



HAL
open science

Nouveaux modèles d'estimation monophone de distance et d'analyse parcimonieuse : Applications sur signaux transitoires et stationnaires bioacoustiques à l'échelle

Yann Doh

► To cite this version:

Yann Doh. Nouveaux modèles d'estimation monophone de distance et d'analyse parcimonieuse : Applications sur signaux transitoires et stationnaires bioacoustiques à l'échelle. Bio-informatique [q-bio.QM]. Université de Toulon, 2014. Français. NNT : 2014TOUL0020 . tel-01358281

HAL Id: tel-01358281

<https://theses.hal.science/tel-01358281v1>

Submitted on 31 Aug 2016

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

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

Nouveaux modèles d'estimation monophone de distance et d'analyse parcimonieuse - Applications sur signaux transitoires et stationnaires bioacoustiques à l'échelle

THÈSE

présentée et soutenue publiquement le 17 décembre 2014

pour l'obtention du

Doctorat de l'Université de Toulon

par

Yann Doh

Composition du jury

Rapporteurs : J. L. Zarader ISIR
J. Pinquier IRIT

Examineurs : O. Gérard DGA
G. Pavan CIBRA
H. Glotin (directeur) LSIS
O. Adam (co-directeur) CNPS
G. Nolibe (co-directeur) CeSigma
J. Razik (co-encadrant) LSIS

Invités : P. Cristini LMA

Laboratoire des Sciences de l'Information et des Systèmes — UMR 7296

Remerciements

Ces 3 années de thèse sont avant tout 3 années de vie intense faite d'imprévus, de défis et de rencontres riches. Mes premiers remerciements s'adressent à ceux qui ont rendu possible cette aventure : mes co-directeurs : Hervé Glotin, Olivier Adam, Gilles Nolibe et mon co-encadrant Joseph Razik. Hervé a cru en moi dès la première rencontre à l'occasion de mon stage de fin d'étude qui s'est soldé par un financement de thèse. Son imagination débordante et sans fin m'ont bien souvent "boosté" et sa disponibilité a été sans faille : le dimanche en famille, sur les pistes de ski, entre 2 talks au Japon ... Olivier m'a impressionné par son dynamisme, son efficacité et son tempérament à rendre les longues séances de travail agréables. Il m'a tout de suite fait partager son univers et m'a impliqué dans ses projets. Gilles a fait le lien entre mon travail de recherche et le monde de l'entreprise. A ses côtés sur certaines missions, j'ai pris de véritables leçons sur ce qu'était le savoir-faire opérationnel. Et parfois même des leçons de russe. Joseph mon co-encadrant en plus d'être mon conseiller à toutes épreuves a essuyé mes joies, mes plaisanteries, mes moments de grâce mais aussi mes peines, mes coups de gueule et mes angoisses.

Je tiens à remercier les infatigables Frères Mayer et le plus téméraire des scaphandriers : Henry Bellon, pour leur soutien et appui, logistique à Madagascar. Vous avez vraiment mis les grands moyens à notre disposition pour la réalisation de BAOBAB : avion, bateau, voiture, moto, plongeurs, matériaux, personnel, nourriture, kayak. Une équipe de choc, impressionnant ! Je remercie de fait l'association CETAMADA, son staff permanent et tous les éco-volontaires qui ont participé au projet ou qui se sont donnés du mal pour nous aider. "Misaotra be" à tous les habitants de Sainte Marie, ce coin de paradis, qui ont croisé ma route. Votre accueil, votre hospitalité et votre gentillesse m'ont fait chaud au cœur.

Merci au membres du PNPC pour leur coopération et leur soutien logistique pour le développement de nos méthodes de suivi acoustique en zone PELAGOS.

Je n'oublie pas Nadège Gandilhon et BREACH : merci de m'avoir accueilli et hébergé en Guadeloupe.

Je remercie mes collègues de bureau qui m'ont soutenu, aidé et inspiré jusqu'à l'organisation de ma soutenance. Je remercie aussi le personnel administratif de l'Université de Toulon ainsi que le personnel technique toujours agréables et disponibles.

Emily, tu as été ma compagne et je garderai toujours une place pour toi dans mon cœur. Enfin j'ai une pensée pour mes proches et ceux qui m'ont quitté et avec qui j'aurais aimé partager cette étape de ma vie.

Sommaire

Partie I Introduction générale

xv**Introduction****1****Chapitre 1****Objectifs de la thèse**

1.1	L'intérêt de l'acoustique passive pour l'observation des cétacés	5
1.2	Le contexte de l'acoustique passive	6
1.3	Les types de signaux émis par les cétacés : dualité signaux transitoires <i>vs</i> voisements	6
1.4	Les principales problématiques	7
1.4.1	La détection	7
1.4.2	La classification	9
1.4.3	La localisation de la source	9
1.4.4	L'estimation de densité de population	10
1.5	Les limites actuelles	11
1.6	Développement de méthodes mono-hydrophone et perspectives en passage à l'échelle	12
1.7	Conclusion	13

Chapitre 2

Etat de l'art des méthodes d'analyse monohydrophone en bioacoustique

2.1	Moyens utilisés pour la réalisation de mesures acoustiques passives	15
2.2	Rappels d'acoustique sous-marine générale	16
2.2.1	Equations fondamentales	17
2.2.2	Atténuation	19
2.3	A propos de la détection	23
2.4	Estimation de distance source-récepteur	24
2.4.1	Concept des hydrophones virtuels	24
2.4.2	Equation du sonar passif	25
2.5	Extraction des caractéristiques sonores des cétacés	28
2.5.1	Sur les transitoires : exemple sur les cachalots	28
2.5.2	Sur les voisements : exemple des unités sonores de baleines à bosse	31
2.6	Conclusion	33

Partie II Contributions 35

Chapitre 3

Proposition d'un modèle d'estimation mono-récepteur de distance de source transitoire

3.1	Base théorique	37
3.2	Représentation de l'estimateur distance ISAT et paramètres importants	40
3.3	Architecture et mise œuvre de l'estimation ISAT	42
3.3.1	De l'échelle d'un clic à l'échelle d'un pulse	42
3.3.2	Dans le domaine fréquentiel	43
3.3.3	Estimations de distance ISAT pour un choix de bandes fixe	44
3.3.4	Adaptation à l'ensemble du domaine spectral par multi-combinaisons de bandes	44
3.3.5	Filtrage des estimations sur des fenêtres d'observations de courtes durées	45
3.3.6	Détermination des combinaisons de bandes optimales à partir de la vérité terrain	46
3.3.7	Sélection des combinaisons de bandes en conditions réelles	46
3.4	Conclusion	47

Chapitre 4**Prédiction de distance par réseau de neurones**

4.1	Rappels sur les perceptrons multicouches	49
4.1.1	Le perceptron	49
4.1.2	Règle d'apprentissage du perceptron	50
4.1.3	Le perceptron multi-couches	50
4.2	Définition des modèles	51
4.3	Conclusion	54

Chapitre 5**Modèle de codage parcimonieux de signaux stationnaires sous-marins**

5.1	Objectifs	55
5.2	Représentation cepstrale	56
5.3	Codage parcimonieux	56
5.4	Corrélation MFCC/codes parcimonieux	58
5.5	Estimation de la complexité du dictionnaire	58
5.6	Mesure de divergence des codes d'une année sur l'autre	60
5.7	Conclusion	61

Chapitre 6**Installations, mesures et matériel disponibles pour valider les modèles**

6.1	Chaîne d'acquisition	63
6.2	Systèmes et données monocapteur	64
6.2.1	Dispositif autonome monocapteur Gualiba	64
6.2.2	Enregistrements et base de données DECAV au PNPC/PELAGOS	66
6.2.3	Installation et enregistrements BOMBYX au sud du PNPC (sanctuaire PELAGOS)	69
6.2.4	Enregistrements de chants de baleine à bosse à Madagascar	70
6.2.5	Enregistrements et base de données "St Pierre et Miquelon" SPM	71
6.3	Systèmes multicapteurs	74
6.3.1	Dispositif autonome multicapteurs BAOBAB	74
6.3.2	Base de données "Bahamas"	78
6.4	Conclusion	83

Chapitre 7**Validation des modèles de prédiction de distance sur signaux transitoires de cachalot**

7.1	Estimation de la distance radiale par le modèle ISAT	88
7.1.1	Optimisation des hyper-paramètres P et N_{best}	88
7.1.2	Estimation de la distance de propagation $ISAT$ sur le jeu de données de test	88
7.1.3	Information mutuelle entre les rapports de bandes d'énergie et la distance source-récepteur	92
7.1.4	Sélection des combinaisons de bandes de fréquence basée sur l'Information Mutuelle	93
7.2	Estimation de distance par régressions LER, ISAR et ISARH	95
7.2.1	Choix des ensembles d'apprentissage, de test et de développement . .	95
7.2.2	Optimisation des Hyper-paramètres	95
7.2.3	Prédiction de distance pour un mélange train/test uniforme	98
7.2.4	Prédiction de distance pour un mélange train/test dissocié dans le temps	100
7.2.5	Erreur et précision des prédictions ISAT, LER, ISAR et ISARH . . .	102
7.2.6	Modèle de fusion entre $ISAT$ et $ISAR$	104
7.3	Test du modèle de fusion ISAT/ISAR sur le jeu de données Bahamas set B .	106
7.3.1	Estimations de distance dans le temps	106
7.3.2	Traitement des prédictions par bi-histogrammes temps/distance . . .	108
7.3.3	Répartition des prédictions	110
7.4	Résultats préliminaires sur les données issues du protocole DECAV	113
7.4.1	Estimations de distance dans le temps	113
7.4.2	Traitement des prédictions par bi-histogrammes temps/distance . . .	114
7.4.3	Répartition des prédictions	114
7.4.4	Synthèse des résultats sur les données DECAV en présence d'un seul cachalot	115
7.4.5	Synthèse des résultats sur les données DECAV en présence de plusieurs cachalots	116
7.5	Résultats préliminaires sur un extrait des données BOMBYX	118
7.6	Conclusion	121

Chapitre 8 Application du codage parcimonieux sur des signaux stationnaires de baleine à bosse

8.1	Conditionnement des descripteurs	125
8.2	Apprentissage du dictionnaire	125
8.3	Extraction des codes de chant	127
8.4	L'évolution des codes de chant	128
8.5	Discussion	129
8.6	Conclusion	132

Partie IV	Discussion et perspectives	133
------------------	-----------------------------------	------------

Chapitre 9 Synthèse, perspectives et conclusion générale

9.1	Discussion sur l'estimation de distance sur les signaux transitoires	135
9.2	Discussion sur le codage parcimonieux de voisements	137
9.3	Conclusion générale	138

Glossaire	141
------------------	------------

Bibliographie	143
----------------------	------------

Publications personnelles	155
----------------------------------	------------

Table des figures

1	Les 4 méthodes d'observation des cétacés	2
2.1	Représentation d'un volume élémentaire de fluide n est le vecteur normal à un élément de surface ds	17
2.2	Passage d'une onde à l'interface liquide/liquide. Représentation dans le plan sagittal [Josso 10].	20
2.3	Profils de célérité typique [Josso 10].	21
2.4	Profil de célérité et tracé de rayons typiques de la mer Méditerranée. La source est située à une profondeur de 100 mètres et est illustrée par un cercle. On peut remarquer la présence de rayons réfractés (traits pleins), de rayons réfractés et réfléchis en surface (traits hachurés) et de rayons réfractés, réfléchis fond et surface (traits pointillés hachurés) [Josso 10].	22
2.5	Modèle de bruit ambiant pour une mer calme, un état de trafic maritime limité et une profondeur d'observation de 100 m [Zimmer 11].	26
2.6	Coefficient α (en $dB.km^{-1}$) en fonction de la fréquence (kHz) selon Thorp [Thorp 67].	26
2.7	Coupe sagittale de la tête du cachalot et trajets issus de la vibration initiale P0 responsables de la structure multipulsée [Laplanche 05].	29
2.8	Différents types de pulses suivant la position de l'hydrophone. On-axis (en haut) et off axis (en bas). Illustration tirée de la thèse de C. Laplanche, 2005 [Laplanche 05].	30
2.9	Représentation du diagramme de directivité des émissions sonores du cachalot dans un plan longitudinal [Zimmer 05].	30
2.10	Diagramme de directivité du rayonnement d'un piston circulaire plan de diamètre 8.5 cm. A gauche pour la fréquence 10 kHz, à droite pour la fréquence 40 kHz. . .	31
2.11	Spectrogramme d'un chant de baleine à bosse. Début : 5 min 40 s, fin : 6 min et en ordonnées de 0 à 22.05 kHz, codé sur 512 bins (fft calculée sur 1024 points sur une fenêtre de 10 ms). Enregistré par Darewin à la Réunion.	33
3.1	Illustration de la déformation spectrale selon la distance de propagation sur un signal idéal de type Dirac, de 1 m à 10 km de haut en bas. La résolution en fréquence est de $\delta f = 24$ Hz.	40
3.2	Représentation de la relation ISAT pour différentes paires de fréquences centrales (F_1^c/F_2^c) (en Hz) des bandes B_1 et B_2 du ratio $\frac{E_1}{E_2}$. La taille des bandes est fixée à $\Delta B = 200$ Hz. A gauche : IIB variable, à droite F_1^c/F_2^c variable.	41
3.3	Clic moyen dans le jeu de données BAHAMAS. Deux échelles différentes de calcul, centrées sur le pulse le plus énergétique.	42

3.4	Extraction du spectre moyen du pulse le plus énergétique et de son écart type (filtré par un passe haut $f_c = 2$ kHz) à partir du jeu de données BAHAMAS pour l'hydrophone H11 et les effets résultant de l'étape de lissage. A gauche : le spectre original du pulse et la région d'intérêt (entre 7 kHz et 22 kHz). A droite : la région d'intérêt du spectre après lissage.	43
4.1	Schématisation du modèle <i>ISAR</i>	52
4.2	Schéma de synthèse des différents modèles considérés. LER, ISAR, ISARH (IFARH et IFAR sont respectivement ISARH et ISAR).	54
5.1	Représentation de la matrice d'autocorrélation entre les vecteurs MFCC du côté gauche, respectivement, pour les vecteurs de code parcimonieux du côté droit. Les résultats sont obtenus sur l'enregistrement de 2009 avec un dictionnaire appris sur l'année 2009. La longueur de fenêtre est de 4 s et le dictionnaire contient 32 vecteurs.	59
5.2	Rapports de corrélations entre espace vectoriel MFCC et espace vectoriel des codes parcimonieux. Les points au-dessus de la diagonale sont plus corrélés dans l'espace de code parcimonieux, les points en dessous sont de moins en moins corrélés dans le code parcimonieux. Enregistrements de 2009, de longueur de fenêtre de 4 s, dictionnaire de 32 vecteurs.	60
6.1	Vu du dispositif gualiba en surface.	65
6.2	Résultats bruts de toutes les sorties du projet DECAV : les cinq cachalots vus ont été mesurables uniquement acoustiquement, comme 16 autres uniquement contactés acoustiquement sur un total de 21 contacts acoustiques de cachalots ("PNPC" : catamaran moteur ou zodiac du PNPC, "RdV" : association Regard du Vivant [Abeille 14]).	67
6.3	Détections de cachalots au large de Toulon/Hyères. Les étiquettes jaunes sont les points de détection. Les lettres G et P désignent la taille du spécimen, respectivement individu "grand" et "petit", "?" indique que la taille n'est pas mesurable du fait d'un Rapport Signal sur Bruit trop faible, ou d'un échantillon trop court [Abeille 14].	68
6.4	Structure BOMBYX sur le pont du bateau opérationnel de IXSURVEY. Crédits photo : Gilles Rougier.	69
6.5	Image de BOMBYX après déploiement (coordonnées GPS : 42.56.205 et 6.19.066). Crédits photo : Gilles Rougier.	70
6.6	Spectrogrammes de quelques segments de chant de Pace et al. 2010, réalisés à partir du même chant que celui utilisé dans la thèse. On aperçoit une même sous-unité (entourée) dans différents contextes de chant.	71
6.7	Schéma du dispositif d'acquisition AURAL utilisé par l'association Frag-îles.	72
6.8	Bilan des enregistrements SPM avec les 2 hydrophones en 2010 et en 2011.	72
6.9	Clics de cachalot. Extraits d'AuralA, A32C0299, 1/9/2010, 04 :37min à 04 :39min.	73
6.10	Situation géographique des 3 sous réseaux d'hydrophones le long de la côte Ouest de Ste Marie.	75
6.11	Module C1.	76
6.12	Module C2.	77
6.13	Module C3.	77
6.14	Trajectoire 2D (dans le plan $x-y$) d'un cachalot isolé et position des hydrophones. Il est observé pendant 25 min (LSIS/DYNI [Giraudet 06b]). L'animal se dirige vers le Sud-Est.	79

6.15	Dynamique de la distance radiale vérité terrain suivant H8, H9, H10 et H11. . . .	79
6.16	Histogramme des distances radiales vérité terrain sur les 25 min.	80
6.17	Histogramme des angles vérité terrain entre l'animal et les hydrophones.	80
6.18	Trajectoire 2D (dans le plan $x - y$) de 4 cachalots et position des hydrophones. Les animaux évoluent séparément durant 20 min vers le Nord-Est (LSIS/DYNI [Glotin 08a]).	81
6.19	Dynamique de la distance radiale de la vérité terrain suivant H5 pour les 4 cachalots.	82
6.20	Dynamique de la distance radiale de la vérité terrain suivant H6 pour les 4 cachalots.	82
7.1	Moyenne de l'erreur relative absolue entre la vérité terrain et l'estimation $N_{best} = 1, 2, \dots, 16$, <i>all</i> (sur $N = 128$ échantillons) en fonction du paramètre de lissage des spectres en fréquence F_p . Moyenne sur les hydrophones H8, H9, H10, H11 de la partition dev.	89
7.2	Exemple d'estimations temporelles finales de la distance animal-récepteur avec le modèle ISAT, en comparaison avec la vérité terrain sur l'hydrophone H11. En bas l'évolution de l'azimut et de l'élévation (issus de la cinématique de la vérité terrain). Les paramètres fixés : $N_{best}^* = 4$, $F_{P_{128}}^* = 8.5$ kHz et $F_{P_{256}}^* = 9$ kHz. . . .	89
7.3	Prédictions temporelles de distance selon le modèle ISAT optimal retenu et vérité terrain. Pour être comparé aux prédictions par réseau de neurones, les données test proviennent de l'ensemble train/test dissocié dans le temps.	91
7.4	Fréquences centrales des bandes mises en jeu pour les 32 plus fortes valeurs d'IM selon l'azimut.	92
7.5	Prédictions temporelles de distance selon le modèle ISAT2 (a et b) basé sur l'IM et la vérité terrain. Les données de test proviennent de l'ensemble train/test dissocié dans le temps.	94
7.6	Courbes d'erreur obtenues lors de la phase d'optimisation des paramètres : nombre d'unités de la couche cachée N_{unit} et nombre d'itérations de la boucle d'apprentissage N_{iter} , pour modèles LER, ISAR.	96
7.7	Prédictions temporelles comparées aux vérités terrain de la distance radiale et des angles de position selon les modèles LER, ISAR, ISARH. Ensemble train/test uniforme et spectres non normalisés. Predictions données sur l'ensemble de test et modèle optimisé sur l'ensemble développement.	99
7.8	Valeurs des poids W1 et W2 appris sur l'ensemble train du mélange train/test dissocié dans le temps.	100
7.9	Prédictions temporelles de distance selon les modèles ISAR et LER optimaux retenus et de la vérité terrain. Les données de test proviennent de l'ensemble train/test dissocié dans le temps et les spectres sont normalisés selon l'équation 7.5. Les prédictions sont données sur l'ensemble de test et le modèle est optimisé sur l'ensemble de développement.	101
7.10	Bi-distribution de l'erreur de prédictions vs l'azimut (au milieu), l'élévation (en bas) sur les données du jeu de données "Bahamas" dédiées au test. En haut, Information mutuelle entre l'erreur de prédiction et les angles de position.	103
7.11	Prédictions temporelles ISAR, ISAT et modèle de fusion comparées aux vérités de terrain distance radiale.	105
7.12	Prédictions temporelles ISAR de distance sur l'hydrophone 6 représentées par des cercles noirs. Les distances "vérité terrain" sont matérialisées par des points rouges.	107
7.13	Prédictions temporelles ISAR de distance sur l'hydrophone 5 représentées par des cercles noirs. Les distances "vérité terrain" sont matérialisées par des points rouges.	108

7.14	Bi-histogrammes temps/distance des prédictions temporelles par fusion ISAT2/ISAR sur le jeu de données BAHAMAS SET B (hydrophone 6).	109
7.15	Bi-histogrammes temps/distance des prédictions temporelles par fusion ISAT2/ISAR sur le jeu de données Bahamas set B (hydrophone 5).	110
7.16	Histogrammes normalisés des prédictions ISAR sur le jeu de données Bahamas set B (hydrophone 6).	111
7.17	Histogrammes normalisés des prédictions ISAR sur le jeu de donnée Bahamas set B (hydrophone 5).	112
7.18	Prédictions temporelles ISAR de distance sur l'enregistrement du 02/08/2011 représentées par des cercles.	113
7.19	Bi-histogrammes temps/distance des prédictions temporelles ISAR sur l'enregistrement du 02/08/2011.	114
7.20	Histogrammes normalisé des prédictions ISAR sur l'enregistrement du 02/08/2011.	115
7.21	Représentation temporelle des prédictions obtenues sur 3 enregistrements consécutifs contenant 1 animal.	116
7.22	Représentation temporelle, bi-histogrammes et histogrammes des résultats obtenus sur les fichiers traités parmi les données DECAV.	117
7.23	Synthèse des prédictions obtenues sur 200 s d'un extrait des données Bombyx.	118
7.24	Spectrogramme long terme de BOMBYX SABIOD représentant 20 détections par tranche de 5 minutes (analyse Dyni SABIOD). La majorité des détections de cachalot apparaît entre 14h et 21h du second jour bombyx en juillet 2014.	119
7.25	Synthèse des prédictions de distance médianes par pas de 20 min sur 41 h des données BOMBYX par les modèles ISAR et fusion ISAT2/ISAR.	120
7.26	Prédictions de distance médianes par pas de 20 min sur 41 h des données BOMBYX par le modèle LER.	121
7.27	Vitesse radiale estimée <i>vs</i> vitesse radiale de référence sur BAHAMAS SET A et B. La régression est effectuée uniquement sur le modèle de fusion.	123
8.1	Dictionnaire de 32 codes, triés par degré de complexité, calculés en utilisant une échelle de temps de 250 ms et appris sur l'union de sous-ensembles de chants de 2008 et de 2009. L'échelle de couleur de la carte est de -1 pour le bleu et 1 pour le rouge.	126
8.2	Valeurs de la complexité des 32 codes (triés) du dictionnaire illustré en figure 8.1 (sur une échelle de temps de 250 ms, appris sur des sous-ensembles de chants de 2008 et de 2009. En abscisse, l'indice des codes, en ordonnée la complexité en bits. La différence entre les codes les plus complexes et les moins complexes est importante.	127
8.3	Histogramme de la distance KL (2008, 2009) calculée sur les 32 vecteurs parcimonieux du dictionnaire. Il démontre que les représentations de courte durée sont plus stables au fil des ans que les plus longues.	128
8.4	Distance entre les chants de 2008 et 2009, en moyenne sur les deux codes les plus complexes, contre les deux les plus simples. En abscisse l'échelle de temps du code, de 1/4 à 4 secondes. En ordonnées, la distance.	129
8.5	Evolution dans le temps pendant 100 ms du coefficient de quatre vecteurs parcimonieux pour le dictionnaire de 16 unités, pour une partie de l'enregistrement de chants de l'année 2008.	130

8.6	Valeur absolue du logarithme du rapport des probabilités des bigrammes des chants de 2008 par rapport aux chants de 2009, 2012, 2013 et 2014. Ceci est illustré pour $K = 16$, échelle de temps de 500 ms. La représentation est convoluée par $[1 \ 1; 1 \ 1]$ pour lisser l'ordre des mots.	131
8.7	Histogramme des bigrammes 2008 <i>vs</i> 2009, 2012, 2013 et 2014. Bin 0.5 = divergence nulle, les bins sont espacés de 1.125 unité de divergence.	131

Première partie
Introduction générale

Introduction

Observer les cétacés dans leur environnement est une tâche complexe pour plusieurs raisons, notamment parce que ces quelques 80 espèces ont des comportements différents, des structures sociales différentes, des habitats différents, des modes alimentaires différents. . . Les motivations principales des études sur les cétacés sont de pouvoir les recenser, les localiser, de définir leur statut de résidence, d'étudier leurs comportements et plus récemment, les scientifiques s'intéressent aux impacts potentiels de certaines activités humaines marines. Le niveau de connaissance varie selon les espèces et l'effort des scientifiques est hétérogène. Globalement, les espèces qui évoluent en surface proche des côtes et les individus résidents, ont été particulièrement observés : nous pouvons citer le cas, par exemple, de certains groupes d'orques, de grands dauphins, de cachalots. Par contre, lorsque les espèces sont furtives, évoluent au large, ou ne sont que de passage, les données sont plus éparpillées et voire, dans certains cas, insuffisantes pour renseigner la population. Nous sommes confronté à ce problème en particulier pour la caractérisation des baleines bleues d'Antarctique.

Il en est de même avec l'évaluation des effets des activités humaines sur ces espèces. Depuis la fin des années 90, les nombreuses observations témoignées dans des articles scientifiques et des conférences internationales n'ont pas permis d'établir des principes généraux, comme par exemple de limiter des sources sonores sous-marines à des seuils acoustiques maximum. La question de l'impact des sons d'origine anthropique devient alors un sujet de recherche très actif. L'ensemble des paramètres est mesurable, sauf la sensibilité auditive qui n'est connue que pour une dizaine d'espèces d'odontocètes (potentiels évoqués auditifs mesurés sur des individus que l'on a mis en bassin pour réaliser les expérimentations). Ainsi, par exemple, pour le *Tursiops truncatus*, Finneran [Finneran 05] a déterminé le seuil de perte auditive temporaire pour une émission sonore de quelque ms provenant d'une source sismique à 224 dB re.1 μPa . En 1997, Ridgway et al. [Ridgway 97] a évalué ce seuil pour un signal artificiel à bande étroite de 1 seconde. Le seuil est atteint pour un niveau de 194 dB re. 1 μPa à 3 kHz, 193 dB re. 1 μPa à 20 kHz, 192 dB re .1 μPa à 75 kHz. Les sons impactent également d'autres parties de l'anatomie que le système auditif, voire engendrent des changements de comportement qui peuvent être à l'origine de la mort (comme peut-être des remontées trop rapides de baleines à bec), des changements de communication, des changements d'habitat, des échouages. Il aura fallu attendre 2010 pour voir se créer un consortium de scientifiques travaillant sur le comportement des cétacés, notamment en présence de sources sonores artificielles. Là encore, la diversité des espèces rend les travaux difficiles et, à ce jour, aucune conclusion ne permet de quantifier ou qualifier l'impact d'une onde acoustique sur un individu en tenant compte de son âge, de son comportement, s'il est seul ou en groupe.

Ces situations font que les politiques de conservation dédiées aux cétacés sont parfois délicates à établir si elles veulent s'appuyer sur la plus grande connaissance que l'on a sur la zone visée, notamment parce que ces connaissances sont dynamiques dans le temps et qu'il est nécessaire de les reconsidérer constamment en fonction des nouvelles observations. C'est ainsi, par exemple, que le

statut des baleines à bosse a évolué sur la liste rouge de l'IUCN devenant "low critical" en 2009, lors de la conférence de la Society of Marine Mammalogy. [Barlow 09] a confirmé la tendance positive de l'état de la population du Pacifique Nord, dépassant les 20000 individus (projet SPLASH, [Calambokidis 10]). Deux études comparables ont été menées pour la population de l'Atlantique Nord (projet YONAH puis projet MOHAH) montrant que la population est également sur une tendance positive, dépassant actuellement les 11000 individus [Stevick 03]. Récemment, Simmonds propose d'utiliser les observations des baleines à bosse comme indicateurs dans des modèles de prédiction du réchauffement climatique de la planète [Simmonds 09, Simmonds 07] (notamment la durée de leur séjour dans les zones d'alimentation).

Les actions de protection de la biodiversité comme la volonté pure de comprendre le vivant repose aussi sur des connaissances fines, à l'échelle de l'espèce, sur les habitudes de vie, le comportement social, les régimes alimentaires. Il existe plusieurs méthodes d'observation des mammifères marins (figure 1) :

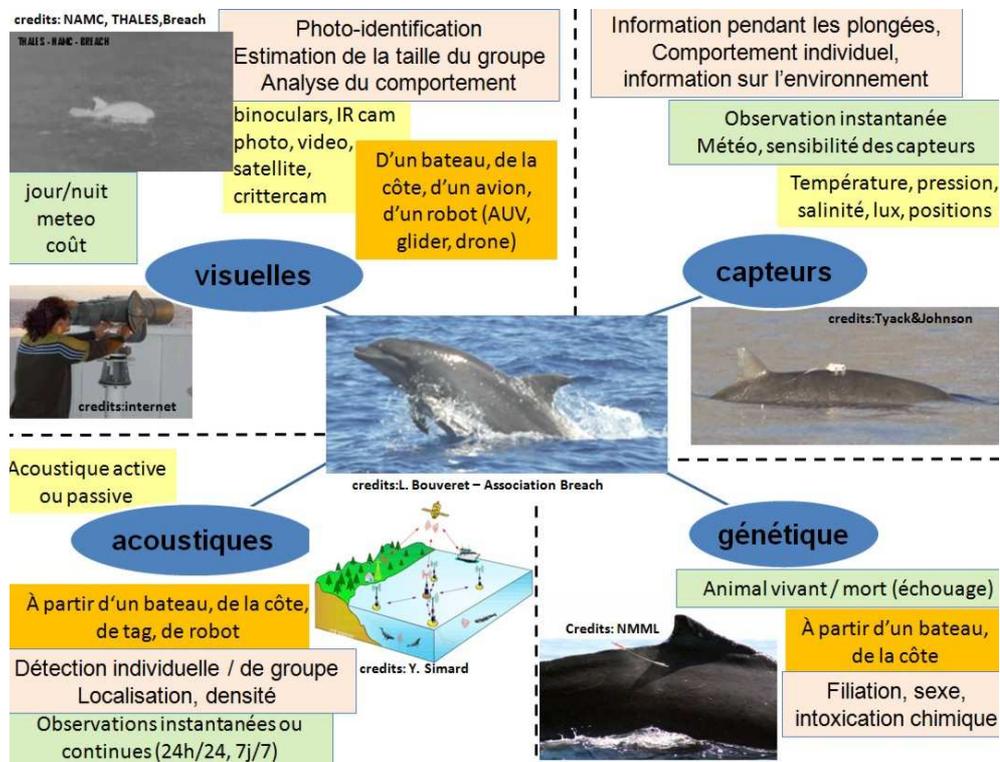


FIGURE 1 – Les 4 méthodes d'observation des cétacés

– Observations visuelles :

les observations visuelles servent à savoir si des cétacés sont présents sur une zone. Elles permettent d'identifier les individus, de donner une estimation de leur nombre et de caractériser leurs comportements. Elles sont utiles aussi pour décrire l'environnement : présence d'autres espèces d'animaux, d'activités humaines, conditions météo... Elles peuvent se faire depuis la côte mais généralement, l'observateur se rend sur zone avec un bateau. Il n'y a pas de matériel spécifique à avoir, mais généralement, les observateurs sont équipés de jumelles et d'appareils photos. De nouvelles approches font état de matériels infra-rouge (IR) [Mobley 08, Welcome 09, Baldacci 05, Mccafferty 07, Cuyler 92] qui ont certains d'inconvénients : il faut que les cétacés

soient visibles en surface, les conditions météo doivent être clémentes, la visibilité doit être excellente. Et généralement, il y a un biais dû à la distance parcourue, les zones observées sont finalement souvent celles proches du port où est amarré le bateau. Les cétacés sont plus ou moins actifs en surface et peuvent être vus lors de leur expiration. Les collectes de données basées sur l'observation physique d'un opérateur humain sont probablement les plus anciennes et robustes. Les données accumulées sont principalement des photos et du texte. Les principales applications sont la photo-identification pour le "matching" des individus et l'estimation de densité de population. Le comptage des individus doit être effectué selon une procédure dite de transect lors de laquelle la trajectoire du mobile ne doit pas être influencée par les observations. Cela afin d'assurer la validité de l'échantillonnage de la surface investiguée. La grande limitation du visuel réside dans le fait que beaucoup d'espèces sont plus rarement à la surface et sont donc indétectables visuellement. De plus le poste d'observation est mouvant à l'instar des cétacés.

– Utilisation de capteurs :

récemment, nous avons vu se développer des modules électroniques étanches embarquant plusieurs capteurs qu'il est possible de poser sur le cétacé. Ces appareils enregistrent la position Argos et/ou GSM, la température, la pression, les accélérations dans les 3 axes et la rotation (cellule inertielle), les émissions sonores, parfois même la salinité et la luminosité. Ces modules sont très intéressants pour étudier les espèces qui passent la majorité de leur temps à grande profondeur, et qui sont donc difficilement observables. On peut ainsi recréer, *a posteriori*, leurs plongées et leurs activités durant leurs plongées. Ces modules électroniques sont encore en développement à l'heure actuelle, avec l'ajout de plus en plus de capteurs, de meilleures performances en termes d'autonomie et de stockage de l'information. Quelques constructeurs leaders proposent des solutions prêtes à l'emploi ("Argos tag", "DTag", "Acousonde"). L'inconvénient est qu'il s'agit d'une technique invasive (faiblement lorsque le module est ventosé et fortement lorsqu'il est accroché par des hameçons), générant probablement un stress chez l'animal. Il est également difficile à déployer sur les individus, certaines espèces étant peu coopératives et les conditions de mer doivent être particulièrement calmes. Nous pouvons citer d'autres difficultés comme l'identification des individus sur lesquels on veut déployer la sonde et la recherche ensuite du matériel lorsque les données ne sont pas transmises à distance (par exemple via une connection Argos). Ces balises peuvent rester de quelques heures à plusieurs jours voire plusieurs semaines fixées à l'animal. La durée dépend du système d'accroche et de la qualité de la méthode de déploiement. Enfin, il s'agit de données décrivant un individu, il est alors risqué de généraliser les résultats à l'espèce ou à un groupe d'individus. Ceci étant dit, ces techniques, déployées en même temps sur plusieurs individus ou de façon répétitive d'une année sur l'autre, offrent la possibilité d'avoir finalement un nombre critique de données représentatives de plusieurs individus, voire de suivre des individus sur de très grandes échelles, par exemple, suivre les baleines à bosse sur une partie de leurs routes migratoires, représentant plusieurs milliers de kilomètres. Les collectes d'informations sur le positionnement relatif de l'animal permettent d'analyser le comportement moteur des animaux et d'en déduire des relations entre le mouvement de l'animal et le coût énergétique.

– Analyse génétique :

la génétique permet de connaître le sexe de l'individu (le dimorphisme sexuel n'étant pas forcément très prononcé chez certaines espèces de cétacés), l'affiliation, le degré d'ouverture génétique d'un groupe d'individus. A partir des biopsies, il est possible d'effectuer des analyses toxicologiques et de mesurer la présence éventuelle de métaux lourds ou de composants chimiques

polluants. Cet aspect est particulièrement intéressant, notamment pour des groupes de résidents qui sont proches de côtes où les activités humaines industrielles sont fortes, par exemple, lors de la surveillance des 300 grands dauphins qui se déplacent entre le Mont St Michel, Cherbourg et les îles anglaises. Les biopsies se font généralement à l'aide d'une flèche propulsée par une arbalète ou un fusil à air comprimé, voire une perche. Des méthodes non invasives sont aussi à l'étude, comme le prélèvement de fécès, de souffles, en prélevant un morceau de peau flottant ou en grattant l'animal. L'inconvénient est qu'il faut trouver l'individu et s'approcher suffisamment prêt pour pouvoir pratiquer le prélèvement.

– Observations acoustiques :

Il est possible d'utiliser des sonars pour repérer les cétacés, en particulier lorsqu'ils sont en groupe et en profondeur. Mais ceci n'est pas la méthode retenue car les cétacés sont actifs vocalement. Compte tenu des propriétés de propagation des ondes sonores dans le milieu sous-marin, le recours à l'acoustique passive est une technique qui a connu un développement spectaculaire ces 10 dernières années, notamment du fait que la communauté internationale s'est structurée rapidement au début des années 2000. Les différents objectifs sont de détecter la présence des cétacés vocaux, de classer leurs émissions sonores et d'estimer, lorsque cela est possible, leurs localisations. Plus récemment, en 2009, une nouvelle thématique a fait son apparition au workshop DCL à Pavie (Italie) : l'estimation du nombre d'individus à partir de la densité acoustique enregistrée. Il y a plusieurs techniques : de façon instantanée avec un ou plusieurs hydrophones déployés d'un bateau ou de façon (quasi-)permanente avec un hydrophone ou un réseau d'hydrophones fixés en surface ou sur le fond marin. L'inconvénient est que généralement ces hydrophones sont déployés proches des côtes, soit pour pouvoir transmettre l'information, soit pour pouvoir aller les récupérer sur zone. L'autre difficulté concerne le contenu des enregistrements : la grande diversité des émissions sonores des différentes espèces, la variabilité des émissions sonores entre individus d'une même espèce, la présence de bruits ambiants stationnaires et non stationnaires, et les déformations dues à la propagation acoustique.

Cette thèse s'inscrit dans la thématique de l'étude des cétacés à partir de leurs émissions sonores. Elle est financée par la région et par partenariat industriel, dans l'objectif initial d'innover des modèles d'analyse pour des stations autonomes, de type bouée, souvent mono-capteur, et à faible capacité de transmission de données. Dans ce contexte, nous avons développé et validé un nouveau modèle d'estimation de distance source-récepteur appliqué aux odontocètes, et nous avons proposé une méthode d'analyse de chant de mysticète qui en permet aussi une compression efficace pour leur gestion sur système autonome. Cette thèse a été soutenue par le projet Scaled Acoustic Biodiversity de la MI CNRS MASTODONS sur les masses de données scientifiques¹. Elle a aussi été soutenue par la fondation TOTAL et par l'Institut Universitaire de France, Toulon Provence Méditerranée, CG83 et l'UTLN.

1. <http://sabiod.org>

1

Objectifs de la thèse

Sommaire

1.1	L'intérêt de l'acoustique passive pour l'observation des cétacés	5
1.2	Le contexte de l'acoustique passive	6
1.3	Les types de signaux émis par les cétacés : dualité signaux transitoires <i>vs</i> voisements	6
1.4	Les principales problématiques	7
1.4.1	La détection	7
1.4.2	La classification	9
1.4.3	La localisation de la source	9
1.4.4	L'estimation de densité de population	10
1.5	Les limites actuelles	11
1.6	Développement de méthodes mono-hydrophone et perspectives en passage à l'échelle	12
1.7	Conclusion	13

1.1 L'intérêt de l'acoustique passive pour l'observation des cétacés

L'observation des cétacés est motivée par le besoin d'émettre des hypothèses quant à leurs modes de société, leurs comportements, leurs préférences d'habitats, leurs régimes alimentaires, leurs interactions avec les autres individus de la même espèce, d'une autre espèce, et les effets potentiels des activités humaines. Pour chaque étude, une ou plusieurs méthodes peuvent être déployées, car elles ont chacune leurs avantages, leurs inconvénients et peuvent être complémentaires. Par exemple, estimer la densité d'une population, suivre la trajectoire d'une baleine sur des échelles locales ou internationale, réaliser du "matching" d'individu d'une année à l'autre sur un site défini. Chaque procédé d'observation comporte ses propres "verrous" scientifiques ou physiques et limites techniques. Ces méthodes diffèrent aussi par le coût de leur mise en œuvre et les moyens humains engagés.

Or, les cétacés sont actifs vocalement. Leurs émissions sonores sont primordiales : elles servent à la communication, la cohésion du groupe, la recherche de nourriture, l'alimentation coopérative, les contacts mère-petit, la reconnaissance individuelle, potentiellement la signature acoustique (montrée chez les grands dauphins), la chasse, la socialisation, la recherche d'un partenaire,

le maintien d'un territoire, l'établissement d'une hiérarchie, la détections de prédateurs et de dangers et l'orientation.

Ainsi, les méthodes d'analyse par acoustique passive sont de plus en plus utilisées dans différents projets scientifiques, portant aussi bien sur le recensement, la répartition géographique ou l'étude des comportements. Il s'agit d'extraire, ordonner et exploiter l'information contenu dans l'environnement acoustique. L'aspect non-invasif rend ces méthodes particulièrement attractives et adaptées à l'étude des mammifères marins. Cependant, le choix du mode passif rend la tâche difficile. En effet, on ne contrôle que peu de paramètres. La plupart d'entre eux reste complètement ou partiellement inconnu comme les conditions initiales du signal émis (date et forme initiale et puissance), le nombre de sources acoustiques, les propriétés physico-chimiques du milieu de transmission et les propriétés acoustiques de la colonne d'eau. L'acquisition des signaux est réalisée par le biais de systèmes d'écoute acoustique mono-capteur ou multi-capteurs (réseaux) fixes ou mobiles. Le développement d'analyses des données bioacoustiques souffre d'un manque de "vérité terrain", pourtant primordial pour calibrer et évaluer les méthodes. On peut définir 4 problématiques majeures auxquelles la communauté scientifique tente de répondre via l'acoustique passive : détection, classification, localisation et estimation de densité. Les scientifiques s'efforcent à résoudre ces problèmes de manière automatisée et algorithmique.

1.2 Le contexte de l'acoustique passive

L'utilisation de l'acoustique passive pour l'étude des cétacés est récente. Son développement et son application explosent à partir des années 2000. En 2003 la communauté scientifique s'organise, en lançant le "International Workshop on Detection, Classification and Localization (DCL) of Marine Mammals using Passive Acoustics", dont une originalité est de rassembler des experts de différentes spécialités scientifiques : des biologistes, des acousticiens, des traiteurs de signaux, des informaticiens, des mathématiciens. A chaque session (une sur 2 ans), une nouvelle base de données est mise à disposition des chercheurs, leur permettant de proposer de nouvelles approches et de comparer leurs performances. L'acoustique passive devient attractive et commence à donner des résultats intéressants.

1.3 Les types de signaux émis par les cétacés : dualité signaux transitoires *vs* voissements

Le milieu marin dans lequel évolue les cétacés est particulièrement adapté à la propagation des ondes acoustiques, donc propice à la transmission de leurs émissions vocales. Cette activité bioacoustique offre l'opportunité de les étudier par le biais de méthodes totalement passives. La grande richesse des émissions acoustiques des cétacés s'exprime sur une large bande de fréquence : de quelque Hz à plusieurs centaines de kHz. Les caractéristiques intrinsèques des signaux (bande fréquentielle, forme, durée, énergie en sortie) varient d'une espèce à l'autre.

On distingue 2 grandes classes de signaux :

1. Les signaux d'une durée suffisamment longue pour engendrer un signal continu dans le temps et à bande de fréquence étroite. Mysticètes et odontocètes sont capables d'émettre ce type de signaux. Les appellations sont nombreuses, sifflements, cris, appels, chirps, voissements...
2. Les signaux de durée brève de type impulsifs dont le support fréquentiel est large tels que les clics ou des trains de clics. Ces signaux sont émis particulièrement par les odontocètes.

La bande de fréquence et l'énergie contenue dans le signal en sortie du générateur sonore de l'animal conditionne la distance sur laquelle se propage l'onde. Les basses fréquences voyagent sur de longues distances par opposition aux ultrasons. Aussi, les portées sur lesquelles certaines espèces sont détectables varient de l'ordre de la dizaine de m à la centaine de km. La forme du signal définit la sensibilité du signal émis aux effets de déformation engendrés par la dispersion, la réflexion et l'effet Doppler. Les propriétés acoustiques et la structure de ces différents types de signaux les rendent donc particulièrement adaptés pour certaines fonctions indispensables à la survie telles que :

- Signaux continus : la communication pour la cohésion de groupe, la coordination de méthodes de chasse en groupe, le contact mère-baleineau, la reconnaissance individuelle, la socialisation, la recherche de partenaire, la délimitation de territoire, l'orientation, l'alerte d'un danger.
- Signaux impulsifs : l'écholocation pour la détection de cible, la caractérisation de leur environnement, l'orientation, la recherche de proies et la communication. Le tableau 1.1 illustre quelques exemples des niveaux sonores à la source et des bandes fréquentielles de quelques espèces communes même si elles ne sont pas restreintes à un seul mode de production sonore. La grande variété des signaux implique autant de diversité dans les techniques de traitement des données bioacoustiques.

Source	SL (dB re μPa at 1 m)	bande de fréquence
Cachalots (clics)	163-223	100-30000
Belouga (clics)	206-225	-
Dauphin à nez blanc (clics)	194-219	-
Dauphin à long bec (burst)	108-115	-
Tursiops (sifflement)	125-173	800-24000
Tursiops (clics)	212-228	100-130000
Dauphin de risso	120	65000
Dauphin bleu et blanc	-	6000-24000
Rorqual commun (cris)	155-186	30-750
Baleine Bleue (cris)	155-188	-
Baleine boréale (cris, tonals, chants)	128-189	-
Baleine à bosse (chants)	144-174	30-8000
Baleine Franche du sud (appels)	172-187	-
Globicéphales	183	5000-20000
Crépitement de crevette	183-189	-

TABLE 1.1 – Caractéristiques de puissance en sortie et de fréquence des signaux pour quelques espèces usuelles (extrait de Oceans of noise, a WDCS Science report, eds : M. Simmonds, S. Dolman and L. Weilgart, 2004)

1.4 Les principales problématiques

1.4.1 La détection

A partir des signaux bruts, la phase de détection vise à notifier l'apparition et la disparition d'un événement acoustique recherché. Il se caractérise en réception par une variation d'énergie

du signal mesuré autour d'une valeur de référence. On assimile cette valeur au niveau de bruit ambiant auquel s'ajoute le bruit généré par le système d'acquisition. On définit le Rapport Signal à Bruit (RSB) [Bellanger 84, Ravier 92] sur le support temporel de l'événement (*ie.* sa durée) par :

$$RSB = \frac{E[s(t)^2]}{E[b(t)^2]} \quad (1.1)$$

où $s(t)$ est le signal temporel, $b(t)$ le bruit et E l'espérance mathématique. Les méthodes de détection sont fortement conditionnées par ce RSB. Au jour d'aujourd'hui, il n'existe pas de détecteur universel mais des détecteurs dédiés à des types de sons particuliers. Leurs performances dépendent en effet du type de signaux à détecter (impulsifs, non stationnaires, dont la durée est brève, signaux de type continus, bande étroite, large-bande...). Les types de détecteurs sont étroitement liés à leur application et aux objectifs de l'étude (extraction des dates de détections ou extraction de la forme de l'évènement acoustique). Il est alors difficile de les caractériser de manière générale.

Les méthodes les plus basiques sont basées sur une analyse et un seuillage de l'énergie (moment d'ordre 2) du signal reçu. Certaines méthodes utilisent les moments d'ordre supérieurs. Dans le cadre de la détection de signaux non gaussiens, Sangfelt a montré que les détecteurs fondés sur les moments d'ordre 4 aboutissaient à de meilleurs résultats [Sangfelt 93]. On peut citer aussi l'utilisation du bispectre (Transformée de Fourier à 2 dimensions de la bicorrélation) pour tester le caractère gaussien, linéaire ou stationnaire d'un signal [hinich 82, hinich 90]. Ces propriétés sont déduites d'une analyse de l'occupation du bispectre. D'autres types de transformations comme les transformées en ondelettes [Frisch 92] ont été utilisées pour la détection. Le principe consiste à comparer les coefficients en ondelettes d'un événement détecté avec ceux d'un modèle. Ces méthodes fonctionnent tant que les caractéristiques du signal réel et du modèle restent similaires.

Les signaux de types impulsifs et transitoires émis par les odontocètes sont large-bandes et de courte durée. Ils induisent localement une forte variation d'énergie sur une large bande de fréquence. Dans ce cas une analyse de l'énergie peut suffire à la détection sans prise en compte de la forme du signal. Dans [Morrissey 06], l'objectif des auteurs est l'extraction des temps de détection de clics de cachalot. Leur détection se déroule en 2 étapes : d'abord, ils appliquent une transformée de Fourier par fenêtre glissante avec un seuillage sur la variation d'énergie par fréquence et par unité de temps (moyenne sur 21.3 ms). Le spectre obtenu est alors binaire. Puis chaque fenêtre est analysée pour déterminer la présence de clics. Le détecteur génère "une carte des clics" au cours du temps. Dans [Lebot 14], l'auteur utilise une Analyse par Récurrence de Phase (RPA) *ie.* une analyse non linéaire qui consiste en une représentation dans l'espace des phases d'un signal mono dimensionnel. Ce procédé permet de faire apparaître les échantillons récurrents dans une série temporelle.

Les signaux continus de type voissements ou sifflements émis par les odontocètes et les mysticètes nécessitent le plus souvent de considérer les contours du signal détecté dans le plan temps-fréquences. Dans [Gillespie 04] l'auteur désire extraire les "calls" de Baleine Franche dans un jeu de données. Il procède dans un premier temps à un lissage des représentations temporelles, puis utilise un algorithme de détection de contours pour extraire les contours de signaux continus. Pour déterminer s'il s'agit de "calls", il compare avec une référence plusieurs paramètres comme la durée, la largeur de bande, la fréquence maximale et la fréquence minimale.

Dans l'article [Mellinger 04], l'auteur compare 2 méthodes de détections des "calls" de Baleines Franques : corrélation de spectre et réseau de neurones artificiels. La première consiste à con-

voluer une fenêtre d'observation du plan temps-fréquence avec un noyau synthétique représentant la forme du signal à détecter. Cette quantité appelée fonction de détection génère des pics pour les dates où les "calls" sont présents. Pour la deuxième méthode, on opère une régression de petites parties du spectrogramme appelées minigramme. Une phase d'apprentissage permet de déterminer les poids à affecter aux éléments d'entrée (minigramme) pour aboutir à la régression optimale. Une phase de test dans laquelle on interroge le réseau de neurones sur des entrées "jamais vues" par le système. La comparaison des courbes COR montre que le réseau de neurone est légèrement plus performant que la corrélation de spectre pour un faible RSB et légèrement moins performant pour un RSB élevé.

1.4.2 La classification

Après l'étape de détection, il peut s'avérer intéressant de déterminer l'origine des sons détectés et de les affecter à une classe. Les objectifs sont multiples : déterminer l'espèce, différencier des individus au sein d'une même espèce ou observer l'évolution dans le temps des chants ou cris. Les problèmes de classification posent la question de la représentation des données (signal ou descripteurs). En effet, il s'agit de projeter le signal dans un espace qui permette de différencier les classes voulues. La plupart des méthodes s'appuie sur celles développées pour les sciences computationnelles historiquement confrontées à ces questions. Les méthodes sont dites supervisées lorsque le nombre de classes est fixé, les données labélisées et que l'on utilise un ensemble d'*a priori* sur les objets à classer. Ce type de méthode nécessite une étape d'apprentissage sur une base de données labélisées. Les méthodes dites non supervisées vont créer automatiquement des classes en fonction du degré de différence entre les formes à classer.

L'extraction de descripteur peut être opérée dans le domaine temporel (durée, énergie, variance, moment d'ordre supérieur...) [Jin 01] mais demeure fortement sensible au niveau de bruit. Pour traiter les signaux de cétacés, les méthodes les plus usuelles sont basées sur des représentations fréquentielles (transformée de Fourier). Pour différencier les types de clics émis par le Grand Dauphin, Houser mesurent la position des pics principaux dans le spectre [Houser 99]. Gillespie propose de combiner descripteurs temporels et fréquentiels pour classer des signaux de mysticètes [Gillespie 04].

Par ailleurs, d'autres types de représentation ont été investigués comme les transformées en ondelettes et les coefficients des paquets d'ondelettes pouvant être de bons candidats pour les classificateurs de transitoires [Learned 92]. D'autres auteurs [Hussain 01] s'intéressent à l'analyse des composantes principales (PCA). Dans [Lopatka 07], l'auteur propose une méthode statistique pour la classification de clics d'odontocètes basée sur l'identification par mise en correspondance. Des caractéristiques sont extraites des signaux via une représentation orthogonale de Schur temps variant. Plusieurs méthodes de classifications dédiées à la bioacoustique sont présentées dans [Glotin 13b, Glotin 13d].

Enfin, l'emploi d'encodeurs neuronaux permet de bonnes classifications sur les signaux stationnaires, comme le démontre le récent article issue de mon équipe DYNI [Halkias 13].

1.4.3 La localisation de la source

Ce thème concerne les moyens possibles pour estimer la position dans l'espace de la source acoustique (émetteur). On retrouve cette information en utilisant les propriétés de propagation des ondes acoustiques combinées à des méthodes de traitement du signal. Les méthodes employées sont souvent liées à la configuration du système de détection : nombre de capteurs utilisés, largeur de la base et forme tridimensionnelle du réseau.

Dans le cas mono-capteur, on ne peut pas utiliser la cohérence spatiale entre les capteurs. En revanche ces méthodes sont transposables pour l'analyse multi-capteurs et dans les cas où la base du réseau est trop espacée pour assurer une cohérence spatiale et temporelle du signal. La première méthode est basée sur l'équation du sonar passif [Zimmer 11, Lurton 98], elle consiste à mesurer l'intensité acoustique à la réception, d'émettre des hypothèses sur l'intensité à l'émission (à 1 m de la source acoustique) et d'utiliser un modèle de propagation acoustique décrivant une loi de perte de puissance pour estimer la distance parcourue par l'onde sonore. Cette approche a été appliquée avec succès pour l'estimation des distances des baleines bleues Antarctique par [Samaran 10]. L'autre méthode est basée sur la théorie des hydrophones virtuels : la configuration entre la source, le capteur réel et les différents trajets acoustiques (directs, réflexions sur fond, réflexions sur la surface, réflexions multiples) permet de replacer géométriquement le problème dans les conditions de détection d'un réseau de plusieurs capteurs. Laplanche et Nosal décrivent comment appliquer cette méthode sur des enregistrements de clics de cachalot [Laplanche 05, Nosal 06]. Ces méthodes mono-capteur seront plus largement abordées dans la section sur l'état de l'art.

Le recours à un réseau d'hydrophones permet une analyse combinatoire entre chaque couple d'hydrophones. Il s'agit de calculer les différences de temps d'arrivée (TDOA) pour une paire d'hydrophones donnés. L'ensemble des points vérifiant un tel TDOA se trouve être une hyperbole ou une nappe hyperboloïde suivant les dimensions de l'espace de résolution. L'intersection des ensembles pour toutes les paires du réseau donne une estimation de la position de l'émetteur acoustique [Nosal 07, Giraudet 06a, Bénard 10b, Bénard 10a]. Une synthèse des estimations de vitesse et tests de différents détecteurs pour le cachalot sont disponibles dans [Bénard 11a]. Cette méthode impose un contrôle des capteurs, une connaissance précise de leur position et une cohérence spatio-temporelle, entre les signaux reçus sur l'ensemble des capteurs. Le cas à plusieurs cachalots est traité dans [Glotin 08a], un brevet a d'ailleurs été déposé [Glotin 07]. Enfin, un récapitulatif des méthodes de trajectographie développées au DYNI est disponible dans [Glotin 13c] et [Glotin 13a].

1.4.4 L'estimation de densité de population

L'objectif est d'estimer le nombre d'individus moyen sur une surface ou dans un volume de prospection. Jusqu'à ce jour, il existe 4 méthodes pour donner une estimation du nombre d'individus dans une zone géographique d'intérêt :

1. La déclinaison de Distance Sampling, méthode statistique définie pour des observations visuelles et consistant à étendre les événements observés dans une zone limitée à l'ensemble de la région. L'approche proposée par Thomas et appliquée par Marques en 2009 [Marques 09], reprend ce concept, à partir des clics de fausses orques. Il définit l'estimateur de densité suivant pour un capteur fixé :

$$\hat{D} = \frac{n_c(1 - \hat{c})}{\pi\omega^2\hat{P}T\hat{r}} \quad (1.2)$$

où n_c est le nombre de clics détectés sur une période T , le paramètre \hat{c} prend en compte le taux de fausses détections. ω est la distance maximale de détection, \hat{P} est la probabilité de détection et \hat{r} la distance de détection.

Toutefois, 2 paramètres sont difficiles à évaluer : ω , l'estimation de la distance de la cible, notamment lorsque celle-ci est éloignée de l'hydrophone et l'estimation du nombre d'émissions sonores par individus (le "call-rate") qui non seulement peut changer d'un individu

à l'autre, mais d'un mois à l'autre, voire d'une saison à l'autre. Si, grâce à l'analyse des émissions sonores, on est capable de trouver des localisations différentes à chaque son émis, alors c'est qu'ils proviennent d'individus différents. Cette technique est attractive même si pour l'instant, elle n'a connu de succès que pour un nombre limité d'espèces (en l'occurrence des baleines à bec. Cette technique peut se révéler également complexe lorsque les individus se déplacent.

2. Une autre méthode consisterait à utiliser la signature acoustique des individus pour les distinguer les uns des autres. Malheureusement, ce travail est difficile et sauf pour une ou deux espèces d'odontocètes (grand dauphins, belugas [Simard 10, Leblanc 08], et probablement orques [Murray 10]), la signature acoustique individuelle est impossible à extraire pour la majorité des espèces de cétacés, et en particulier les mysticètes.
3. Une autre méthode consiste à estimer les positions des individus à partir de leurs émissions sonores, et de les dénombrer ainsi [Glotin 08a], [Giraudet 06b].
4. Enfin, la distribution statistique des émissions sonores peut aider à calculer le nombre d'individus. Les cétacés émettent des sons simultanément et leurs sons sont souvent émis à un rythme particulier, en séquence et donc leur distribution contient une information sur le nombre d'individus émetteurs. Cette méthode a donné des résultats préliminaires intéressants pour l'estimation du nombre de baleines bleues Antarctique [Gonzalez 11].

1.5 Les limites actuelles

En terme de détection, les systèmes voient leurs performances limitées (fort taux de fausses détections) à cause de la grande diversité des sons émis intra et inter espèces de cétacés ; au bruit de mer ; mélange de bruits naturels, biologiques, bruits d'origine anthropique et bruits générés lors de la transmission des données. Pour notre étude, ce qui n'est pas émis par l'animal est considéré comme du bruit. La classification inter-espèces et l'extraction de signature individuelle ne sont absolument pas abouties. Les localisations sont parfois approximatives (grande surface/volume d'incertitude) voire rendues impossibles par les conditions de propagation (chenal de surface, réflecteurs aléatoires...). En 2011, il n'existait toujours pas de logiciel temps-réel .

Les méthodes d'estimation de densité requièrent de bonnes performances à la détection et à la classification. Elles nécessitent en plus une estimation de distance source-récepteur. Les modèles actuels sont très controversés et les résultats d'estimation de densité ne sont pas satisfaisant même s'il existe des pistes intéressantes.

L'impact des activités humaines (notamment acoustique) reste difficile à estimer. Faisons référence à Finneran en 2012, qui propose de regrouper les espèces de cétacés en 3 groupes suivant les bandes de fréquences de leurs émissions sonores et donc potentiellement de leurs bandes auditives. Cette avancée est intéressante mais il reste des progrès à faire, notamment descendre au niveau de l'espèce, de l'individu, et considérer également ses caractéristiques propres (âge, santé, comportement). Toutefois, Finneran introduit la notion de correction auditive en proposant des courbes de correction tenant compte de l'audiogramme, comme cela a été fait pour considérer l'impact du bruits sur notre audition. Les travaux en cours de Moretti (US Navy) sont particulièrement intéressants, puisqu'il introduit la notion de fonction risque, calculée sur les observations visuelles et acoustiques de baleines à bec évoluant dans un milieu bruité [Moretti 14]. Moretti se pose la question de la réponse auditive d'un individu (âge, sexe, comportement) et d'un groupe.

1.6 Développement de méthodes mono-hydrophone et perspectives en passage à l'échelle

La majorité des études est souvent effectuée par le biais de réseaux d'hydrophones. Même si ces méthodes sont souvent employées avec succès sur les mysticètes [Martin 11] et odontocètes [Nosal 07, Giraudet 06b] avec un haut niveau de précision, celles-ci nécessitent l'utilisation de matériel lourd et coûteux. Le nombre d'hydrophones à déployer et contrôler pour couvrir des zones étendues comme le sanctuaire PELAGOS (AMP en Méditerranée), est très important.

Cependant, l'utilisation d'un unique hydrophone mobile, bon marché et rapide à déployer, pourrait satisfaire certaines applications qui ne nécessitent pas une telle précision. Les points d'écoute mobiles, les systèmes anti-collision, ne requièrent qu'une estimation de la distance source-récepteur. Or, une grande quantité d'information est contenue dans un enregistrement mono-voie : 1) L'information nécessaire à la détection et à la classification. 2) L'information sur la distance de propagation source-récepteur et les angles de positions de l'animal au travers des pertes d'énergie et de la déformation du signal. L'analyse mono-capteur pourrait donc aussi constituer un moyen d'étude du comportement et aider à réaliser l'estimation de distance requise dans les problèmes d'estimation de densité de population. Par leur simplicité d'installation et leur faible coût, les dispositifs monocapteurs rendent possible la multiplication des points de mesure sur l'ensemble d'une zone géographique. L'ajout de capacités de transmission en temps réel offrent de véritables perspectives de passage à l'échelle pour le suivi des cétacés. De plus, la mise sur le marché de systèmes de mono-capteurs pour le grand public à travers les "smartphones" par exemple constitue une ressource importante favorisant le développement de la science participative et qui permettra le passage à l'échelle d'études en tous genres.

Il existe déjà plusieurs travaux fructueux. En 2010, l'étude sur les baleines bleues Arctiques [Samaran 10] est basée sur une estimation de distance mono-capteur. Il s'agit de retrancher le niveau reçu à un niveau à la source théorique fixe et de relier cette quantité à la distance de propagation par le biais d'un modèle de propagation adéquate. Cette méthode fonctionne particulièrement dans le cadre des baleines bleues dont les vocalises sont particulièrement stéréotypées. Pour la plupart des espèces, la variabilité des émissions sonores, rend cette méthode inutilisable. Dans les travaux [Laplanche 05, Nosal 06, Josso 10], les auteurs développent des méthodes visant à utiliser la cohérence spatio-temporelle des différents trajets acoustiques et recréent ainsi un réseau d'hydrophones virtuels. Nous reprendrons ces méthodes plus en détail dans le chapitre suivant.

Cette thèse s'inscrit dans un financement de type région PACA/partenaire industriel CeSigma. Le sujet alors validé au moment du lancement englobait la conception de systèmes d'écoute acoustique passive, la transmission et le stockage des données et le traitement des signaux pour accomplir une tâche de suivi de l'écosystème du littoral PACA.

Par la suite, nous avons précisé nos objectifs en fonction du matériel déjà existant tant sur le plan technologique que sur le plan théorique. L'aspect mono-capteur nous est alors paru primordial. Ce travail de thèse est focalisé en partie sur la thématique des méthodes d'analyse pour les systèmes mono-hydrophones via des représentations originales des données acoustiques. Il vise à apporter des solutions bons marchés pour d'une part, inciter les autorités et les acteurs de la protection à prendre des mesures pour le contrôle et le maintien des populations de cétacés ; et d'autre part dans une perspective de passage à l'échelle, il s'agit de préparer le développement et la conception de systèmes de mesures et d'analyse "tout terrain" commercialisables. En terme de dispositifs, nous avons travaillé avec CeSigma sur des systèmes simples et des structures pouvant être déployées rapidement avec des moyens logistiques limités. En terme d'analyse des données,

nous proposons plusieurs approches pour la caractérisation spatiale de la source. Dans un premier temps, nous avons développé des méthodes d'estimation de distance mono-capteur suivant un modèle théorique dérivant des lois de propagation de l'acoustique [Doh 13a, Glotin 11a, Doh 11b]. Puis nous nous sommes intéressés à un modèle d'apprentissage basé sur la neuromimétique pour la prédiction de distance source-récepteur [Doh 13b]. Le passage à l'échelle des analyses est aussi dépendant de la représentation des données, de leur segmentation et de leur compression en vue de leur transmission et de leur stockage. Nous aborderons cet aspect à travers la notion de représentation parcimonieuse ("Sparse Coding") des chants de Baleine à Bosse pour observer l'évolution dans le temps de leur structure [Doh 13c, Razik 11, Razik 13].

1.7 Conclusion

Nous avons introduit les moyens usuels d'étude des cétacés et positionné le travail de cette thèse parmi les techniques d'observations acoustiques. L'importance du caractère passif et non invasif a été souligné pour ne pas perturber les animaux.

Après avoir défini les problématiques principales qui motivent la communauté scientifique dans le domaine de l'analyse bioacoustique : détection, classification, localisation et estimation de densité, nous avons abordé les limites actuelles et les difficultés que soulèvent chacune de ces tâches :

- la détection d'événements dans un milieu bruité,
- la classification inter-espèce encore insatisfaisante et l'extraction de signature individuelle inaboutie,
- l'incertitude sur les localisations lorsque les conditions de propagation dégradent parfois les détections,
- l'aspect controversé de certains modèles d'estimation de densité.

Enfin, Nous avons présenté les intérêts de l'utilisation de capteurs isolés en terme de coût économique et de facilité de déploiement. Le mono-capteur permet de multiplier les points de mesures pour favoriser le passage à l'échelle des analyses. Nous avons donc précisé l'apport de cette thèse dans le développement des méthodes d'analyse mono-capteur. Un premier ensemble de contributions concerne l'estimation de distance source-récepteur sur signaux impulsifs. Un autre volet propose une représentation originale des signaux continus pour leur caractérisation, basée sur une décomposition par codage parcimonieux du signal.

Etat de l'art des méthodes d'analyse monohydrophone en bioacoustique

Sommaire

2.1	Moyens utilisés pour la réalisation de mesures acoustiques passives	15
2.2	Rappels d'acoustique sous-marine générale	16
2.2.1	Equations fondamentales	17
2.2.2	Atténuation	19
2.3	A propos de la détection	23
2.4	Estimation de distance source-récepteur	24
2.4.1	Concept des hydrophones virtuels	24
2.4.2	Equation du sonar passif	25
2.5	Extraction des caractéristiques sonores des cétacés	28
2.5.1	Sur les transitoires : exemple sur les cachalots	28
2.5.2	Sur les voisements : exemple des unités sonores de baleines à bosse	31
2.6	Conclusion	33

2.1 Moyens utilisés pour la réalisation de mesures acoustiques passives

Comme toute captation, les enregistrements sous-marins méritent une attention particulière sur la chaîne de mesure complète et sur l'environnement à surveiller. Cela signifie qu'il faut tenir compte en particulier de la sensibilité de l'hydrophone mais aussi du placement dans l'espace de celui-ci, comme lors d'un enregistrement de concert dans une salle de spectacle ou un amphithéâtre romain : les enregistrements sonores n'auront pas la même texture suivant l'emplacement du microphone par rapport à l'acteur. Le décor, le volume, les courants d'air, le bruit ambiant, etc., tout cela va venir impacter l'enregistrement d'une façon ou d'une autre. Il en est de même pour les acquisitions sous-marines, auxquelles il faut ajouter un mode de propagation acoustique différent de l'aérien.

L'acoustique sous-marine passive consiste à placer un ou plusieurs hydrophones dans une colonne d'eau pour enregistrer les ondes sonores environnantes. Plusieurs choix peuvent être faits concernant le nombre d'hydrophones, la profondeur à laquelle les placer, le plan du réseau (dans le cas de plus d'un hydrophone), les paramètres d'enregistrement.

Le recours à des installations en surface ou placées sur le fond marin est motivé principalement par les possibilités technologiques pour recueillir les données et alimenter les matériels. Ainsi, même si la disposition des systèmes sur le fond marin a des avantages certains, cela n'est possible que lorsqu'on est proche des côtes si l'on souhaite pouvoir poser des câbles d'alimentation et de transfert de données. Se positionner loin des côtes implique que l'on accepte d'accéder aux données *a posteriori* et non en temps réel.

La question du recours à un ou plusieurs hydrophones est plus délicat. Motivée originalement par des questions de coût financier, l'utilisation d'un seul hydrophone a été longtemps privilégiée, notamment pour des déploiements instantanés à partir du bateau, soit pour aider à la recherche de cétacés lorsque ces derniers n'étaient pas visibles à la surface, soit pour étudier des émissions sonores notamment les chants des baleines à bosse ou les sons émis par les groupes d'orques. C'est plus récemment que l'installation d'hydrophones sur zone a fait son apparition : à la fin des années 90, Pop-up de Cornell², et apparition des T-Pod³, puis au début des années 2000 pour l'Aural de MultiElectronique⁴. Ces hydrophones sont totalement autonomes, déposés sur le fond marin ou suspendus à une bouée en surface. Ils enregistrent pendant une période déterminée à l'avance, focalisés sur l'étude d'une espèce particulière avec les contraintes des capacités de stockage et de l'autonomie énergétique. A notre connaissance, au début de cette thèse, aucune station permanente équipée d'un hydrophone n'est disponible sur le marché ou utilisée par les chercheurs. Les raisons sont principalement techniques avec les contraintes suivantes :

- la transmission des données : le recours à des traitements *in situ* pour trier les enregistrements où un cétacé est détecté des enregistrements où il n'y a que du bruit : le recours à des algorithmes de compression ne suffisent pas
- la gestion de l'alimentation électrique et les capacités de stockage.

Les détails sur le type de chaîne d'acquisition sont donnés dans la partie matériel. Ce chapitre vise à faire le point sur le matériel retenu dans le cadre de mon travail de thèse, en terme de dispositif et de base de données.

2.2 Rappels d'acoustique sous-marine générale

De leur production à leur propagation, les données bio-acoustiques que nous traitons ont une origine physique. Le traitement mono-capteur vise à utiliser certaines propriétés issues des phénomènes entrant en jeu dans la propagation du son entre sa source et notre récepteur. Dans cette section, nous rappelons les équations fondamentales de l'acoustique régissant la propagation du son. Une onde acoustique est une onde mécanique se propageant dans un milieu. Nous nous placerons dans le cadre de l'acoustique linéaire : le passage de l'onde n'induit que de faibles variations d'amplitude des fonctions décrivant l'état d'un fluide compressible, autour d'une position d'équilibre mécanique (hypothèse des petits déplacements et des petites déformations). L'acoustique linéaire ne peut pas expliquer les phénomènes d'onde de choc dépendants de termes non-linéaires qu'on ne peut pas négliger. Le fluide est supposé parfait. Les effets de la viscosité, la dissipation d'énergie associée et ses aspects non linéaires sont donc négligés. Cette considération importante va nous permettre de linéariser les équations fondamentales de l'acoustique. Dans notre développement, l'état du fluide en un point x et à l'instant t est caractérisé par la vitesse particulière $v(x, t)$, la densité de masse $\rho(x, t)$ et la pression $p(x, t)$.

2. <http://www.birds.cornell.edu>

3. <http://www.t-pod.co.uk>

4. <http://multi-electronique.com>

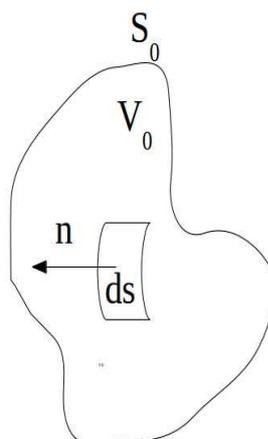


FIGURE 2.1 – Représentation d'un volume élémentaire de fluide n est le vecteur normal à un élément de surface ds .

2.2.1 Equations fondamentales

Equations fondamentales de l'acoustique

– loi de continuité ou de conservation de la masse

On considère un volume V_0 de l'espace et S_0 la surface l'entourant (figure 2.1). La variation temporelle de masse de fluide totale contenue dans ce volume est : $-\partial_t \left(\iiint_{V_0} \rho dV \right)$:

En appliquant la loi de conservation de la masse, nous trouvons que cette quantité est égale à la quantité de masse sortante par unité de temps à travers la surface S_0 entourant V_0

$$\int_{S_0} \rho v dV = \iiint_{V_0} \text{div}(\rho v) dV \quad (2.1)$$

Nous obtenons l'équation de continuité faisant intervenir la vitesse particulière v dont les composantes sont indicées par i et l'opérateur divergence div . Sous sa forme locale l'équation 2.1 devient :

$$\partial_t \rho + \partial_{v_i} \rho v_i = 0 \quad (2.2)$$

– loi de conservation de la quantité de mouvement

En utilisant la relation fondamentale de la dynamique appliquée à un volume élémentaire de fluide et l'équation de continuité, nous obtenons :

$$\rho (\partial_t v_j + v_i \partial_i v_j) = -\partial_j P \quad (2.3)$$

où P est la pression.

– équation d'état

L'équation d'état thermodynamique, reliant la pression à la densité constitue la troisième équation du système. On suppose que l'onde acoustique induit de faibles variations de pression autour d'une pression p_0 à l'équilibre. Du point de vue thermodynamique, on considère que les cycles de compression-dilatation des éléments de fluide sont isentropiques. Ce qui permet d'admettre que la pression ne dépend pas de l'entropie. Elle est fonction uniquement de la

densité. Dans ces conditions, nous pouvons effectuer un développement limité de p en p_0 autour de ρ_0 jusqu'à l'ordre 2 :

$$p = p_0 + \frac{\partial P}{\partial \rho}(\rho - \rho_0) + \frac{1}{2} \frac{\partial^2 P}{\partial \rho^2}(\rho - \rho_0)^2 \quad (2.4)$$

où le terme $\frac{\partial P}{\partial \rho}$ est homogène à une vitesse au carré, il sera noté C_0^2 par la suite.

Linéarisation des équations constitutives

Les équations que nous venons d'établir comportent des termes non linéaires. Elles peuvent être linéarisées sous certaines conditions. L'onde acoustique ne provoque que de faibles variations des grandeurs physiques par rapport à leur valeur à l'équilibre, elles sont désignées par l'indice a . Le fluide est supposé initialement au repos, la vitesse est donc nulle. Nous posons alors :

$$\begin{aligned} \rho &\triangleq \rho_0 + \rho_a \\ \vec{v} &\triangleq 0 + \vec{v}_a \\ p &\triangleq p_0 + p_a \end{aligned} \quad (2.5)$$

Tel que :

$$\begin{aligned} \frac{\rho_a}{\rho_0}, \frac{p_a}{p_0}, \frac{v_a}{C_0} &\approx O(\epsilon) \\ \frac{\rho_a}{\rho_0}, \frac{p_a}{p_0}, \frac{v_a}{C_0} &\ll 1 \end{aligned} \quad (2.6)$$

La dernière condition fait intervenir le nombre de Mach acoustique $\frac{v_a}{C_0}$ où C_0 est la célérité de l'onde dans le fluide. La linéarisation des équations est raisonnable si le nombre de Mach est inférieur à 10^{-3} . Suivant ces hypothèses on peut négliger les termes des équations constitutives d'ordre supérieur à ϵ .

– Linéarisation de l'équation d'Euler

$$\rho_0 \partial_t v_a = -\partial_j p_a \quad (2.7)$$

– Linéarisation de l'équation de continuité

$$\partial_t \rho_a + \rho_0 \partial_i v_{ai} = 0 \quad (2.8)$$

– Linéarisation de l'équation d'état

$$p_a = C_0^2 \rho_a \quad (2.9)$$

Equations de propagation

L'équation de propagation (*i.e.* l'équation d'Helmholtz) se déduit par une combinaison de la divergence de l'équation d'Euler et la dérivée temporelle de l'équation de continuité. Elle se présente ici en fonction de la pression mais la vitesse et le potentiel sont aussi solution de l'équation d'onde. Par la suite on omettra l'indice a par souci d'allègement d'écriture :

$$\nabla^2 p - \frac{1}{C_0^2} \frac{\partial^2 p}{\partial t^2} = 0 \quad (2.10)$$

∇^2 représente l'opérateur Laplacien. Dans le cas où la dépendance temporelle de p est de type $\exp(j\omega t)$ (i. e. en régime harmonique de pulsation ω), l'opérateur $\frac{1}{C_0^2} \frac{\partial^2}{\partial t^2}$ peut être remplacé par le vecteur d'onde $k^2 = \frac{\omega^2}{C_0^2}$

Nous cherchons à présent l'expression d'une solution dans l'espace 3D. D'abord, réécrivons l'équation de propagation en coordonnées sphériques selon r le rayon :

$$\left[\frac{1}{r} \frac{\partial}{\partial r} r^2 \frac{\partial}{\partial r} + k^2 \right] p(r) = 0 \quad (2.11)$$

La solution de cette équation différentielle se présente sous la forme :

$$p(r) = A \frac{\exp(jkr)}{r} \quad (2.12)$$

Il s'agit d'une onde sphérique dont l'amplitude présente une décroissance typique en $1/r$.

2.2.2 Atténuation

Phénomènes d'atténuation de l'énergie en transmission

Lors de la propagation, l'onde acoustique subit des pertes d'énergie d'origines multiples.

– Les pertes par divergence géométrique.

L'onde sphérique, au fil de sa propagation, voit son énergie se répartir sur des surfaces de front d'onde de plus en plus grandes. Pour des raisons de conservation de l'énergie, on observe une décroissance de celle-ci en $1/r^2$. Ce qui est en accord avec le terme obtenu pour la solution de l'équation d'onde en 3D.

– Les pertes par atténuation fréquentielle.

Elles sont dues aux interactions entre l'onde et le milieu de propagation. Plusieurs modèles basés sur des relations empiriques (Leroy, Thorp, Francois et Garisson) décrivent cet amortissement en fonction des paramètres physico-chimique du milieu (salinité, température, PH, pression...).

Atténuation au passage des interfaces

Au cours de sa propagation dans l'océan, l'onde acoustique rencontre des interfaces du guide d'onde formé par la surface (liquide/liquide) et le fond (liquide/solide). Les phénomènes qui en résultent sont alors différents (conversion onde P en onde S par exemple dans le cas de l'interface liquide/solide). Nous ne traiterons que le cas liquide/liquide. On désigne respectivement par p_i , p_t et p_r , la pression des ondes incidentes, transmises et réfléchies.

$$\begin{aligned} p_i &= \exp\left(j \frac{\omega}{C_1} (x \cos \theta_1 + z \sin \theta_1)\right) \exp(j\omega t) \\ p_t &= T \exp\left(j \frac{\omega}{C_1} (x \cos \theta_2 + z \sin \theta_2)\right) \exp(j\omega t) \\ p_r &= R \exp\left(j \frac{\omega}{C_1} (x \cos \theta_1 - z \sin \theta_1)\right) \exp(j\omega t) \end{aligned} \quad (2.13)$$

Nous appliquons les conditions de continuité à l'interface pour la pression et la vitesse en $z = 0$.

$$\begin{aligned} [p] &= 0 \\ \left[\frac{1}{\rho} \frac{\partial p}{\partial z} \right] &= 0 \end{aligned} \quad (2.14)$$

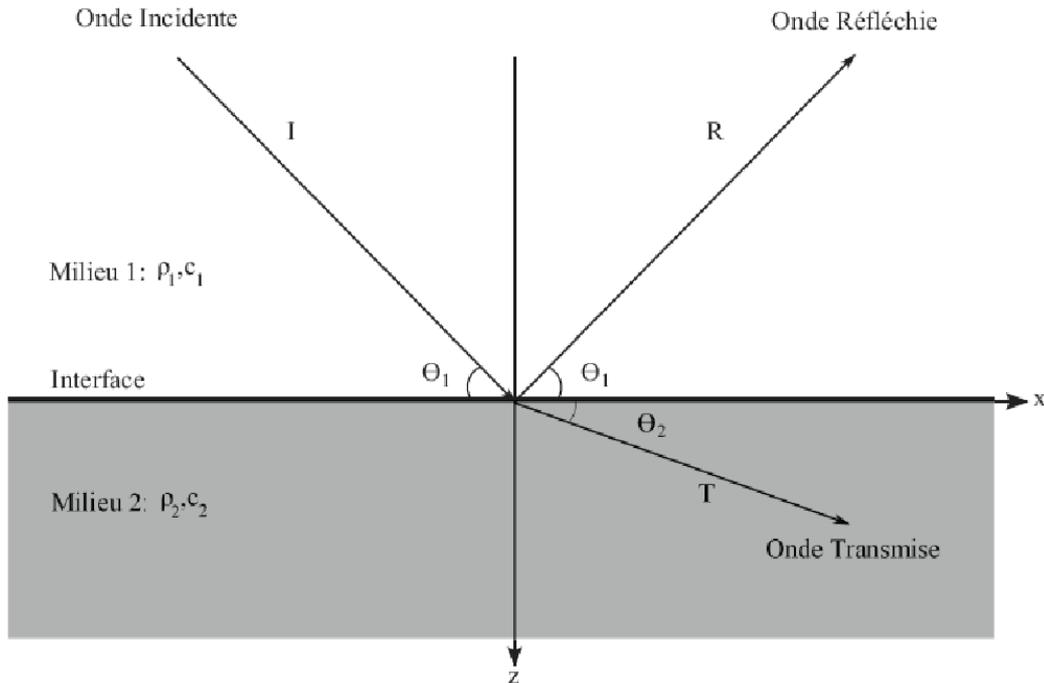


FIGURE 2.2 – Passage d'une onde à l'interface liquide/liquide. Représentation dans le plan sagittal [Josso 10].

Nous introduisons l'impédance acoustique du milieu homogène par $Z_0 = -\frac{P}{v} = \frac{\rho_0 C_0}{\cos \theta_0}$

Après quelques calculs pour la résolution de ce système, nous pouvons établir l'expression de la loi de Snell-Descartes :

$$\frac{\cos \theta_1}{C_1} = \frac{\sin \theta_2}{C_2} \quad (2.15)$$

Nous en déduisons l'expression des coefficients de réflexion R et de transmission T :

$$\begin{aligned} R &= \frac{Z_2 - Z_1}{Z_2 + Z_1} \\ T &= \frac{2Z_2}{Z_2 + Z_1} \\ T &= 1 + R \end{aligned} \quad (2.16)$$

Les réflexions sur la surface de l'eau sont caractérisées par une importante rupture d'impédance acoustique entre l'eau de mer (célérité : 1500 m/s, masse volumique : 1000 kg/m³) et l'air (célérité : 340 m/s, masse volumique : 1.2 kg/m³). Les coefficients de réflexion et de transmission deviennent $R \approx -1$ et $T \approx 0$. Peu d'énergie est transmise dans l'air, l'essentiel est réfléchi sous l'eau et l'onde subit un déphasage de π . L'interface eau/air agit comme un miroir quasi-parfait face aux ondes acoustiques. Notons que cette approximation n'est vérifiée que lorsque la mer est calme, les réflexions sur le fond sont plus complexes à traiter. En effet, les impédances acoustiques sont souvent du même ordre. Une partie de l'énergie incidente est donc transmise et réfléchi en proportion de R et T suivant les angles décrits par la relation de Snell-Descartes. Il existe

toutefois un angle d'incidence critique pour lequel le coefficient de réflexion devient complexe. Alors l'essentiel de l'énergie est réfléchi, on dit qu'il y a réflexion totale. L'angle critique est donné par la relation $\theta_c = \arccos \frac{C_1}{C_2}$.

Propagation du son en milieu océanique

La vitesse du son dépend surtout de la température, de la pression et de la salinité. Or, en milieu océanique, ces paramètres varient en fonction de la profondeur et subissent l'influence des saisons. La célérité du son suit donc un profil selon z , propre à chaque région. On observe souvent 4 zones particulières. La première est appelée couche de surface. Elle est exposée au rayonnement solaire, aux variations de températures nocturne/diurne et chahutée par les vagues, bulles, vent... La célérité du son y est très fluctuante sur de courtes échelles de temps. La hausse de pression influence principalement la célérité. Celle-ci augmente avec la profondeur. La deuxième est la thermocline saisonnière. Cette couche subit encore l'influence des saisons. La troisième est la thermocline principale. La température décroît avec la profondeur et surpasse l'influence de la pression. La célérité décroît jusqu'à la couche isothermale profonde. La température est alors constante pendant que la pression augmente toujours. La célérité augmente de nouveau jusqu'au fond.

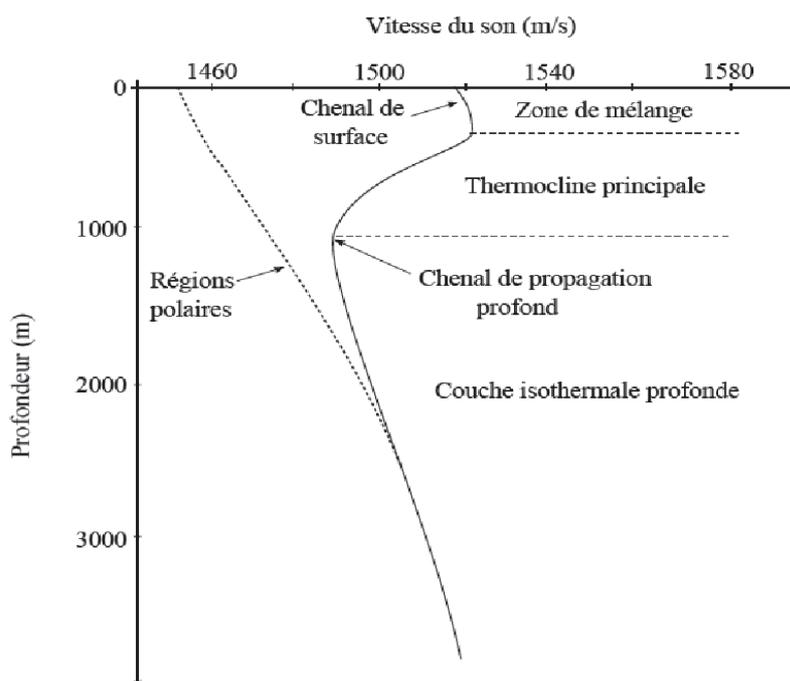


FIGURE 2.3 – Profils de célérité typique [Josso 10].

L'onde acoustique se propage donc dans un milieu dont la célérité varie continûment. Cette situation peut être traitée partiellement en introduisant l'expression de la pression telle que l'amplitude A et la phase S sont réels et dépendent de la distance de propagation $p(r) = A(r) \exp(iS(r))$

- En injectant cette expression dans l'équation d'Helmholtz, nous obtenons deux équations :
- l'équation de transport régissant l'amplitude des ondes en milieux d'indice variable,
 - l'équation Eikonale décrivant le comportement des ondes dans un tel milieu.

Après quelques calculs et approximation sur la deuxième équation, nous aboutissons à :

$$\frac{\cos \theta(z)}{c(z)} = \text{constante} \quad (2.17)$$

où $\theta(z)$ est l'angle entre le rayon de l'onde et la surface à la profondeur z , $c(z)$ la célérité du son à la profondeur z .

Cette relation correspond à la généralisation de la loi de Snell-Descartes. En effet, cette situation peut être traitée en considérant que l'onde se propage dans un milieu à n interfaces d'indice variable (milieux stratifiés). Selon la loi de Snell-Descartes, chacune d'elles va dévier la direction d'un rayon. La propagation d'un rayon n'est plus rectiligne. Toutefois l'approximation rectiligne reste valable par petits fonds, c'est-à-dire lorsque $h < 10\lambda$, avec λ la longueur d'onde. Les situations de profil de célérité présentant des minima locaux engendrent des chenaux de propagation. Ce sont de véritables guides d'onde, où l'énergie acoustique est canalisée sur de longues distances. Le profil décrit en figure 2.3 présente 2 extrema créant un chenal de surface et un chenal de propagation profond. La figure 2.4 présente un exemple de profil de célérité et une simulation de lancer de rayon dans ces conditions.

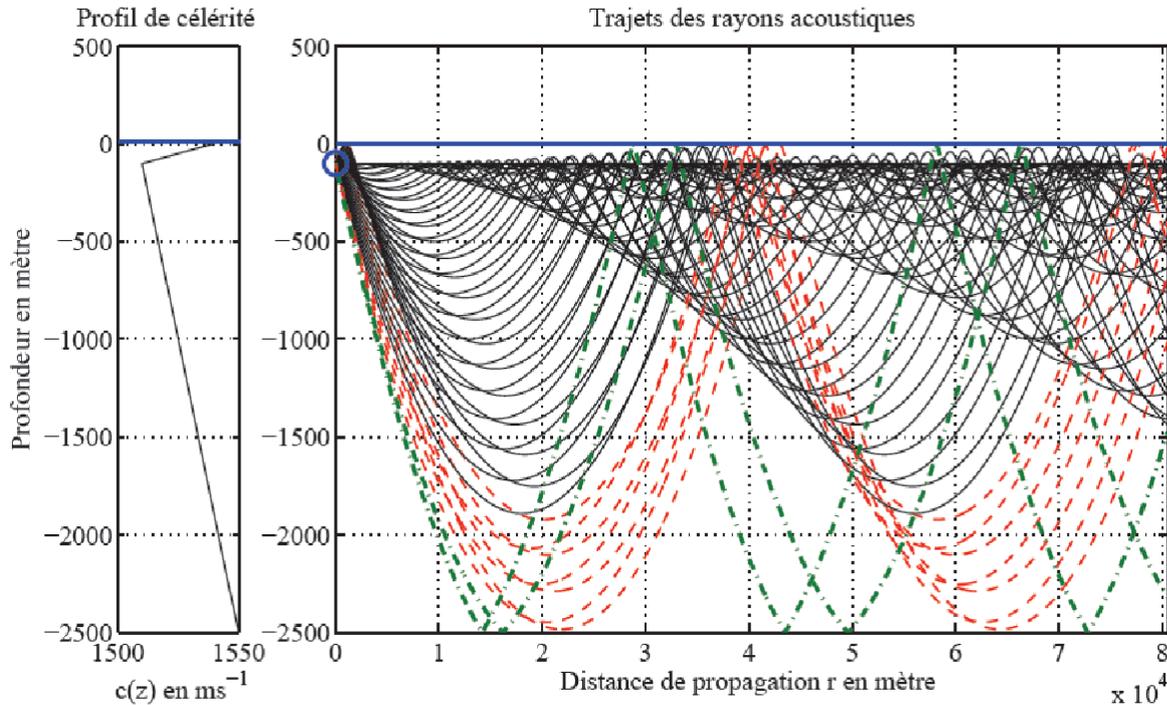


FIGURE 2.4 – Profil de célérité et tracé de rayons typiques de la mer Méditerranée. La source est située à une profondeur de 100 mètres et est illustrée par un cercle. On peut remarquer la présence de rayons réfractés (traits pleins), de rayons réfractés et réfléchis en surface (traits hachurés) et de rayons réfractés, réfléchis fond et surface (traits pointillés hachurés) [Josso 10].

Notion d'énergie acoustique

L'énergie acoustique se définit comme la variation d'énergie induite par le passage de l'onde dans un fluide homogène et au repos. La notation $\langle \rangle$ désigne l'opérateur moyenne. La densité

totale d'énergie est donnée par :

$$\langle \delta W \rangle = \frac{1}{2\rho_0 C_0^2} \langle p^2 \rangle + \frac{1}{2} \rho_0 \langle \vec{v}\vec{v} \rangle \quad (2.18)$$

On reconnaît à gauche, un premier terme d'énergie cinétique liée à la vitesse des particules de fluides autour de leur position d'équilibre. Le deuxième terme exprime l'énergie potentielle liée aux effets de compressibilité du fluide. Pour une onde sphérique, la relation $\vec{v} = \frac{p}{\rho_0 C_0} \vec{n}$, permet de monter l'équipartition de l'énergie potentielle et cinétique :

$$\langle \delta W \rangle = \frac{1}{\rho_0 C_0^2} \langle p^2 \rangle = \rho_0 \langle \vec{v}\vec{v} \rangle \quad (2.19)$$

L'énergie acoustique est donc proportionnelle au carré de la pression. L'amplitude du signal mesuré est liée à la pression par une relation linéaire. L'énergie du signal reçu par l'hydrophone est donc proportionnelle à la pression au carré. Par conséquent, l'énergie acoustique est accessible via l'énergie du signal par une relation de proportionnalité.

2.3 A propos de la détection

La détection consiste à repérer une information d'intérêt dans un processus. Dans ce travail, il s'agit d'identifier, dans des enregistrements acoustiques continus, les émissions sonores de certaines espèces de cétacés sélectionnées à l'avance suivant le ou les objectifs scientifiques de l'étude. On peut également restreindre la détection à un type de son particulier émis par l'espèce (par exemple détecter uniquement les clics des dauphins et non leurs sifflements).

Le plus simple détecteur consiste à calculer l'énergie du signal sur une fenêtre temporelle glissante et de faire l'hypothèse qu'une émission sonore aura une amplitude plus importante que celle du bruit de fond, le reste étant considéré comme du bruit. Il est possible de décliner cette approche dans le domaine fréquentiel, offrant la possibilité dans ce cas, de choisir une ou plusieurs bandes fréquentielles en fonction des caractéristiques des émissions sonores de l'espèce ou des espèces de cétacés visées. L'intérêt de cette technique est qu'elle ne nécessite aucun *a priori* sur le signal à détecter. Il est possible d'améliorer les performances en utilisant un seuil adaptatif qui réglé automatiquement par rapport au niveau de bruit ambiant (par exemple, en le réajustant chaque heure). Il est également possible de comparer des énergies sur 2 bandes fréquentielles différentes : celle où l'on cherche les émissions sonores d'une espèce et une autre à l'extérieur de cette bande fréquentielle. Ainsi, lorsque ces 2 valeurs énergétiques ne sont pas constantes ou s'éloignent l'une de l'autre, la présence d'un événement sonore d'intérêt est soupçonné. Plus récemment, il a été introduit pour la détection des transitoires dans les enregistrements, le recours à l'opérateur de Teager-Kaiser (Kaiser, 1990, 1993). Ce détecteur a été appliqué avec succès sur la détection de clics de cachalots (Kandia and Stylianou, 2005, Lopatka et al., 2008). Malheureusement, ces techniques basées sur l'énergie dépendent du rapport RSB et leurs performances ont tendance à décliner lorsque le rapport signal-sur-bruit décroît notamment lorsqu'il atteint des valeurs négatives inférieures à -5 voire -10 dB.

Une autre approche consiste à tenir compte des informations *a priori* sur les émissions sonores des cétacés. L'ensemble des espèces émettent des signaux stéréotypés ayant des caractéristiques temporelles et fréquentielles déterministes. Les plus usuelles sont la durée, la forme, l'amplitude, la fréquence fondamentale, la présence d'harmoniques voire de formants. Il est également possible d'utiliser ces caractéristiques pour créer des signaux de référence (template), puis d'utiliser ces références pour créer un détecteur basé sur la corrélation temporelle ou fréquentielle

[Mellinger 00]. Les caractéristiques peuvent également être basées sur l'extraction de paramètres purement mathématiques n'ayant pas obligatoirement de signification physique claire ou immédiate mais issus de transformées mathématiques : les transformées en ondelettes, la représentation de Schur, les modes intrinsèques issus de la transformée de Hilbert-Huang, La récurrence de phase [Lopatka 05, Adam 05, Adam 08, Lelandais 08, Lebot 14]... Ces paramètres peuvent être utilisés ensuite pour la classification des sons détectés, notamment leur attribution à une espèce.

2.4 Estimation de distance source-récepteur

2.4.1 Concept des hydrophones virtuels

Dans l'océan, la propagation des ondes acoustiques se fait dans un volume contraint par deux limites : la surface et le fond marin. Sur ces deux dioptries, les coefficients de réflexion et d'absorption ne sont pas les mêmes. Alors que pour la surface, la réflexion des ondes sonores de fréquence inférieure à 10 kHz est considérée comme quasi-totale, il est en différément pour le fond marin. Dans ce cas, plusieurs paramètres sont à prendre en compte : les différents types de sédiments (et notamment la couche du premier sédiment), la bathymétrie, le relief... Toutefois, dans certains cas, les hydrophones disposés dans la colonne d'eau peuvent recevoir une émission sonore via un trajet direct et via des trajets plus complexes, issus d'une ou de plusieurs réflexions avec la surface et/ou le fond marin.

Suivant la profondeur de l'hydrophone, les propriétés de réflexion du fond et l'état de la surface, les hydrophones peuvent recevoir le signal issu du trajet direct suivi des trajets indirects dûs aux réflexions simples et aux réflexions multiples.

Par des relations de géométrie, on peut montrer que cette situation est équivalente à un réseau d'hydrophones unidimensionnel dont 1 est réel et les autres virtuels [Josso 10, Au 08, Mouy 67]. Ce réseau est constitué d'autant d'hydrophones virtuels qu'il y a d'échos du aux réflexions. Cette méthode a été appliquée avec succès pour la localisation de cachalots au cours de leur plongée [Laplanche 05]. Les clics de cachalot, émis à très fortes intensité (>190 dB re1uPa à 1 m) et de bande passante large (avec des pics fréquentiels autour de 10 kHz) sont de très bons candidats pour générer des échos facilement détectables, notamment par grand fonds et lorsque la nature du fond est particulièrement réfléchante (type roche par exemple). Ainsi, dans le cas favorable où le cachalot est relativement proche de l'hydrophone (l'intensité reçue en trajet direct reste supérieure à 150 dB) et que les conditions favorisent les réflexions multiples (fond marin réfléchissant, type roche et mer calme en surface minimisant la dispersion), il est possible d'enregistrer plusieurs échos. Si de plus, l'animal est dans l'axe de l'hydrophone ou qu'il pointe verticalement vers le bas ou vers le haut, les échos sont plus énergétiques et donc plus facilement détectables dans les enregistrements. Si le signal acquis comporte le trajet direct, une réflexion fond et une réflexion, alors la profondeur peut être déduite ainsi que la distance horizontale de l'animal [Laplanche 05].

$$z_s = \frac{4z_b\tau_{ds}(z_b - z_h) - c_0^2\tau_{ds}\tau_{db}\tau_{sb}}{4(z_b\tau_{ds} + z_h\tau_{sb})} \quad (2.20)$$

$$c_0\tau_d = \frac{4z_s z_h - (c_0\tau_{ds})^2}{2C_0\tau_{ds}} \quad (2.21)$$

$$x_s = \pm \sqrt{(c_0\tau_d)^2 - (z_s - z_h)^2} \quad (2.22)$$

où z_b est la profondeur du fond, z_h la profondeur de l'hydrophone, τ_{ds} , τ_{db} et τ_{sb} sont les temps de retard respectivement direct/réfléchi surface, direct/réfléchi fond et refléchi surface/réfléchi

fond. τ_d est le temps de propagation du trajet direct et c_0 la célérité du son dans la colonne d'eau.

En pratique, le signal reçu ne comporte souvent qu'une partie de ces différents trajets acoustiques ce qui rend difficile l'utilisation systématique de cette méthode.

2.4.2 Equation du sonar passif

Lorsqu'une onde acoustique se propage dans milieu, son énergie est atténuée. L'atténuation dépend en partie de la distance parcourue. Il est donc possible de relier les pertes d'énergie à la distance source-récepteur. Etablissons tout d'abord l'expression, en dB, de l'énergie du signal E reçu à la distance r de la source en fonction du niveau d'énergie à la source E_{SL} et des pertes par transmission (TL) [Viala 07, Au 08] pour une fréquence donnée. Nous nous plaçons dans le cadre d'une onde sphérique divergente omnidirectionnelle (le modèle de divergence cylindrique peut être considéré). Nous admettons aussi que l'émetteur et le récepteur sont omnidirectionnels et qu'il n'y a pas de gain apporté par le capteur ou le pré-traitement. L'énergie reçue s'exprime alors ainsi :

$$E(r, f) = E_{SL}(f) - TL(r, f) - NL(f), \quad (2.23)$$

où NL est le niveau de bruit ambiant ou bruit de masquage. Ce bruit provient de 4 contributions majeures [Zimmer 11] :

1. Le bruit de surface généré par l'agitation des masses d'eau et l'éclatement des bulles en surface. Ce terme est dépendant de la fréquence, de l'état de la surface (vent) et de la profondeur du capteur. Notons qu'il s'agit de la contribution la plus large bande [Lurton 98].
2. Le bruit de bateaux issu du rayonnement des bruits de machines via la structure rigide des coques. A cela s'ajoute les bruits de cavitation générés par la rotation de l'hélice. L'énergie spectrale est principalement limitée à la bande [10; 1000] Hz.
3. Le bruit de turbulence océanique sensible à très basse fréquence (< 10 Hz).
4. Le bruit thermique produit par l'agitation aléatoire des molécules du milieu. La gamme spectrale impactée est très haute fréquence (> 100 kHz).

En figure 2.5, le spectre du bruit est représenté en échelle logarithmique. La courbe en trait plein foncé est la somme des 4 contributions décrites ci-dessus. Le bruit ambiant est principalement basse fréquence, à 10 kHz le niveau d'énergie spectral est de 30 dB. Cette courbe n'est qu'une estimation théorique du bruit basée sur des modèles empirique mais ne remplace pas des mesures *in situ*.

Le terme TL peut être décomposé comme :

$$TL(r, f) = 20 \log_{10}(r) + \alpha(f)r, \quad (2.24)$$

où r est la distance de propagation (en m), f est la fréquence (en Hz) et α est le coefficient d'atténuation fréquentielle en ($dB.m^{-1}$). Le premier terme de l'équation 2.24 est dû aux pertes par divergence géométrique. Le deuxième terme est induit par les interactions entre l'onde et les particules du milieu. On peut considérer plusieurs modèles : Leroy [Leroy 65], Thorp [Thorp 67]... Nous donnons ici l'expression du modèle de Thorp adapté à la description de l'océan atlantique exprimant la dépendance fréquentielle de l'atténuation :

$$\alpha(f) \triangleq \frac{f^2}{0.9144} \left(\frac{0.1}{1 + f^2} + \frac{40}{4100 + f^2} \right). \quad (2.25)$$

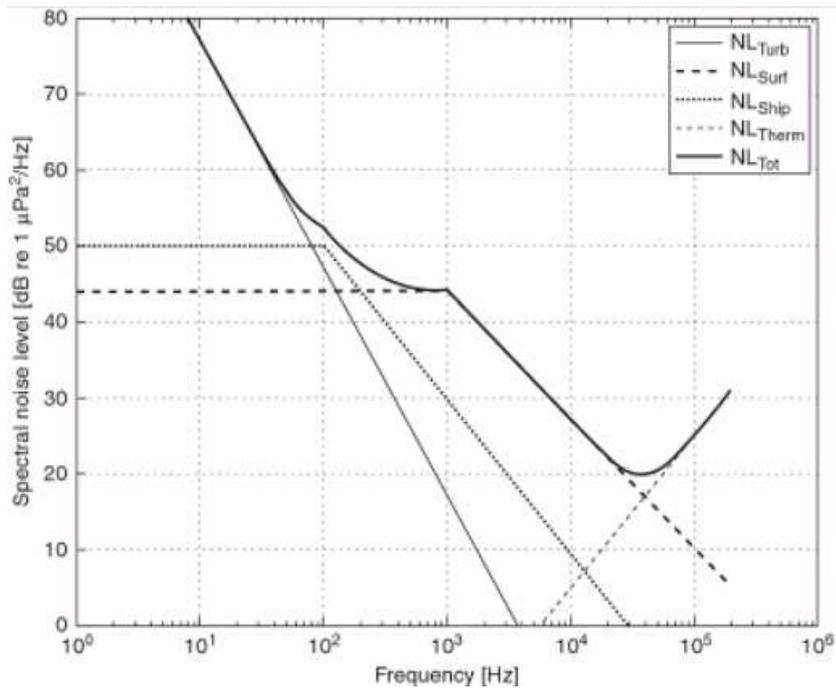


FIGURE 2.5 – Modèle de bruit ambiant pour une mer calme, un état de trafic maritime limité et une profondeur d'observation de 100 m [Zimmer 11].

Le coefficient d'atténuation fréquentielle α calculé selon Eq. (2.25) est illustré en Fig. 2.6. On remarque que l'atténuation n'est pas linéaire et que les hautes fréquences s'atténuent plus fortement que les basses fréquences.

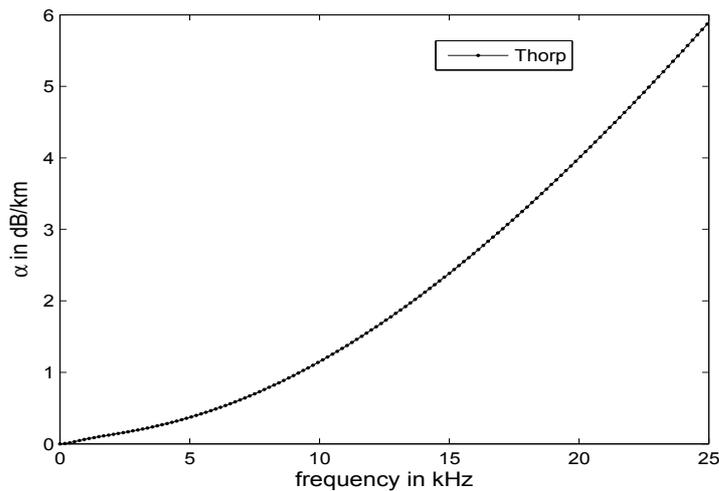


FIGURE 2.6 – Coefficient α (en dB.km^{-1}) en fonction de la fréquence (kHz) selon Thorp [Thorp 67].

En première approximation, on peut négliger l'atténuation fréquentielle devant les pertes par divergence. La dépendance en fréquence des TL devient négligeable, ce qui nous permet de raisonner à partir de l'énergie totale du signal E_0 . Le problème $r = F(E_0)$ nécessite alors un modèle théorique ou statistique du SL. l'équation sonar est solvable dans le cas de sources acoustiques contrôlées dont le niveau à la source (SL) est stable et dont le diagramme de rayonnement est connu. Dans le cas des sources acoustiques bioacoustiques (non-contrôlées), cette méthode devient beaucoup plus difficile à mettre en œuvre du fait de la directivité spécifique des biosonars et du caractère arbitraire des productions sonores.

2.5 Extraction des caractéristiques sonores des cétacés

2.5.1 Sur les transitoires : exemple sur les cachalots

Mode production sonore : le bipôle acoustique

Les odontocètes disposent d'un appareil de production sonore associé à un système écholocatif très développé contenu dans le complexe naso-frontal [Au 93]. Cette aptitude à utiliser les ondes acoustiques tant pour la chasse que pour la communication entre individus est probablement à l'origine de leur succès écologique. En effet, ils peuplent tous les océans, la plupart des mers et certaines régions fluviales. Parmi les odontocètes, le cachalot présente une taille disproportionnée et le plus haut niveau de perfectionnement de l'organe de production acoustique (1/3 de la masse et de la taille totale de l'animal).

Les lèvres phoniques sont composées de tissus conjonctifs en communication avec les poumons et l'évent. En régime de plongée, les lèvres phoniques fonctionnent en circuit fermé avec les poumons et peuvent initier une vibration acoustique orientée vers l'arrière du complexe naso-frontal. Ce système est composé d'un premier module graisseux appelé organe à spermaceti, d'un réflecteur acoustique nommé sac frontal et d'un second module graisseux, le junk, faisant office de focaliseur acoustique. L'animal projette un signal directif devant lui de puissance variable. Le complexe naso-frontal est un véritable dipôle acoustique biologique [Laplanche 05]. L'écho est reçu et transmis au cerveau par le biais de la mâchoire inférieure dissociée de la partie supérieure sur le plan osseux. Une représentation schématique du complexe naso-frontal est donnée en figure 2.7.

Le cachalot émet des signaux transitoires appelés clics. Certains sont destinés à la communication entre individus d'autres servent à l'écholocation. L'intervalle entre les clics est appelé Inter Clic Interval (ICI). Celui-ci varie au cours du temps en fonction de l'activité de l'animal : il diminue à mesure qu'il se rapproche de sa proie (phase de « burst »). On remarque qu'une réduction d'ICI s'accompagne d'une diminution de l'énergie des signaux émis. En effet, l'animal est limité physiologiquement par la constante de temps de son organe de production sonore et le fait qu'il fonctionne en circuit fermé.

Nous décrivons ici le mécanisme de production des clics de cachalot (la figure 2.7 illustre les étapes) :

1. la contraction des lèvres phoniques initialise une vibration acoustique (P0) faiblement directive de puissance moyenne.
2. une partie de cette onde se propage dans le milieu marin.
3. l'autre est transmise à travers l'organe à spermaceti vers l'arrière du complexe naso-frontal et se réfléchit sur le sac frontal.
4. à ce stade, le signal effectue 3 trajets acoustiques différents :
 - une partie rayonne directement dans le milieu (PF1).
 - une autre est transmise dans le junk, composé de masses graisseuses et assurant un rôle de lentille acoustique ; le signal y subit une focalisation le rendant plus énergétique et directif avant d'être émis vers l'avant de l'animal (PJ1).
 - une dernière partie du signal réfléchi sur le sac frontal traverse l'organe à spermaceti vers le sac distal.
5. une partie rayonne dans l'environnement (P1) l'autre est réfléchi de nouveau vers l'arrière et accomplit un nouveau cycle PF2, PJ2, P2... [Laplanche 05]. Les signaux de type PJ ne seraient observables que dans le cas où l'hydrophone est placé proche de l'axe du cachalot (« on axis »).

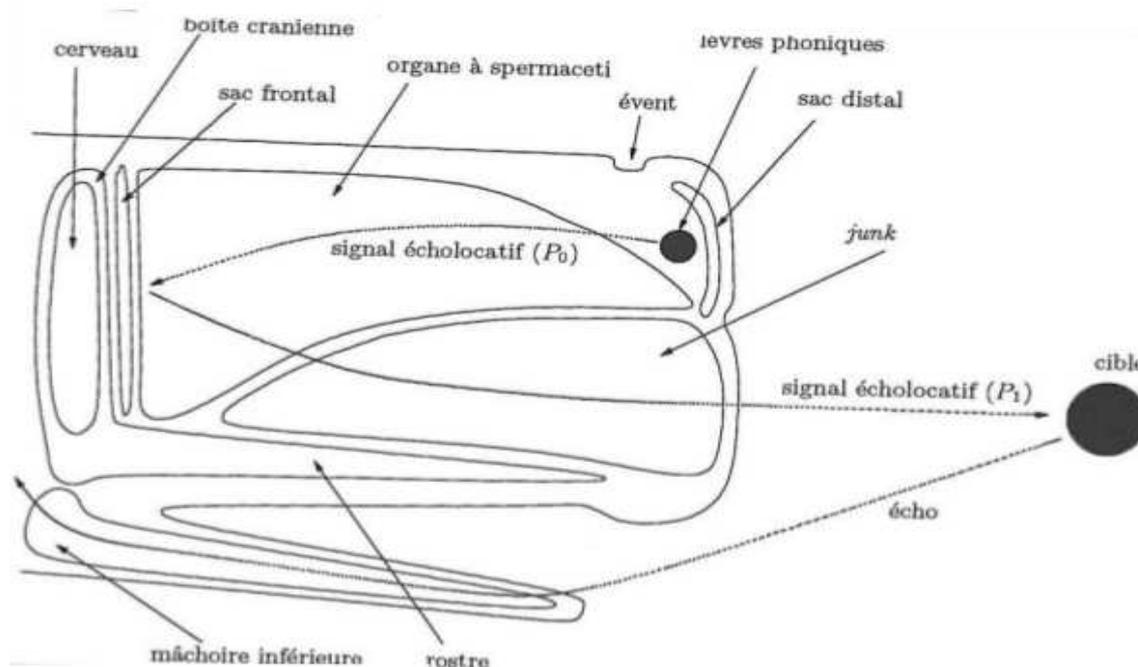


FIGURE 2.7 – Coupe sagittale de la tête du cachalot et trajets issus de la vibration initiale P_0 responsables de la structure multipulsée [Laplanche 05].

Les clics présentent une structure typique dite multipulsée, conséquence du mode de production sonore ainsi décrit. L'ordre des pulses en termes de temps d'arrivée varie selon la position du récepteur par rapport à l'axe antéro-postérieur du cachalot [Laplanche 05]. La structure temporelle d'un clic "vu" dans l'axe (on axis) diffère de la structure d'un clic "vu" hors de cet axe. La figure 2.8 représente ce cas avec les pulses labélisés.

Comme nous l'avons vu précédemment les pulsations mesurées proviennent de réflexions successives dans la tête du cachalot. Les différences de temps d'arrivée entre P_0 et P_1 , P_1 et P_2 sont constantes et correspondent au temps mis par le son pour parcourir un aller retour dans la tête. La mesure de ces retards et la connaissance de la célérité du son dans le spermacéti peuvent permettre d'évaluer la taille de la tête. Ce retard est baptisé Inter Pulses Interval (IPI). En supposant que la célérité du son dans le spermacéti est proche de celle du son dans l'eau de mer C_0 , la taille de la tête est donnée par : $IPI.C_0/2$. Une loi d'échelle permet de retrouver la taille de l'individu (abaques) [Goold 96]. Une proposition de mesure automatique est proposée par la thèse de Abeille [Abeille 13]

Directivité et contenu fréquentiel

Les pulsations issues de ce mode de production diffèrent par leur contenu fréquentiel, leur directivité et leur intensité. La pulsation directive $PJ1$ présente une fréquence centroïde d'environ 15 kHz pour un niveau de 240 dB re : $1 \mu\text{Pa}$ [Mohl 00]. Les autres pulsations peu directives montrent une centroïde de l'ordre de quelques kHz pour un niveau d'environ 180 dB re : $1 \mu\text{Pa}$ [Zimmer 04].

L'efficacité des signaux émis en terme d'écholocation est, entre autres, conditionnée par leur fréquence. En effet, celle-ci doit correspondre à une longueur d'onde inférieure à la taille

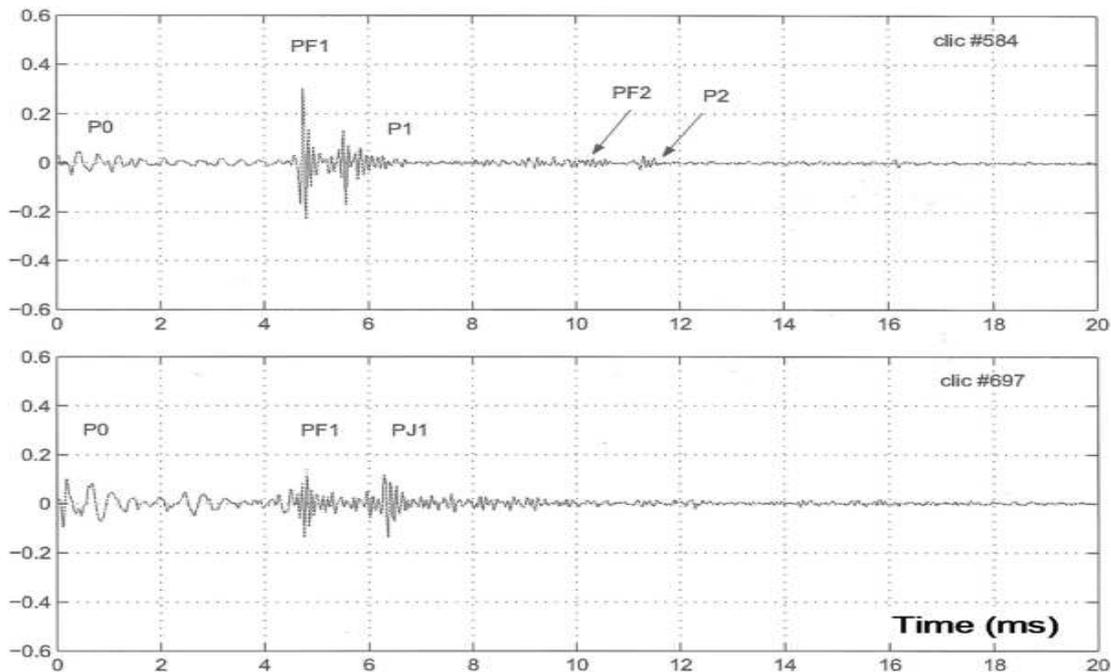


FIGURE 2.8 – Différents types de pulses suivant la position de l'hydrophone. On-axis (en haut) et off axis (en bas). Illustration tirée de la thèse de C. Laplanche, 2005 [Laplanche 05].

des proies. Un intervalle de fréquence de 1 à 15 kHz permet de localiser des proies de 10 cm à 1,5 m, ce qui semble correspondre à son régime alimentaire. Le niveau sonore global en sortie de l'animal présente un diagramme de directivité représenté en figure 2.9.

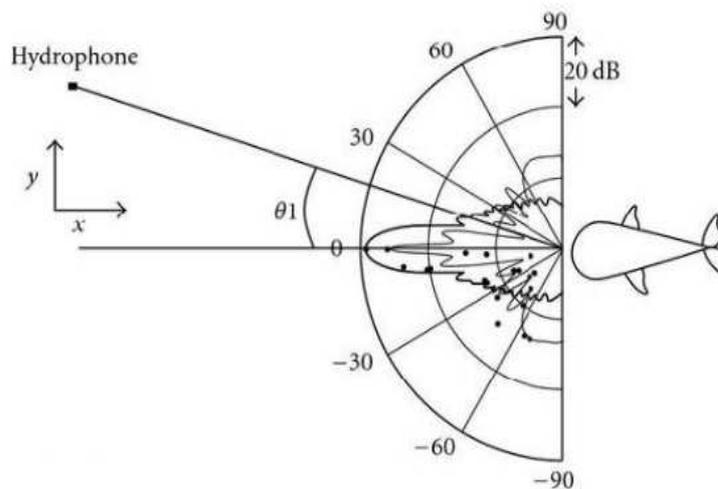


FIGURE 2.9 – Représentation du diagramme de directivité des émissions sonores du cachalot dans un plan longitudinal [Zimmer 05].

Dans [Au 93], l'auteur propose d'utiliser le modèle acoustique baffé du piston circulaire plan pour simuler la directivité des émissions de clics d'odontocètes. L'équation 2.26 donne l'expression de la pression p rayonnée à une distance r de la source en fonction de l'angle θ de réception :

$$p(\theta) = (j/2) \frac{\rho a c k}{r} 2 \frac{J_1(ka \sin(\theta))}{ka \sin(\theta)}; \quad (2.26)$$

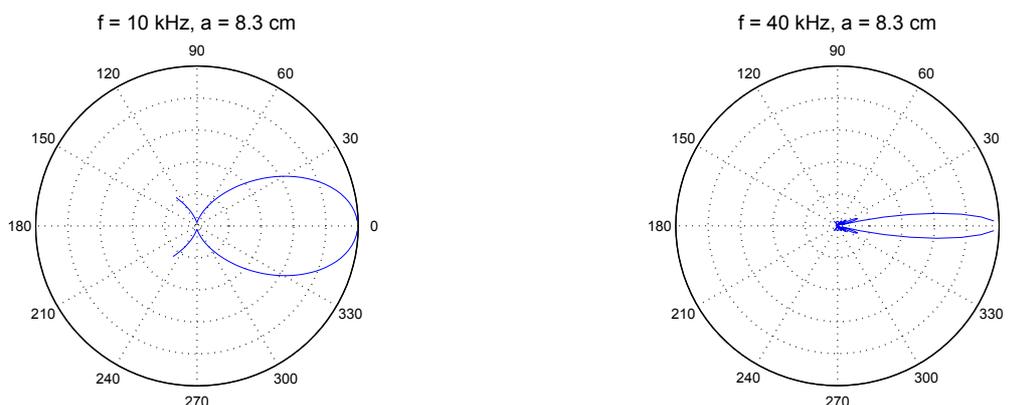


FIGURE 2.10 – Diagramme de directivité du rayonnement d'un piston circulaire plan de diamètre 8.5 cm. A gauche pour la fréquence 10 kHz, à droite pour la fréquence 40 kHz.

où ρ est la masse volumique du milieu de propagation, a le diamètre du piston, c la célérité du son dans le milieu, et le vecteur d'onde associé à une pulsation de fréquence f telle que $k = \frac{2\pi}{\lambda}$ et $\lambda = \frac{c}{f}$. La figure 2.10 représente le module de la pression à 1 m pour des fréquences de $f = 10$ kHz et $f = 40$ kHz. On constate que le diagramme est d'autant plus directif que la fréquence est élevée.

2.5.2 Sur les voissements : exemple des unités sonores de baleines à bosse

Au cours de l'automne et de l'hiver, les interactions entre les différentes baleines à bosse sont particulièrement intenses. Il s'agit de la saison de reproduction et cette espèce de mysticète déploie des stratégies individuelles complexes, basées sur des actions d'attractions mâles/femelles, d'intimidation entre les mâles et de délimitations des territoires. Dans ces échanges, les activités vocales occupent une place très importante. Ces baleines émettent une large diversité de sons, de nature pulsée et de nature harmonique [Cazau 13]. En particulier, certains mâles émettent des vocalises organisées dans le temps, sous forme de séquences et de phrases constituant le leitmotiv d'un chant [Payne 71]. Ces vocalises, nommées « unités sonores », sont de complexités variables : de durée limitée, bornées par 2 silences, de forte intensité acoustique (supérieure à 160 dB re 1uPa à 1 m), avec une fondamentale basse fréquence (de l'ordre de 100 Hz), avec ou sans harmonique ou formant pouvant s'étendre au-delà de 10 kHz (figure 2.11). Leurs caractéristiques temporelles et fréquentielles sont liées à l'anatomie de leur zone laryngée et à leurs motivations comportementales [Reidenberg 07, Adam 13, Cazau 13]. Certaines variations de ces caractéristiques ont été signalées et peuvent contenir de l'information [Au 05]. L'hypothèse principale est que ces chants jouent un rôle dans l'attraction des femelles [Winn 78, Medrano 94] et éventuellement dans la défense territoriale [Tyack 81]. Différentes équipes ont montré que la distance entre deux mâles qui vocalisent est plus grande que celle entre deux mâles silencieux

[Frankel 95, Darling 01]. Baker et Herman ont émis l'hypothèse que les chants pouvaient permettre la synchronisation de l'ovulation [Baker 84]. Ces chants sont prédominants dans la zone de reproduction mais ont également été enregistrés lors de la migration et de temps en temps dans la zone d'alimentation [Clapham 90, Clark 04]. Ce manque de connaissance est expliqué par le fait que les zones d'alimentation ne sont pas adaptées pour les observations visuelles et acoustiques, contrairement aux zones de reproduction.

Miksis indique que « la structure des chants de baleines à bosse de l'hémisphère nord a été étudiée davantage que celle des populations de l'hémisphère sud » [Miksis-Olds 08]. Noad évoquent la copie de chants entre les mâles de la côte Est australienne et ceux de la côte Ouest australienne [Noad 00b]. C'est la première fois qu'une « révolution culturelle » a été montrée chez cette espèce.

Même si dans les années 1980 Payne a précisé la structuration des chants, ils sont toujours étudiés et la grande variété de ces unités sonores décrite ci-avant rend leur analyse très difficile pour les classificateurs automatiques. Les chants sont cycliques et composés d'une séquence continue et structurée de sons pouvant être répétés plusieurs fois de façon continue. Les chants sont décomposés en unités sonores (un son continu entre deux silences) ; plusieurs unités formant une séquence, les séquences formant une phrase, plusieurs phrases formant un thème de chant [Payne 71]. Celui-ci dure en moyenne 2 à 4 minutes, parfois plus de 10 minutes. Trois à neuf thèmes de chant peuvent former un chant. Les chants évoluent au cours d'une saison, d'une année à l'autre, avec des modifications substantielles des caractéristiques de certaines unités sonores, la disparition de certaines unités sonores et l'apparition de nouvelles. Les chants sont également différents d'une région à l'autre. La figure 2.11 est un exemple de représentation temps fréquence de chant de baleine à bosse. La structuration automatique de ce chant a été posée en challenge internationale par l'équipe DYNi au workshop NIPS4B associé à NIPS 2013 [Glotin 13d], dans le cadre du projet SABIOD de la MI CNRS MASTODONS, les détails et le signal complet sont en ligne sur <http://sabioid.org>. On distingue environ 8 unités sonores.

Plusieurs méthodes ont été suggérées pour la détection des unités sonores, leur segmentation puis leur classification. Le filtre adapté et la corrélation de spectrogramme [Mellinger 00, Abbot 12] sont, encore actuellement largement utilisés par la communauté des chercheurs en bioacoustique. Plus récemment, des techniques issues du traitement des images ont été déclinées pour analyser des représentations temps-fréquences, comme notamment la détection de contours [Gillespie 04, Mallawaarachchi 08, Madhsudhana 08] et le Pitch tracking [Oswald 07, Shapiro 09, Baumgartner 11]. Urazghildiiev et Clark (2006) ont proposé une approche statistique basée sur le test du maximum de vraisemblance [Urazghildiiev 06]. La mesure de l'entropie a également été appliquée avec succès [Suzuki 06]. Plusieurs méthodes utilisées pour analyser la parole humaine ont été appliquées aux baleines à bosse. En effet, on observe de nombreuses similitudes, notamment la présence de vocalisations de type voisé et non voisé telles que définies dans [Mercado III 98]. Les unités sonores des baleines à bosse ont été analysées en utilisant le codage par prédiction linéaire (LPC) [Mercado III 98], le contenu énergétique sur une fenêtre de temps spécifique [Rickwood 08], l'analyse spectrographique [Suzuki 06], les coefficients cepstraux à échelle Mel (MFCC) [Helweg 96, Mazhar 08, Picot 08, Glotin 08b, Pace 10], l'affinité de propagation [Glotin 08b], Kmeans [Pace 10], la classification par cartes auto-organisatrices (SOM) [Mercado III 98, Suzuki 06], les modèles de Markov cachés (HMM) [Pace 11], la détermination de la longueur de fenêtre glissante par estimation de l'entropie (SWML) [Suzuki 06] et les réseaux neuronaux. La grande variété des méthodes utilisées par les chercheurs pour analyser les vocalisations des baleines à bosse reflète la grande diversité des caractéristiques intrinsèques de ces sons [Harris 06].

Une des difficultés dans l'analyse automatique des chants de baleines à bosse est que ceux-ci

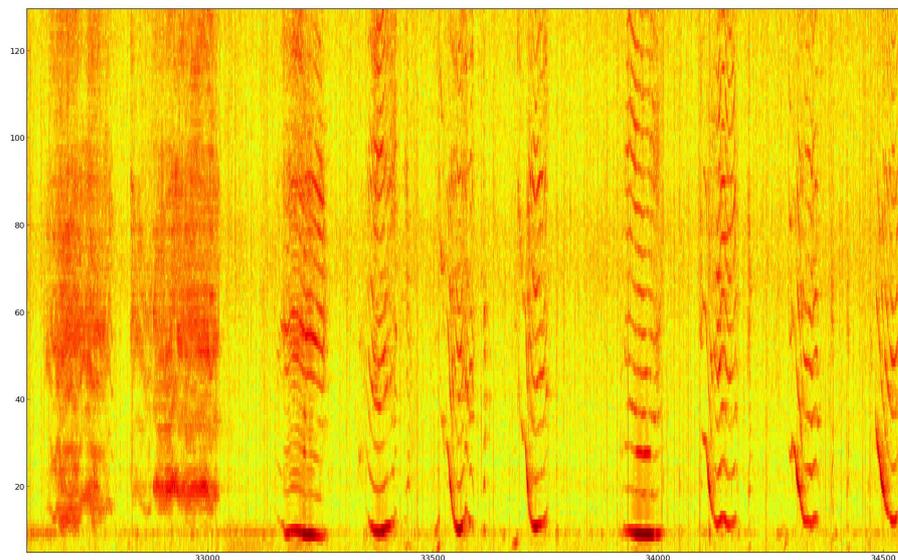


FIGURE 2.11 – Spectrogram d’un chant de baleine à bosse. Début : 5 min 40 s, fin : 6 min et en ordonnées de 0 à 22.05 kHz, codé sur 512 bins (fft calculée sur 1024 points sur une fenêtre de 10 ms). Enregistré par Darewin à la Réunion.

sont évolutifs dans le temps, laissant penser que le nombre d’unités sonores est illimité. Pour pallier cette situation, et dans le but de mieux analyser les chants d’une année sur l’autre et sur des populations de différentes régions, nous avons défini le concept de sous-unité [Glotin 08b, Pace 09]. Nous suggérons qu’une ou plusieurs sous-unités sont présentes pour composer une unité sonore. L’intérêt de cette approche est de montrer que le nombre de sous-unités est limité voire restreint et qu’elles pourraient être utilisées pour caractériser les unités des chants, non limitées, mais construites à partir des combinaisons de sous-unités temps-fréquences.

Néanmoins, la sous-unité et le concept de l’unité doivent encore être améliorés et entièrement automatisés. Les travaux de [Pace 09], ne proposent pas un codage non supervisé sous contrainte de parcimonie. L’analyse et le regroupement des vecteurs MFCC sont effectués par une simple analyse de voisinage (à différentes échelles de temps), sans extraire les éléments codant le plus efficacement les signaux (chants, articulations...). L’équipe de Picot construit des classes d’unité de chant sur des mesures de fondamentale et de prosodie, suivies d’un regroupement en six classes avec un critère de Bouldwin [Picot 08]. Mais ce regroupement est dépendant de l’initialisation, problème connu des algorithmes kNN. Dans l’article de [Glotin 13e], il est proposé de regrouper des unités de chant extraites de décomposition parcimonieuses pour obtenir des représentations plus efficaces en suivi d’animaux et analyse des chants.

2.6 Conclusion

Dans ce chapitre, nous avons commencé par décrire les moyens principaux pour la mise en œuvre de mesures acoustiques et l’organisation générale de la chaîne de mesure. Puis, nous avons

rappelé les éléments théoriques classiques de l'acoustique sous-marine permettant d'aboutir à des modèles de perte d'énergie lors de la propagation du signal.

Par la suite nous effectuons une présentation des méthodes d'analyse mono-hydrophone déjà existantes en lien avec les contributions de cette thèse. Pour l'estimation de distance, nous décrivons les méthodes basées sur le concept des hydrophones virtuels ainsi que celles utilisant l'équation du sonar passif.

Nous terminons ce chapitre en précisant l'état des connaissances sur les caractéristiques des signaux émis par les espèces choisies pour nos études :

- le cachalot pour les signaux transitoires,
- la baleine à bosse pour les signaux continus de type voisements.

Deuxième partie
Contributions

3

Proposition d'un modèle d'estimation mono-récepteur de distance de source transitoire

Sommaire

3.1	Base théorique	37
3.2	Représentation de l'estimateur distance ISAT et paramètres importants	40
3.3	Architecture et mise œuvre de l'estimation ISAT	42
3.3.1	De l'échelle d'un clic à l'échelle d'un pulse	42
3.3.2	Dans le domaine fréquentiel	43
3.3.3	Estimations de distance ISAT pour un choix de bandes fixe	44
3.3.4	Adaptation à l'ensemble du domaine spectral par multi-combinaisons de bandes	44
3.3.5	Filtrage des estimations sur des fenêtres d'observations de courtes durées	45
3.3.6	Détermination des combinaisons de bandes optimales à partir de la vérité terrain	46
3.3.7	Sélection des combinaisons de bandes en conditions réelles	46
3.4	Conclusion	47

3.1 Base théorique

Le modèle théorique d'atténuation intra-spectrale "ISAT" que nous proposons a pour objectif, à partir du signal enregistré à l'hydrophone, d'estimer la distance parcourue par l'onde acoustique depuis la source qui l'a émise. Le facteur de perte par transmission TL dépend de la fréquence et de la distance parcourue. Le partie appelée atténuation fréquentielle est dépendante de la fréquence de manière non linéaire. Ce terme a tendance à favoriser les basses fréquences par rapport aux hautes fréquences (effet passe-bas). La conséquence est donc une déformation du spectre original. Celle-ci est d'autant plus importante que la distance de propagation est grande. Une analyse à l'échelle intra-spectre permettra de relier cette déformation à la distance source récepteur afin de l'estimer. Notre première approche se veut théorique et tente de fournir à partir de la physique, une relation simple entre un descripteur spectral et la distance à estimer. Le but

de cette expression est de se défaire des termes décrivant la divergence géométrique et surtout du niveau à la source (SL) moyennant des simplifications. L'originalité de cette approche est d'exploiter le ratio des énergies de 2 bandes de fréquences. Dans ce raisonnement, nous considérons uniquement le signal issu du trajet direct, selon un modèle de propagation géométrique en milieu isotrope. Nous n'intégrons donc pas les pertes d'énergie dues aux réflexions en se plaçant en espace libre. Cette étude vise à valider des estimations de distance compte tenu des hypothèses simplificatrices utilisées pour établir ce modèle théorique.

Soit $x(t; r)$ le signal temporel reçu à l'instant t , à la distance r de la source, sur une durée d'observation T et $X(f; r)$ sa transformée de Fourier. Selon le théorème de Parseval, l'énergie du signal peut s'exprimer à partir du spectre par :

$$E \triangleq \int |X(f; r)|^2 df. \quad (3.1)$$

Les signaux considérés sont large-bandes, c'est-à-dire de durée brève devant la fenêtre d'observation de durée T et peuvent être décrits comme une somme de signaux monochromatiques. L'intensité totale reçue résulte de la contribution de toutes les composantes fréquentielles, subissant chacune leur propre loi de décroissance.

Soit $\epsilon(r, f)$ l'énergie élémentaire exprimé en dB à partir de l'équation 2.24 :

$$\begin{aligned} \epsilon(r, f) &\triangleq \epsilon_0(f) - TL(r, f) \\ &= \epsilon_0(f) - 20 \log_{10}(r) - \alpha(f)r, \end{aligned} \quad (3.2)$$

où $\epsilon_0(f)$ est l'énergie élémentaire en sortie de la source et α est le coefficient d'atténuation fréquentielle selon Thorp décrit dans l'équation 2.25 et représenté en figure 2.6. Alors l'énergie totale en dB s'exprime comme l'intégrale de toutes les contributions :

$$E_{dB}(r) \triangleq \int \epsilon(r, f) df. \quad (3.3)$$

Nous supposons qu'il est possible d'approximer l'équation 3.3 à partir du signal mesuré par un passage en décibel de l'équation 3.1. On a alors :

$$E_{dB} = 10 \log_{10} \left(\frac{E}{E_{ref}} \right), \quad (3.4)$$

où E_{ref} est le niveau d'énergie de référence. Puis, nous procédons de la manière suivante : dans l'équation 3.3, nous remplaçons i) Le terme de gauche par sa forme intégrale et ii) le terme de droite par l'expression donnée par l'équation 3.1. Nous obtenons donc :

$$\begin{aligned} &\int \left(\epsilon_0(f) - 20 \log_{10}(r) - \alpha(f)r \right) df \\ &= 10 \log_{10} \left(\frac{\int |X(f; r)|^2 df}{E_{ref}} \right). \end{aligned} \quad (3.5)$$

En pratique l'équation 3.5 n'est pas solvable car le terme ϵ_0 est inconnu. Nous proposons alors de mesurer le ratio des énergies de 2 bandes de fréquences, ce qui est équivalent à une différence en dB . La relation donnée par l'équation 3.5 peut être étendue à une bande de fréquence. Nous réécrivons alors l'équation 3.5 pour une bande de fréquence $B_1 = [F_1, F_2]$ telle que $F_1 < F_2$. L'expression de l'énergie en décibel de cette bande est donnée ci-dessous :

$$\begin{aligned}
 & \int_{F_1}^{F_2} \left(\epsilon_0(f) - 20 \log_{10}(r) - \alpha(f)r \right) df \\
 &= 10 \log \left(\frac{\int_{F_1}^{F_2} |X(f; r)|^2 df}{E_{ref}} \right).
 \end{aligned} \tag{3.6}$$

Introduisons une seconde bande de fréquence $B_2 = [F'_1, F'_2]$ avec $F_1 < F_2 < F'_1 < F'_2$ sans recouvrement entre les bandes. En écrivant Eq. 3.6 pour B_2 et en considérant la différence entre les 2 bandes d'énergie, nous obtenons :

$$\begin{aligned}
 & \int_{F_1}^{F_2} \left(\epsilon_0(f) - 20 \log_{10}(r) - \alpha(f)r \right) df \\
 & - \int_{F'_1}^{F'_2} \left(\epsilon_0(f) - 20 \log_{10}(r) - \alpha(f)r \right) df \\
 &= 10 \log_{10} \left(\frac{E_1}{E_2} \right),
 \end{aligned} \tag{3.7}$$

où $\frac{E_1}{E_2}$ est le ratio des énergies intégrées sur leur bande respective. De manière à simplifier l'équation 3.7, nous proposons 2 hypothèses :

- hypothèse 1 (hyp.1) : nous choisissons arbitrairement des bandes de taille égale $\Delta B_1 = F_2 - F_1$ et $\Delta B_2 = F'_2 - F'_1$, ce qui permet d'agir sur les bornes d'intégration. Nous obtenons alors :

$$\begin{aligned}
 (F_2 - F_1) &= (F'_2 - F'_1) \\
 \int_{F_1}^{F_2} 20 \log_{10}(r) df &= \int_{F'_1}^{F'_2} 20 \log_{10}(r) df.
 \end{aligned} \tag{3.8}$$

- hypothèse 2 (hyp.2) : pour des signaux d'odontocètes de type clic, de durée brève, large-bandes, nous faisons l'approximation que l'énergie en sortie de l'animal est constante pour toutes les fréquences. Le spectre est donc plat d'amplitude ϵ_0 . Les clics sont alors modélisés par des sources impulsives idéales (Dirac). Cette hypothèse permet de faire disparaître le terme dépendant du SL dans l'équation 3.7 car :

$$\int_{F_1}^{F_2} \epsilon_0(f) df = \int_{F'_1}^{F'_2} \epsilon_0(f) df.$$

Sous la considération de hyp.1 et hyp.2, l'équation 3.7 peut être réduite à :

$$r \left(\int_{F'_1}^{F'_2} \left(\alpha(f) \right) df - \int_{F_1}^{F_2} \left(\alpha(f) \right) df \right) = 10 \log_{10} \left(\frac{E_1}{E_2} \right). \tag{3.9}$$

Nous obtenons alors une expression théorique de la distance de propagation r en m (relation *ISAT* :

$$r(B_1, B_2) = \frac{10 \log_{10} \left(\frac{E_1}{E_2} \right)}{\int_{F_1'}^{F_2'} \alpha(f) df - \int_{F_1}^{F_2} \alpha(f) df}. \quad (3.10)$$

Dans cette expression, r ne dépend ni des pertes par divergence, ni du niveau d'énergie à la source mais uniquement de l'atténuation fréquentielle. Nous espérons observer un effet significatif compte tenu des faibles ordres de grandeur de ce coefficient pour des fréquences moyennes (0.05 dB/km pour 1 kHz [Leroy 65, Thorp 67]).

La figure 3.1 illustre les conséquences de la propagation sur un spectre de signal synthétique idéal de type impulsion de Dirac selon un modèle de perte par divergence sphérique et d'atténuation fréquentielle (Thorp). L'énergie totale est affectée mais le spectre subit une déformation. Les hautes fréquences s'atténuent plus fortement que les basses fréquences pouvant évoquer l'équivalent d'un comportement de filtre passe-bas généré par ce modèle de propagation acoustique. En accord avec les références bibliographiques, le niveau d'énergie en sortie de l'animal (à 1 m) a été fixé à 220 dB.

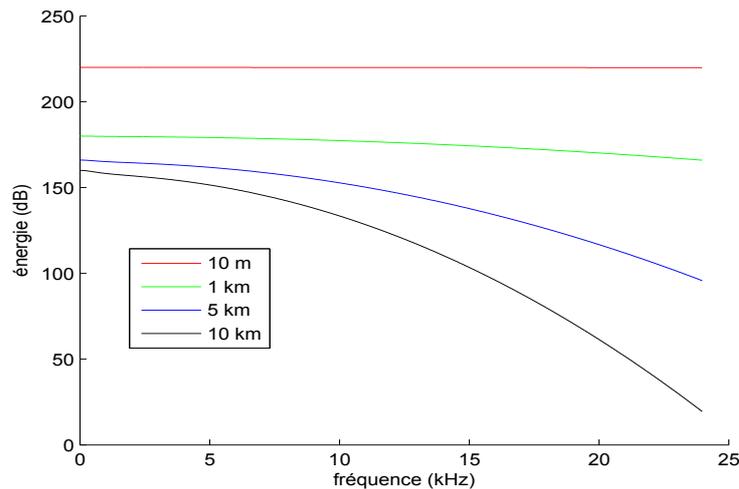


FIGURE 3.1 – Illustration de la déformation spectrale selon la distance de propagation sur un signal idéal de type Dirac, de 1 m à 10 km de haut en bas. La résolution en fréquence est de $\delta f = 24$ Hz.

3.2 Représentation de l'estimateur distance ISAT et paramètres importants

L'objectif est d'utiliser la relation *ISAT* sur des signaux réels de cachalot. Si l'hypothèse hyp.2 est une simplification extrême, elle nous permet de passer outre notre manque d'information sur la source acoustique, c'est-à-dire la forme réelle ou le profil spécifique du spectre en sortie. Le spectre des clics de cachalot a été analysé, montrant son caractère complexe et coloré dû notamment à la forte directivité de certains pulses [Zimmer 05, Madsen 02, Mohl 03]. Si cette information était

disponible, l'équation 3.10 pourrait être agrémentée d'un terme correspondant à la différence d'énergie entre les 2 bandes d'énergie, avant propagation. Cette donnée est autrement dit la forme du spectre initial. L'équation (3.10) peut être réécrite ainsi :

$$r_{\epsilon_0}(B_1, B_2) = \frac{10 \log_{10} \left(\frac{E_1}{E_2} \right) + \int_{F'_1}^{F'_2} \epsilon_0(f) df - \int_{F_1}^{F_2} \epsilon_0(f) df}{\int_{F'_1}^{F'_2} \alpha(f) df - \int_{F_1}^{F_2} \alpha(f) df} . \quad (3.11)$$

La représentation de la distance r en fonction des valeurs du ratio $\frac{E_1}{E_2}$, selon le modèle simplifié, est donnée figure 3.2 pour différentes fréquences centrales de bande $F_1^c = \frac{F_2 - F_1}{2}$ et $F_2^c = \frac{F'_2 - F'_1}{2}$ associées respectivement aux bandes B_1 et B_2 . Les estimations de r définies par la relation ISAT, sont dépendantes du processus de génération de bandes. Trois paramètres importants méritent d'être considérés : la largeur de bande, l'Intervalle Inter-Bandes (IIB) et des fréquences centrales considérées.

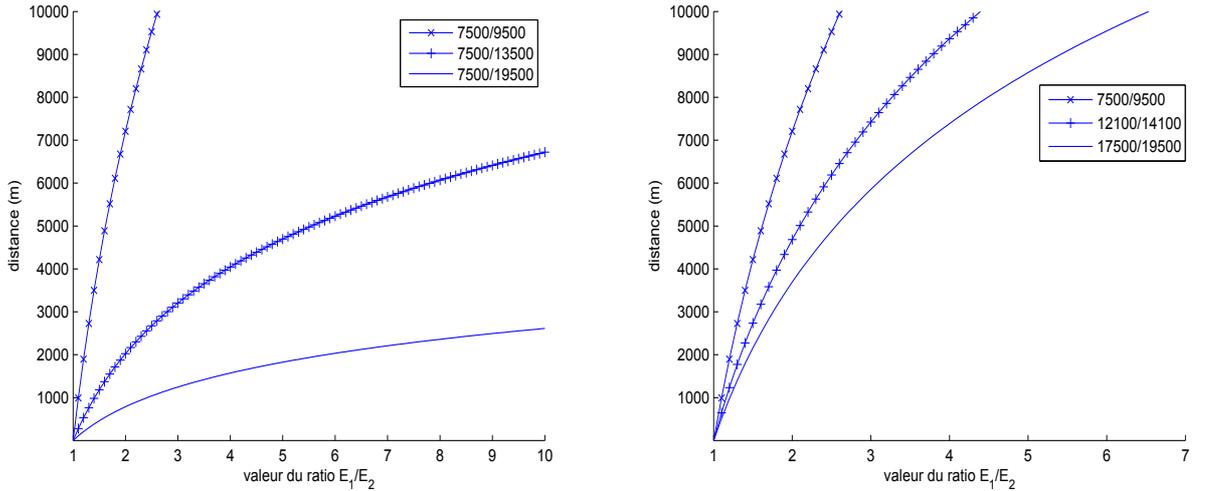


FIGURE 3.2 – Représentation de la relation ISAT pour différentes paires de fréquences centrales (F_1^c/F_2^c) (en Hz) des bandes B_1 et B_2 du ratio $\frac{E_1}{E_2}$. La taille des bandes est fixée à $\Delta B = 200$ Hz. A gauche : IIB variable, à droite F_1^c/F_2^c variable.

Le comportement des courbes (figure 3.2) est conditionné par le terme :

$$\left(\int_{F'_1}^{F'_2} \alpha(f) df - \int_{F_1}^{F_2} \alpha(f) df \right)^{-1} . \quad (3.12)$$

La figure 3.2 montre l'impact de l'IIB (à gauche) et du choix de F_1^c/F_2^c (à droite). Lorsque l'IIB est élevé, la courbe s'aplatit à partir de faibles valeurs de ratio. L'estimateur de distance qui en résulte devient peu sensible aux variations de ratio et conduirait à des prédictions similaires

pour des ratios mesurés différents. En choisissant de faibles valeurs d'IIB, nous générons des estimateurs robustes pour des distances comprises entre 0 et 15 km. Les courbes présentées en figure 3.2 illustrent bien cet effet. La courbe paramétrée par le plus grand IIB paraît constante. A IIB fixe, le choix des fréquences centrales mises en jeu dans le ratio semble impacter de manière moindre l'allure des courbes. Les courbes générées montrent des caractéristiques du même ordre.

3.3 Architecture et mise œuvre de l'estimation ISAT

3.3.1 De l'échelle d'un clic à l'échelle d'un pulse

Selon le modèle de production sonore "bent horn" [Zimmer 04], un clic de cachalot est composé de son impulsion initiale (P0) et de ses répliques (P1, P2, PF1, PF2, ..., PJ1, PJ2, ...) telle que le présente la figure 2.8. Les clics présentent ainsi une structure temporelle complexe dite "multipulsée" [Laplanche 05] résultant de réflexions multiples et de différents chemins acoustiques dans la tête de l'animal. La directivité des pulses, et finalement du clic généré, joue un rôle important dans la forme du signal reçu [Laplanche 05].

Pourtant, l'hypothèse hyp.2 implique de travailler sur un signal similaire à une impulsion de Dirac, ou tout du moins large-bande sur les bandes B_1 et B_2 considérées. Aussi, nous choisissons de focaliser nos calculs sur un pulse du clic, plus proche d'une impulsion de type Dirac. C'est pour cela que nous considérons uniquement le pulse le plus énergétique. Cela est arbitraire, car l'énergie reçue à l'hydrophone dépend de l'orientation de l'axe cachalot-hydrophone du fait de la directivité. La durée de cet événement est assez courte, environ 5 ms, au regard de la durée d'un clic complet, plus de 20 ms.

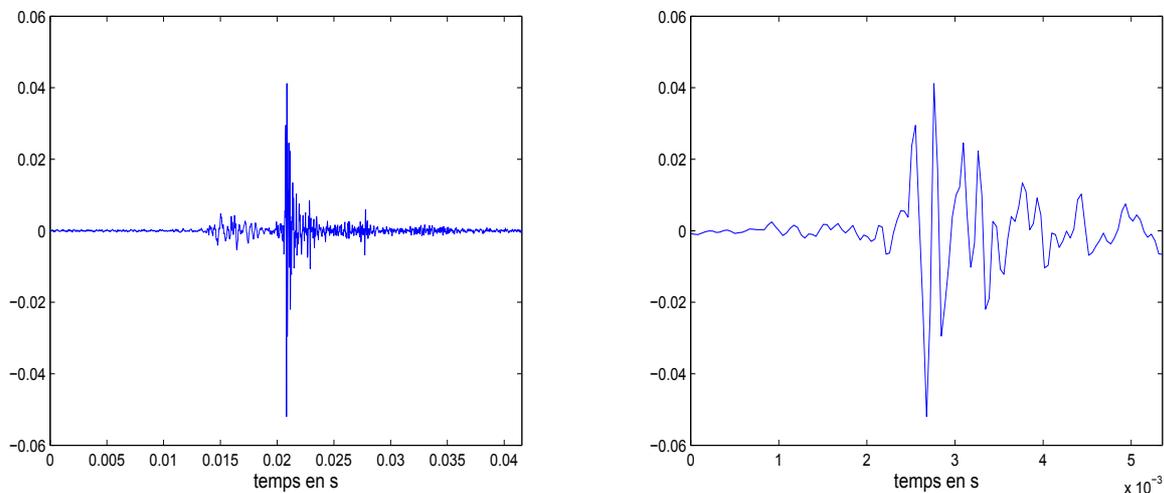


FIGURE 3.3 – Clic moyen dans le jeu de données BAHAMAS. Deux échelles différentes de calcul, centrées sur le pulse le plus énergétique.

Sur la partie gauche de la figure 3.3, nous représentons sur une fenêtre d'observation de 40 ms, la forme temporelle moyenne des clics du jeu de données BAHAMAS, échantillonnés à 48 kHz faisant apparaître la forme typique multipulsée. La partie de droite correspond à une échelle intermédiaire de 5.3 ms (soit 256 échantillons) centrée sur le pulse le plus énergétique et incluant le second pulse énergétique du clic. L'échelle d'observation la plus courte utilisée pour le développement d'ISAT est de 2.7 ms (soit 128 échantillons) centrée sur le pulse le plus

énergétique. A ce stade, nous conserverons ces 3 échelles d'observation, nous n'en retiendrons qu'une *a posteriori*.

Segmenter automatiquement un pulse dans un clic spécifique n'est pas une tâche aisée. Une identification exhaustive des pulses est une procédure complexe due : i) à un fort niveau de bruit de mer, ii) à la présence d'effet de fenêtrage, iii) à de possibles permutations du pulse le plus énergétique selon l'angle de visé de l'animal par rapport à l'hydrophone [Zimmer 04, Laplanche 05], celui-ci variant au cours du déplacement de l'animal [Le Cadre 97]. De plus, les configurations "on axis" et "off axis" conduisent à des structures de clic différentes [Zimmer 05]. Elles sont illustrées en figure 2.8. Par conséquent, nous sélectionnons un pulse dont la nature et les caractéristiques spectrales diffèrent d'un clic à l'autre. Nous nous attendons à ce que l'ensemble de ces facteurs impacte fortement les estimations de distance ISAT. Pour la phase de développement nous conserverons les échelles $N = 128$ et $N = 256$ échantillons pour le calcul de ISAT.

3.3.2 Dans le domaine fréquentiel

Comme nous l'avons indiqué précédemment, nous devons calculer le spectre des pulses. En se référant au spectre moyen (coté gauche de la figure 3.4), l'intervalle de fréquences compris entre 5 kHz et 7 kHz ne semble pas adéquat pour l'application de la méthode, étant donné les fortes modulations risquant de contredire nos hypothèses de départ. Aussi, de manière à éviter de travailler sur les basses fréquences polluées par le bruit de mer ambiant (détaillé dans l'état de l'art), nous considérons alors toutes les fréquences entre 7 kHz et 24 kHz (pour une fréquence d'échantillonnage $F_s = 48$ kHz).

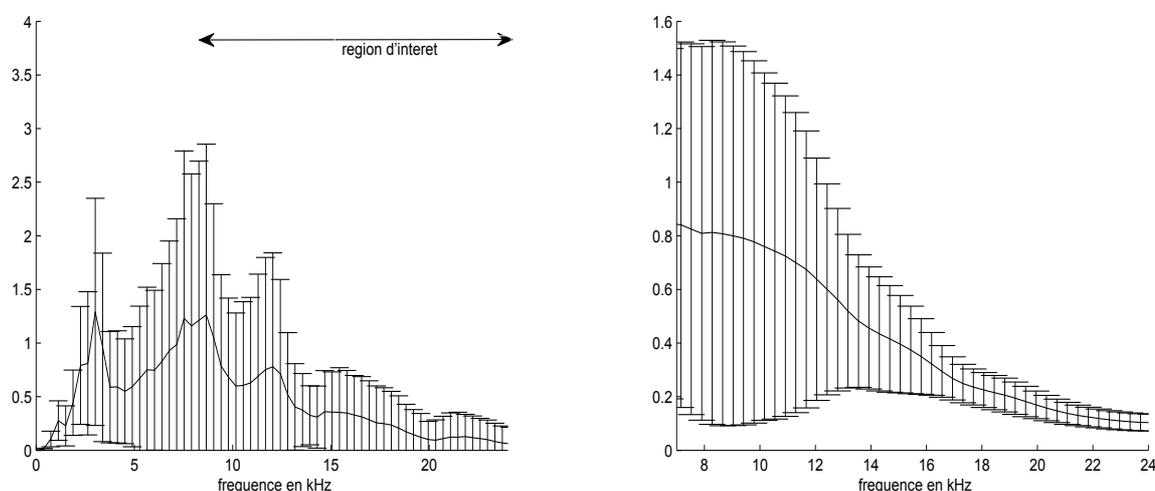


FIGURE 3.4 – Extraction du spectre moyen du pulse le plus énergétique et de son écart type (filtré par un passe haut $f_c = 2$ kHz) à partir du jeu de données BAHAMAS pour l'hydrophone H11 et les effets résultant de l'étape de lissage. A gauche : le spectre original du pulse et la région d'intérêt (entre 7 kHz et 22 kHz). A droite : la région d'intérêt du spectre après lissage.

Toutefois, sur cette région, les spectres montrent d'importantes variations locales en désaccord avec notre approximation large bande. Le modèle ne prend pas en compte la forme précise du spectre en sortie du cachalot. Pour concorder avec le modèle, nous devons transformer les spectres de manière à rendre leur forme plus monotone. Nous effectuons donc un lissage des spectres sur

la bande considérée par le biais d'un simple filtre de moyenne sur $2P+1$ points avec la contrainte $1 < 2P+1 < \frac{N}{2}$, N étant le nombre d'échantillons du signal. Une discussion sur le choix de P sera abordé.

Sur la partie gauche de la figure 3.4, nous avons calculé le spectre moyen (moyenne sur l'ensemble des clics détectés dans le jeu de données BAHAMAS [Doh 13b, Doh 13a]) pour $N = 256$ échantillons. Un filtre passe-haut a été utilisé pour la détection des clics, expliquant le "vide" avant 2 kHz. Sur la partie droite est affiché le spectre, sur le domaine de fréquence retenu pour le calcul, après lissage de ce dernier. La transformation a affecté de manière significative la forme originale du spectre mais conserve la pente moyenne du spectre.

3.3.3 Estimations de distance ISAT pour un choix de bandes fixe

Sur une période d'analyse de durée T , pour une fenêtre W , nous définissons un estimateur de la distance de propagation selon l'équation 3.10. L'énergie E_1 et E_2 des bandes B_1 and B_2 est calculée à partir du module des spectres des pulses (FFT). L'espace des fréquences est donc discrétisé selon $f_k \triangleq k\Delta F$ où $\Delta F = \frac{F_s}{N}$ et $k = 0, 1, \dots, \frac{N}{2} - 1$. Les fréquences positives sont décrites par les $\frac{N}{2}$ premiers points du spectre obtenu par FFT.

Un estimateur brut $r(B_1, B_2)$ pour chaque clic détecté $j = 1, \dots, N_W$ sur la période T peut être calculé par :

$$\hat{r}_j = \frac{10 \log_{10} \left(\frac{\hat{E}_{1,j}}{\hat{E}_{2,j}} \right)}{\sum_{k=K'_1}^{K'_2} \alpha(k\Delta F) - \sum_{k=K_1}^{K_2} \alpha(k\Delta F)}, \quad (3.13)$$

où

$$\hat{E}_{1,j} = \sum_{k=K_1}^{K_2} |\hat{X}_j(k\Delta F)|^2,$$

$$\hat{E}_{2,j} = \sum_{k=K'_1}^{K'_2} |\hat{X}_j(k\Delta F)|^2.$$

où K_1, K_2, K'_1 et K'_2 sont les indices des extrémités des bandes B_1 et B_2 respectivement. $\hat{X}_j(k\Delta F)$ est la valeur discrète du spectre associée au $j^{i\text{ème}}$ clic détecté. Dans le jeu de données, le nombre de clics N_W , sur 1 min est borné par $N_{clickmax} = 60$.

3.3.4 Adaptation à l'ensemble du domaine spectral par multi-combinaisons de bandes

Jusqu'à maintenant, le choix des bandes était déterminé manuellement, arbitrairement et demeurait fixe pour toute l'analyse [Glotin 11a, Doh 11b]. Cependant, au regard de la variabilité des clics dans le domaine fréquentiel, expliquée précédemment par les variations des angles de position de l'animal [Mohl 00, Laplanche 05], il semble judicieux d'adapter et de réactualiser le choix de bande au cours du temps. Pour suivre cette approche, nous considérons plusieurs combinaisons de bandes B_1, B_2 (*i.e.* plusieurs paires de bandes pour le calcul des ratios).

Une combinaison de bandes $\theta \in N_{\text{band}} \subset N^+$ est définie par 2 paires de fréquences $[F_{1,\theta}, F_{2,\theta}]$ et $[F'_{1,\theta}, F'_{2,\theta}]$ telles que respectivement $F_{1,\theta} < F_{2,\theta} < F'_{1,\theta} < F'_{2,\theta}$. En admettant que i) nous n'avons aucune connaissances *a priori* de la région d'intérêt dans le spectre, ii) Pour un ΔB fixe, nous pouvons générer toutes les combinaisons possible $B_{1,\theta}, B_{2,\theta}$ entre 7 et 22 kHz. Nous fixons la largeur des bandes à $\Delta B = 500$ Hz. Selon cette configuration, nous devons générer $N_{\text{band}} = \binom{30}{2} = 435$ combinaisons de bandes possibles (sans recouvrement). Pour une combinaison donnée θ (*i.e.* pour les bandes $(B_{1,\theta}, B_{2,\theta})$), l'équation. 3.13 peut être réécrite ainsi :

$$\hat{r}_{\theta,j} = \frac{10 \log_{10} \left(\frac{\hat{E}_{1,\theta,j}}{\hat{E}_{2,\theta,j}} \right)}{\sum_{k=K_{1,\theta}}^{K_{2,\theta}} \alpha(k\Delta F) - \sum_{k=K'_{1,\theta}}^{K'_{2,\theta}} \alpha(k\Delta F)}, \quad (3.14)$$

où

$$\hat{E}_{1,\theta,j} = \sum_{k=K_{1,\theta}}^{K_{2,\theta}} |\hat{X}_j(k\Delta F)|^2,$$

et

$$\hat{E}_{2,\theta,j} = \sum_{k=K'_{1,\theta}}^{K'_{2,\theta}} |\hat{X}_j(k\Delta F)|^2,$$

où $K_{1,\theta}, K_{2,\theta}, K'_{1,\theta}$ et $K'_{2,\theta}$ sont les indices de fréquences définissant l'intervalle des bandes $B_{1,\theta}$ et $B_{2,\theta}$ respectivement.

3.3.5 Filtrage des estimations sur des fenêtres d'observations de courtes durées

Dans la section précédente, nous avons établi l'expression d'estimateurs de distance locaux (pour chaque clic) par $\hat{r}_{\theta,j}$, $\theta \in N_{\text{band}} \subset N^+$ associés au $j^{\text{ième}}$ clic détecté à la date $t_j \in [0, T]$. Chaque combinaison conduit à une distance estimée. L'objectif de cette étape est de retenir les distances estimées les plus cohérentes. Une première tentative consisterait à chercher cette ou ces combinaisons fiables pour la totalité de la durée du signal T . Or, il n'y a aucune raison de maintenir ce choix de combinaison θ constant sur T , compte tenu de la variabilité des spectres dans le temps.

Nous décidons alors de procéder à une analyse par fenêtres de courte durée sans recouvrement. L'enregistrement est donc divisé en L portions de durées égales telles que $T = LT^W$. La durée T^W de la fenêtre d'analyse reste un sujet ouvert. Il s'agit d'établir un compromis entre des variations acceptables du spectre et le nombre de clics impliqués sur cette durée. Parmi les multiples facteurs de variabilité, la directivité du bio-sonar est souvent citée [Au 08, Zimmer 04]. Du fait de la spécificité de la nage du cachalot (lent, trajectoire rectiligne et prédictible), nous considérons que les variations de positions et d'angles de l'animal sont faibles sur une durée de 1 minute, ce qui correspond en moyenne à 60 clics (environ 1 clic/seconde).

3.3.6 Détermination des combinaisons de bandes optimales à partir de la vérité terrain

Nous obtenons donc des estimations de distances $\{\hat{r}_{\theta,j}^l\}$ grâce à l'équation 3.14 pour L fenêtres temporelles non recouvrante W^l indexées par $l = 1, \dots, L$ et de durée T^W . L'objectif est de sélectionner une unique combinaison de bandes θ^l pour chaque W^l .

Dans le cas idéal où la vérité terrain $\{d_j^l\}_{j=1, \dots, N_{W^l}}$ est disponible, il est possible d'optimiser le choix de θ^l directement par minimisation d'une fonction de coût particulière (sans régularisation) sur les N_{W^l} clics détectés. Plusieurs types de fonction coût peuvent être investiguées, elles sont notées ℓ . En pratique on utilise les normes de type ℓ_2 ou ℓ_1 .

$$\hat{\theta}^{\star,l} = \arg \min_{\theta^l \in \Theta} \left\{ \sum_{j=1}^{N_{W^l}} \ell(d_j^l, \hat{r}_{\theta^l,j}^l) \right\}. \quad (3.15)$$

Par la suite, l'équation 3.15 est optimisée avec $\ell_1(x, y) \triangleq |x - y|$ étant donnée $\Theta = N_{band} \subset \mathbb{N}^+$ avec $N_{band} = 435$.

Pour chaque fenêtre W^l , nous obtenons une estimation de distance en résolvant l'équation 3.15 et nous définissons la séquence $\{\hat{r}_{\theta^{\star,l},j^l}^l\}$ avec j^l tel que $\forall t_{j^l} \in [(l-1)T^W, lT^W[$. Nous retenons les distances optimales obtenues pour chaque fenêtre $\{W^l\}$. Sur la durée totale T , la séquence de distance optimale \hat{r}^* minimisant la fonction de coût de chaque fenêtre W^l , est défini par :

$$\hat{r}^* = \bigcup_{l=1}^L \bigcup_{j^l=1}^{N_{W^l}} \{\hat{r}_{\theta^{\star,l},j^l}^l\}, \quad j^l \text{ tel que } \forall t_{j^l} \in [(l-1)T^W, lT^W[. \quad (3.16)$$

3.3.7 Sélection des combinaisons de bandes en conditions réelles

En conditions réelles, la vérité terrain n'est généralement pas disponible et l'intérêt est justement d'estimer la distance entre l'animal et l'hydrophone. Nous devons optimiser le choix des combinaisons indépendamment des $\{d_j^l\}$. Sur une durée d'une minute, la variation de distance radiale ne devrait pas excéder 100 mètres au regard de la dynamique de l'animal (environ 1 m/s). On admet donc une faible dispersion des estimations de distance sur une minute. Notre critère de minimisation est la variance des estimations sur la durée de W^l . Une forte dispersion peut résulter d'un rapport de bandes instable dans le temps.

Donc, pour une fenêtre donnée, nous proposons de calculer \hat{s}_θ^l , l'écart-type dans le temps des estimations locales $\{\hat{r}_{\theta,j^l}^l\}$ résultantes de chaque combinaison de bandes θ . On définit σ l'opérateur écart-type dans le temps :

$$\hat{s}_\theta^l = \sigma(\hat{r}_{\theta,j^l}^l), \quad j^l \text{ tel que } \forall t_{j^l} \in [(l-1)T^W, lT^W[. \quad (3.17)$$

Puis nous procédons à une étape de tri des valeurs \hat{s}_θ^l par ordre de variance ascendante. soit $m = 1, 2, \dots, N_{band}$ le rang d'une combinaison θ_m^l après le tri. Nous calculons l'estimateur suivant :

$$\hat{r}_{j^l}^l = \frac{1}{N_{best}} \sum_{m=1}^{N_{best}} \hat{r}_{\theta_m^l,j^l}^l, \quad j^l \text{ tel que } \forall t_{j^l} \in [(l-1)T^W, lT^W[, \quad N_{best} = 1, 2, 8, 16, \dots, N_{band}. \quad (3.18)$$

Seuls $N_{best} \ll N_{band}$ combinaisons présentant les plus faibles variances contribuent à l'estimation. En combinant les résultats de chaque fenêtre d'analyse, nous construisons sur l'ensemble de l'enregistrement l'estimation finale ISAT basée sur une minimisation de la variance.

$$\hat{\mathbf{r}} = \bigcup_{l=1}^L \bigcup_{j^l=1}^{N_{W^l}} \{\hat{r}_{j^l}^l\}, \quad j^l \text{ tel que } \forall t_{j^l} \in [(l-1)T^W, lT^W]. \quad (3.19)$$

On obtient ainsi une estimation par clic détecté.

3.4 Conclusion

Dans ce chapitre, nous proposons un modèle d'estimation de distance basé sur les lois d'atténuation acoustique. Ce modèle tente de relier la déformation du spectre du signal reçu à la distance de propagation (source-récepteur). Cette déformation est évaluée par le biais de rapports d'énergie entre de 2 bandes fréquentielles du signal. Nous donnons donc les éléments permettant d'établir l'expression théorique de cet estimateur appelé "Intra Spectral Attenuation" (ISAT).

Puis nous adaptons le calcul des estimations à un processus informatique et donc numérique. Résumons ici les différentes étapes du calcul ISAT.

- sélection d'un pulse pour chaque clic,
- calcul de spectres *via* transformation de Fourier (FFT) puis lissage des spectres selon le paramètre P ,
- création des bandes de fréquences entre 7 kHz et 22 kHz. Largeur de bande fixée à $B = 500$ Hz pour générer N_{band} ratios,
- estimation d'une distance par clic et par combinaison $\hat{r}_{\theta^l, j^l}^l$,
- réduction le long de la dimension des combinaisons pouvant être effectuée par fenêtre d'analyse de courte durée (1 min) si besoin. Nous proposons 2 méthodes :
 - i) si la vérité terrain est disponible, par minimisation d'une quantité basée sur une fonction de coût de type ℓ_1 donnée par l'équation (3.15),
 - ii) si la vérité terrain n'est pas disponible (cas pratique), par minimisation de la variance des estimations locales. Les combinaisons sont classées par ordre de variance croissante et intégrées aux estimations selon l'équation (3.18) et le paramètre N_{best} .

Prédiction de distance par réseau de neurones

Sommaire

4.1 Rappels sur les perceptrons multicouches	49
4.1.1 Le perceptron	49
4.1.2 Règle d'apprentissage du perceptron	50
4.1.3 Le perceptron multi-couches	50
4.2 Définition des modèles	51
4.3 Conclusion	54

4.1 Rappels sur les perceptrons multicouches

4.1.1 Le perceptron

Les réseaux neuromimétiques sont basés sur le principe d'apprentissage, qui consiste à minimiser un critère d'erreur entre des sorties calculées et des sorties désirées. Les neurones ont au moins une entrée x à p dimensions pondérée par des coefficients W , appelés poids, et une sortie y à n dimensions, qui est fonction de la somme pondérée des entrées. Pour un élément k de cette sortie tel que $1 < k < n$, on a :

$$y_k = \sum_{i=1}^p W_{ki}x_i, \quad (4.1)$$

Les perceptrons multi-couches ont une architecture particulière, avec une organisation des neurones en couche cachée. Un ou plusieurs neurones sont regroupés en couche. Les réseaux ont donc au moins une couche d'entrée et une couche de sortie, auxquelles on ajoute une ou plusieurs couches cachées, placées entre les couches d'entrée et de sortie. La couche d'entrée accueille les données et la couche de sortie fournit les valeurs calculées par le réseau. La ou les couches cachées organisent l'information pertinente à travers le réseau. La fonction neuronale doit être définie, continue et dérivable. Elle peut être linéaire ou non-linéaire, suivant le rôle que l'on veut faire jouer au neurone : originalement, cette fonction était linéaire pour la présentation d'une valeur de sortie dans un intervalle donnée, de type sigmoïde pour réaliser une classification binaire. W est donc la matrice de poids $n \times p$ qui, mise à jour régulièrement à la présentation des exemples

contenus dans la base d'apprentissage, permet de faire converger l'état des sorties des neurones vers les valeurs souhaitées.

4.1.2 Règle d'apprentissage du perceptron

L'algorithme de la rétro-propagation du gradient est basé sur une fonction de coût de l'erreur : il s'agit généralement de la valeur instantanée ou moyenne de l'erreur quadratique sur un, plusieurs ou tous les exemples de la base d'apprentissage. En minimisant cette fonction, les valeurs des poids du réseau de neurones sont mises à jour.

Partons du cas le plus simple *i.e.* à une couche de neurones [Rosenblatt 60] dont l'opérateur de type échelon sera noté σ . La réponse du neurone s'écrit alors $\sigma(p - \theta)$ où p est un potentiel et θ un seuil. Pour un pas d'apprentissage donné, l'erreur quadratique entre la sortie obtenue et la sortie "vraie" s'exprime simplement par : $E = (y - y^*)^2$. Minimiser ce terme revient à calculer le gradient de l'erreur selon les poids W_{ki} .

$$\frac{\partial E}{\partial W_k} = 2(y - y^*) \frac{\partial y}{\partial W_k} \quad (4.2)$$

Or $y = \sum_j W_j x_j$, donc $\frac{\partial y}{\partial W_k} = x_k$. Le gradient de l'erreur vaut donc :

$$\frac{\partial E}{\partial W_k} = 2(y - y^*) x_k \quad (4.3)$$

La mise à jour des poids peut se faire à chaque itération (c'est-à-dire à chaque présentation d'un exemple de la base d'apprentissage) ou après plusieurs exemples voire tous les exemples compris dans la base d'apprentissage. Dans notre application, pour des raisons de temps de calcul, parce que les exemples de la base de données peuvent avoir des différences importantes ou pour privilégier le caractère adaptatif, la mise à jour s'effectue à chaque itération. Nous nous intéressons à la convergence des sorties vers les sorties désirées, en fonction de la valeur du gain, constant ou éventuellement dont la valeur est ajustée itérativement.

4.1.3 Le perceptron multi-couches

Pour un perceptron multicouches, la réponse du neurone à un potentiel p est de type sigmoïde et peut s'adapter au type de sortie voulue [Minski 69, Lecun 85]. On définit :

$$y = \sigma(p - \theta) = \frac{1}{1 + \exp(-k(p - \theta))} \quad (4.4)$$

Etablissons l'expression du gradient de l'erreur quadratique dans ce cas. D'abord réécrivons la sortie d'un neurone i en fonction de la matrice de poids W :

$$y = \sigma(p_i - \theta) = \sigma\left(\sum_{j=1}^p W_{ij} x_j - \theta\right) \quad (4.5)$$

L'erreur quadratique associée à l'ensemble du réseau est la somme sur tout les neurones des erreurs obtenues.

$$E = \frac{1}{2} \sum_{j=1}^n (y_j - y_j^*)^2 \quad (4.6)$$

Le gradient de l'erreur vaut donc :

$$\frac{\partial E}{\partial W_{ik}} = \sum_{j=1}^n (y_j - y_j^*) \frac{\partial (y_j - y_j^*)}{\partial W_{ik}} \quad (4.7)$$

Notons que le terme $\frac{\partial (y_j - y_j^*)}{\partial W_{ik}}$ est non nul uniquement pour $i = j$. Alors l'expression du gradient permettant de réévaluer les poids à chaque itération est :

$$\frac{\partial E}{\partial W_{ik}} = (y_j - y_j^*) \frac{\partial (y_i)}{\partial W_{ik}} = (y_j - y_j^*) x_k \sigma(p_i - \theta)' \quad (4.8)$$

4.2 Définition des modèles

Le problème d'estimation de la distance source-récepteur r que l'équation sonar tente de résoudre est difficilement solvable sans un modèle théorique précis ou un modèle statistique du niveau d'énergie à la source. Pourtant, ce cas peut être vu comme un problème de régression $f(e) = r$ sur e l'énergie reçue. Lorsqu'il y a possibilité de réunir la vérité terrain et les données, cette fonction f peut être approximée par apprentissage en utilisant un réseau de neurones [Duda 00]. Les Perceptrons Multi Couches (MLP) paraissent être de bons candidats pour apprendre des relations complexes que la physique peine à formaliser car ce sont des régresseurs universels [Cybenko 89, Hornik 91]. Le principal défaut de ce type de régression, est que la fonction approximée est empirique et dépend fortement des données d'apprentissage. Est-il possible d'apprendre une fonction de régression sur une base de données valide sur des données issues d'un autre contexte acoustique? La problématique de généralisation est un aspect usuel des réseaux de neurones.

Plusieurs modèles ont été créés afin d'étudier quelles informations spectrales sont le plus reliées à la distance source-récepteur. Il s'agit de tester différents descripteurs en entrée du MLP pour une même sortie : la distance radiale vérité terrain r^* normalisée, c'est-à-dire dont les valeurs sont comprises entre 0 et 1. Une fonction logarithmique peut être appliquée sur l'espace d'entrée de manière à compresser la distribution des éléments d'entrée. Le MLP constitué de 2 couches cachées permet d'attribuer des poids aux éléments de l'entrée décrivant au mieux la régression. Nous optimiserons 3 paramètres : le nombre d'itérations lors de la phase d'apprentissage, le nombre d'unités de la couche cachée et le taux d'apprentissage. Les modèles réalisés sont les suivant (ils sont récapitulés en figure 4.2) :

- Le modèle LER (Learned Energie Regression) est basé sur l'échelle la plus globale en terme d'information spectrale. Il s'agit d'une régression de l'énergie totale du clic e .

$$\hat{r} = h(e; \mathbf{W}), \quad (4.9)$$

où \mathbf{W} représente la matrice des poids déterminés par le MLP lors de la phase d'apprentissage.

En reprenant l'équation 4.8 dans le cas où l'entrée et la sortie sont des variables scalaires et où seul le nombre de neurones des couches cachées peut être supérieur à 1. Nous aboutissons à l'équation permettant de mettre à jour W à chaque itération :

$$\delta W_i = -\alpha \delta_i e \quad (4.10)$$

avec $\delta_i = (\hat{r} - r^*) \sigma(p_i - \theta)'$ où α est un gain adaptatif, \hat{r} est l'estimation de distance et r^* est la distance fournie par la vérité terrain.

- A l'échelle intra-clic, le modèle ISAR (Intra Spectral Attenuation Regression) est une régression des composantes du spectre du clic j $|X_j(k\Delta F)|$.

$$\hat{r} = h(|X(f; r)|; \mathbf{W}), \quad (4.11)$$

Dans ce cas l'entrée est un vecteur à autant de dimensions que de canaux fréquentiels 128, 256,... La sortie reste un scalaire. L'équation de mise à jour des poids devient :

$$\delta W_{ik} = -\alpha \delta_i x_k \quad (4.12)$$

Une représentation schématisée de ce réseau de neurone est donnée en figure 4.1. $X(k_1), \dots, X(k_n)$ sont les éléments du module du spectre appliqués en entrée du MLP et r la distance source-émetteur.

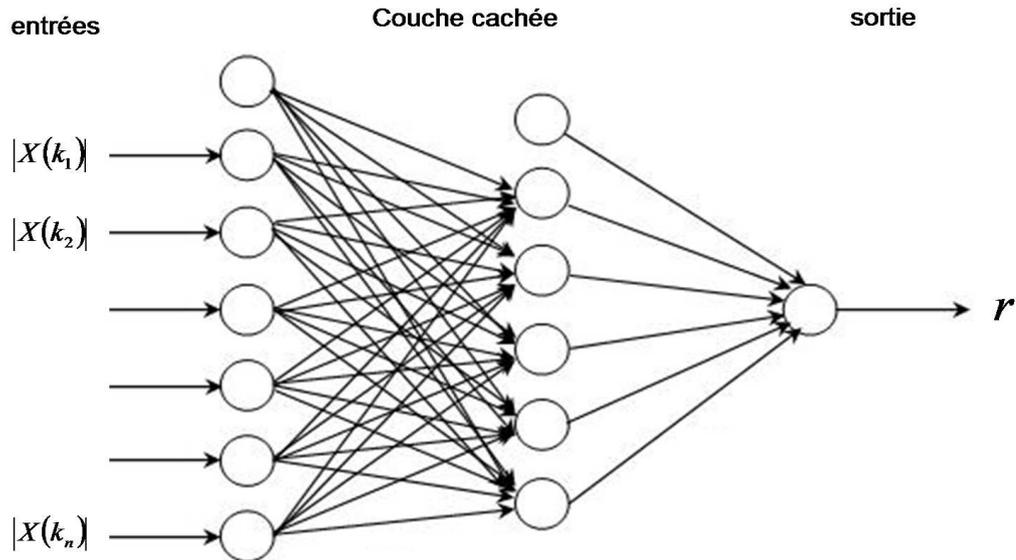


FIGURE 4.1 – Schématisation du modèle ISAR.

Si l'on applique la fonction logarithmique en base 10 à l'espace d'entrée, ce modèle peut être relié à l'estimateur théorique de distance ISAT dans lequel la largeur des bandes est à la limite de la résolution fréquentielle soit 1 échantillon fréquentiel bin. L'équation 3.10 peut être réécrite comme suit :

$$\begin{aligned} \hat{r} &= A_{12} (\log_{10}(|X_1(f; r)|^2) - \log_{10}(|X_2(f; r)|^2)) \\ &= A_{12} \sum_{i=1}^2 \log_{10} \left(E_i^{(-1)^{i+1}} \right) \end{aligned} \quad (4.13)$$

où \hat{r} représente l'estimation de distance en un instant donné (date de détection du clic) et où le terme A_{12} remplace la constante dépendante du choix de bande (bande 1/ bande 2)

et du modèle d'atténuation fréquentielle choisi. Ici ce terme dépend de manière globale du couple 1, 2.

Par analogie, une sortie y du réseau de neurones *ISAR* à N_{hu} unités cachées et une entrée à N dimensions (typiquement 128) peut s'écrire :

$$\begin{aligned}\hat{r} &= \sum_{j=1}^{N_{hu}} [A_{j1} \log_{10}(|X_1(f; r)|) + A_{j2} \log_{10}(|X_2(f; r)|) + \dots + A_{jN} \log_{10}(|X_N(f; r)|)] \\ &= \sum_{j=1}^{N_{hu}} \sum_{i=1}^N A_{ji} \log_{10}(\sqrt{E_i})\end{aligned}\tag{4.14}$$

La sortie du réseau de neurones montre des similarités avec l'estimateur ISAT. Elle peut être vu comme une version avancée et complétée d'ISAT pour lequel la largeur des bandes serait d'un canal fréquentiel. La valeur des poids A étant négative ou positive, cet estimateur n'est plus construit sur le ratio entre 2 bandes mais sur une combinaison produit/ratio de toutes les bandes (réduite à 1 échantillon fréquentiel). De plus A dépend des bandes séparément et non de toutes comme pour ISAT. Cette constante exprime maintenant une dépendance complexe à plusieurs paramètres du milieu d'apprentissage comme le canal de propagation acoustique, les caractéristiques intrinsèque de l'émetteur, celles du récepteur et de la chaîne d'acquisition. Ce qui renvoie au problème de la généralisation d'un tel modèle. On peut aussi émettre l'hypothèse que la forme du spectre est prise en compte dans cette constante. Celle-ci avait été écarté dans ISAT avec l'hypothèse du signal de type Dirac.

- Dans la mesure où nous possédons une vérité terrain riche et précise sur les données d'apprentissage, nous avons voulu vérifier si de l'information sur les angles de position pouvait être extraite des spectres de clics, nous proposons le modèle *ISARH* (Intra Spectral Attenuation Regression Hierarchical). L'objectif est d'améliorer le modèle ISAR en associant 2 modèles de prédiction en "cascade". Le premier apprend une régression des composantes du spectre pour l'estimation des angles décrivant la position de l'animal : l'off axis o , l'azimut a et élévation el . Un second modèle procède à une régression sur le spectre complété des angles estimés : \hat{o} , \hat{a} et \hat{el} pour l'estimation de distance, tel que :

$$\hat{r} = h(|X(f; r)|; \hat{o}; \hat{a}; \hat{el}; \mathbf{W}).\tag{4.15}$$

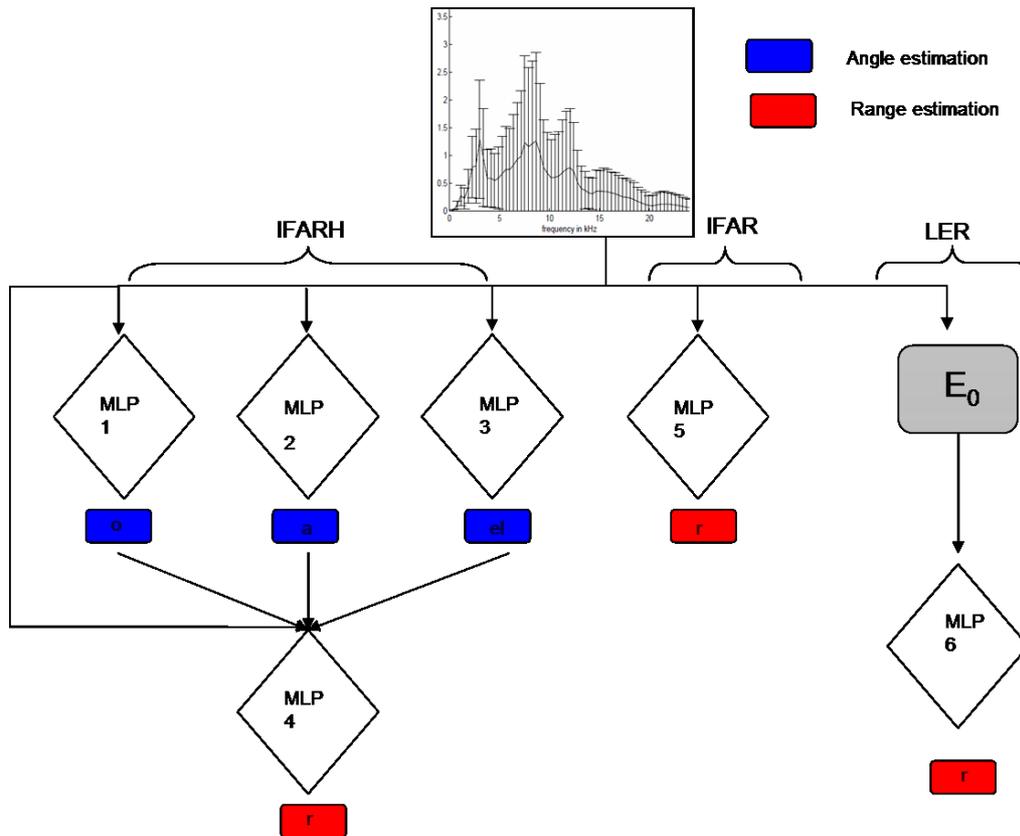


FIGURE 4.2 – Schéma de synthèse des différents modèles considérés. LER, ISAR, ISARH (IFARH et IFAR sont respectivement ISARH et ISAR).

4.3 Conclusion

Nous avons introduit les réseaux neuromimétiques en tant qu'approximateurs universels de fonctions. Nous présentons l'expression de la sortie d'un perceptron simple couche puis multicouches ainsi que la règle d'apprentissage par rétro propagation du gradient de l'erreur.

Nous proposons donc d'utiliser un perceptron (MLP) à 2 couches cachées pour apprendre la relation empirique entre distance source-récepteur et le spectre du signal reçu. Nous avons construit plusieurs modèles que nous récapitulons.

- le modèle LER est une régression de l'énergie totale du clic (échelle globale)
- le modèle ISAR est régression des composantes du spectre (échelle intra-spectre)
- le modèle ISARH est une association en cascade de plusieurs MLP faisant intervenir une régression intermédiaire des angles de position puis une régression des composantes du spectre (échelle intra-spectre)

Nous proposons une analogie entre les sorties du premier estimateur ISAT et l'estimateur ISAR. Ce dernier est vu comme une version avancée et complétée d'ISAT pour lequel la largeur des bandes serait d'un canal fréquentiel et dont les combinaisons incluraient des produits et des ratios entre les bandes d'énergie.

Modèle de codage parcimonieux de signaux stationnaires sous-marins

Sommaire

5.1	Objectifs	55
5.2	Représentation cepstrale	56
5.3	Codage parcimonieux	56
5.4	Corrélation MFCC/codes parcimonieux	58
5.5	Estimation de la complexité du dictionnaire	58
5.6	Mesure de divergence des codes d'une année sur l'autre	60
5.7	Conclusion	61

5.1 Objectifs

Depuis 2007, notre équipe travaille sur l'étude des chants des baleines à bosse. Ces sons émis par les mâles au cours de la période de reproduction joue un rôle important, soit pour attirer les femelles [Winn 78, Herman 80, Tyack 81, Medrano 94], soit pour délimiter des territoires [Darling 01, Cholewiak 08]. Plusieurs autres hypothèses évoquaient notamment la possibilité que ces chants soient issus des mâles exclus des groupes actifs [Clapham 96]. Toutefois, les baleines mâles créent un chœur permanent à plusieurs individus émettant des chants simultanément dans la même zone géographique, produisant ainsi une cacophonie puissante importante.

Les chants sont organisés en une succession d'unités sonores [Payne 71], elles-mêmes structurées en sous-phrases, en phrases, définissant ainsi des leitmotifs. Ces derniers sont fonction de l'année et du lieu géographique.

A ce jour, le nombre d'unités sonores n'est pas clairement établi, allant de 9 [Au 06] à une 60e [Dunlop 07]. Cette situation s'explique du fait de modifications, parfois subtiles, dans les unités qui composent un chant, rendant difficile la classification. Les travaux montrent également, le plus souvent, une classification subjective, plutôt que le recours à une approche automatique basée sur des critères mathématiques. Pour pallier cet inconvénient, notre démarche est basée sur les deux propositions suivantes :

- introduction du concept de sous-unités sonores : une ou plusieurs sous-unités constituent une unité sonore. En changeant d'échelle, on fait les hypothèses que, d'une part, le nombre de classes de sous-unités est inférieur ou égal au nombre d'unités sonores et d'autre part, que le nombre de sous-unités peut être limité, même si celui des unités ne l'est pas.

- recours à une méthode robuste de classification pour aborder ces chants. Pour la première fois, nous introduisons un codage parcimonieux entièrement automatique des chants de baleines à bosse afin de déterminer les composantes stables par rapport à celles qui évoluent à des échelles de temps différents.

Dans ce chapitre, nous proposons également une définition de la complexité du code qui peut clairement séparer les composantes du chant de celles du bruit de mer. Nous illustrerons notre méthode sur un chant préalablement étudié et analysé. Nous avons également comparé ces résultats avec d'autres méthodes tant sur les unités que sur les sous-unités.

Nous avons calculé les coefficients cepstraux à échelle Mel (MFCC) des enregistrements sonores à partir du programme Spro [Gravier 10]. Cette approche est issue de l'analyse de la parole humaine. Ce choix est motivé par plusieurs raisons : premièrement, l'information pertinente des unités sonores est principalement portée par la fréquence fondamentale et les premières harmoniques (quand elles existent), c'est-à-dire contenue dans la bande des basses fréquences, entre 50 Hz et 4 kHz. Deuxièmement, les harmoniques à plus haute fréquence sont moins énergétiques, moins stables, et donc plus sensibles au bruit. Troisièmement, les hautes fréquences sont davantage déformées lors de la propagation acoustique. Le recours à l'échelle Mel nous permet de focaliser notre approche sur les basses fréquences et donc de tenir compte de ces 3 caractéristiques.

Par ailleurs, les MFCC sont des traits classiques suffisamment invariants pour permettre d'uniformiser les canaux. Il est intéressant de noter à cet égard que les MFCC correspondent aux premiers étages de traitement des scalograms, et en sont finalement une approximation.

5.2 Représentation cepstrale

Parmi les différentes méthodes classiques de représentation des unités sonores (coefficients de Fourier, AR, LPC, MFCC), le recours au MFCC semble la plus efficace [Davis 80, Rabiner 93, Pace 09, Pace 10]. Cinq ensembles de coefficients cepstraux à échelle Mel ont été calculés, avec un décalage de fenêtre de 10 ms et avec, respectivement, une longueur de fenêtre d'analyse différente de : 250 ms, 500 ms, 1 s, 2 s et 4 s. Nous calculons seulement les 12 premiers coefficients statiques M_1, M_2, \dots, M_{12} et l'énergie, ce qui donne des vecteurs MFCC de dimension 13. Le terme statique est employé par opposition aux coefficients dérivés (différence des coefficients de deux fenêtres consécutives) et dérivés secondes (différence des coefficients dérivés).

Sur ces vecteurs résultants, nous appliquons une normalisation par soustraction de la moyenne cepstrale (CMS). Enfin, par pas de 10 ms, nous concaténons un certain nombre de vecteurs MFCC consécutifs en un seul pour former un "super-vecteur" MFCC. L'idée est de prendre en compte dans chaque « super-vecteur » les données relatives à une échelle plus importante. Par exemple pour analyser les sous-unités de 500 ms, nous prenons 50 MFCC consécutifs, d'où des vecteurs résultants de tailles 13x50 composantes.

5.3 Codage parcimonieux

Pour décrire des signaux, deux approches opposées peuvent être suivies : tenir compte soit de l'information intrinsèque totale, soit d'une partie de cette information. La première méthode est risquée notamment lorsque le nombre de classes n'est pas connue et éventuellement non limité. Le critère d'attribution à une classe se heurte à une ambiguïté entre deux ou plusieurs classes et le seuil choisi sera toujours sans doute sous-optimal (soit pour une classification robuste, soit pour un temps de calcul pertinent). Dans la deuxième méthode, l'hypothèse est faite que les

différents éléments d'une même classe partagent un socle commun et l'objectif est donc de baser la classification sur les informations de ce socle.

L'utilisation du codage parcimonieux (*sparse coding*) provient de deux de leurs principaux avantages : d'une part, les vecteurs de code parcimonieux sont plus efficaces que les vecteurs complets lorsqu'ils doivent être classifiés ; d'autre part, ils engendrent une faible erreur de reconstruction. Le premier avantage est important car nous voulons analyser et classer plus précisément des unités et des sous-unités sonores. Le deuxième avantage assure qu'en utilisant des vecteurs parcimonieux, nous ne pratiquons pas une approximation trop grossière par rapport aux vecteurs des paramètres initiaux complets. Par ailleurs, le codage parcimonieux peut être utilisé de manière non supervisée, sans avoir besoin d'aucune connaissance sur les données. Pour une faible erreur de reconstruction, le codage parcimonieux permet aussi une bonne généralisation pour des données inconnues.

L'objectif de la classification est de regrouper les données disponibles afin de les classer par classes prédéfinies. La première méthode qui peut être utilisée est une quantification vectorielle (VQ) [Rabiner 93]. Un dictionnaire de K vecteurs est appris et chaque classe est représentée par un seul vecteur du dictionnaire (appartenant aux données d'apprentissage). Tout vecteur de nouvelles données est alors représenté par le vecteur le plus proche du dictionnaire.

Soit \mathbf{X} la matrice de dimensions $n \times N$ extraite à partir de N données audio à n dimensions d'entrée, *i.e.* $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_N] \in \mathbb{R}^{n \times N}$.

Soit \mathbf{D} le dictionnaire composé de K vecteurs tels que $\mathbf{D} = [\mathbf{d}_1, \dots, \mathbf{d}_K] \in \mathbb{R}^{n \times K}$ et formé à partir des données. A partir de ce dictionnaire \mathbf{D} , dans l'approche VQ chaque vecteur \mathbf{x}_i de \mathbf{X} est affecté à un seul \mathbf{d}_j tel que :

$$\mathbf{d}_j = \arg \min_{k=1, \dots, K} \|\mathbf{x}_i - \mathbf{d}_k\|_2. \quad (5.1)$$

Soit $\mathbf{C} = [\mathbf{c}_1, \dots, \mathbf{c}_N] \in \mathbb{R}^{K \times N}$ la matrice pour laquelle chaque vecteur \mathbf{c}_i n'a qu'une seule composante $c_i^j \neq 0$, correspondant au vecteur \mathbf{d}_j du dictionnaire. Le problème d'optimisation VQ est formulé comme suit :

$$\arg \min_{\mathbf{D}, \mathbf{C}} \sum_{i=1}^N \|\mathbf{x}_i - \mathbf{D}\mathbf{c}_i\|_2^2 \quad s.t. \quad \|\mathbf{c}_i\|_{\ell_0} = 1, \forall i, \quad (5.2)$$

où $\|\mathbf{x}\|_{\ell_0}$ désigne la pseudo norme-zéro, *i.e.* un seul élément de \mathbf{x} est égal à 1, les autres sont égaux à 0. Les matrices \mathbf{D} et \mathbf{C} doivent être optimisées conjointement par un algorithme tel que K -means par exemple.

Comme tout vecteur des données d'entrée est uniquement représenté par un vecteur du dictionnaire, cette approximation est trop forte dans la plupart des cas et il en résulte des erreurs de classification. L'idée du codage parcimonieux est de relâcher la contrainte $\|\mathbf{x}\|_{\ell_0} = 1$ afin d'exprimer un vecteur d'entrée non par un seul vecteur du dictionnaire mais par une combinaison linéaire de quelques vecteurs du dictionnaire. Habituellement, le dictionnaire \mathbf{D} est sur-complet, *i.e.* qu'il y a plus de vecteurs que de classes. Le problème à résoudre s'exprime alors par l'équation suivante :

$$\arg \min_{\mathbf{D}, \mathbf{C}} \sum_{i=1}^N \|\mathbf{x}_i - \mathbf{D}\mathbf{c}_i\|_2^2 + \lambda \|\mathbf{c}_i\|_{\ell_1} \quad s.t. \quad \|\mathbf{d}_j\|_2 = 1, \forall j. \quad (5.3)$$

Le terme de régularisation λ couplé à la norme ℓ_1 garantit la parcimonie des codes optimisés. Cependant, comme cette optimisation jointe n'est pas convexe, sa résolution se fait en procédant

de manière itérative jusqu'à convergence : (i) la mise à jour du dictionnaire courant par exemple *via* descente de blocs de coordonnées (à partir des codes parcimonieux courants), (ii) la mise à jour des codes parcimonieux selon le dictionnaire courant par exemple *via* l'algorithme LARS pour résoudre le problème du LASSO (implémentation basée sur les codes sources de J. Mairal [Mairal 09]).

La partie duale de la formation du dictionnaire \mathbf{D} et du calcul des projections parcimonieuses \mathbf{C} est la reconstruction. C'est-à-dire, comment à partir d'un vecteur de code parcimonieux et connaissant \mathbf{D} nous retrouvons une approximation du vecteur MFCC d'entrée. Aussi, pour un vecteur MFCC $\mathbf{x} \in \mathbb{R}^n$ et le vecteur code parcimonieux associé $\mathbf{c} \in \mathbb{R}^K$, la reconstruction du vecteur MFCC $\hat{\mathbf{x}}$ est la combinaison linéaire des vecteurs parcimonieux \mathbf{d}_i du dictionnaire selon les valeurs c_i du code parcimonieux \mathbf{c} . Plus formellement, $\hat{\mathbf{x}}$ est donné par l'équation suivante :

$$\hat{\mathbf{x}} = \mathbf{D} \cdot \mathbf{c} = \sum_{i=1}^K \mathbf{d}_i \cdot c_i. \quad (5.4)$$

5.4 Corrélation MFCC/codes parcimonieux

A chaque vecteur MFCC correspond un vecteur de code parcimonieux (décomposition sur le dictionnaire). Ainsi, nous analysons la corrélation entre les vecteurs MFCC et les vecteurs de code parcimonieux. Nous nous attendons à ce que les structures qui apparaissent par autocorrélation des MFCC apparaissent toujours avec l'autocorrélation de vecteurs de code parcimonieux. Sur l'enregistrement de 2009, la figure 5.1 montre qu'il y a bien conservation des propriétés de corrélation entre les domaines cepstral et parcimonieux. A gauche l'autocorrélation des vecteurs MFCC et à droite l'autocorrélation sur les vecteurs de codes parcimonieux (sur 400 échantillons choisis au hasard). Etant donné que la corrélation est symétrique, nous n'affichons qu'une partie triangulaire.

La figure 5.2 montre qu'entre l'espace des MFCC et l'espace des codes parcimonieux, la corrélation est plus forte surtout dans l'espace parcimonieux. Cela signifie que l'information présente dans le domaine MFCC est renforcée dans le domaine parcimonieux. Toutefois, le principal inconvénient est qu'il est possible qu'une corrélation non informative dans le domaine MFCC soit surestimée et induise des erreurs d'interprétation.

5.5 Estimation de la complexité du dictionnaire

De nombreuses fonctions ont été proposées pour estimer la complexité dans un plan temps-fréquence, dont les entropies de Shannon et Renyi [Flandrin 94]. Afin d'analyser le dictionnaire que nous générons dans le domaine des cepstres, nous étendons la définition de la complexité temps-fréquence à la complexité du modèle cepstral. Intuitivement, si une composante a une concentration d'énergie dans le plan temps-fréquence, nous allons supposer la même chose dans le plan des cepstres mais cette notion est encore difficile à traduire dans un concept quantitatif. Plutôt que de s'attarder à la question de ce qu'est une composante cepstrale, nous étudions une mesure quantitative de la complexité inspirée de travaux précédents [Flandrin 94]. Cette mesure est intimement liée à l'hypothèse que les signaux d'une grande complexité (et donc à haut contenu d'information) doivent être construits à partir d'un grand nombre de composantes élémentaires. Nous utilisons donc comme mesure de complexité des vecteurs parcimonieux \mathbf{d}_i du

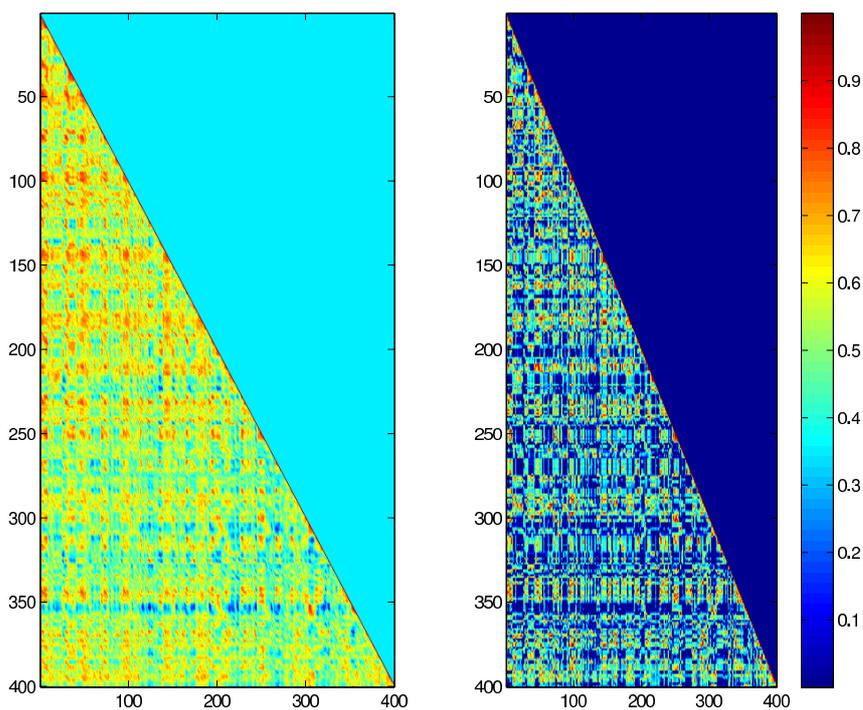


FIGURE 5.1 – Représentation de la matrice d'autocorrélation entre les vecteurs MFCC du côté gauche, respectivement, pour les vecteurs de code parcimonieux du côté droit. Les résultats sont obtenus sur l'enregistrement de 2009 avec un dictionnaire appris sur l'année 2009. La longueur de fenêtre est de 4 s et le dictionnaire contient 32 vecteurs.

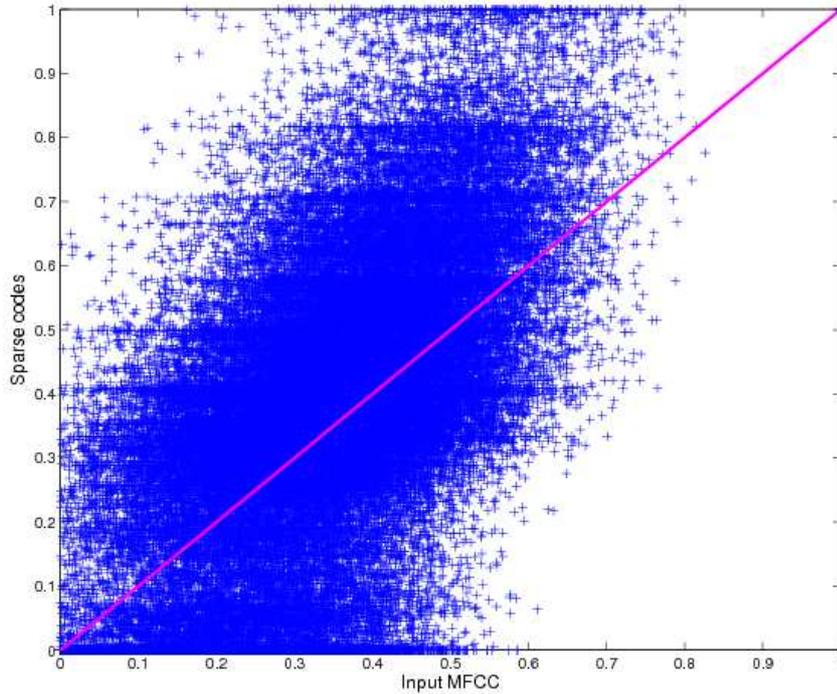


FIGURE 5.2 – Rapports de corrélations entre espace vectoriel MFCC et espace vectoriel des codes parcimonieux. Les points au-dessus de la diagonale sont plus corrélés dans l’espace de code parcimonieux, les points en dessous sont de moins en moins corrélés dans le code parcimonieux. Enregistrements de 2009, de longueur de fenêtre de 4 s, dictionnaire de 32 vecteurs.

dictionnaire \mathbf{D} l’entropie de Shannon :

$$H(\mathbf{d}_i) = - \sum_{t,j} p(\mathbf{d}_i(t, M_j)) \cdot \log(p(\mathbf{d}_i(t, M_j))), \quad (5.5)$$

où $p(\mathbf{d}_i(t, M_j))$ est l’estimation de la distribution d’énergie dans la cellule au temps t et pour le coefficient cepstral M_j .

5.6 Mesure de divergence des codes d’une année sur l’autre

La théorie de l’information [Shannon 48] détermine la structure et l’organisation dans un système de communication. Elle donne un critère objectif pour comparer et mettre en contraste les systèmes de communication. La distance de Kullback-Liebler (KL) [Kullback 51] est la distance de référence entre deux distributions de variables aléatoires. C’est donc la métrique la plus directe pour notre sujet, sans rentrer dans une modélisation plus fine de ces chants, qui pourrait induire des *a priori*. Nous proposons donc d’estimer la divergence de chant avec la distance de Kullback-Liebler des composantes du chant. Afin d’obtenir une analyse diachronique, *i.e.* afin de déterminer quel code est plus ou moins utilisé d’une année à l’autre, nous calculons entre chaque code, leur divergence. Nous supposons que la moyenne de la distance de Kullback-Leibler sur un sous-ensemble du code entre la distribution des chants de 2008 et celle de 2009 mesure un changement de structure à différents niveaux. On peut penser que ce changement reflète une évolution

des chants, même s'il n'y a pas de déduction de distance évidente entre les distributions. En supposant également que plus cette distance est grande, plus les chants ont évolué d'une année à l'autre. Par conséquent, la distance de chant est défini comme suit :

Soit $A_{\mathbf{d}_i}$ (resp. $B_{\mathbf{d}_i}$) la distribution de probabilité discrète sur R ensembles $r = \{1, \dots, R\}$ des codes parcimonieux \mathbf{C} de 2008 pour le vecteur parcimonieux \mathbf{d}_i (resp. pour 2009). Alors la distance pour le vecteur parcimonieux \mathbf{d}_i est la suivante :

$$dist_{KL}(A_{\mathbf{d}_i}, B_{\mathbf{d}_i}) = \sum_{r=1}^R (A_{\mathbf{d}_i}^r - B_{\mathbf{d}_i}^r) \cdot \log_2(A_{\mathbf{d}_i}^r / B_{\mathbf{d}_i}^r). \quad (5.6)$$

Enfin, la distance KL de chant finale est la moyenne de $dist_{KL}$ sur le sous-ensemble de codes cibles.

5.7 Conclusion

Dans ce chapitre, nous avons défini l'origine de nos motivations pour la classification des signaux de type voisements en particulier sur les chants de baleine à bosse. Afin d'évaluer, au sein d'une population d'individu, la stabilité des séquences de chant en fonction de différentes échelles de temps, nous proposons une représentation des signaux selon un codage parcimonieux automatique et robuste. L'un des atouts majeurs de cette technique de codage est de fournir des descripteurs de l'objet codé invariants par transformation spatiale des vecteurs d'entrée. Nous proposons également une évaluation de la complexité des codes afin de séparer les chants du bruit de mer. Inspirés du traitement de la parole humaine, nous utilisons comme représentation de départ, les vecteur MFCC à 13 composantes (dont l'énergie). Nous avons rappelé les équations régissant la méthode de codage parcimonieux et aboutissant à un problème de LASSO que nous résolvons à l'aide de l'algorithme LARS. L'autocorrélation des vecteurs parcimonieux d'une part et des vecteurs MFCC de l'autre, montre que l'utilisation des "sparse codes" renforcent l'information présente dans les entrées MFCC. Nous représentons aussi la corrélation entre les codes et les entrées MFCC.

Puis, nous mesurons la complexité des vecteurs parcimonieux en calculant l'entropie de Shannon. Celle-ci est élevée lorsque le signal codé est complexe et faible lorsque le signal codé est aléatoire.

Enfin, nous proposons d'analyser la stabilité des codes parcimonieux pour des séquences de chant d'échelles différentes en étudiant la divergence des segments d'une année à l'autre. La métrique utilisée est la distance de Kullback-Liebler car elle quantifie l'écart entre 2 distributions de variables aléatoires.

6

Installations, mesures et matériel disponibles pour valider les modèles

Sommaire

6.1	Chaîne d'acquisition	63
6.2	Systèmes et données monocapteur	64
6.2.1	Dispositif autonome monocapteur Gualiba	64
6.2.2	Enregistrements et base de données DECAV au PNPC/PELAGOS	66
6.2.3	Installation et enregistrements BOMBYX au sud du PNPC (sanctuaire PELAGOS)	69
6.2.4	Enregistrements de chants de baleine à bosse à Madagascar	70
6.2.5	Enregistrements et base de données "St Pierre et Miquelon" SPM	71
6.3	Systèmes multicapteurs	74
6.3.1	Dispositif autonome multicapteurs BAOBAB	74
6.3.2	Base de données "Bahamas"	78
6.4	Conclusion	83

6.1 Chaîne d'acquisition

Procéder à des enregistrements sous-marins implique le recours à une chaîne d'acquisition classique incluant un capteur, une mise en forme des signaux avant les étapes de traitement incluant la détection et l'extraction d'informations. L'ensemble de cette chaîne doit être homogène en qualité pour optimiser les analyses. Nous reprenons ici l'ensemble des étages de la chaîne d'acquisition. Le capteur est un hydrophone, suivi de son pré-amplificateur et d'un filtre passe-bas pour couper avant la fréquence de Nyquist, avant de procéder à la numérisation de l'enregistrement analogique. Au début des années 2000, ce conditionnement du signal se faisait à distance de l'hydrophone, c'est-à-dire après transmission par câble des différences de pression captées par l'hydrophone. L'inconvénient de cette approche est que du bruit venait se rajouter lors de la transmission par le câble. Ainsi, pour nos applications, nous avons privilégié le pré-traitement et la numérisation directement en sortie de l'hydrophone. Ce dernier est donc fixé sur un caisson qui est lui-même totalement immergé à la profondeur choisie. Le signal numérique est remonté à la surface pour transmission ou stockage. Cela permet de conserver la sensibilité de l'hydrophone et de son pré-amplificateur, sans perdre de performance avec l'ajout de bruit via le câble. Une fois le signal numérisé, il peut être stocké totalement sur un disque dur pour archivage

et traitement ultérieur ou transmis à une station à proximité, sur la côte ou sur un bateau. Nous avons retenu ces 2 solutions dans le matériel développé au cours de ma thèse. Enfin, viennent les étapes de traitement du signal et reconnaissance des formes qui consistent à filtrer les signaux pour réaliser la détection des émissions sonores des cétacés et pour extraire les informations qui permettront de les classer et d'indexer les détections.

Pour ma thèse, je me suis concentré sur 2 approches de captation différentes : monocapteur et multicapteurs. Le travail sur le système monocapteur GUALIBA a été largement initié par le partenaire industriel CeSigma et mes directeurs de thèse avant le début de ma thèse.

6.2 Systèmes et données monocapteur

6.2.1 Dispositif autonome monocapteur Gualiba

L'objectif est de lever les difficultés techniques liées à la captation et de proposer un système autonome permettant d'enregistrer différentes espèces de cétacés évoluant dans une zone spécifiques. Le cahier des charges que se sont fixées les acteurs du projet était le suivant : en terme d'objectif biologiste, il s'agissait tout d'abord de détecter la présence d'émissions sonores de cétacés présents dans une zone autour de l'hydrophone. Il est à noter que nos zones d'intérêt (Méditerranée, Antilles, Madagascar) sont fréquentées par différentes espèces de cétacés incluant des mysticètes et des odontocètes. Les intensités émises par les différentes espèces vont d'un peu moins de 100 dB à plus de 190 dB re/1 μ Pa à 1 m. Les bandes fréquentielles sont plutôt larges allant de quelques Hz à plus de 100 kHz. Toutefois, l'objectif de détection nous autorise à être un peu plus lâche sur les contraintes, notamment en termes de gain et de bande passante. Cela signifie qu'il est possible de s'autoriser a) des saturations temporaires, par exemple lorsque le cétacé est trop proche ou lors de la présence exceptionnelle d'une activité humaine et b) une bande passante plus réduite, étant entendu que les espèces de cétacés visées émettent une partie énergétique de leurs sons dans cette bande passante. Il en est bien sûr autrement si l'on souhaite faire l'analyse des émissions sonores de ces espèces. Dans ce cas, au contraire, il faut prêter une attention tout à fait particulière au gain (en privilégiant les cétacés les plus proches et à la bande fréquentielle. Celle-ci doit être assez large pour englober l'ensemble de la répartition fréquentielle des sons émis par l'espèce visée).

En terme d'objectif écologiste, il était souhaitable de disposer d'un matériel qui puisse servir à la détection d'espèces évoluant à distance de l'hydrophone. Le choix s'est donc porté sur un gain maximum de la chaîne d'acquisition, afin de détecter la présence des individus qui émettent des sons les plus faibles en intensité, ou les individus les plus éloignés. En terme d'autonomie, il a été retenu le fait que le matériel devait embarquer l'énergie électrique suffisante pour assurer son fonctionnement. En terme de données, celles-ci sont conservées brutes en l'état, aucun traitement embarqué n'est considéré alors que la mode des capteurs intelligents montrait un regain d'intérêt au milieu de ma thèse. Ce choix a été motivé principalement par le fait que les enregistrements acquis étaient à destination de travaux de recherche et que, par conséquent, il était important de conserver toutes les informations qu'ils contenaient et de ne pas commencer à faire des choix *a priori* pouvant détruire ou altérer des informations qui pourraient s'avérer utiles voire cruciales dans la suite des analyses. D'autre part, nous nous intéressons également à l'environnement et, notamment aux activités humaines bruyantes. Aussi, les enregistrements sur des périodes où il n'y a pas forcément d'émissions sonores de cétacés sont également importants dans nos travaux d'observation.

En terme de transmission, les ingénieurs ont été confrontés à des choix technologiques difficiles, d'une part par rapport aux distances à la côte auxquels ils souhaitaient déployer le matériel

et, d'autre part, par rapport à l'énergie demandée pour cela. Donc le choix retenu a été a) de stocker tous les enregistrements sur le matériel via un disque dur ou une mémoire SSD, b) d'envisager une transmission locale (soit par échange de disque dur externe ou de SSD, soit via une connection Wifi, et c) de recourir à une transmission par câble si besoin.

Le module autonome de surface a été déployé pour la première fois en Guadeloupe. Appelé Gualiba, il était constitué de 2 parties distinctes : une bouée de surface et le module d'acquisition immergé en profondeur :

- La bouée de surface est destinée à recevoir :
 1. L'alimentation électrique : celle-ci est assurée par 3 panneaux solaires de 20W (monocristallin norme EN61215, 450/425mm) doublés d'une éolienne de 100W à 20 nœuds (Marine Aérogène 4, 12V) visibles sur la figure 6.1, le tout couplé à 2 batteries gel (Optima Blue Top, 75Ah) de stockage. Nous avons retenu 2 sources de génération d'électricité car a) les endroits retenus pour le déploiement d'un tel système sont variés, b) ces 2 systèmes sont complémentaires, et c), afin de palier le manque de luminosité ou le manque de vent à certaines périodes de la journée.
 2. L'ordinateur : il est installé dans un compartiment étanche à l'intérieur de la bouée. Il s'agit d'une carte Mini ITX intégrant un processeur basse consommation ATOM D510 (1,6 GHz, double cœur et 1 Go de RAM). Les 2 unités de stockage sont des SSD Kingston de 512 MB et 256 MB, extractible pour permettre leurs échanges et de récupérer les données. Les SSD se sont imposés d'une part pour leur plus faible consommation électrique que les disques durs externes et d'autre part parce qu'ils sont plus résistants aux vibrations.
 3. Le système de parafoudre, placé en haut du mat.
 4. Le système de transmission, via une borne Wifi.

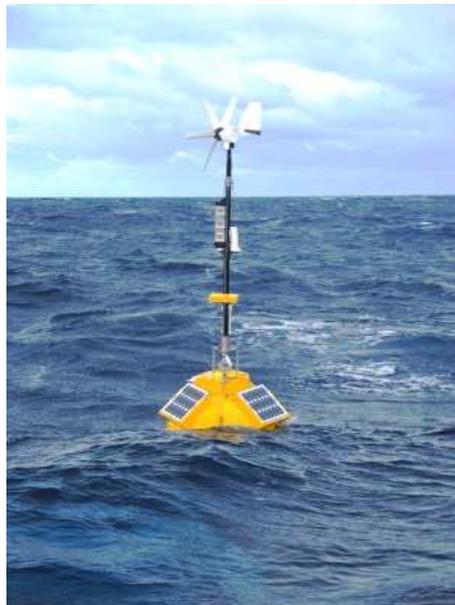


FIGURE 6.1 – Vu du dispositif gualiba en surface.

La bouée a été dessinée par PLK Marine et une quille d'équilibrage a finalement été ajoutée pour assurer sa stabilité et éviter qu'elle ne coule par forte houle.

- Le module immergé est constitué d'un hydrophone :

L'hydrophone choisi par mon entreprise est le Cetacean CR55 pour ses caractéristiques larges et ses performances. Il s'agit d'un matériel sensible (-185 dB re1V/ μ Pa et -165 dB re1V/ μ Pa lorsqu'il est alimenté par son pré-ampli) et robuste. Il est omnidirectionnel et sa bande passante est de 10 Hz à 30 kHz. Son pré-amplificateur a été disposé en sortie directe de l'hydrophone pour éviter toute captation de bruit par câble.

- La carte d'acquisition :

Une carte d'acquisition CAPNG a été développée par l'entreprise CeSigma pour numériser les enregistrements sur 24 bits avec une fréquence d'échantillonnage pouvant aller jusqu'à 250 kHz. Il est également possible d'ajouter un gain sur cette carte d'acquisition. L'ensemble de cette instrumentation est placé dans un conteneur immersible à plus de 50 m.

Le module immergé est relié par une corde à la bouée. Il y a 2 difficultés : s'affranchir des mouvements de la bouée en surface et, notamment, le fait qu'elle tire sur la corde en permanence avec de la houle. Pour éviter cela, un système antipilonnement (en fait, un détournement du câble) a été ajouté permettant d'atténuer voire d'éliminer complètement par petite houle (< 2 m) les oscillations générées par la bouée. Il faut éviter que les câbles transmettant l'énergie électrique et les données s'emmêlent dans la corde notamment lors de tensions et de rotations dues aux mouvements désordonnés entre le module immergé et le module de surface. Un système de gainage accroché à la corde est alors retenu pour éviter tout emmêlement.

6.2.2 Enregistrements et base de données DECAV au PNPC/PELAGOS

DECAV est un projet DYNI/UTLN de suivi bioacoustique des cétacés dans le Parc National de Port-Cros (PNPC), en charge de l'animation de la partie française du Sanctuaire Pelagos (Barcelo et Boudouresque, 2012) et financé par PELAGOS 2010-2013. Les enregistrements sont très riches en termes d'événements acoustiques aussi bien d'origine biologique qu'anthropique. Dans le cadre de cette thèse nous nous sommes intéressés plus particulièrement aux détections de signaux impulsifs de cachalot. Le laboratoire, dans ce cadre, a élaboré un nouvel algorithme d'analyse de la structure multipulsée des clics de cachalots pour l'estimation de la taille des individus, l'intervalle de temps entre les pulses (IPI) dépendant entre autres paramètres de la taille de l'animal [Abeille 14].

Ce corpus de données a été enregistré sur 26 sorties entre mai 2011 et octobre 2012, depuis un zodiac ou un bateau avec le PNPC ou pour certaines sorties avec l'Association Regard du Vivant. Les enregistrements sont mono-hydrophone de durée de 5 min. L'hydrophone utilisé est le modèle CR55 de Cetacean Research SA Seattle, ayant une réponse plate sur la bande d'intérêt. L'hydrophone est déployé rapidement et immergé à environ -20 m (câble de 30 m à 50 m). La fréquence d'échantillonnage est de 48 kHz, le codage sur 16 bits, le tout acquis par la carte son PC. Le bilan des sorties effectuées et les résultats des estimations des tailles sont donnés en table 6.2. Les points d'écoutes sont représentés géographiquement sur la carte donnée en figure 6.3.

Date des 27 sorties	Bateau PNPC ou 'Regard du Vivant'	Nbr contact visuel (Tot. 5)	Nbr contact acoustique (Tot. 21)	Horaire du contact	IP +/- 0,01 ms (si mesurable)	Taille (m) suivant Growcott	Zone du contact
06/05/2011	PNPC	0	2	11h57	4,88 et ?	11,88	Tête du canyon des Stoechades
19/05/2011	PNPC	0	1	10h49	5,66	12,86	4 km au Sud de Port-Cros
07/06/2011	PNPC	0	0	-	-	-	-
12/07/2011	PNPC	0	0	-	-	-	-
28/07/2011	PNPC	0	0	-	-	-	-
02/08/2011	PNPC	1	1	9h45	4,54	11,45	8 km au Sud-Est du Levant
31/08/2011	PNPC	0	1	10h52	5,38	12,50	7,5 km au Sud de Porquerolles
16/09/2011	PNPC	0	0	-	-	-	-
26/01/2012	PNPC	0	2	10h47	4,35 et 4,55	11,21 et 11,46	9 km au Sud de Porquerolles
16/03/2012	PNPC	0	0	-	-	-	-
26/03/2012	PNPC	0	0	-	-	-	-
04/05/2012	PNPC	0	0	-	-	-	-
30/05/2012	PNPC	1	2	10h53 et 11h27	4,40 et 4,45	11,27 et 11,33	Canyon des Stoechades
06/06/2012	PNPC	0	1	9h20	6,3*	13,66*	3 km à l'Est du Levant
07/06/2012	PNPC	0	0	-	-	-	-
15/06/2012	PNPC	0	0	-	-	-	-
28/06/2012	PNPC	0	0	-	-	-	-
10/07/2012	RdV	1	2	9h52	4,86 et ?	11,85	Large Cap Sicié
23/07/2012	PNPC	0	0	-	-	-	-
02/08/2012	RdV	0	2	10h16	4,90 et 4,78	11,90 et 11,74	8 km au Sud-Est du Levant
23/08/2012	PNPC	0	1	-	?	-	-
09/09/2012	RdV	0	0	-	-	-	-
16/09/2012	RdV	0	0	-	-	-	-
17/09/2012	RdV	1	2	14h04 et 14h09	4,60 et 4,66	11,52 et 11,60	Cassis Sud - Bec de l'aigle
23/09/2012	RdV	0	1	13h58	5,72	12,93	Sud Cassis - Cap de l'aigle
06/10/2012	RdV	1	2	17h23 et 17h56	4,44 et 5,62	11,32 et 12,80	Large Cap Sicié
28/11/2013	RdV	0	1	15h30	5,5*	12,65*	Cassis Sud - Bec de l'aigle

FIGURE 6.2 – Résultats bruts de toutes les sorties du projet DECAV : les cinq cachalots vus ont été mesurables uniquement acoustiquement, comme 16 autres uniquement contactés acoustiquement sur un total de 21 contacts acoustiques de cachalots ("PNPC" : catamaran moteur ou zodiac du PNPC, "RdV" : association Regard du Vivant [Abeille 14]).

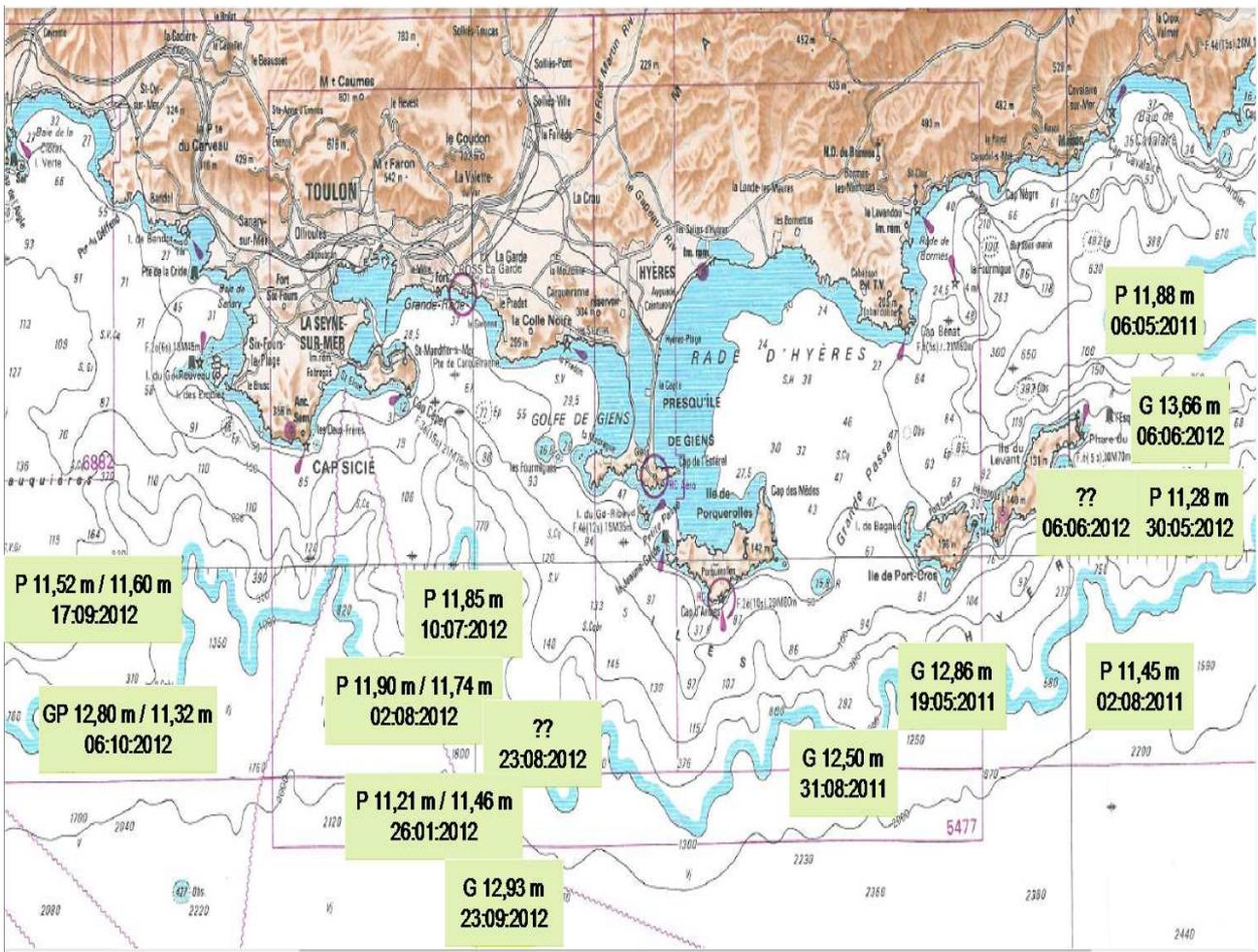


FIGURE 6.3 – Détections de cachalots au large de Toulon/Hyères. Les étiquettes jaunes sont les points de détection. Les lettres G et P désigne la taille du spécimen, respectivement individu "grand" et "petit", " ?? " indique que la taille n'est pas mesurable du fait d'un Rapport Signal sur Bruit trop faible, ou d'un échantillon trop court [Abeille 14].

6.2.3 Installation et enregistrements BOMBYX au sud du PNPC (sanctuaire PELAGOS)

Le projet BOMBYX a consisté en l'installation d'une bouée multimodale pour l'étude de la biodiversité et de la physique des océans dans les eaux du Parc National de Port-Cros (PNPC). Le projet Bombyx est piloté par LSIS Glotin en collaboration avec l'UMR MIO. Il a été cofinancé par TPM, CG, IUF et UTLN, et est activement soutenu par le PNPC depuis que le projet a été posé en 2010 à la Fondation de Monaco et non financé alors, puis reposé en 2011 à TPM. Il aura fallu deux ans pour le réaliser, et nous remercions tous les acteurs ayant participé à sa réussite. La bouée Bombyx est en subsurface (figure 6.5) à -26 m, ancrée par 280 m de fond, sur une crête qui surplombe 2500 m d'abysse, entre Porquerolles et Port-Cros, à 1 mile au Sud. C'est une zone à fort courant, fréquentée par une faune très riche, dont des cachalots. La structure (figure 6.4) est équipée d'un courantomètre, d'une mini caméra et d'un module acoustique équipé d'un hydrophone. La partie batterie a été conçue par CÉSIGMA, la Bouée par IXSURVEY et l'hydrophone par CETACEAN RESEARCH. L'acquisition est assurée par une "Beagle board" via un script conçu par l'équipe DYNI/LSIS [Descudé 14]. Ce projet profite aux UMR CNRS LSIS (porteur du projet) et au MIO, ainsi qu'au PNPC pour la surveillance des cétacés sur la zone. j'ai participé à la conception sur papier et en atelier ainsi qu'au déploiement sur le terrain de la structure. J'ai pu traiter les premières données acoustiques dans mon programme de thèse qui a duré exactement 36 mois. Une première session de relève des données a été effectuée fin octobre 2014 et nous comptons traiter les signaux en partie selon les méthodes développées dans cette thèse pour évaluer l'éloignement à l'hydrophone des détections, segmenter des signaux de types voisement et compresser les données. BOMBYX est l'aboutissement du travail de cette thèse en région PACA.



FIGURE 6.4 – Structure BOMBYX sur le pont du bateau opérationnel de IXSURVEY. Crédits photo : Gilles Rougier.

Une analyse préliminaire a montré que BOMBYX a fonctionné en continu et valide la qualité de l'infrastructure (très peu de bruit généré par la structure). La richesse des enregistrements en terme de source bioacoustiques confirme notre choix de la localisation géographique du dispositif. En effet nous obtenons un contact acoustique avec des cachalots et des dauphins dès les premières



FIGURE 6.5 – Image de BOMBYX après déploiement (coordonnées GPS : 42.56.205 et 6.19.066).
Crédits photo : Gilles Rougier.

heures de l'installation. Des échantillons sonores sont disponibles sur le site <http://sabiod.org>.

6.2.4 Enregistrements de chants de baleine à bosse à Madagascar

Des enregistrements de chanteurs ont été réalisés dans le canal de Sainte Marie (Madagascar) en 2007, 2008 et 2009, fournissant plus de 50 h de chants. Ce travail de terrain se fait en étroite collaboration avec l'association Cetamada (Madagascar⁵). Nous avons utilisé un hydrophone Colmar Italia GP280 (omnidirectionnel sur $[5 \text{ Hz}; 90 \text{ kHz}]$, sensibilité $-170 \text{ dB-re } 1 \text{ V}/\mu\text{Pa}$, fiche technique sur www.colmaritalia.it), déployé à partir d'un bateau à moteur (moteur éteint) placé devant les chanteurs ($\approx 100 \text{ m}$) à 20 m de profondeur (la profondeur de la colonne d'eau se situait entre 40 m à 50 m). Les données ont été numérisées par l'enregistreur Tascam HD-P2 avec une fréquence d'échantillonnage de 44,1 kHz sur 16 bits. La figure 6.6 illustre la succession de 4 unités sonores. Le spectrogramme montre la présence de la fréquence fondamentale (120 Hz), puis des harmoniques (F1 à 800 Hz, F2 à 1600 Hz, F3 à 2400 Hz...).

5. <http://cetamada.com>

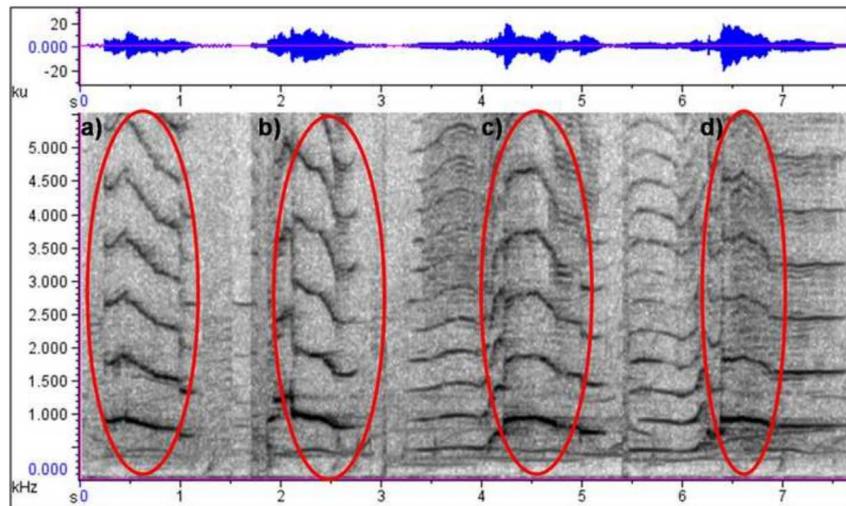


FIGURE 6.6 – Spectrogrammes de quelques segments de chant de Pace et al. 2010, réalisés à partir du même chant que celui utilisé dans la thèse. On aperçoit une même sous-unité (entourée) dans différents contextes de chant.

6.2.5 Enregistrements et base de données "St Pierre et Miquelon" SPM

Nous avons eu accès à des enregistrements provenant de l'association SPM Frag'iles dont l'une des missions est de contribuer aux suivis des différentes populations de cétacés évoluant autour de l'archipel de St Pierre et Miquelon. Le choix de l'association SPM Frag'iles s'est porté sur un dispositif à déployer sur le fond marin pour au moins 2 raisons : d'une part les conditions météo dans la zone d'intérêt incitent à ne pas disposer de matériel en surface, et d'autre part l'accès temps réel aux données n'étaient pas une motivation de cette étude. Il s'agit d'un dispositif fixe d'acquisition de type AURAL-M2 .

Caractéristiques principales de l'AURAL-M2

L'AURAL-M2 est un enregistreur acoustique autonome (figure 6.7). L'hydrophone unique est relié à une carte d'acquisition qui convertit les intensités acoustiques en données numériques, alors enregistrées sur un disque dur. L'alimentation électrique est assurée par 128 piles placées dans le compartiment étanche.

L'hydrophone utilisé pour l'AURAL-M2 est un HTI-96-MIN. Cet hydrophone a une sensibilité de $-165dBV/\mu Pa$. Lors de son utilisation, il a été ajouté un gain de 22 dB. La formule donnant l'intensité acoustique (en décibel), à partir des intensités en V est : $IntR = Vrec - (-165) - 22 = 10 \log 10(x^2) + 143dB$

Ainsi, si le signal est saturé ($x = 2V$), on aura en intensité max enregistré : 149 dB. Autrement dit, un signal d'intensité supérieure à 149 dB fera saturer l'hydrophone.

D'autre part, le constructeur indique qu'il faut tenir compte d'un bruit électronique qui masquera tous les sons d'intensité inférieure à 59 dB.

La base de données et contenu bioacoustique des enregistrements

La base de données correspond à 8509 fichiers acquise par les 2 hydrophones au cours de 2 saisons d'enregistrements, correspondant, en cumulé, à 353 jours d'observations acoustiques, soit



FIGURE 6.7 – Schéma du dispositif d'acquisition AURAL utilisé par l'association Frag-iles.

5158 h d'enregistrements (non continus) (voir tableau 6.8).

2010	Date du premier fichier	Date du dernier fichier	Durée d'enreg. / fichier	Nb total de fichiers	Nb total de jours	Nb d'heures cumulées
AURALA	19/8 à 17h	2/11 à 23h	45min	1807	75	1355h15 min
AURALB	19/8 à 17h	2/11 à 23h	45min	1807	75	1355h15 min
			sous-total	3614	150	2710h30 min
2011						
AURALA	23/4 à 13h01	23/7 à 12h01	30min	2184	91	1092h
AURALB	25/4 à 20h	16/8 à 18h	30min	2711	112	1355h30 min
			sous-total	4895	203	2447h30 min
			Total	8509	353	5158h

FIGURE 6.8 – Bilan des enregistrements SPM avec les 2 hydrophones en 2010 et en 2011.

Les enregistrements acoustiques contiennent les intensités des sources sonores autour des hydrophones. Ils contiennent des sons d'origine naturelle et biologique et potentiellement d'activités humaines. Ils contiennent également des sons d'origine diverse, dont ceux liés aux systèmes d'enregistrements.

Pour analyser ces enregistrements, nous avons segmenté la bande fréquentielle exploitable

délimitée par les caractéristiques de l'hydrophone et la fréquence d'échantillonnage sélectionnée en 3 sous-bandes :

- 1ère sous-bande : $[10 \text{ Hz}; 1 \text{ kHz}]$. Cette première sous-bande est dédiée à la recherche de sons émis par des mysticètes, en particulier les grands rorquals (baleines bleues). Un filtre de Butterworth passe-bas d'ordre 12 avec une fréquence de coupure de 1kHz a été utilisé.
- 2ème sous-bande : $[1 \text{ kHz}; 8 \text{ kHz}]$. Cette 2e sous-bande est utilisée pour mettre en évidence des sons harmoniques, type vocalises de rorquals, d'odontocètes, voire certains sifflements. Deux filtres consécutifs de Butterworth ont été appliqués : tout d'abord, un filtre passe-haut d'ordre 12 de fréquence de coupure 1 kHz, puis un filtre passe-bas d'ordre 12 de fréquence de coupure 8 kHz.

Sur les enregistrements continus pendant ces années, nous retrouvons aussi bien au nord qu'au sud des émissions sonores de différentes espèces de cétacés. La base de données collectée par les 2 AURAL sur les 2 années consécutives est riche en émissions sonores de cétacés (tableau 6.1).

		total fichiers	orque	cachalot	odontocète clics	odontocète buzz	odontocète sifflements	Mysticète
2010	A	1807	1	28	27	180	668	12
	B	1807	0	118	3	32	268	7
2011	A	2184	0	71	7	29	154	37
	B	2711	1	164	7	30	227	31

TABLE 6.1 – Émissions sonores des cétacés.

Dans le cadre de cette thèse nous retenons particulièrement les détections de cachalots dont un extrait est donné en figure 6.9.

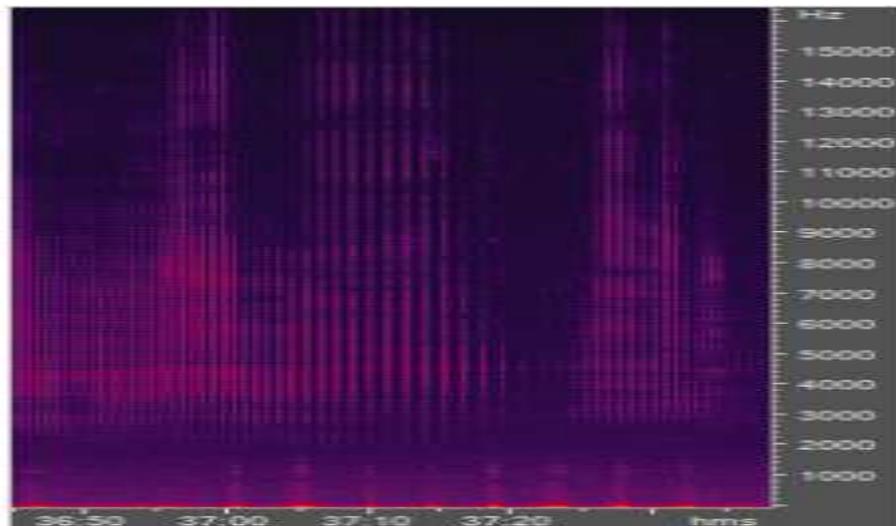


FIGURE 6.9 – Clics de cachalot. Extraits d'AuralA, A32C0299, 1/9/2010, 04 :37min à 04 :39min.

6.3 Systèmes multicateurs

6.3.1 Dispositif autonome multicateurs BAOBAB

Pour observer une zone large ou procéder à des estimations de localisation de sources sonores, la solution du réseau d'hydrophones peut être privilégiée.

Entre 2007 et 2011, au cours de la saison de reproduction, des baleines à bosse mâles chanteurs étaient enregistrées au large de l'île de Ste Marie (Côte Nord Est de Madagascar) à l'aide d'un unique hydrophone mobile (ColmarItalia GP0280) et numérisées par un enregistreur numérique TASCAM HD-P2 (1 voie, 44 kHz, 16 bits). L'intérêt de cette approche était de fournir des enregistrements de chanteurs isolés malgré la nécessité d'utiliser un bateau pour se rendre sur zone, et de réserver un temps d'effort de recherche pour localiser le chanteur et s'approcher au plus près (< 100 m) pour procéder à l'enregistrement. Dans le but de couvrir la partie Sud du Canal Ste Marie (hot-spot de la côte Est de Madagascar particulièrement fréquentée par les baleines à bosse pendant la période de reproduction), nous avons l'intention de déployer un réseau d'hydrophones. Cette initiative s'est insérée dans le volet acoustique du projet Balise et Acoustique pour l'Observation des Baleine A Bosse (BAOBAB) financé par la Fondation Total. Plusieurs plans ont été proposés :

1. Nous pouvions recourir aux enregistreurs conçus par CeSigma, et même s'ils procèdent à des enregistrements de façon indépendante, il est possible de procéder à la synchronisation des différentes voies, soit via un son relativement fort (détonation) émis à distance connue, soit *a posteriori* en utilisant l'horloge universelle GPS. Cette solution a été déployée en 2012, d'une part pour s'assurer de la faisabilité de cette approche et d'autre part pour commencer une base de données de ces chants à distance.
2. Nous avons également envisagé de déployer un module central équipé de 5 hydrophones en étoile placé à 10 km de distance de la côte ouest de l'île Ste Marie. Ce plan, même s'il était attractif, présentait l'inconvénient de nécessiter de se rendre sur place régulièrement pour procéder à la récupération des données et éventuellement à s'assurer de l'alimentation électrique. La solution par câble nous semblait difficile à déployer (déployer 10 km de câble nécessite un bateau spécial), représentait un coût financier important et nécessiterait un entretien constant qui lui aussi pourrait finalement se montrer onéreux à moyen terme.

Finalement, nous avons retenu les motivations suivantes : a) une faible bathymétrie, des propriétés de propagation acoustique de petits fonds et la possibilité de couvrir une grande partie du canal (< 30 km), b) un système qui n'est pas temps réel, c) une autonomie du système de 3 jours, nécessitant une intervention pour la récupération des données d) éviter la perte de la bouée de surface compte tenu des mauvaises conditions de mer durant l'hiver austral (saison de présence des animaux), e) éviter les collisions avec les embarcations même si le trafic est faible dans le canal, principalement des pirogues locales et bateaux de taille moyenne destinés au "whale watching".

En 2013, nous avons décidé de déployer une station permanente constituée de 3 sous réseaux d'hydrophones le long de la côte ouest de l'île de Ste Marie. Les réseaux sont indépendants, autonomes et équipés chacun de 3 hydrophones autour d'un système central alimenté par liaison câblée vers la côte (voir figure 6.10). L'ensemble est placé sur le fond aux profondeurs respectives pour les 3 sites de 18 m, 12 m et 8 m (du sud vers le Nord). la distance maximale à la côte est de 700 m et les 3 sous réseaux sont espacés de 20 km et 30 km. Les hydrophones sont reliés au boîtier central par des câbles de 100 m de long. Les données précises sont récapitulées dans la table 6.2.

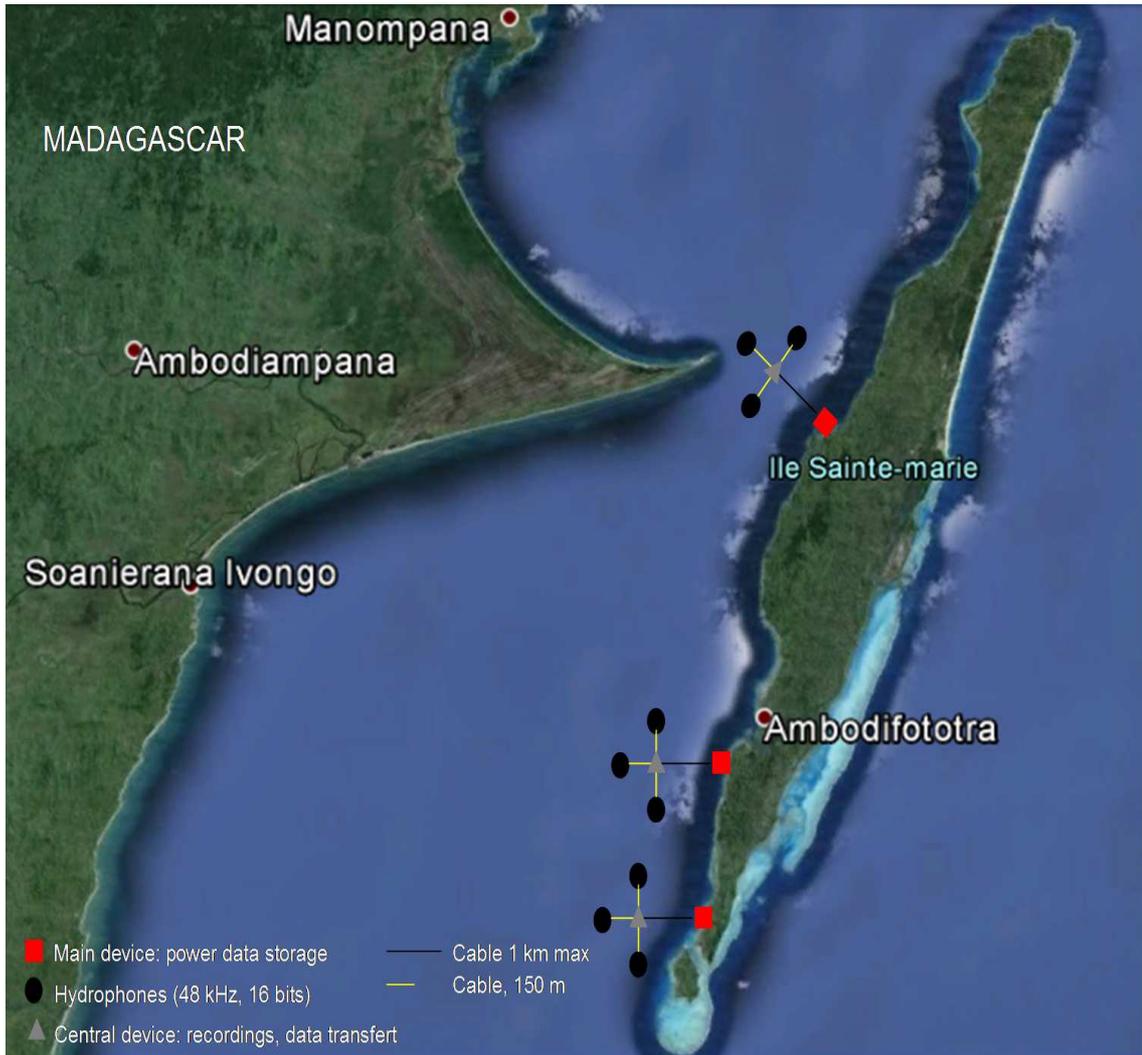


FIGURE 6.10 – Situation géographique des 3 sous réseaux d’hydrophones le long de la côte Ouest de Ste Marie.

situation	système central		hydro Sud		hydro Ouest		hydro Nord	
	GPS	distance à la côte	GPS	distance système central	GPS	distance système central	GPS	distance système central
du Sud au Nord	17 ° 04'48.8"	700		100		100		100
	49 ° 48'43.5"							
	17 ° 00'36.2"	300	17 ° 00'39.7"	100	17 ° 00'34.0"	100	17 ° 00'33.7"	100
	49 ° 50'02.4"		49 ° 50'02.2"		4 ° 49'59.3"		49 ° 50'03.7"	
	16 ° 52'22.26"	275	16 ° 52'20.88"	100	16 ° 52'19.08"	100	16 ° 52'21.18"	100
	49 ° 53'35.52"		49 ° 53'32.46"		49 ° 53'35.4"		49 ° 53'37.92"	

TABLE 6.2 – Données géographiques des 3 sous réseaux d’hydrophones.

Chaque sous élément du réseau a la même architecture électronique. Pour minimiser les risques d'interruption de l'ensemble du dispositif au cas où une défaillance surviendrait sur l'un des sous réseaux, chacun est indépendant vis à vis de l'alimentation et du stockage des données. Ils sont constitués de 3 sous parties ou modules :

1. C1 (voir figure 6.11), le dispositif central, contenu dans un boîtier étanche est articulé autour d'une carte d'acquisition, d'une alimentation électrique et d'un modem de transmission de données numériques synchronisées. La Carte comprend un pré-amplificateur à faible émission de bruit, le gain peut être fixé manuellement entre -95 dB et $+31.5$ dB. La fréquence d'échantillonnage maximale est de 250 kHz. Les données sont numérisées sur 16 ou 24 bits. La bande de fréquence utilisée par les baleines à bosse est inférieure à 20 kHz, donc nous échantillons le paysage acoustique à $F_s = 48$ kHz. Les hydrophones sont connectés à ce boîtier *via* des prises SubConn. Une prise de plus a été prévue pour brancher un quatrième hydrophone.
2. C2 (voir figure 6.12) est dédié à l'alimentation énergétique en courant alternatif et comprend un système WIFI 802.11g. Ce module est connecté d'un coté au dispositif central (700 m maximum), de l'autre coté au module principal C3 (150 m maximum).
3. C3 (voir figure 6.13), le module principal, comprend un convertisseur AC/ DC, un ordinateur (2 Go RAM, 2 ports SATA, 1 liaison Ethernet 1 Gbps, 1 carte graphique), un GPS pour dater et synchroniser les fichiers avec le signal PPS (précision de synchronisation $< 1 \mu s$, un disque dur externe de 1 To pour le stockage des données. Le disque dur peut être facilement changé lorsqu'il est plein. 1 To représente le volume de données acquises sur environ 30 jours. Ce module comprend aussi une alimentation permanente (onduleur UPS) de 600 W. Cette mesure est prise pour faire face aux variations de tension délivrée par le courant de secteur public à ste Marie. Le transfert de données entre C1 (boîtier sous marin) et le module principal C3 sur la côte est opéré via le câble d'alimentation par protocole Ethernet (22 Mbits/s). Nous avons aussi ajouté une connection WIFI (150 Mbps) en alternative de la connection Ethernet. Le logiciel DRS 2000 installé sur un système d'exploitation Linux organise le stockage des données en fichiers bruts (.dat), la communication et les configurations des différents modules.

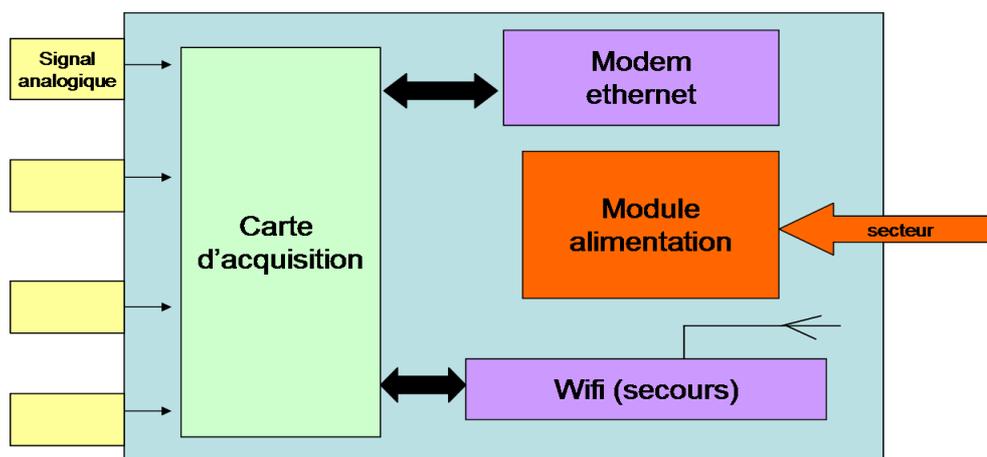


FIGURE 6.11 – Module C1.

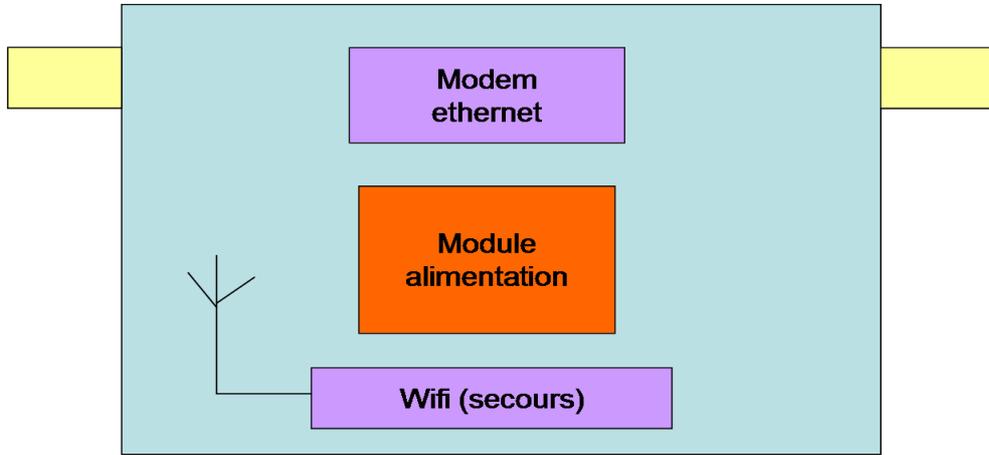


FIGURE 6.12 – Module C2.

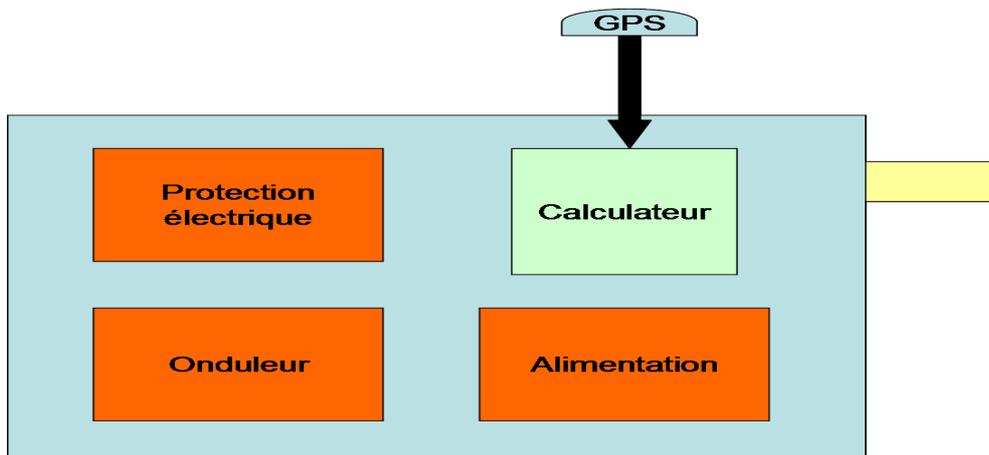


FIGURE 6.13 – Module C3.

6.3.2 Base de données "Bahamas"

Origine des données et conditionnement du signal

Le jeu de données "Bahamas" a été distribué par l'AUTEC à l'occasion du second workshop DCL à Monaco en 2005. Il est composé de 2 sets. Le set A est composé d'enregistrements provenant de 5 Hydrophones déployés au large des Bahamas. Pendant 25 minutes, un cachalot émettant des clics évolue à travers le réseau d'hydrophones. Le set B provient de 20 minutes d'enregistrements de 6 autres hydrophones lors desquels 4 animaux se déplacent en cliquant simultanément. Les caractéristiques de la chaîne d'acquisition sont inconnues hormis la fréquence d'échantillonnage fixée à 48 kHz.

La détection des clics et l'extraction des temps d'arrivée sont effectuées à l'aide d'un détecteur quadratique. Le signal brut est de qualité variable et présente parfois de faibles RSB suivant l'hydrophone considéré. En milieu marin, les causes de bruit sont multiples : bruit de fond, trafic maritime, houle, poissons, bruit de chaîne et coque... On admet que le bruit de fond est principalement basse fréquence. Nous filtrons donc le signal par un filtre passe-haut dont la fréquence de coupure est comprise entre 1000 Hz et 5000 Hz dans le but de faire suffisamment émerger les clics du bruit. Notre stratégie pour la reconnaissance d'un clic consiste à parcourir le signal par fenêtres glissantes de durée d'environ 40 ms et à y calculer le niveau d'énergie. La comparaison avec un niveau d'énergie seuil et un étalement spectral de référence valide la détection ou non. Le clic est recentré dans la fenêtre par rapport à son amplitude maximale, puis stocké avec son temps d'arrivée. Le signal brut montre alternativement un premier clic suivi d'un deuxième clic régulier d'amplitude 4 à 5 fois plus faible. Il présente une différence de temps d'arrivée en moyenne d'environ 0.5 s ce qui correspond à une différence de marche des trajets acoustiques d'environ 800 m. Nous l'interprétons comme une réflexion du signal sur la surface de l'eau ou sur le fond. Ces événements sont volontairement retirés des détections mais peuvent être récupérés si besoin.

Set A à 1 cachalot

La trajectoire calculée par l'équipe LSIS-DYNI (figure 6.14) [Giraudet 06b, Bénard 11b] concorde avec celles trouvées par d'autres membres de la communauté scientifique ayant participé au challenge [Morrissey 06, Nosal 06]. Cette trajectoire et les données cinématiques associées seront retenues comme vérité terrain. Une illustration supplémentaire est disponible avec l'animation 3D de la trajectoire en ligne sur internet⁶. Les fichiers contenant les positions $[x, y, z, t]$ et les angles de position $[az, el, offaxis, t]$ sont également disponibles pour chaque hydrophone.

Pour l'ensemble des hydrophones H8, H9, H10 et H11, l'intervalle des distances entre source et récepteur s'étale entre 2500 m et 5500 m avec une distribution donnée par la figure 6.16. La figure 6.15 montre la dynamique de la distance radiale sur 25 minutes.

Les angles précis azimut a et élévation el associés à la cinématique de la trajectoire ont été extraits. Leur répartition est présentée en figure 6.17. Nous notons que les distributions ne sont pas uniformes. La configuration "on axis" semble plus fréquente dans ce jeu de données. Les valeurs moyennes et écart types des distributions sont récapitulés en table 6.3.

Le set A constitue un jeu de données simple (1 seul animal) et sera principalement pour le développement de nos méthodes d'estimation de distance radiale. Nous partitionnerons par la suite les données. Une première partition sera consacrée à l'apprentissage, une deuxième partition sera dédiée au développement et à l'optimisation des paramètres régissant les modèles proposés.

6. <http://glotin.univ-tln.fr/oncet> et <http://www.youtube.com/watch?v=0Szo3gdiTRk>

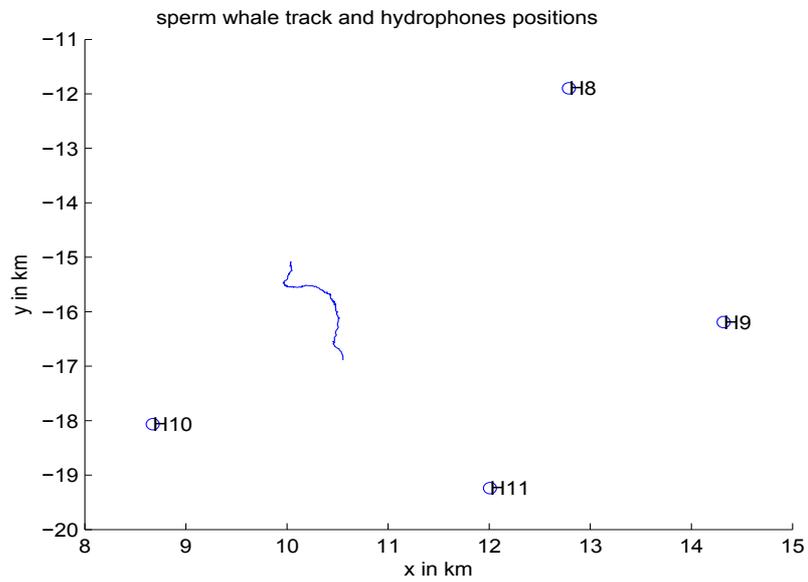


FIGURE 6.14 – Trajectoire 2D (dans le plan $x-y$) d'un cachalot isolé et position des hydrophones. Il est observé pendant 25 min (LSIS/DYNI [Giraudet 06b]). L'animal se dirige vers le Sud-Est.

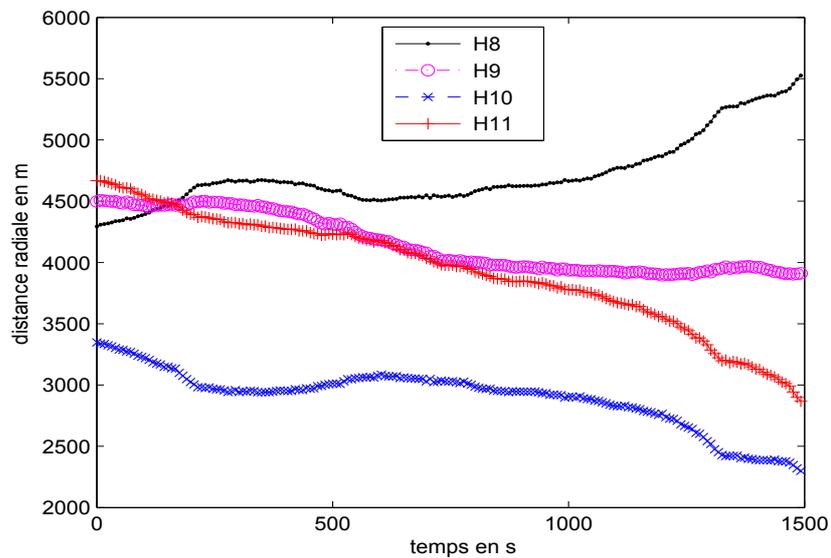


FIGURE 6.15 – Dynamique de la distance radiale vérité terrain suivant H8, H9, H10 et H11.

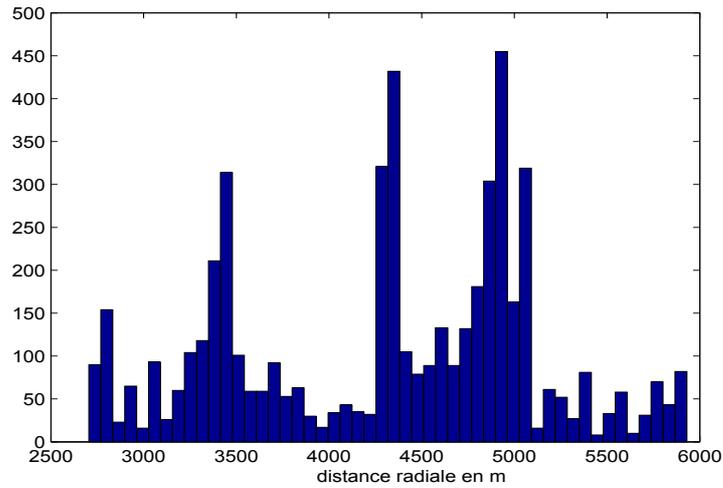


FIGURE 6.16 – Histogramme des distances radiales vérité terrain sur les 25 min.

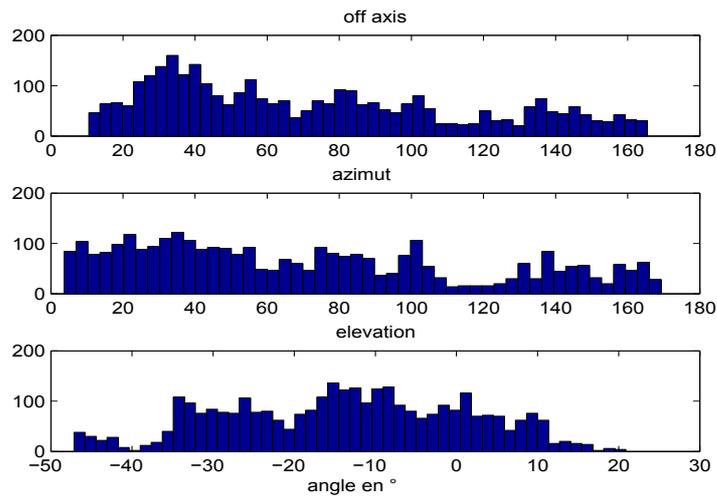


FIGURE 6.17 – Histogramme des angles vérité terrain entre l'animal et les hydrophones.

	H11	H10	H9	H8
profondeur hydrophone	-1522 m	-1361 m	-1553 m	-1556 m
off axis moyen	36 degrees	60 degrees	74 degrees	125 degrees
azimut moyen	29 degrees	56 degrees	73 degrees	129 degrees
elevation moyenne	-14 degrees	-15 degrees	-14 degrees	-13 degrees
distance moyenne	3937 m	2900 m	4150 m	4716 m
écart type distance	452 m	242 m	234 m	283 m
vitesse radiale moyenne	-1.13 m/s	-0.7 m/s	-0.4 m/s	0.8 m/s

TABLE 6.3 – synthèse des données terrain : distance moyenne, azimuth, élévation, off axis, profondeur des hydrophones.

Enfin une dernière partition sera réservée aux estimations/prédictions finales.

Set B à 4 cachalots simultanés

Nous présentons en figure 6.18, la trajectoire définitive proposée par l'équipe LSIS-DYNI [Glotin 08a] 3 ans après la mise en ligne des données. La trajectoire a été comparée avec les résultats obtenus par le SOEST (Hawai), seul laboratoire à avoir proposé une solution aboutie. Les trajectoires des 2 laboratoires concordent fortement et donnent des vitesses moyennes de déplacement de l'ordre de 10 km/h. Les 4 animaux cliquent simultanément et s'orientent vers le Nord-Est en empruntant des trajectoires clairement dissociées. La problématique de séparation de sources au niveau des clics constitue toujours un réel challenge pour les participants. Les hydrophones sont fixés sur le fond océanique. Le tableau 6.4 donne les profondeurs des hydrophones.

	H1	H2	H3	H4	H5	H6
profondeur hydrophone	-1530 m	-1556 m	-1553 m	-1361 m	-1522 m	1530 m

TABLE 6.4 – Profondeur des hydrophones pour le set B

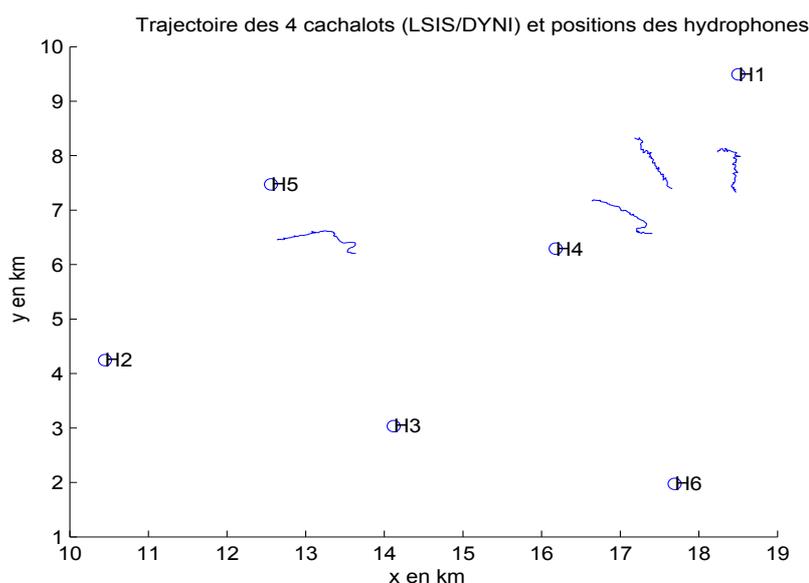


FIGURE 6.18 – Trajectoire 2D (dans le plan $x - y$) de 4 cachalots et position des hydrophones. Les animaux évoluent séparément durant 20 min vers le Nord-Est (LSIS/DYNI [Glotin 08a]).

Les figures 6.19 et 6.20 présentent les dynamiques des distances radiales extraites des trajectoires retenues selon les hydrophones 5 (de 2000 à 5000 m) et 6 (de 4500 à 7000 m). Même si les trajectoires 3D paraissent bien séparées, les trajectoires dans le plan temps/distance peuvent se retrouver presque confondues. Pour cette étude nous nous baserons sur les hydrophones H5 et H6. Le tableau 6.5 donne les distance moyennes et les vitesses radiales moyennes entre l'hydrophone et l'individus (4 baleines). Elles sont négatives sur H5, les cachalots s'approchent de l'hydrophone. Elles sont positives sur H6, les animaux s'en éloignent.

	SW1	SW2	SW3	SW4
H5	-0.41 m/s	-0.83 m/s	-0.65 m/s	-0.36 m/s
H6	0.7 m/s	0.54 m/s	0.85 m/s	0.58 m/s

TABLE 6.5 – Vitesse radiale moyenne entre les hydrophones et chaque cachalot (SW).

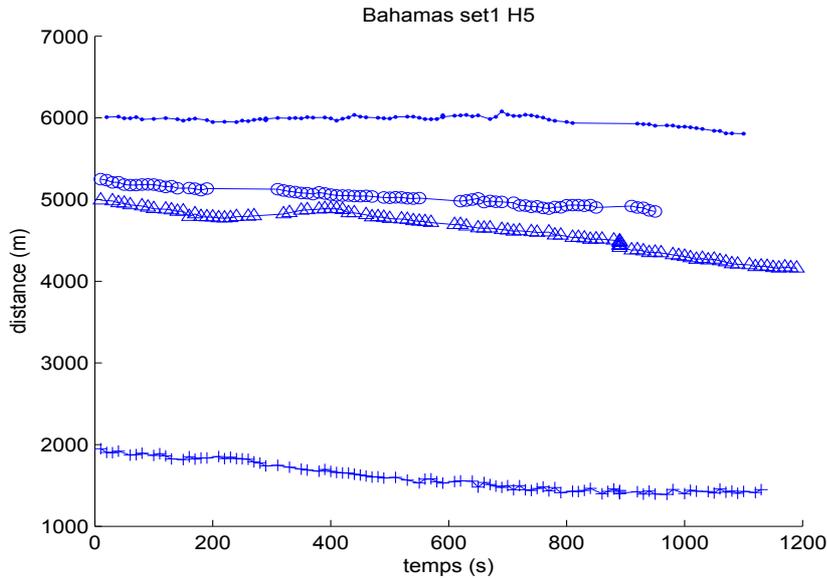


FIGURE 6.19 – Dynamique de la distance radiale de la vérité terrain suivant H5 pour les 4 cachalots.

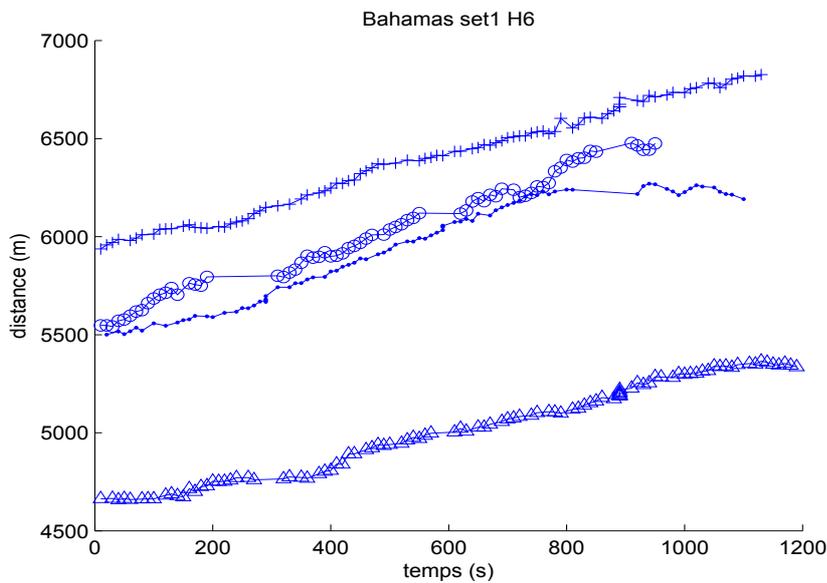


FIGURE 6.20 – Dynamique de la distance radiale de la vérité terrain suivant H6 pour les 4 cachalots.

6.4 Conclusion

Dans ce chapitre nous avons décrit les systèmes de mesure acoustique développés par l'équipe de recherche et le partenaire industriel CeSigma. Un premier projet en Guadeloupe (GUALIBA), initié avant le début de la thèse, avait pour but la conception et le déploiement d'une bouée mono-hydrophone autonome. Un second projet (BAOBAB) sur lequel j'ai participé pleinement concernait le suivi acoustique des baleines à bosses dans le canal de Sainte Marie (Madagascar) par le déploiement de réseaux multi-hydrophones conçus par CeSigma. Ce projet est unique dans l'océan indien est continue actuellement en 2014. Un dernier projet (BOMBYX) constitue le support d'application de ce travail de thèse en région PACA. Les données ont été relevées fin octobre 2014 et ne sont donc pas toutes traitées dans ce manuscrit.

Enfin, nous avons décrit le corpus de données dont nous disposons pour valider les modèles d'estimation de distance et notre méthode de représentation de voisements mises au point lors de ma thèse. La base de données BAHAMAS SET A et B est précieuse pour nous dans la mesure où nous possédons les trajectoires des baleines estimées par l'équipe LSIS/DYNI via un procédé de trajectographie multicapteurs et validées par d'autres équipes de recherche. Nous listons ici (tableau 6.6), les sites et caractéristiques de ces données (BAOBAB compris).

base de données	lieux	caractéristiques	espèces présentes
BAOBAB 2007-2009	Madagascar	1 voie (TASCAM) $F_s = 44.1$ kHz 16 bits	baleine à bosse
BAOBAB 2012	Madagascar	2voies $F_s = 48$ kHz 16/24 bits	baleine à bosse
BAOBAB 2013	Madagascar	3×3 voies $F_s = 48$ kHz 16/24 bits	baleine à bosse
BAHAMAS SET A	Bahamas	5 voies $F_s = 48$ kHz -	cachalot (1 animal isolé)
BAHAMAS SET B	Bahamas	5 voies $F_s = 48$ kHz -	cachalot (4 animaux simultanés)
DECAV	PNPC	1 voie (LINUX) $F_s = 48$ kHz 16 bits	cachalot
BOMBYX	PNPC	1 voie (LINUX) $F_s = 44.1$ kHz 16 bits	cachalot/dauphin
SPM	st Pierre et Miquelon	2×1 voies $F_s = 32.768$ kHz 16 bits	mysticètes/odontocètes

TABLE 6.6 – listes des bases de données disponibles.

Troisième partie

Validations expérimentales et
discussions

Validation des modèles de prédiction de distance sur signaux transitoires de cachalot

Sommaire

7.1	Estimation de la distance radiale par le modèle ISAT	88
7.1.1	Optimisation des hyper-paramètres P et N_{best}	88
7.1.2	Estimation de la distance de propagation $ISAT$ sur le jeu de données de test	88
7.1.3	Information mutuelle entre les rapports de bandes d'énergie et la distance source-récepteur	92
7.1.4	Sélection des combinaisons de bandes de fréquence basée sur l'Information Mutuelle	93
7.2	Estimation de distance par régressions LER, ISAR et ISARH .	95
7.2.1	Choix des ensembles d'apprentissage, de test et de développement .	95
7.2.2	Optimisation des Hyper-paramètres	95
7.2.3	Prédiction de distance pour un mélange train/test uniforme	98
7.2.4	Prédiction de distance pour un mélange train/test dissocié dans le temps	100
7.2.5	Erreur et précision des prédictions ISAT, LER, ISAR et ISARH . .	102
7.2.6	Modèle de fusion entre $ISAT$ et $ISAR$	104
7.3	Test du modèle de fusion ISAT/ISAR sur le jeu de données Bahamas set B	106
7.3.1	Estimations de distance dans le temps	106
7.3.2	Traitement des prédictions par bi-histogrammes temps/distance . .	108
7.3.3	Répartition des prédictions	110
7.4	Résultats préliminaires sur les données issues du protocole DECAV	113
7.4.1	Estimations de distance dans le temps	113
7.4.2	Traitement des prédictions par bi-histogrammes temps/distance . .	114
7.4.3	Répartition des prédictions	114
7.4.4	Synthèse des résultats sur les données DECAV en présence d'un seul cachalot	115
7.4.5	Synthèse des résultats sur les données DECAV en présence de plusieurs cachalots	116

7.5	Résultats préliminaires sur un extrait des données BOMBYX	118
7.6	Conclusion	121

7.1 Estimation de la distance radiale par le modèle ISAT

7.1.1 Optimisation des hyper-paramètres P et N_{best}

Le modèle ISAT a été testé sur les données "Bahamas set A" dont la vérité terrain est disponible. Pour chaque hydrophone, nous avons estimé les distances animal-récepteur selon le processus décrit dans la partie contributions. En travaillant successivement sur des fenêtres de $N = 256$ et $N = 128$ échantillons, nous obtenons les estimations suivantes :

- sans utilisation de la vérité terrain : \hat{r} avec $N_{best} = 1, 2, 4, 8, 16$.
- sans utilisation de la vérité terrain : toutes les combinaisons sont sélectionnées $N_{best} = N_{band}$ (noté *all*).
- avec utilisation de la vérité terrain : \hat{r}^* (noté *optimal*).

Pour l'étape d'optimisation, nous utilisons les données de la partition dédiée au développement (ensemble "dev"). Nous avons calculé l'erreur relative entre la vérité terrain et l'estimation de distance pour différentes valeurs des paramètres. L'erreur moyenne (sur tous les hydrophones) la plus faible détermine le jeu d'hyper-paramètres optimal (P, N_{best}) . Notons que la dispersion de cette erreur est un facteur à prendre en compte. Nous considérons que l'erreur et sa dispersion sont acceptables si leur valeur reste inférieure à 25%. Le paramètre de lissage de spectre P est optimisé selon la moyenne de l'erreur relative absolue dont l'expression suit ci-dessous :

$$\epsilon_r(P) = \sum_{l=1}^L \sum_{j^l=1}^{N_{Wl}} \frac{|d_{j^l}^l - \hat{r}_{j^l}^l(P)|}{d_j} \quad (7.1)$$

Notons que P peut être converti en fréquence par $F_P = P\Delta f$.

Pour les estimations $N_{best} = N_{band}$, il n'y donc pas d'application du critère de minimisation de la variance. Cela nous permet de vérifier l'amélioration apportée par ce critère. En se référant à la figure 7.1, l'erreur est en moyenne 2 à 3 fois plus faible avec application du critère.

Les estimateurs employés ici admettent différentes valeurs optimales de P . Plus le nombre de candidats est élevé, plus ce paramètre doit être petit. Le moyennage sur plusieurs combinaisons semble aussi stabiliser la courbe d'erreur alors que son comportement paraît moins monotone (alternance de phases croissantes et décroissantes) pour $N_{best} = 1$.

L'estimateur qui présente la plus faible erreur sur la partition 1 est obtenu pour $N_{best}^* = 4$. La courbe d'erreur (figure 7.1) affiche un paramètre de lissage commun proche de $F_{P_{128}}^* = 8.5$ kHz. Les estimations réalisées à partir de $N = 256$ échantillons n'aboutissent pas à des erreurs plus faibles. Nous constatons une légère translation de l'extremum. Le paramètre de lissage optimal est donné pour $F_{P_{256}}^* = 9$ kHz.

7.1.2 Estimation de la distance de propagation ISAT sur le jeu de données de test

Nous présentons ici les estimations \hat{r}^* de distance finale dans le temps entre l'hydrophone 11 et le cachalot. Le calcul a été réalisé sur la partition de l'enregistrement dédiée au test de prédiction (ensemble test) selon les paramètres F_P^* et N_{best}^* optimisés sur le jeu de données de développement utilisé dans la section précédente.

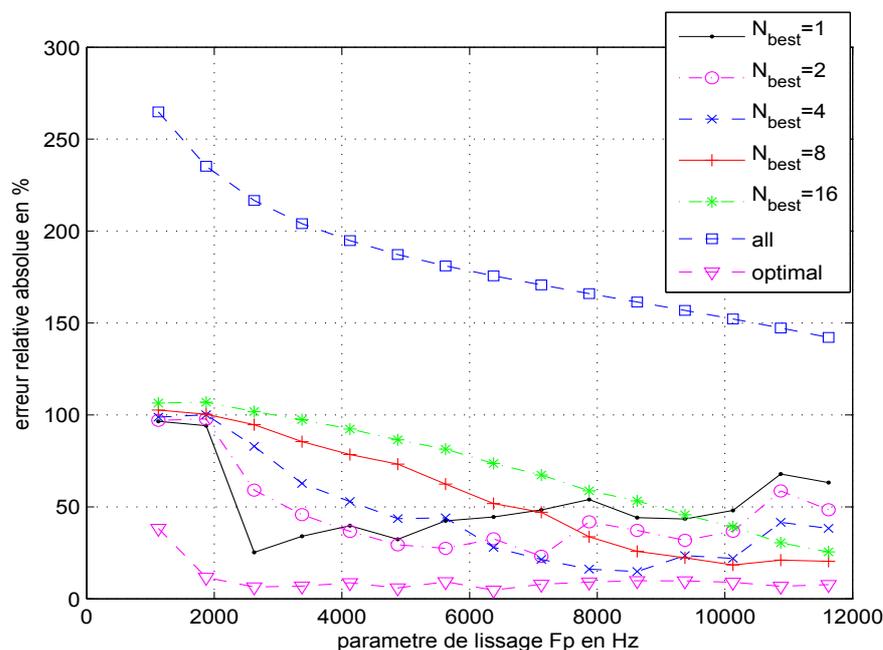


FIGURE 7.1 – Moyenne de l'erreur relative absolue entre la vérité terrain et l'estimation $N_{best} = 1, 2, \dots, 16, all$ (sur $N = 128$ échantillons) en fonction du paramètre de lissage des spectres en fréquence F_p . Moyenne sur les hydrophones H8, H9, H10, H11 de la partition dev.

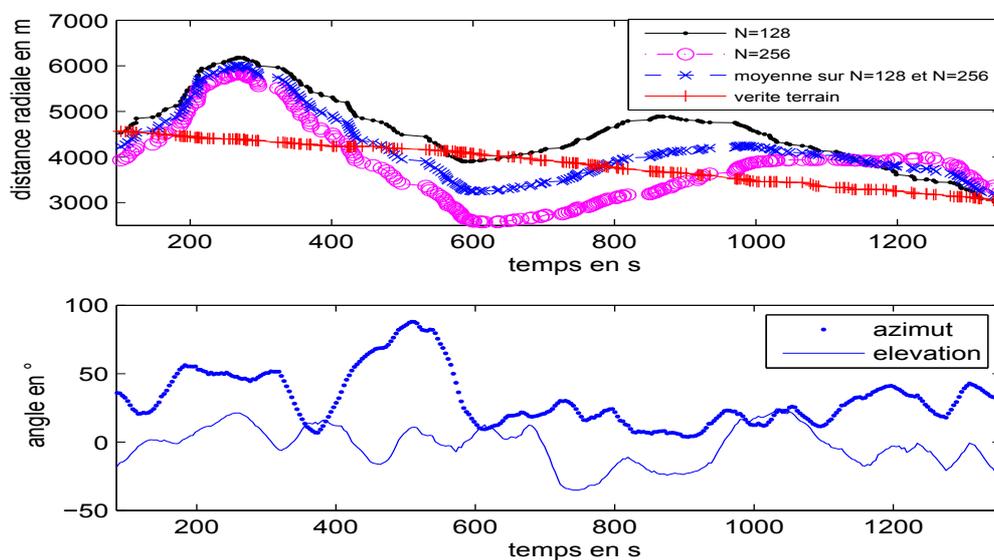


FIGURE 7.2 – Exemple d'estimations temporelles finales de la distance animal-récepteur avec le modèle ISAT, en comparaison avec la vérité terrain sur l'hydrophone H11. En bas l'évolution de l'azimut et de l'élévation (issus de la cinématique de la vérité terrain). Les paramètres fixés : $N_{best}^* = 4$, $F_{P_{128}}^* = 8.5$ kHz et $F_{P_{256}}^* = 9$ kHz.

	H11	H10	H9	H8	MEAN
ISAT $N = 128$	21	16	14	41	23
ISAT $N = 256$	20	21	9	16	16.5
moyenne sur les modèles $N = 128$ et $N = 256$	17	16	11	28	18
erreur moyenne relative	19	17	11	28	-

TABLE 7.1 – Moyenne de l’erreur relative absolue des prédictions pour les hydrophones H8, H9, H10 et H11 selon le modèle ISAT. L’erreur est en %.

Sur la figure 7.2, la dispersion des 2 courbes semble du même ordre. Utiliser $N = 128$ échantillons engendre des distances surestimées alors qu’utiliser $N = 256$ échantillons les sousstime. L’importante dispersion des estimations ne peut pas être due aux variations de distance physique. Elle pourrait être liée aux variations de l’angle de position de l’animal (off axis). Pour $N = 256$ échantillons, l’erreur semble s’accumuler sur les sections où l’azimut est élevé (> 20 deg). Lorsque l’hydrophone est dans l’axe de l’animal (on axis), les estimations ISAT semblent converger vers la vérité terrain. D’autre part, pour $N = 256$ échantillons, les erreurs les plus faibles sont localisées sur les portions où l’azimut est fort. Il en résulte de meilleures estimations pour les configurations off axis. La moyenne des 2 estimateurs $N = 128$ et $N = 256$ montre un comportement uniforme face à l’azimut. Le tableau 7.1 reprend les erreurs relatives absolues obtenus sur les hydrophones H8, H9, H10 et H11.

La pente globale des estimations de distance est décroissante dans le temps ce qui est en accord avec la vérité terrain. Le cachalot se rapproche de l’hydrophone H11. L’application de la méthode ISAT a permis de déterminer le sens de la trajectoire de l’animal.

Afin de pouvoir comparer les résultats avec les méthodes de régression par réseau de neurones, nous donnons en figure 7.3, les résultats détaillés pour tous les hydrophones à partir des données de l’ensemble train/test dissocié dans le temps (introduit dans section suivante). Les modèles utilisés sont paramétrés avec $N = 256$, $N_{best}^* = 4$ et $F_{P_{256}}^* = 9$ kHz. Deux lissages sont proposés. La différence majeure avec l’expérience précédente est que l’intégration des modèles se fait sur une dynamique de temps discontinue et plus courte (200 s contre 1500 s). On observe des effets de saut plus importants.

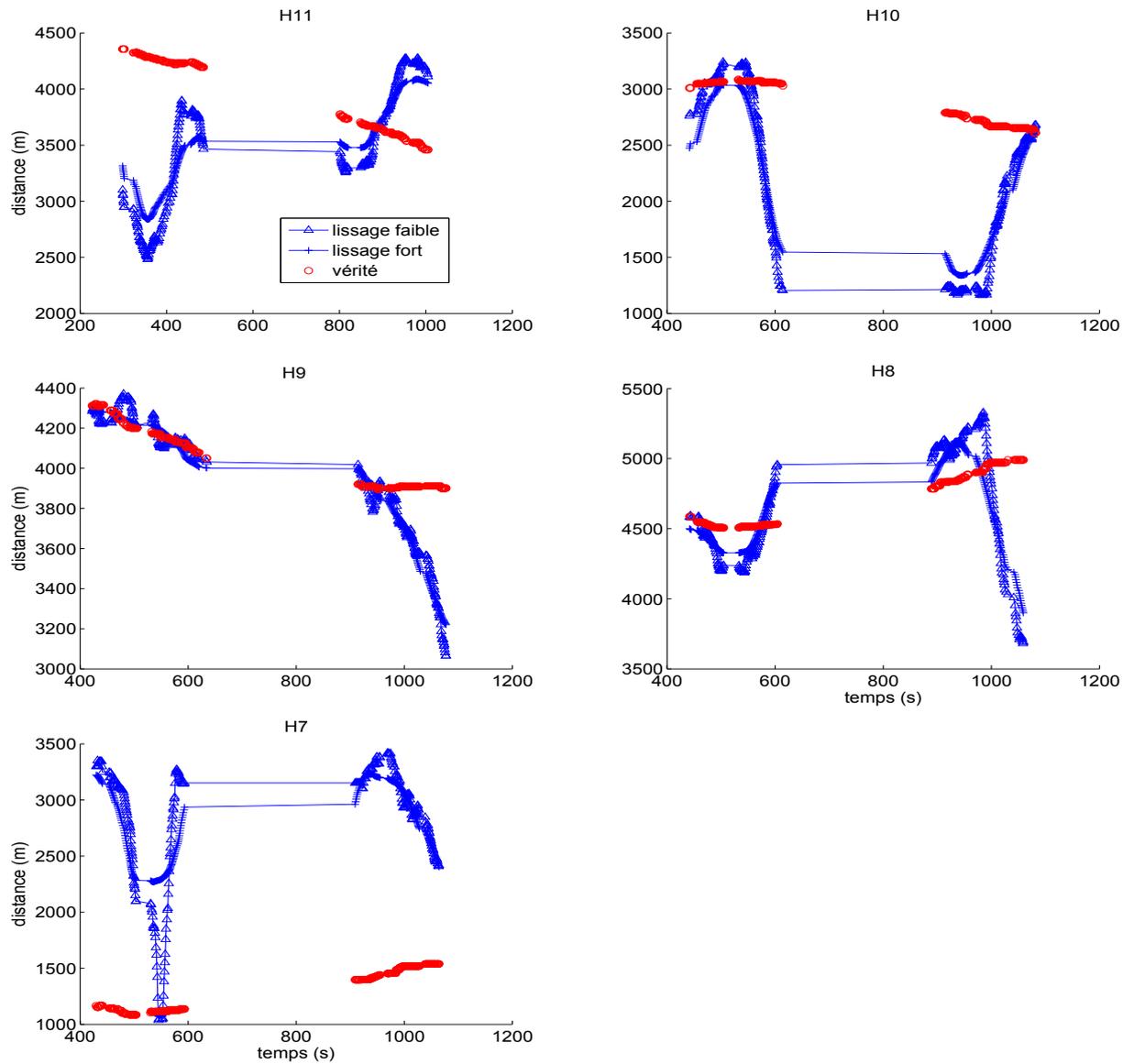


FIGURE 7.3 – Prédictions temporelles de distance selon le modèle ISAT optimal retenu et vérité terrain. Pour être comparé aux prédictions par réseau de neurones, les données test proviennent de l'ensemble train/test dissocié dans le temps.

7.1.3 Information mutuelle entre les rapports de bandes d'énergie et la distance source-récepteur

Afin d'évaluer la quantité d'information partagée par les rapports de bandes d'énergie et la distance de propagation, il est possible de calculer l'information mutuelle moyenne (IM) pour chaque combinaison $\theta \in N_{\text{band}}$. On considère l'espace des probabilités jointes $X^\theta Y$ tel que $X^\theta = \{\log 10(\mu_1^\theta), \log 10(\mu_2^\theta), \dots, \log 10(\mu_J^\theta)\}$ où μ_j^θ représente le ratio d'un couple θ de bandes d'énergie pour le $j^{\text{ième}}$ clic (issu de l'équation 3.14) et $Y = \{d_1, d_2, \dots, d_J\}$ l'ensemble des distances source-récepteur "vérité terrain" associé à la collection de clics. Les densités de probabilité jointes sont estimées par une répartition 2D normalisée entre les 2 variables. L'information mutuelle moyenne s'exprime alors ainsi :

$$IM(X^\theta Y) = \sum_{k=1}^J \sum_{j=1}^J P(\log 10(\mu_k^\theta), d_j) \log_2 \left(\frac{P(\log 10(\mu_k^\theta), d_j)}{P(\log 10(\mu_k^\theta)) P(d_j)} \right) \quad (7.2)$$

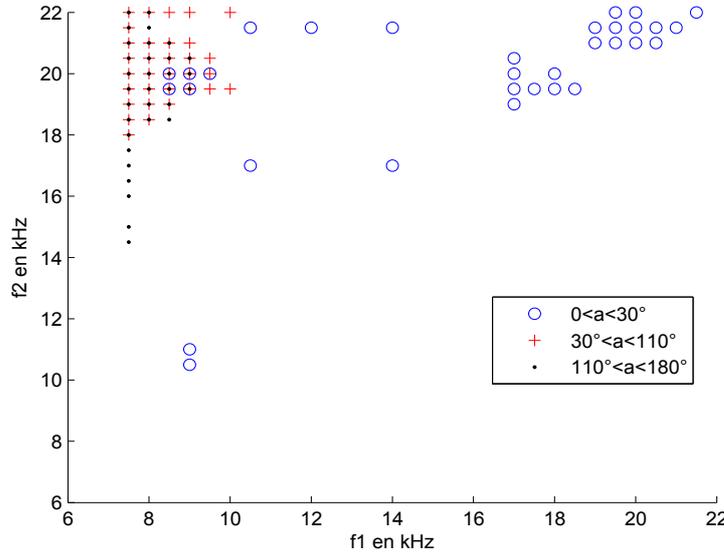


FIGURE 7.4 – Fréquences centrales des bandes mises en jeu pour les 32 plus fortes valeurs d'IM selon l'azimut.

Le calcul de l'IM est réalisé pour tous les ratios f_1/f_2 tels qu'ils ont été déterminés pour *ISAT* entre 7 kHz et 22 kHz avec $f_1 < f_2$. Nous nous intéressons aux 32 valeurs d'IM les plus fortes. La figure 7.4 permet de visualiser les bandes de fréquences impliquées dans les ratios concernés. Afin de discuter de l'impact des angles de position, 3 partitions du jeu de données en fonction de l'azimut sont construites : p1. "on axis" : $0 \text{ deg} < a < 30 \text{ deg}$, p2. "exposition latérale" : $30 \text{ deg} < a < 110 \text{ deg}$, p3. "exposition dorsale" : $110 \text{ deg} < a < 180 \text{ deg}$.

La première constatation est que les combinaisons partageant le plus d'IM avec la distance sont différentes selon la partition. Pour les clics "on axis", les rapports importants concernent des bandes dispersées sur l'ensemble du spectre avec 3 groupes : basses (BF 9 kHz), moyennes (MF 15 kHz) et hautes fréquences (HF 22 kHz). On trouve des rapports dont les bandes sont proches dans l'espace des fréquences (faible intervalle inter bandes (IIB)) pour toutes les régions

globalement avec une majorité sélectionnée vers les hautes fréquences. Notons que l'hypothèse de spectre plat avant propagation est moins erronée pour ce type de combinaisons.

De plus, on constate certaines combinaisons à fort IIB dont f_1 est basse fréquence et f_2 est haute fréquence. L'effet de déformation en fonction de la propagation est d'autant plus fort que les bandes sont écartées. La configuration de dos implique uniquement des rapports dont l'écart inter bandes est élevé tels que $7 \text{ kHz} < f_1 < 11 \text{ kHz}$ (BF/HF) et $14 \text{ kHz} < f_1 < 22 \text{ kHz}$ (MF/HF). La configuration latérale semble hybride et décale légèrement f_1 et f_2 vers les moyennes fréquences. "On axis" l'information est partagée par les faibles IIB. Dans l'article [Zimmer 04], les auteurs montrent que l'énergie est beaucoup plus uniformément répartie entre 1 kHz et 16 kHz pour les clics "on axis". A l'inverse de dos, le contenu basse fréquence est plus important. En effet, nous "perdons" les contributions HF des pulses très directifs et nous remarquons que les combinaisons communes aux 3 partitions sont dans cette région.

Ces résultats mettent en évidence les variations de la structure fréquentielle des clics avec l'orientation de l'animal, l'aspect directif du contenu HF et le caractère omnidirectionnel des composantes BF. En terme d'IM pour les clics hors d'axe, les ratios semblent tirer profit de cette disparité HF/BF.

7.1.4 Sélection des combinaisons de bandes de fréquence basée sur l'Information Mutuelle

Notre analyse de l'information mutuelle entre la distance et les bandes de fréquence nous a permis de constater que certaines combinaisons de bandes semblent indépendantes des angles de position du cachalot. Cela nous conduit naturellement à réviser la sélection des combinaisons de bandes de fréquences lors du calcul d'ISAT. Reprenons l'équation 3.14 qui décrit les estimations brutes $\hat{r}_{\theta,j}$ où j est l'indice du clic et θ est l'indice de la combinaison. Nous ne conservons alors que les K_{best} combinaisons θ^* telles que $8.5 \text{ kHz} < f_1 < 9.5 \text{ kHz}$ et $19 \text{ kHz} < f_2 < 20 \text{ kHz}$

Nous proposons donc une estimation moyenne sur l'ensemble des combinaisons retenues *ISAT2a*. Nous définissons :

$$\hat{r}_j = \frac{1}{K_{best}} \sum_{k=1}^{K_{best}} \hat{r}_{\theta_k^*,j} \quad (7.3)$$

Nous proposons également une solution constitué de la combinaison aboutissant à la variance minimale des estimations de distance que nous nommons *ISAT2b* :

$$\hat{r}_j = \hat{r}_{\theta^{**},j} \quad (7.4)$$

tel que : $\theta^{**} = \arg \min \sigma(\hat{r}_{\theta^*,j})$

Les représentations temporelles des prédictions sont présentées en figure 7.5 et les erreurs "Root Mean Square Error" (RMSE) obtenues par les 3 versions d'ISAT sont récapitulés dans le tableau 7.2. Les versions 2a et 2b d'*ISAT* donnent des scores RMSE similaires, voire inférieur alors que le temps de calcul est allégé d'environ 50%.

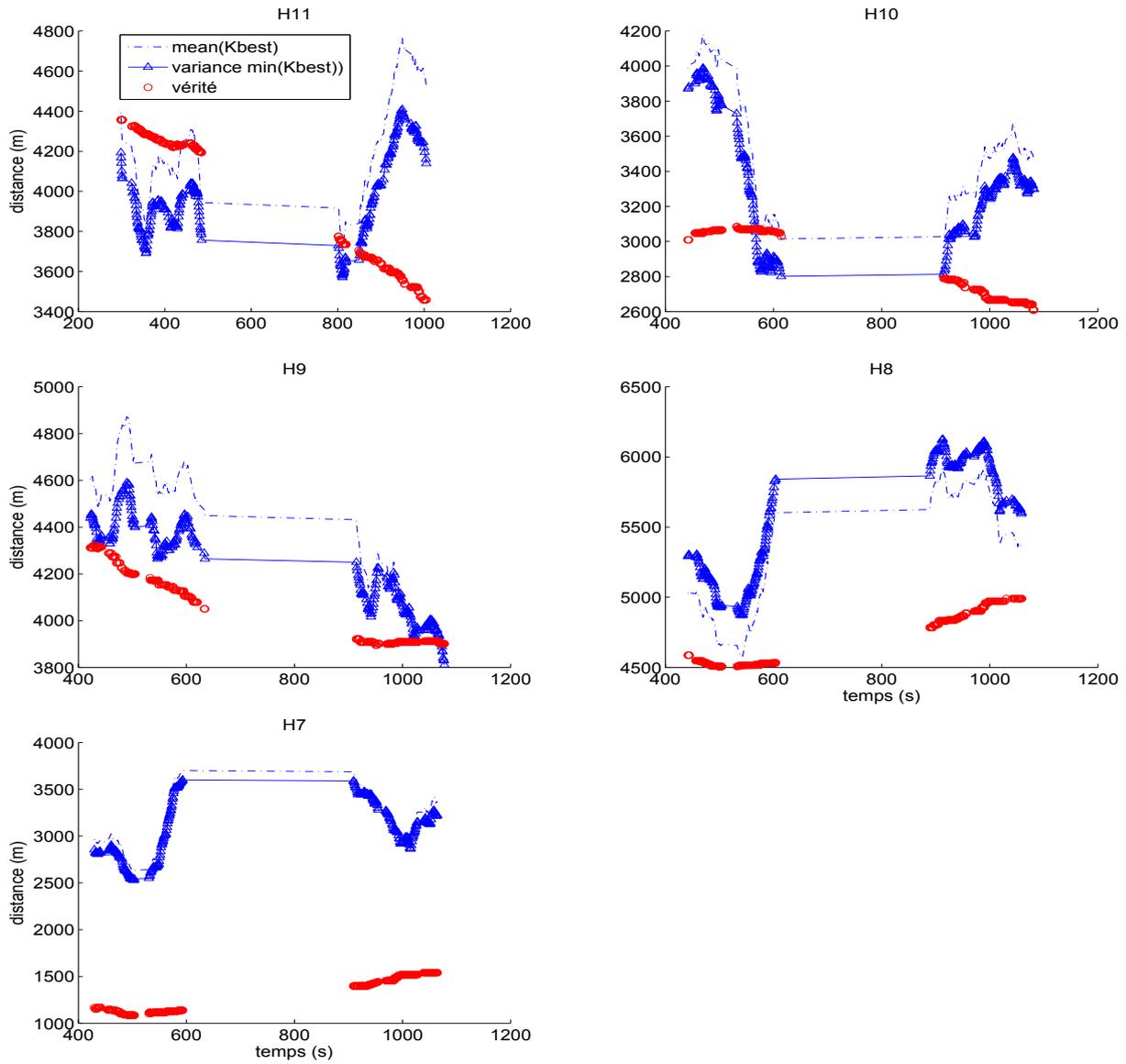


FIGURE 7.5 – Prédictions temporelles de distance selon le modèle ISAT2 (a et b) basé sur l’IM et la vérité terrain. Les données de test proviennent de l’ensemble train/test dissocié dans le temps.

	H11	H10	H9	H8	H7	moyenne
<i>ISAT</i>	809	838	251	327	1722	789
<i>ISAT2a</i>	576	725	353	672	1864	838
<i>ISAT2b</i>	461	555	197	885	1780	775

TABLE 7.2 – RMSE en mètre pour les distances prédites par les modèles *ISAT*, *ISAT2a* et *ISAT2b* sur l’ensemble de test (train/test dissociés).

7.2 Estimation de distance par régressions LER, ISAR et ISARH

7.2.1 Choix des ensembles d'apprentissage, de test et de développement

L'utilisation de méthodes de régression par apprentissage est fortement conditionnée par la ressemblance entre les données du test et de l'apprentissage. Nous possédons un unique jeu de données dont la vérité terrain est disponible afin d'entraîner le système et de vérifier la précision des prédictions de distance. Cela nous contraint donc à effectuer l'apprentissage et le test sur des données obtenues dans le même milieu et provenant d'un même animal. Dans ces conditions, il est nécessaire d'envisager différentes manières de partitionner les données. De plus la question de la normalisation des entrées (spectres) se pose dans la mesure où la plupart du temps nous ne connaissons pas les caractéristiques des gains engendrés par la chaîne d'acquisition : nous renonçons alors à l'information sur la distance qu'apporte l'énergie totale du signal, faute de calibration précise. L'essentiel de l'information sera extrait via l'effet de déformation spectrale induit par la propagation acoustique. Ce qui est l'approche proposée par le modèle théorique ISAT. Le processus d'apprentissage est donc réalisé sur une partition "train" des données "Bahamas" set A. Les descripteurs ne sont pas présentés dans l'ordre temporel en entrée du MLP. Nous testons 2 types de partitionnement pouvant répondre à différentes configurations ou conditions d'acquisition des données d'apprentissage :

1. Les données sont mélangées uniformément dans le temps puis séparées en deux ensembles train/test. Un élément du test a donc un voisin temporel direct dans l'espace d'apprentissage. Les espaces de train et de test sont assez similaires et "échantillonnent" de la même manière les variables influentes du problème ($r, a, el \dots$). L'enchevêtrement des données train/test est une condition idéale. Les spectres ne sont pas normalisés (système calibré et données complètes). Ce travail a été notamment publié lors de la conférence NIPS4B 2013 [Doh 13b].
2. Afin de sortir du cadre idéal d'un mélange uniforme, les données sont séparées de manière à ce que les éléments utilisés pour l'apprentissage ne soient pas voisins temporellement des éléments utilisés pour les prédictions. Les espaces de train et de test sont donc relativement différents. On introduit une forte variabilité entre ces 2 espaces. Les éléments du spectre $|S[k]|$ sont normalisés tels que :

$$|S_{norm}[k]| = \frac{|S[k]|}{\max|S|}, \quad (7.5)$$

(système non calibré et données éparées). Nous nous attendons à une dégradation de la précision des estimations.

Ces résultats permettraient soit d'envisager la généralisation d'un modèle appris sur un site spécifique à un ensemble de sites, soit d'utiliser un modèle de prédiction propre à la zone d'apprentissage.

7.2.2 Optimisation des Hyper-paramètres

Lors de la phase de développement, nous optimisons la qualité des prédictions en fonction de plusieurs paramètres : la structure du réseau de neurones (le nombre d'unités de la couche cachée N_{unit} et l'ensemble des valeurs possibles prises par les poids ($[-1 : 1]$ ou $[0 : 1]$)) et les contrôles de l'apprentissage (le nombre d'itérations de la boucle d'apprentissage N_{iter} et le taux d'apprentissage L_r). Les prédictions sont générées sur la partition test du jeu de données de la même manière que pour les estimations ISAT. Aucun des éléments de cette partition n'a été appris

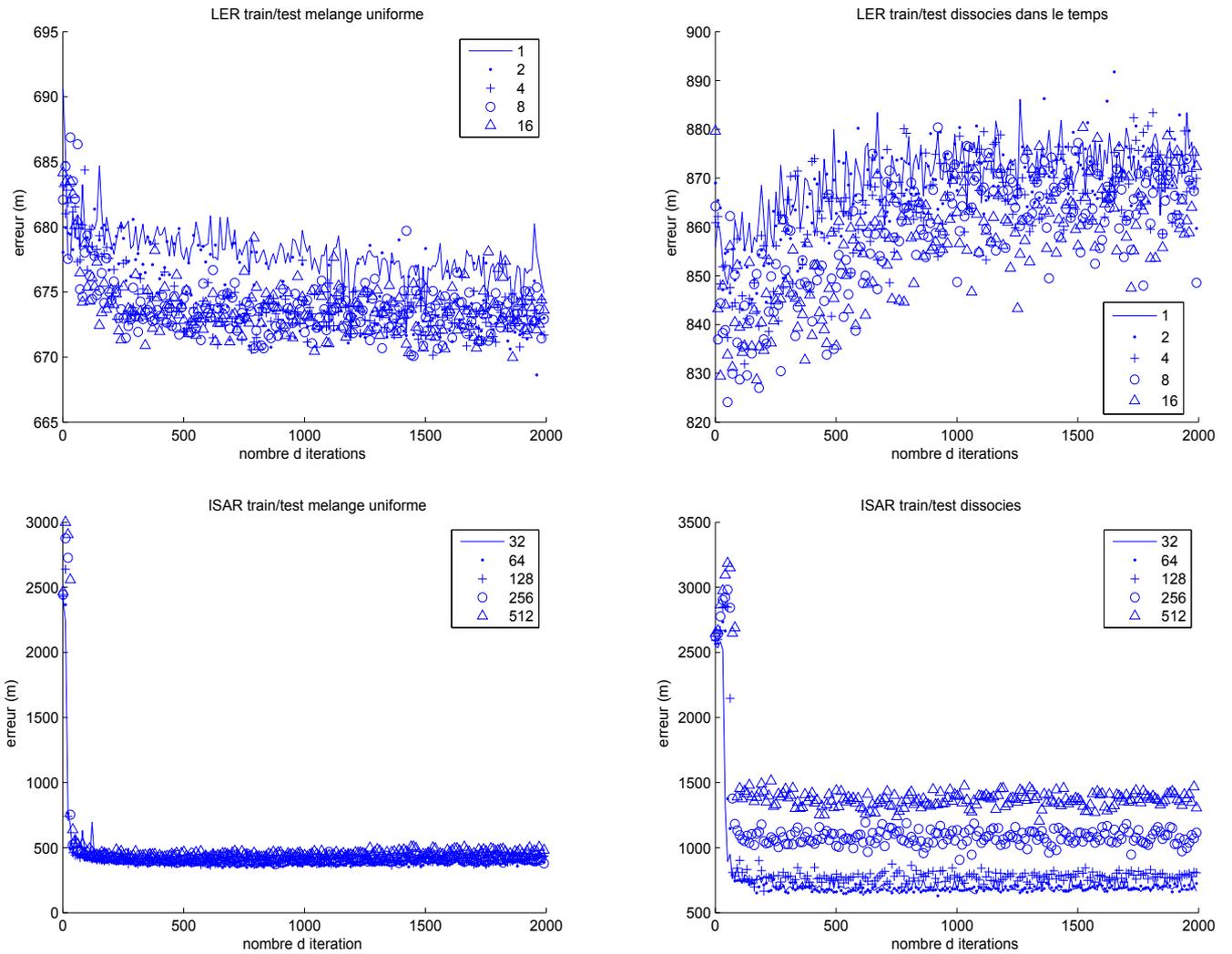


FIGURE 7.6 – Courbes d’erreur obtenues lors de la phase d’optimisation des paramètres : nombre d’unités de la couche cachée N_{unit} et nombre d’itérations de la boucle d’apprentissage N_{iter} , pour modèles LER, ISAR.

par le MLP. Les courbes présentées figure 7.6 représentent l’erreur générée sur les prédictions en fonction du nombre d’itérations à l’apprentissage pour différentes valeurs de N_{unit} .

Pour le modèle *ISAR*, la relation entre les descripteurs présentés et la distance source-récepteur est apprise sur un nombre d’itérations relativement faible. Moins de 100 itérations sont suffisantes pour atteindre un taux d’erreur acceptable (<800 m) et réduire l’erreur initiale d’un facteur 4. Nous trouvons un palier assez stable autour de 300 itérations. Afin de ne pas se retrouver en situation de sur-apprentissage, nous maintenons volontairement le nombre d’itérations au plus bas. De même pour le nombre d’unités de la couche cachée, 64 unités semblent suffisantes pour aboutir à une erreur faible, *i.e.* la moitié de la dimension des éléments d’entrée. Nous constatons que le nombre d’itérations requis pour approximer la régression augmente avec N_{unit} . Pour le mélange dissocié, l’erreur augmente comme nous l’avions prévu (hausse supérieure à 50 %), et l’écart entre les différentes configurations du nombre d’unités de la couche cachée est

aussi amplifié. Nous retiendrons les valeurs optimales communes aux deux types de séparation train/test : $N_{unit} = 64$ et $N_{iter} = 300$. Le modèle ISAR est le plus performant en terme d'erreur de prédiction : autour de 500 m d'erreur en moyenne dans le cas du mélange uniforme.

Pour le modèle LER, l'apprentissage semble nécessiter plus d'itérations. Nous remarquerons que le gain d'efficacité obtenu d'itération en itération est faible : environ 20 m. A l'instar de ISAR, le passage en mélange train/test dissocié détériore les résultats (hausse de 25 % de l'erreur). Mais cette fois nous constatons dans cette condition que l'erreur augmente avec le nombre d'itérations. Ce phénomène n'intervient pas pour le mélange uniforme. Cette configuration pourrait révéler un défaut de convergence de la solution. Les valeurs $N_{unit} = 8$ et $N_{iter} = 350$ seront retenues.

Les tableaux 7.3 et 7.4 récapitulent les paramètres optimaux et l'erreur associée pour les modèles LER et ISAR dans le cadre du mélange train/test dissocié sur l'ensemble test. Nous avons aussi testé le passage en logarithme de l'entrée. Celui-ci semble profitable au modèle LER contrairement au modèle ISAR. Le choix de l'ensemble des valeurs possibles des poids est peu influent sur les performances du modèle LER. En revanche pour ISAR, contraindre les poids à des valeurs comprises dans $[-1; 1]$ réduit légèrement l'erreur ($\approx 50, 100$ m). Le choix de cet ensemble et la compression logarithmique de l'entrée permettent de simplifier la structure du réseau en terme d'unités cachées en conservant $N_{unit} < 6$.

	poids Z	poids D
log	(2,1,2)=878 p1(9,22,2)=1141 p2(1,6,1)=910 p3(9,32,3)=873	(8,4,2)=872 p1(9,6,4)=1048 p2(9,2,1)=854 p3(8,19,4)=739
min	(9,20,4)=750 p1(9,12,3)=960 p2(1,48,5)=559 p3(7,2,1)=630	(8,24,5)=747 p1(2;24,1)=1010 p2(1,28,5)=558 p3(7,25,4)=772
lin	(9,18,1)=790 p1(9,22,2)=1119 p2(9,4,1)=669 p3(8,20,4)=841	(8,4,5)=797 p1(9,6,4)=1005 p2(9,42,3)=611 p3(9,35,5)=762
min	(9,3,3)=736 p1(8,1,1)=993 p2(9,4,3)=624 p3(1,5,2)=585	(9,29,5)=744 p1(1,44,5)=997 p2(9,47,4)=539 p3(6,1,5)=640

TABLE 7.3 – Récapitulatif des RMSE (en m) obtenues pour les hyperparamètres optimaux N_{unit} , N_{iter} et L_r , le taux d'apprentissage pour le modèle LER. $RMSE = (N_{unit}, N_{iter}, L_r)$. Sur l'ensemble test (train/test dissociés). p1, p2 et p3 sont les partitions selon les expositions de l'animal.

	poids Z	poids D
log	(3,21,4)=456 p1(5,7,4)=934 p2(7,25,5)=484 p3(7,49,4)=1125	(8,4,2)=872 p1(9,6,4)=1048 p2(9,2,1)=854 p3(8,19,4)=739
min	(3,3,5)=415 p1(8,50,3)=773 p2(3,40,3)=513 p3(7,44,5)=924	(8,24,5)=747 p1(2;24,1)=1010 p2(1,28,5)=558 p3(7,25,4)=772
lin	(9,33,3)=469 p1(6,45,2)=1015 p2(3,40,3)=513 p3(7,44,5)=924	(9,19,3)=540 p1(6,12,4)=1050 p2(7,10,4)=392 p3(9,16,5)=887
min	(8,26,4)=442 p1(8,35,5)=665 p2(6,38,4)=412 p3(5,41,1)=393	(9,33,2)=458 p1(7,26,5)=737 p2(6,11,4)=388 p3(8,13,5)=467

TABLE 7.4 – Récapitulatif des RMSE (en m) obtenues pour les hyperparamètres optimaux N_{unit} , N_{iter} et L_r , le taux d'apprentissage pour le modèle ISAR. $RMSE = (N_{unit}, N_{iter}, L_r)$. Sur l'ensemble test (train/test dissocié). p1, p2 et p3 sont les partitions selon les expositions de l'animal.

7.2.3 Prédiction de distance pour un mélange train/test uniforme

Dans ce paragraphe, nous proposons une prédiction temporelle de distance selon les régressions *LER*, *ISAR* et *ISARH* sur tous les hydrophones H11, H10, H9, H8 dans le cadre d'un mélange train/test uniforme. Les dimensions respectives sont 2048/2048.

En figure 7.7, nous pouvons comparer la fidélité des prédictions à la vérité terrain (ces résultats sont récapitulés table 7.5). Tout d'abord, les prédictions *ISAR* et *ISARH* sont nettement plus précises que les prédictions *LER*. Cela démontre l'intérêt de ne pas seulement considérer l'énergie totale du signal mais de travailler à l'échelle intra-spectre via une relation entre plusieurs canaux fréquentiels du spectre. Pour les 3 modèles, l'erreur la plus élevée est obtenue sur l'hydrophone H10.

Le modèle *LER* aboutit à des résultats proches du modèle théorique en terme d'erreur moyenne. On observe certaines transitions qui conduisent à un effondrement des prédictions et ne semblant pas affecter les autres modèles.

Le biais entre prédiction et vérité terrain semble s'accroître sur les sections où l'azimut est le plus faible (< 20 degrés). Ce comportement peut s'expliquer par les carences de configurations "on axis" dans notre jeu de données. La configuration majoritaire est latérale. La structure des signaux dépend fortement de la position relative du récepteur du fait de l'aspect directif des émissions sonores du cachalot [Laplanche 05]. Le réseau de neurones semble apprendre la forme caractéristique du spectre pour une orientation donnée.

De plus, l'erreur moyenne du modèle *ISARH* est proche de celle du modèle *ISAR*. Pourtant la dispersion des résultats de prédiction est légèrement améliorée pour *ISARH* qui comprend une étape intermédiaire d'évaluation des angles de position. Les résultats de ces prédictions

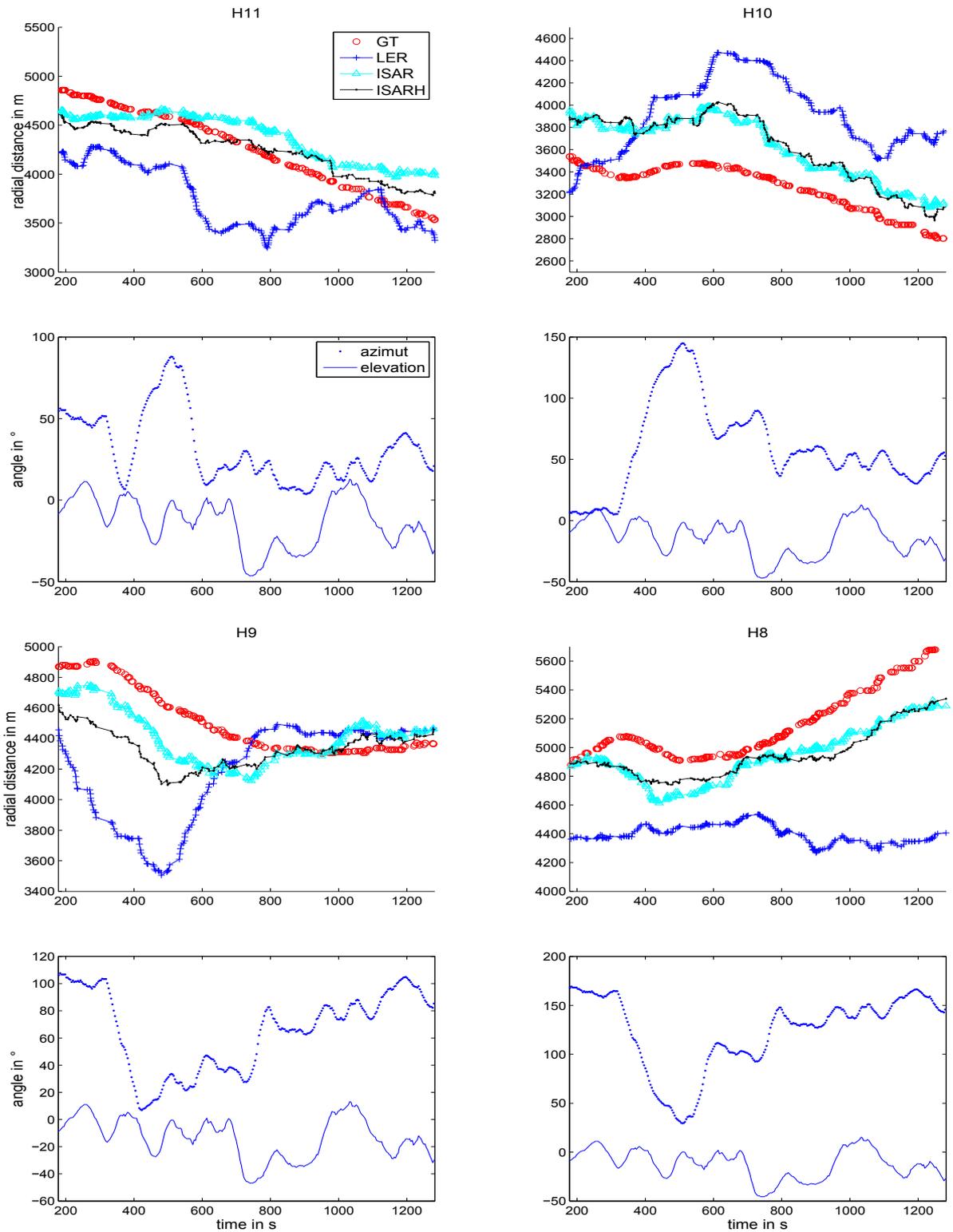


FIGURE 7.7 – Prédiction temporelle comparées aux vérités terrain de la distance radiale et des angles de position selon les modèles LER, ISAR, ISARH. Ensemble train/test uniforme et spectres non normalisés. Prédiction données sur l'ensemble de test et modèle optimisé sur l'ensemble développement.

d'angles sont bien moins précis que les prédictions de distance. Qualitativement, les prédictions d'azimut sont plus précises que celles de l'élévation, ce qui suggère que la forme du spectre est impactée en grande partie par l'azimut. De manière analogue à une source acoustique de type piston circulaire, le cône de directivité de l'animal dans l'espace 3D présenterait une symétrie de révolution autour de l'axe antéro-postérieur. Une autre piste pourrait être investiguée : Glotin et al. proposent dans [Glotin 09] un modèle d'estimation d'offaxis monocapteur pour le cachalot basé sur les rapports d'énergie des pulses composant le clic. Un couplage de cette méthode avec nos modèles intra-spectraux pourrait être envisagée.

	H11	H10	H9	H8	moyenne
<i>LER</i> moyenne	551	580	332	707	542
<i>ISAT</i> moyenne	784	609	373	754	630
<i>ISAR</i> moyenne	157	406	124	188	218
<i>ISARH</i> moyenne	157	319	166	188	208

TABLE 7.5 – RMSE des distances prédites par les modèles *LER*, *ISAR*, *ISARH* pour le mélange train/dev/test uniforme, comparée au modèle *ISAT*.

7.2.4 Prédiction de distance pour un mélange train/test dissocié dans le temps

Dans cette section nous présentons les estimations de distance résultantes des modèles *LER* et *ISAR* entraînés sur le mélange train/test dissocié dans le temps. Les dimensions respectives des ensembles train/test sont 1840/2050. Les courbes sont présentées en figure 7.9 et un récapitulatif des résultats est donné dans le tableau 7.6.

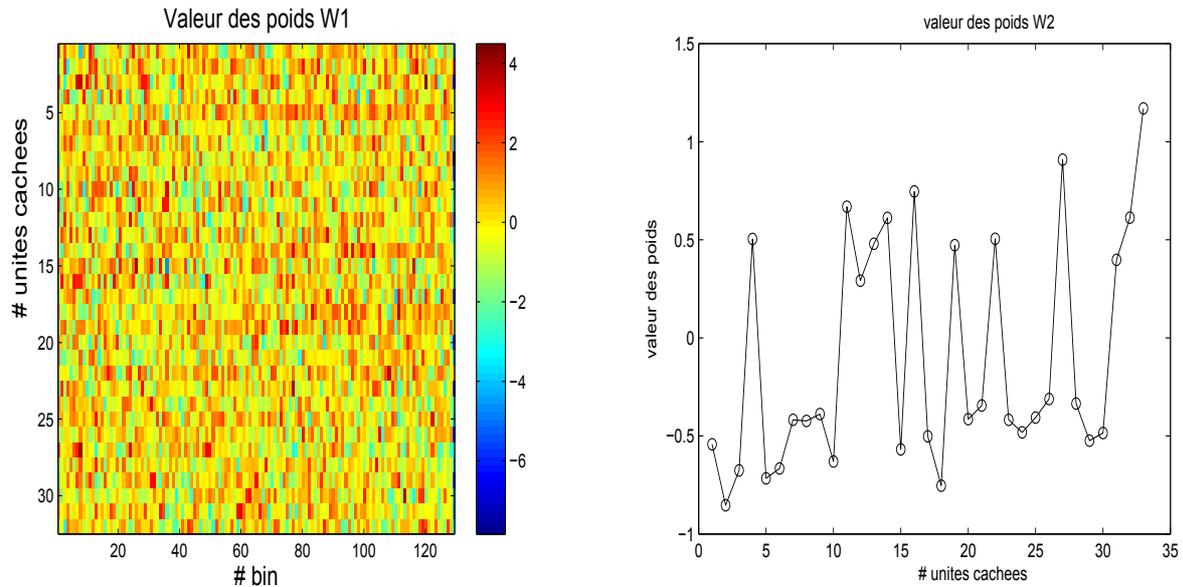


FIGURE 7.8 – Valeurs des poids W1 et W2 appris sur l'ensemble train du mélange train/test dissocié dans le temps.

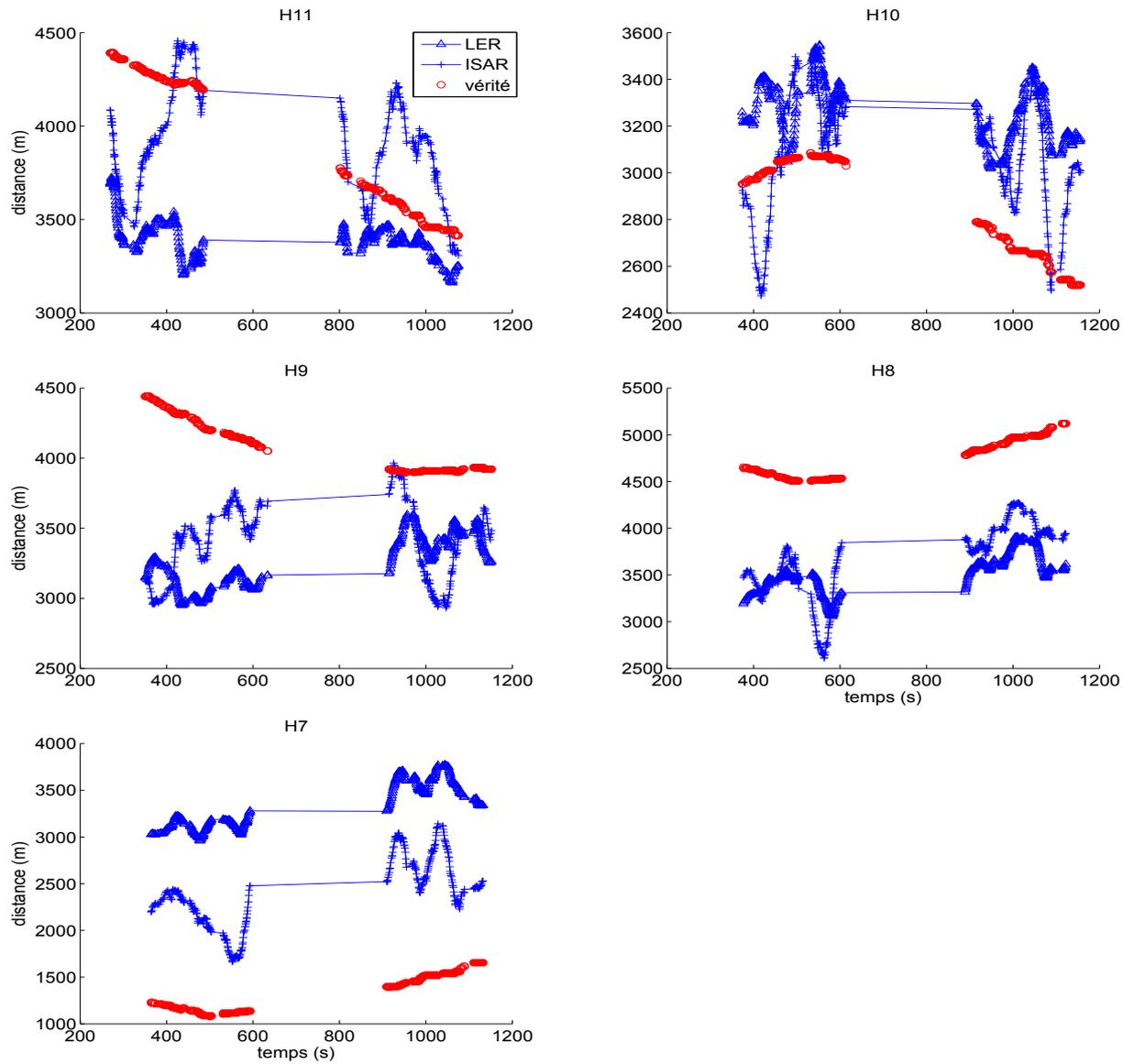


FIGURE 7.9 – Prédictions temporelles de distance selon les modèles ISAR et LER optimaux retenus et de la vérité terrain. Les données de test proviennent de l'ensemble train/test dissocié dans le temps et les spectres sont normalisés selon l'équation 7.5. Les prédictions sont données sur l'ensemble de test et le modèle est optimisé sur l'ensemble de développement.

La qualité des estimations se trouve relativement impactée par le passage au mélange train/test dissocié. Les courbes présentent de fortes variations et paraissent "bruitées". Le modèle *ISAR* montre un apport sur *LER* en améliorant de plus de 25% la précision des prédictions. Les prédictions sur H7 et H8 sont particulièrement erronées. Remarquons qu'il s'agit des hydrophones pour lesquels les collections de clics sont associées aux distances extrêmes, situées sur les queues de distribution. Lors de la phase d'apprentissage, H7 fournit les distances les plus petites et H8 les distances les plus grandes. Ces distances et les spectres associés sont donc statistiquement peu appris par le réseau qui peine à restituer des prédictions fiables dans ces échelles de distances.

La figure 7.8 montre des valeurs de poids aussi bien négatives que positives. La sortie est bien une combinaison de produits et ratios du logarithme des entrées comme le suggère l'équation 4.14. Nous avons construit une analogie entre les estimateurs *ISAR* et *ISAT*. De plus, ces résultats nous permettent de constater que *ISAT* est plus fiable que *ISAR* sur les grandes distances (> 4000 m). À l'inverse *ISAR* semble mieux adapté à la prédiction de petites distances (< 4000 m). *ISAT* et *ISAR* semblent être complémentaires.

	H11	H10	H9	H8	H7	moyenne
<i>ISAT</i>	809	838	251	327	1722	789
<i>LER</i>	631	452	883	1243	2017	1045
<i>ISAR</i>	397	350	761	1259	1110	775

TABLE 7.6 – RMSE en m, entre les distances prédites par les modèles *LER* et *ISAR*. Des prédictions aléatoires de distance donneraient une RMSE d'environ 1800 m.

7.2.5 Erreur et précision des prédictions *ISAT*, *LER*, *ISAR* et *ISARH*

Selon les résultats récapitulés (tables 7.1, 7.5), l'erreur obtenue avec le modèle *ISAT* a tendance à varier en fonction du choix de l'hydrophone. Toutefois, le modèle *ISAT* et les modèles basés sur des réseaux de neurones *ISAR* and *ISARH* conduisent au même constat : les estimations sont améliorées si l'on considère les propriétés inter-spectrales au lieu de l'énergie totale du signal. Celles-ci semblent plus sensibles aux variations naturelles et volontaires du niveau d'énergie à la source.

À propos de l'échelle des pulses intervenant dans le traitement *ISAT*, les calculs réalisés sur $N = 256$ échantillons comparés à $N = 128$, donnent des estimations plus précises. La plus petite échelle semble être encore plus dépendante des configurations à fort azimuth (table 6.3). Quelle que soit l'échelle, nous observons que les estimations les plus précises sont obtenues sur l'hydrophone 9. Notons que les enregistrements de cet hydrophone correspondent à des situations dont l'azimut est intermédiaire (73 degrés, voir table 6.3). Comme nous l'avons suggéré précédemment, nous savons que le récepteur acquiert différentes structures de clic suivant sa position relative à l'animal. Se situer devant l'animal "on axis" favorise les clics hautement directifs composés de pulses de type PJ. Hors de l'axe de l'animal, sont favorisés les clics composés de pulses de type P [Laplanche 05]. Donc, les configurations de l'azimut pour H9 pourraient fournir en moyenne autant de clics composés de pulse PJ que clics constitués de pulses P.

Les modèles *ISARH*, *ISAR*, *ISAT* et *LER* se comportent différemment face à l'azimut et à l'élévation. Les performances du modèle *ISAT* semblent plus impactées par l'azimut. Ce qui n'est pas surprenant compte tenu que nous avons développé ce modèle sans considérer la directivité des productions sonores de l'animal. De futurs travaux pourraient nous permettre d'intégrer cette

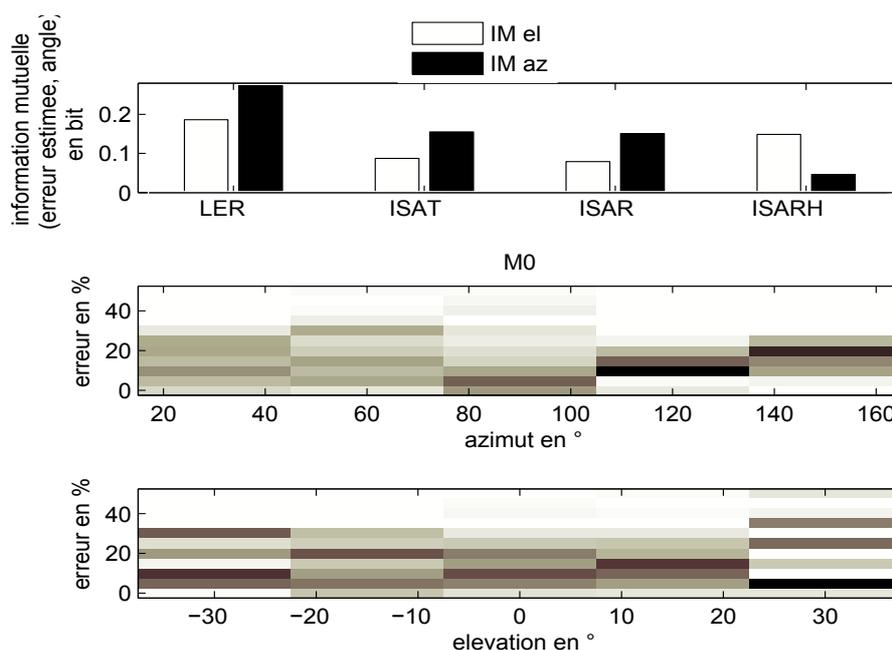


FIGURE 7.10 – Bi-distribution de l'erreur de prédictions vs l'azimut (au milieu), l'élévation (en bas) sur les données du jeu de données "Bahamas" dédiées au test. En haut, Information mutuelle entre l'erreur de prédiction et les angles de position.

variable dans l'expression de l'estimateur de distance. Pourtant, comme nous nous y attendions, le MLP a probablement exprimé cette dépendance malgré peu de données "on axis" présentés lors de la phase d'apprentissage.

La figure 7.10 présente la distribution jointe de l'erreur et des angles de position de l'animal ainsi que l'information mutuelle (IM) résultant des modèles utilisés dans cette étude. Les résultats confirment que les modèles intra-fréquences (par opposition au modèle *LER*) ont réduit l'IM. *ISAT* et *ISAR* obtiennent des valeurs similaires. *ISARH* aboutit à la valeur de l'IM la plus faible et, de plus, les fortes valeurs d'IM sont obtenues pour l'azimut par tous les modèles en dehors de *ISARH*. Cela suggère et confirme une plus forte dépendance des estimations à l'azimut. La relation erreur *vs* angles de position semble être complexe et s'exprimer selon différents modes.

Toutefois, l'étude des résultats moyens (sur tous les hydrophones) conduit à la conclusion que *ISAR* affiche les meilleurs résultats sur de petites distances (< 4000 m) et *ISAT* fournit les meilleures estimations sur de grandes distances (> 4000 m). Une explication de nature physique liée à la propagation sous-marine pourrait justifier cette différence de comportement face aux distances à estimer. La propagation sur de telles distances pourrait engendrer un "gommage" de la forme du spectre (lié à l'angle de tir de l'animal) favorable à l'estimateur *ISAT* qui conduit à considérer un spectre plat en sortie. En revanche, sur de faibles distances, la forme des spectres est très liée aux angles de position et cette relation forme du spectre/angle est apprise par le MLP *ISAR*. Cela conduit à élaborer un modèle basé sur ces 2 types d'estimateurs.

7.2.6 Modèle de fusion entre *ISAT* et *ISAR*

Afin de tirer parti des estimations obtenues par le modèle théorique *ISAT* et le modèle basé sur un MLP *ISAR*, nous construisons un modèle simple de fusion entre ces 2 estimateurs. Nous effectuons alors une moyenne pondérée dont l'équation est :

$$d_{fusion} = (1 - Q)d_{ISAR} + Qd_{ISAT} \quad (7.6)$$

d_{fusion} est le résultat de la fusion, d_{ISAR} et d_{ISAT} respectivement les estimations des 2 estimateurs et Q une simple moyenne dans l'intervalle $[0; 10km]$ telle que :

$$Q = \frac{1}{2}(\min(d_{ISAR}, 10000) + \min(d_{ISAT}, 10000)).$$

Les résultats sont présentés en figure 7.11 et synthétisés dans le tableau 7.7. La fusion permet d'améliorer les résultats moyens sur l'ensemble des hydrophones et la robustesse des estimations face aux distances à prédire. Nous observons un gain de 150 m de RMSE, soit environ 25%.

	H11	H10	H9	H8	H7	moyenne
<i>ISAT2b</i>	461	555	197	885	1780	775
<i>ISAR</i>	397	350	767	1260	1099	774
<i>Fusion</i>	386	396	369	388	1294	566

TABLE 7.7 – RMSE (en m) des distances prédites par les modèles *ISAT2b*, *ISAR* et le modèle de fusion.

L'effet notable du modèle de fusion est de diminuer l'erreur RMSE et d'uniformiser la RMSE par hydrophone autour de sa valeur moyenne (566 m). Le modèle de fusion est donc plus stable face aux différentes échelles de distance et aux effets de directivité des émissions sonores.

Le tableau 7.8 donne les vitesses radiales moyennes estimées par le modèle de fusion. En comparaison avec le tableau 6.3, les vitesses obtenues et leur module sont cohérents à l'exception de l'hydrophone H11. *ISAT* a montré de mauvais résultats sur cet hydrophone particulièrement "on axis" et dont la trajectoire de l'animal est presque radiale à l'hydrophone.

	H11	H10	H9	H8
vitesse radiale moyenne	0.25 m/s	-0.64 m/s	-0.42 m/s	0.71 m/s
vitesse radiale vraie	-1.13 m/s	-0.7 m/s	-0.4 m/s	0.8 m/s

TABLE 7.8 – Vitesses radiales moyennes obtenues par le modèle de fusion *ISAT/ISAR*.

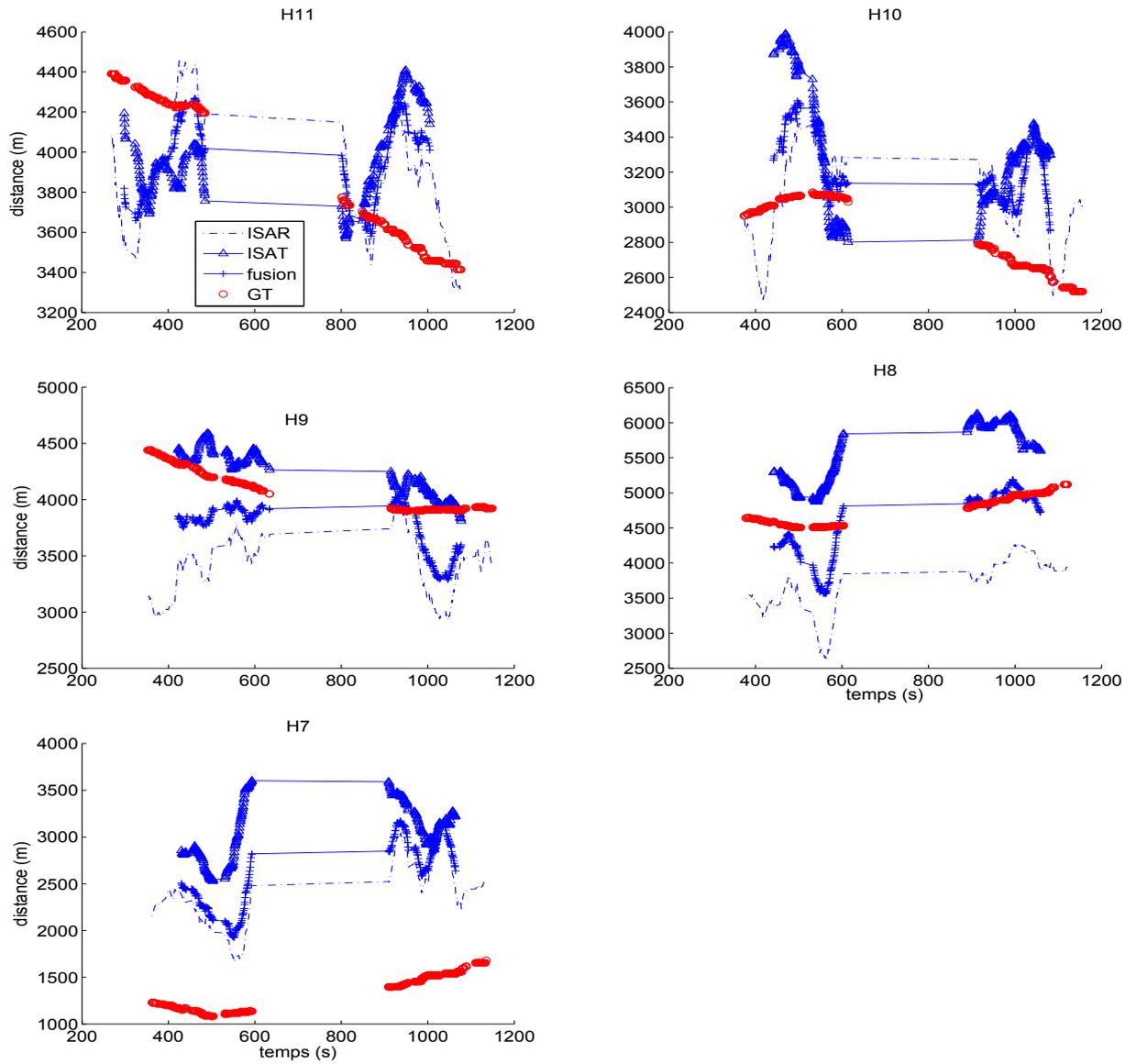


FIGURE 7.11 – Prédictions temporelles ISAR, ISAT et modèle de fusion comparées aux vérités de terrain distance radiale.

7.3 Test du modèle de fusion ISAT/ISAR sur le jeu de données Bahamas set B

Dans cette section, nous présentons les résultats des prédictions temporelles de distance sur le jeu de données Bahamas set B (hydrophones 6 et 5) où 4 animaux évoluent simultanément. Rappelons que H6 est positionné en moyenne "off axis" voire vers le dos des animaux et que H5 est en moyenne face aux individus ("on axis"). Nous avons retenu le modèle de fusion ISAR/ISAT2. Le Modèle ISAR a été développé et optimisé sur le jeu de données Bahamas set A. Même si les propriétés acoustiques de la colonne d'eau et la bathymétrie sont similaires au set A, les individus et les caractéristiques des hydrophones ne sont pas identiques et donc inconnus du réseau de neurones. Les modèles reçoivent tous en entrée des spectres normalisés.

Les résultats sont présentés de la manière suivante :

- à gauche, le modèle est appris sur la partition train/développement/test dissociée
- à droite, le modèle est appris sur toutes les données
- en haut, les prédictions du modèle de fusion ISAT2a/ISAR
- en bas, les prédictions du modèle de fusion ISAT2b/ISAR

7.3.1 Estimations de distance dans le temps

Les figures 7.12 et 7.13 représentent les résultats bruts des estimations de distance en fonction des dates de détection des clics. Dans ce cas à plusieurs animaux les résultats ne peuvent pas être lissés au risque d'associer des prédictions d'animaux différents. En revanche, en considérant l'hypothèse que des estimations successives sur un individu doivent être stables dans le temps et que les clics des différents animaux ne sont pas toujours imbriqués, nous effectuons un filtrage des estimations dit "de voisinage". Il consiste à ne retenir que les estimations cohérentes entre elles sur une courte durée (3 clics). Ce filtrage va favoriser les régions dans lesquelles des estimations successives se sont accumulées.

Soit $d_{fusion}(t)$ l'estimation brute pour un clic reçu à l'instant t et δd_{seuil} le seuil des variations de distances estimées consécutivement. Alors, l'estimation après filtrage $d_{fusion}^*(t)$ est défini par :

$$\begin{aligned}
 d_{fusion}^*(t) &= d_{fusion}(t) \\
 &si \quad |d_{fusion}(t) - d_{fusion}(t-1)| < \delta d_{seuil}, |d_{fusion}(t) - d_{fusion}(t+1)| < \delta d_{seuil} \\
 &= \emptyset \\
 &sinon
 \end{aligned}
 \tag{7.7}$$

Ces résultats montrent qu'il est difficile d'observer une accumulation de points selon 4 traces distinctes. Sur H6, il apparaît 2 traces et les prédictions sont, pour la plupart, contenues entre 3000 m et 7000 m. La vérité terrain est située entre 4500 m et 6500 m et 3 trajectoires de distances sont relativement confondues. Sur H5, les estimations sont denses entre 2000 m et 7500 m et nous retrouvons globalement 3 traces distinctes, la vérité se situe entre 1500 m et 6000 m et donne au moins 3 distances distinctes. Nous constatons donc que les prédictions obtenues sur H6 sont moins étalées que sur H5. Cette tendance est confirmée quelque soit le type de modèle utilisé. Nous remarquons que la dynamique des traces obtenus concorde avec les dynamiques suggérées par la vérité : plus importante sur H5 que sur H6.

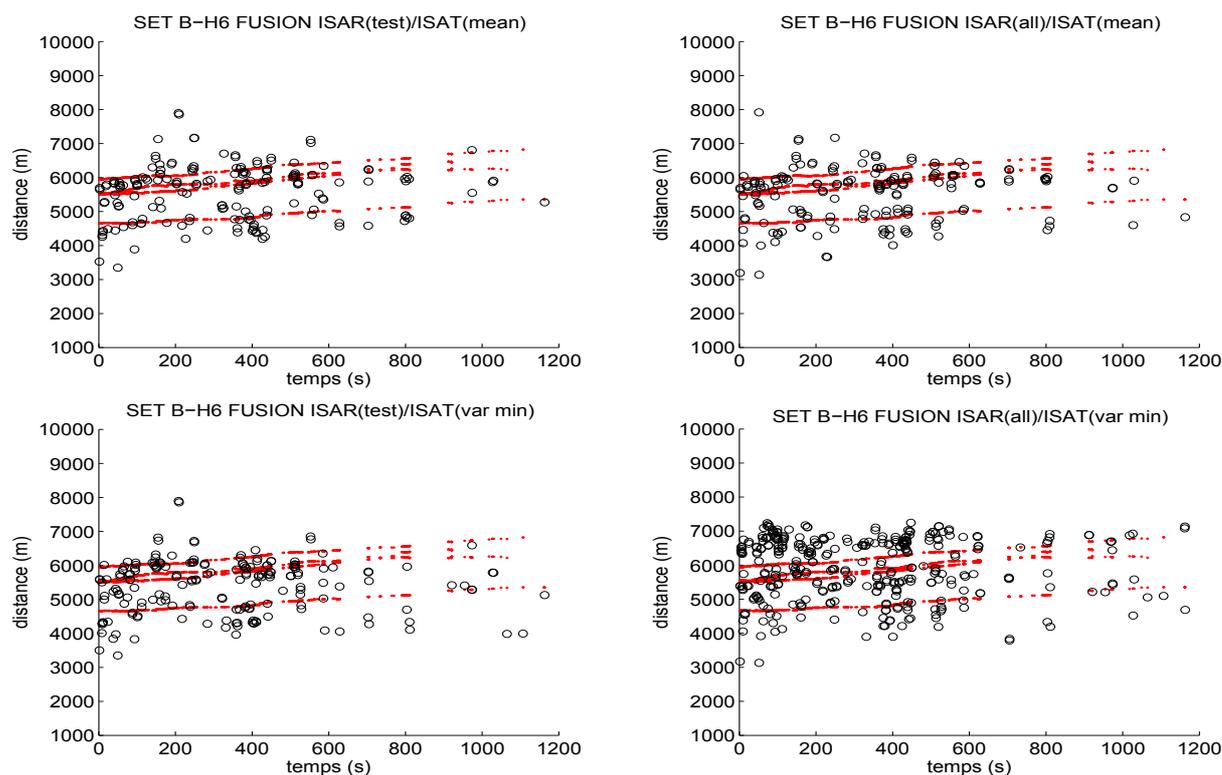


FIGURE 7.12 – Prédictions temporelles ISAR de distance sur l'hydrophone 6 représentées par des cercles noirs. Les distances "vérité terrain" sont matérialisées par des points rouges.

Par ailleurs, nous remarquons le peu d'influence de l'espace d'apprentissage sur la répartition des prédictions (de gauche à droite sur les figures 7.12 et 7.13). Le modèle appris sur la totalité des données semble engendrer des distances plus stables que le modèle appris sur le mélange de données train/dev/test dissocié. Cela s'observe sur H6 comme sur H5 où 3 traces de distances sont distinctes.

De plus le choix ISAT2a dans le modèle de fusion semble favoriser la concentration et l'accumulation de points sur H6. Ce qui ne s'observe pas sur H5 où il y a très peu de différences entre *ISAT2a* et *ISAT2b*.

Sur les représentations les plus claires, nous estimons la vitesse radiale issue de nos estimations pour chaque trace de distance obtenue via le coefficient directeur d'une régression linéaire. Les valeurs des vitesses et leur signe concordent avec la référence.

	SW1	SW2	SW3	SW4
H5 estimation	-0.63 m/s	-0.81 m/s ★	-0.81 m/s ★	-0.18 m/s
H5 référence	-0.41 m/s	-0.83 m/s	-0.65 m/s	-0.36 m/s
H6 estimation	1 m/s	0.66 m/s	0.41 m/s ★	0.41 m/s ★
H6 référence	0.7 m/s	0.54 m/s	0.85 m/s	0.58 m/s

TABLE 7.9 – Vitesse radiale moyenne entre les hydrophones et chaque cachalot (SW) selon l'estimation du modèle de fusion filtré. Nous notons ★ lorsque les trajectoires de distances sont confondues.

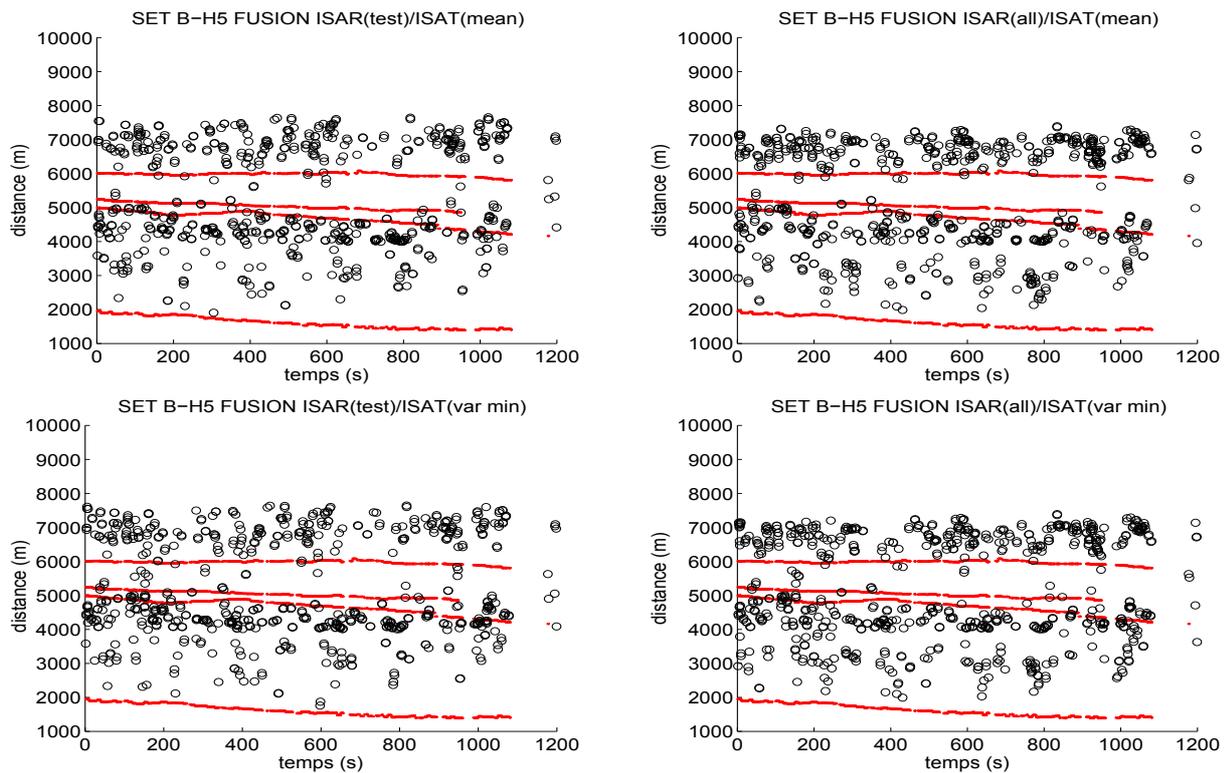


FIGURE 7.13 – Prédiction temporelle ISAR de distance sur l'hydrophone 5 représentées par des cercles noirs. Les distances "vérité terrain" sont matérialisées par des points rouges.

7.3.2 Traitement des prédictions par bi-histogrammes temps/distance

Dans cette section nous réalisons un histogramme à 2 dimensions des estimations fournies par le modèle de fusion ISAT2/ISAR dans le temps. L'espace temps/distance est découpé en pavés de dimension 30 s par 250 m. Ce type de représentation permet de visualiser l'évolution de la répartition des prédictions dans le temps. L'observation des figures 7.14 et 7.15 confirme les tendances générales dégagées par les estimations temporelles brutes.

Sur la figure 7.14 les bi-histogrammes font apparaître des accumulations de points sur certaines zones pour un instant donné. Cela reflète le caractère multisource des signaux enregistrés. Par exemple, à gauche, à la date $t \approx 400$ s, nous constatons simultanément une forte densité de points aux distances 4500 m et 5500 m pour le modèle appris sur la partition train/développement/test dissociée ; à droite, 4250 m et 5750 m pour le modèle appris sur toutes les données. La vérité terrain nous fournit les distances 6100 m, 5600 m, 5650 et 4600 m. Sur cet hydrophone les distances sont relativement similaires entre-elles.

Sur la figure 7.15, les bi-histogrammes sont plus étalés en distance en plus de faire apparaître des accumulations de points. Par exemple, en haut, à la date $t \approx 800$ s, nous constatons simultanément que les prédictions s'étalent entre 2000 et 7000 m avec une forte densité de points aux distances 3000 m, 4250 m et 6750 m pour le modèle appris sur la partition train/dev/test dissociée, 2500 m, 4000 m et 7000 m pour le modèle appris sur toutes les données. La vérité terrain nous fournit les distances 6000 m, 5000 m, 4700 m et 1900 m. Pour cet hydrophone et à cette résolution, il est difficile de distinguer 4 groupes de distances distincts car les 2 trajectoires intermédiaires dans le plan temps/distance sont confondues. Le modèle de droite fait apparaître

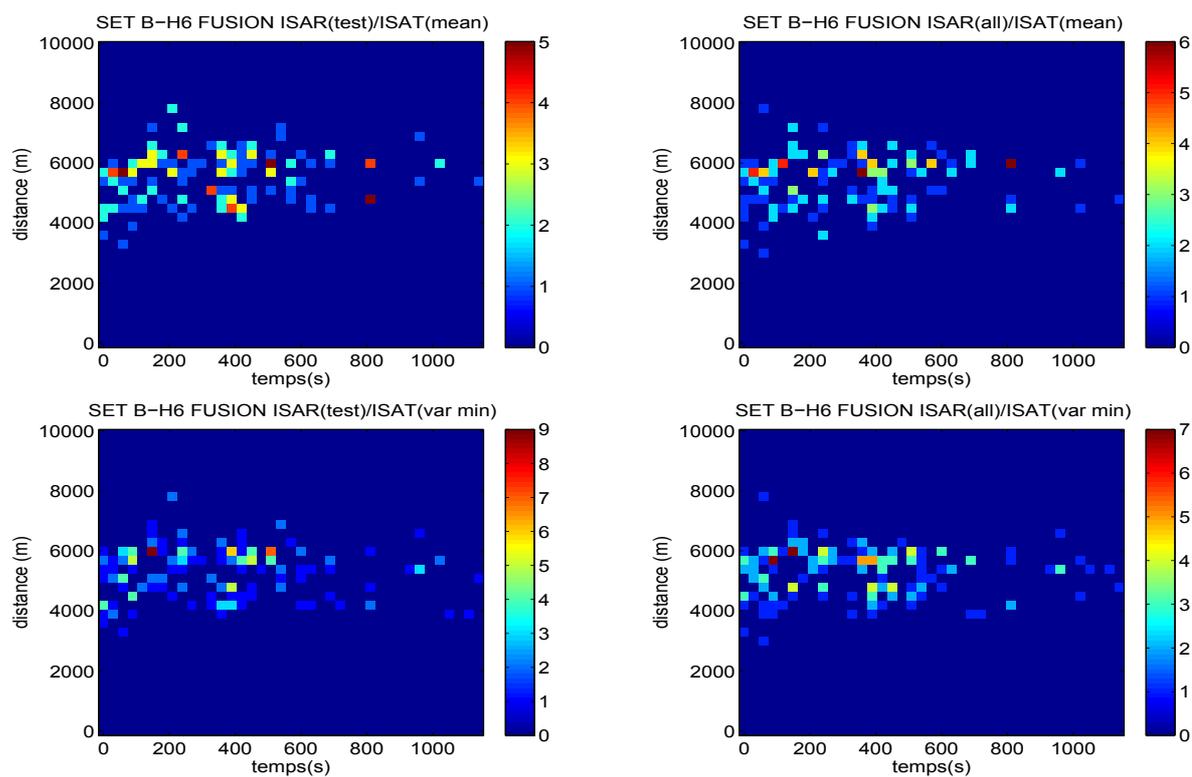


FIGURE 7.14 – Bi-histogrammes temps/distance des prédictions temporelles par fusion ISAT2/ISAR sur le jeu de données BAHAMAS SET B (hydrophone 6).

3 traces distinctes comprises dans un intervalle de distance cohérent avec la vérité terrain.

Pour H6 comme pour H5, les différences entre les figures hautes et basses sont peu significatives. Nous confirmons alors, le peu de différences entre l'emploi de *ISAT2a* ou *ISAT2b*.

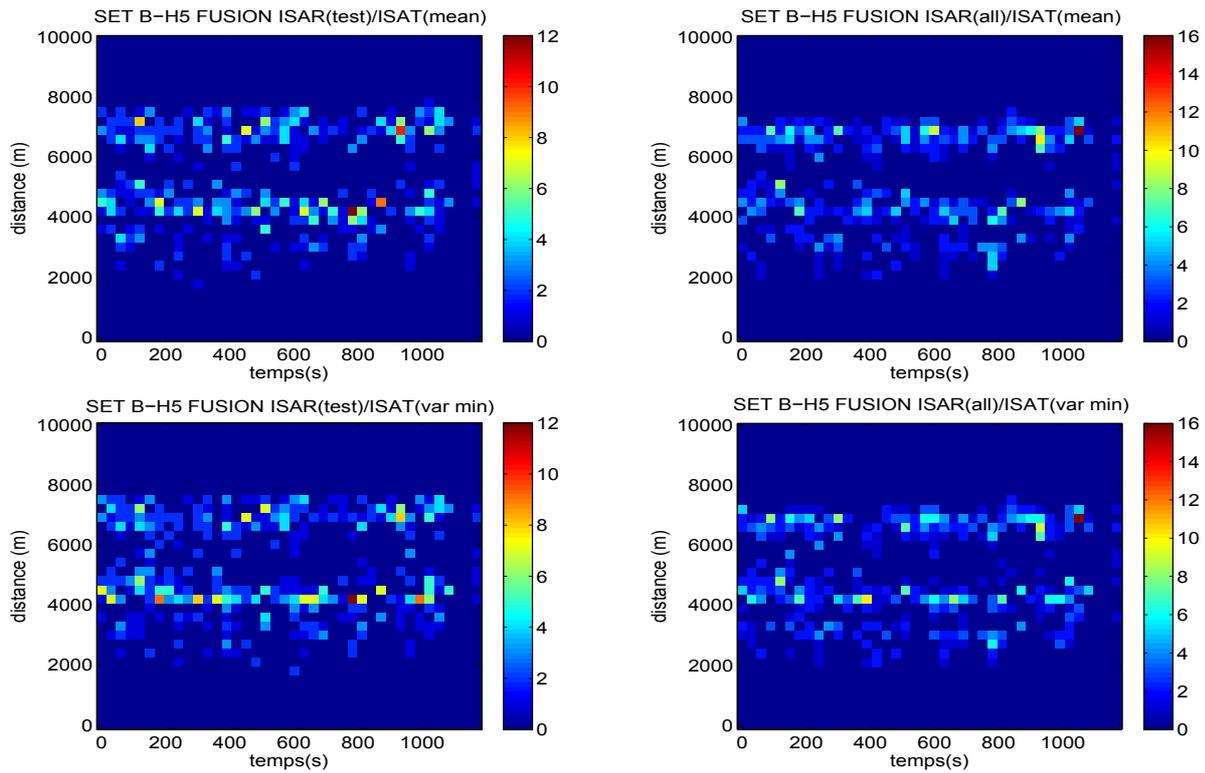


FIGURE 7.15 – Bi-histogrammes temps/distance des prédictions temporelles par fusion ISAT2/ISAR sur le jeu de données Bahamas set B (hydrophone 5).

7.3.3 Répartition des prédictions

Nous présentons dans cette section la répartition des prédictions en terme de distance. Les histogrammes sont normalisés et donc homogènes à une probabilité de distance. Le nombre de classes N_{classe} est défini par rapport à l'effectif total N_{clie} tel que : $N_{classe} = \sqrt{N_{clie}}$. Les figures 7.16 et 7.17 présentent ces densités de probabilité.

Sur la figure 7.16, à gauche, pour le modèle train/dev/test : les distances dont la probabilité est maximale sont de l'ordre de 5900 m (17%) et un second pic émerge distinctement (7%) autour de 4500 m. A droite, pour le modèle appris sur toutes les données, le pic d'amplitude maximale se situe vers 5800 m (20%) et le pic secondaire à 4500 m (7%). Selon la vérité terrain, il y a des pics de distances probables autour de 4700, 5500 m et 6000 m (10%). Ces deux derniers apparaissent confondus sur nos estimations. Seul l'amplitude élevé du pic suggère que plusieurs animaux contribuent à ces distances. Nous remarquons (figure du bas), que *ISAT2b* fait apparaître des variations locales significatives des pics. Sur les 4 représentations, l'ensemble des distances les plus probables (>3%) est compris entre 4000 m et 6500 m, ce qui concorde avec les distances "vraies" s'étalant de 4500 m à 6800 m.

Sur la figure 7.17, à gauche, pour le modèle train/dev/test : les pics de distances apparaissent autour de 3000 m (3%), 4100 m (10%) et 7000 m (8%). A droite, pour le modèle appris sur toutes les données, les pic sont situés autour de 3000 m (3%), 4500 m (7%) et 7000 m (14%) mais semblent plus dissociés. La vérité terrain nous donne 3 pics importants : en 1500 m (8%), en 5000 m (10%) et en 6000 m (20%). Les pics secondaires sur nos estimations sont d'autant plus marqués qu'on utilise le modèle appris sur toutes les données. L'ensemble des distances les plus

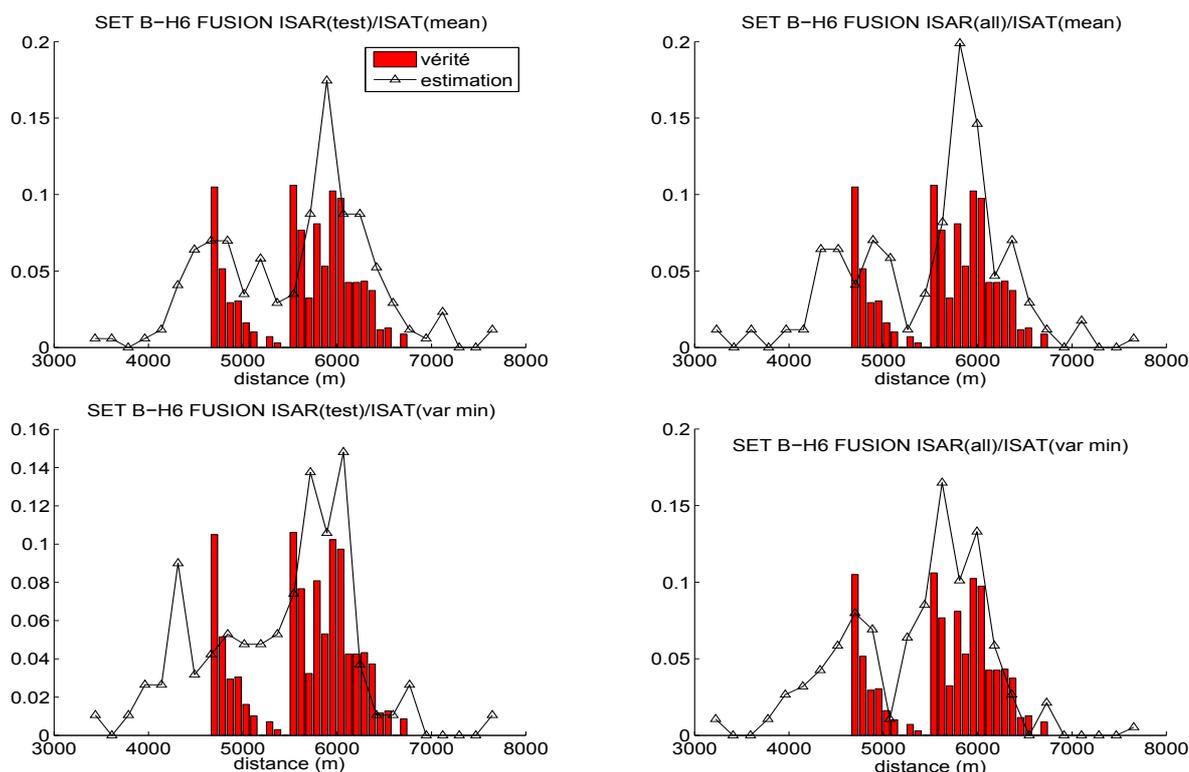


FIGURE 7.16 – Histogrammes normalisés des prédictions ISAR sur le jeu de données Bahamas set B (hydrophone 6).

probables ($> 3\%$) est compris entre 2800 m et 5700 m (10%). Les distances "vraies" s'étalent entre 1700 m et 6000 m.

L'erreur de prédiction est plus importante lorsque les distances à estimer sont extrêmes, celles-ci ont tendance à être surestimées. Enfin, les écarts entre prédiction et vérité sont plus élevés dans le cas de H5. Cet effet pourrait être relié à l'angle de visé de l'animal par rapport à l'hydrophone majoritairement "on axis" pour les 4 cachalots. Or nous savons que 1) le modèle *ISAR* a appris surtout sur des clics vus "off axis" et 2) la structure fréquentielle des pulses reçus de face peut s'écarter de la structure plate envisagée par *ISAT*.

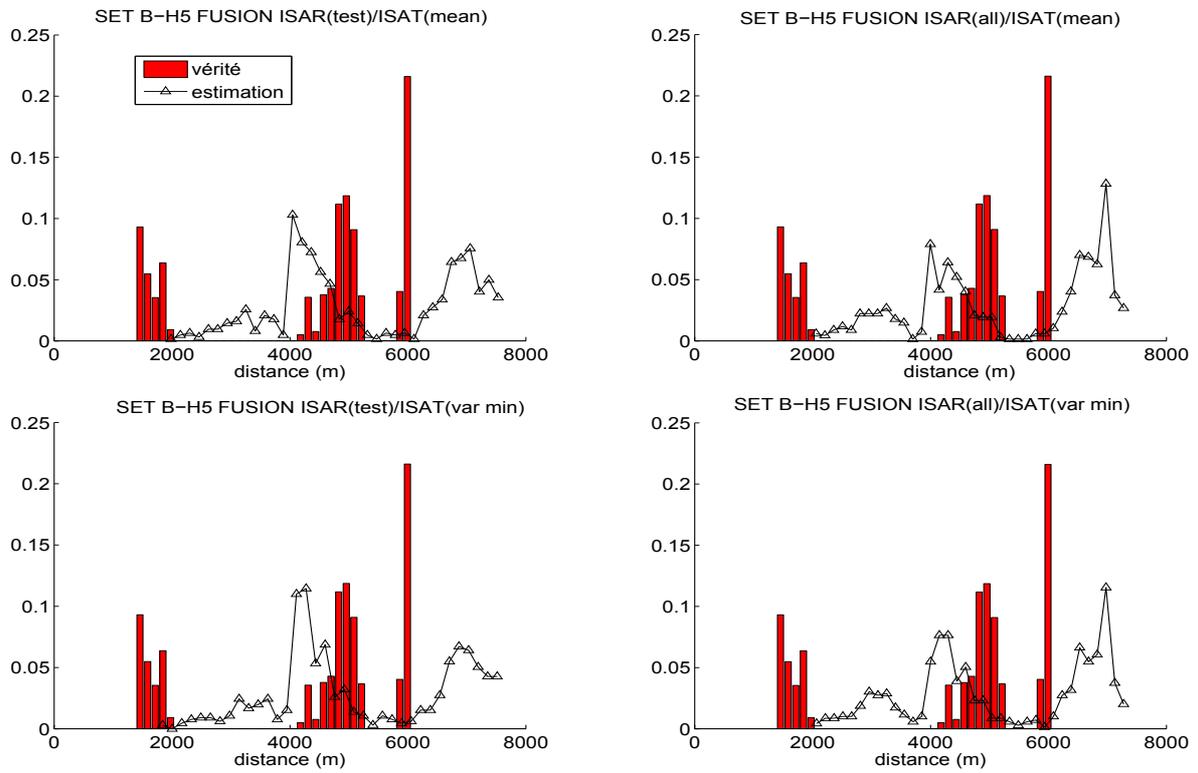


FIGURE 7.17 – Histogrammes normalisés des prédictions ISAR sur le jeu de donnée Bahamas set B (hydrophone 5).

7.4 Résultats préliminaires sur les données issues du protocole DECAV

Nous utilisons les mêmes modèles qu'en section précédente. Les résultats sont présentés de manière identique. Par contre, nous n'avons pas d'information sur la vérité terrain. Les portions utilisées durent 5 min et proviennent de mesures embarquées sur un zodiac du PNPC. Nous appliquons donc le modèle dans un milieu dont la réponse impulsionnelle est totalement différente de celle du milieu d'apprentissage :

- la chaîne d'acquisition est totalement différente de celle employée pour les données Bahamas
- l'hydrophone est proche de la surface (< 30 m de profondeur), contrairement au cas de des données de Bahamas dont les hydrophones sont placés sur le fond océanique à environ 1500 m de profondeur. Le fait que les mesures soient réalisées dans le canal de surface génère un important parasitage auquel notre modèle n'a jamais été confronté
- le milieu est la Mer méditerranée alors que l'apprentissage a été effectué en Atlantique.

7.4.1 Estimations de distance dans le temps

La représentation temporelle des estimations de distance en figure 7.18 révèle des accumulations de points autour de 4000 m et 8000 m et suggère la présence de 2 animaux. Notons que l'emploi de *ISAT2b* accentue la densité des prédictions autour de 8000 m. De plus nous confirmons que le choix des ensembles d'apprentissage influe peu sur les prédictions (plus faibles distances sur un apprentissage uniforme). D'après [Abeille 14], il y avait bien plusieurs cachalots sur zone lors de cette mesure (un contact visuel et un contact acoustique).

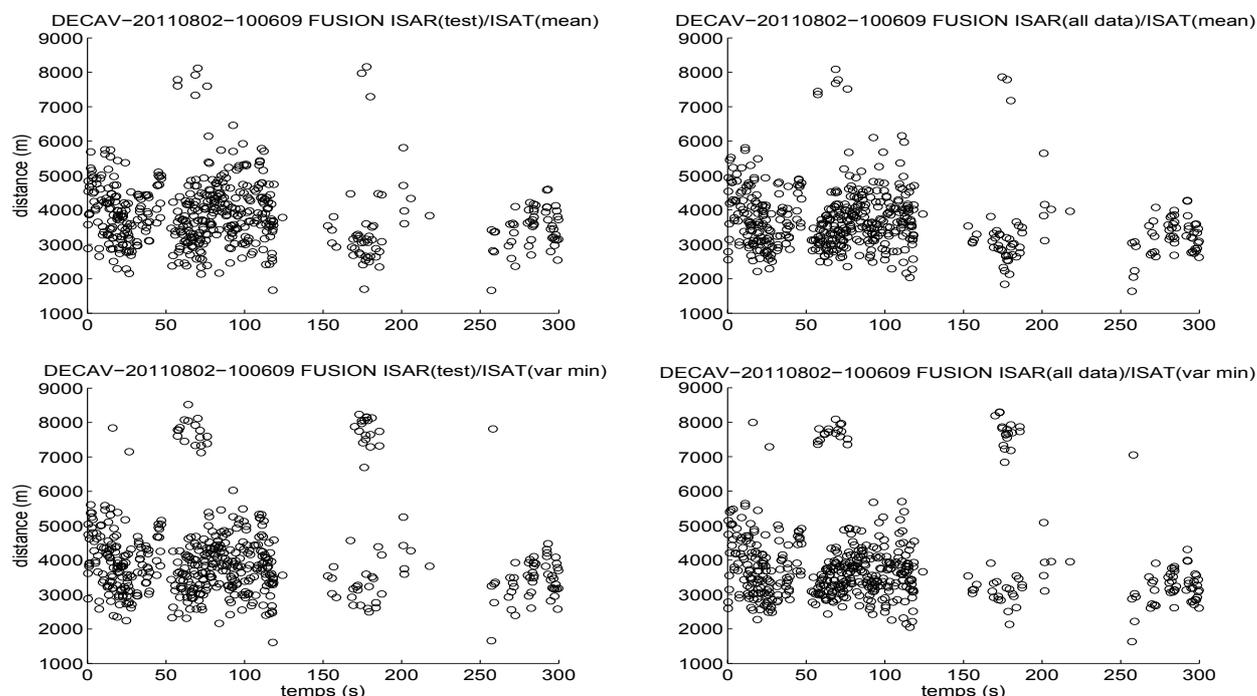


FIGURE 7.18 – Prédictions temporelles ISAR de distance sur l'enregistrement du 02/08/2011 représentées par des cercles.

7.4.2 Traitement des prédictions par bi-histogrammes temps/distance

Sur une durée de 5 min (dynamique de distance faible de l'animal), nous obtenons des prédictions relativement stables et localisées en distance. Nous retrouvons sur le bi-histogramme, en figure 7.19, 2 pics important en 4000 m et 4500 m pour le mélange train/test dissociés, en 3000m et 3500 m et 4750 m pour l'apprentissage sur toutes les données. *ISAT2b* donne des pics significatifs autour de 8000 m. Ces représentations semblent indiquer 3 individus.

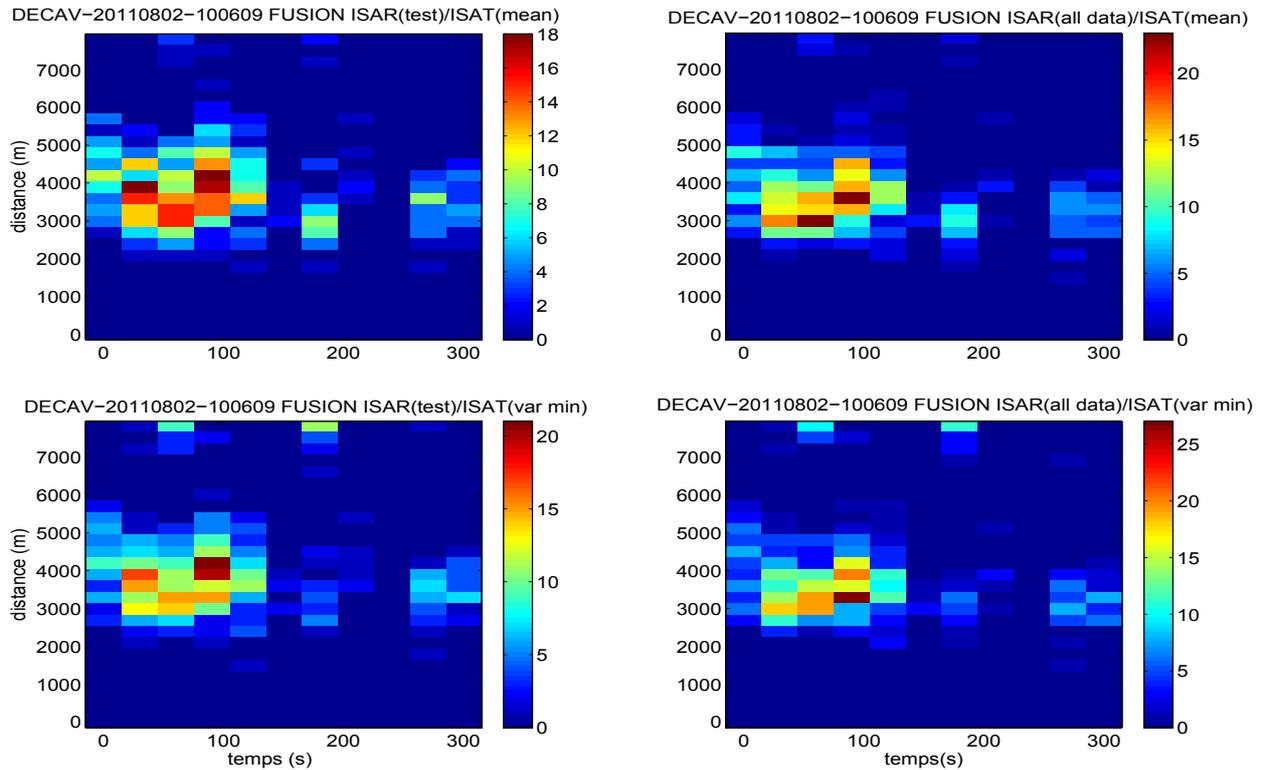


FIGURE 7.19 – Bi-histogrammes temps/distance des prédictions temporelles ISAR sur l'enregistrement du 02/08/2011.

7.4.3 Répartition des prédictions

Les 4 histogrammes 7.20 montrent des blocs clairement distincts. Le plus important ($>5\%$) s'étale entre 3000 m et 5000m, le second concerne les distances entre 7500 m et 8000 m. Ce dernier est plus marqué (environ 4%) par l'emploi de *ISAT2b*. L'analyse des histogrammes confirme le caractère multisource de cet enregistrement : entre 2 et 3 animaux. Le modèle utilisant ISAR appris sur l'ensemble des données test donne des distance inférieures à celle prédites par le modèle appris sur le mélange dissocié. Ce modèle à tendance aussi à faire ressortir un troisième pic autour 5000 m.

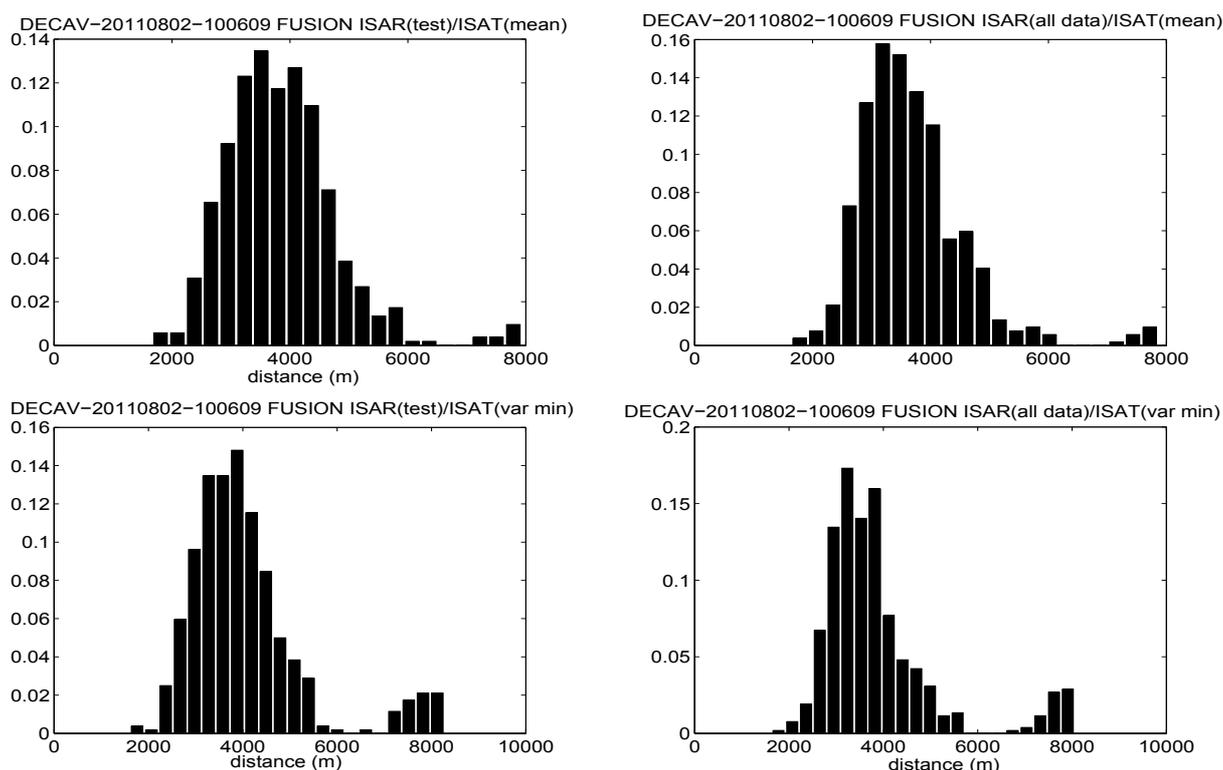


FIGURE 7.20 – Histogrammes normalisé des prédictions ISAR sur l’enregistrement du 02/08/2011.

7.4.4 Synthèse des résultats sur les données DECAV en présence d’un seul cachalot

Les fichiers issus de la campagne de mesures du 23/09/2012 contiennent 1 seul cachalot. Plusieurs opérateurs humains le confirment après écoute. Nous représentons en figure 7.21 une moyenne glissante (lissage) des estimations afin d’adoucir les variations de la courbe. Une regression linéaire est alors effectuée sur le lissage. Le coefficient directeur de la droite (en rouge sur la figure 7.21) représente alors la vitesse radiale moyenne de l’animal au regard de nos estimations. Celle-ci doit être inférieure à la vitesse de l’animal pour une trajectoire quelconque, égale pour une trajectoire radiale.

Sur ces 3 enregistrements consécutifs dans le temps sur une durée de 15 min, nous obtenons des estimations cohérentes entre elles. L’animal est situé à environ 3000 m et les vitesses radiales moyennes sont respectivement de -0.7, 0.85, -2 m/s ce qui donne une moyenne de -0.61 m/s. Cette vitesse est cohérente avec la vitesse moyenne d’un cachalot qui est de l’ordre de 1 m/s suivant la littérature. Cette vitesse est négative, ce qui suggère que l’animal s’est éloigné sur 15 min. Le détail des estimations fichier par fichier permet de conjecturer que l’animal serait à l’origine à 3000 m et se rapproche de l’hydrophone durant les 5 premières minutes jusqu’à 2700 m, puis s’éloigne jusqu’à 3000 m pendant les 5 min suivantes. Enfin sur le dernier fichier, nos estimations aboutissent à une vitesse radiale négative importante (2 m/s) et pourrait révéler une accélération de l’animal due à un comportement de chasse ou de fuite par exemple.

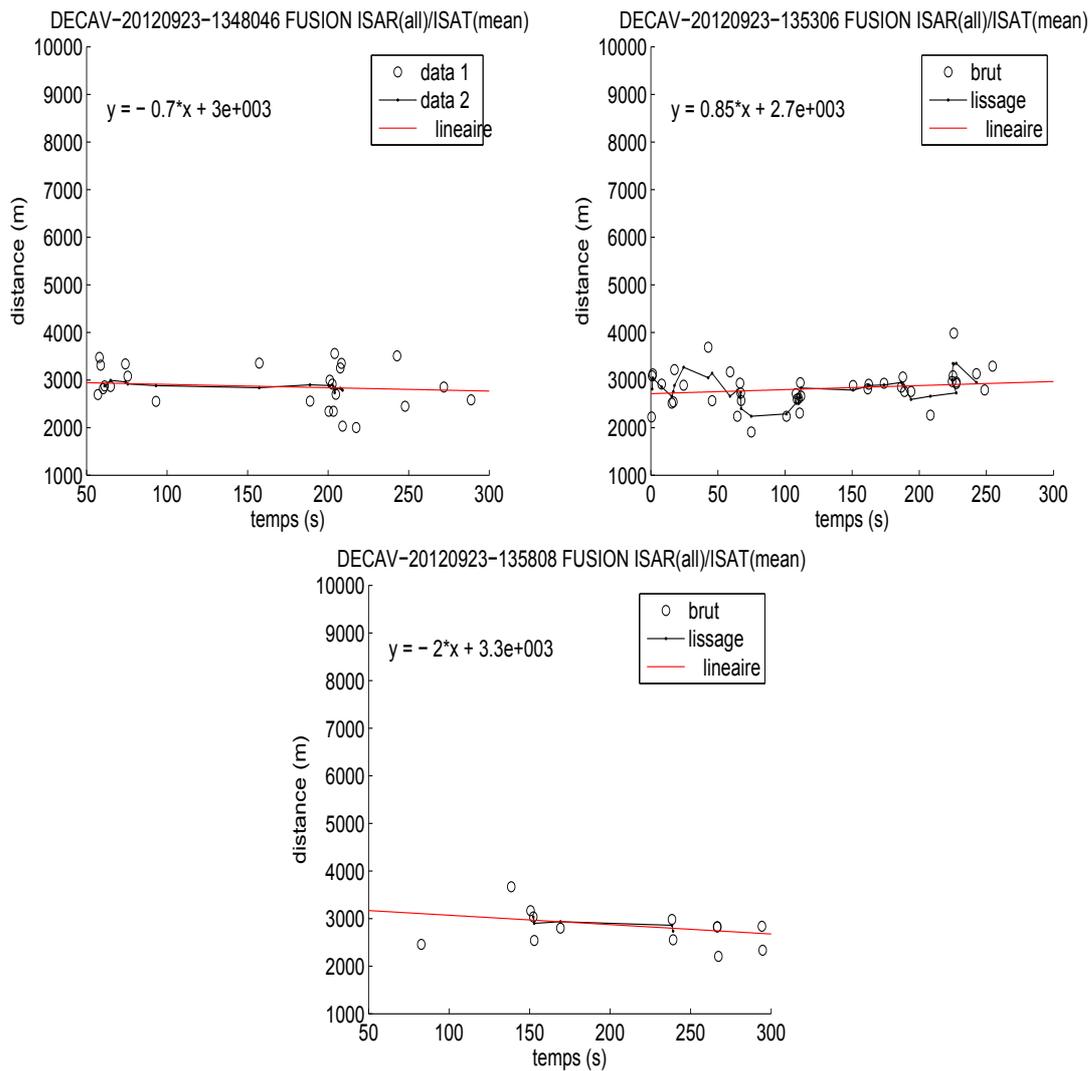


FIGURE 7.21 – Représentation temporelle des prédictions obtenues sur 3 enregistrements consécutifs contenant 1 animal.

7.4.5 Synthèse des résultats sur les données DECAV en présence de plusieurs cachalots

Trois autres fichiers sont traités avec le modèle de fusion appris sur l'ensemble des données BAHAMAS SET A. Le tableau 7.10 donne les résultats après analyse manuelle des histogrammes des prédictions et synthétise les données apportées par l'article [Abeille 14]. Les représentations graphiques des prédictions sont données en figure 7.22. Les \hat{r}_i sont les distances probables source-récepteur.

7.4. Résultats préliminaires sur les données issues du protocole DECAV

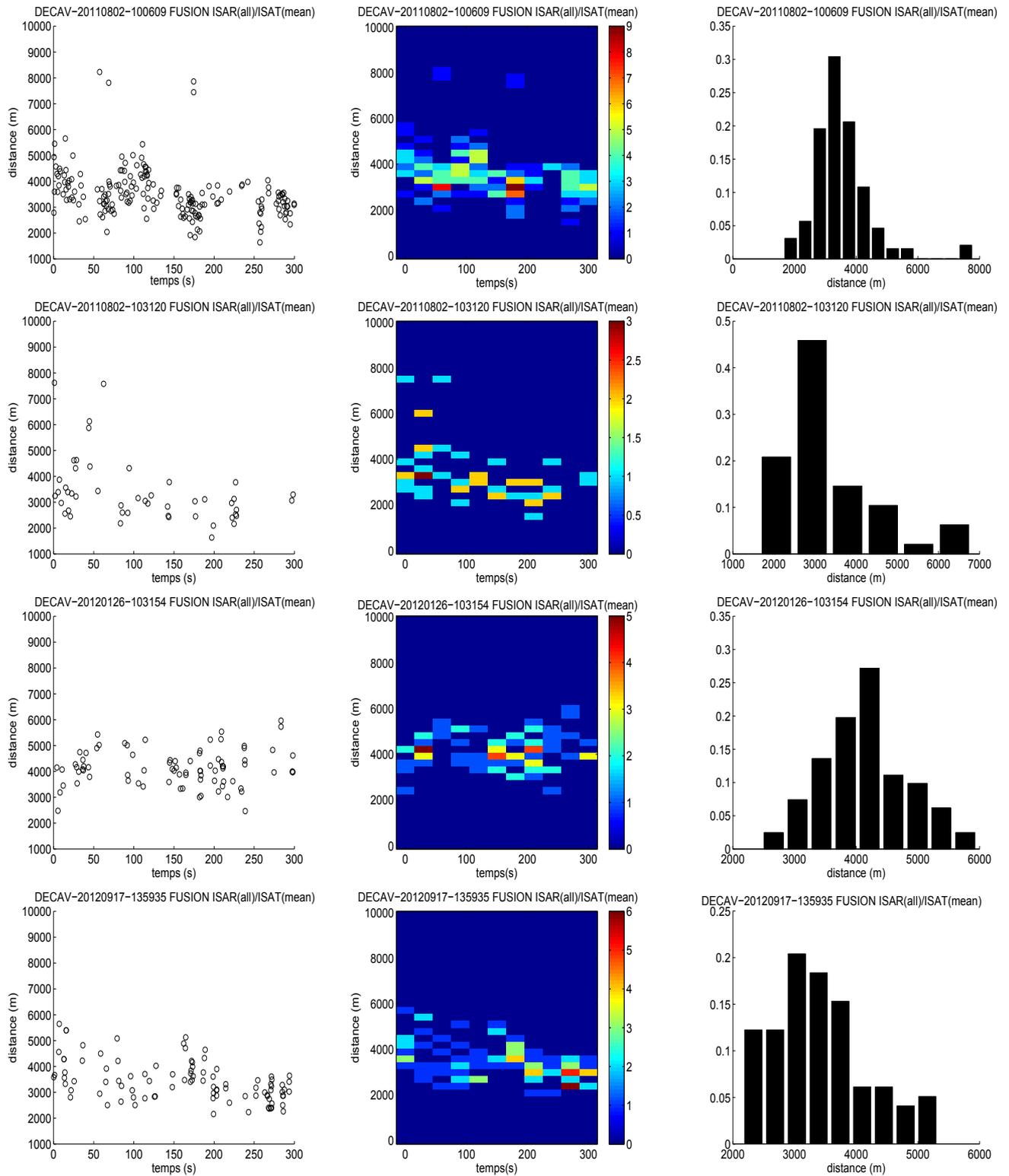


FIGURE 7.22 – Représentation temporelle, bi-histogrammes et histogrammes des résultats obtenus sur les fichiers traités parmi les données DECAV.

fichier	taille (m)	\hat{r}_1	\hat{r}_2	\hat{r}_3	\hat{r}_4
20110802-100609	11.45	3000	5000	8000	–
20110802-103120	11.45	3500	6000	7500	–
20120126-103154	11.21/11.46	4200	5000	–	–
20120917-135935	11.52/11.60	2500	3000	3500	5000
20120923-1348-1358	12.93	3000	–	–	–

TABLE 7.10 – Synthèse des résultats sur les données DECAV. Les distance sont en mètre.

7.5 Résultats préliminaires sur un extrait des données BOMBYX

Nous avons procédé à des estimations de distance selon le modèle de fusion ISAT2/ISAR sur un ensemble de clics de cachalots consécutifs sélectionnés sur un extrait des enregistrements de BOMBYX d'environ 3 min. Nous notons que les clics sont de bonne qualité mais chevauchés par des sifflements de dauphins pouvant impacter la forme des spectres. Les résultats sont présentés en figure 7.23.

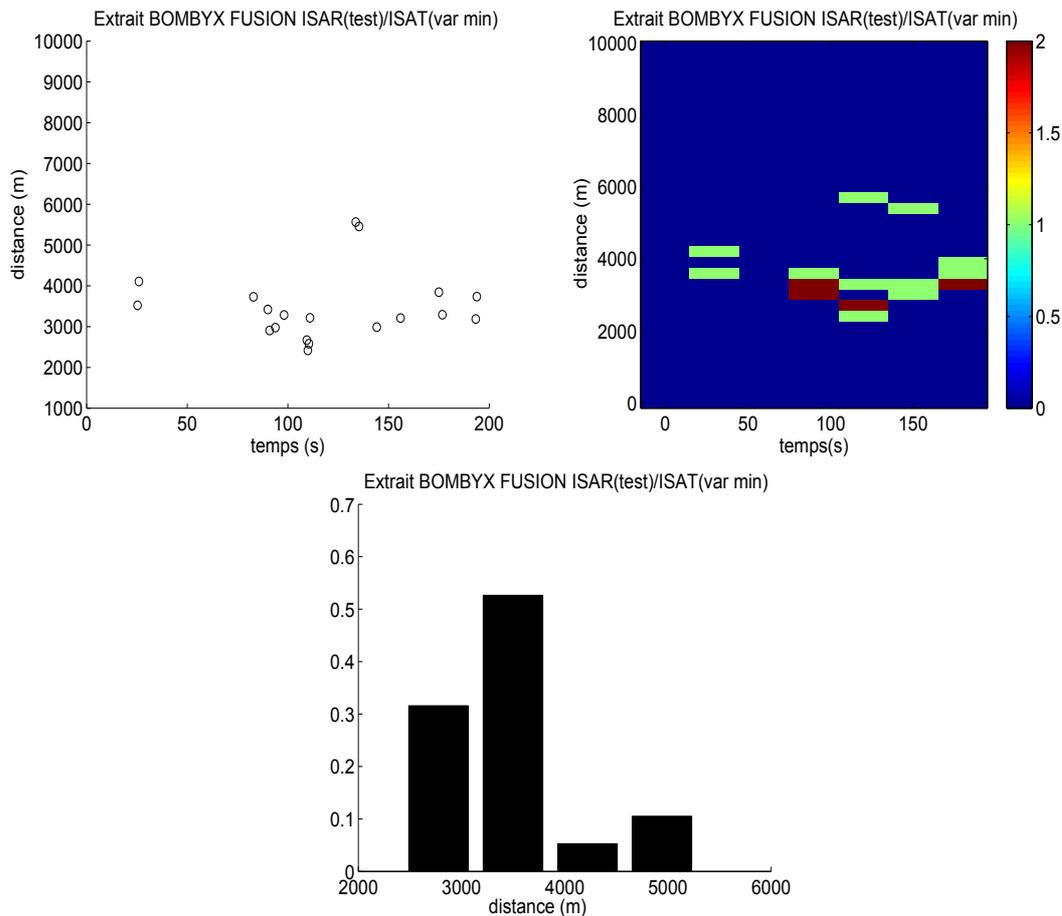


FIGURE 7.23 – Synthèse des prédictions obtenues sur 200 s d'un extrait des données Bombyx.

Nous obtenons des estimations fortement localisées formant 2 traces dont la dynamique est faible sur 200 s. Les histogrammes et bi-histogrammes donnent deux pics clairement dissociés ce qui suggèrent la présence de deux individus : le plus proche situé entre 3000 et 3500 m et le plus éloigné entre 5500 m et 6000 m.

Enfin, nous avons procédé à un calcul d'estimation de distance sur un ensemble de détection provenant de 41 heures consécutives d'enregistrement dont le spectrogramme à long terme est représenté en figure 7.24. Nous représentons sur la figure 7.25, les estimations de distances médianes par fenêtres temporelles de 20 min. Jusqu'à la date $t_{20} = 70$ soit 14h. La majorité des détections est composée de sifflements de dauphins et de bruits impulsifs dus à la manipulation de la structure. Ces signaux, dont les spectres sont éloignés des spectres de pulses de cachalot, induisent des estimations quasi constantes (autour de 4800 m pour le modèle de fusion, autour de 5000 m pour le modèle ISAR). Entre $t_{20} = 70$ et $t_{20} = 110$ soit 3h15, les détections sont partiellement composées de clics de cachalots. Nous observons que les prédictions varient entre 3700 m et 6200 m pour ISAR et le modèle de fusion.

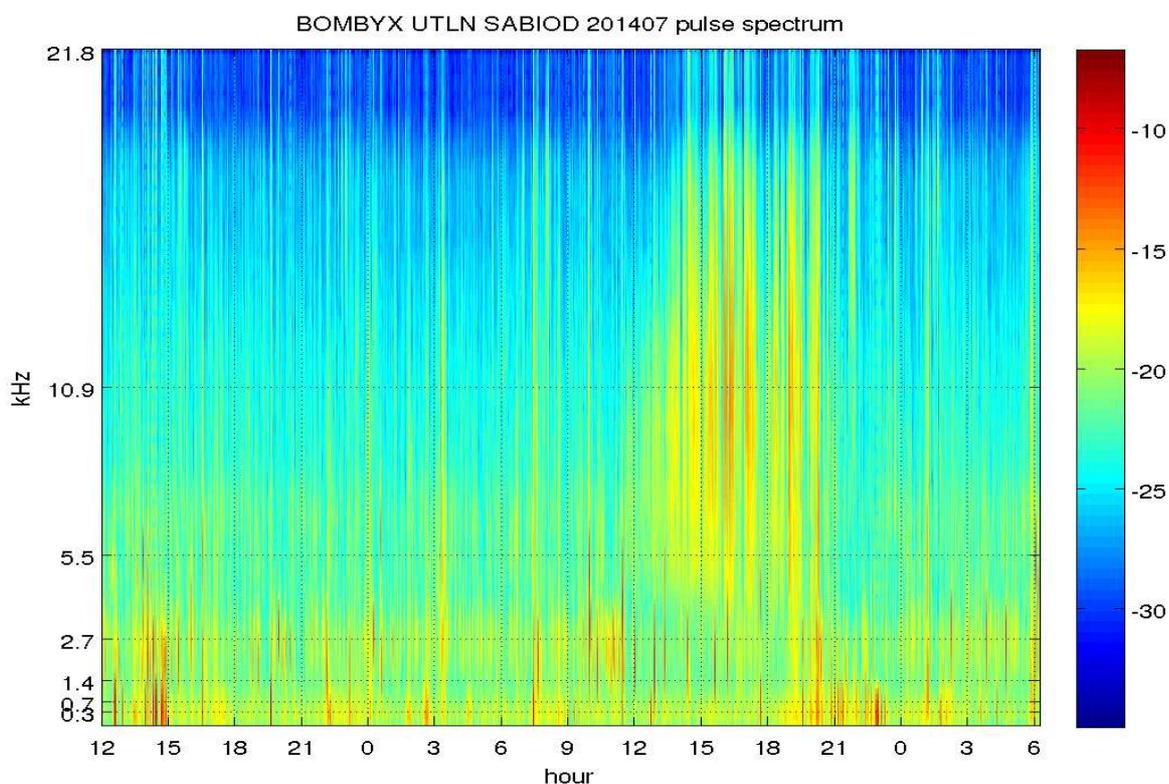


FIGURE 7.24 – Spectrogramme long terme de BOMBYX SABIOD représentant 20 détections par tranche de 5 minutes (analyse Dyni SABIOD). La majorité des détections de cachalot apparaît entre 14h et 21h du second jour bombyx en juillet 2014.

Les allures des courbes générées par les 2 modèles sont en accord entre elles et semblent cohérentes avec une phase d'approche de l'animal jusqu'au point le plus proche à 19h00, soit environ 3900 m de l'hydrophone selon ISAR et le modèle de fusion, puis une phase d'éloignement progressif. Le modèle de fusion ISAT2/ISAR montre toutefois plus de variations que ISAR, cela pourrait être lié à des effets de directivité du biosonar de l'animal que le modèle ISAT2 ne contrôle pas. Ces résultats nous permettent de tester le comportement de notre estimateur face à des signaux différents des clics de cachalots. La poursuite de cette analyse avec un niveau de résolution temporel supérieur pourrait confirmer ces prédictions.

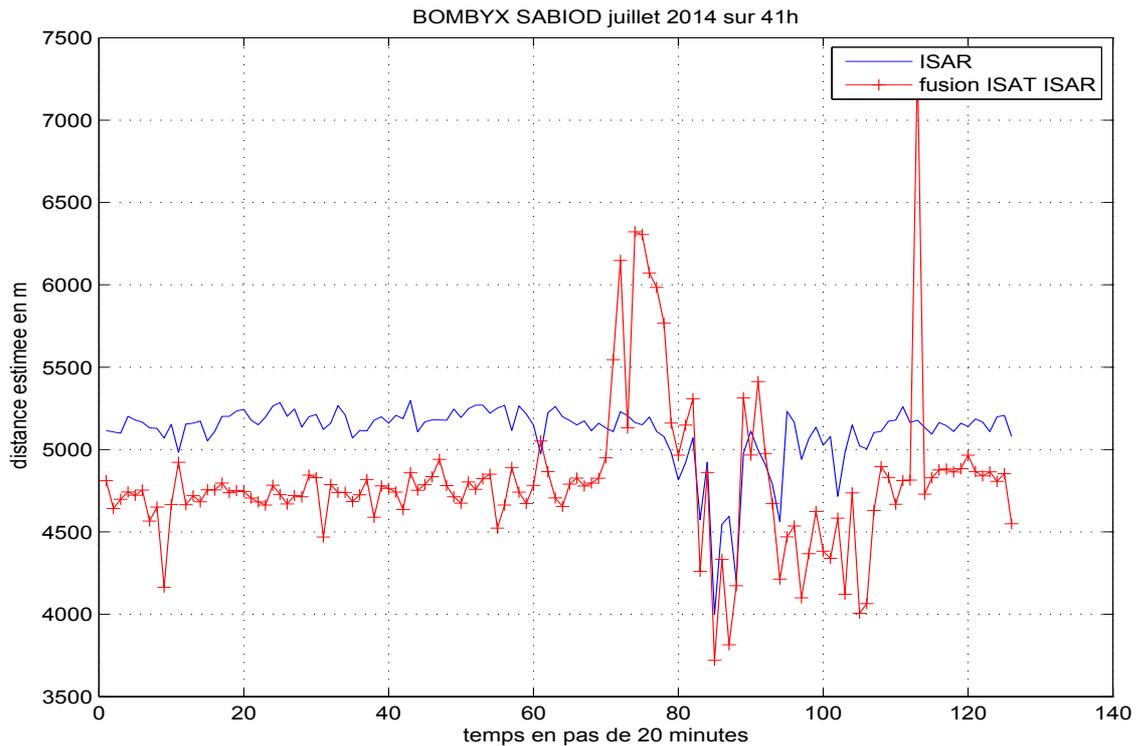


FIGURE 7.25 – Synthèse des prédictions de distance médianes par pas de 20 min sur 41 h des données BOMBYX par les modèles ISAR et fusion ISAT2/ISAR.

Les résultats d'estimation de distance pour le modèle LER sont donnés à titre indicatif en figure 7.26. La dynamique de la courbe est similaire aux deux modèles précédent. une réponse stable autour de 4500 m en l'absence de détection de clic et une réponse variant entre 2000 m et 4500 m en présence de détection. Le modèle LER confirme sa tendance à prédire des distances faibles avec une estimation en 2000 m au point le plus proche.

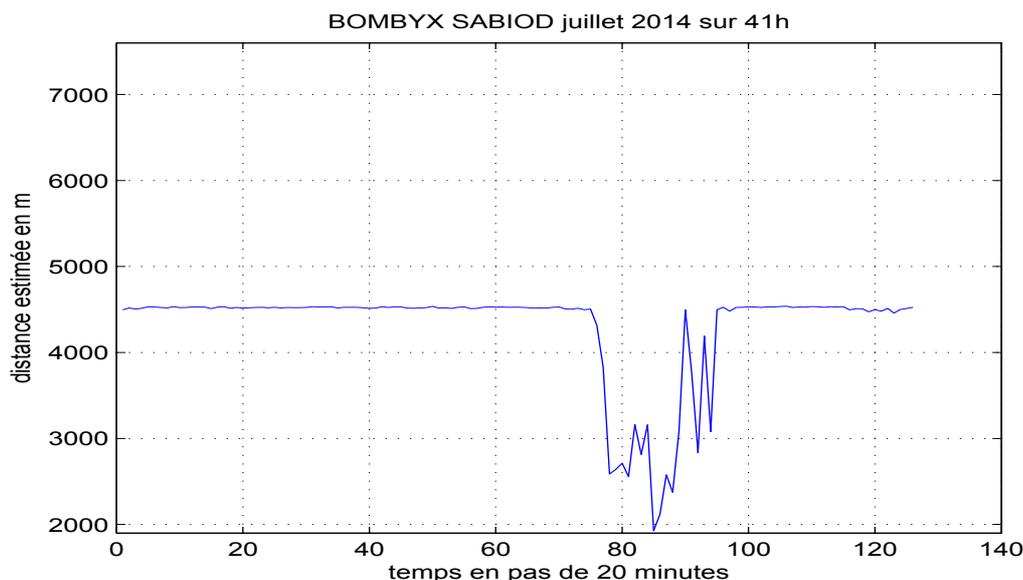


FIGURE 7.26 – Prédications de distance médianes par pas de 20 min sur 41 h des données BOMBYX par le modèle LER.

7.6 Conclusion

Ce chapitre présente les résultats obtenus sur le volet estimation de distance appliquée au cachalot. Nous avons abordé ce problème en proposant 2 approches différentes.

1. La première approche est basée sur une relation acoustique physique qui relie l'atténuation fréquentielle, responsable en partie de la déformation intra-spectre à la distance de propagation (ISAT, ISAT2a, ISAT2b). Les résultats sont organisés de la manière suivante :
 - optimisation, sur les données BAHAMAS SET A dédiées au développement, des paramètres de lissage de spectre et du nombre de combinaisons retenues pour le calcul d'ISAT. Les valeurs retenues sont respectivement : $N_{best}^* = 4$ et $F_{P_{256}}^* = 9$ kHz,
 - estimations temporelles ISAT sur l'ensemble dédié au test de BAHAMAS SET A. La performance est mesurée par la RMSE entre les estimations et la vérité terrain. Nous avons obtenu une valeur de 790 m,
 - analyse de l'information mutuelle entre les rapports d'énergie et les distances de référence. Nous donnons donc les résultats issus d'une sélection empirique des combinaisons de bandes d'énergie (ISAT2a/ISAT2b). Les résultats sont de qualité similaire avec un temps de traitement informatique fortement diminué (50 %).
2. La deuxième approche est basée sur l'emploi d'un réseau de neurones (MLP) organisé pour apprendre la relation entre le spectre du signal reçu et la distance source-récepteur (LER, ISAR, ISARH). Les expériences sont présentées ainsi :
 - définition, à partir des données BAHAMAS SET A, des espaces d'apprentissage, de développement et de test. D'une part les exemples des 3 espaces sont uniformément enchevêtrés et les entrées sont les modules bruts du spectre. D'autre part un mélange dont les exemples des 3 espaces sont temporellement dissociés et dont les entrées sont normalisées,

- résultats de la phase de développement lors de laquelle nous avons optimisé les hyperparamètres du MLP : N_{unit} le nombre d'unités de la couche cachée, N_{iter} le nombre d'itérations pour la phase d'apprentissage et L_r le taux d'apprentissage. Les valeurs sont variables suivant le type de mélange train/développement/test, l'espace numérique des poids et le passage du logarithme sur les entrées,
- résultats des estimations sur le mélange uniforme : LER aboutit à une RMSE de 542 m, 218 m pour ISAR et 208 m pour ISARH. Mais les conditions d'apprentissage sont idéales et le modèle n'est pas généralisable,
- résultats sur le mélange dissocié et entrées normalisées : ISAR fournit une RMSE de 775 m et 1045 m pour LER. La RMSE est fortement dégradée par les conditions d'apprentissage mais le modèle paraît généralisable. Dans ces conditions ISAR donne une RMSE proche d'ISAT. ISAR est toutefois plus performant à petite distance que ISAT.

Après le constat de ces premiers résultats, nous avons proposé un estimateur résultant de la fusion des estimations ISAT2 et ISAR par une moyenne pondérée. La pondération favorise ISAT à grande distance et ISAR à courte distance :

- résultats sur le mélange dissocié, le modèle aboutit à une erreur d'environ 566 m,
- observation d'une uniformisation de la RMSE par hydrophone.

. Le modèle de fusion ISAT2/ISAR est testé sur les données de BAHAMAS SET B (4 animaux émettant simultanément). Les résultats sont représentés selon 3 modes : la représentation classique temporelle des estimations, un bi-histogramme temps-distance et un histogramme des distances estimées. Nous avons alors montré que notre modèle permettait de dissocier entre 2 et 3 individus selon les cas et prédisait des distances cohérentes malgré un apprentissage sur une base de données restreinte. Une estimation de la vitesse radiale est calculée à partir d'une régression linéaire des estimations. Les vitesses radiales obtenues sur le SET A comme sur le set B coïncident en amplitude et en signe avec les vitesses de référence. La figure 7.27 montre la corrélation entre les vitesses radiales estimées et les vitesses radiales de références pour les résultats de BAHAMAS SET A et B. Nous notons un coefficient de l'ordre de 80 %. A titre indicatif nous avons disposé les estimations fournies par le modèle basique LER ne construisant pas de relation intra-spectre. L'amélioration apportée par le modèle de fusion intra-spectre est notable. Ces modèles d'estimation de distance et d'analyse de séquence permettront de construire des index de grandes collections bioacoustiques sous marines comme ceux discutés dans [Bénard 09].

Puis, nous poursuivons par une application du modèle de fusion ISAT2/ISAR sur les données DECAV, le milieu totalement différent du milieu d'apprentissage puisqu'il s'agit des eaux du PNPC. Nous ne disposons pas de données fiables faisant office de référence pour ce corpus, nous ne possédons que des données terrains de surface. Sur 7 sessions de mesures acoustiques positives aux clics de cachalot, nous estimons des distances comprises entre 2500 et 8000 m. De plus, nous estimons être en présence de plusieurs individus sur une majorité des sessions de mesures. Un cas où nous pensons que le cachalot est isolé, est traité séparément. Nous obtenons une estimation stable sur 15 min consécutives et une vitesse radiale cohérente comprise entre -0.7 et 2 m/s.

Enfin nous terminons par une application préliminaire du modèle sur un court extrait des données BOMBYX relevées fin octobre. D'une part, nous avons traité un échantillon de 200 s sur des détections de clic, nous obtenons 2 tracés clairement dissociés et nous estimons la présence de 2 cachalots situés à une distance de l'hydrophone comprise entre 3000 m et 6000 m. D'autre part, nous avons procédé à une estimation de distance "à l'aveugle" sur l'ensemble des détections obtenues sur 41 h. Les premiers résultats mettent en évidence la différence de comportement d'ISAR et du modèle de fusion face à des signaux autres que ceux émis par un cachalot (dauphins, bruits d'origine anthropiques...). De plus, ils coïncident avec une phase

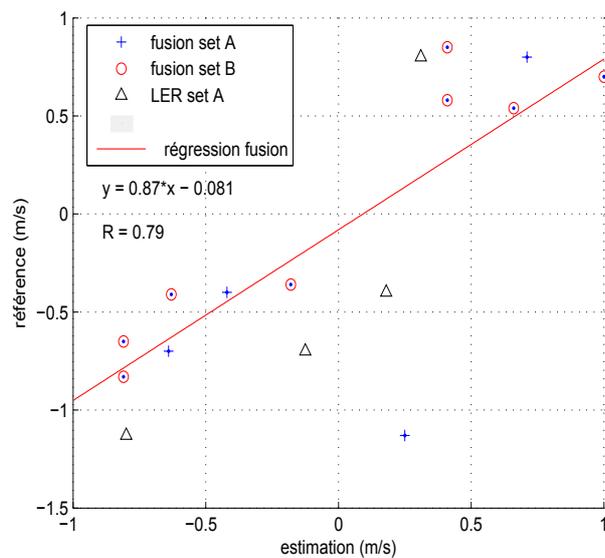


FIGURE 7.27 – Vitesse radiale estimée *vs* vitesse radiale de référence sur BAHAMAS SET A et B. La régression est effectuée uniquement sur le modèle de fusion.

d'approche de l'animal puis d'éloignement malgré un certain biais qui peut-être induit par effets de directivité du biosonar de l'animal. L'intégration d'un modèle du rayonnement acoustique pourrait réduire ces perturbations.

Application du codage parcimonieux sur des signaux stationnaires de baleine à bosse

Sommaire

8.1	Conditionnement des descripteurs	125
8.2	Apprentissage du dictionnaire	125
8.3	Extraction des codes de chant	127
8.4	L'évolution des codes de chant	128
8.5	Discussion	129
8.6	Conclusion	132

8.1 Conditionnement des descripteurs

Comme nous l'avons déjà mentionné, le vecteur d'entrée MFCC que nous utilisons n'est pas simplement les 13 coefficients statiques mais la concaténation de x vecteurs MFCC consécutifs. En effet, nous avons calculé les 13 coefficients statiques toutes les 10 ms (fenêtre d'analyse) mais nous voulons analyser une séquence plus longue. Ainsi, nous concaténons x occurrences entre-elles dans un « super-vecteur » MFCC. Chacun des ces vecteurs va alimenter le codeur parcimonieux. Nous obtenons alors un modèle à $10x$ ms. Cependant, l'échelle de temps des modèles concaténés dans ce « super-vecteur » varie de 250 ms, 500 ms, 1 s, 2 s à 4 s. Ces super-vecteurs sont aussi calculés toutes les 10 ms, au lieu d'un vecteur d'entrée MFCC à 13 dimensions, nous travaillons avec un vecteur MFCC de dimension $13 \times 50 = 650$ MFCC.

8.2 Apprentissage du dictionnaire

Pour cette étude, nous avons travaillé sur les séries d'enregistrements de 2008 et de 2009. Notre objectif est de travailler avec une représentation plus discriminante que les données des vecteurs complets. Au lieu d'utiliser directement les vecteurs MFCC usuels, nous introduisons un codage parcimonieux. Ainsi, nous avons appris un dictionnaire D sur des ensembles de 2008 et de 2009 de manière non supervisée.

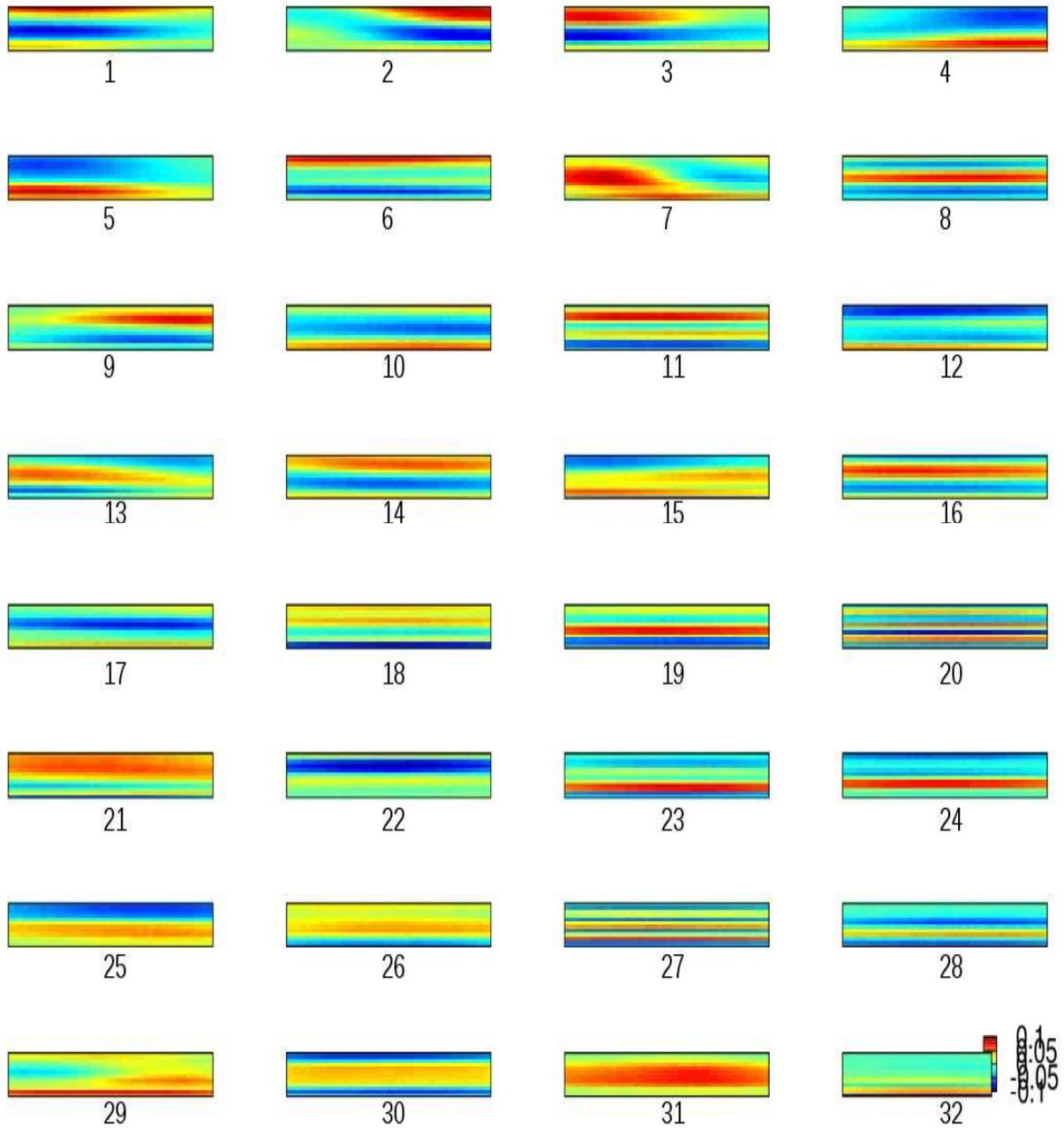


FIGURE 8.1 – Dictionnaire de 32 codes, triés par degré de complexité, calculés en utilisant une échelle de temps de 250 ms et appris sur l’union de sous-ensembles de chants de 2008 et de 2009. L’échelle de couleur de la carte est de -1 pour le bleu et 1 pour le rouge.

Un des inconvénients du codage parcimonieux est que la taille du dictionnaire doit être fixée manuellement. Cette taille devrait être supérieure au nombre de classes attendues après le regroupement, le dictionnaire sera ainsi *sur-complet*. Dans cette expérience nous avons appris deux dictionnaires : un avec $K = 16$ vecteurs et l'autre avec $K = 32$ vecteurs. Dans ce qui suit, nous appellerons un vecteur de ces dictionnaires le vecteur parcimonieux (sparse vector SV) (figure 8.1). Ainsi, un code parcimonieux est la projection d'un vecteur MFCC d'entrée sur les vecteurs d'un dictionnaire selon la contrainte de régularisation.

Après l'apprentissage du dictionnaire, nous pouvons projeter les vecteurs MFCC de 2008 et de 2009 sur le dictionnaire et calculer les codes parcimonieux associés. Le dictionnaire a été appris sur l'union des ensembles de 2008 et 2009.

A partir de ces paramètres, nous allons calculer et analyser la distribution statistique des codes parcimonieux associés aux unités des chants de baleines. Ensuite, nous allons comparer les changements et l'évolution entre les enregistrements de 2008 à 2009 en utilisant les métriques de la théorie de l'information définies à travers le chapitre 5.

8.3 Extraction des codes de chant

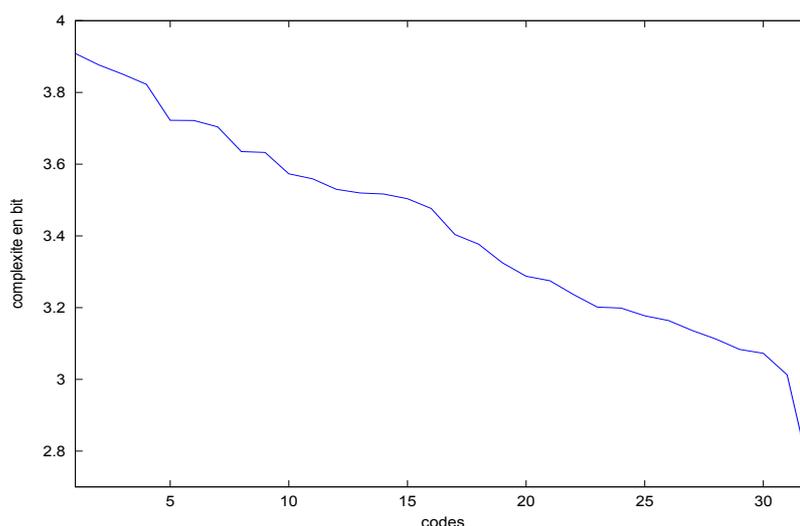


FIGURE 8.2 – Valeurs de la complexité des 32 codes (triés) du dictionnaire illustré en figure 8.1 (sur une échelle de temps de 250 ms, appris sur des sous-ensembles de chants de 2008 et de 2009). En abscisse, l'indice des codes, en ordonnée la complexité en bits. La différence entre les codes les plus complexes et les moins complexes est importante.

Nous avons analysé simplement la complexité d'un mot par son entropie sur tout le plan temps-quérence comme mentionné en section 5.5. Dans la figure 8.1, nous donnons les 32 vecteurs parcimonieux (du plus complexe au moins complexe, leur complexité étant sensiblement différente que celle illustrée dans la figure 8.2) du dictionnaire appris. Nous voyons clairement que la structure des vecteurs parcimonieux du haut est plus complexe que celle des premiers codes en partant du bas. Ces vecteurs correspondent au bruit ambiant, signaux entre 2 vocalises successives du chant analysé : mélange de bruits d'origine naturel, environnemental, de mesure et très rarement d'activités humaines, le trafic maritime dans cette zone est très réduit. Par ailleurs, le code le

plus complexe doit être composé d'une autre source. En effet, comme les données sont composées de bruit de mer ou bien de chants plus du bruit de mer, nous concluons que les codes les plus complexes correspondent aux composantes des chants.

La section suivante permet de comparer l'évolution entre 2008 et 2009 de la composition des chants en fonction de ces éléments.

8.4 L'évolution des codes de chant

Nous calculons la distance KL entre les codes de 2008 et de 2009. Il en résulte 32 distances. L'histogramme de ces 32 distances est donné dans la figure 8.3. Il montre clairement que des représentations de courte durée (250 ms) sont plus stables au fil des ans que des représentations plus longues.

Afin de déterminer quel code est en évolution, nous calculons la distance de code entre 2008 et 2009, en moyenne sur les 2 codes plus complexes, par rapport aux 2 codes moins complexes. Cette analyse de divergence (*cf.* figure 8.4) montre que les ensembles de chants de 2008 et de 2009 contiennent de la même manière les codes les plus simples. Ceux-ci correspondent aux parties sans unités sonores, c'est-à-dire présentant principalement du bruit de mer. Nous voyons clairement que le code le plus simple est presque toujours à la même distance (faible). Il est fort probable que le codage du bruit de fond soit semblable car les enregistrements sont pris exactement au même endroit chaque année. Au contraire, les codes les plus complexes sont éloignés pour une échelle de temps d'une seconde, ce qui représente certainement l'échelle « culturelle » des chants. Par ailleurs, les codes utilisés les plus complexes, sont semblables à courtes échelle de temps (la distance KL est faible pour les 250 ms), mais différent à plus longue échelle de temps. La plus grande différence est observée pour une échelle de temps de 1s.

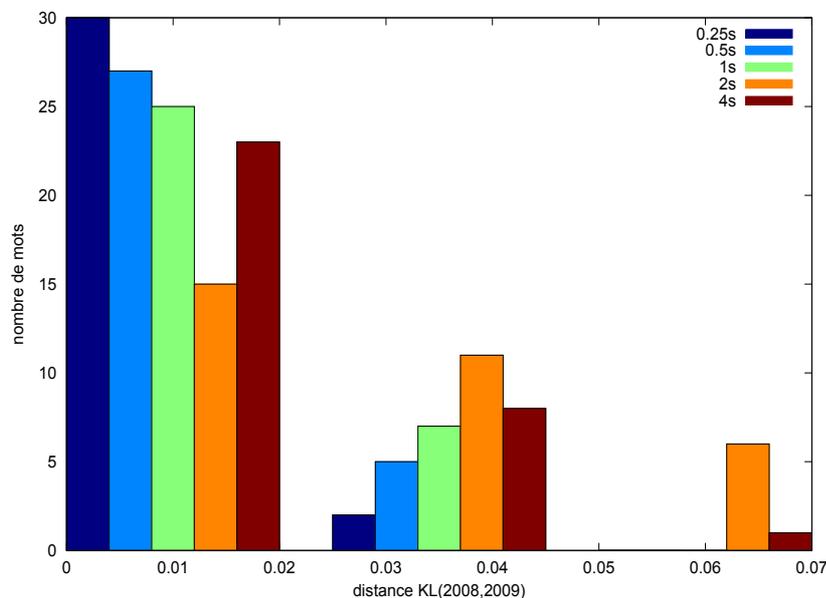


FIGURE 8.3 – Histogramme de la distance KL (2008, 2009) calculée sur les 32 vecteurs parcimonieux du dictionnaire. Il démontre que les représentations de courte durée sont plus stables au fil des ans que les plus longues.

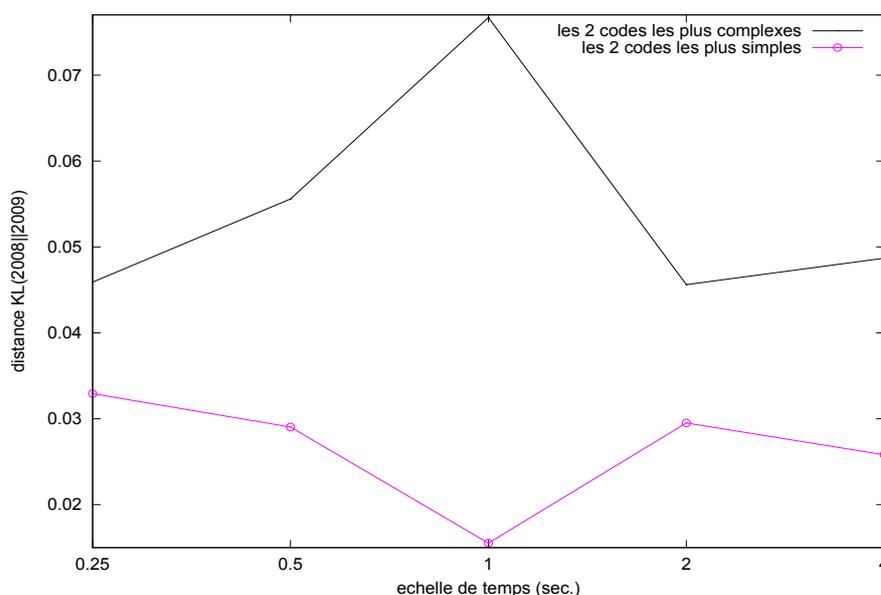


FIGURE 8.4 – Distance entre les chants de 2008 et 2009, en moyenne sur les deux codes les plus complexes, contre les deux les plus simples. En abscisse l'échelle de temps du code, de 1/4 à 4 secondes. En ordonnées, la distance.

Ceci suggère que les codes calculés sur l'échelle de 1 seconde dépendent des années et peuvent être composés de codes calculés sur des échelles de temps plus courtes. Ce résultat est compatible avec le concept de sous-unité [Pace 10], la sous-unité serait codée à l'échelle de 250 ms, tandis que les unités seraient codées à partir d'une échelle de 1 s. Les échelles de temps les plus longues (2 et 4 secondes) sont moins divergentes, probablement en raison du fait que cette échelle de temps est relative à la structure globale du chant. Les variations sont moins importantes qu'au niveau de l'unité.

Notez que le calcul des distances directement sur les coefficients MFCC bruts, donne comme prévu des distances insignifiantes, toutes semblables, quelle que soit la taille ou l'année des unités. Cet effet est connu sous le nom de « malédiction des grandes dimensions ». En effet, la dimension des vecteurs MFCC est de 650 (13 x 50) et selon [Beyer 99] tout calcul simple de distance entre n'importe quelle paire de vecteurs dans cet espace multidimensionnel donne une distance similaire.

8.5 Discussion

Nous montrons dans cette étude que la représentation parcimonieuse est efficace pour décrire des structures de signaux acoustiques complexes.

Pour des échelles de courtes durées, cette méthode peut générer un dictionnaire de sous-unités. Pour des échelles de durées plus longues, elle peut générer un dictionnaire d'unités qui varie d'une année à l'autre. En figure 8.5, nous observons l'activité de la variation du code (« mot »). Par exemple à 60 ms, SV 2 et 3 apparaissent en même temps pour générer une configuration complexe. La variation maximale des chants entre 2008 et 2009 est donnée pour un SV de 1 sec. Cette durée peut être considérée comme étant le niveau significatif de « l'unité » des chants si

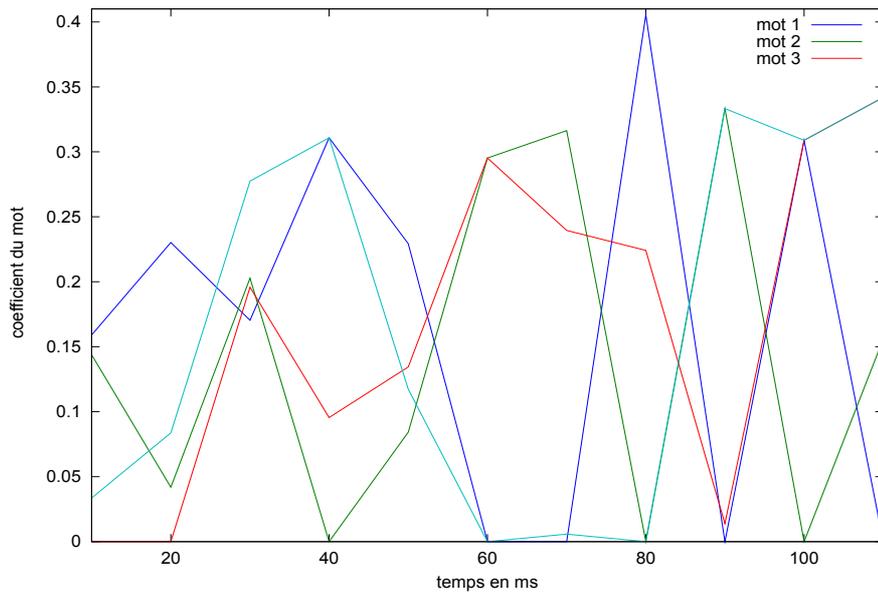


FIGURE 8.5 – Evolution dans le temps pendant 100 ms du coefficient de quatre vecteurs parcimonieux pour le dictionnaire de 16 unités, pour une partie de l’enregistrement de chants de l’année 2008.

l’on considère que l’ensemble des SV est le support des variations de chants sur plusieurs années. En contrepartie, les « sous-unités » qui sont stables au fil des années seraient générées à l’échelle de 0,25 s.

Afin d’analyser la composition des chants et leurs variations à partir des vecteurs résultant du code parcimonieux (également appelé mot), nous avons calculé la probabilité d’occurrence de chaque paire consécutive de vecteur parcimonieux. Notez que cette probabilité pour la paire (w_1, w_2) , notée $P(w_1, w_2)$, peut être égale à $P(w_2, w_1)$ dans le cas d’un système aléatoire. Les résultats montrent qu’au contraire elles diffèrent. De plus, si nous calculons le logarithme du rapport des probabilités des chants de 2008 par rapport à ceux de 2009, nous obtenons alors des variations claires de l’association des sous-unités à travers les années. Ceci est illustré dans la figure 8.6 pour $K = 16$ et une échelle de 500 ms. Les mots sont triés par entropie croissante et nous observons un gradient de la divergence bigrammique en fonction de la complexité des mots. Par exemple entre 2008 et 2009, nous voyons que la paire (6, 7) est plus fréquente en 2008 alors que la paire (6, 2) est moins fréquente en 2008. De plus, nous remarquons que le contraste est d’autant plus marqué que les années sont éloignées : entre 0 et 1.1 pour 2008/2009, entre 0 et 5 pour 2008/2014. Ces résultats sont en adéquation avec la bibliographie démontrant par mesure d’experts cétologues sur spectrogramme la dérive des chants de baleine [Noad 00a].

L’histogramme 8.7 représente la répartition de la divergence pour les 256 combinaisons de mots possibles et pour les 4 couples d’années (bigramme). Nous pouvons observer que le couple 2008/2014 montre davantage de valeurs élevées par rapport aux années rapprochées comme 2008/2009. L’allure des histogrammes confirme bien que les associations de mots évoluent en divergeant d’années en années.

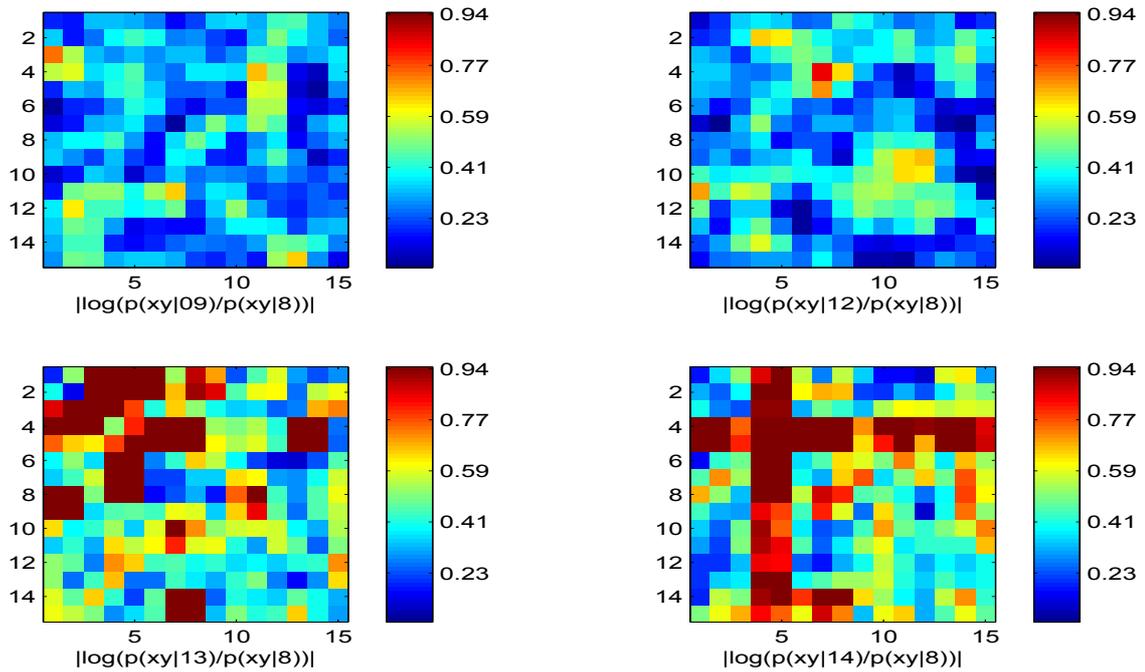


FIGURE 8.6 – Valeur absolue du logarithme du rapport des probabilités des bigrammes des chants de 2008 par rapport aux chants de 2009, 2012, 2013 et 2014. Ceci est illustré pour $K = 16$, échelle de temps de 500 ms. La représentation est convoluée par $[1 \ 1; 1 \ 1]$ pour lisser l'ordre des mots.

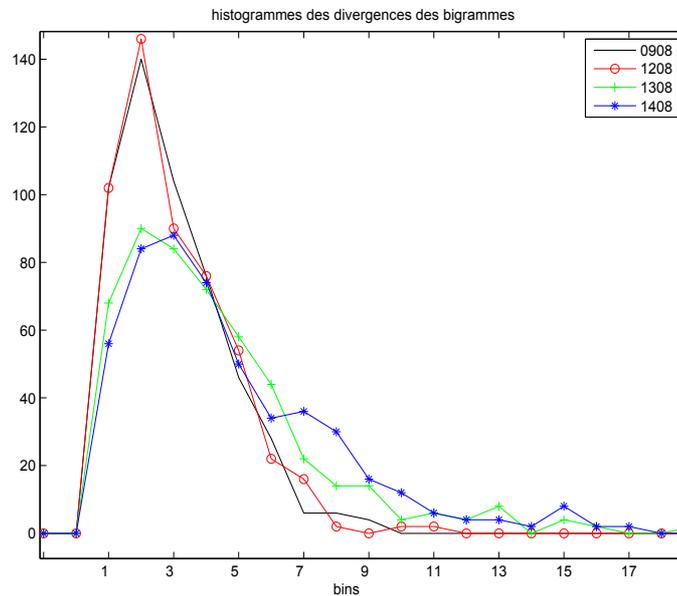


FIGURE 8.7 – Histogramme des bigrammes 2008 *vs* 2009, 2012, 2013 et 2014. Bin 0.5 = divergence nulle, les bins sont espacés de 1.125 unité de divergence.

8.6 Conclusion

Nous avons présenté un algorithme pour créer, à partir d'un dictionnaire non supervisé, un apprentissage d'un proto-lexique du chant des baleines à bosse à des échelles de temps différentes. En effet, dans l'étude précédente, la variation d'échelle de temps induisait des variations de dimension de la représentation dues, par exemple, à la concaténation de vecteurs MFCC. De plus, toute analyse par la méthode du plus proche voisin est biaisée sous l'effet de la dimension des espaces de vecteurs sur les métriques.

Différents types de codes ont été extraits des vecteurs MFCC sur des séquences monolocuteurs. A court terme, des codes peuvent être associés à des sous-unités, stables de 2008 à 2009, et peuvent composer de longues unités, elles-mêmes instables. De longues unités (1 sec) génèrent des variations de chants de baleines d'une année sur l'autre. Ces travaux sont l'illustration d'un changement des variations de chants.

Notre approche n'a besoin d'aucune connaissance sur la source : le codage parcimonieux reconstruit en priorité les événements les plus fréquents. En outre, le code résultant est parcimonieux : un petit nombre de vecteurs creux est utilisé pour chaque fenêtre temporelle. On peut donc supposer qu'un petit nombre de vecteurs creux construit chaque unité.

Quatrième partie

Discussion et perspectives

Synthèse, perspectives et conclusion générale

Sommaire

9.1	Discussion sur l'estimation de distance sur les signaux transitoires	135
9.2	Discussion sur le codage parcimonieux de voisements	137
9.3	Conclusion générale	138

9.1 Discussion sur l'estimation de distance sur les signaux transitoires

Nous avons proposé un modèle théorique d'estimation de distance source-récepteur basé sur l'atténuation inter spectrale (*ISAT*). Nous avons testé plusieurs processus de sélection des combinaisons de bandes impliqués dans le calcul de cet estimateur. Dans un premier temps, il s'agit d'une simple minimisation de la variance des estimations par fenêtres glissantes, puis, en contraignant un choix fixe sur l'ensemble de la séquence, des combinaisons assurant une forte valeur de l'information mutuelle entre ratio et distance quelque soit l'orientation de l'animal (*ISAT2*). On observe un gain combiné en précision et en de temps de calcul de l'ordre de 50%. La rapidité d'exécution permet d'envisager des estimations en temps réel.

Le recours à des réseaux de neurones a été particulièrement bénéfique. L'emploi du MLP sur le set A a permis d'apprendre la relation empirique entre le spectre des pulses des clics et la distance de propagation sur le jeu de données "Bahamas" sur un nombre d'itérations particulièrement raisonnable (apprentissage sur environ 300 itérations). Dans le cas d'un mélange uniforme train/dev/test, nous avons montré que *ISAT* pouvait estimer des distances avec environ 630 m de RMSE et que les modèles basés sur le MLP comme *ISAR* et *ISARH*, la réduit à environ 200 m. Ce gain significatif montre non seulement une meilleure robustesse de l'approche, mais également une possibilité de discriminer des individus actifs vocalement qui évoluent à des distances proches. Dans ce cas, parce que les spectres ne sont pas normalisés, il est probable que le MLP tire de l'information sur la distance à partir de l'énergie totale du signal. Ce type de méthode engendre des résultats assez précis mais ne peut s'appliquer uniquement dans la mesure où les prédictions sont réalisées dans le même cadre (mêmes espèces, paramètres de propagation acoustique et chaîne d'acquisition) que l'étape d'apprentissage. De plus la base de données

d'apprentissage doit être relativement exhaustive en terme de distance et de configuration azimut/élévation.

L'utilisation du mélange train/dev/test dissocié dans le temps et la normalisation des spectres en entrée du MLP dégradent la précision des prédictions. L'erreur RMSE est alors de 775 m. *ISAT* donne une RMSE similaire mais semble donner les meilleures estimations à grandes distances (> 4500 m) et *ISAR* est plus performant sur les petites distances. Nous avons donc amélioré la robustesse de l'estimateur en fusionnant *ISAT2* et *ISAR*. L'effet de cette fusion est de diminuer la RMSE moyenne sur tous les hydrophones et de stabiliser la RMSE de chaque hydrophone autour de la valeur moyenne soit 556 m.

Ce niveau de précision dans les estimations reste raisonnable pour la majorité des applications dont le but est surtout de donner un ordre de grandeur des distances plutôt qu'une distance précise. Finalement, l'objectif était de construire un modèle de prédiction de distance appris sur un site spécifique et exportable sur un site différent avec un matériel d'acquisition quelconque. L'application de ce modèle sur plusieurs sites dont les conditions initiales diffèrent plus ou moins fortement des conditions d'apprentissage, a permis de valider cette hypothèse.

Le modèle de fusion *ISAT2/ISAR* a été utilisé sur le set B (plusieurs cachalots). Les caractéristiques acoustiques de la colonne d'eau sont identiques à ceux de la phase d'apprentissage et les hydrophones sont placés sur le fond. Même si on observe des accumulations de points en certaines régions du plan temps/distance, il reste difficile d'obtenir des trajectoires distinctes. Les prédictions sont toutefois étalées en distance dans le cas de l'hydrophone 5 qui présente bien une dynamique de distance large (de 1500 m à 6000 m). Si cette méthode ne permet pas de séparer les individus, elle met toutefois en évidence le caractère multisources des séquences analysées et une échelle de distance minimale et maximale des sources acoustiques. Ces résultats pourraient permettre d'évaluer la densité de sources acoustique sur la zone de détection.

Lorsqu'il est possible d'observer une trajectoire que l'on peut attribuer à un individu, nous avons montré que la pente moyenne des estimations temporelles, liée à la vitesse radiale de l'animal, était cohérente en amplitude et en signe avec celles des données de référence. Le modèle semble donc assez fiable pour prédire si l'animal se rapproche ou s'éloigne du récepteur. Une vitesse radiale élevée suggère une trajectoire radiale à l'hydrophone.

L'application du modèle de fusion sur les données DECAV dont les caractéristiques acoustiques diffèrent totalement des conditions d'apprentissage n'a pas pu être comparée à la réalité (non disponible) mais semble donner des résultats exploitables. Les résultats obtenus sur l'animal isolé sont cohérents avec des estimations assez localisées et stables ainsi qu'une estimation de vitesse radiale en accord avec la vitesse moyenne d'un cachalot. Il est alors possible de déterminer si l'animal se rapproche de l'hydrophone.

Dans les cas d'acquisition mono-hydrophone en présence de plusieurs animaux, une amélioration de la résolution des estimations pourrait permettre de discriminer tous les clics détectés et de les attribuer à un individu. Une telle représentation pourrait répondre aux problématiques de séparation de sources dans la mesure où leur origine spatiale est distincte.

Pour les réseaux de plusieurs hydrophones, la relation entre distance et spectre sur l'ensemble des hydrophones pourrait être utilisée pour la localisation. En particulier dans les cas où les hydrophones sont très espacés et où la cohérence temporelle des voies n'est pas assurée. L'approche mono-hydrophone permet d'éviter le traitement de "cross corrélation" pour extraire les TDOA pouvant se révéler complexe dans certains cas. Les estimations de distance source-récepteur sur chaque capteur conduisent à des cercles dont les intersections représentent des positions estimées.

Par ailleurs, nous avons vu que les travaux sur l'estimation de densité des espèces peuvent s'appuyer sur une estimation de la distance source-récepteur souvent dans le cadre de détections monocapteur. Les méthodes de prédiction de distance automatique développées dans cette thèse

répondent à la problématique notamment lorsque l'animal émetteur est furtif ou qu'il n'apparaît pas en surface à proximité d'observateurs visuels en poste.

Notre modèle *ISAT* propose une approximation de la distance selon une expression analytique simple. Toutefois, les hypothèses admises pour y parvenir sont fortes voir abusives. Celles-ci pourraient être révisées. En particulier l'hypothèse du "dirac" (hyp. 2) avec une incrémentation dans l'estimateur du terme : $\int_{F_1'}^{F_2'} \epsilon_0(f)df - \int_{F_1}^{F_2} \epsilon_0(f)df$. Ce terme permettrait d'intégrer la forme réelle des pulses contenue dans les clics émis par l'animal. Cette information pourrait être approchée par des connaissances statistiques à la source (via le recours à la pose d'Acousondes par exemple) ou bien par l'élaboration d'un modèle de production sonore de l'animal [Doh 11a].

Le modèle *ISAT* produit des estimations locales de distance, une étape de filtrage particulière [Sanjeev Arulampalam 02] pourrait être envisagée pour améliorer la cohérence des résultats pulse par pulse.

Cette version avancée de *ISAT* pourrait être adaptée à d'autres types de signaux bioacoustiques tels que les sifflements de dauphin et les vocalises de mysticètes.

De plus, nous n'avons pas pris en compte les spécificités de la propagation acoustique sous-marine dans la colonne d'eau. L'approximation de propagation géométrique est un facteur de sous-estimation en sur-évaluant les longueurs des chemins acoustiques. Les échos surface/fond résiduels dans nos détections peuvent aussi induire en erreur l'estimateur. Une solution consisterait à traiter les échos comme une source d'information complémentaire plutôt que de les éliminer des détections.

Après une étape d'apprentissage, il serait possible d'appliquer les modèles basés sur les MLP pour d'autres espèces de la bioacoustique terrestre, utilisant des biosonars comme les chiroptères, musaraignes, certains oiseaux cavernicoles...

Enfin, Il serait intéressant d'adapter la méthode d'estimation *ISAT*, basée sur la déformation spectrale, sur d'autres signaux large-bande d'origine anthropique militaires ou civils tels que les canons à air utilisés pour la prospection pétrolière, les bruits impulsifs issus de travaux sous-marin (pile diving). Ce type de méthodes pourrait enrichir les techniques de tomographie par acoustique passive.

9.2 Discussion sur le codage parcimonieux de voisements

Cette étude montre l'intérêt de la méthode de codage parcimonieux pour la classification des chants de baleines à bosse. D'abord le dictionnaire parcimonieux est dynamique, appris automatiquement à partir des enregistrements et prend en compte dans un nombre limité d'éléments toutes les variations dans les unités sonores/sous-unités. Cette approche est optimale pour analyser des signaux avec des caractéristiques communes mais avec une variabilité due à chaque chanteur et au bruit ambiant. En somme :

1. la méthode est non supervisée et présente clairement un avantage pour analyser la variabilité importante (en temps et en fréquence et pour différents chanteurs) des unités sonores,
2. le dictionnaire est construit à partir d'un ensemble de données et le critère utilisé par le codage parcimonieux oblige à limiter le nombre des éléments de ce dictionnaire,
3. grâce à l'algorithme LASSO, le codage parcimonieux met l'accent sur les principales caractéristiques des unités sonores et permet d'éliminer les composantes du bruit,
4. cette approche pourrait être appliquée sur les unités et sous-unités,
5. nous démontrons l'évolution du comportement d'une année à l'autre pour les sous-unités et les unités.

Les résultats confirment l'intérêt du codage parcimonieux pour classifier les chants de baleines à bosse.

Si l'importance des chants dans la régulation mâle-femelle est connue depuis de nombreuses années, les recherches s'appuyaient sur des méthodes d'analyse dites non automatiques. La méthodologie présentée ici est entièrement automatique et révèle la complexité de chaque composante du chant, d'année en année.

Notre étude encourage également le concept de sous-unités sonores. Nous montrons en effet que les unités les plus courtes (sous-unités) sont les plus stables, survenant avec une fréquence similaire sur deux années consécutives, tandis que les unités les plus longues sont clairement différentes d'une année à l'autre. Une analyse systématique basée sur la théorie de l'information sera utilisée dans des travaux futurs. Une autre application pourrait se trouver dans la modélisation de l'identité vocale des baleines, conduisant à une identification du chanteur ou à l'analyse de dialecte, *ie.* leurs règles de combinaison qui pourraient dépendre de l'individu ou du groupe qu'elles fréquentent la plupart du temps.

Les travaux à venir consisteront à appliquer cette approche à d'autres ensembles de données provenant d'autres régions et nous aimerions analyser les sons provenant d'autres espèces. L'article [Paris] vise à utiliser le codage parcimonieux sur des signaux transitoires de cachalot pour estimer la distance entre l'hydrophone et l'animal. Enfin Glotin a montré que ce type représentation peut servir à l'analyse de TDOA et résoudre des problèmes de localisation [Glotin 11b].

9.3 Conclusion générale

Ce travail de thèse a permis de proposer des approches originales dans le traitement des émissions sonores émis par certaines espèces de cétacés (cachalots, baleines à bosse) et dans leur localisation par acoustique passive. Le tableau 9.1 synthétise les méthodes et les applications résultantes de cette étude.

Contribution	Application	commentaires
Analyse spectrale (LER)	Estimation de distance	peu précis : à 1 km près peu adapté aux variations du SL
Analyse Intra-Spectrale (fusion ISAR/ISAT)	Estimation de distance	500 m d'erreur
	Dénombrement de source	Vitesse radiale si animal isolé Densité de sources acoustiques
Sparse Coding	Compression	Codes parcimonieux
	Segmentation	Bonne corrélation entrée/sortie Dictionnaire non supervisé Analyse séquences de chant Représentation invariante sur signaux transitoires [Paris]

TABLE 9.1 – Synthèse des méthodes et applications des contributions apportées lors de la thèse

Des travaux complémentaires pourraient montrer facilement leur généralisation à d'autres espèces, dont les émissions sonores présentent des caractéristiques voisines (signaux transitoires de type clics pour tous les odontocètes, signaux présentant une fréquence fondamentale forte chez les mysticètes). Il s'agit ainsi de venir compléter les études de détection par des estimations sur la localisation dynamique des individus présents autour de l'hydrophone. Même si la détection est

une étape essentielle dans l'observation des cétacés (elle renseigne sur leur présence et il est donc possible de recenser l'espèce, et éventuellement de caractériser leurs comportements voire leurs habitats), la localisation permet de les suivre à distance et de rendre compte de leurs activités, notamment de plongée. Cela est intéressant pour toutes les espèces, mais en particulier pour les espèces dites "grand plongeur" telles que les cachalots, les globicéphales, les baleines à bec... Ces espèces passent très peu de temps à la surface et sont particulièrement discrètes, rendant difficile les observations visuelles (sans parler des mauvaises conditions météorologiques ou des observations nocturnes).

Ces outils ont pour vœux d'aider les suivis des cétacés, notamment dans des zones spécifiques, comme les aires marines protégées ou des espaces côtiers où les activités humaines peuvent être importantes, et pourraient impacter, à différents niveaux, des cétacés, qu'ils soient résidents ou de passage.

Il semble que les nuisances voire les pollutions issues des activités humaines sont en croissance. Mais il y a aussi de plus en plus d'observateurs. Et mon travail fait partie des outils qui pourraient être mis à disposition de ces observateurs pour les aider dans leurs travaux, contribuer à la création de connaissances, et influencer sur les mesures indispensables à prendre pour protéger les cétacés et défendre leur environnement.

Outre ces aspects écologiques majeurs, cette thèse ouvre également des perspectives pour des modèles d'analyses acoustiques sous-marines utiles en lutte sous-marine et en stratégie de défense.

Glossaire

AMP : Aire Marine Protégée en méditerranée

BAOBAB : Balises et Acoustique pour l'Observation des Baleines A Bosse

BOMBYX : Bouée Multimodale pour l'étude de la Biodiversité et de OcéanophYsique

COR : Caractéristique Opérationnelle du Récepteur

DCL : international workshop on Detection, Classification and Localization of marine mammals

ICI : Inter Clic Interval

IIB : Interval Inter Band

IM : Information Mutuelle

IPI : Inter Pulse Interval

ISAR : Intra Spectral Attenuation Regression

ISARH : Intra Spectral Attenuation Regression Hierarchical

ISAT : Intra Spectral ATtenuation

KL : Kullback-Liebler

LARS : Least Angle Regression

LASSO : Least Absolute Shrinkage and Selection Operator

LER : Lervised Energy Regression

MFCC : Mel Frequency Cepstral Coeficient

MLP : Multi Layer Perceptron

PCA : Analyse des Composantes Principales

PELAGOS : Aire Marine Protégée en méditerranée

PNPC : Parc National de Port-Cros

RMSE : Root Mean Square Error

RSB : Rapport Signal à Bruit

TDOA : Time difference Of Arrival

TL : Transmission Loss

Bibliographie

- [Abbot 12] T. A. Abbot, V. E. Premus, P. A. Abbot et O. W. Mayer. *Receiver operating characteristic for a spectrogram correlator-based humpback whale detector-classifier*. J. Acoust. Soc. Am., vol. 132, no. 3, pages 1502–1510, 2012.
- [Abeille 13] R. Abeille, P. Giraudet et H. Glotin. *Acoustic observations for automatic size estimation of whales*. In Annals of Telecommunications and Radio Science Bulletin, 2013. IEEE OCOSS2013.
- [Abeille 14] R. Abeille, Y. Doh, P. Giraudet, H. Glotin, J.-M Prevot et C. Rabouy. *Estimation robuste par acoustique passive de l'Intervalle-Inter-Pulse des clics de Physeter macrocephalus : méthode et application sur le Parc national de Port-Cros*. Sci. Rep. Port-Cros natl. Park, vol. 28, pages 28–38, 2014.
- [Adam 05] O. Adam, M. Lopatka, C. Laplanche et J. F. Motsch. *Sperm whale signal analysis : comparison using the autoregressive model and the wavelet transform*. International Journal of Information Technology, vol. 2(1), no. 1–8, 2005.
- [Adam 08] O. Adam. *Segmentation of killer whale vocalisations using the Hilbert Huang Transform*. EURASIP Journal on Advances in Signal Processing, vol. 2008, page 10p, 2008.
- [Adam 13] O. Adam, D Cazau, N. Gandilhon, B. Fabre, J. T. Laitman et JS Reidenberg. *New acoustic model for humpback whale sound production*. Journal of Applied Acoustics, vol. 74(10), pages 1182–1190, 2013.
- [Au 93] W. Au. The sonar of dolphins. New York Springer-Verlag, 1993.
- [Au 05] W. W. L. Au, M. O. Lammers, A. Stimpert et M. Schotten. *The temporal characteristics of humpback whale songs*. J. Acoust. Soc. Am., vol. 118, no. 3, page 1940, 2005.
- [Au 06] W. W. L. Au, A. A. Pack, M. O. Lammers, L. M. Herman, M. H. Deakos et K. Andrews. *Acoustic properties of humpback whale songs*. J. Acoust. Soc. Am., vol. 120(2), pages 1103–1110, 2006.
- [Au 08] W. Au et M. Hastings. Principles of marine bioacoustics, modern acoustics and signal processing. Springer Science + business Media, LLC, New York, 2008.
- [Baker 84] C.S. Baker et L.M. Herman. *Aggressive behavior between humpback whale (Megaptera novaeangliae) wintering in Hawaiian waters*. Canadian Journal of Zoology, vol. 62, pages 1922–1937, 1984.

- [Baldacci 05] A. Baldacci, M. Carron et N. Portunato. *Infrared detection of marine mammals. NURC*. Rapport technique, NATO Undersea Research, 2005.
- [Barlow 09] J. Barlow, J. Calambokidis, E. A. Falcone, C. S. Baker, A. M. Burdin, P. J. Clapham, J. K. B. Ford, C. M. Gabriele, R. G. LeDuc, D. K. Mattila, T. J. Quinn, L. Rojas-Bracho, J. M. Straley, B. L. Taylor, R. J. Urban, P. Wade, D. W. Weller, B. H. Witteveen et M. Yamaguchi. *Humpback whale abundance in the North Pacific estimated by photographic capture-recapture with bias correction from simulation studies*. In 18th Biennial Conference on the Biology of Marine Mammals, 2009.
- [Baumgartner 11] M. F. Baumgartner et S. E. Mussoline. *A generalized baleen wale call detection and classification system*. J. Acoust. Soc. Am., vol. 129, pages 2889–2902, 2011.
- [Bellanger 84] M. Bellanger. *Digital processing of signals : Theory and practice*. New York, Wiley-Interscience, 1984.
- [Bénard 09] F. Bénard et H. Glotin. *Sperm whales records indexation using passive acoustics localization*. In CBMI 2009, pages 207–212, june 2009.
- [Bénard 10a] F. Bénard et H. Glotin. *Automatic indexing for content analysis of whale recordings and xml representation*. EURASIP Journal on advances in Signal Processing, vol. 10, 2010.
- [Bénard 10b] F. Bénard, H. Glotin et P. Giraudet. *Whale 3D monitoring using astrophysic NEMO ONDE two meters wide platform with state optimal filtering by Rao-Blackwell Monte Carlo data association*. Journal of Applied Acoustics, vol. 71, pages 994–999, november 2010.
- [Bénard 11a] F. Bénard, H. Glotin et P. Giraudet. *Highly defined whale group tracking by passive acoustic Stochastic Matched Filter*. In Advances in Sound Localization, ISBN 978-953-307-581-5. february 2011.
- [Bénard 11b] F. Bénard, H. Glotin et P. Giraudet. *Highly defined whale group tracking by passive acoustic stochastic matched filter*. In Advances in Sound Localization, chapitre 28, pages 527–544. InTech, Rijeka, Croatia, 2011.
- [Beyer 99] K. S. Beyer, J. Goldstein, R. Ramakrishnan et U. Shaft. *When Is "Nearest Neighbor" Meaningful?* In Proc. of the 7th International Conference on Database Theory (ICDT), pages 217–235. Springer-Verlag London, 1999.
- [Calambokidis 10] J. Calambokidis. *Final Report and Recommendations, Symposium on the results of the SPLASH humpback whale study 2009*. Rapport technique, Cascadia research, Canada, 2010.
- [Cazau 13] D. Cazau, O. Adam, J. T. Laitman et J. S. Reidenberg. *New Understanding the intentional acoustic behavior of humpback whale : a production-based approach*. J. Acoust. Soc. Am., vol. 143(3), pages 2268–2273, 2013.
- [Cholewiak 08] D. Cholewiak. *Evaluating the role of song in the humpback whale (Megaptera novaeangliae) breeding system with respect to intra-sexual interactions*. PhD thesis, Cornell University, Ithaca, NY., 2008.

-
- [Clapham 90] P. J. Clapham et D. K. Mattila. *Humpback whale songs as indicators of migration routes*. Marine Mammal Science, vol. 6, no. 2, pages 155–160, 1990.
- [Clapham 96] P. J. Clapham. *The social and reproductive biology of humpback whales : an ecological perspective*. Mammal Review, vol. 26, pages 27–49, 1996.
- [Clark 04] C. W. Clark et P. J. Clapham. *Acoustic monitoring on a humpback whale (Megaptera novaeangliae) feeding ground shows continual singing into late spring*. Proceedings - Royal Society of London. Biological sciences, vol. 271, no. 1543, pages 1051–1057, 2004.
- [Cuyler 92] L. C. Cuyler, R. Wiulsrod et N. A. Oritsland. *Thermal infrared radiation from free living whales*. Marine Mammal Science, vol. 8(2), pages 120–134, 1992.
- [Cybenko 89] G. Cybenko. *Approximations by superpositions of sigmoidal functions*. Mathematics of Control, Signals and Systems, vol. 2(4), pages 303–314, 1989.
- [Darling 01] J.D. Darling et M. Bérubé. *Interactions of singing humpback whales with other males*. Marine Mammal Science, vol. 17, no. 3, pages 570–584, 2001.
- [Davis 80] S. Davis et P. Nermelstein. *Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences*. IEEE Trans ASSP, vol. 28, pages 357–366, 1980.
- [Descudé 14] J. Descudé, J. M. Prévot, E. Bruno et H. Glotin. *AMR UltraSonoc 4x Wifi Recorder Implemented on BeagleBoard and Dodotronic UM 250kHz*. Rapport technique, LSIS/UTLN, 2014.
- [Doh 11a] Y. Doh et H. Glotin. *Modélisation of Physeter Macrocephalus Pulsed Sound Emission by Green Function*. In 5th International Workshop on Detection, Classification, Localization and Density Estimation of Marine Mammals using Passive Acoustics, USA, august 2011.
- [Doh 11b] Yann Doh. *A model of inter frequency attenuation (ifa) in order to estimate the distance source receiver*. Master’s thesis, Universités d’Aix Marseille/Centrale Marseille, 2011.
- [Doh 13a] Y. Doh, H. Glotin, J. Razik et S. Paris. *An Inter-frequency attenuation model for estimation of click distance : toothed whale example*, 2013. Soumis à JASA, LSIS research report RR.2013.01 en resoumission.
- [Doh 13b] Y. Doh, H. Glotin, J. Razik et S. Paris. *Mono-channel spectral attenuation modeled by hierarchical neural net estimates hydrophone-whale distance*. In Proc. of Neural Information Processing Scaled for Bioacoustics, from Neurons to Big Data. NIPS Int. Conf., South lake Thaeo, Nevada, USA, pages 88–96, december 2013.
- [Doh 13c] Y. Doh, J. Razik, S. Paris, O. Adam et H. Glotin. *Décomposition parcimonieuse des chants de cétacés pour leur suivi*. Traitement du Signal, vol. 30 (3-4-5), pages 219–242, 2013.

- [Duda 00] R. O. Duda, P. E. Hart et D. G. Stork. Pattern classification. Wiley-Interscience, New York, 2000.
- [Dunlop 07] R. A. Dunlop, M. J. Noad D. H. Cato et D. Stokes. *The social vocalization repertoire of east Australian migrating humpback whales (Megaptera novaeangliae)*. J. Acoust. Soc. Am., vol. 122, pages 2893–2905, 2007.
- [Finnereran 05] J. J. Finnereran, D. A. Carder, C. E. Schlundt et S. H. Ridgway. *Temporary threshold shift in bottlenose dolphins (Tursiops truncatus) exposed to mid-frequency tones*. J. Acoust. Soc. Am., vol. 18(4), page 2696–2705, 2005.
- [Flandrin 94] P. Flandrin, R. G. Baraniuk et O. Michel. *Time-Frequency Complexity and Information*. In IEEE International Conference on Acoustics, Speech, and Processing, volume 3, pages 329–332, april 1994.
- [Frankel 95] A.S. Frankel, C.W. Clark, L.M. Herman et C.M. Gabriele. *Spatial distribution, habitat utilization and social interactions oh humpback whales, Megaptera novaeangliae, of Hawaii, determined using acoustic and visual techniques*. Canadian Journal of Zoology, vol. 73, pages 1134–1146, 1995.
- [Frisch 92] M. Frisch et H. Messer. *The use of Wavelet Transform in the Detection of Unknown Transient Signal*. IEEE, Transactions on Information theory, vol. 38(2), pages 892–897, 1992.
- [Gillespie 04] Douglas Gillespie. *Detection and classification of right whale calls using an 'edge' detector operating on a smoothed spectrogram*. Canadian Acoustics, vol. 32, no. 2, pages 39–47, 2004.
- [Giraudet 06a] P. Giraudet et H. Glotin. *Echo-Robust and Real-Time 3D Tracking of Marine-Mammals using their Transient Calls recorded by Hydrophones Array*. In IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), volume IV, pages 1161–1164, may 2006.
- [Giraudet 06b] P. Giraudet et H. Glotin. *Real-time 3D tracking of whales by precise and echo-robust TDOAs of clicks extracted from 5 bottom-mounted hydrophones records of the AUTECH*. Applied Acoustics, vol. 67, pages 1106–1117, 2006.
- [Glotin 07] H. Glotin, O. Giraudet et F. Bénard. *Procédé de trajectographie temps réel de plusieurs cétacés par acoustique passive*, 2007. Déposé à l'Institut National de la Propriété Intellectuelle, INPI, publication USA, numéro 07/06162.
- [Glotin 08a] H. Glotin, F. Bénard et P. Giraudet. *Whales cocktail party : a real-time tracking of multiple whales*. International Journal Canadian Acoustics, vol. 36, no. 1, pages 139–145, march 2008.
- [Glotin 08b] H. Glotin, L. Gauthier, F. Pace, F. Benard et O. Adam. *New automatic classification for Humpback whale songs*. In Pavia university et ONR, editeurs, IEEE PASSIVE 08, page 93, 2008.
- [Glotin 09] H. Glotin, F. Bénard et P. Julie. *Whale axis angle variation estimation according to clic energy conservation*. In 4th International

-
- Workshop on Detection, Classification and Localization of Marine Mammals using Passive Acoustics, september 2009.
- [Glotin 11a] H. Glotin, Y. Doh, R. Abeille et A. Monnin. *Physeter distance estimation using sub-band Leroy transmission loss model*. In 5th International Workshop on Detection, Classification, Localization and Density Estimation of Marine Mammals using Passive Acoustics, USA, page 49, august 2011.
- [Glotin 11b] H. Glotin, J. Razik, P. Giraudet, S. Paris et F. Bénard. *Sparse coding for fast minke whale tracking with Hawaiian bottom mounted hydrophones*. In International Workshop on Detection, Classification, Localization and Density Estimation of Marine Mammals using Passive Acoustics August, 2011, Portland, USA, supported by ONR Dpt of the Navy and Acoustical Society of America (ASA), page 30, 22 august 2011.
- [Glotin 13a] H. Glotin. *Soundscape Semiotics - Localisation and Categorisation*. ISBN 980-953-307-687-9. march 2013.
- [Glotin 13b] H. Glotin, C. Clark, Y. LeCun, P. Dugan, X. Halkias et J. Sueur. *Proc. 1st workshop on Machine Learning for Bioacoustics - ICML4B*. USA, 2013. ICML int. Conf. <http://sabiod.univ-tln.fr>.
- [Glotin 13c] H. Glotin, P. Giraudet, J. Razik, S. Paris, X. Halkias, F. Chamroukhi, J. M. Prévot, J. Patris, F. Bénard, A. Monin, F. Lelandais, R. Abeille, Y. Doh, O. Dufour, C. Rabouy et A. Mishchenko. *Tracking multiple marine mammals by shortly or widely spaced hydrophones*. In Detection Classification localization of Marine Mammals using passive acoustics, volume ISBN 978-2-7466-6118-9, pages 71–92. 2013.
- [Glotin 13d] H. Glotin, Y. LeCun, T. Artières, S. Mallat, O. Tchernichovski et X. Halkias. *Proc. Neural Information Processing Scaled for Bioacoustics, from Neurons to Big Data*. USA, 2013. NIPS Int. Conf. <http://sabiod.org/nips4b>.
- [Glotin 13e] H. Glotin, J. Sueur, T. Artière an O. Adam et J. Razik. *Sparse coding for scaled bioacoustics : From Humpback whale songs evolution to forest soundscape analyses*. J. Acoust. Soc. Am., vol. 133(5), page 3311, 2013.
- [Gonzalez 11] R. P. Gonzalez, M. C. Valsero Blanco, F. Samaran et O. Adam. *Estimate the density of Antarctic blue whales and fin whales around the Crozet Islands using a passive acoustic sensor*. In 19th Biennial Conference on the Biology of Marine Mammals, Florida, USA, 2011.
- [Goold 96] J.C. Goold. *Signal processing techniques for acoustics measurement of sperm whale body lengths*. J. Acoust. Soc. Am., vol. 100, pages 3431–3441, 1996.
- [Gravier 10] G. Gravier. *SPro : a free speech signal processing toolkit*, 2010. Vers. 5.0. <https://gforge.inria.fr/projects/spro>.
- [Halkias 13] X. Halkias, H. Glotin et S. Paris. *Classification of Mysticete sounds using machine learning techniques*. J. Acoust. Soc. Am., vol. 134, no. 5, page 3496, november 2013.

- [Harris 06] J. G. Harris et M. D. Skowronski. *Automatic speech processing methods for bioacoustics signal analysis : a case study of cross-disciplinary acoustic research*. In ICASSP, may 2006.
- [Helweg 96] D. A. Helweg. *Geographic and temporal variation in songs of humpback whales*. J. Acoust. Soc. Am., vol. 100, no. 4, page 2609, 1996.
- [Herman 80] L. M. Herman et W. Tavolga. *The communication systems of cetaceans*. In L. M. Herman, ed. Cetacean behavior : Mechanisms and function. John Wiley and Sons, New York, NY. 1980.
- [hinich 82] M.-J. hinich. *Testing for gaussianity and linearity of stationary time series*. journal of time series analysis, vol. 3(3), pages 169–176, 1982.
- [hinich 90] M.-J. hinich. *Detecting a transient by bispectral analysis*. IEEE, Transactions on Acoustics, Speech and Signal Processing, vol. 38(7), pages 1277–1283, 1990.
- [Hornik 91] K. Hornik. *Approximation Capabilities of Multilayer Feedforward Networks*. vol. 4(2), pages 251–257, 1991.
- [Houser 99] D. S. Houser, D. A. Helweg et P. W. Moore. *Classification of dolphin echolocation clicks by energy and frequency distributions*. J. Acoust. Soc. Am., vol. 106(3), pages 1579–1585, 1999.
- [Hussain 01] A. Hussain, M. H. Sukairi, A. Mohamed et R. Mohamed. *Automatic detection of power quality disturbances and identification of transient signals*. In Int. Symposium on Signal Processing and its applications (ISSPA), Kuala Lumpur, Malaysia, pages 462–465, 2001.
- [Jin 01] J. Jin et J. Shi. *Automatic feature extraction of waveform signals for in-process diagnostic performance improvement*. J. Intelligent Manufacturing, vol. 12, pages 257–268, 2001.
- [Josso 10] N. Josso. *Characterizing the underwater environment using moving sources by opportunity*. PhD thesis, Université de Grenoble, 2010.
- [Kullback 51] S. Kullback et R.A. Leibler. *On Information and Sufficiency*. Annals of Mathematical Statistics, vol. 22, no. 1, pages 79–86, 1951.
- [Laplanche 05] C. Laplanche. *Studies by passive acoustics of the hunting behaviour of the sperm whale*. PhD thesis, Université Paris XII Val-de-Marne, 2005.
- [Le Cadre 97] J.P. Le Cadre et C. Jauffret. *Discrete-time observability and estimability for bearings-only target motion analysis*. IEEE Trans. on Aerospace and Electronic Systems, vol. 33(1), pages 178–201, 1997.
- [Learned 92] R. E. Learned, W. C. Karl et A. S. Willsky. *Wavelet packet based transient signal classification*. In *Time-Frequency and Time-Scale Analysis*. In Proc. of the IEEE-SP International Symposium, Canada, pages 109–112, 1992.
- [Leblanc 08] E. Leblanc, M. Bahoura et Y. Simard. *Comparison of automatic classification methods for beluga whales vocalizations*. In Acoustics' 2008. Paris. 29 June-4 July. J. Acoust. Soc. Am., volume 123, page 3772, 2008.
- [Lebot 14] O. Lebot. *Détection, localisation, caractérisation de transitoires acoustiques sous-marins*. PhD thesis, Université de Grenoble, 2014.

-
- [Lecun 85] Y Lecun. *A learning scheme for asymmetric threshold network*. COGNITIVA, vol. 85, pages 599–604, 1985.
- [Lelandais 08] F. Lelandais et H. Glotin. *Mallat's Matching Pursuit of Sperm Whale Clicks in Real-Time using Daubechies 15 wavelets*. In IEEE explorer, editeur, PASSIVE'08, page 5p, october 2008.
- [Leroy 65] C. Leroy. *Sound attenuation between 200 and 10000 cps measured along single paths*. Rapport technique 43, Saclant ASW Research Center, 1965.
- [Lopatka 05] M. Lopatka, O. Adam, C. Laplanche, J. Zarzycki et J. F. Motsch. *An Attractive Alternative For Sperm Whale Click Detection Using the Wavelet Transform in Comparison to the Fourier Spectrogram*. Aquatic Mammals, vol. 31(4), no. 463–467, 2005.
- [Lopatka 07] M. Lopatka. *Détection et reconnaissance des signaux stochastiques transitoires application à l'identification des mammifères marins*. PhD thesis, Université Paris 12 Val de Marne, 2007.
- [Lurton 98] X. Lurton. Acoustique sous-marine. IFREMER, 1998.
- [Madhsudhana 08] S. K. Madhsudhana, E. M. Oleson, M. S. Soldevilla, M. A. Roch et J. A. Hildebrand. *Frequency based algorithm for robust contour extraction of blue whale B and D calls*. In Proc. of the IEEE Oceans, Kobe, Japan, volume 3, page pp 8, 2008.
- [Madsen 02] P. Madsen, R. Payne, N. U. Kristiansen, I. Waltherberg M. and Kerr et B. Mohl. *Sperm whale sound production studied with ultrasound time/depth-recording tags*. The Journal of Experimental Biology, vol. 205, pages 1899–1906, 2002.
- [Mairal 09] J. Mairal, F. Bach, J. Ponce et G. Sapiro. *Online dictionary learning for sparse coding*. ICML, pages 689–696, 2009.
- [Mallawaarachchi 08] A. Mallawaarachchi, S. H. Onga, M. Chitre et E Taylor. *Spectrogram denoising and automated extraction of the fundamental frequency variation of dolphin whistles*. J. Acoust. Soc. Am., vol. 124, pages 1159–1970, 2008.
- [Marques 09] T. A. Marques, L. Thomas, J. Ward, N. DiMarzio et P. L. Tyack. *Estimating cetacean population density using fixed passive acoustic sensors : An example with Blainville's beaked whales*. J. Acoust. Soc. Am., vol. 125, page 1982, 2009.
- [Martin 11] S. W. Martin, T. Norris, E. M. nosal, D. K. Mellinger, R. P. morrissey et s. Jarvis. *Automatic localization of individual Hawaiian minke whales from boing vocalization*. J. Acoust. Soc. Am., vol. 129, page 2506, 2011.
- [Mazhar 08] S. Mazhar, T. Ura et R. Bahl. *An analysis of Humpback whale songs for individual classification*. J. Acoust. Soc. Am., vol. 123, no. 5, page 3774, 2008.
- [Mccafferty 07] D. J. Mccafferty. *The value of infrared thermography for research on mammals : previous applications and future directions*. MMammal Rev., vol. 37(2), pages 207–223, 2007.

- [Medrano 94] L. Medrano, M. Salinas, I. Salas, P. Ladron De Guevara, A. Agayo, J. Jacobsen et C. S. Baker. *Sex identification of humpback whales, Megaptera novaeangliae, on the wintering grounds of the pacific ocean*. Canadian Journal of Zoology, vol. 72, pages 1771–1774, 1994.
- [Mellinger 00] D. K. Mellinger et C. W. Clark. *Recognizing transient low-frequency whale sounds by spectrogram correlation*. J. Acoust. Soc. Am., vol. 107, pages 3518–3529, 2000.
- [Mellinger 04] D. K. Mellinger. *A comparison of methods for detecting right whale calls*. Canadian Acoustics, vol. 32, no. 2, pages 55–65, 2004.
- [Mercado III 98] E. Mercado III et A. Kuh. *Classification of humpback whale vocalizations using a self-organizing neural network*. In IEEE World Congress on Computational Intelligence, volume 2, pages 1584–1589, 1998.
- [Miksis-Olds 08] J.L. Miksis-Olds, J.R. Buck, M.J. Noad, D.H. Cato et M.D. Stokes. *Information theory analysis of Australian humpback whale song*. J. Acoust. Soc. Am., vol. 124, pages 2385–2393, 2008.
- [Minski 69] M. Minski et S. Papert. *Perceptrons*. In MIT Press. 1969.
- [Mobley 08] J.R. Mobley et R. Uyeyama. *Potential Impact of a Large Capacity Ferry on Marine Mammals of Hawaii*. In Submitted to Belt Collins Hawaii, page 38, 2008.
- [Mohl 00] B. Mohl, M. Wahlberg et P.T. Madsen. *Sperm whale clicks, directionality and source level revisited*. J. Acoust. Soc. Am., vol. 107, pages 638–648, 2000.
- [Mohl 03] B. Mohl, M. Wahlberg et P. T. frazer. *The monopulsed nature of sperm whale clicks*. J. Acoust. Soc. Am., vol. 114(2), pages 1143–1154, 2003.
- [Moretti 14] D. Moretti, L. Thomas, T. Marques, J. Harwood, A. Dilley, B. Neales, J. Shaffer, E. McCarthy, L. New, S. Jarvis et R. Morrissey. *A risk function for behavioral disruption of Blainville’s beaked whales (Mesoplodon densirostris) from mid-frequency active sonar*. Plos One, vol. 9(1), 2014.
- [Morrissey 06] R. Morrissey, J. Ward, N. DiMarzio, S. Jarvis et D. Moretti. *Passive acoustic detection and localization of sperm whales (physeter macrocephalus) in the tongue of the ocean*. Applied Acoustics, vol. 67, pages 1091–1105, 2006.
- [Mouy 67] X. Mouy et D. Hannay. *Tracking of Pacific walruses in the Chukchi Sea using a single hydrophone*. J. Acoust. Soc. Am., vol. 42(1), pages 1349–1358, 1967.
- [Murray 10] S. O. Murray, E. Mercado et H. L. Roitblat. *Characterizing the graded structure of false killer whale (pseudorca crassidens) vocalizations*. J. Acoust. Soc. Am., vol. 104, pages 1679–1688, 2010.
- [Noad 00a] Michael J Noad, Douglas H Cato, MM Bryden, Micheline-N Jenner et K Curt S Jenner. *Cultural revolution in whale songs*. Nature, vol. 408, no. 6812, pages 537–537, 2000.

-
- [Noad 00b] M.J. Noad, D.H. Cato, M.M. Bryden, M.N. Jenner et K.C.S. Jenner. *Cultural revolution in whale songs*. Nature, London, vol. 408, page 537, 2000.
- [Nosal 06] E.-M. Nosal et L. Frazer. *Track of a sperm whale from delays between direct and surface-reflected clicks*. Applied Acoustics, vol. 67, pages 1187–1201, 2006.
- [Nosal 07] E.-M. Nosal et L.N. Frazer. *Sperm whale three-dimensional track, swim orientation, beam pattern, and click levels observed on bottom-mounted hydrophones*. J. Acoust. Soc. Am., vol. 122(4), pages 1969–1978, 2007.
- [Oswald 07] J. N. Oswald, S. Rankin, J. Barlow et M. O. Lammers. *A tool for real-time acoustic species identification of delphinid whistles*. J. Acoust. Soc. Am., vol. 122, pages 587–595, 2007.
- [Pace 09] F. Pace, P. White et O. Adam. *Characterisation of sound subunits for humpback whale song analysis*. In 4th International Workshop on Detection and Localization of Marine Mammals using Passive Acoustics, page 56, 2009.
- [Pace 10] F. Pace, F. Benard, H. Glotin, O. Adam et P. White. *Subunit definition and analysis for humpback whale classification*. Journal of Applied Acoustics, vol. 71, november 2010.
- [Pace 11] F. Pace, P. R. White et O. Adam. *Classification of humpback whale (Megaptera novaeangliae) calls using Hidden Markov Models*. In 5th International Workshop on Detection, Classification, Localization, and Density Estimation of Marine Mammals using Passive Acoustics, page 29, 2011.
- [Paris] S. Paris, Y. Doh, H. Glotin, X. Halkias et J. Razik. *Physeter catodon localization by sparse coding*. In ICML4B in ICML 2013 conference, Atlanta, Georgia, USA, june.
- [Payne 71] R. S. Payne et S. McVay. *Songs of Humpback Whales*. Science, vol. 173, no. 3997, pages 585–597, 1971.
- [Picot 08] G. Picot, O. Adam, M. Bergounioux, H. Glotin et FX Mayer. *Automatic prosodic clustering of humpback whales song*. In New Trends for Environmental, PASSIVE 08, october 2008.
- [Rabiner 93] L. Rabiner et B. H. Huang. Fundamentals of speech recognition. Prentice Hall, Englewood Cliffs, NJ, 1993.
- [Ravier 92] P. Ravier. *Détection de transitoires par ondelettes adaptées - critères d'adaptation fondés sur les statistiques d'ordre supérieur*. PhD thesis, Institut National Polytechnique de Grenoble, 1992.
- [Razik 11] J. Razik, H. Glotin, S. Paris et O. Adam. *Humpback whale song sparse coding and information theory analysis*. In International Workshop on Detection, Classification, Localization Density Estimation of Marine Mammals using Passive Acoustics August, 2011, Portland, USA, supported by ONR Dpt of the Navy and Acoustical Society of America (ASA), page 41, 22 august 2011.

- [Razik 13] J. Razik, H. Glotin, S. Paris, O. Adam et Y. Doh. *Sparse coding of Humpback whale songs determines their evolution*, 2013. Soumis à JASA, LSIS research report RR.2013.02 en resoumission.
- [Reidenberg 07] J. S. Reidenberg et J. T. Laitman. *Discovery of a low frequency sound source in Mysticeti (baleen whales) : Anatomical Establishment of a Vocal Fold Homolog*. *Anat. Rec.*, vol. 290, pages 745–759, 2007.
- [Rickwood 08] P. Rickwood et A. Taylor. *Methods for automatically analyzing humpback song units*. *J. Acoust. Soc. Am.*, vol. 123, no. 3, pages 1763–1772, 2008.
- [Ridgway 97] S. Ridgway. *Who are the whales?* *Bioacoustics*, vol. 8(1-2), pages 3–20, 1997.
- [Rosenblatt 60] F. Rosenblatt. *Perceptron simulation experiments*. In *Proceeding of the IRE*, 3, 48, 1960.
- [Samaran 10] F. Samaran, O. Adam et C. Guinet. *Detection range modeling of blue whale calls in Southwestern Indian Ocean*. *Applied Acoustics*, vol. 11(71), pages 1099–1106, 2010.
- [Sangfelt 93] E. Sangfelt et L. Persson. *Experimental performance of some higher-order cumulant detectors for hydroacoustics transients*. In *Proc. of the IEEE Signal Processing workshop on Higher-Order Statistics*, south Lake Tahoe, CA, USA, pages 182–186, 1993.
- [Sanjeev Arulampalam 02] M. Sanjeev Arulampalam, S. Maskell et N. Gordon. *A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking*. *IEEE, Transactions on Signal Processing*, vol. 50, pages 174–188, 2002.
- [Shannon 48] C.E. Shannon. *A mathematical theory of communication*. *Bell System Technical Journal*, vol. 27, pages 379–423,623–656, 1948.
- [Shapiro 09] A. D. Shapiro et C. Wang. *A versatile pitch tracking algorithm : from human speech to killer whale vocalizations*. *J. Acoust. Soc. Am.*, vol. 126, pages 451–459, 2009.
- [Simard 10] Y. Simard, N. Roy, S. Giard, C. Gervaise, M. Conversano et N. Ménard. *Estimating whale density from their whistling activity : example with St. Lawrence beluga*. *Applied Acoustics*, vol. 71, pages 1081–1086, 2010.
- [Simmonds 07] M. Simmonds et S. Isaac. *The impact of climate change on marine mammals : early signs of significant problems*. *Oryx*, vol. 41(01), pages 19–26, 2007.
- [Simmonds 09] M. Simmonds et W. J. Elliott. *Climate change and cetaceans : concerns and recent developments*. *Mar. Biol. Assoc. UK.*, vol. 89, pages 203–210, 2009.
- [Stevick 03] P. T. Stevick, J. Allen, M. Bérubé, P. J. Clapham, S. K. Katona, F. Larsen, Lien, Jon, D. K. Mattila, P. J. Palsboll, J. Robbins, J. Sigurjonsson, T. D. Smith, N. Oien et P. S. Hammond. *Segregation of migration by feeding ground origin in North Atlantic humpback whales (Megaptera novaeangliae)*. *Journal of Zoology*, vol. London 259, pages 231–237, 2003.

-
- [Suzuki 06] P. Suzuki, J. R. Buck et P. L. Tyack. *Information entropy of humpback whale songs*. J. Acoust. Soc. Am., vol. 119, no. 3, pages 1849–1866, 2006.
- [Thorp 67] W.H. Thorp. *Analytic description of the low frequency attenuation coefficient*. J. Acoust. Soc. Am., vol. 42(1), page 270, 1967.
- [Tyack 81] P. Tyack. *Interactions between singing Hawaiian humpback whales and conspecifics nearby*. Behavioral Ecology and Sociobiology, vol. 8, no. 2, pages 105–116, 1981.
- [Urazghildiiev 06] I. R. Urazghildiiev et C. W. Clark. *Acoustic detection of North Atlantic right whale contact calls using the generalized likelihood ratio test*. J. Acoust. Soc. Am., vol. 120, pages 1956–1963, 2006.
- [Viala 07] C. Viala. *Real time inversion in geoacoustic of wideband signals in deep water*. PhD thesis, Université de Toulon et du Var, 2007.
- [Welcome 09] A. Welcome. *Night vision and thermal imaging : Hawaii ferry makes good use of night navigation technology*. In Passenger Ship Technology (December/January 2008/2009), page 3p, 2009.
- [Winn 78] H.E. Winn et L.K. Winn. *The song of the humpback whale Megaptera novaeangliae in the West Indies*. Mar. Biol., vol. 47, pages 97–114, 1978.
- [Zimmer 04] W.M. Zimmer, P.L. Tyack, M.P Johnson et P. T. madsen. *Three-dimensional beam pattern of regular sperm Whale clicks confirms bent-horn hypothesis*. J. Acoust. Soc. Am., vol. 117(3), pages 1473–1485, 2004.
- [Zimmer 05] W. M. X. Zimmer, P. T. Madsen, V. Teloni, M. P. Johnson et P. L. Tyack. *Off-axis effects on the multipulse structure of sperm whale usual clicks with implication for sound production*. J. Acoust. Soc. Am., vol. 118(5), pages 3337–3345, 2005.
- [Zimmer 11] W.M.X. Zimmer. *Passive acoustic monitoring of cetaceans*. Passive Acoustic Monitoring of Cetaceans. Cambridge University Press, 2011.

Publications personnelles

- [Abeille 13] R. Abeille, F. Chamroukhi, Y. Doh, O. Dufour, P. Giraudet, X. Halkias, H. Glotin, J. M. Prévot, C. Rabouy et J. Razik. *Détection et classification sur transect audiovisuels de population de cétacés du Nord Pelagos- Iles d'Or*, 2013.
- [Abeille 14] R. Abeille, Y. Doh, P. Giraudet, H. Glotin, J.-M. Prevot et C. Rabouy. *Estimation robuste par acoustique passive de l'Intervalle-Inter-Pulse des clics de *Physeter macrocephalus* : méthode et application sur le Parc national de Port-Cros*. Sci. Rep. Port-Cros natl. Park, vol. 28, pages 28–38, 2014.
- [Adam 11] O. Adam et Y. Doh. *Mesure du bruit ambiant sous-marin dans le Grand Cul-de-Sac Marin (Guadeloupe)*. Rapport technique, Paris11/UTLN, 2011.
- [Doh 11a] Y. Doh, R. Abeille et H. Glotin. *Can we estimate the Radial Velocity of a *Physeter Macrocephalus* from Doppler Effect?* In 5th International Workshop on Detection, Classification, Localization and Density Estimation of Marine Mammals using Passive Acoustics, USA, 2011.
- [Doh 11b] Y. Doh et H. Glotin. *Modelisation of *Physeter Macrocephalus* Pulsed Sound Emission by Green Function*. In 5th International Workshop on Detection, Classification, Localization and Density Estimation of Marine Mammals using Passive Acoustics, USA, 2011.
- [Doh 13a] Y. Doh, H. Glotin, J. Razik et S. Paris. *An Inter-frequency attenuation model for estimation of click distance : toothed whale example*, 2013. Soumis à JASA, LSIS research report RR.2013.01 en resoumission.
- [Doh 13b] Y. Doh, H. Glotin, J. Razik et S. Paris. *An Inter-Frequency Attenuation Model for Estimation of Click Distance with a Single Hydrophone : Sperm Whale Example*. In International workshop on Cetacean echolocation and outer space neutrinos : ethology and physics for an interdisciplinary approach to underwater bioacoustics and astrophysical particles detection, Erice, Sicilia, 2013.
- [Doh 13c] Y. Doh, H. Glotin, J. Razik et S. Paris. *Mono-channel spectral attenuation modeled by hierarchical neural net estimates hydrophone- whale distance*. In Proc. of Neural Information Processing Scaled for Bioacoustics, from Neurons to Big Data. NIPS Int. Conf., South lake Thaeo, Nevada, USA, pages 88–96, 2013.
- [Doh 13d] Y. Doh, J. Razik, S. Paris, O. Adam et H. Glotin. *Décomposition parcimonieuse des chants de cétacés pour leur suivi*. Traitement du Signal, vol. 30 (3-4-5), pages 219–242, 2013.
- [Glotin 11] H. Glotin, Y. Doh, R. Abeille et A. Monnin. **Physeter* distance estimation using sub-band Leroy transmission loss model*. In 5th International Workshop on Detection, Classification, Localization and Density Estimation of Marine Mammals using Passive Acoustics, USA, page 49, August 2011.

- [Glotin 12] H. Glotin, Y. Doh et R. Abeille. *Suivi et analyse du comportement des grands cétacés par acoustique passive*. In SERENADE 2012 pour la session caractérisation du vivant marin, Grenoble, France, 2012.
- [Glotin 13a] H. Glotin, Y. Doh et O. Adam. *Avancées en bioacoustique des cétacés : de la détection à la trajectographie 3D. Illustrations sur des sons d'odontocètes (cachalot projet BOUSSOLE) et de mysticètes (baleine à bosse, projet BAOBAB)*. In Biologging, bio-téléométrie et ingénierie écologique au sein des ZA : mise en commun des informations et des ressources, Strasbourg, France, 2013.
- [Glotin 13b] H. Glotin, P. Giraudet, J. Razik, S. Paris, X. Halkias, F. Chamroukhi, J. M. Prévot, J. Patris, F. Bénard, A. Monin, F. Lelandais, R. Abeille, Y. Doh, O. Dufour, C. Rabouy et A. Mishchenko. *Tracking multiple marine mammals by shortly or widely spaced hydrophones*. In Detection Classification localization of Marine Mammals using passive acoustics, volume ISBN 978-2-7466-6118-9, pages 71–92. 2013.
- [Paris 13] S. Paris, Y. Doh, H. Glotin, X. Halkias et J. Razik. *Physeter catodon localization by sparse coding*. In ICML4B in ICML 2013 conference, Atlanta, Georgia, USA, 2013.
- [Razik 13] J. Razik, H. Glotin, S. Paris, O. Adam et Y. Doh. *Sparse coding of Humpback whale songs determines their evolution*, 2013. Soumis à JASA, LSIS research report RR.2013.02 en resoumission.
- [Samaran 13] F. Samaran, N. Gandilhon, Y. Doh, F. Pace, D. Cazau, C. Laplanche M. Lopatka, H. Glotin, P. White, J. Zarzycki, J. F. Motsch et O. Adam. *Inside the sounds emitted by some cetacean species*. In Detection Classification localization of Marine Mammals using passive acoustics, volume ISBN 978-2-7466-6118-9, pages 267–287. 2013.

Résumé

Les ondes acoustiques subissent peu de dispersion dans le milieu marin, comparé au milieu aérien. Certaines espèces de cétacés communiquent ainsi à grande distance, d'autres utilisent leurs émissions sonores pour s'orienter. La bioacoustique consiste à étudier ces espèces à partir de l'analyse de leurs sons, c'est-à-dire à les détecter, classer, localiser. Cela peut se faire via un réseau d'hydrophones au déploiement fastidieux. Afin de contribuer au passage à l'échelle de la bioacoustique, cette thèse propose des modèles originaux mono-hydrophone pour l'analyse de ces signaux stationnaires ou transitoires.

Premièrement, nous dérivons un nouveau modèle d'estimation de la distance entre une source impulsive (ex. biosonar) et un hydrophone. Notre modèle théorique, l'Intra Spectral ATténuation (ISAT), dérive des lois acoustiques de déformation spectrale du signal transitoire induite par l'atténuation durant sa propagation. Ce modèle relie les rapports énergétiques des bandes de fréquences pondérés par un modèle de perte par atténuation fréquentielle (Thorp ou Leroy) à la distance de propagation. Nous approximations aussi ISAT par un modèle neuromimétique. Ces deux modèles sont validés sur le sonar du cachalot (*Physeter macrocephalus*) enregistré avec notre bouée acoustique autonome BOMBYX et notre système d'acquisition DECAV en collaboration avec le Parc National de Port-Cros et le sanctuaire Pelagos pour la protection des mammifères marins en Méditerranée. Les mesures d'erreur (RMSE) d'environ 500 mètres sur nos références du centre d'essai OTAN aux Bahamas présentent un intérêt opérationnel.

Deuxièmement, nous proposons une analyse original de l'évolution des voisements de cétacé par codage parcimonieux. Notre encodage des cepstres par apprentissage non supervisé d'un dictionnaire met en évidence l'évolution temporelle des bigrammes des chants que les baleines à bosse mâles émettent durant la période de reproduction. Nous validons ce modèle sur nos enregistrements du canal de Sainte-Marie à Madagascar entre 2008 et 2014, via notre réseau d'hydrophones BAOBAB qui constitue une première dans l'Océan Indien.

Nos modèles s'inscrivent dans le projet Scaled Bioacoustics (SABIOD, MI CNRS) et ouvrent de nouvelles perspectives pour les passages à l'échelle temporelle et spatiale de la bioacoustique.

Mots-clés: estimation de distance source-récepteur ; codage parcimonieux ; bioacoustique ; acoustique sous-marine ; atténuation fréquentielle ; signaux stationnaires ; signaux transitoires ; mono-hydrophone ; suivi des cétacés ; passage à l'échelle ; baleine à bosse ; cachalots ; BAOBAB ; BOMBYX ; DECAV ; Pelagos ; NPC ; Madagascar.

Abstract

Acoustic waves show low dispersion due to the underwater propagation, compared to the propagation in the air. Some species of cetaceans communicate at long distance, others use their sound production for orientation. The goal of the scientific area called bioacoustics is to study animal species based on the analysis of their emitted sound. Their sounds can be used to detect, to classify and to locate the cetaceans. Recordings can be done with a passive acoustic array of multiple hydrophones, but this method is expensive and difficult to deploy. Thus, in order to scale this approach, we propose in this Phd thesis several original single hydrophone models to analyze these stationary or transient signals.

Firstly, we provide a new theoretical model to estimate the distance between the impulsive source (ex. biosonar of the cetacean) and the hydrophone. Our model, the Intra Spectral ATtenuation (ISAT), is based on the spectral signal alteration due to the underwater acoustic propagation, especially the differences in different frequency bands. We also approximated ISAT by an artificial neural network. Both models are validated on clicks emitted by sperm whales (*Physeter macrocephalus*) recorded by our sonobuoy BOMBYX and our data-acquisition system DECAV developed in collaboration with the National Park of Port-Cros (France) and the Pelagos sanctuary for the protection of marine mammals in the Mediterranean sea (France). The error (RMSE) measures on the recordings of the NATO test center in the Bahamas are about 500 meters, promising further real applications.

Secondly, we worked on the variations of the cetacean vocalizations using the sparse coding method. The encoding of the cepstrums by unsupervised learning of a dictionary shows bigrammic time changes of the songs of humpback whales (*Megaptera novaeangliae*). We validate this model on signals recorded in the Ste Marie Channel (Madagascar) between 2008 and 2014, through our network of hydrophones BAOBAB which is the first passive acoustic array deployed in the Indian Ocean.

Our models are part of the Scaled Bioacoustics project (SABIOD, MI CNRS) and open perspectives for temporal and spatial scaling of bioacoustics.

Keywords: source-receiver range estimation, sparse coding, bioacoustics, underwater acoustics, frequency attenuation, stationary signals, transient signals, single hydrophone, cetacean survey, large scaling, humpback whale, sperm whale, BAOBAB, BOMBYX, DECAV, Pelagos, PNPC, Madagascar