fenêtre w.

Plusieurs fonctions de pondération $w[n]$ sont proposées dans la littérature ; celles utilisées dans ce travail sont reportées en annexe A. Selon la nature du signal à analyser, il est possible de modifier la taille, K, et le pas d'avancement, L, de la fenêtre glissante.

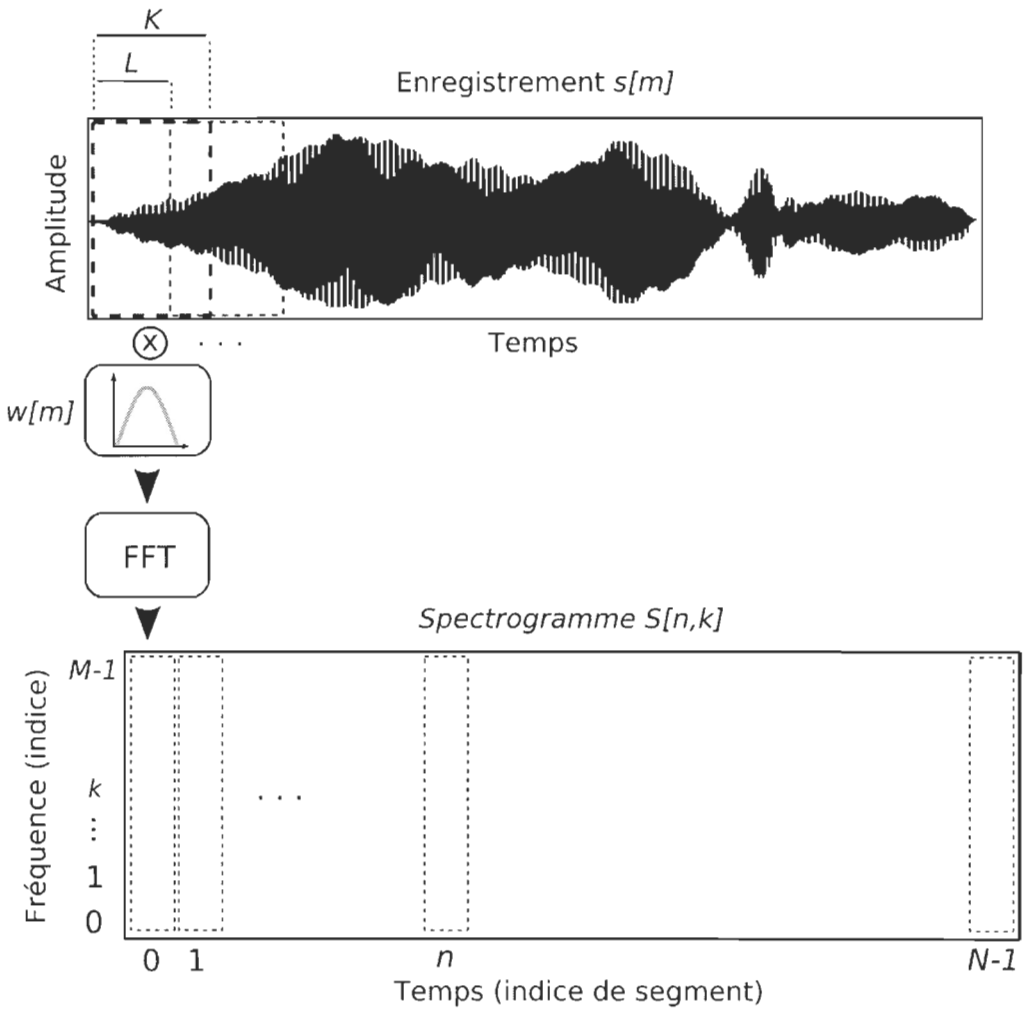

Figure 2.3 – Schéma du principe du calcul d'un spectrogramme

Plus K est petit, plus la précision temporelle du spectrogramme est accrue et la précision fréquentielle basse. Au contraire, plus K est grand, plus la précision temporelle du spectrogramme est basse et la précision fréquentielle accrue. L'amélioration de la résolution temporelle se fait donc au détriment de la résolution fréquentielle et vice versa. Ce compromis de précision en temps et en fréquence est appelé *dualité d'Heisenberg*. Deux techniques permettent cependant de *contourner* cette dualité afin d'obtenir un spectrogramme avec à la fois une bonne précision en temps et en fréquence. Tout d'abord, le choix d'un pas d'incrémentation L inférieur à la taille de la fenêtre K (chevauchement des segments d'analyse) permet d'améliorer la précision temporelle du spectrogramme sans affecter la précision fréquentielle. Plus le pas sera petit, meilleure sera la précision temporelle. Ensuite, calculer la FFT sur une durée supérieure à celle des segments ($M > K$) en ajoutant des zéros à la fin du signal fenêtré (*zero padding*), permet d'obtenir une résolution fréquentielle plus grande sans pour autant dégrader la précision temporelle. Il est à noter que cette opération n'ajoute aucune information au signal. Cependant, elle agit comme une interpolation du spectre et permet d'obtenir une image temps-fréquence plus précise. Il est à noter que ces deux techniques augmentent de façon importante la quantité de calcul.

Les vocalises auxquelles on s'intéresse ici ayant des caractéristiques de durée et de fréquence différentes, il n'est pas possible de les représenter toutes avec précision sur un même spectrogramme (dualité d'Heisenberg). Plusieurs jeux de paramètres sont alors utilisés pour le calcul des spectrogrammes, un premier pour les vocalises A et B, un second pour les vocalises D et un troisième pour les impulsions de 20 Hz (*cf.* tableaux A.1 et A.2 en annexe).

2.3.2 Réduction du bruit

Les enregistrements sont très pollués par le bruit intense dû majoritairement à la navigation. La performance des méthodes de détection automatique est directement liée aux conditions de bruit. Il est alors important d'effectuer un pré-traitement de l'information

utilisée pour la détection afin de rehausser les composantes du signal utile et d'éliminer le maximum de bruit parasite. Étant donné que la représentation temps-fréquence d'un signal peut être considérée comme une image, il est possible de travailler directement sur cette représentation en utilisant des méthodes de traitement d'image. Trois méthodes sont appliquées successivement, l'égalisation, le lissage et le seuillage.

2.3.2.1 Égalisation

La première étape du processus de réduction du bruit est l'égalisation du spectrogramme (Van-Trees, 1968; Mellinger, 2004). Cette technique consiste, dans un premier temps, à calculer un spectrogramme moyenné en temps, $\overline{S}[n,k]$, en appliquant une moyenne mobile à chaque ligne de la matrice du spectrogramme (*i.e.* pour chaque bande de fréquence du spectrogramme, *cf.* section 2.3.1). On a alors

$$\overline{S}[n,k] = \frac{1}{\Delta t} \sum_{i=0}^{\Delta t-1} S[n+i,k], \tag{2.2}$$

avec $n = 0, 1, \cdots, N-1,\ k = 0, 1, \cdots, M-1$ et où Δt est la longueur de la fenêtre mobile. Dans un second temps, le spectrogramme moyenné obtenu est soustrait au spectrogramme original. Le spectrogramme égalisé, $S_{eq}[n,k]$, est alors défini par

$$S_{eq}[n,k] = S[n,k] - \overline{S}[n,k]. \tag{2.3}$$

Ce procédé a pour effet de supprimer du spectrogramme les raies spectrales produites par les navires (figure 2.4b). La durée correspondant à la fenêtre, Δt, utilisée pour la moyenne mobile est un paramètre important qui définit la durée minimale des bandes spectrales à supprimer. Des valeurs différentes de Δt ont été choisies selon le type de vocalises visé (*cf.* annexe A).

2.3.2.2 Lissage du spectrogramme

L'étape suivante consiste à lisser le spectrogramme en utilisant un noyau gaussien (Gillespie, 2004). Ceci revient à convoluer l'image de l'enregistrement (*i.e.* matrice issue du spectrogramme) avec un masque bidimensionnel, G (l lignes par c colonnes), représentant une gaussienne, ce qui revient à écrire

$$S_{lis} = S_{eq} * G, \tag{2.4}$$

où S_{lis} est le spectrogramme lissé. Les vocalises se trouvent alors rehaussées et les bruits ponctuels abaissés (*i.e.* augmentation du contraste, figure 2.4c). Différentes dimensions de masque ont été utilisées selon les vocalises à détecter (*cf.* tableaux A.1 et A.2 en annexe). L'utilisation de masques non-carrés permet de lisser le spectrogramme selon une direction privilégiée, permettant ainsi d'améliorer la continuité de certaines vocalises. Cependant, il est à noter qu'une dimension de masque favorisant de façon trop importante une direction peut conduire à rehausser des parties de l'image ne correspondant pas à des vocalises.

2.3.2.3 Seuillage

La dernière étape a pour but de retenir uniquement les composantes du spectrogramme correspondant aux vocalises en supprimant celles de plus faible énergie. Pour ce faire, une série de seuillages est mise en place.

Tout d'abord, afin de détecter les périodes qui contiennent potentiellement des vocalises, l'énergie moyenne résiduelle, $\overline{e}[n]$, pour chaque pas de temps, n, du spectrogramme est calculée. Elle est définie par

$$\overline{e}[n] = \frac{1}{M} \sum_{k=0}^{M-1} S_{lis}[n,k], \qquad n = 0, 1, \cdots, N-1 \tag{2.5}$$

où M est l'indice des fréquences le plus élevé du spectrogramme lissé S_{lis} (*cf.* section 2.3.2.2). Pour faire ressortir d'avantage ces périodes (*i.e.* pics plus démarqués), un lissage

par moyenne mobile de longueur de fenêtre d_1 (*cf.* tableaux A.1 et A.2 en annexe) est effectué sur la courbe $\overline{e}[n]$ (*cf.* courbe en bleu sur la figure 2.4d). La courbe $\overline{e}_{lis}[n]$ alors obtenue s'exprime par

$$\overline{e}_{lis}[n] = \frac{1}{d_1} \sum_{i=0}^{d_1-1} \overline{e}[n+i], \qquad n = 0, 1, \cdots, N-1. \tag{2.6}$$

Deux seuils adaptatifs, T_1 et T_2, sont ensuite calculés en utilisant une fonction intermédiaire, $E[n]$, définie par (Renevey et Drygajlo, 2001)

$$E[n] = \lambda E[n-1] + (1-\lambda)\overline{e}_{lis}[n], \qquad n = 0, 1, \cdots, N-1 \tag{2.7}$$

où λ est un facteur d'adaptivité (*cf.* tableaux A.1 et A.2 en annexe). Les seuils T_1 et T_2 sont alors définis par

$$T_1[n] = E[n] + \delta_1 E[n], \qquad n = 0, 1, \cdots, N-1 \tag{2.8}$$

et

$$T_2[n] = E[n] + \delta_2 E[n], \qquad n = 0, 1, \cdots, N-1 \tag{2.9}$$

où δ_1 et δ_2 sont des constantes comprises dans l'intervalle $[0, 1]$. Les parties de la courbe $\overline{e}_{lis}[n]$ qui excèdent T_2 (en noir sur la figure 2.4d) sont considérées correspondre à des périodes d'activité vocale. Le seuil T_2 a donc un rôle de détection. Le seuil T_1 (en rouge sur la figure 2.4d) permet de définir les temps de début et fin des occurrences détectées. Les tranches temporelles du spectrogramme qui sont détectées sont conservées, les autres sont ignorées.

Enfin, pour chaque trame du spectrogramme, un dernier seuil adaptatif, T_3 est défini par

$$T_3[n] = \overline{e}[n] + \delta_3 \sigma[n], \qquad n = 0, 1, \cdots, N-1 \tag{2.10}$$

où $\sigma[n]$ est l'écart-type des valeurs du spectrogramme lissé au temps n et où δ_3 est une constante comprise dans l'intervalle $[0, 1]$. Seules les valeurs supérieures à ce seuil sont

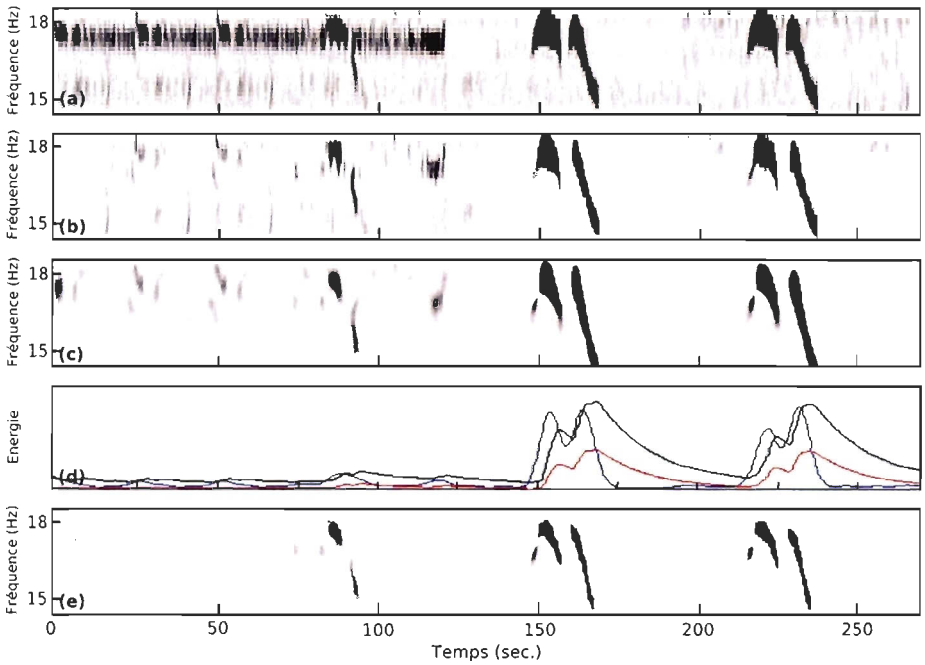

Figure 2.4 – Étapes d'atténuation du bruit. (a) Spectrogramme original, $S[n,k]$, contenant trois phrases AB de rorqual bleu, (b) spectrogramme, $S_{eq}[n,k]$, issu de l'étape d'égalisation, (c) spectrogramme, $S_{lis}[n,k]$, issu de l'étape de lissage gaussien, (d) courbe $\overline{e}_{lis}[n]$, en bleu, le seuil T_1, en rouge, le seuil T_2, en noir, (e) spectrogramme final, $\widehat{S}[n,k]$, après le seuillage. L'axe des temps en abscisse est obtenu par : $t = nL/Fs$.

conservées. Le spectrogramme final issu de l'étape de seuillage est noté $\widehat{S}[n,k]$ (figure 2.4e).

2.3.3 Détection par coïncidence des spectrogrammes

La coïncidence des spectrogrammes consiste à retrouver dans un spectrogramme inconnu une vocalise précise à partir d'un patron (image) de sa représentation temps-fréquence. Cette technique, couramment utilisée en traitement de l'image, fait partie des premières méthodes robustes appliquées pour la reconnaissance de vocalises d'animaux, plus spécialement de baleines (Mellinger et Clark, 1996, 2000). L'opération permettant l'association du patron avec le spectrogramme peut varier selon les études. Ici l'opération utilisée est l'opération logique *AND*.

Tout d'abord, le spectrogramme issu de l'étape d'atténuation du bruit (*cf.* section 2.3.2) est binarisé. Les valeurs du spectrogramme supérieures à zéro sont fixées à un, les autres sont fixées à zéro.

Ensuite, une image binaire du patron temps-fréquence de la vocalise à détecter est créée par l'expérimentateur. Elle est définie par un segment de droite (*i.e.* une image de *chirp*) caractérisé par les paramètres suivants : la fréquence de début, f_1, la fréquence de fin, f_2, la durée, D_{voc}, l'épaisseur en fréquence, Δf et la durée qui précède et succède l'image de la vocalise, D_{ini}. Ces paramètres sont illustrés sur la figure 2.5. Les valeurs de ces paramètres pour les différentes vocalises sont reportées dans le tableau A.2 en annexe. Enfin, en calculant le taux de superposition du modèle créé, à chaque pas de temps du spectrogramme inconnu grâce à l'opération *AND*, une fonction de détection est obtenue. Une valeur de 100% indique une correspondance parfaite des zeros et des uns. Un seuil T_{cs} est défini. Les pics de la fonction de détection qui excèdent ce seuil définissent les positions temporelles des vocalises détectées. La figure 2.6 illustre de façon schématique le processus de détection.

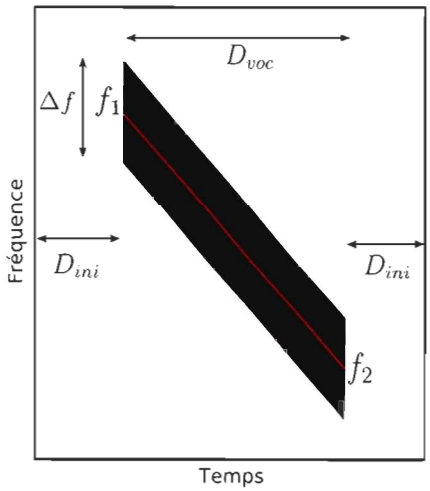

Figure 2.5 – Illustration des paramètres d'un patron modèle de vocalise dans le plan temps-fréquence.

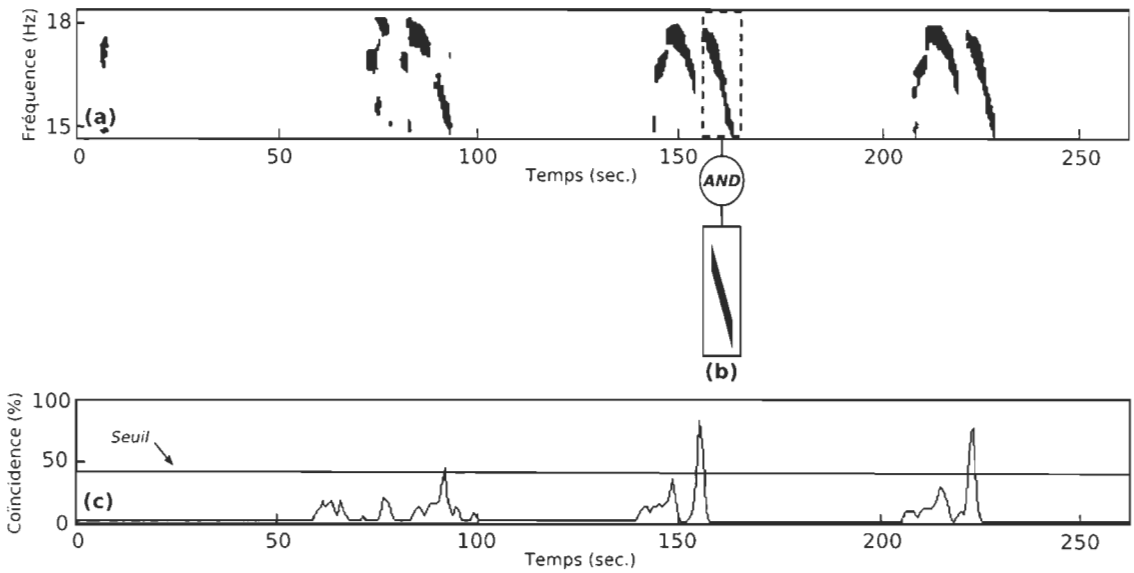

Figure 2.6 – Schéma du principe de la coïncidence des spectrogrammes. (a) Spectrogramme (figure 2.4e) binarisé, (b) patron modèle d'une vocalise B, (c) fonction de détection.

2.3.4 Reconnaissance des contours des vocalises

La première de cette approche consiste à extraire les contours des vocalises du spectrogramme filtré. Deux méthodes de classification sont ensuite explorées pour identifier chacun des contours, la déformation temporelle dynamique (« Dynamic Time Warping »), notée DTW et la quantification vectorielle (« Vector Quantization »), notée VQ.

2.3.4.1 Extraction des contours

L'extraction des contours temps-fréquence a fait l'objet de plusieurs études dans divers domaines. On notera Brown et Zhang (1991), Brown (1992), pour les signaux musicaux ; Rabiner (1977), Rabiner *et al.* (1976), pour les signaux de parole et Halkias et Ellis (2006), Datta et Sturtivant (2002), Sturtivant et Datta (1995a,b), Brown *et al.* (2006), pour les vocalises d'odontocètes. Dans notre cas, le choix d'un algorithme doit tenir compte à la fois de la qualité d'extraction mais aussi du temps de calcul, un des objectifs étant d'implanter l'algorithme dans un système temps-réel (*cf.* chapitre 1). La méthode utilisée ici est grandement inspirée des travaux de Halkias et Ellis (2006). L'extraction s'articule en trois étapes, l'identification des maxima locaux, la création de fragments et la connection des fragments. Dans un premier temps les maxima locaux du spectrogramme sont extraits. Ils sont définis comme une, ou un ensemble, de valeur(s), pour chaque colonne du spectrogramme, pour lesquelles il est impossible de trouver une valeur plus élevée sans passer par une valeur inférieure d'abord. En d'autres termes, les maxima locaux correspondent aux pics du spectrogramme (figure 2.7a).

L'objectif est ensuite de chercher la cohérence entre ces points en formant des structures organisées. Les points contigus en temps et en fréquence sont alors connectés pour former des fragments. Un certain écart est toléré pour la contiguïté en fréquence. Deux points peuvent alors se connecter seulement si ils sont adjacents en temps et séparés de moins de quatre points en fréquence. Pour chaque fragment, un lissage temporel par moyenne mobile de longueur de fenêtre d_2 (*cf.* tableaux A.1 et A.2 en annexe), est effectué. Un exemple de fragments est illustré à la figure 2.7b.

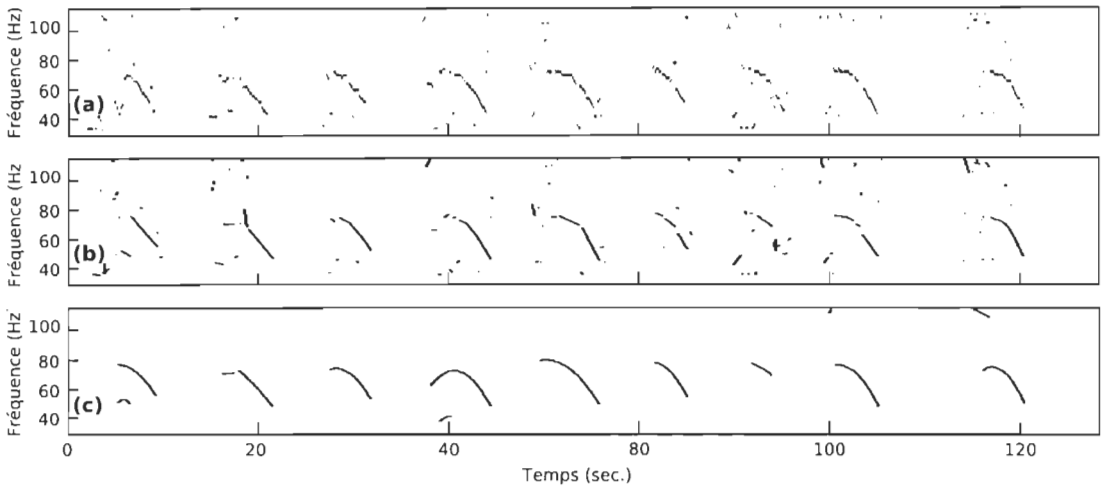

Figure 2.7 – Extraction des contours de vocalises (neuf vocalises D de rorqual bleu). (a) Maximums locaux extraits du spectrogramme débruité, (b) fragments extraits puis lissés, (c) contours formés par connexion des fragments puis sélectionnés par durée.

Certaines vocalises sont représentées par un seul fragment, cependant d'autres sont constituées de plusieurs fragments (contour fragmenté). Il s'avère donc nécessaire de connecter certains de ces fragments afin de reconstituer l'intégrité des contours.
Chaque paire, x_i, de fragments suffisamment proches en temps (inférieurs à T_{seg} secondes) est caractérisée par un vecteur, α_i, contenant les pentes, α_{i1} et α_{i2} (figure 2.8), des bouts [1] des fragments adjacents et par la distance fréquentielle minimale de cassure, β_i. Deux distances de cassure sont mesurées pour chaque connexion. L'une, β_{i2}, correspond à l'écart de fréquence mesuré lorsque la fin du premier fragment est prolongée linéairement jusqu'au début du deuxième fragment. L'autre, β_{i1}, correspond à l'écart de fréquence mesuré lorsque le début du deuxième fragment est prolongé linéairement jusqu'à la fin du premier (figure 2.8). La distance β_i correspond à la plus petite des deux.

La connexion des fragments est décidée en utilisant des modèles de probabilités (Halkias et Ellis, 2006). La probabilité de connexion a été modélisée en extrayant, de l'ensemble de la base de données d'apprentissage, deux séries d'observations, L_α et L_β, correspondant

[1] les bouts sont définis comme étant les points aux extrémités d'un fragment dont la pente calculée pour chaque point contigus est de même signe

respectivement aux paramètres α et β calculés pour N paires de fragments appartenant à des mêmes contours. Ces deux séries d'observations peuvent alors s'écrire

$$L_\alpha = \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \\ \alpha_{i1} & \alpha_{i2} \\ \vdots & \vdots \\ \alpha_{N1} & \alpha_{N2} \end{bmatrix} \quad et \quad L_\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_i \\ \vdots \\ \beta_N \end{bmatrix}.$$

On définit ainsi deux distributions normales $\Theta_\alpha(\Sigma_\alpha, \mu_\alpha)$ (figure 2.9a, c) et $\Theta_\beta(\sigma_\beta, \mu_\beta)$ (figure 2.9b) pour modéliser les séries d'observations L_α et L_β, avec Σ_α, μ_α, σ_β et μ_β, respectivement la matrice de covariance et le vecteur de moyennes de L_α, et, l'écart-type et la moyenne de L_β. Une connexion de paire de fragments inconnus, x, de paramètres α_x et β_x, peut ainsi être évaluée par sa vraisemblance, $P(x)$, exprimée par

$$P(x) = P(x|\Theta_\alpha)P(x|\Theta_\beta), \tag{2.11}$$

où $P(x|\Theta_\alpha)$ et $P(x|\Theta_\beta)$ sont respectivement les vraisemblances que la connexion x puisse être engendrée par les modèles Θ_α et Θ_β, et sont définies par

$$P(x|\Theta_\alpha) = \frac{1}{(2\pi)\left|\Sigma_\alpha\right|^{1/2}} e^{-\frac{1}{2}(\alpha_x - \mu_\alpha)^T \Sigma_\alpha^{-1}(\alpha_x - \mu_\alpha)}, \tag{2.12}$$

et

$$P(x|\Theta_\beta) = \frac{1}{\sqrt{2\pi}\sigma_\beta} e^{-\frac{1}{2}\left(\frac{\beta_x - \mu_\beta}{\sigma_\beta}\right)^2}. \tag{2.13}$$

Il est à noter que l'équation (2.11) n'est vérifiée que si L_α et L_β sont indépendantes.

Pour des raisons de commodité de calcul, les deux distributions, Θ_α et Θ_β, ont été normalisées à un (*i.e.* divisées par leurs maximums respectifs). Les fragments sont connectés seulement si $P(x)$ excède une valeur, T_c, fixée de façon empirique (*cf.* tableau A.1 en annexe). Pour chaque bout de fragment on peut avoir le choix entre plusieurs connexions ; la connexion retenue est celle dont la vraisemblance est maximale. Les connexions sont

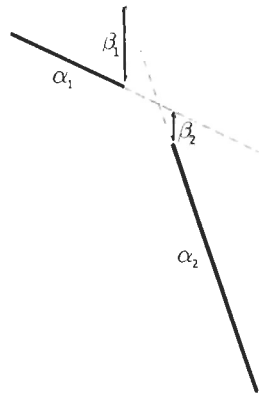

Figure 2.8 – Paramètres, α_1, α_2, β_1 et β_2, mesurés pour évaluer la connexion de deux fragments (exemple d'une vocalise D de rorqual bleu).

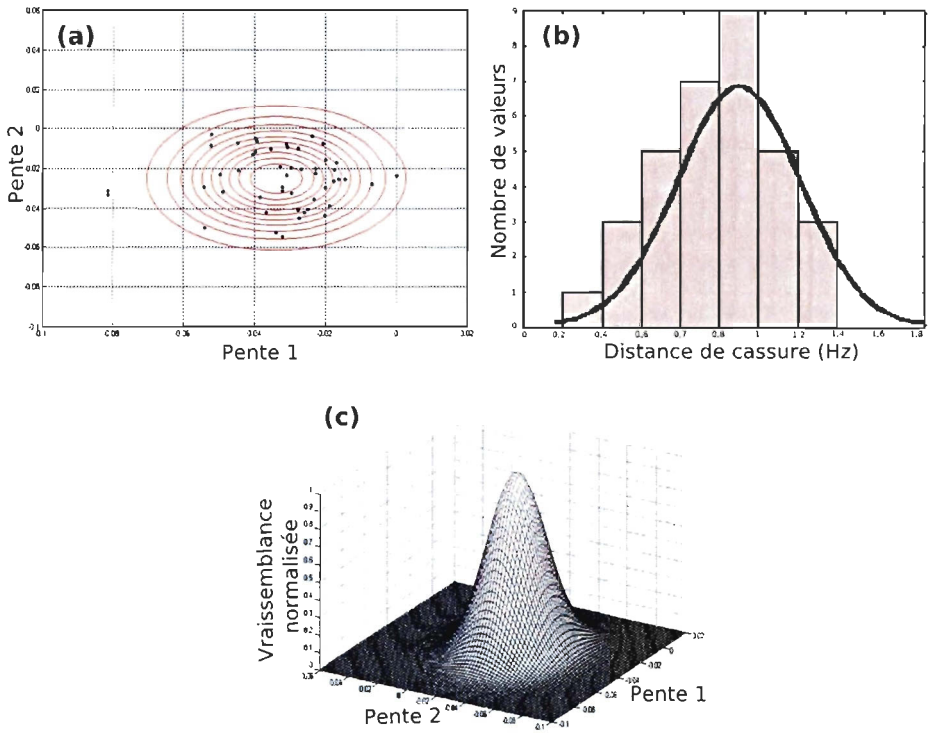

Figure 2.9 – Représentation graphique des modèles de connexion de fragments. (a) Représentation sur un plan des vecteurs de pentes de la série d'observation L_α (points noirs) et des lignes de niveau du modèle Θ_α qui lui est associée (lignes rouges). (b) distribution des distances de cassure de la série d'observation L_β (barres grisées) et représentation du modèle Θ_β qui lui est associé (courbe noire), (c) visualisation en trois dimensions du modèle de pente Θ_α.

réalisées par interpolation par un polynôme du second degré ajusté sur les fragments adjacents. Seuls les contours résultants dont la durée est comprise entre T_{min} et T_{max} sont conservés (figure 2.7c). Les valeurs des différents paramètres utilisés sont rapportées en annexe A.

2.3.4.2 Classification des contours par déformation temporelle dynamique

La déformation temporelle dynamique (DTW) est une méthode de classification initialement développée et utilisée dans le domaine de la parole pour la reconnaissance de mots isolés (Rabiner et Juang, 1993). Cette méthode fut ensuite introduite pour la classification des vocalises stéréotypées d'animaux, comme par exemple les vocalises de dauphins (Buck et Tyack, 1993), d'épaulards (Brown *et al.*, 2006) et d'oiseaux (Ito *et al.*, 1996; Anderson *et al.*, 1996).

L'algorithme consiste à reconnaître une vocalise inconnue, T, en la comparant à des modèles de vocalises connues d'un *dictionnaire*. Il permet, lors de cette comparaison, de tenir compte des compressions et des extensions temporelles des vocalises, engendrées soit lors de la production du son (modulation par la baleine elle même, *e.g.* vocalise D, Berchok *et al.*, 2006) soit lors de la propagation des ondes sonores dans le médium (*e.g.* échos, trajets multiples, *cf.* figure 1.2b). Le dictionnaire est constitué de k modèles de références, R_k, de vocalises connues sélectionnées dans la base de données d'apprentissage par l'expérimentateur. Le but est de définir une mesure de dissemblance (distance) entre la vocalise inconnue T et chaque modèle R_k.
Chaque vocalise est représentée à chaque pas de temps n, par trois paramètres : $f_{inst}[n]$, la fréquence instantanée et, $v[n]$ et $a[n]$ respectivement la vitesse et l'accélération de $f_{inst}[n]$ définis par

$$v[n] = f_{inst}[n] - f_{inst}[n-1], \tag{2.14}$$

et

$$a[n] = v[n] - v[n-1]. \tag{2.15}$$

Une vocalise T ou R_k est alors caractérisée par une suite de vecteurs à trois dimensions :

$$T : [T[i]; \qquad i = 1, 2, \cdots, I]$$
$$R_k : [R_k[j]; \qquad j = 1, 2, \cdots, J_k]$$

où I et J_k représentent respectivement le nombre de trames de temps de la vocalise T et de la vocalise de référence R_k. Une matrice X, de dimension I x J_k, représentant les distances euclidiennes, $d(T(i), R_k(j))$, séparant les vecteurs de T de ceux de R_k, est définie (figure 2.10). Calculer la dissimilarité entre T et R_k revient à définir le parcourt w, de longueur C, progressant de façon monotone du point $(1, 1)$ au point (I, J_k) de la matrice X. Il est défini par

$$w : [i(c), j(c)] \quad ; \quad c = 1, 2, \cdots, C \tag{2.16}$$

avec $i(1) = 1$, $j(1) = 1$, $i(C) = I$, $j(C) = J_k$, et tel que la distance qui lui est associée (Boite *et al.*, 2000),

$$D_w(T, R_k) = \frac{\sum_{c=1}^{C} d\{T[i(c)], R_k[j(c)]\}\ g(c)}{N(g)}, \tag{2.17}$$

soit minimale. Dans l'équation (2.17), $g(c)$ est une fonction de pondération définie par des contraintes locales (figure 2.11) et $N(g)$ un facteur de normalisation dépendant de la fonction g, défini par

$$N(g) = \sum_{c=1}^{C} g(c). \tag{2.18}$$

Le chemin w résultant est en fait une image de la distorsion temporelle qui existe entre les deux séquences comparées. Pour deux séquences parfaitement identiques, w serait confondu avec la diagonale de la matrice X.

Au lieu de chercher le chemin w de façon exhaustive (très coûteux en temps de calcul), l'algorithme DTW calcule de façon récursive la distance accumulée minimale pour chaque point en tenant compte de certaines contraintes locales (Vintsyuk, 1968; Ney, 1984). La contrainte locale utilisée ici est représentée à la figure 2.11. Il existe d'autres contraintes

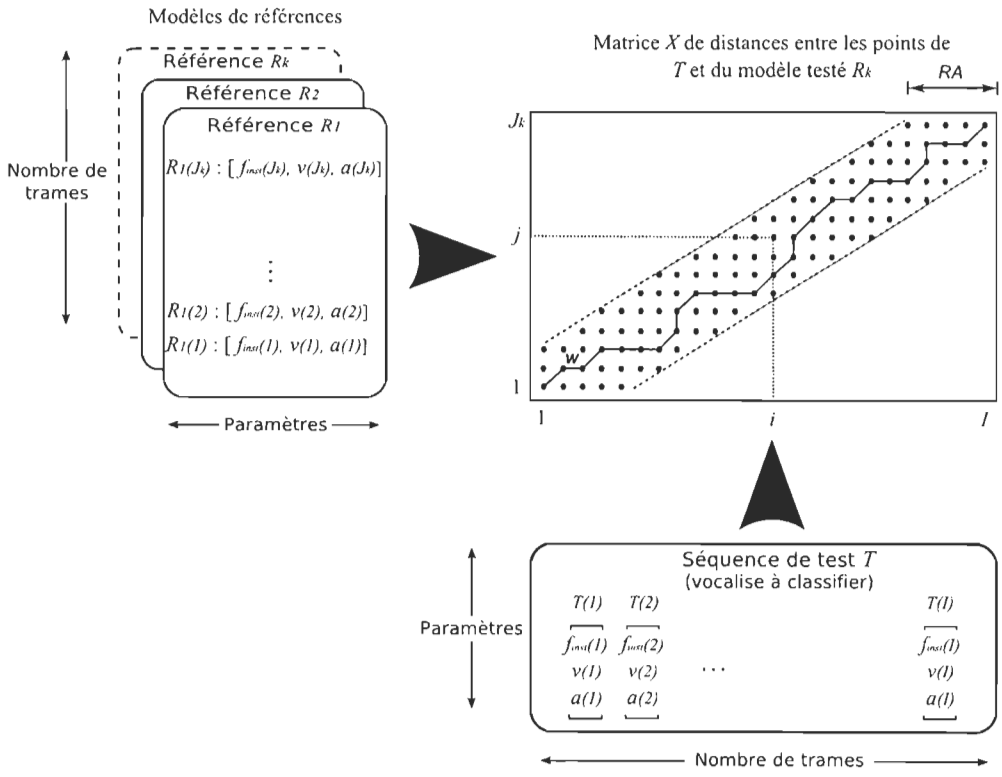

Figure 2.10 – Schéma du principe de l'algorithme DTW.

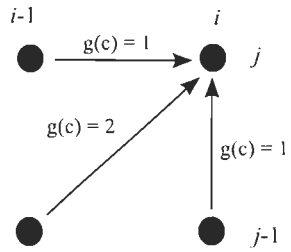

Figure 2.11 – Contraintes locales et pondérations

locales. Pour plus de précision se reporter au travaux de Myers *et al.* (1980) et Rabiner et Juang (1993). Pour chaque point (i, j) de X avec une distance locale $d(i, j)$, la distance accumulée, $\tilde{D}(i, j)$ est obtenue comme suit (Sakoe et Chiba, 1978) :

$$\tilde{D}(i,j) = d(i,j) + min \begin{cases} \tilde{D}(i-1,j) \\ \tilde{D}(i-1,j-1) + d(i,j) \\ \tilde{D}(i,j-1) \end{cases} . \tag{2.19}$$

La distance minimale cherchée entre les vocalises R_k et T s'écrit donc

$$D_w(R_k, T) = \frac{\tilde{D}(I, J_k)}{N(g)}. \tag{2.20}$$

Afin d'éviter les distorsions exagérées et de réduire le temps de calcul, la zone de recherche du chemin w est limitée par deux droites parallèles à la diagonale (figure 2.10) dont les positions sont fixées par un paramètre RA (dans notre étude : $RA = 5$) tel que (Sakoe et Chiba, 1978)

$$|i - j| \leq RA. \tag{2.21}$$

Une autre contrainte sur la zone de recherche existe (*i.e.* un parallélogramme), se reporter à Itakura (1975) et Myers *et al.* (1980) pour plus d'information. Comme déjà évoqué, la distance $D_w(T, R_k)$ relate de la dissimilarité entre de la vocalise T et chaque modèle R_k du dictionnaire. Ainsi, plus cette distance est petite, plus les séquences (contours) comparées sont similaires. La séquence T est ainsi identifiée au modèle R_k pour lequel la distance

$D_w(T, R_k)$ est minimale. Ce qui revient à écrire

$$T \triangleq \arg\min_k D_w(T, R_k). \tag{2.22}$$

Les références sélectionnées pour créer le dictionnaire représentent cinq types de vocalises, les vocalises de type A, B, D, 20 Hz et les phrases AB (*cf.* chapitre 1). L'ajout d'une classe de bruit (*i.e.* mauvais contours extraits) n'est pas envisageable. En effet, l'infinité de contours possible pour le bruit rend sa modélisation (*cf.* section 2.3.4.1) très difficile, affaiblissant ainsi de façon importante la performance de l'identification. Les contours pour lesquels la distance avec le modèle de référence le plus proche excède une valeur fixée, T_{DTW} (*cf.* tableau A.1 en annexe), sont donc considérés trop « dissimilaires » et sont rejetés.

Le choix des modèles de références est l'élément qui conditionne le plus la performance de l'identification. C'est pourquoi une procédure de sélection semi-supervisée des modèles de référence est mise en place (Rabiner et Juang, 1993). L'ensemble des modèles de référence sont issus de la base de données d'entraînement et sont déterminés comme suit. Premièrement, en phase d'initialisation, un modèle de chaque type de vocalise à reconnaître est sélectionné par l'utilisateur. Deuxièmement, chaque contour extrait de la base de données d'apprentissage est comparé aux modèles de références grâce à l'algorithme DTW, pour être identifié. Si l'identification est correcte, cela implique que le contour testé est déjà bien représenté, il n'est alors pas ajouté aux références. Si l'identification n'est pas correcte (jugement de l'expérimentateur), le contour testé est ajouté aux modèles de références. Cette routine est répétée jusqu'à la dernière vocalise de la base de donnée d'apprentissage. Cette procédure permet ainsi de minimiser les redondances. Le dictionnaire final est composé de 12 références de vocalises A, 17 de vocalises B, 7 de phrases AB, 19 de vocalises D et 2 de vocalises de 20 Hz.

2.3.4.3 Classification des contours par quantification vectorielle

La quantification vectorielle est à la base une technique de groupement qui peut aussi être utilisée comme une méthode de classification en faisant référence à des modèles. Elle a été développée principalement reconnaissance de la parole pour le codage et l'identification du locuteur (Pan *et al.*, 1985; Soong *et al.*, 1985), mais est utilisée dans divers problèmes de classification comme par exemple les sons respiratoires (Bahoura et Pelletier, 2003) et les caractères manuscrits (Camastra et Vinciarelli, 2001). La méthode opère en deux phases, l'apprentissage et la classification.

L'extraction des descripteurs consiste à caractériser chaque contour de vocalise extrait (*cf.* section 2.3.4.1) par un vecteur à quatre dimensions composé de la fréquence minimale, de la fréquence maximale, de la durée et de la différence de fréquence entre le début et la fin du contour. Parmi les autres descripteurs testés[2] (résultats non reportés dans ce document), la combinaison citée ci-dessus permet d'obtenir la meilleure discrimination.

Lors de la phase d'apprentissage, chaque vocalise extraite de la base de données d'apprentissage est représentée par un point dans un espace des descripteurs à quatre dimensions (figure 2.12). Chaque classe de vocalise, k, est donc représentée par un nuage de N_k points $X_k = \{x_{k1}, x_{k2}, \cdots, x_{kN_k}\}$, où N_k est le nombre d'exemplaires de cette vocalise. Le processus d'apprentissage, Q, consiste à faire correspondre à chacun de ces nuages de points, un nombre restreint, M, de régions dans le même espace (Linde *et al.*, 1980). Chaque région, peut être représentée par son centroïde. Pour une classe, k, de vocalises données, ces centroïdes, constituent le *dictionaire* $C_k = \{c_{k1}, c_{k2}, \cdots, c_{kM}\}$ de cette classe. On peut noter ce processus par

$$Q : X_k \mapsto C_k. \tag{2.23}$$

Chaque dictionnaire est construit grâce à l'algorithme *LBG* (Linde *et al.*, 1980). Bien que non utilisé pour cette étude, l'algorithme *k-means* peut également être utilisé (Rabiner et Juang, 1993).

[2] *i.e.* Fréquence de début, fréquence de fin, position de la fréquence maximale, position de la fréquence minimale

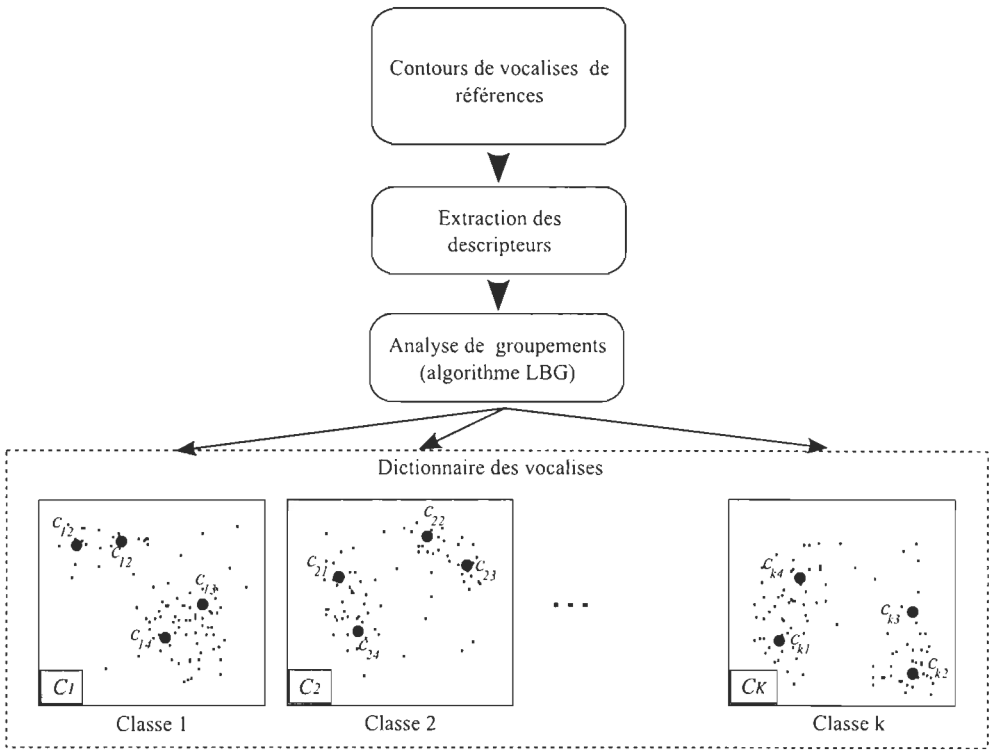

Figure 2.12 – Schéma du principe de l'étape d'apprentissage de l'algorithme VQ (illustration pour un espace à deux dimensions avec un nombre de régions $M = 4$). Chaque point (X_k) représente une vocalise extraite de la base de données d'apprentissage. Les gros points (c_{ki}) sont les centroïdes des différentes régions.

En phase de classification, le contour d'une vocalise inconnue, T, est caractérisé puis représenté dans l'espace des descripteurs. La distance

$$D(T, C_k) = \frac{1}{M} \sum_{i=1}^{M} d(T, c_{ki}), \tag{2.24}$$

séparant T de C_k est calculée pour l'ensemble des classes de vocalises. Le terme $d(T, c_{ki})$ dans l'équation (2.24) est la distance euclidienne séparant T du centroïde c_{ki}. La séquence T est identifiée à la classe pour laquelle la distance $D(T, C_k)$ est minimale. Ce qui revient à écrire

$$T \stackrel{\triangle}{=} \arg\min_{k} D(T, C_k). \tag{2.25}$$

Comme pour la méthode *DTW*, le dictionnaire est composé des cinq classes de vocalises, A, B, phrase AB, D et vocalise de 20 Hz. Lorsque la plus petite distance $D(T, C_k)$ excède la valeur T_{VQ} (*cf.* tableau A.1 en annexe), T est considéré comme étant du bruit et est rejeté.

2.3.5 Ajustement des paramètres

L'ajustement des paramètres des méthodes utilisées dans cette étude est fait de façon empirique. Chaque méthode est testée pour chaque type de vocalise sur la base de données d'apprentissage (*cf.* section 2.2). Les paramètres sont ajustés manuellement un par un. La combinaison de paramètres choisie est celle qui permet d'obtenir le moins de fausses détections et de vocalises oubliées sur la base de données. Les valeurs de tous les paramètres sont reportées en annexe A.

2.4 Évaluation de la performance

Afin d'évaluer la performance des différentes méthodes exposées précédemment, chacune d'elle est testée sur la base de données de test (*cf.* section 2.2). Deux critères sont évalués : la performance de la reconnaissance et la rapidité d'exécution.

2.4.1 Performance de la reconnaissance

Deux grandeurs relatant la performance de la méthode sont alors mesurées, les faux négatifs et les faux positifs. Les faux négatifs correspondent aux vocalises présentes dans la base de données de test contenant les vocalises qui ne sont pas détectées par la méthode. Pour un type donné de vocalise, le taux de faux négatifs, TFN, est exprimé par le rapport du nombre de vocalises « oubliées », N_d, sur le nombre total de vocalises de la base de données, N_{voc}. Ce qui revient à écrire

$$TFN = \frac{N_d}{N_{voc}} \tag{2.26}$$

Les vocalises de la base de données sont préalablement identifiées et positionnées dans le temps par un opérateur expérimenté (*cf.* section 2.2), le calcul du taux de faux négatifs peut ainsi s'effectuer de façon automatique.

Les faux positifs correspondent aux fausses détections, c'est à dire à des périodes de bruit détectées comme étant des vocalises. Pour chaque type de vocalise, la méthode est testée sur la base de données contenant uniquement du bruit (*cf.* section 2.2), le nombre de faux positifs est défini comme le nombre de détections total. Le résultat obtenu est divisé par la durée des enregistrements de bruit (*i.e.* 14 heures) pour être exprimé en nombre de faux positifs par heure.

Pour une description approfondie de la performance des méthodes testées, les indices cités ci-dessus sont évalués pour différentes conditions de bruit. Pour chaque vocalise de la base de données de test, un rapport signal sur bruit (noté RSB) est estimé. Théoriquement, ce rapport est décrit par l'expression

$$RSB_{dB} = 10 \log \left(\frac{P_{voc}}{P_{bruit}} \right), \tag{2.27}$$

où P_{voc} et P_{bruit} sont respectivement la puissance de la vocalise et la puissance du bruit. L'équation (2.27) implique que l'on a accès à la puissance de la vocalise seulement et à la

puissance du bruit seulement, ce qui n'est pas notre cas. P_{voc} et P_{bruit} sont donc estimés en suivant la démarche décrite par Mellinger (2004) et Mellinger et Clark (2006). P_{voc} est alors définie, à partir du spectrogramme, comme étant la puissance moyenne dans l'intervalle de temps de la vocalise dans la bande de fréquence 15-18 Hz pour les vocalises A et B, 18-26 Hz pour les vocalises de 20 Hz et 30-100 Hz pour les vocalises D. P_{bruit} est défini comme étant la puissance moyenne T_{voc} secondes avant et T_{voc} secondes après la vocalise dans la même bande de fréquence. La valeur de T_{voc} est définie par la durée de la vocalise observée. Il est à noter que ce calcul est effectué sur le spectrogramme initial S (*i.e.* sans toutes les étapes d'atténuation du bruit). Les indices de performance sont évalués séparément pour les vocalises avec un RSB inférieur à 0 dB, celles avec un RSB allant de 0 à 5 dB, celles avec un RSB allant de 5 à 10 dB et celles avec un RSB supérieur à 10 dB.

2.4.2 Rapidité d'exécution

Afin d'estimer la rapidité d'exécution des différentes méthodes un indice de temps-réel, I_r est calculé. Il est défini comme le rapport de la durée, T_p, du processus de reconnaissance d'un enregistrement et de la durée T_r du même enregistrement. On peut alors écrire

$$I_r = \frac{T_p}{T_r}. \tag{2.28}$$

Grâce à cet indice, on peut définir si une méthode, pour un type de vocalise donné, fonctionne en temps-réel ($I_r \leq 1$) ou non ($I_r > 1$).

La rapidité d'exécution des méthodes de reconnaissance est très dépendante du matériel informatique utilisé. Le calcul de l'indice de temps-réel permet donc principalement de comparer la rapidité des méthodes entre elles. L'ordinateur utilisé pour cette étude est équipé d'un processeur Intel pentium® 4 cadencé à 3.20 GHz avec une mémoire vive de 1 Gb et fonctionne avec Windows XP professionnel (version 2002, service pack 2).

Chapitre 3

Résultats

3.1 Performance de reconnaissance

Les trois méthodes utilisées sont testées sur les bases de données, décrites par intervalles de rapports signal sur bruit (*RSB*), sur les graphiques 3.1a, 3.2a, 3.3a et 3.4a. Les enregistrements des bases de données de vocalises A, B, D et 20 Hz constituent respectivement une durée totale de 17, 17, 7 et 3 heures. Les vocalises A et B les plus fréquentes ont un *RSB* compris entre 0 et 10 dB. Pour les vocalises de 20 Hz, le nombre de vocalises est réparti de façon égale dans les intervalles de *RSB* supérieurs à 0 dB et est négligeable en dessous. Les vocalises D dont le *RSB* est compris entre 0 et 5 dB constituent la majorité de la base de données tandis que celles supérieures à 10 dB sont négligeables.

Les figures 3.1b, 3.1c, 3.1d et 3.2b, 3.2c, 3.2d, présentent respectivement les pourcentages de faux négatifs obtenus pour la vocalise A et B pour chacune des méthodes. Pour chaque méthode, le taux de faux négatifs diminue quand le *RSB* augmente. Pour ces 2 types de vocalises, le taux de faux négatifs total le plus faible est obtenu par la coïncidence des spectrogrammes (8,5 % pour la vocalise A et 19 % pour la vocalise B). Le nombre minimum de fausses détections par heure sur les données ne contenant que du bruit (tableau 3.1), est réalisé par la coïncidence des spectrogrammes pour la vocalise A (trois par heure) et par la méthode *DTW* pour les vocalises B (une par heure).

Les graphiques 3.3b, 3.3c et 3.3d présentent les résultats de faux négatifs de la vocalise D pour les trois méthodes testées. Contrairement aux vocalises A et B, le taux de faux négatifs augmente pour les vocalises avec un *RSB* supérieur à 10 dB pour les méthodes *DTW* (figure 3.3a) et *VQ* (figure 3.3b). Le comportement des taux de faux négatifs pour la coïncidence des spectrogrammes est similaire à celui déterminé pour les vocalises A et B (*i.e.* décroissance du taux de faux négatifs lorsque le *RSB* augmente). La méthode *VQ* permet une reconnaissance des vocalises D avec un taux global de faux négatifs légèrement plus faible que les autres méthodes (33,5 %). *DTW* est la méthode la plus performante pour le taux de fausses détections par heure (tableau 3.1).

Les graphiques 3.4b, 3.4c et 3.4d présentent les taux de faux négatifs des trois méthodes de cette étude, pour la vocalise de 20 Hz du rorqual commun. L'ensemble des résultats obtenus suit le comportement attendu (*i.e.* une décroissance du taux de faux négatifs quand le *RSB* augmente). La technique utilisant l'identification *VQ* est celle qui détecte le plus de vocalises de 20 Hz, les taux globaux de faux négatifs étant respectivement 33,3 %, 27,9 % et 35,8 % pour les méthodes *DTW*, *VQ* et la coïncidence des spectrogrammes. La méthode *VQ* présente cependant trois fois plus de fausses détections sur un signal bruité que la méthode de coïncidence des spectrogrammes (tableau 3.1), la plus performante par ce critère.

Tableau 3.1 – Nombre de faux positifs par heure pour les trois méthodes utilisées

	Voc. A	Voc. B	Voc. D	Voc. 20 Hz
DTW	4	1	11	63
VQ	4	3	12	74
Coïnc. spec.	3	2	22	23

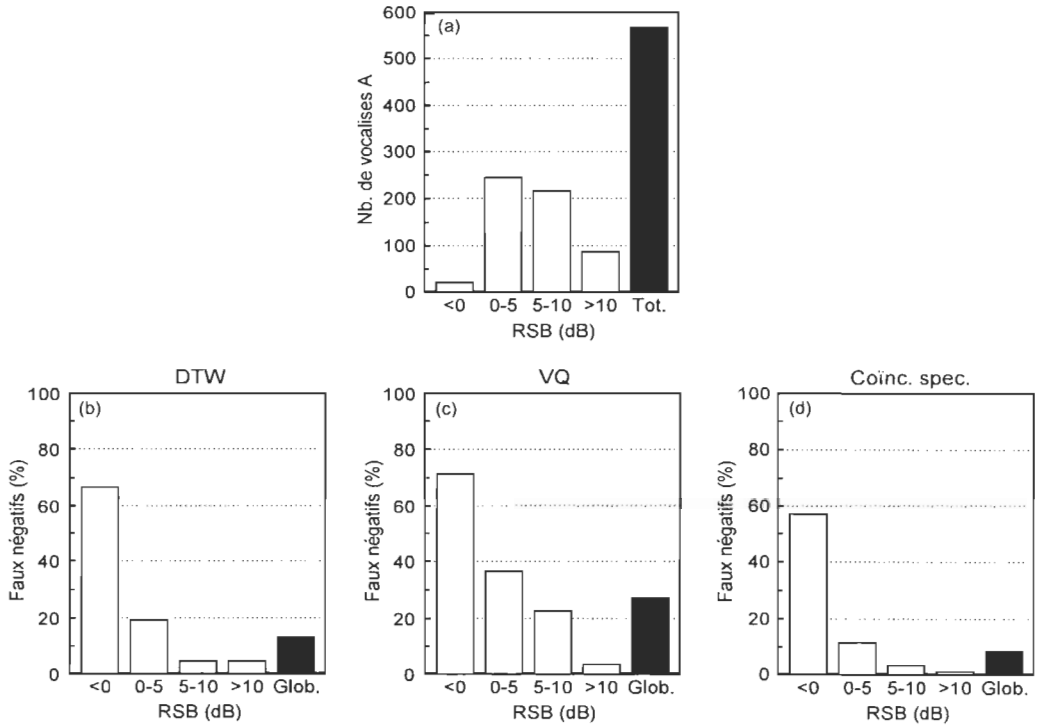

Figure 3.1 – Performance des trois méthodes pour la vocalise A du rorqual bleu. (a) Répartition des vocalises de la base de données par intervalles de *RSB*, (b) taux de faux négatifs par intervalles de *RSB* obtenus avec la méthode *DTW*, (c) la méthode *VQ* et (d) la méthode de coïncidence des spectrogrammes. Les barres noires indiquent le taux de faux négatifs global (*i.e.* sans distinction de *RSB*).

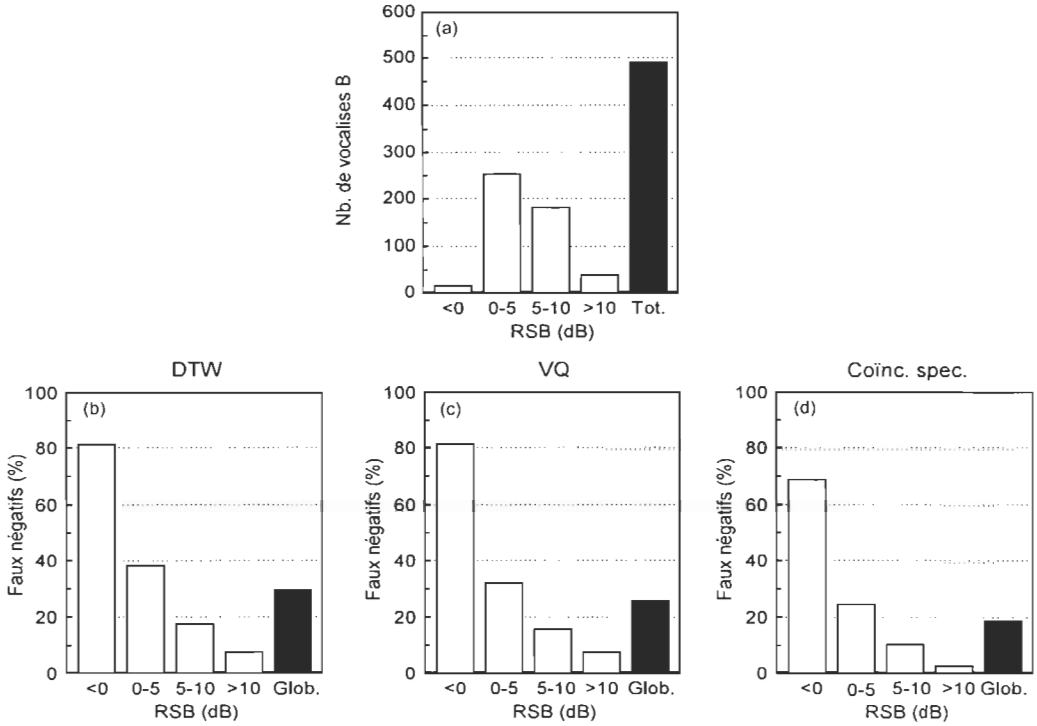

Figure 3.2 – Performance des trois méthodes pour la vocalise B du rorqual bleu. (a) Répartition des vocalises de la base de données par intervalles de *RSB*, (b) taux de faux négatifs par intervalles de *RSB* obtenus avec la méthode *DTW*, (c) la méthode *VQ* et (d) la méthode de coïncidence des spectrogrammes. Les barres noires indiquent le taux de faux négatifs global (*i.e.* sans distinction de *RSB*).

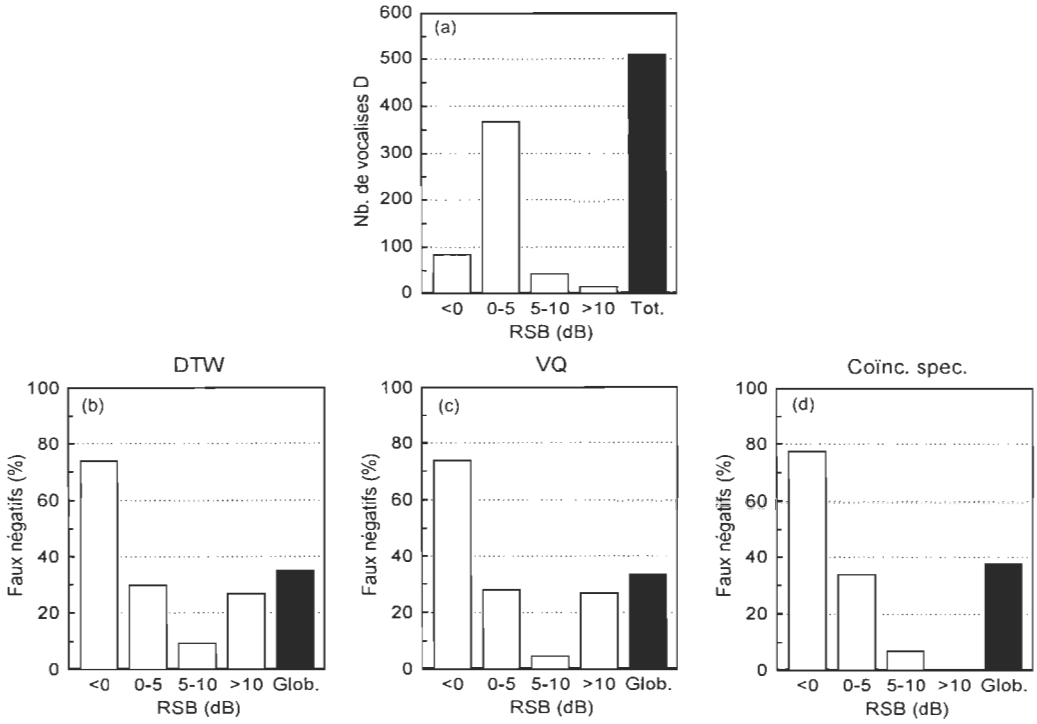

Figure 3.3 – Performance des trois méthodes pour la vocalise D du rorqual bleu. (a) Répartition des vocalises de la base de données par intervalles de *RSB*, (b) taux de faux négatifs par intervalles de *RSB* obtenus avec la méthode *DTW*, (c) la méthode *VQ* et (d) la méthode de coïncidence des spectrogrammes. Les barres noires indiquent le taux de faux négatifs global (*i.e.* sans distinction de *RSB*).

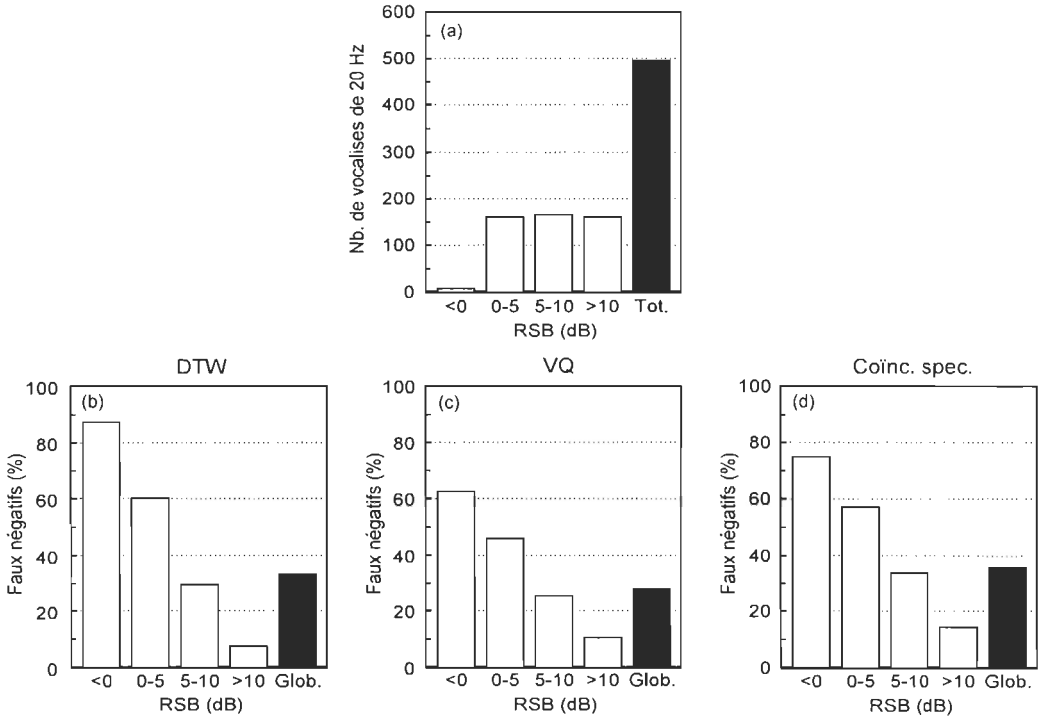

Figure 3.4 – Performance des trois méthodes pour la vocalise de 20 Hz du rorqual commun. (a) Répartition des vocalises de la base de données par intervalles de RSB, (b) taux de faux négatifs par intervalles de RSB obtenus avec la méthode DTW, (c) la méthode VQ et (d) la méthode de coïncidence des spectrogrammes. Les barres noires indiquent le taux de faux négatifs global (*i.e.* sans distinction de RSB).

3.2 Rapidité d'exécution

L'indice de temps-réel est calculé à chaque enregistrement (fichier) de la base de données de test (*cf.* section 2.2) pour chaque catégorie de vocalises. Ainsi, pour une catégorie de vocalise donnée, on obtient autant d'indices de temps-réel qu'il y a d'enregistrements dans la base de données. Pour les vocalises A et B, la reconnaissance s'effectue sur le même spectrogramme (*cf.* section 2.3.1), c'est à dire en même temps, il est donc impossible de distinguer le temps d'exécution de la reconnaissance de la vocalise A de celui de la vocalise B. L'indice de temps-réel est donc calculé pour les trois catégories suivantes : les vocalises A et B, les vocalises D et les impulsions de 20 Hz.

Les figures 3.5a, 3.5b et 3.5c présentent la distribution des indices de temps-réel sous forme de graphiques moustaches (*Box-plots*) pour chaque catégorie de vocalise et pour chaque méthode de reconnaissance. Le temps de calcul de la coïncidence des spectrogrammes (figure 3.5a) est, pour l'ensemble des vocalises, très petit comparé aux autres méthodes. Les temps de calcul obtenus avec la méthode *VQ* (figure 3.5c) sont légèrement plus faibles qu'avec la méthode *DTW* (figure 3.5b). Il est à noté que pour les méthodes *DTW* et *VQ*, l'indice de temps-réel est plus variable pour la vocalise D. Quelque soit le type de vocalise à reconnaître, toutes les méthodes de reconnaissances opèrent plus rapidement que le temps-réel (pour l'ordinateur utilisé dans cette étude, *cf.* section 2.4.2).

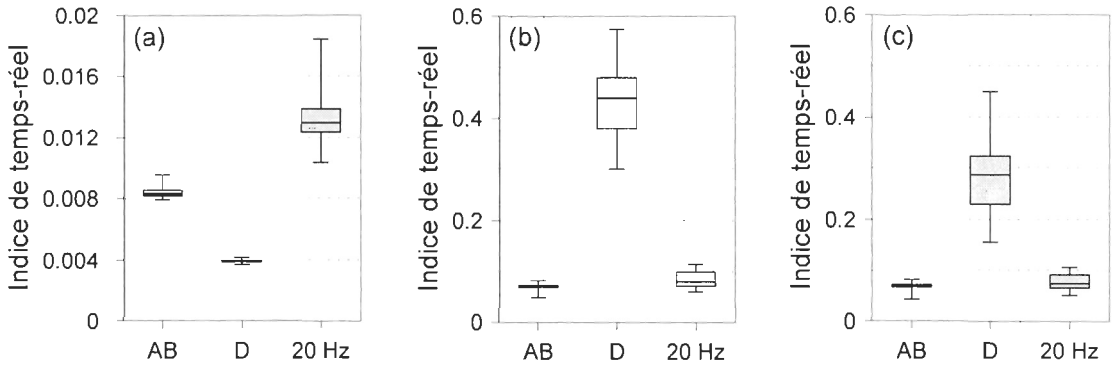

Figure 3.5 – Temps d'exécution des méthodes de reconnaissance pour chaque type de vocalises. Distribution des indices de temps-réel obtenus avec (a) la méthode de coïncidence des spectrogrammes, (b) la méthode *DTW* et (c) la méthode *VQ*. Les graphiques indiquent la médiane (barres horizontales à l'intérieur des boites), les quartiles supérieurs et inférieurs (limites des boites), ainsi que les minimums et maximums (barres) des valeurs de l'indice de temps-réel obtenues pour chaque type de vocalises.

Chapitre 4

Discussion

4.1 Analyse des résultats

4.1.1 Comparaison des méthodes

Deux types d'approches de détection/identification des vocalises ont été testés au cours de cette étude. L'une, la coïncidence des spectrogrammes, est basée sur une approche classique qui a été utilisée dans de nombreuses études (Wiggins *et al.*, 2005; Sirovic *et al.*, 2004), l'autre basée sur l'extraction de contours présente une approche plus nouvelle essayant de palier aux faiblesses de l'approche classique.

Pour les vocalises A et B du rorqual bleu, la coïncidence des spectrogrammes est la méthode la plus performante. La différence de performance entre les méthodes est marquée principalement pour les taux de non-détection. Les vocalises non détectées sont majoritairement dues à des erreurs commises lors de l'étape d'extraction du contour des vocalises. Il arrive que certaines vocalises soient tronquées en plusieurs fragments non connectables. Dans le cas où ces fragments sont trop courts alors ils sont tous supprimés (*cf.* contrainte de temps T_{min}, section 2.3.4.1), la vocalise n'est donc pas considérée. Dans le cas où certains fragments atteignent une longueur suffisante, ils sont présentés à l'algorithme d'identification (DTW ou VQ) mais sont ignorés (*i.e.* identifiés comme du bruit) à cause

d'une fragmentation trop importante. Pour la vocalise de 20 Hz du rorqual commun, la coïncidence des spectrogrammes se démarque très nettement par son taux de fausses détections beaucoup moins élevé que les deux autres méthodes. En effet, même si les taux de faux négatifs sont semblables, voir meilleurs pour les méthodes *DTW* et *VQ* (figure 3.4), la coïncidence des spectrogrammes obtient plus de trois fois moins de fausses détections (tableau 3.1). La majorité des fausses détections est liée à la présence de bruits impulsifs en basses fréquences (*e.g.* battements des hélices d'un navire), identifiés comme des vocalises de 20 Hz. Pour la vocalise D, l'approche par extraction de contours de vocalises est la plus performante. Le modèle figé utilisé lors de la coïncidence des spectrogramme ne permet pas de s'adapter à toutes les variations de durées, de pentes et de bandes de fréquences des vocalises D (Berchok *et al.*, 2006; Oleson *et al.*, 2007). D'autre part, la coïncidence des spectrogrammes s'avère être très vulnérable aux bruits de vibration du système de mouillage (*strumming*) générant un nombre plus important de fausses détections (tableau 3.1).

La différence de performance des algorithmes d'identification *VQ* et *DTW* n'est pas très importante. L'algorithme *DTW* est légèrement plus performant que l'algorithme *VQ*. Sur l'ensemble des vocalises testées, les taux de faux négatifs sont similaires, les différences de performance sont surtout notables sur les fausses détections (tableau 3.1). Ces légères différences de performance sont probablement dues à un mauvais ajustement manuel du seuil de classification T_{VQ} (*cf.* section 2.3.4.3). Une méthode d'ajustement automatique des paramètres comme celle proposée dans la section 4.2.1 permettrait de pallier à cette faiblesse.

Comme attendu, la performance des méthodes utilisées (*i.e.* le taux de faux négatifs) augmente en fonction du rapport signal sur bruit. Une exception cependant est à noter pour les méthodes utilisant les algorithmes *VQ* et *DTW* pour identifier la vocalise D du rorqual bleu (figure 3.3). En effet, l'augmentation du taux de faux négatifs pour les vocalises D avec un *RSB* supérieur à 10 dB semble plutôt être expliquée par les distorsions des vocalises liées à la propagation sonore dans le milieu que par le bruit lui même.

L'effet des trajets multiples succédant au trajet direct des vocalises peut engendrer une déformation des contours extraits sur le spectrogramme qui les rend non-identifiables lors de la phase de classification (*i.e.* reconnus comme du bruit). Le nombre restreint de vocalises D de la base de données avec un *RSB* supérieur à 10 dB (15 vocalises) ne permet cependant pas de tirer de conclusion générale quant à la performance de ces deux méthodes pour des vocalises dans cet intervalle de *RSB*.

L'approche classique de coïncidence par spectrogrammes présentée dans la littérature (Mellinger et Clark, 1996, 2000) a été modifiée en ajoutant un processus de réduction du bruit adapté aux conditions acoustiques du Saint-Laurent (*cf.* 2.3.2). Ce processus joue un rôle majeur pour la reconnaissance des vocalises. De par la simplicité de l'opération logique *AND*, l'étape de coïncidence est très rapide et permet de répondre à la contrainte de temps-réel. Par contre, l'ajustement des paramètres du modèle temps-fréquence des vocalises est laborieux et n'est pas optimisé (*i.e.* ajustement par essais-erreurs supervisé par un expérimentateur, *cf.* 2.3.3). La méthode de coïncidence des spectrogrammes permet d'obtenir de bon résultats de détections pour les vocalises avec un patron temps-fréquence peu variable (vocalises A et B) mais est moins adaptée pour les vocalises plus variables.

L'approche par extraction puis identification des contours temps-fréquence de vocalises a été mis en place afin d'améliorer la reconnaissance des vocalises variables en durée et fréquence. Le temps de calcul de l'étape d'extraction est variable et est dépendant de la complexité des enregistrements. Cependant, les vocalises auxquelles nous nous intéressons ici ont une structure relativement élémentaire (*i.e.* contours simples sans harmoniques), la durée du calcul n'est donc pas importante. L'étape d'identification n'est elle non plus pas exigeante en temps de calcul. On notera que l'algorithme *DTW* demande un temps de calcul, dépendant du nombre de références du dictionnaire, plus importante que l'algorithme *VQ* (figure 3.5). L'apprentissage des modèles de connexions et des dictionnaires est aisé et robuste puisqu'il s'effectue de façon semi-supervisée sur une base de données de vocalises. La performance des algorithmes de connexion des segments et d'identification est donc directement dépendante des vocalises présentes sur la base de données. Celle-ci

Tableau 4.1 – Synthèse de la performance des méthodes pour les conditions du Saint-Laurent

	DTW	*VQ*	*Coïnc.spec.*
Qualité de l'identification			
Voc. A	***	**	****
Voc. B	**	**	***
Voc. D	***	***	**
Voc. 20-Hz	*	*	**
Rapidité d'exécution	***	***	****
Facilité de l'entraînement	***	***	* [1]

**** Très bonne *** Bonne ** Moyenne * Médiocre

doit donc être la plus représentative possible des conditions acoustiques du milieu d'intérêt. Cette remarque implique que cette approche peut être adaptée à d'autres vocalises et/ou d'autres milieux. Cette étude montre donc que cette approche a un potentiel intéressant pour la reconnaissance des vocalises de mammifères marins. Ses performances sont encore légèrement inférieures à celles obtenues par l'approche classique pour les vocalises stéréotypées, cependant quelques modifications pourraient être envisagées pour permettre de les améliorer (*cf.* section 4.2.1).

Le tableau 4.1 synthétise les performances des trois méthodes utilisées dans cette étude.

4.1.2 Utilisation comme outils de monitorage

Qu'en est il de la performance des méthodes testées dans le contexte concret d'un suivi d'individus ? Les résultats présentés précédemment (cf. chapitre 3) relatent la performance de chaque méthode. Ils sont très utiles pour évaluer, améliorer et comparer les méthodes entres elles, cependant, lorsque l'on s'intéresse à un problème concret tel que le monitorage d'une espèce, ils ne doivent pas être interprétés de la même manière. Ainsi, pour bien

[1] Il est à noter qu'en utilisant un modèle empirique (*i.e.* créé à partir de plusieurs vocalises d'un jeu de données) ce résultat serait différent.

Tableau 4.2 – Nombre de détections anticipées sur une heure en assumant que l'animal vocalise régulièrement sans interruptions.

	# vocalises	# détections	# vraies	# fausses
Vocalise A				
Coïnc. spec.		50	47	3
DTW	51	48	44	4
VQ		41	37	4
Vocalise B				
Coïnc. spec.		44	42	2
DTW	51	37	36	1
VQ		41	38	3
Vocalise 20 Hz				
Coïnc. spec.		253	231	22
DTW	360	251	240	11
VQ		272	260	12

interpréter les résultats, il est intéressant de les replacer dans le contexte rythmique de chacune des vocalises.

Les vocalises A et B du rorqual bleu sont répétées approximativement toutes les 70 secondes (*cf.* chapitre 1), donc si une baleine bleue vocalisait pendant une heure sans interruptions, il y aurait environ 51 vocalises A et 51 vocalises B. Si l'on se reporte aux vocalises oubliées (figures 3.1 et 3.2) et aux fausses détections (tableau 3.1) engendrées par la coïncidence des spectrogrammes, il y aurait alors pour la vocalise A, trois fausses détections et quatre vocalises oubliées. Autrement dit, le système trouverait 50 détections dont trois sont fausses. En répétant ce scénario pour la vocalise B, on aurait sur les 51 vocalises, 44 détections dont 2 sont fausses. Le tableau 4.2 regroupe le même type de données obtenues en suivant le même raisonnement pour toutes les méthodes utilisées et pour les vocalises A, B et 20 Hz.

Pour le rorqual bleu, quelle que soit la méthode utilisée, on obtient globalement une quarantaine de vraies détections de vocalises A et de vocalises B par heure, ce qui permet un échantillonnage dans le temps relativement suffisant pour estimer avec précision la position de l'animal étant donné sa faible vitesse moyenne de nage (*i.e.* entre 1 et 5

m/s pour un rorqual commun selon Goldbogen *et al.* (2006)). Il est à noter que pour la localisation des individus, plusieurs capteurs sont nécessaires ; ainsi un filtrage des fausses détections peut être effectué. Pour éliminer les fausses détections, une condition peut être mise en place, stipulant qu'une détection est validée seulement si elle est présente sur trois capteurs au moins par exemple. Ainsi, si le réseau de capteurs est assez large (*i.e.* conditions de bruit différentes sur chaque capteur), les fausses détections peuvent être largement diminuées. Si jamais une série de fausses détections entraînait de fausses localisations, un algorithme de suivi (« *tracking* ») serait en mesure d'éliminer ces points aberrants (*e.g.* Seebaruth, 2006). Pour le rorqual commun, la même remarque s'applique, quelles que soient les méthodes utilisées, plus de 200 vraies détections sont présentes sur une durée d'une heure. Malgré le nombre plus élevé de fausses détections, l'abondance de vraies données par heure s'avère suffisante pour assurer le monitorage de cet animal via un algorithme de localisation.

Il est moins pertinent de prévoir la performance des méthodes de détections pour le suivi de la baleine bleue en utilisant la vocalise D car elles ne sont pas répétées de façon régulières dans le temps. Les méthodes utilisées au cours de ce travail s'avèrent donc utilisables dans un processus de suivi d'animaux (*i.e.* détection puis localisation).

Quelques éléments sont à prendre en considération quand à l'interprétation des résultats qui est faite ici. Premièrement, les résultats sont interprétés pour le cas où l'animal vocalise régulièrement sans interruptions pendant une heure. Bien que les vocalises A, B et 20 Hz soient connues pour être répétées régulièrement sur de longues périodes (Berchok *et al.*, 2006; Mellinger et Clark, 2003; Samaran, 2004), il est possible de voir des séquences de vocalises répétées de façon irrégulière (Oleson *et al.*, 2007). Deuxièmement, le taux de fausses détections est estimé sur une base de données de bruit de 14 heures (*cf.* section 2.2) et est ensuite ramené en fausses détections par heure (*i.e.* une moyenne). Cet indice donne un bon aperçu de la quantité de fausses détections, mais il est à noter que les fausses détections ne sont pas réparties de façon homogène dans les enregistrements. Elles sont très dépendantes du type de bruit présent. Il est donc possible d'observer des

périodes avec un nombre plus important de fausses détections que spécifié dans le tableau 3.1, comme il est également possible d'en observer moins, voir aucune.

4.2 Perspectives

4.2.1 Améliorations

Dans cette section sont évoquées plusieurs suggestions d'améliorations envisageables afin d'accroître la performance des méthodes de reconnaissance (uniquement les méthodes utilisant l'approche par extraction de contours sur le spectrogramme).

La méthode de classification *DTW* est plus coûteuse en temps de calcul que la méthode *VQ* (figure 3.5). Ceci est dû principalement au fait que chaque classe est représentée par plusieurs versions de vocalises (*i.e.* plusieurs références pour une même vocalise). La distance est calculée entre la vocalise inconnue et chacune des séquences de référence de chaque classe de vocalise. Afin de limiter le calcul, Rabiner et Juang (1993), proposent plusieurs méthodes de groupement (*clustering*) permettant de modéliser une seule référence par classe qui décrit la variabilité de l'ensemble des versions de vocalises. Un tel procédé permettrait de réduire de façon importante le temps de calcul sans diminuer la performance de la classification.

La représentation précise des vocalises A, B, D et de 20 Hz dans le plan temps-fréquence implique l'utilisation de trois jeux différents de paramètres de spectrogrammes afin d'obtenir la résolution nécessaire dans la bande de fréquence et la durée de la vocalise. Un pour les vocalises A et B, un autre pour la vocalise D, et un dernier pour la vocalise de 20 Hz. Pour la reconnaissance par extraction de contours, il n'est donc pas possible d'extraire puis d'identifier les contours de toutes les vocalises simultanément. Le processus de reconnaissance doit être répété sur chaque spectrogramme. Afin d'alléger ce processus, il serait intéressant de représenter toutes les vocalises sur un même plan temps-fréquence. D'autres méthodes de représentation temps-fréquence telle que la transformée de Wigner-

Ville et ses dérivées (Flandrin, 1993) pourraient ainsi être explorées. Cette transformation a l'avantage de fournir une décomposition temps-fréquence sans aucune restriction sur les résolutions temporelles et fréquentielles et ne nécessite pas d'*a priori* sur le signal lui-même. Cette distribution est de ce fait adaptée à l'analyse des signaux non stationnaires et a déjà été exploitée pour des problèmes de classification de signaux (Boashash et O'Shea, 1988; Chouvarda *et al.*, 2003; Caimi et Hassan, 2000).

Le choix des paramètres pour l'optimisation des méthodes de reconnaissance est un autre point important à améliorer. Toutes les méthodes de reconnaissance sont paramétrées par un nombre plus ou moins important de variables. Une étape importante au développement d'une méthode est de bien choisir la valeur de chacun des paramètres afin que sa performance soit maximale. Le choix manuel de ces valeurs (*i.e.* ajustement des paramètres un par un) est très laborieux, coûteux en temps et surtout ne permet pas d'optimiser la méthode testée. Le choix exhaustif des combinaisons possibles pour les valeurs des paramètres serait une alternative permettant de trouver la combinaison maximisant la performance, cependant irréalisable étant donné le très grand nombre de combinaisons possibles. Une des solutions possible est l'utilisation de méthodes statistiques telles que les plans d'expériences (*design of experiment*) (Lochner et Matar, 1990). Cette méthode, surtout utilisée en contrôle de la qualité pour les entreprises, permet à partir d'un petit échantillon de toutes les combinaisons possibles de paramètres, de déterminer l'influence de chaque paramètre par rapport aux autres pour ensuite définir statistiquement une combinaison maximisant la sortie de la fonction testée. L'utilisation de cette méthode pourrait ainsi permettre d'optimiser les méthodes décrites dans cette étude pour accroître leur performance.

4.2.2 Autres applications

Il peut être envisageable d'utiliser l'approche par extraction de contour afin de reconnaître d'autres vocalises d'animaux. Certaines vocalises d'odontocètes (Belikov et Bel'kovich, 2006; Buck et Tyack, 1993) ou d'oiseaux (Laiolo *et al.*, 2000) suivent une même

forme en temps et fréquence mais avec des durées variables et/ou produites à différentes hauteurs fréquentielles.

Concernant le premier point, on a vu que les algorithmes DTW et VQ étaient adaptés à la reconnaissance de vocalises contractées ou dilatées en temps. Ainsi, si les vocalises ont seulement des différences de durées, la méthode consistant en l'extraction des contours et l'identification par DTW ou VQ est pertinente.

Concernant le second point, la variabilité de tonalité observée sur certaines vocalises implique qu'il est nécessaire de faire quelques modifications sur les contours extraits, pour que la méthode soit applicable à ce type de vocalises. Une possibilité serait de soustraire la moyenne fréquentielle à chaque contour. Chaque vocalise serait alors représentée par son contour en fréquence centré à zéro. Deux vocalises de hauteurs différentes se verraient alors toutes les deux ramenées au même référentiel fréquentiel relatif.

Des premiers essais ont été effectués sur des vocalises de bélugas (*Delphinapterus leucas*) et de phoques barbus (*Erignatus barbatus*) enregistrées dans l'Arctique canadien (Simard *et al.*, 2006a) et les résultats, après quelques modifications pour l'étape d'atténuation du bruit spécifique à cet environnement, se montrent encourageants.

Annexe A

Paramètres utilisés pour les méthodes de détection et de reconnaissance

Tableau A.1 – Valeur des paramètres utilisés pour l'approche par extraction des contours

Paramètres	Voc. A	Voc. B	Voc. D	Voc. 20 Hz
F_s (Hz)	2000	2000	400	400
Spectrogramme				
Taille de fenêtre (pts)	4096	4096	64	128
Fonction d'apodisation	Kaiser (4)	Kaiser (4)	Hamming	Hamming
Recouvrement (%)	93	93	75	87,5
Taille FFT (pts)	12288	12288	192	384
Résolution fréq. (Hz)	0,16	0,16	2,08	1,04
Résolution temp. (s)	0,14	0,14	0,04	0,04
f_{min} (Hz)	15	15	30	18
f_{max} (Hz)	18,5	18,5	115	30
Atténuation du bruit				
Δl (s)	30	30	15	35
l	4	4	3	3
c	8	8	11	11
δ_1	-0,16	-0,16	-0,16	-0,16
δ_2	0,7	0,7	0,5	0,9
λ	0,996	0,996	0,992	0,993
d_1 (s)	5	5	1	1
δ_3	0,25	0,25	0,6	0,7
Extraction des contours				
d_2 (s)	3	3	1	0,3
T_{seg} (s)	6,5	6,5	2,5	0,5
T_c (%)	20	20	30	70
T_{min} (s)	3,5	3,5	1	0,5
T_{max} (s)	35	35	9	3
Classification DTW				
T_{DTW}	0,2	0,2	5	1,3
Classification VQ				
m	16	16	16	16
T_{VQ}	1	1	18	4

Tableau A.2 – Valeur des paramètres utilisés pour l'approche par coïncidence des spectrogrammes

Paramètres	Voc. A	Voc. B	Voc. D	Voc. 20 Hz
F_s (Hz)	200	200	200	200
Spectrogramme				
Taille de fenêtre (pts)	512	512	128	64
Fonction d'apodisation	Hamming	Hamming	Hamming	Hamming
Recouvrement (%)	87,5	87,5	75	87,5
Taille FFT (pts)	4608	4608	128	192
Résolution fréq. (Hz)	0,04	0,04	1,56	1,04
Résolution temp. (s)	0,32	0,32	0,16	0,04
f_{min} (Hz)	15,5	15,5	30	18
f_{max} (Hz)	18,5	18,5	100	26
Atténuation du bruit				
Δt (s)	55	15	55	15
l	3	3	3	3
c	9	6	6	5
δ_1	-0,06	-0,06	-0,1	-0,06
δ_2	0,9	0,6	0,6	1,2
λ	0,99	0,986	0,97	0,994
d_1 (s)	6	9	1	1
δ_3	0,92	0,85	0,92	0,92
Coïncidence des spectrogrammes				
f_1 (Hz)	18,1	17,5	80	26
f_2 (Hz)	17,8	15,8	40	20
D_{voc} (s)	7	9	3	1
Δ_f (Hz)	0,3	0,3	4	1
D_{ini} (s)	2	0,7	1	1
T_{CS} (%)	77	75	67	74

Annexe B

Durées des vocalises de la base de données de test

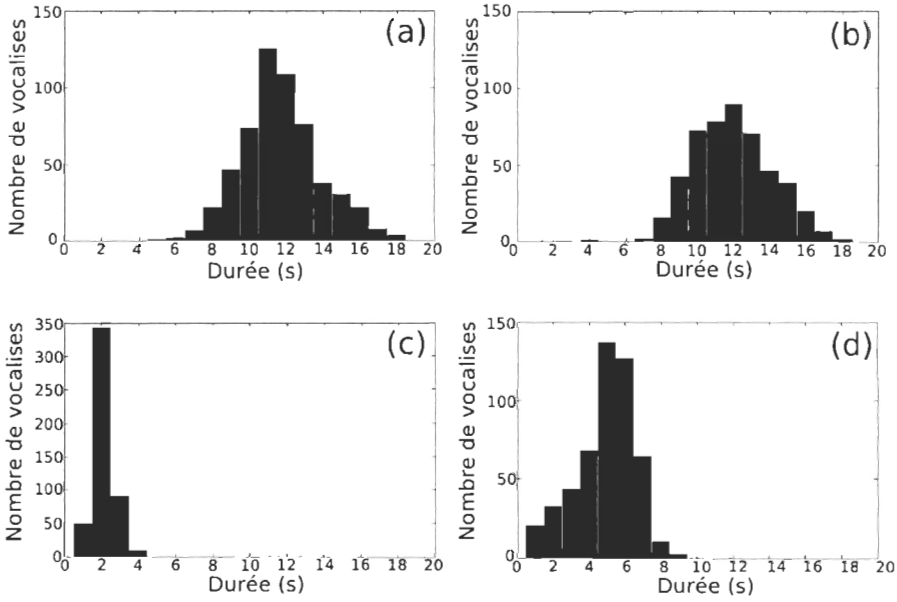

Figure B.1 – Durées des vocalises de la base de données de test. Histogrammes des durées des vocalises A (a), des vocalises B (b), des vocalises de 20 Hz (c) et des vocalises D (d).

Références

Anderson, S. E., Dave, A. S., and Margoliash, D. : Template-based automatic recognition of birdsong syllables from continuous recordings. *J. Acoust. Soc. Am.*, 100:1209–1219, 1996.

Andrew, R. K., Howe, B. M., Mercer, J. A., and Dzieciuch, M. A. : Ocean ambient sound : Comparing the 1960s with the 1990s for a receiver off the California coast. *ARLO*, 3 (2):65–70, 2002.

Anonyme : Draft recovery plan for the fin whale (*Balaenoptera physalus*). , NOAA, National Marine Fisheries Service, Silver Spring, MD., 2006.

Bahoura, M. and Pelletier, C. : New parameters for respiratory sound classification. *In Canadian Conference on Electrical and Computer Engineering, 2003. IEEE CCECE 2003.*, vol. 3, pages 1457–1460, 2003.

Bannister, R. W., Guthrie, K. M., Kay, J. S., Bold, G. E. J., Johns, M. D., Tan, S. M., and Tindle, C. T. : ATOC-New Zealand receiver site survey and acoustic test. *J. Acoust. Soc. Am.*, 93(4):2380–2380, 1993.

Barlow, J. : Abundance of cetaceans in California waters : I. Ship surveys in summer/fall 1991. *Fish. Bull.*, 93:1–14, 1995.

Belikov, R. and Bel'kovich, V. : High-pitched tonal signals of beluga whales (*Delphinapterus leucas*) in a summer assemblage off Solovetskii Island in the White Sea. *Acoustical Physics*, 52(2):125–131, 2006.

Berchok, C. L., Bradley, D. L., and Gabrielson, T. B. : St. Lawrence blue whale vocalizations revisited : Characterization of calls detected from 1998 to 2001. *J. Acoust. Soc. Am.*, 120(4):2340–2354, 2006.

Boashash, B. and O'Shea, P. : Time-frequency analysis applied to signaturing of underwater acoustic signals. *In International Conference on Acoustics, Speech, and Signal Processing. ICASSP-88*, pages 2817–2820 vol.5, 1988.

Boite, R., Bourlard, H., Dutoit, T., Hancq, J., and Liech, H. : *Traitement de la parole.* Presses Polytechniques Romandes, 2000.

Brown, J. C. : Musical fundamental frequency tracking using a pattern recognition method. *J. Acoust. Soc. Am.*, 92(3):1394–1402, 1992.

Brown, J. C., Hodgins-Davis, A., and Miller, P. J. O. : Classification of vocalizations of killer whales using dynamic time warping. *J. Acoust. Soc. Am.*, 119(3):EL34–EL40, 2006.

Brown, J. C. and Zhang, B. : Musical frequency tracking using the methods of conventional and "narrowed" autocorrelation. *J. Acoust. Soc. Am.*, 89(5):2346–2354, 1991.

Buck, J. R. and Tyack, P. L. : A quantitative measure of similarity for *Tursiops truncatus* signature whistles. *J. Acoust. Soc. Am.*, 94(5):2497–2506, 1993.

Caimi, F. M. and Hassan, G. A. : Pattern classification approach to underwater acoustic communications based on the Wigner-Ville distribution. *In Proc. SPIE Int. Soc. Opt. Eng.*, vol. 4045, pages 167–174, Orlando, FL, USA, 2000.

Calambokidis, J. : Underwater behavior of blue whales using a suction-cup attached CRITTERCAM. Grant Number : N00014-00-1-0942, ONR Final technical report, 2002. URL `http://www.cascadiaresearch.org/reports/rep-ONR.pdf`.

Calambokidis, J. : Underwater behavior of blue whales examined with suction-cup attached tags. Grant Number : N00014-02-1-0849, ONR report, 2003. URL `http://www.cascadiaresearch.org/reports/ONR-Rep-03.pdf`.

Calambokidis, J., Steiger, G., Cubbage, J., Balcomb, K., Ewald, C., Kruse, S., Wells, R., and Sears, R. : Sightings and movements of blue whales off central California 1986–88 from photo-identification of individuals. *Rep. Int. Whal. Comm.*, 12:343–348, 1990.

Camastra, F. and Vinciarelli, A. : Cursive character recognition by learning vector quantization. *Pattern Recogn. Lett.*, 22(6-7):625–629, 2001. ISSN 0167-8655.

Charif, R. A., Mellinger, D. K., Dunsmore, K. J., Fristrup, K. M., and Clark, C. W. : Estimated source levels on fin whale (*Balaenoptera physalus*) vocalizations : Adjustment for surface interference. *Mar. Mamm. Sci.*, 18(1):81–98, 2002.

Chouvarda, I., Maglaveras, N., Boufidou, A., Mohlas, S., and Louridas, G. : Wigner-Ville analysis and classification of electrocardiograms during thrombolysis. *Med. Biol. Eng. Comput.*, 41(6):609–617, 2003.

Clark, C. W., Charif, R., Mitchell, S., and Colby, J. : Distribution and behavior of the bowhead whale, *Balaena mysticetus*, based on analysis of acoustic data collected during the 1993 spring migration off Point Barrow, Alaska. *Rep. Int. Whal. Comm.*, 46:541–552, 1996.

Clark, C. W. and Ellison, W. T. : *Advances in the Study of Echolocation in Bats and Dolphins*, chap. Potential use of low-frequency sounds by baleen whales for probing the environment : Evidence from models and empirical measurements. Plenum, New York, 2003.

Clark, C. : Whale voices from the deep : Temporal patterns and signal structures as adaptations for living in an acoustic medium. *J. Acoust. Soc. Am.*, 103:2957, 1998.

Clark, C. and Fristrup, K. : Whales 95 : A combined visual and acoustic survey of blue and fin whales off southern California. *Rep. Int. Whal. Comm.*, 47:583–600, 1997.

Clemins, P. J. and Johnson, M. T. : Generalized perceptual linear prediction features for animal vocalization analysis. *J. Acoust. Soc. Am.*, 120(1):527–534, 2006.

Clemins, P. J., Johnson, M. T., Leong, K. M., and Savage, A. : Automatic classification and speaker identification of african elephant (*Loxodonta africana*) vocalizations. *J. Acoust. Soc. Am.*, 117(2):956–963, 2005.

COSEPAC : Espèces canadiennes en péril. , Comité sur la situation des espèces en péril au Canada, 2004. URL `http://www.cosepac.gc.ca`. 65 p.

Costa, D. : The secret life of marine mammals, novel tools for studying their behavior and biology at sea. *Oceanography*, 6(3):120–128, 1996.

Croll, D. A., Clark, C. W., Calambokidis, J., Ellison, W. T., and Tershy, B. R. : Effect of anthropogenic low-frequency noise on the foraging ecology of balaenoptera whales. *Animal Conservation*, 4:13–27, 2001.

Croll, D. A., Clark, C. W., Acevedo, A., Tershy, B., Flores, S., Gedamke, J., and Urban, J. : Only male fin whales sing loud songs. *Nature*, 417(6891):809, 2002.

Cummings, W. C. and Holliday, D. V. : Passive acoustic location of bowhead whales in a population census off Point Barrow, Alaska. *J. Acoust. Soc. Am.*, 78(4):1163–1169, 1985.

Cummings, W., Thompson, P., and Ha, S. : Sounds from bryde, *Balaenoptera edeni*, and finback, *B. Physalus*, whales in the gulf of California. *Fish. Bull.*, 84:359–370, 1986.

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

Datta, S. and Sturtivant, C. : Dolphin whistle classification for determining group identities. *Signal Processing*, 82(2):251–258, 2002. ISSN 0165-1684.

Edds, P. L. and MacFarlane, J. A. F. : Occurrence and general behavior of balaenopterid cetaceans summering in the St. Lawrence estuary, Canada. *Can. J. Zool.*, 65(6):1363–1376, 1987.

Edds, P. : Vocalizations of the blue whale, *Balaenoptera musculus*, in the St. Lawrence River. *J. Mamm.*, 63:345–347, 1982.

Edds, P. : Characteristics of finback (*Balaenoptera physalus*) vocalizations in the St. Lawrence estuary. *Bioacoustics*, 1:131–149., 1988.

Fisher, J. : Synopsis mammalium. *Cottae, Stuttgart*, 1829.

Flandrin, P. : *Temps-Fréquence*. Traité des Nouvelles Technologies, série Traitement du Signal. Hermès, Paris, 1993.

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

Goldbogen, J. A., Calambokidis, J., Shadwick, R. E., Oleson, E. M., McDonald, M. A., and Hildebrand, J. A. : Kinematics of foraging dives and lunge-feeding in fin whales. *J. Exp. Biol.*, 209(7):1231–1244, 2006.

Halkias, X. C. and Ellis, D. P. : Call detection and extraction using bayesian inference. *Applied Acoustics*, 67(11-12):1164–1174, 2006.

Hoyt, E. : Whale watching 2001 : Worldwide tourism numbers, expenditures and expanding socioeconomic benefits. , International Fund for Animal Welfare, Yarmouth Port, MA, USA, 2001.

Ichihara, T. : *Whales, dolphins and porpoises*, chap. The pygmy blue whale *Balaenoptera musculus brevicauda*, a new subspecies from the Antarctic, pages 79–113. Berkeley & Los Angeles : Univ. California Press, 1966.

Itakura, F. : Minimum prediction residual principle applied to speech recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 23(1):67–72, 1975. ISSN 0096-3518.

Ito, K., Mori, K., and Iwasaki, S. : Application of dynamic programming matching to classification of budgerigar contact calls. *J. Acoust. Soc. Am.*, 100(6):3947–3956, 1996.

Jahoda, M., Lafortuna, C. L., Biassoni, N., Almirante, C., Azzellino, A., Panigada, S., Zanardelli, M., and Sciara, G. N. : Mediterranean fin whale's (*Balaenoptera physalus*) response to small vessels and biopsy sampling assessed through passive tracking and timing of respiration. *Mar. Mamm. Sci.*, 19(1):96–110, 2003.

Jensen, A. and Silber, G. : Large whale ship strike database. , U.S. Department of Commerce, NOAA Technical Memorandum. NMFS-OPR-, 2003. URL `http://www.nmfs.noaa.gov/pr/overview/publicat.html`.

Johnson, M. and Tyack, P. : A digital acoustic recording tag for measuring the response of wild marine mammals to sound. *IEEE Journal of Oceanic Engineering*, 28(1):3–12, 2003. ISSN 0364-9059.

Kibblewhite, A. C., Denham, R. N., and Barnes, D. J. : Unusual low-frequency signals observed in New Zealand waters. *J. Acoust. Soc. Am.*, 41(3):644–655, 1967.

Ko, D., Zeh, J., Clark, C., Ellison, W., Krogman, B., and Sonntag, R. : Utilization of acoustic location data in determining a minimum number of spring-migrating bowhead whales unaccounted for by the ice-based visual census. *Rep. Int. Whal. Comm.*, 36:325–338, 1986.

Kogan, J. A. and Margoliash, D. : Automated recognition of bird song elements from continuous recordings using dynamic time warping and hidden markov models : A comparative study. *J. Acoust. Soc. Am.*, 103(4):2185–2196, 1998.

Lagerquist, T. B. A., Stafford, K. M., and Mate, B. R. : Dive characteristics of satellite-monitored blue whales (*Balaenoptera musculus*) off the central California coast. *Mar. Mamm. Sci.*, 16(2):375–391, 2000.

Laiolo, P., Palestrini, C., and Rolando, A. : A study of choughs' vocal repertoire : variability related to individuals, sexes and ages. *Journal für Ornithologie*, 141(2):168–179, 2000.

Laist, D. W., Knowlton, A. R., Mead, J. G., Collet, A. S., and Podesta, M. : Collisions between ships and whales. *Mar. Mamm. Sci.*, 17(1):35–75, 2001.

Linde, Y., Buzo, A., and Gray, R. : An algorithm for vector quantizer design. *IEEE Trans. Comm.*, 28(1):84–95, 1980.

Linnæus, C. : *Systema naturæ per regna tria naturæ, secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis. Tomus I. Editio decima, reformata. - pp. [1-4], 1-824. Holmiæ. (Laurentii Salvii).* 1758.

Lochner, R. H. and Matar, J. E. : *Designing for Quality.* Quality Resources, 1990.

McDonald, M. A., Mesnick, S. L., and Hildebrand, J. A. : Biogeographic characterisation of blue whale song worldwide : using song to identify populations. *J. Cetacean Res. Manage.*, 8(1):55–65, 2006a.

McDonald, M., Hildebrand, J., and Webb, S. : Blue and fin whales observed on a seafloor array in the Northeast Pacific. *J. Acoust. Soc. Am.*, 98:712–721, 1995.

McDonald, M. A., Calambokidis, J., Teranishi, A. M., and Hildebrand, J. A. : The acoustic calls of blue whales off California with gender data. *J. Acoust. Soc. Am.*, 109 (4):1728–1735, 2001.

McDonald, M. A., Hildebrand, J. A., and Wiggins, S. M. : Increases in deep ocean ambient noise in the Northeast Pacific west of San Nicolas Island, California. *J. Acoust. Soc. Am.*, 120(2):711–718, 2006b.

Mellinger, D. K. : A comparison of methods for detecting right whale calls. *Canadian Acoustics*, 32(2):55–65, 2004.

Mellinger, D. K. and Clark, C. W. : Methods for automatic detection of mysticete sounds. *Mar. Fresh. Behav. Physiol.*, 29:163–181, 1996.

Mellinger, D. K. and Clark, C. W. : Recognizing transient low-frequency whale sounds by spectrogram correlation. *J. Acoust. Soc. Am.*, 107(6):3518–3529, 2000.

Mellinger, D. K. and Clark, C. W. : Blue whale (*Balaenoptera musculus*) sounds from the North Atlantic. *J. Acoust. Soc. Am.*, 114(2):1108–1119, 2003.

Mellinger, D. K. and Clark, C. W. : Mobysound : A reference archive for studying automatic recognition of marine mammal sounds. *Applied Acoustics*, 67(11-12):1226–1242, 2006.

Munger, L., Mellinger, D., Wiggins, S., Moore, S., and Hildebrand, J. : Performance of spectrogram correlation in detecting right whale calls in long-term recordings from the Bering Sea. *Canadian Acoustics*, 33(2):25–34, 2005.

Myers, C., Rabiner, L., and Rosenberg, A. : Performance tradeoffs in dynamic time warping algorithms for isolated word recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 28(6):623–635, 1980.

National Research Council : *Ocean noise and marine mammals.* National Academy Press, Washington, D.C., 2003.

Ney, H. : The use of a one-stage dynamic programming algorithm for connected word recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 32(2):263–271, 1984. ISSN 0096-3518.

Oleson, E. M., Calambokidis, J., Burgess, W. C., McDonald, M. A., LeDuc, C. A., and Hildebrand, J. A. : Behavioral context of call production by eastern North Pacific blue whales. *Mar. Ecol. Prog. Ser.*, 330:269–284, 2007.

Pan, K.-C., Soong, F., and Rabiner, L. : A vector-quantization-based preprocessor for speaker-independent isolated word recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 33(3):546–560, 1985. ISSN 0096-3518.

Payne, R. and Webb, D. : Orientation by means of long range acoustic signaling in baleen whales. *Ann. N. Y. Acad. Sci.*, 188:110–141, 1971.

Perry, S. L., DeMaster, D. P., and Silber, G. K. : The great whales : History and status of six species listed as endangered under the U.S. Endangered Species Act of 1973. *Mar. Fish. Rev. (special issue)*, 61(1):1–82, 1999.

Potter, J. R., Mellinger, D. K., and Clark, C. W. : Marine mammal call discrimination using artificial neural networks. *J. Acoust. Soc. Am.*, 96(3):1255–1262, 1994.

Rabiner, L. : On the use of autocorrelation analysis for pitch detection. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 25(1):24–33, 1977. ISSN 0096-3518.

Rabiner, L., Cheng, M., Rosenberg, A., and McGonegal, C. : A comparative performance study of several pitch detection algorithms. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 24(5):399–418, 1976. ISSN 0096-3518.

Rabiner, L. and Juang, B.-H. : *Fundamentals of Speech Recognition.* Prentice Hall signal processing series, 1993.

Rankin, S., Ljungblad, D., Clark, C. W., and Kato, H. : Vocalisations of blue whales, *Balaenoptera musculus intermedia*, recorded during 2001-2002 and 2002-2003 IWC-SOWER circumpolar cruises, Area V, Antarctica. *J. Cetacean Res. Manage.*, 7(1):13–20, 2005.

Reeves, R., Clapham, P., Brownell, R., and Silber, G. : Recovery plan for the blue whale (*Balaenoptera musculus*). , NOAA, National Marine Fisheries Service, Silver Spring, MD., 1998. URL `http://www.nmfs.noaa.gov/pr/pdfs/recovery/whale_blue.pdfs`. 42 pp.

Renevey, P. and Drygajlo, A. : Entropy based voice activity detection in very noisy conditions. *In EUROSPEECH-2001*, 2001.

Richardson, W. J., Greene, C. J., Malme, C., and Thomson, D. : *Marine Mammals and Noise.* Academic Press, New York, 1995.

Roman, J. and Palumbi, S. R. : Whales before whaling in the North Atlantic. *Science*, 301(5632):508–510, 2003.

Sakoe, H. and Chiba, S. : Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 26 (1):43–49, 1978. ISSN 0096-3518.

Salvado, C., Kleiber, P., and Dizon, A. : Optimal course by dolphins for detection avoidance. *Fish. Bull.*, 90:417–420, 1992.

Samaran, F. : Détectabilité des vocalisationsde rorquals communs (*Balaenoptera physalus*) à partir d'une station côtière dans la voie maritime de l'estuaire du Saint-Laurent. Mémoire de Master, Université du Québec à Rimouski, 2004.

Schevill, W., Watkins, W., and Backus, R. : The 20-cycle signal and balaenoptera (fin whales). *Marine Bio-Acoustics*, 1:147–152, 1964.

Sears, R. and Calambokidis, J. : Rapport de situation du COSEPAC sur le rorqual bleu (*Balaenoptera musculus*) au Canada - Évaluation et rapport de situation du COSEPAC sur le rorqual bleu (*Balaenoptera musculus*) au Canada - Mise à jour. , Comité sur la situation des espèces en péril au Canada. Ottawa, 2002. URL `http://dsp-psd.tpsgc.gc.ca/Collection/CW69-14-287-2003F.pdf`. Pages 1-38.

Seebaruth, D. : Algorithmes de localisation de baleines pour le Saint-Laurent. Mémoire de Master, Dept. de génie électrique et de génie informatique (GEGI), Université de Sherbrooke, 2006.

Selin, A., Turunen, J., and Tanttu, J. T. : Wavelets in recognition of bird sounds. *EURASIP Journal on Advances in Signal Processing*, 2007:1–9, 2007.

Simard, Y., Bédard. C., Mouy, X., Weise, H., , and Fortier, L. : Passive acoustic detection of bowhead, belugas and seals at Mackenzie Shelf break in Beaufort Sea during fall 2004. *In Arctic Net NSERC RCE Annual Science Meeting. Victoria, Canada, 12-15 Dec.*, 2006a. URL `http://www.arcticnet-ulaval.ca/pdf/talks2006/simard_yvan.pdf`.

Simard, Y. and Lavoie, D. : The rich krill aggregation of the Saguenay - St. Lawrence Marine Park : hydroacoustic and geostatistical biomass estimates, structure, variability, and significance for whales. *Can. J. Fish. Aquat. Sci.*, 56(7):1182–1197, 1999.

Simard, Y., Roy, N., and Gervaise, C. : Shipping noise and whales : World tallest ocean liner vs largest animal on earth. *In Proceedings of OCEANS'06 MTS/IEEE- Boston. IEEE, Piscataway, NJ, USA.*, 2006b.

Sirovic, A., Hildebrand, J. A., Wiggins, S. M., McDonald, M. A., Moore, S. E., and Thiele, D. : Seasonality of blue and fin whale calls and the influence of sea ice in the Western Antarctic Peninsula. *Deep Sea Research Part II : Topical Studies in Oceanography*, 51 (17-19):2327–2344, 2004.

Soong, F., Rosenberg, A., Rabiner, L., and Juang, B. : A vector quantization approach to speaker recognition. *In IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP '85*, vol. 10, pages 387–390, 1985.

Stafford, K. M., Fox, C. G., and Clark, D. S. : Long-range acoustic detection and localization of blue whale calls in the northeast Pacific Ocean. *J. Acoust. Soc. Am.*, 104 (6):3616–3625, 1998.

Stafford, K. M., Bohnenstiehl, D. R., Tolstoy, M., Chapp, E., Mellinger, D. K., and Moore, S. E. : Antarctic-type blue whale calls recorded at low latitudes in the Indian and eastern Pacific Oceans. *Deep Sea Research Part I : Oceanographic Research Papers*, 51 (10):1337–1346, 2004.

Sturtivant, C. and Datta, S. : The isolation from background noise and characterisation of bottlenose dolphin (*Tursiops truncatus*) whistles. *J. Acoust. Soc. India*, 23(4):199–205, 1995a.

Sturtivant, C. and Datta, S. : Techniques to isolate dolphin whistles and other tonal sounds from background noise. *Acoust. Lett.*, 18(10):189–193, 1995b.

Thode, A. M., D'Spain, G. L., and Kuperman, W. A. : Matched-field processing, geoacoustic inversion, and source signature recovery of blue whale vocalizations. *J. Acoust. Soc. Am.*, 107(3):1286–1300, 2000.

Thompson, P. O., Findley, L. T., and Vidal, O. : 20-Hz pulses and other vocalizations of fin whales, *Balaenoptera physalus*, in the Gulf of California, Mexico. *J. Acoust. Soc. Am.*, 92(6):3051–3057, 1992.

Thompson, P. O. and Friedl, W. A. : A long term study of low frequency sound from several species of whales off Oahu, Hawaii. *Cetology*, 45:1–19, 1982.

Thompson, P., Winn, H., and Perkins, P. : *Behavior of marine animals, vol. 3 : Cetaceans*, chap. Mysticete sounds. Plenum Press, NY, 1979.

Urick, R. J. : *Principles of Underwater Sound 3rd Edition*. McGraw-Hill, 1983.

Van-Trees, H. : *Detection, Estimation, and Modulation Theory. Part I.* John Wiley & Sons, 1968.

Vintsyuk, T. K. : Speech discrimination by dynamic programming. *Cybernetics and Systems Analysis*, 4(1):52–57, 1968.

Walker, R. A. : Some intense, low-frequency, underwater sounds of wide geographic distribution, apparently of biological origin. *J. Acoust. Soc. Am.*, 35(11):1880–1880, 1963.

Watkins, W., Tyack, P., Moore, K., and Bird, J. : The 20-Hz signals of finback whales (*Balaenoptera physalus*). *J. Acoust. Soc. Am.*, 82:1901–1912, 1987.

Watkins, W. A., Daher, M. A., George, J. E., and Rodriguez, D. : Twelve years of tracking 52-Hz whale calls from a unique source in the North Pacific. *Deep Sea Research Part I : Oceanographic Research Papers*, 51(12):1889–1901, 2004.

Weston, D. E. and Black, R. I. : Some unusual low-frequency biological noises underwater. *Deep Sea Research*, 12(3):295–296, 1965.

Wiggins, S. : Autonomous acoustic recording packages (ARPs) for long-term monitoring of whale sounds. *MTS Journal*, 37(2):13–22, 2003.

Wiggins, S. M., Oleson, E. M., McDonald, M. A., and Hildebrand, J. A. : Blue whale (*Balaenoptera musculus*) diel call patterns offshore of southern California. *Aquat. Mamm.*, 31(8):161–168, 2005.