Study of a biosonar based on the modeling of a complete chain of emission-propagation-reception with validation on sperm whales

Maxence Ferrari

To cite this version:

Maxence Ferrari. Study of a biosonar based on the modeling of a complete chain of emission-propagation-reception with validation on sperm whales. General Mathematics [math.GM]. Université de Picardie Jules Verne, 2020. English. NNT : 2020AMIE0006 . tel-03626254

HAL Id: tel-03626254
https://theses.hal.science/tel-03626254

Submitted on 31 Mar 2022

Thèse de Doctorat

Mention : Mathématiques
Spécialités : Mathématiques, Informatique

présentée à *l'Ecole Doctorale en Sciences Technologie et Santé (ED 585)*

de l'Université Picardie Jules Verne
par
Maxence Ferrari

pour l'obtention du grade de Docteur de l'Université Picardie Jules Verne

Study of a Biosonar Based on the Modeling of a Complete Chain of Emission-Propagation-Reception with Validation on Sperm Whales

Soutenue le 30 septembre 2020, après avis des rapporteurs, devant le jury d'examen :

M. Mark ASCH	**LAMFA UMR CNRS 7352** **Université de Picardie - Jules Verne**	**Directeur de thèse**
M. Hervé GLOTIN	**LIS UMR CNRS 7020** **Université de Toulon**	**Co-directeur de thèse**
M. Marxer RICARD	**LIS UMR CNRS 7020** **Université de Toulon**	**Co-encadrant de thèse**
M. Ronan FABLET	**MT Atlantique/Lab-STICC** **UMR CNRS**	**Rapporteur**
Mme. Marie ROCH	**San Diego State University**	**Rapporteur**
M. Gianni PAVAN	**Centro Interdipartimentale CIBRA** **Università degli Studi di Pavia**	**Président du jury**
M. Yannis STYLIANOU	**University of Crete**	**Examinateur**
Mme. Amélie BARAZZUTTI	**Innovation Défense Lab**	**Invitée**
Mme. Odile GERARD	**Direction Générale de l'Armement**	**Invitée**
M. François SARANO	**Longitude 181**	**Invité**

Acknowledgements

I would like to first thanks Agence de l'innovation de défense and Région Hauts-de-France for funding. I would also like to thank my supervisors, Hervé Glotin, Mark Asch, and Ricard Marxer, for helping me throughout the three years of writing this manuscript. Amélie Barazzutti and Odile Gerard also deserved my gratitude for monitoring the progression of my thesis and the suggestion they have made.

I would also like to thanks my friends and colleagues, Guillaume Sanchez, Marion Poupard, Paul Best, for helping me stay motivated and bring an alternative view on my research. All the discussions we had stimulated new ideas. They spring into my mind and lead to new papers.

I would also like to acknowledge the distinguished scientists that are Marie Roch, Ronan Fablet, Gianni Pavan, and Yannis Stylianou, for accepting to be the members of my jury.

Thanks to all the souls of the bâtiment X, and the Ph.D. students of the LAMFA for the time we shared. I thank the glorious sun for bringing me vitamin D to stay motivated and productive.

While a thesis does not mark the end of learning, it is still a significant milestone in the accumulation of knowledge gained through life. Thus I would like to thanks the teachers that I had. I would most like to thank the teachers from the Legendre primary school and the teachers of Chaptal, where I have learned the most.

A special thanks go to François Sarano for the data he acquired and share, and for his patience with the ongoing evolution of the antenna and the last-minute fixes. Another one goes to the team of Seaproven, for without them, there would not be the Sphyrna drones. They also have my gratitude for helping and coping with us during the two long missions.

I would like to thank Marina Oger for the work she has done on labeling the IPIs, and Nicolas Thellier for his work on the Sphyrna. Finally, I would also like to thanks Vincente, Julien, Vincent, and Jean-Marc.

Résumé en français

Contexte et motivation

Les cachalots (*Physeter macrocephalus, Pm*) ont le bio-sonar le plus puissant du règne animal (230 dB re: 1µPa rms). Les clics produits par ce sonar ne sont pas seulement utilisés pour l'écholocalisation lors des plongées, mais aussi dans les interactions sociales. Lors de leurs plongées, les cachalots émettent des trains de clics, un peu comme ceux des chauves-souris, alors que pour la socialisation, ils émettront de petites séries rythmiques de clics.

Il a été largement admis que le cachalot crée une impulsion initiale à l'avant de sa tête, dans le "museau de singe", qui rebondira alors et en avant dans sa tête, passant à travers plusieurs sacs d'huile, avant de sortir. Cependant, les détails d'un tel mécanisme et les paramètres sur lesquels le cachalot peut agir restent inconnus.

L'étude des cachalots est un vaste sujet, qui couvre des champs disciplinaires multiples et variés, de l'éthologie à la conservation en passant par la bioacoustique. Étant donné que ce manuscrit ne peut intégrer un sujet aussi vaste, il traitera de quelques questions liées à un élément central du cachalot, qui est son sonar. Comprendre le mécanisme qui régit le sonar de cachalot aidera à étudier ces diverses domaines, car il s'agit d'un élément clé dans la vie des cachalots.

Les cachalots sont des espèces sauvages qui, contrairement à d'autres espèces qui vivent également en captivité, ne peuvent pas être équipées d'instruments lourds ou entraînées pour tester leurs capacités de sonar. Sans possibilité d'expérimentation contrôlée sur de vrais cachalots, une option est d'expérimenter sur un cachalot virtuel, conduisant au premier objectif de cette thèse, la création d'un simulateur biosonar de cachalot. Cependant, un simulateur à lui seul est inutile, car on ne peut pas savoir à quel point il est éloigné de la réalité.

Ainsi, le deuxième objectif de cette thèse, traite de la façon d'améliorer les simulations à l'aide de mesures. Ce deuxième objectif est en outre divisé en deux parties, la méthode de couplage et l'acquisition desdites mesures.

Le dernier objectif de cette thèse est l'utilisation de l'apprentissage automatique, pour améliorer la vitesse d'extraction de données utiles à partir d'enregistrements pour apprendre les fonctionnalités qui caractérisent leurs clics acoustiques.

Le manuscrit est composé de quatre parties successives. Dans la partie I, après avoir présenté les bases des techniques de traitement du signal qui seront utilisées tout au long de notre étude, trois bases de données différentes sont décrites: deux en champ lointain et une en champ proche, qui ont été acquises et / ou traitées pendant la thèse. Ces bases de données d'enregistrements de cachalots sont à la base de la deuxième (Machine Learning) et de la dernière (Coupling) parties.

La partie II concerne l'apprentissage automatique appliqué à la bioacoustique. Il utilise ces bases de données pour entraîner les réseaux de neurones afin de découvrir ce qu'est un clic, qui après l'entraînement peut être utilisé pour l'analyse de nouvelles bases de données.

La troisième partie du manuscrit présente la théorie de la propagation des ondes et les méthodes numériques afin de créer une simulation de la propagation des ondes à l'intérieur de la tête du cachalot.

Enfin, la partie IV concerne le couplage, utilisant la théorie des problèmes inverses pour améliorer les estimations cruciales des paramètres matériels du modèle.

Ces quatre parties répondent aux objectifs de la thèse exprimés précédemment et ont permis d'obtenir les résultats suivants:

- Modification de l'algorithme BIVIP pour estimer l'IPI.
- Proposition d'un processus de détection de clics.
- Proposition de la méthode d'estimation Geometric SRP TDOA.
- Application de la distribution ellipsoïde TDOA à la fois pour l'estimation de la forme de l'antenne et le filtrage.
- Développement d'une méthode de synchronisation audio entre un signal compressé avec perte et un signal dilaté.
- Estimation des trajectoires des cachalots sur une antenne à deux hydrophones.
- Proposition de DBSCAN comme méthode d'estimation des pistes à des fins de filtrage.
- Estimation des trajectoires de plongée des cachalots.
- Application d'un réseau de segmentation d'images à des vidéos de cachalots.
- Localisation de source 3D et traçage vidéo.
- Création d'une base de données de clics avec classification par individus.
- Création du jeu de données de clic et défi odontocetes, DOCC10.
- Formulation de l'architecture du modèle UpDim.
- Premier modèle de classification biosonar efficace de bout en bout.

- Développement d'un auto-encodeur pour débruiter le signal et estimer le TDOA.
- Création d'un modèle 3D d'une géométrie d'organe nasal de cachalot adulte.
- Application d'un simulateur de propagation d'onde FDTD sur le modèle.
- Formulation d'une méthode de couplage mesure-simulation pour la résolution de problèmes inverses en propagation d'onde.
- Dérivation d'une formulation adjointe complète pour un modèle de propagation d'onde pleine avec des conditions aux limites absorbantes.

Acoustique de la tête du cachalot

Avant de répondre aux objectifs de la thèse, il faut d'abord avoir connaissance des théories permettant d'expliquer la formation du clic multi-pulsée du cachalot. La répétition d'une impulsion dans un clic de cachalot est expliquée par le modèle de la corne courbée, ou le modèle de la corne courbée avec fuites, qui ajoute le chemin acoustique pour les fausses impulsions.

Une fois que l'impulsion a été émise dans le *museau de singe*, l'impulsion se propage vers l'arrière de la tête à travers le spermaceti. Une partie de l'énergie libérée lors de la création de l'impulsion fuit directement, qui sera enregistré en premier et est noté P_0. L'impulsion ayant traversé le spermaceti rencontre alors le sac frontal, qui agit comme un miroir , redirigeant le pouls vers l'avant. À partir de ce point, l'impulsion peut suivre trois chemins. Une partie de l'énergie s'échappera ainsi à travers l'eau. Une autre partie passera par le junk pour atteindre l'eau. Le reste de l'énergie retournera vers les lèvres du singe à travers le spermaceti, qui après avoir rebondi sur le sac d'air distal, redémarrera le cycle. L'énergie va donc rebondir entre le sac frontal et distal, tandis qu'une partie de celle-ci partira dans l'eau et dans le junk. Les impulsions quittant la tête du cachalot à travers le junk, et qui contiennent la plupart de l'énergie et sont nommées $P_1, P_2, P_3, \ldots$. Les impulsions qui fuient dans l'eau sont appelées demi-impulsions, car elles ne traversent que la moitié des organes, et sont notées $P_{1/2}, P_{3/2}, P_{5/2}, \ldots$. Elles sont également nommées fausses impulsions, donc parfois nommées $P_{f1}, P_{f2}, P_{f3}, \ldots$.

L'intervalle entre les impulsions (IPI) est l'intervalle de temps entre la réception de deux impulsions P_n et P_{n+1} et est noté P_nP_{n+1}. Puisque $\forall n \in \mathbb{N}^*$, P_{n+1} parcourt le même chemin acoustique que P_n plus un aller-retour dans l'organe spermaceti, l'IPI est constant entre chaque paire d'impulsions (P_n, P_{n+1}) (voir figure 2.4). La valeur de l'IPI P_0P_1 est proche de celle de P_1P_2. Au lieu d'une différence liée à un aller-retour dans le spermaceti, P_0P_1 est dû à un aller dans le spermaceti et à un retour dans le junk, et à la différence dûe aux zones de sortie de la tête. La comparaison entre la valeur de P_0P_1 et celle de P_1P_2 conduit en pratique à une différence de 10%. Des exemples d'enregistrements que nous avons faits de clics de cachalots sont illustrés dans les figures 2.5, 2.6 et 2.7. Le clic enregistré à 96 kHz sur la première figure est un exemple d'un IPI de 4,042 ms (probablement une femelle car il se traduit par une longueur de corps comprise entre 9 et 11 mètres) où θ est compris entre $-90°$ et $90°$. La différence des chemins acoustiques peut être vue dans ce clic, car P_0P_1 n'est que de 3,60 ms. La deuxième paire de la figure montre un clic enregistré à 600 kHz avec un IPI 3,27 ms. Ce clic a été enregistré lors de la session d'enregistrement de Maurice 2018 (voir section 6.1). Puisqu'il a été enregistré près du cachalot et que le clic faisait partie d'un coda 2 + 5, ses impulsions ultérieures peuvent être facilement vues.

Techniques du traitement de signal

Afin d'étudier l'acoustique du cachalot, il faut d'abord enregistrer leurs signaux, puis les traiter pour au final les analyser. Ici sont ainsi présentées les différentes techniques de traitement du signal qui seront utilisées avec les bases de donnée présentées par la suite. Ce processus d'analyse est souvent composé de deux étapes principales: la détection des clics et la localisation de la source.

Détection de clics

Les signaux d'acoustiques passives proviennent souvent d'enregistrement long terme et ne contiennent que très peu de signal biologique de l'espèce observée, par rapport à la quantité des autres signaux. L'expertise humaine étant trop coûteuse par rapport à la taille des enregistrements, il faut recourir à un détecteur automatique de clics pour extraire la position de chaque clic de cachalot présent dans le signal. Une méthode simple et classique pour extraire les clics est d'utiliser le filtre Teager-Kaiser (TK), aussi appelé Teager-Kaiser Energy Operator. Ce filtre se comporte comme un détecteur de pics de Dirac. Après avoir utilisé un filtre passe-bande afin de ne garder que la bande de fréquence contenant l'énergie des clics de cachalot, on applique le filtre TK. Puis l'on extrait les maximums locaux par fenêtre glissante de 20 ms. Avec une estimation du niveau de bruit on peut estimer un niveau au-dessus duquel tous les maximums appartiennent à des clics.

Cette méthode de détection n'est pas parfaite car toutes autres sources produisant des clics (bateaux, électricité, autres espèces, etc.) avec un peu d'énergie dans la même bande de fréquence seront aussi détectées. Ces faux clics seront filtrés plus tard dans l'analyse grâce à d'autres informations. Après ce second filtrage, la base de données de clics obtenue peut servir à entraîner un réseau de neurones servant de meilleur détecteur de clics.

Estimation des TDOA

Les TDOA ou différences de temps à l'arrivée en français, sont utiles pour estimer la position de la position d'un cachalot qui a émit le clic. Cela vient du fait que les hydrophones qui mesurent le signal sont placés à différents endroits dans l'espace. Un clic arrivera donc à différents moments pour chacun de ces hydrophones.

Une méthode pour estimer les TDOA est d'utiliser la corrélation. La corrélation est le produit scalaire de deux signaux, pour différents décalages temporels. Si les deux signaux sont identiques à l'exception d'un décalage temporel, le maximum de leur corrélation sera pour le décalage opposé. Cependant, le bruit reçu par chaque hydrophone n'est pas exactement le même. Par conséquence estimer un TDOA avec une corrélation peut donner des résultats aléatoires. Afin de lever ce problème, il est possible de combiner l'estimation de plusieurs TDOA. pour N hydrophones, il y a $\binom{N}{2}$ corrélations et TDOA différent, mais seulement $N-1$ TDOA sont indépendants. Il est donc possible d'estimer simultanément les TDOA en cherchant les $N-1$ TDOA indépendants qui maximisent les $\binom{N}{2}$ corrélations, ou plutôt qui maximisent une valeur représentant cette maximisation. Soit l'on maximise la somme des valeurs de chaque corrélation, ce qui est fait dans la méthode SRP, soit l'on maximise le produit des valeurs de chaque corrélation, ce qui est fait dans la méthode geometric-SRP. L'estimation des TDOA réalisée dans la suite de cette thèse utilise la seconde méthode. L'avantage de ces méthodes par rapport à une méthode d'estimation de TDOA qui prendrait en compte la géométrie de l'antenne, c'est que l'on peut filtrer les clics dont les TDOA ne correspondent pas a la géométrie de l'antenne.

Estimation de la taille de la géométrie de l'antenne

Même si 4 hydrophones suffisent théoriquement pour récupérer la distance, le calcul n'est pas robuste et toute petite erreur de mesure peut conduire à de grandes variations dans l'estimation de la distance. Afin améliorer ce calcul, le TDOA peut être utilisé pour estimer la géométrie de l'antenne. Pour une antenne à 4 hydrophones, et des clics provenant d'une source lointaine (au moins 5 fois la taille de l'antenne), les 3 TDOA indépendants de ces clics seront tous dans une ellipsoïde, dont les paramètres dépendent des TDOA indépendants choisis et de la position des hydrophones. On peut donc, soit estimer la taille de l'antenne en inversant les équations donnant les paramètres de l'ellipse en fonction des positions des hydrophones, soit filtrer les TDOA n'étant pas sur cette ellipse à l'aide des positions des hydrophones.

BOMBYX

BOMBYX est une bouée installée en mer Méditerranée, près de l'île de Porquerolles (42°56 N et 6°19 E), au sud de Hyères, dans le sud-est de la France, comme le montre la figure 4.1. Elle est positionnée à 27 mètres de profondeur, et enregistre à 50 kHz avec deux hydrophones espacés de 1,83 m. Bombyx est orienté vers le sud, ce qui signifie que l'évolution du TDOA permet de savoir si un groupe de cachalots va d'est en ouest ou d'ouest en est. Puisque BOMBYX est complètement émergé et à 27 mètres de profondeur, il réduit le bruit de surface généré et enregistré. La carte son a été fabriquée par OSEAN et les deux hydrophones sont des Neptune D / 70 - (voir Figure4.4). Le protocole d'enregistrement a évolué au fil des années (de l'enregistrement entièrement continu à 5 minutes toutes les 20 minutes, variant entre des encodages en 16 ou 24 bits), avec des sessions d'enregistrement d'une durée d'environ 3 mois. Une équipe de plongeurs est alors envoyée pour changer les piles et récupérer le disque dur contenant les enregistrements. Avec une estimation de 400 cachalots en mer Méditerranée, Bombyx sert de station de surveillance pour étudier cette espèce menacée. C'est aussi pourquoi Bombyx est situé dans le sanctuaire Pelagos qui vise à protéger les mammifères marins.

Sphyrna

Contrairement au chapitre précédent qui présentait une antenne fixe bi-hydrophone, avec une basse fréquence à enregistrer pendant des mois, la base de données qui sera présentée ici est une antenne mobile à cinq hydrophones.

Nous avons conçu une carte son à taux d'échantillonnage élevé et un algorithme acoustique passif robuste pour un petit réseau non uniforme de 5 hydrophones montés directement sous un véhicule de surface autonome (ASV) appelé Sphyrna. Cet ASV a été construit par Sea Proven en France et est illustré à la figure 5.1. Cette configuration est compliquée en raison de la petite taille de l'antenne et de l'enregistrement du son près de la surface, au-dessus de la thermocline. Les enregistrements sont réalisés à 16 bits à 600 kHz avec notre carte son haute résolution disponible JASON (Figure 5.3), conçue dans notre plateforme scientifique SMIoT. Nous proposons donc une méthode de filtrage efficace des transitoires et autres bruits générés dans ces conditions extrêmes, afin de surveiller les cétacés butinant en profondeur comme le cachalot qui passe plus de 70% de son temps jusqu'à 1 km de profondeur. Dans ce qui suit, nous illustrerons nos algorithmes pour traiter les impulsions d'écholocalisation de Pm qu'il utilise pour l'orientation et la localisation des proies. L'analyse de leur mouvement peut également aider à déduire leur comportement, aidant à la création d'un programme de conservation. Tel qu'il a été expliqué précédemment, le détecteur de clics produit des faux positifs. Puisque que le Sphyrna possède 5 hydrophones, la position de la source pour chaque clic a pu être estimée. Les clics peuvent donc être regroupés en pistes de clics, correspondant à la trajectoire des cachalots. Les faux clics ne correspondant à aucune trajectoire, seront ainsi filtrés. Ces pistes sont calculées à l'aide de l'algorithme de clustering DBSCAN, sur les TDOA, car l'estimation de distance de la source est imprécise, créant des trajectoires plus oscillantes. L'algorithme DBSCAN connecte entre eux tous les points ayant suffisamment de voisins, s'ils sont voisins eux même. Dans notre cas, le nombre de voisins nécessaires est de 2, car l'on cherche à faire des pistes (les voisins étant le point suivant et le point précédent dans la piste), et un point est voisin s'il est suffisamment proche en temps et TDOA. Les paramètres convenables pour filtrer les faux clics font qu'un seul cachalot aura plusieurs petites pistes, mais cela peut être résolut en appliquant l'algorithme une deuxième fois, ou en utilisant une autre technique une fois que les faux clics ont été filtrés.

Base de données de l'île Maurice

Après avoir décrit deux bases de données qui enregistrent les clics de cachalots loin de la source, nous présentons ici une base de données où les cachalots ont été enregistrés à courte distance. Nous proposons également un ensemble de méthodes pour utiliser TDOA afin de récupérer la position source et l'ajout de traitements vidéo afin d'identifier la source.

Sous l'autorité de la Marine Megafauna Conservation Organization dirigée par H.Vitry et, dans le cadre du programme mondial Maubydick, une équipe dirigée par F. Sarano a mené une étude longitudinale sur le même groupe de 27 cachalots au large de la côte ouest de l'île Maurice, depuis 2013. L'objectif principal est de comprendre la relation entre les individus au sein du groupe familial et la dynamique de la population de cachalots mauriciens. La principale originalité est que, depuis 2017, le protocole de données est renforcé par une collaboration avec H.Glotin par l'utilisation d'une matrice d'hydrophones à taux d'échantillonnage élevé, Blue JASON, de SMIoT et LIS DYNI, qui permet d'enregistrer leurs comportements acoustiques le plus intime tout en minimisant leur perturbation. Depuis 2017, nous avons pu améliorer l'antenne acoustique en lui ajoutant chaque année un hydrophone supplémentaire. L'antenne possède aussi une GoPro. Puisque l'on arrive à localiser la position de la source avec les hydrophones de l'antenne, on peut positionner la source dans la vidéo fournie par la gopro, et ainsi identifier l'individu qui a émit ces clics. Ceci permet de créer une base de données de clics et d'individus, mais aussi d'étudier les comportements résultant après une séquence de clics.

Réseau de neurones

La surveillance acoustique passive est aujourd'hui une approche courante pour la surveillance de la biodiversité. Son efficacité repose sur un vaste ensemble de données, et donc une détection automatique fiable des espèces. Cet article traite d'un type particulier d'émission, les transitoires des odontocètes, qui sont des impulsions à large bande de courte durée. Nous présenterons une étude de cas, le projet CARI'MAM, et décrirons comment un ensemble de données de références pourrait être construit pour un tel suivi. Ensuite, nous proposons une nouvelle approche pour la classification des clics basée sur un modèle CNN de bout en bout. Le projet CARI'MAM vise à créer un réseau de gestionnaires d'aires marines protégées répartis dans toute la mer des Caraïbes pour la conservation des mammifères marins. Afin d'étudier la répartition des mammifères marins, un système mono-hydrophone devait être déployé ce printemps pendant 40 jours dans 20 endroits différents, mais le déploiement a été retardé. La quantité de données collectées sera trop importante pour être analysée manuellement. Pour préparer cette analyse, nous avons créé un premier jeu de données constitué de clics des différentes espèces présentes dans les Caraïbes. L'ensemble de données proposé contient 10 des 30 espèces que le projet CARI'MAM vise à étudier. Ce premier corpus nous permettra de tester les différentes techniques d'analyse semi ou entièrement automatisée ainsi que de former des modèles préliminaires d'apprentissage profond pour résoudre la tâche de classification. Cet ensemble de données est également distribué comme référence pour la classification des clics dans le défi DOCC10 (Dyni Odontocete Click Classification). Pour créer un ensemble de données suffisamment grand pour former des réseaux de neurones, nous avons collecté des données à partir de différentes sources: i) le défi DLCDE 2018, créé par HILDEBRAND John , GLOTIN Hervé, ROCH Marie et al., et ii) des clics de cachalot de l'expédition Sphyrna Odyssey 2018. Ces ensembles existants contiennent de longues séquences audio avec des annotations approximatives des régions temporelles avec des clics. Notre objectif est de produire un ensemble avec des clics individuels associés à une espèce particulière. Dans ce travail, nous présentons notre méthodologie pour extraire les clics et les étiqueter avec l'identité de l'espèce. Nous présentons également une analyse préliminaire du corpus résultant, une répartition des données utile pour le benchmarking et un modèle d'apprentissage profond pour classer les clics servant de point de départ pour comparer les performances des autres réseaux. Même si notre méthode d'extraction de clics et d'étiquettes peut induire un certain bruit d'étiquette, il s'agit d'une situation rencontrée dans un scénario réel, augmentant ainsi la validité écologique du jeu de données. De plus, cela permet d'explorer l'utilisation de techniques traitant spécifiquement de ces problèmes, comme l'apprentissage négatif. Nous avons donc décidé d'augmenter le nombre d'échantillons, au prix d'une éventuelle augmentation des erreurs d'étiquetage.

Autoencoder

Dans le domaine bioacoustique, il existe de nombreux signaux inconnus ou mal connus. Les approches habituelles basées sur les réseaux de neurones pour les signaux sonores sont faites pour la parole humaine et utilisent les connaissances sur la façon dont ce signal est créé, comme les phonèmes existants ou même la structure du langage, pour améliorer les performances. Certaines recherches ont été effectuées sur les signaux audio bruts. Les réseaux de neurones sur les formes d'ondes brutes ont également montré qu'ils permettaient d'obtenir de meilleures capacités de débruitage, en particulier par rapport à d'autres méthodes qui copient simplement la phase d'origine. Sans cette connaissance d'un dictionnaire de signaux à classer, on pourrait essayer un apprentissage non supervisé, comme un auto-encodeur. Dans cet article, nous étudions les jeux de données stéréo de bruits transitoires. Nous montrons qu'en ajoutant une deuxième branche dans le décodeur, il est possible de réaliser un meilleur démêlage des codages. Cette étude a été motivée par l'analyse d'un ensemble de données de clics de cachalots non étiquetés.

Méthode numérique

Pour des géométries simples et des matériaux homogènes, des solutions analytiques au modèle de propagation acoustique peuvent être trouvées, qui donnent un aperçu de ce qui se passe et des mécanismes impliqués. Cependant, dans la plupart des applications du monde réel avec des géométries irrégulières et des milieux inhomogènes, aucune solution analytique ne peut être trouvée. Il faut donc recourir à des méthodes numériques pour résoudre le modèle. Même si les méthodes numériques ne donnent qu'une approximation de la solution, elles ont l'avantage de pouvoir s'adapter facilement à tout changement de paramètre (géométrie, matériaux, etc.). De nombreuses méthodes numériques existent, chacune répondant à des besoins différents. On choisira ainsi la méthode appropriée en fonction des besoins (précision, type de matériel) et des ressources disponibles, telles que le temps, ou l'utilisation CPU, GPU et mémoire. Trois grands types de méthode existent : les différences finies, les éléments finis, et les tirs de rayons.

Les différences finies sont un groupe de méthodes qui approximent les dérivées par des différences de valeur de la fonction. Les différences finies sont moins précises que d'autres méthodes, mais plus rapides à implémenter. Les éléments finis sont un autre type de méthodes qui découpe le domaine Ω en élément finis sur lesquels l'on va approximer le système d'équations différentiel partiel en système d'équations linaires. Ces équations sont résolues aux nœuds du maillage, et la solution est ensuite extrapolée sur l'ensemble du domaine. Un dernier type de méthodes est le tir de rayons. Cette technique est similaire à la lumière que l'on peut modéliser avec des rayons.

Propagation d'ondes dans la tête du cachalot

Parmi les méthodes précédemment décrites, la méthode des différences finies a été choisie pour construire la simulation pour sa facilité de mise en œuvre, et la rapidité de la simulation, permettant plus de pas de temps pour une discrétisation spatiale donnée. Dans ce chapitre 11, cette méthode est utilisée dans le domaine temporel afin de modéliser la propagation des ondes à l'intérieur de la tête du cachalot.

Depuis les années 90, les scientifiques modélisent la propagation des ondes sonores vocalisées dans la tête des mammifères marins. La capacité à modéliser la propagation des ondes chez les mammifères marins permet une meilleure compréhension de l'interaction entre tous les organes responsables de la création sonore, ou du moulage de l'onde sonore, pour obtenir le diagramme de faisceaux très directifs de ces espèces. A notre connaissance, ces types de simulations n'ont pas été réalisés sur le biosonar des cachalots, la plupart de ces simulations sont basées sur des données anatomiques dérivées de tomodensitométrie (CT). Ces informations permettent de construire la géométrie du modèle, et d'obtenir les paramètres mécaniques de chaque matériau et leur emplacement (jusqu'à la résolution du scanner). Cependant, la plupart des scans employés ont été effectués sur des personnes post mortem. Dans [186], les données ont été comparées entre les spécimens morts et vivants et leurs effets sur les simulations. Les spécimens morts sont susceptibles d'introduire des artefacts dans le modèle, tels que des vaisseaux sanguins remplis d'air, mais ne souffriront pas d'erreurs de numérisation dues au mouvement d'un spécimen vivant. Cependant, ces écarts ne changeront probablement pas les paramètres mécaniques des différents tissus, et donc l'unité de Hounsfield que le scanner va mesurer, s'est avérée bien corrélée à la densité et à la vitesse du son.

Problème inverse

Cette dernière partie de la thèse pose les bases théoriques et algorithmiques pour résoudre les problèmes inverses d'identification des paramètres, qui combinent données de mesure et modèles de propagation. Comme on l'a vu dans toutes les parties précédentes, on ne peut jamais avoir un modèle complètement précis de l'acoustique. Ainsi, l'objectif de l'approche inverse est d'utiliser les données enregistrées pour emph améliorer les modèles, puis ces modèles améliorés peuvent être utilisés à leur tour pour mieux analyser, identifier et prédire les sons eux-mêmes. En particulier, nous pourrions alors étudier les effets des externalités sur le comportement et la sécurité des cétacés.

Cette partie n'est pas complète, faute de temps, mais fait l'objet d'efforts de recherche en cours et futurs.

Dans ce chapitre 12 nous présentons les bases théoriques du couplage des mesures avec les modèles de propagation. L'approche que nous utilisons est basée sur une équation adjointe et est étroitement liée à l'assimilation des données. S'il y a du bruit dans les mesures et que nous souhaitons obtenir une caractérisation complète de l'incertitude, alors une approche bayésienne peut être utilisée. D'autres couplages sont possibles entre l'apprentissage automatique et les modèles PDE. En particulier, nous pourrions utiliser le modèle pour générer des réalisations pour l'algorithme ML. Une autre possibilité est d'insérer les équations différentielles dans le réseau neuronal, formant une méthode dite "d'équation différentielle neuronale" qui couple les deux. Cela fera l'objet de projets de recherche à plus long terme.

Faisabilité et implémentation du couplage

Afin de montrer que le couplage est possible, il faut d'abord commencer par faire une expérience synthétique sur une version simplifiée du modèle du cachalot, avec la supposition que les mesures sont disponibles dans le domaine de l'eau qui entoure ce modèle simplifié de la tête du cachalot. Cette supposition est ensuite simplifiée en ne prenant qu'un nombre limité de points de mesure. Du bruit artificiel est rajouté aux mesures. Les mesures sont réalisées sur les résultats d'une première simulation avec des paramètres arbitraires qui représentera la vérité. Le couplage est ainsi testé en partant de paramètres différents des paramètres de la vérité. La façon dont les paramètres initiaux convergent vers les paramètres vérité permet d'évaluer le couplage. Cependant, à causse d'un manque de temps les résultats de ces expériences ne sont pas rapportés ici.

Conclusion

Le processus complet d'analyse développée tout au long de ce manuscrit vise à améliorer nos connaissances sur les mécanismes qui régissent le sonar de cachalot. Avec les cachalots, les expériences classiques qui sont effectuées sur des chauves-souris ou des dauphins pour qualifier pleinement leur sonar dans tous les aspects ne peuvent pas être faites, car les cachalots sont des espèces sauvages et de grande taille. Cela signifie qu'il n'y a pas de mesures de leur réponse nerveuse, de tests de leur capacité auditive, de mesures précises de leur modèle de faisceau, ou de toute autre expérience qui oblige l'animal à porter un gros équipement ou à être entraîné à avoir un comportement spécifique pour une expérience. D'où la nécessité d'une simulation du sonar de cachalot.

Comme décrit dans le chapitre 1, le besoin a été divisé en trois objectifs:

- La réalisation de la simulation.
- Le couplage de la mesure avec la simulation.
- L'utilisation de l'apprentissage automatique pour classer et améliorer les mesures.

Les résultats de ces objectifs qui ont été explorés par cette thèse seront maintenant résumés, puis quelques perspectives de recherches prometteuses seront présentées.

Acquisition et traitement des données

Trois bases de données ont été présentées, chacune ayant des caractéristiques spécifiques:

- BOMBYX, une base de données constituée de deux hydrophones, à 50 kHz, lors de sessions de longue durée, présentée au chapitre 4.
- Sphyrna, une base de données constituée de cinq hydrophones, à 384 kHz, enregistrée sur ASV, présentée au chapitre 5.
- Sarano, une base de données constituée de deux à quatre hydrophones, à 300 kHz, enregistrée près du cachalot, présentée au chapitre 6.

Ces bases de données ont été analysées en utilisant les techniques de traitement du signal introduites au chapitre 3, suivant le processus de détection, d'estimation TDOA et de filtrage.

Détection L'étape de détection qui a été proposée et utilisée pour les trois bases de données est également constituée d'étapes successives, commençant par un filtre passe-bande, suivi d'un filtre TK, et se terminant par l'utilisation d'un seuil. Un autre procédé de détection est également mentionné, utilisant l'architecture de réseau neuronal du chapitre 7 pour classer dix espèces de mammifères marins en fonction des enregistrements bruts de leurs clics.

Estimation des TDOA La méthode SRP géométrique a été présentée pour estimer le TDOA et a été comparée au SRP standard. Les deux méthodes ont la même complexité et présentent les mêmes avantages par rapport à une seule paire d'estimations TDOA de canal, lorsqu'elles sont utilisées sur un système avec la même quantité de bruit et le même niveau de signal sur tous les canaux. Cependant, ils diffèrent par le fait que SRP utilise l'addition, tandis que SRP géométrique utilise la multiplication. Si un clic n'est pas présent sur un canal, SRP produira toujours un bon TDOA pour les autres canaux, tandis que le SRP géométrique produira un TDOA aléatoire avec une valeur de corrélation inférieure. Lorsque tous les TDOA sont nécessaires pour estimer la position de la source, ces deux éléments facilitent le filtrage de ces clics. Puisque BOMBYX ne possède que deux hydrophones, le SRP géométrique n'a pas été utilisé pour son analyse, mais il a été utilisé pour les deux autres bases de données.

Filtrage Deux méthodes de post-filtrage ont été présentées. Dans un premier temps, la méthode de l'ellipse a été utilisée pour récupérer la forme de l'antenne, mais aussi pour filtrer les clics parasites non générés par les signaux provenant de l'eau, donc avec TDOA éloigné de l'ellipsoïde. L'autre méthode de filtrage était le clustering DBSCAN présenté dans la section 5.5.1. Partant de l'idée qu'un point à l'intérieur d'une piste avec ses deux voisins peut être vu comme un cluster de trois points, la piste étant une succession de clusters connectés, DBSCAN est la méthode de clustering qui correspond à cette définition exacte. Avec BOMBYX n'ayant que deux canaux, les clics avec TDOA aléatoire ont plus de chances d'être considérés comme faisant partie d'un cluster et de ne pas être filtrés. Cependant, cette méthode était toujours utilisée pour les trois bases de données car elle donnait de bons résultats même sur BOMBYX.

Machine learning

Dans le chapitre 7, nous introduisons un nouveau défi de détection des clics *Odontoceti*, nommé DOCC10. Il répond au besoin d'avoir d'une grande base de données de mammifères marins transitoires dans l'intention d'avoir suffisamment d'exemples pour entraîner un réseau de neurones, ou une autre méthode d'apprentissage automatique, afin de pouvoir classer chaque clic en fonction de son espèce. La présentation de DOCC10 a été introduite avec une autre base de données nommée DOCC7, une version de DOCC10 où trois classes ont été supprimées, car elles croisent d'autres classes, ou sont trop différentes entre le train et les ensembles de test.

Une architecture de réseaux neuronales a également été proposée parallèlement à la base de données DOCC10. À ce jour, il a atteint un score de pointe sur DOCC10 lorsque seul la base de données d'entraînement est utilisée. Ce modèle, qui utilise directement le signal brut comme entrée, peut également être utilisé comme détecteur.

Perspectives

Comme expliqué dans le chapitre 13, les implémentations du couplage dans les cas synthétiques et réels sont en cours, et n'ont pas été intégrées à ce manuscrit en raison de contraintes de temps. L'implémentation dans le cas réel ajustera les paramètres du modèle, améliorant le clic synthétique généré. À ce stade, les mécanismes qui apparaissent dans le modèle peuvent être étudiés pour émettre des hypothèses sur les mécanismes réels qui régissent le sonar. Cependant, le modèle pourrait encore être amélioré à ce stade. En effet, la méthode numérique qui a été choisie a été le FDTD en raison de sa simplicité et de sa vitesse d'itération, avec pour inconvénient d'avoir une précision moindre. Avec les paramètres améliorés, le modèle et la méthode de couplage pourraient être mis en œuvre avec une méthode numérique plus précise, telle que la méthode des éléments finis spectraux. L'objectif ultime est de combiner cette modélisation de propagation avec les techniques d'apprentissage automatique utilisées pour l'analyse des signaux. C'est un travail en cours.

Comme indiqué dans le chapitre 12, l'apprentissage automatique peut également être utilisé pour estimer les paramètres d'un modèle en utilisant l'observation comme entrée. Cet usage ainsi que la possibilité d'incorporer les équations différentielles directement dans le réseau de neurones sont deux aspects clés des réseaux de neurones qui pourraient aider à résoudre des problèmes inverses, ce qui en fait des axes de recherche à fort potentiel.

Un autre point d'amélioration est l'utilisation de la théorie des ondes non linéaires. L'utilisation de la théorie des ondes linéaires est utile comme première approche et facilite l'interprétation des résultats générés par le modèle. Pourtant, nous sommes bien conscients que les clics plus forts émis par les cachalots ont une courte distance de choc, ce qui signifie que les effets non linéaires ne peuvent être négligés, même à l'intérieur de sa tête.

La base de données créée à la fin du chapitre 6 associait les clics au cachalot qui les avait émis. Cette base de données a l'avantage d'explorer les fonctionnalités partagées entre les clics d'un même cachalot, qui si elles existent et sont suffisamment discriminatoires (excluant ainsi l'IPI) conduiraient à un classifieur capable de distinguer les cachalots en utilisant uniquement l'acoustique. De plus, l'interaction sociale entre plusieurs individus peut être étudiée, puisque le canal vidéo nous permet de voir le comportement résultant de divers échanges de coda, tandis que la base de données vidéo nous permet de savoir quel individu a émis quel coda.

Ces trois perspectives de recherche montrent que la quantité de travail restante n'est pas faible, mais conduira sûrement à des résultats intéressants.

Notation

ABC	Absorbing boundary condition
CFL	Courant-Friedrichs-Lewy
CNN	Convolutional neural network
DOA	Direction of arrival
GCC	Generalised cross correlation
ICI	Inter click interval
IPI	Inter pulse interval
FDM	Finite difference method
FDTD	Finite difference time domain
FEM	Finite element method
FFT	Fast Fourier transform
FOV	Field of view
MLDB	Monkey lips dorsal bursae
PML	Perfectly matched layers
RNP	Right nasal passage
SNR	Signal to noise ratio
SRP	Steered response power
TDOA	Time difference of arrival
TK filter	Teager-Kaiser filter
TLM	Tangent linear model
TOA	Time of arrival

Contributions of this Thesis

- Chapter 2
    - Modification of the BIVIP algorithm to estimate the IPI.
- Chapter 3:
    - Proposal of a click detection pipeline.
    - Proposal of the Geometric SRP TDOA estimation method.
    - Application of the TDOA ellipsoid distribution for both antenna shape estimation and filtering.
    - Development of an audio synchronisation method between a lossy compressed signal and a dilated signal.
- Chapter 4:
    - Estimation of sperm whale trajectories on a two-hydrophone antenna.
- Chapter 5:
    - Proposal of DBSCAN as a track estimation method for filtering purposes.
    - Estimation of sperm whale diving trajectories.
- Chapter 6:
    - Application of an image segmentation network to videos of sperm whales.
    - 3D source localisation and video plotting.
    - Creation of a click database with classification by individuals.
- Chapter 7:
    - Creation of the odontocetes click dataset and challenge, DOCC10.
    - Formulation of the UpDim model architecture.
    - First End to End efficient biosonar classification model.
- Chapter 8

  - Development of an autoencoder to denoise the signal and estimate the TDOA.

- Chapter 11:
  - Creation of a 3D model of an adult sperm whale nose organ geometry.
  - Application of a FDTD wave propagation simulator on the model.

- Chapter 13:
  - Formulation of a measurement-simulation coupling method for the solution of inverse problems in wave propagation.
  - Derivation of a complete adjoint formulation for a full wave propagation model with absorbing boundary conditions.

Contents

List of Figures

List of Tables

Chapter 1

Context and Motivation

1.1 Sperm whales and their sonar

Sperm whales (*Physeter macrocephalus*, *Pm*) have the most powerful bio-sonar in the animal kingdom (230 dB re: 1 μPa rms, [1]). The clicks produced by this sonar are not only used for their echolocation during dives, but also in their social interactions. During dives, sperm whales emit trains of clicks, much like those of bats, whereas for socialization, they will emit small rhythmic series of clicks. Since [2] first theorized the way their sonar worked, it has been broadly accepted that *Pm* creates an initial pulse at the front of its head, in the *"museau de singe"* (aka. monkey lips), which will then bounce back and forth in its head, passing through multiple oil sacs, before exiting. However, the details of such a mechanism and which parameters the sperm whale can act on, remain unknown.

1.2 Human impact

1.2.1 Whaling

While they were already hunted in other parts of the world, American whalers caught their first sperm whales around the year 1720 [3]. With this event started two large periods of sperm whale hunting [4, 5, 6]. The first period ended in the 1870's when the petrochemical industry started to produce products that could replace the spermaceti oil, and the second starting in the 1950's when the other large baleen whale populations declined. While the in-between hunting period seemed to have helped the sperm whale population recover, [6] estimated that the population was around 1 110 000 in the pre-hunting era, 788 000 in 1880, and only 360 000 remaining in 1999. The end of the hunting did not entirely protect the sperm whale population from human interference, as progress brought modern threats.

1.2.2 Sound pollution

With the arrival of motors, marine vessels started to emit noises. With the increase of ferries, cruise ships, and other goods transportation, the noise level followed. This rise in noise level has a high impact on marine mammals who then suffer from an increase of stress [7, 8, 9]. The background noise also affects the range of echolocation for clicks of the same energy, making it harder to detect preys.

Sound pollution is also produced by seismic surveys, which mostly use air guns which blast loud, low frequency pulses. Seismic surveys seldom use explosives to create these kinds of pulses. While the use of air guns does not seem to bother sperm whales at first [10], sperm whales leave the survey areas, without returning to them even days after the survey has ended [11]. Note that sound pollution does not necessarily stop after the initial survey, since the goal of seismic surveys is to find digging sites, whose underwater mining will generate additional acoustic pollution [12], [13].

1.2.3 Sea pollution

Another common modern threat is that of sea pollution of varying kinds that accumulates inside sperm whales' stomachs, such as ingestion of plastics, slowly killing them [14, 15].

1.2.4 Collisions

Anther major threat faced by sperm whales, is collision with ferries [16]. As demonstrated in [17], sperm whales might not be able to perform escape maneuvers in time, making this a major threat if not the primary one, as in other cetaceans species [18, 19].

1.3 Importance of passive bioacoustics

Passive bioacoustics records and processes the audio environment, while active bioacoustics will emit sounds and record the echoes produced by that sound, in a similar way as a sonar works, except that the recording antenna can be dissociated from the source. Control over the source (time of emission, power, signal, ...) increases the amount of information obtained with active bioacoustics, allowing to gather some key data (e.g. distance to the target). However active bioacoustics has the disadvantage of emitting sound. This implies additional sound pollution, potential perturbation of animal behaviour, higher energy consumption and will overshadow simultaneous animal emissions, rendering it useless for many studies. It should be noted that amongst passive bioacoustic methods we find the use of DTAGs, which are small recording device with suction cups that are placed on the animals, however this technique may potentially also perturb the animal behaviour.

Underwater visibility is limited to a range of between 20 m to 60 m, while some species like sperm whales, can be heard 5 or 10 km away. Such facts make passive bioacoustics an ideal tool to monitor many marine mammals. In the case of sperm whales, who emit clicks during both echolocation and social interaction, passive bioacoustics allows us to obtain information about their hunting behaviour, record the sequence of clicks they emit to communicate, and estimate their size.

1.4 The challenge of data analysis

The drawback of passive acoustics is the tremendous amount of data that is generated by the recordings, with most of it filled with background noise. Thus the data needs to be curated in order to greatly reduce the effort in further in-depth analysis. This will both reduce the amount of time of these analyses, and improve the results. Let us take the recording of a sperm whale clicking every second as an example. These clicks can be contained within intervals of 20 ms, with some background noise included. This means that only 2% of the signal could be kept for further analysis. While periods when the sperm whale is close to its prey will see this number increase, or be null during breathing time without social interaction, the main hypothesis is still that a sperm whale is present, which is not true most of the time on stationary recording stations.

The typical data analysis tasks done on bioacoutics databases are detection, classification, localisation, and density estimation, with the latter usually using the previous analyses.

Between detection, classification, and localisation, the classification tasks can be considered the hardest. While conventional signal processing techniques exist to handle these tasks, they are still prone to errors. Thus multiple biologists choose to manually annotate their databases [20, 21], sometimes even without using a detector to reduce the amount of preliminary work. In this setting the amount of work requires more than one expert and takes large amounts of time, of the same order as the time it took to record the database in the first place.

1.4.1 Deep learning

With the increase of computing power and the appearance of large databases, neural networks have started since the years 2010 to be applied to many practical cases. They have the potential to become the state of the art in terms of methods to use in various hard perceptual tasks. With the previously mentioned classification tasks being the kind of task where neural networks obtain better results than classical methods, they started to be applied to bioacoustic signals in [22, 23, 24], with current results proving that the classification task can be solved with them, potentially removing the need of large amounts of experts' hours.

1.5 Simulation

While simulations do not simulate all the existing mechanisms present in reality, they are still an alternative to real experiments. With simulation, the whole environment is observable and all the parameters can be modified. The observation is also done without the use of any measurement devices, removing any artefacts that they could introduce. This does come at the cost of not being a perfect copy of reality, which not only means imperfect results, but also means that unrealistic behaviour can arise from the simulation. Hence, at the end, the simulation results need to be compared with real measurements.

In the case of marine mammals, simulations have been used to simulate the propagation of acoustic waves inside the heads of multiple species, such as *Ziphius cavirostris*, *Delphinus delphis*, *Tursiops truncatus*, *kogia breviceps* or *Lipotes vexillifer*. Simulating these waves in marine mammals allows us to estimate multiple characteristics about these species, such as their beam patterns or the click shapes after emission.

1.6 Goals and Plan of the Manuscript

The study of sperm whales is a vast subject, that covers multiple and varied disciplinary fields, from ethology to conservation, passing through bioacoustics. This is even wider when the methods need to be adapted to this subject are considered. Since this manuscript cannot integrate such a vast subject, it will deal with a few issues linked to a core element of the sperm whale, which is its sonar.

Sperm whales are wild species, which unlike other species that also live in captivity, cannot be equipped with heavy instrumentation or trained in order to test their sonar capabilities. Without the possibility of controlled experimentation on real sperm whales, an option is to experiment on a virtual sperm whale, leading to the first goal of this thesis, the creation of a sperm whale biosonar simulator. However, a simulator by itself is useless, since one cannot know how far from reality it is. Hence, the second aim of this thesis, deals with how to improve the simulations using measurements. This second goal is further divided into two parts, the coupling method, and the acquisition of said measurements. The last aim of this thesis is the use of machine learning, to improve the speed of extracting useful data from recordings to learn features that characterise their acoustic clicks.

The manuscript is built of four successive Parts. After introducing the basic grounds of the signal processing techniques that will be used throughout our study, Part I describes three different databases: two far-field and one near-field, that were acquired and/or processed during the thesis. These databases of sperm whale recordings are the foundation of the second (Machine Learning) and last (Coupling) parts. Part II is about Machine Learning applied to bioacoustics. It uses these databases to train neural networks to discover what is a click, which after training can be used for the analysis of new databases. Part III of the manuscript introduces wave propagation theory and numerical methods in order to create a simulation of wave propagation inside the head of the sperm whale. Finally, Part IV is about coupling, using inverse problem theory to improve the crucial estimations of material parameters of the model.

Figure 1.1 shows how the chapters are connected together, not in reading order, but in a data flow manner, creating two feedback loops. The right loop is the model improvement loop, while the left loop is the data acquisition loop, itself feeding into the first loop. The data acquisition loop is made of Part I, in which the signal processing techniques

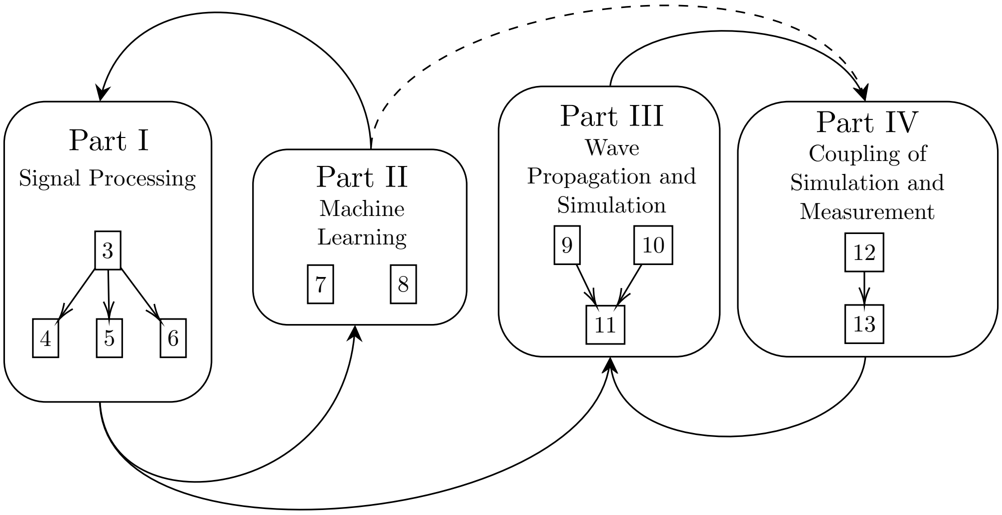


Figure 1.1: Flow of the chapters, with their feedback loops.
The Arabic numbers represent the individual chapters of each Part.

described in Chapter 3 are applied to the three databases, described in Chapters 4, 5, and 6. The first two databases are similar in the fact that they are both composed of recordings of sperm whale clicks far from the individual, but still have many differences. The first database described in Chapter 4 is made using an immersed recorder, with a low sampling rate, during long recording sessions. The database described in Chapter 5 was recorded by a surface drone, with a high sampling rate and multiple hydrophones, and with human labelling indicating periods when sperm whales were present. The last database described in Chapter 6 is different as it is made of hand-held recordings taken close to the sperm whales. Chapter 6's database also contained a video channel used to identify the sperm whales associated with their clicks. The products of the first part are then used as training data for the neural networks presented in Part II.

In Part II, Chapter 7 is dedicated to a classifier network to classify clicks from multiple marine mammals. Chapter 8 is dedicated to auto-encoders, used to cluster clicks in order to learn what kind of information can be extracted from them. The neural network can then be used to reanalyse the database of Part I and other new databases, which can then be used as new training data to improve the model performance, thus creating a feedback loop.

The second loop of the thesis structure starts with Part III. Chapter 9 describes the linear theory of wave propagation, while Chapter 10 describes and compares the numerical methods that can be used to solve a system of partial differential equations along with a description of the geometry. Chapters 9 and 10 lead to a wave propagation simulation, used in Chapter 11 to simulate wave propagation inside the head of the sperm whale. Chapter 11 starts with an in-depth description of the finite difference in time domain method used for the simulation, followed by the description of the geometry and mechanical parameters that will be used in the simulation, and ends with the first results obtained with this model. Particular attention is paid to the absorbing boundary conditions on the outer boundary of the simulation domain. Part IV then starts with Chapter 12 which details the base of inverse problem theory. Inverse problem theory is used to implement the coupling between the simulation and the measurements, tuning the parameters of the simulation to reduce the discrepancy between simulated clicks and recorded ones. Chapter 13 introduces a toy task used to demonstrate how this theory works on a practical example. The simulations of Part III and the coupling of Part IV thus form the model improvement loop, which, as explained earlier, is itself fueled with the improved measurement of the data acquisition loop.

Chapter 2

Use of Acoustics by Cetaceans

Unlike terrestrial mammals, which evolve in air, cetaceans move through water which reduces the usefulness of vision. Indeed, in clean air a human can see up to 250 km [25] whereas underwater visibility varies from none up to 60 m [26]. The latter is also limited by the lack of light at greater depth, at which some species, such as *Physeter Macrocephalus* or *Ziphius cavirostris* dive [27, 28]. While primates evolved their vision [29], cetaceans have evolved to exploit sound, which unlike light, propagates much more efficiently in water [30, 31]. This allows some species to communicate over multiple kilometers, and others to sense their environment using echolocation, similarly to bats. While multiple dissimilarities exist in the echolocation of bats and cetaceans, they are often compared since they share major points in common.

2.1 Hearing in cetaceans

The cetaceans' middle ear is largely derived from the mammalian middle ear [32]. Yet it differs mainly in the impedance of the soft body tissue, which is close to that of water, meaning that the acoustic energy can easily flow between these two media. The cetaceans' middle ear is made of massive and dense mineralized ossicles. Their joints are stiffened by membranous sheets and ligaments. In sperm whales, the tympanic membrane is replaced by a thin bony tympatic plate. The middle ear epithelium is thick and highly vascular. The middle ear cavity is filled with air. The eustachian tubes are large and serve in pressure equilibration. The skull and attached bones also serve as isolation between the whole hearing complex and the sound production complex. A model by [33] suggests that the cetaceans' middle ear amplifies the particle velocity, unlike the terrestrial mammal middle ear which amplifies the pressure.

2.1.1 Dolphin auditory processing

Unlike sperm whale, dolphins who also live in captivity have been largely studied, including their auditory system. One point of interest is to find the dolphin's capability to discriminate between two pairs of pulses with different interval. Crossing a threshold that seems to be between 0.2 ms and 0.3 ms showed multiple effects that indicate that the integration time of the dolphin is $0.2 - 0.3$ ms [34, 35, 36], which is an order of magnitude smaller than that of humans. Integration time can be understood as the temporal resolution. This value has also been confirmed with the evoked potential (EP) experiments that have been conducted on dolphins [37]. The EP response is nonlinearly proportional to the power of the sound [38, 39], which follows the principle of Stevens's power law. Using this fact combined with double click and gap detection experiments, the same authors estimated a temporal transfer function [40, 41, 42]. A temporal transfer function is a function that gives the gain of a system along time. This function has an initial part of 0.3 ms (which is in agreement with the integration time) and a long tail with a slope of 40 dB per time decade. If we consider that the dolphin can hear the emission of the signal, then the 40 db slope matches the spreading loss of spherical propagation of the wave.

The inferior colliculus (IC) has been shown to be one of the main centers in the auditory sensory system in terrestrial mammals. As expected, an immunocytochemical study [43] showed that the IC in echolocating mammals is enlarged relative to the whole brain. The study also showed that the IC in *odontoceti* cetaceans as well as in insectivorous bats showed a modular arrangement whereas it is homogeneous in non-echolocating mammals.

2.2 Acoustics for socializing

Sperm whales form a matriarchal society [44]. Made of groups of a dozen sperm whales, themselves composed of adult female sperm whales and juveniles, each group spreads over a vast hunting territory [45, 46]. At some point (between 10 and 15 years of age), male sperm whales leave the group to go hunting alone for the rest of their lives, only returning for short periods of time to reproduce with another female group that can be in another ocean [47, 48].

Within a sperm whale group, individuals will socialize using acoustics or tactile stimuli. They socialise acoustically using coda, creaks and clangs. Codas are sequences of clicks [49, 50]. They are classified by their rhythm. For example 2+6 design codas start with 2 clicks, followed by 6 regularly spaced clicks. Figure 2.1 shows an example of a 2+8 coda. The frequency of each type of coda used is different for each group of sperm whales, with some types of codas being exclusive to some regions. Some codas can be quite rare [51], while others such as the 5R are more commonly used [52].

Creaks or buzzes are long series of clicks with a small inter click interval (ICI) (less than a tenth of a second) with slowly varying ICI and click amplitude. Creaks can last up to a minute, and may have a few clicks missing. Examples are shown in Figure 2.2.

Clangs are sharp metallic sounds emitted by male sperm whales, with long ICI, usually from 5 to 10 seconds. Other signals are emitted by sperm whales, but are seldom used, such as trumpets or meows. An example is shown in Figure 2.3.

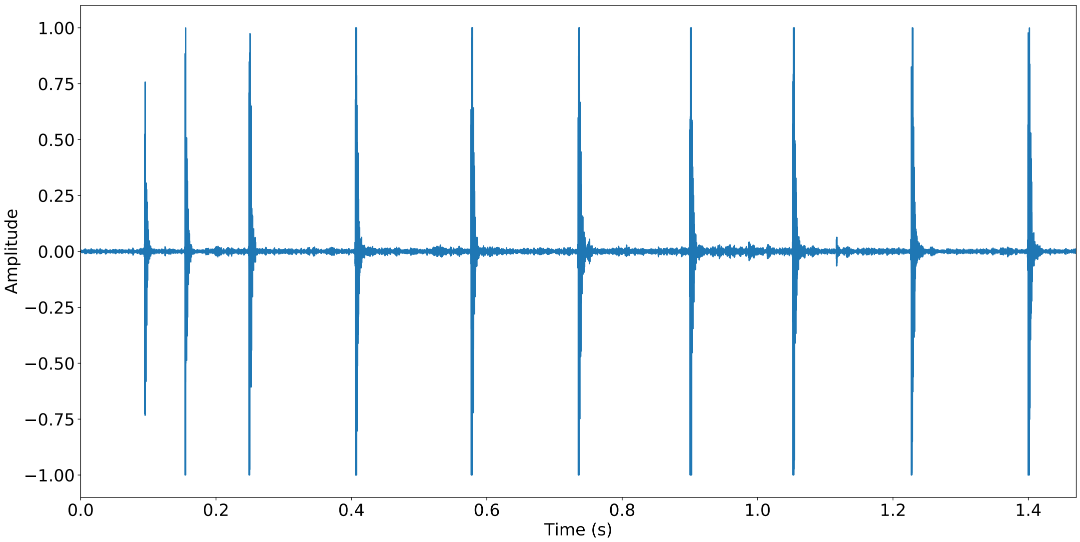


Figure 2.1: Example of a 2+8 coda.

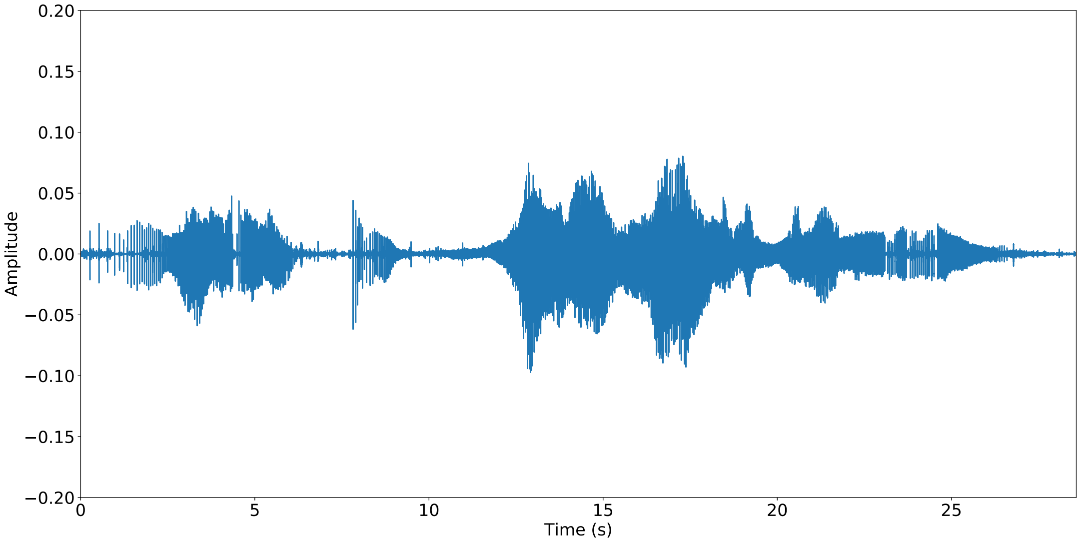


Figure 2.2: Example of creaks.

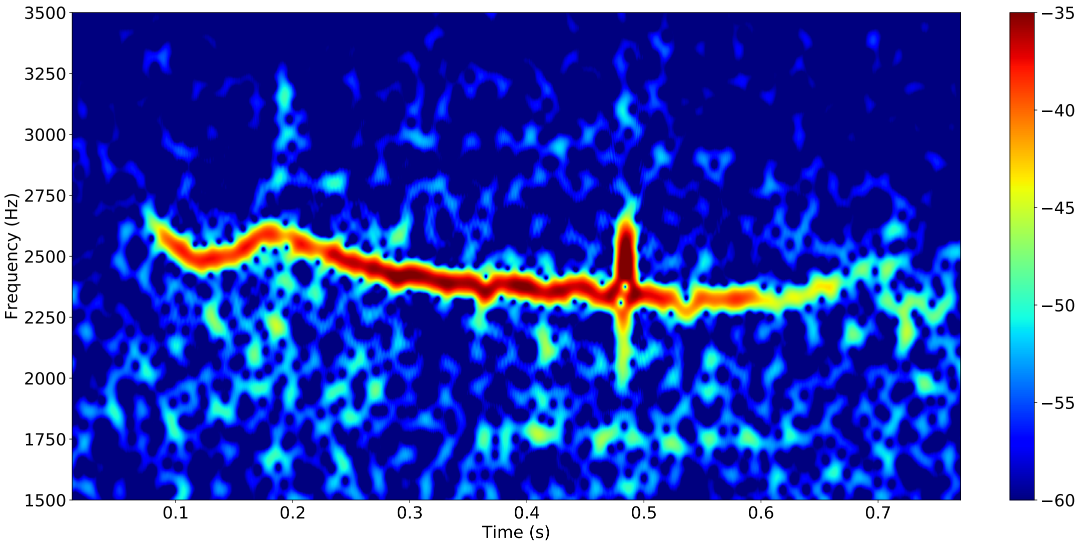


Figure 2.3: Example of a meow.

2.3 Acoustics for sensing

Many cetaceans will use an echolocation sonar to hunt preys or view their environment, as some species such as sperm whales are nearly blind, with some specimens totally blind yet still healthy [53]. The principle of echolocation is to send a short pulse. The pulse will echo from an obstacle such as a wall or a prey, and will be heard by the emitter. The time between the emission and the hearing will tell the emitter how far the obstacle is.

Research on echolocation consists of passive acoustic measurements and experimentation requiring training or equipping a specimen with some material. The latter type of research allows to study a particular point of interest, but requires a captive animal. Hence it has mostly been done on bats and dolphins.

Even though the sperm whale sonar's scale and complexity are different when compared with the one of the dolphins, the underlying mechanisms remain the same. In the same way that a plethora of things separate dolphins and bats (nasal air sacs vs larynx, water vs air, size, ...) yet their sonar capabilities share similarities, and work done on these other species that are easier to study can serve as a baseline for the sperm whale sonar. Even though this baseline may be quite far from reality, it still sheds light on some of the sperm whale sonar mechanisms.

Bats and dolphins both use echolocation to interact with their environment and to track their preys. Before starting a comparison between the two, it should be highlighted that bats have evolved in air, while dolphins live in water, in which sounds travels 5 times faster than in air, and also travel further away due to the difference in sound attenuation.

It should also be noted that dolphins have a larger head and emit a narrower beam.

Bats emit 3 types of signals: upward and downward frequency sweeps, and a constant frequency (CF) one [54]. The CF ones are of length from 1 to 300 ms, while the frequency modulated (FM) ones have lengths from 0.5 to 10 ms. FM signals are Doppler tolerant, whereas CF signals carry Doppler information, which is used by some species of bats, like the *Rhinolophus* to measure the relative speed of their preys.

Dolphins emit two types of signals, whistles and short broadband clicks, with a linear relation between the center frequency and the signal strength [55]. The center frequency usually goes from 30 to more than 100 kHz for high intensity signals.

Multiple experiments have been conducted to estimate the echolocation capability of bats and dolphins, and to compare them without forgetting their differences. The echolocation characteristics are spread among the following tasks: presence of target with and without external noises, closest object selection, recognition of an object with depth singularities, vertical angular precision and the detection threshold with multiple echoes [56].

Presence of target in a noiseless environment This experiment consists of detecting the presence of a sphere with the distance to the sphere as a parameter. To put bats and dolphins on the same level, the comparisons between the two have been done by calculating the energy of the echo, which is simply done by using the usual energy equation of wave propagation,

$$EE = S - L + G, \tag{2.1}$$

where EE is the echo energy at the emitter, S is the source energy, L is the loss and G is the gain, all of them in dB. Here the loss is composed of the spreading loss which is $20\log(R^2)$ where log is the logarithm in base 10, and the absorption loss which is $2\alpha(f)R$, where α is the absorption coefficient of the medium for a specific frequency f, and R is the radius of the sphere. The gain comes from the ratio between the energy of the emitted wave on the target surface and the energy re-emitted in the echo, which is in our case $20\log(\frac{a}{2})$ with a the diameter of the target, if it was considered as a perfect sphere [57]. We obtain the energy expression,

$$EE = S - 40\log(R) - 2\alpha(f)R + 20\log(\frac{a}{2}). \tag{2.2}$$

Species chosen for this task were *E. fucus* and *P. phocenea* [58, 59]. Since the *E. fucus* emits a large broadband signal and absorption in the air is stronger for higher frequencies, it is simpler to replace the source term with a source term that already takes into account the absorption,

$$AS = 10\log(\frac{1}{\rho c}\int_0^T \mathfrak{F}^{-1}(\mathfrak{F}[p_s(t)]e^{-2\eta(f)R})^2\,\mathrm{d}t), \tag{2.3}$$

where $\mathfrak{F}[p_s(t)]$ is the Fourier transform of the instantaneous acoustic pressure $p_s(t)$ and $\mathfrak{F}^{-1}$ is the inverse Fourier transform. $\eta(f)$ is the absorption coefficient of the medium but in Neper/m. Then (2.3) is basically the source energy flux density formula where the acoustic pressure is modified with the absorption inside the frequency domain. Equation (2.3) substituted in (2.2) gives the following equation for Echo Energy for bats,

$$EE = AS - 40\log(R) + 20\log(\frac{a}{2}). \tag{2.4}$$

The case of *P. phocenea* is much simpler since they emit narrow-band clicks. The absorption coefficient $\alpha(f)$ is almost constant on that band, which means the EE can be calculated by using the absorption coefficient for the center frequency of the clicks.

Results show that the echo energy flux density threshold for a 75% correct answer is 5 dB higher for the *E. fucus* compared to the *P. phocenea.* This small gap in sensitivity might be caused by variation in methodologies, or sources used in calculation, meaning that the echo detection sensitivity is either identical for the two species, or a notch better for the dolphin.

Presence of target in a noisy environment The same types of experiments have been conducted with added background noise. The aim was to evaluate the echo energy to noise ratio threshold to have a 75% positive answer. For *T. trucatus*, the estimated SNR for this task was 7.8 dB [60]. For bats the SNR needed seemed to be much higher as the first experiments estimated a threshold 36 dB and 50 dB on *Eptesicus serotinus* and *Pipistrellus pipistrellus* respectively [61, 62]. To explain this difference in performance compared to dolphins, the supposition was made that the speaker used to play back the echo had too much clutter. This supposition is supported by another experiment on bats (on *E. fuscus*) which had very little clutter, and showed a much lower SNR threshold of only 11.8 dB, which is much closer to the dolphin threshold, and coherent with the dolphin/bat results in noiseless environments.

Closest object selection The next task is the evaluation of the minimal distance between two targets that is needed for the animal to know which one is the closest. Two targets were positioned and the distance between them was progressively reduced. Results show that bats and dolphins share a 75% correct response threshold when the targets are separated by 1.5 cm. However, sound travel speed differs in the two media, therefore the results are to be interpreted in terms of arrival time difference instead of distance difference. *T. trucatus* can differentiate two signals delayed from each other by 2μs when the total travel time is 1 ms, and need 40 μs when the travel time goes up to 9.2 ms [63]. On the bats' side, *E. fucus* needs 77 μs whether the total travel time is 3 ms or 15 ms [64], while other bats such as *Rhinolophus ferrumequinum* also have a precision which is relative to the travel time [65].

Recognition of an object with depth singularities The aim of this experiment is to determinate how precisely two echoes can be discriminated in regards to time position of glints in their signal. For bats, these glints where generated with two Plexiglas planes with an array of holes [66]. The difference depth of the holes between the two planes made the difference of time position between the glints. For dolphins, the experimentation was conducted with two cylinders which had different wall thicknesses [67]. The supposition was that the echo is first created with the reflection on the outer wall of the cylinder and that the second emission came from the reflection on the inner wall of the cylinder(on the opposite side, meaning that the wave has to travel through the inner water cylinder).

For both experiments, the results can be expressed as time difference between the glint or as frequency difference between ripples produced by glint. If the echo is expressed as the sum of two reflections as follows, $e(t) = r(t)+a*r(t+\tau)$, where a is the attenuation of the second reflection in regards to the former, and τ is the delay between the two, we can calculate the Fourier transform of this echo [60] which is

$$\|\mathfrak{F}[e(t)]\|_1 = \sqrt{1+a^2+2a\cos(2\pi f\tau)}\|S(f)\|_1. \tag{2.5}$$

The delay of the glint will increase the frequency at which the cosine term produces ripples in the spectrum. Bats and dolphins might take cues of the frequencies delta between ripples of two objects with distinct hole depth or wall thickness. The final results showed that dolphins can differentiate glints with a $-0.5\,\mu$s and $0.6\,\mu$s resolution, while bats have a threshold of $-1.0\,\mu$s and $1.3\,\mu$s.

Angular precision The next experiment was to evaluate the angular resolution. To do so, two arrays of rods where used, with one array having smaller angular spreading of its rods. Subjects were then trained to locate the array with the smaller angle. The study concluded that bats have a angular resolution of $3.5°$ vertically [68] and $1.5°$ horizontally [69], while dolphins have a resolution of $0.7°$ vertically and $0.9°$ horizontally [70]. As mentioned earlier, the better results for the dolphins might come from its narrower beam.

Detection threshold with multiple echoes The final characterization of the echolocation was based on the threshold of detection when only one, and then two echoes where emitted, with the time between two echoes as a parameter. The results for dolphins and bats show that the integration time for the second echo seems to be at $264\,\mu$s for dolphins [71] and 2.4 ms for bats [72]. The threshold is also 3 dB lower when the second echo arrives before the integration time limit than when it arrives after.

2.3.1 Application to sperm whales

Equation (2.4) can be applied to sperm whales in order to estimate their detection range. For an echolocation click with its maximum of energy at a frequency of 12.5 kHz, a sufficient echo energy can be estimated at 66 dB in order to have a sufficient SNR [73]. At this frequency, for a salinity of 35 ppt, a depth of 1 km and a temperature of $8\,°C$, the absorption range from 1.028 dB/km to 1.338 dB/km [74, 75, 76]. Note that changing the parameters used to obtain the absorption coefficient will still provide values close to 1 dB/km, which will not have much influence on the final result since the absorption term is not the predominant term. Finally, for a target size of 1 m and a source level of 180 dB, equation (2.4) predicts a range of 470 m. Note that without the absorption term, the range will be 501 m. With the increase of the source level, the range will also increase, but the hypothesis of linear acoustics would not hold anymore.

2.4 Sound production

Bats adapts their sonar signal to the echoes they receive. CF bats will lower their emission frequency so that the echo frequency stays constant [77]. Using artificially frequency shifted echoes, they found that accuracy and consistency vary among species. As an example, Horseshoe bats will try to keep their echoes at 83.3 kHz by lowering their emission up to 6 kHz. Bats will also emit shorter signals when they get closer to their prey, and increase their repetition rate. FM Bats will also show changes in the bandwidth of their signal [78]. It has been shown that the speed of the frequency sweep and its maximum frequency increase with the developmental stage of the FM Bats [79, 80, 81]. As bats become adults they will also emit much shorter signals (see [82] as an example).

2.4.1 Pulse production in *Odontoceti*

The sonar source location of *odontoceti* is still a source of debate. Most studies have been focusing on locating the source, but the assumption of a single source has rarely been put in question. Knowing the latter might help in solving the former. The location of the source(s) has been tracked down to be in the nasal apparatus, which differs from bats and most mammals where the sonar signal originates from the larynx. The most supported theory is that the source location is in the structural monkey lips dorsal bursae (MLDB) complex. Note monkey lips are also called phonic lips since they do not look like monkey lips in smaller *odontoceti*. However, the MLDB is complex and the role of each of its components is not well understood. Basically, the primary air space conducts the air stream which will pass through the MLDB, to go to the inferior vestibule. Sperm whales show the particularity of having a single MLDB complex, while other *odontoceti* have two, one associated with each nasal passage.

Possible pulse production mechanism in *Odontoceti* Three production mechanisms have been proposed for the *Odontoceti*: friction/stiction, cavitation, and pneumatic-mechanical. However, the friction/stiction hypothesis has been invalidated by [83].

The cavitation hypothesis was proposed by K.S. Norris during one of his lectures in 1986 at U.C. Santa Cruz. One of the perks of this mechanism is that it is an efficient way to produced high intensity sound. A way that this mechanism might work is that micro bubbles are formed near the phonic lips during a low-pressure phase. The bubbles then collapse (cavitation) to form a powerful pulse. However, this requires that the bubble formation and collapsing be uniform and synchronized, which might put into question the possibility of this mechanism being used by *odontoceti*.

The last mechanism proposed is a simple pneumatically driven tissue generator [84]. The sound generation tissues are set into vibration and reach a state of self-oscillation. The sound generation could then be when tissues impact on one another.

Physical and mathematical model of the click production mechanism [85] The following physical model has been proposed: a rubber tube with a rubber ring on it. The rubber ring has a smaller diameter than the tube, thus the ring pinches the tube, preventing air from passing through. Yet, the ring is soft enough that when the air pressure in the inlet starts increasing, the ring expands and lets the air flow through the tube. The air flow then reduces the air pressure (Bernoulli's theorem states that $p + \frac{1}{2}\rho v^2$ is constant, which means that an increase in v means an decrease in p), which makes the ring close the tube again, which start a new cycle. The parameters that set up the cycle characteristics are the tube and ring stiffness (one for each) and the static pressure in the inlet. This setting produces clicks whose amplitudes are linked with the acceleration of the ring.

The mathematical model does not model the displacement of the ring but the acoustic pressure caused by it. The ring is assumed to have no weight except a small spherical part of it. Then the acoustic sound pressure can be found using the the acoustic sound pressure formula for a spherical acoustic dipole in the frequency domain [86],

$$p(r,t) = \rho\omega^2 a^3 \xi(\omega) \frac{(ikr-1)\cos(\alpha)\exp^{-iak}\exp^{ikr-i\omega t}}{(2-2ika-(ka)^2)r^2}, \tag{2.6}$$

where a is the diameter of the sphere, $\xi(\omega)$ is the displacement for a given frequency, k is the wave number, r is the distance to the center of the sphere, α is the angle between the direction of the displacement and the direction where the acoustic pressure is calculated. When a triangular displacement is input, results are similar to that of the physical model, the main difference coming from the input signal which is not a perfect triangle in real life.

2.5 Pulse structure of sperm whale clicks

The repetition of a pulse in a sperm whale click is explained by the bent horn model [1, 2], or the leaky bent horn model which adds the acoustic path for false pulses. [87]

Once the pulse has been emitted in the *museau de singe* (monkey lips), the pulse will propagate toward the rear of the head, which is represented by the black path in Figure 2.4. Part of the energy released when creating the pulse leaks directly, represented here as the orange path, which will be recorded first and is denoted as P_0. The pulse that follows the black path then encounters the frontal sac, which acts as a mirror, redirecting the pulse towards the front. From this point the pulse can follow three paths. Part of the energy will thus leak through the water following the red path. Another part will go through the junk to reach the water, following the magenta path into the cyan path. The rest of the energy will go back toward the monkey lips following the purple path, which after bouncing on the distal air sac, will restart the cycle by following the black

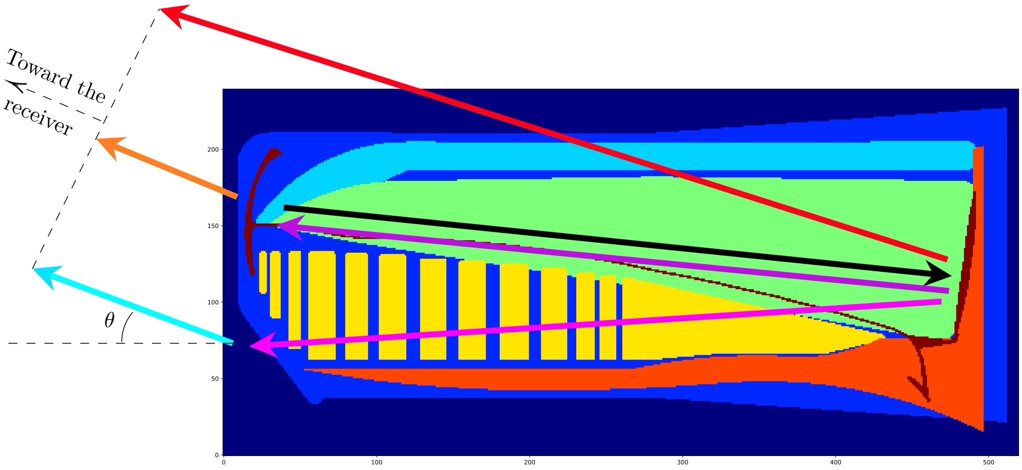


Figure 2.4: Pulse path in the head of the sperm whale in the leaky bent horn model.

path once again. The energy will thus bounce back and forth between the frontal and distal sac, while part of it will leave through the magenta and red paths. The pulses leaving through the cyan path contain most of the energy and are named $P_1, P_2, P_3, \dots$. The pulses that leak trough the red path are called half pulses, as they only go through half of the organs, and are denoted $P_{1/2}, P_{3/2}, P_{5/2}, \dots$. They are also named false pulses, thus sometimes named $P_{f1}, P_{f2}, P_{f3}, \dots$.

When θ is below 0°, the false pulses will not be recorded since they leak from the top of the body. Finally, when θ is above 90°, all the pulses come directly from the distal sac/*museau de singe* region, with successive pulses still spaced by the time to travel through the black and purple paths. In this configuration, P_0 will be the pulse with the most of the energy.

However other mechanisms must influence the production of clicks, since sperm whale clicks are different when they echolocate compared to when they socialize [88]. Echolocation clicks usually have most of their energy around 12.5 kHz with few subsequent pulses (at most P_0, P_1 and P_2 are above the background noise), while socializing clicks have most of their energy at 3.13 kHz or 6.25 kHz and display a lot of subsequent pulses.

2.5.1 An acoustic valve within the nose of sperm whales Physeter macrocephalus

In [89], they theorized a new hypothesis on how sperm whales could control the decay rate of their clicks, and thus explain the difference between echolocation and socializing clicks. Sperm whales have their spermaceti organ and junk, which are the main acoustic parts of their sonar, separated by the right nasal passage (RNP) that can be seen in [90] CT scan. [89] supposed that sperm whales can collapse their RNP, which when collapsed allows the sound to go through. When the RNP is not collapsed and filled with air, the high impedance mismatch between the air and the flesh makes the RNP act as a mirror which reflects some of the signal back to the distal sac. With this added reflection, part of the energy is conserved between the two mirrors, which are the frontal and distal sacs, and distributed to the subsequent pulses P_2, P_3, ..., and so on. They supposed that sperm whales collapse their RNP using their maxillonasolabialis muscle, which would pressurize the RNP and bend the head a little, which leads to the collapsing of the RNP.

2.5.2 Estimating the length of sperm whales

The inter pulse interval (IPI) is the time interval between the reception of two pulses P_n and P_{n+1} and is noted P_nP_{n+1}. Since $\forall n \in \mathbb{N}^*$, P_{n+1} is the same acoustic path as P_n plus one round trip in the spermaceti organ, the IPI is constant between each pair of pulses (P_n, P_{n+1}). The value of the IPI P_0P_1 is close to the one of P_1P_2. Instead of a difference linked to one round trip in the spermaceti (purple + black path of Figure 2.4), P_0P_1 is due to the black and the magenta paths, and the difference between the orange and cyan paths, which is proportional to the angle θ. The comparison between the value of P_0P_1 and that of P_1P_2 hence comes down to the difference between the purple and magenta paths, and the difference between the cyan and orange paths, which in practice leads to a difference of 10%.

Examples of recordings that we have made of sperm whale calls are shown in Figures 2.5, 2.6 and 2.7. The click recorded at 96 kHz in the first figure is an example of an IPI of 4.042 ms (probably a female since it translates to a body length between 9 and 11 meters) where θ is between $-90°$ and $90°$. The difference of acoustic paths can be seen in this click, as P_0P_1 is only 3.60 ms. The second pair of the figure shows a click recorded at 600 kHz with an IPI 3.27 ms. This click was recorded during the 2018 Mauritius recording session (see Section 6.1). Since it was recorded near the sperm whale and the click was part of a 2+5 coda, its subsequent pulses can be easily seen.

Since the IPI is related to the length of the spermaceti organ, and its value is stable [91, 92], it is possible to retrieve this length using the IPI. With the relationship between the length of the head and the length of the whole body established by [93], a relation between the length of a sperm whale L and its IPI can be found, assuming that the spermaceti organ is close in length to the head. Here are five formulas that have been proposed:

$$L = 0.76 + 4.64\,\mathrm{IPI} - 0.259\,\mathrm{IPI}^2, \tag{2.7}$$

$$L = 9.75 - 0.521\,\mathrm{SL} + 0.068\,\mathrm{SL}^2 + 0.057\,\mathrm{SL}^3, \tag{2.8}$$

$$L = 4.833 + 1.453\,\mathrm{IPI} - 0.001\,\mathrm{IPI}^2, \tag{2.9}$$

$$L = 17.120 - 2.189\,\mathrm{IPI} + 0.251\,\mathrm{IPI}^2, \tag{2.10}$$

$$L = 5.736 + 1.258\,\mathrm{IPI}. \tag{2.11}$$

For all the equations, the IPI needs to be in milliseconds to obtain a length L in meters. SL, the spermaceti organ length, also needs to be expressed in meters. The first equation

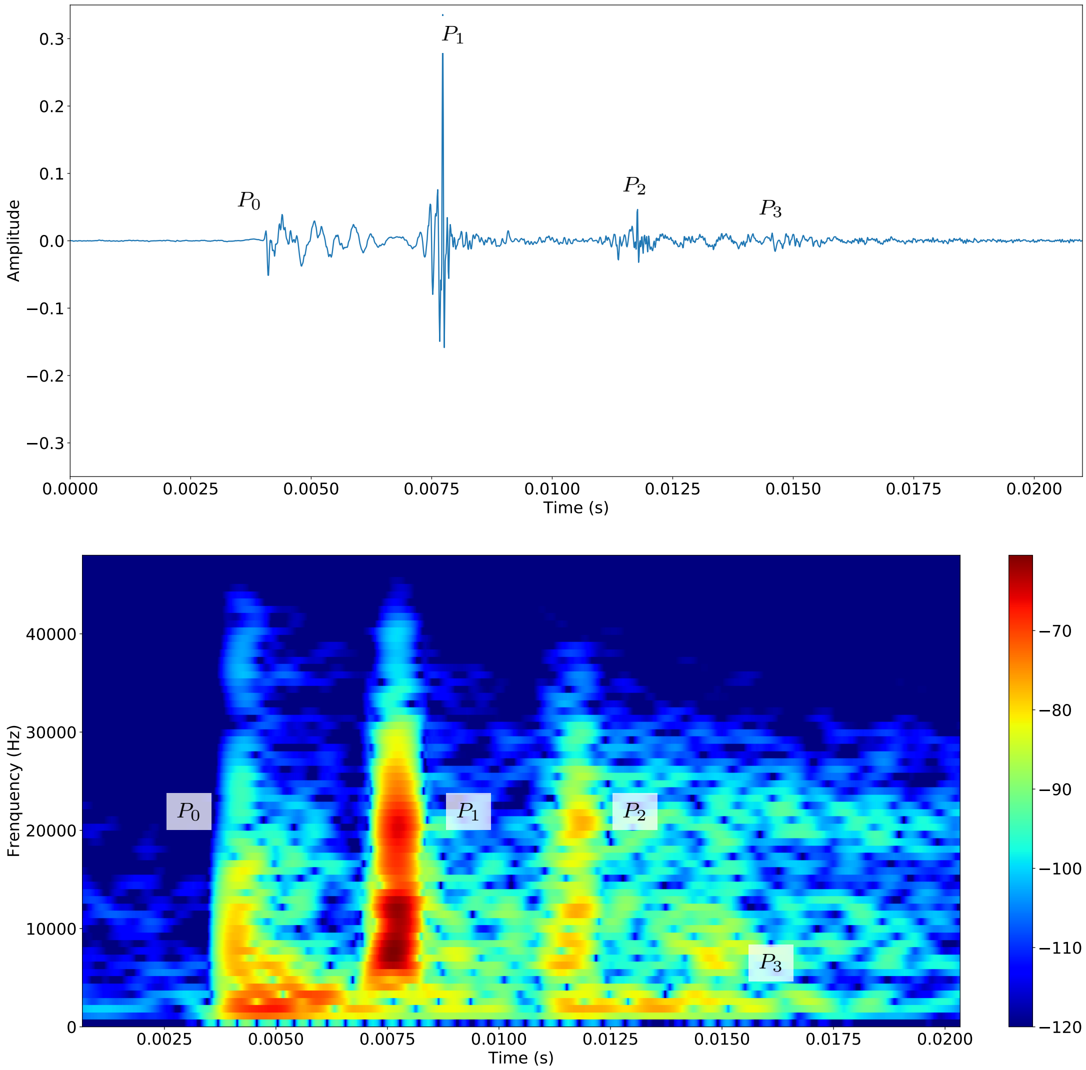


Figure 2.5: Sperm whale click and its spectrogram

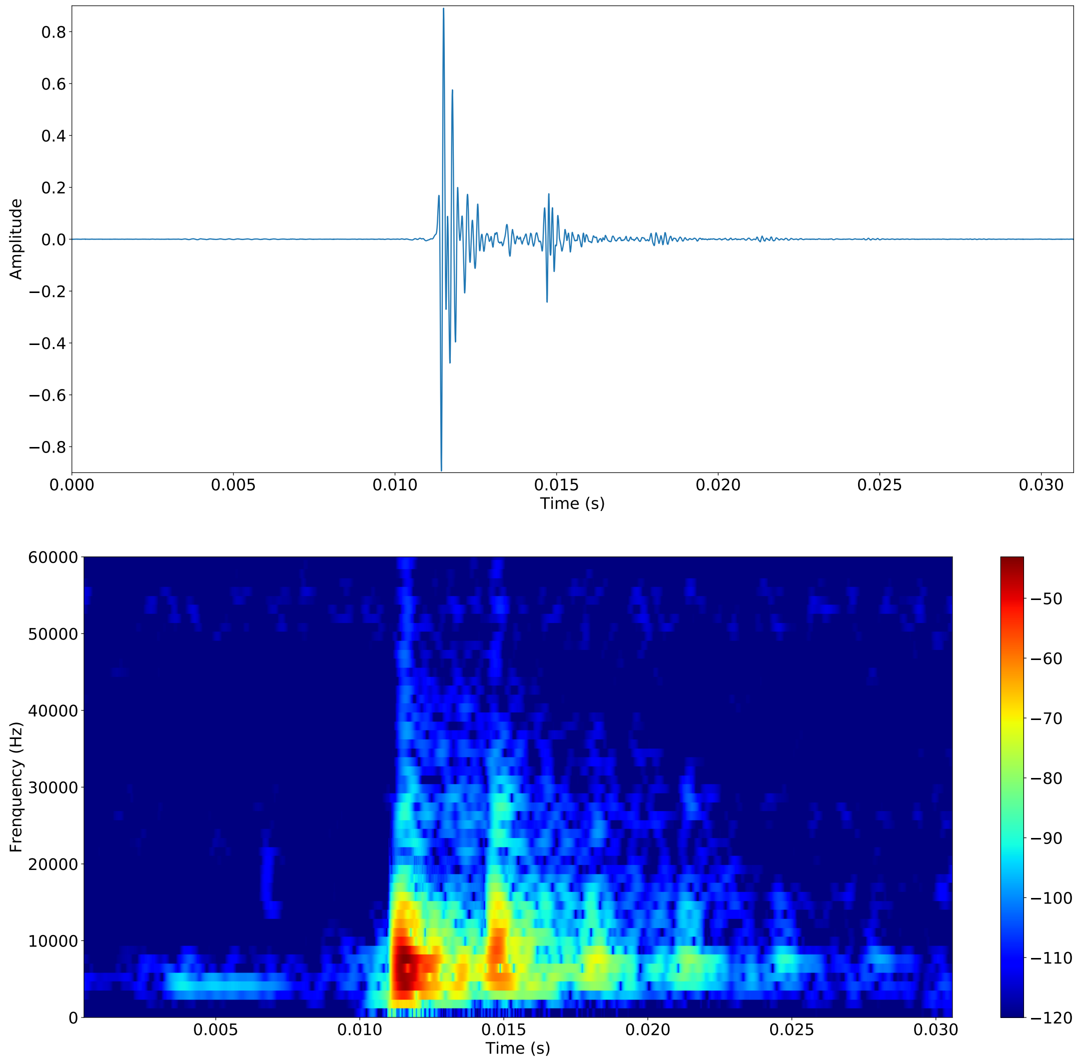


Figure 2.6: Sperm whale click and its spectrogram.

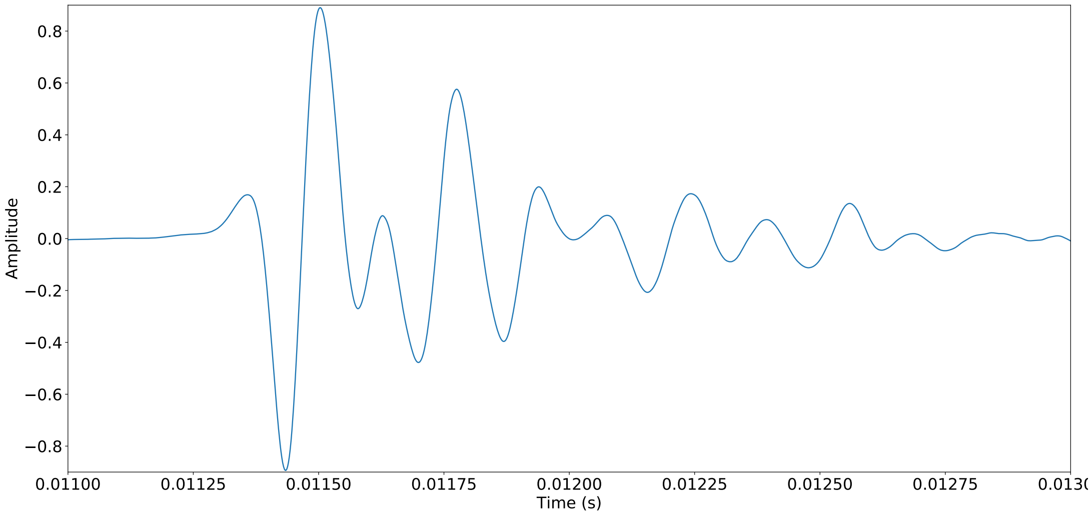


Figure 2.7: Zoom on the pulse of Figure 2.6

was established by [94] using all the relationships described above. However, as noted by [95], this equation has multiple flaws. The main one is that it is based on a speed of sound given in [2], which was two times faster than the correct speed of sound in the spermaceti oil. A corrected version of this equation should thus use half the IPI as an input. The second and third equations were derived by [95], where SL (spermaceti organ length) itself can be computed from the IPI using SL $= (c_{\text{spermaceti}}/2)$IPI where the speed of sound in the spermaceti organ is $c_{\text{spermaceti}}$. The equation (2.8) was obtained by fitting a polynomial to the total body length and the spermaceti length found in [96], while (2.9) was obtained by directly fitting a polynomial between the IPI and the total body length of Sri Lanka and Azores whales recorded and whose lengths have been photographically measured. Similarly the fourth equation was obtained by [97], but this time done on whales from New Zealand. Finally, the last equation was published in [98], also using whales from New Zealand.

As seen in Figure 2.8, which combines all the previous equations, the various estimations differ in their predictions. Multiple factors explain this disparity. First, some studies contain very few individuals; 5, 11 and 12 individuals for formulas (2.8), (2.9) and (2.10) respectively [98]. Also, most of the polynomial fittings were done on groups of sperm whales with similar sizes within each study, but different between studies, meaning that the equations extrapolate the value for the rest of the valid range of IPI. For example, all individuals but one in (2.9) were smaller than 12 m whereas (2.10) only contains males larger than 12 m. Equation (2.11) can be seen as the best model, as it was made on more individuals, and predicts values close to (2.9) and (2.10) over their respective fitting intervals.

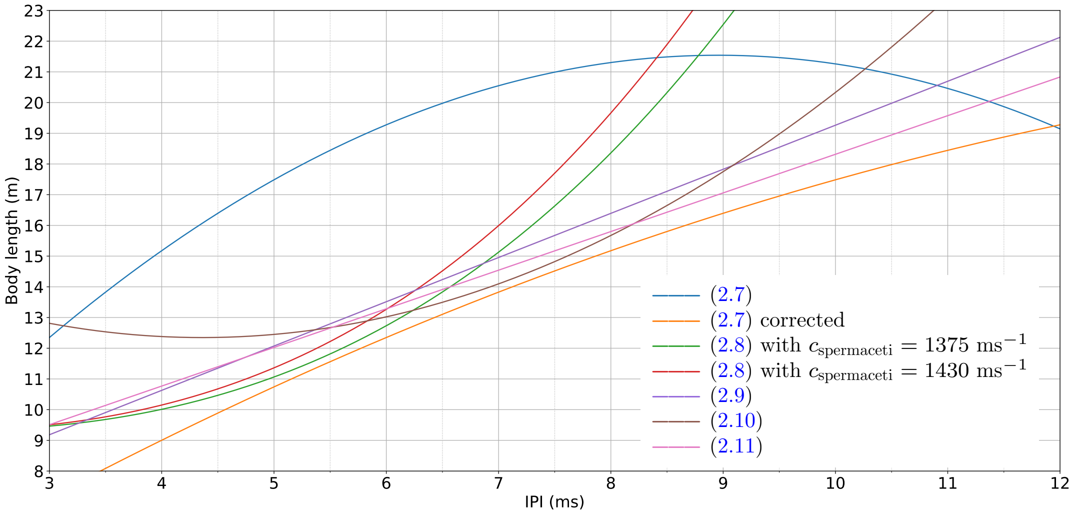


Figure 2.8: Comparison of the different total body length estimation equations.

IPI estimation

Multiple techniques exist to automatically estimate the value of the IPI. A first set of methods use the autocorrelation to perform this estimation. Since a click is made of a repetition of one pulse, the autocorrelation should show local maxima with a regular spacing corresponding to the IPI. The autocorrelation can be done directly on the signal, or on an estimation of the waveform envelope, that can be obtained using the Hilbert transform for example. The advantages of the waveform is that it will prevent the confusion between a periodic pattern due to the pulse repetition, and a periodic pattern due to the pulse itself. The other popular IPI estimation method is based on the cepstrum introduced in [99]. If a signal contains periodic structure, then its cepstrum will have a local maximum at its period. All these methods are often improved by averaging the results over multiple clicks coming from the same sperm whale. These different methods' performances are compared in [100]. In [101], another method called BIVIP is proposed. In this method, all possible candidates for P_0 and P_2 are extracted, and the pair (P_0, P_2) that gives similar values of the IPI P_0P_1 and P_1P_2 is kept. The joint distribution of these IPI obtained from these extractions done on multiple clicks is then used to estimate the IPI. Since the pairs made of the true P_0 and P_2 should be extracted more often than a random pulse due to background noise, the joint distribution should have a maximum at the right IPI. If multiple whales are present, and their IPI's are spread enough such that the joint distribution consists of multiple modes, then we propose that each IPI can be estimated without attributing a click to a sperm whale first. An alternative to this method is also possible where instead of one pair each candidate is kept. In this version the histogram of the joint distribution of every possible pair for each click is weighted regarding how relatively far the two IPI differ from each other. For example,

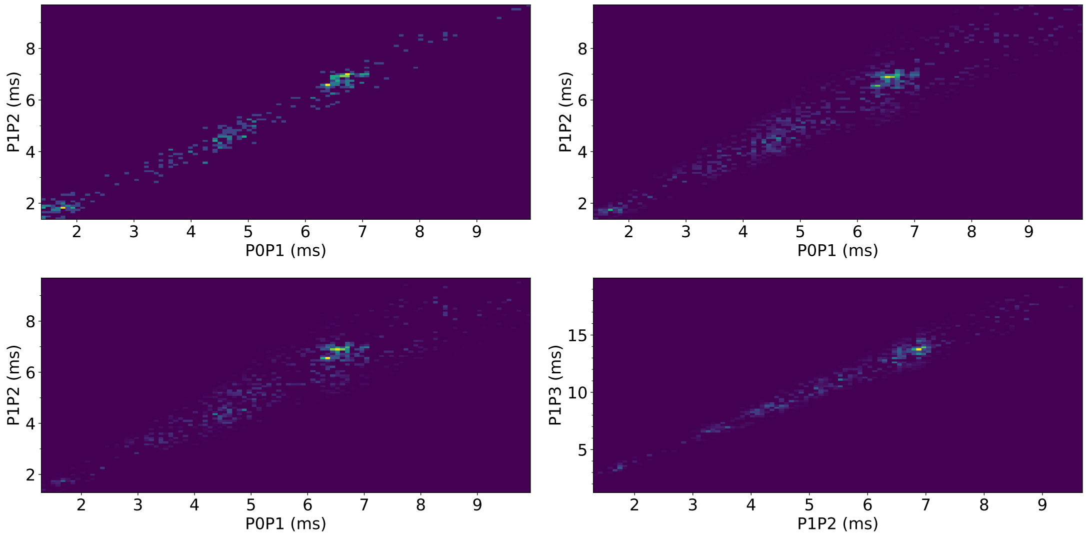


Figure 2.9: Joint distribution of IPI
Top left: BIVIP method. Top right: weighted BIVIP. Bottom: 3D weighted BIVIP3.

the following equation can be used for $(P_0P_1,\ P_1P2)$,

$$\mathcal{W}(P_0P_1, P_1P_2) = e^{\frac{-(P_1P_2 - P_0P_1)^2}{0.02\left(\frac{P_1P_2+P_0P_1}{2}\right)^2}}. \tag{2.12}$$

For $(P_1P_2,\ P_2P3)$, or $(P_1P_2,\ \frac{P_1P3}{2})$, a smaller standard deviation should be used, since the difference in these two IPI's should only arise from errors in the estimation of the pulse's position. When using multiple pairs of IPI, the multiple weight formulas can be multiplied together. Figure 2.9 shows an example of the last three methods applied on a 5 minute recording, containing a click with a 7 ms IPI.

2.5.3 Sound transmission in the nose of the sperm whale Physeter macrocephalus. A post mortem study

In 1997, [102] had the rare opportunity to experiment on a freshly dead sperm whale who stranded at Rømø. They managed to conduct acoustic experiments one hour after it had last shown signs of life. Usually they only arrived between eight and twenty two hours after the presumed time of death, which they say is too late to obtain any result due to excessive rotting. However, with only one hour of delay, the animal was still well enough conserved to simulate clicks in the way that [2] theorized, in which the first pulse (P_0) is generated, while the others are its multiple reflections produced by the two mirrors that are the distal and frontal sacs. Since they were in a rush to get to the dead sperm whale as quickly as possible, they were not able to take good equipment. They had a sampling rate of 48 kHz with some aliasing, and were only able to emit FM downsweep of 25 ms from 30 kHz to 10 kHz. This led them to believe that the experiment was not working, since they were not able to see a signal with multiple pulses on an oscilloscope (the measured IPI of 7.5 ms is a lot shorter than the 25 ms sweep ”pulses”). They however did at least three measurements with multiple takes. For each of them, they use a hydrophone next to the emitter. The measurements were made at the distal sac (m1), at the center of the front of the junk (m2), and at the top of the center of the spermaceti organ (m3). To compute the IPI, they did an autocorrelation on the waveform and took its envelope using the Hilbert transform, and smoothed it using a 4 kHz lowpass filter. They did not measure any reflection with m3, but did obtain the multipulse pattern with m1 and m2. The 7.5 ms IPI corresponds to the 15.1 m body length using the [95] algorithm. Their results also showed a broadening of P_2, which is not present in sperm whales recordings, and some other reflections between the usual pulses.

2.6 Acoustic focusing

In order to detect a target, the echo energy needs to be high enough, above a specific threshold. As shown in (2.2), the only parameter that can be of influence is the source. Instead of emitting a click uniformly in every direction, the energy emitted can be redirected in one direction through acoustic lenses and acoustic mirrors. With the energy concentrated in a beam, the detection range is increased, but limited to a small solid angle $8° - 10°$ in front of the animal [103], meaning that the animal will need to adopt a foraging behaviour in order to inspect the whole volume of water. The odontoncetes evolve to have a sac of oil in their head, called the melon, to do this beamforming. As an example, a model of this effect can be seen in the *Kogia* in [104].

Sperm whales were thought not to beamform [105], until it was disproven in [106] and further analysed in [107] and [108]. Considering the size and the complexity of the sperm whale sonar, it would have been surprising that its sonar possesses no particular directionality. In [107], they predicted a directivity index of 26.7 dB for the P_1 pulse, while P_0 only seems to have a directivity index of 7.4 dB. This difference of directivity index shows the importance of the complex acoustic path of the sperm whale’s sonar in terms of added directivity.

2.7 Conclusion

Unlike other click emitting cetaceans, sperm whales emit multi-pulsed clicks, spaced by 2 ms to 10 ms depending on their length. This multi-pulsed click is explained by the bent horn model that models its complex sonar structure. With this sonar they communicate to form complex societies and they echolocate to find their prey.

The contents of this chapter form the basis for the rest of the thesis, where the characteristics of the clicks will be used as a basis for the modelling of sound generation, propagation and reception. The models we will describe next, cover a range of approaches, from signal processing to machine learning, and finally direct and inverse propagation models whose objective is to reproduce these sounds as faithfully as possible. We hope that this will provide better understanding of sperm whale behaviour and help in the monitoring and surveillance of endangered populations.

Part I

Signal Processing for Passive Bioacoustics

In this part, that consists of four chapters, we study the recorded bioacoustic signals and develop new methods for analyzing their characteristics. These methods cover both near- and far field recordings.

Chapter 3

Signal Processing Techniques

As presented in the previous section, acoustics is a vast subject when it comes to sperm whales. In order to study sperm whale acoustics, their signals need to be recorded, processed, and analysed. This chapter presents the signal processing techniques that are used throughout this thesis. The main topics tackled here are the click detection and the TDOA estimation used for the source localisation, which itself is developed and used in Section 6.2.

3.1 Click Detection

Bioacoustics signals are often unannotated, or weakly labeled—e.g. a large window of time where the animal should be present, but might not be heard continuously,—which renders them not readily usable with a neural network or other data analysis techniques. These signals usually come from long recording sessions, meaning that an expert labeling of these data would be extremely costly, both in terms of salary and time. In the case of sperm whale clicks, one may use a click detector to automatically extract the precise time positions of each click. As this detector will be the first stage of a data analysis pipeline, its recall should be high, even though a high recall usually comes with many false positive detections.

3.1.1 Teager-Kaiser approach

The most used detector in this thesis is based on a Teager-Kaiser filter [109, 110]. When we are aiming for one particular species, or a number of animals that emit clicks in the same frequency range, we start by cross-correlating the signal with one period of a sinusoid which acts as a band-pass filter. The sinusoid frequency should be around the centroid frequency of the aforementioned clicks. This pre-filter is followed by a TK filter. The TK filter can be seen as a Dirac detector. For the n^{th} sample x_n of a signal, the TK filter will return the value tk_n according to

$$tk_n = x_n^2 - x_{n+1}x_{n-1}. \tag{3.1}$$

The last step is the extraction of local maxima in windows that are slightly larger than the size of the studied clicks, and shorter than the minimum ICI. In the case of sperm whales, this would mean a window of 20 ms, twice the maximum possible IPI. All the maxima, and most importantly their position, that are above a threshold set with a method that will fit the problem, are retrieved. All the steps of this detector have complexity in $\mathcal{O}(n)$, where n is the number of samples of the signal studied, except for the first cross-correlation which has a complexity of $\mathcal{O}(mn)$ where m is the number of samples of the sinusoidal wave (usually m is less than 10). These different steps are shown in Figure 3.1.

3.1.2 Setting the threshold

Before finding the right threshold value, the maxima are converted to dB. Energy distributions are usually log-normally distributed (or normally distributed in dB), yet these maxima reflect the local energy of the signal. The signal is composed of three types of maxima. Ones that belong to animal clicks, others where nothing was emitted, belonging to the background noises, and the last ones that belong to neither (e.g. boat noises). The bottom scatter plot of Figure 3.1 shows an example where the three types of maxima can be seen. Here the bottom dense line is made of points belonging to the noise distribution. The sparse curve is made of sperm whale maxima. The remaining points that do not belong to any of the two described distributions are the boat noise points. The aim of the threshold is to filter out any clicks that belong to the noise distribution. The third type of maxima cannot be simply filtered out with this method, and should be filtered in another step further down the data analysis pipeline.

Constant threshold

In situations where the noise level is constant, or approximately piece-wise constant, the threshold value is set as a constant, chosen from empirical results. This prevents any difficulty that could arise in estimating the background noise during periods of dense click emission.

Gaussian fitting

We then convert the maxima's values into dB. The maxima usually form two distributions: one that emanates from the actual animal clicks and non-animal clicks, and one that emanates from the maxima that are between clicks, which are maxima created by white noise. We thus filter out the noise by fitting two Gaussians on the mixed distribution, and only keep values that are above three times the standard deviation [111] of the Gaussian with the smallest mean, as it should be the Gaussian that describes the distribution of the noisy maxima. However this technique possesses some flaws. First of all, the noise level is assumed to be constant during each period where the two Gaussian distributions are fitted. A sudden change in noise level (e.g. an antenna recording in air then in water at the start of a dive), will usually lead to fitting the two Gaussians

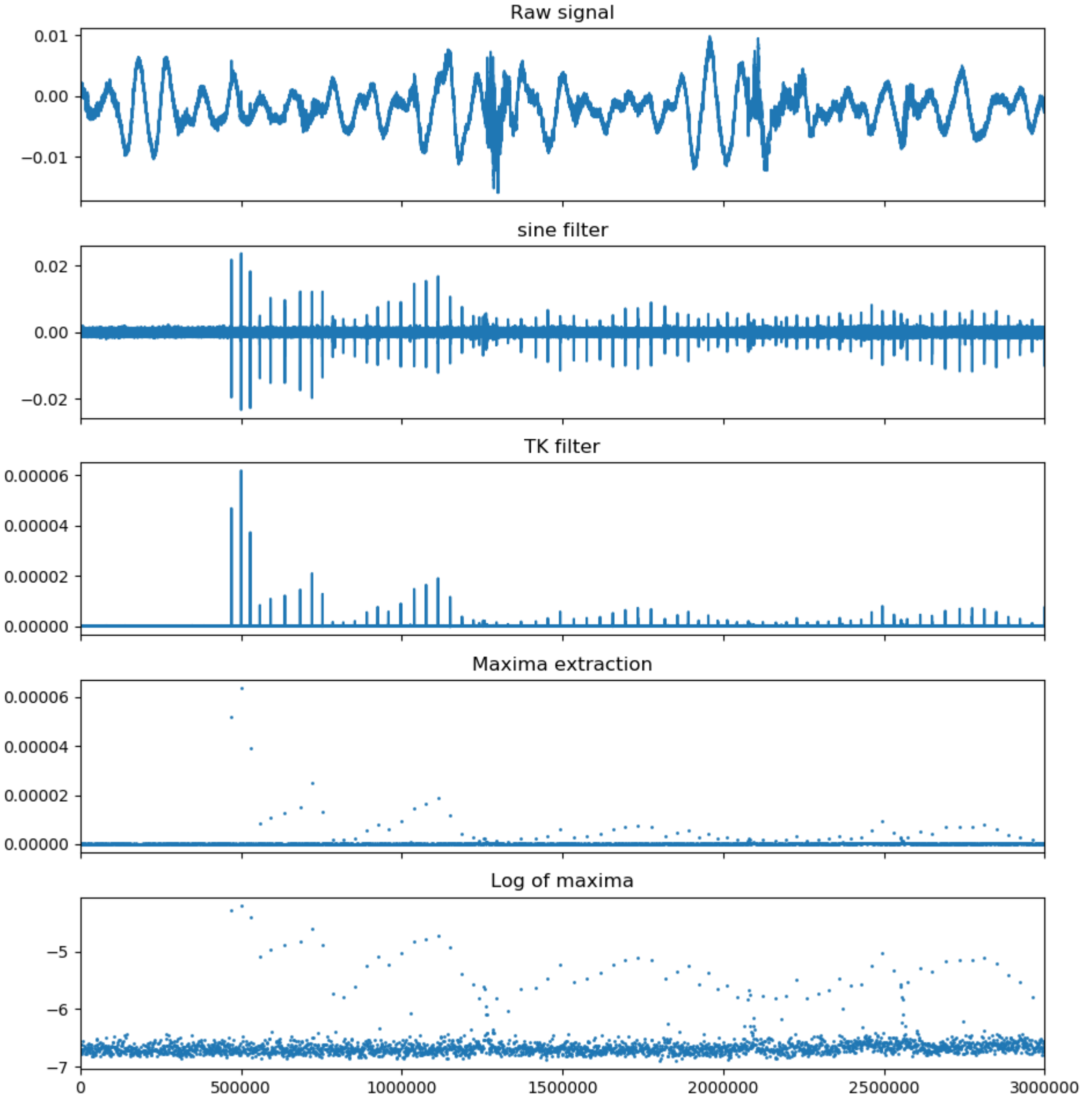


Figure 3.1: Example of each step of the Teager-Kaiser approach.

on these two noise distributions. This will cause part of the second noise distribution to be above the threshold, flooding the detection with false positives. This issue can be addressed by shortening the time window used to compute the Gaussian distributions. However, a too short time window will not incorporate enough points to evaluate precisely the distributions.

A second issue appears if most of the maxima do not belong to the noisy class. This may happen when the animal emits clicks with a small ICI. In this case, the clicks will be the ones that the two Gaussian distributions will fit, thus leading to a majority of clicks being classified as noise. In order to fix this, one could try to reduce the time window used for the extraction of the local maxima, so that some noisy maxima could be found between two clicks. But this time window should not be reduced too much, as the detector will then be able to detect unwanted maxima, such as other pulses of the same click.

What is described above, is the approach in which each Gaussian estimation is done on non-overlapping windows. A more sophisticated method could be to use a rolling Gaussian estimation.

Rolling median

In the case that noisy maxima are rare, an alternate method can be used. One can compute a rolling median. In this case, the value obtained from the median should be close to the mean of the noisy distribution. The threshold is then set to be the rolling median plus another value. This latter value can be a constant, if the noise distribution has a constant width. Otherwise, a rolling standard deviation can be computed, if the density of noisy maxima is large enough so that the standard deviation of the global distribution (distribution that contains the three classes) is the same as the one of the noise distribution. However the rolling standard deviation will not be able to cope with sudden changes in the noise level, whereas a constant value will, which gives in this particular but not unusual case, an advantage to this method against the Gaussian estimation.

Rolling median estimation

In the event that a fast threshold needs to be computed, and that the noise level is quite stable without any sudden jumps, an estimation of the rolling median can be used.

This can be computed by the following code snippet.

```
float median = 0.0f;
float average = 0.0f;
// for each sample
{
    average += ( sample - average ) * 0.1f; // rough running average.
    median  += _copysign( average * 0.01, sample - median );
}
```

3.1.3 Deep learning

An alternative to the TK filtered approach that was described above to detect sperm whale clicks, is to use a neural network. Please refer to Part II for a description of some of the machine learning approaches used in bioacoustics. As an example of a neural network detector, the architecture used in Chapter 7 can be applied with two classes (noises and sperm whale) or one class (sperm whale), and was used to detect the clicks that are present in the figure of the next chapter.

3.2 TDOA computation

When a sound emanates from a point source, it will propagate through space at the speed of sound of the medium. The sound will then reach various points in space at various times, the time of arrival (TOA). A recording station with multiple recording sensors, hydrophones in the case of underwater acoustics, will thus record the sound delay differently on each channel. As the TOA depends on the speed of sound and the distance between the source and the hydrophones, some information about the source position can be retrieved. However, it is quite unlikely to have the time at which the signal was emitted in passive acoustics. The time difference of arrival (TDOA), which is just the difference between two TOAs, can be used instead of the TOA to retrieve the same kind of information about the source position. Let us note T_i the TOA of hydrophone H_i, and $\tau_{ij} = T_i - T_j$ the TDOA of hydrophone H_i relative to hydrophone H_j. For N hydrophones, there are $\binom{N}{2}$ possible pairs of TDOA, but only $N-1$ independent TDOA, limited to a maximum of three if the equations of the antenna shape are taken into account. If we choose the $\{\tau_{i1}\}_{i\in[\![2,N]\!]}$ to be the independent TDOA, then the other TDOA could be constructed as follows:

$$\forall (i,j) \in [\![1,N]\!]^2, \quad \begin{aligned} \tau_{ij} &= T_i - T_j \\ &= T_i - T_1 + T_1 - T_j \\ &= \tau_{i1} - \tau_{j1}. \end{aligned} \tag{3.2}$$

Each independent TDOA can be used to obtain one more information about the position of the source. With two hydrophones (one TDOA), the azimuth of the source is usually computed, with the addition of the elevation with three hydrophones, and finally the distance with four hydrophones. Any other additional hydrophones will make the problem over-determined. Even if four hydrophones are theoretically enough to compute the exact 3D position of the source, the distance of the source is too unreliable with only four hydrophones.

3.2.1 Cross correlation

A way to compute the TDOA of a stereo signal is to compute the correlation between the two channels. Since the maximum of an autocorrelation is at 0, the maximum of correlation between a signal and a lagged version of itself (with or without attenuation a), will be at the lag. This is because

$$(s * a\tau_{t_0} s)(t) = a(s * s)(t - t_0), \tag{3.3}$$

where s is the source sound, a is the attenuation, t_0 is the lag, τ_{t_0} is a translation of t_0, and $*$ is the correlation. This technique works well if there is only one source at a time, if the source has a sharp autocorrelation around zero and a maximum that can be discriminated from other extrema, and if the noise is uncorrelated between the two channels.

A cross correlation can also be computed using the Fourier transform, since the Fourier transform of a convolution is the multiplication of the Fourier transforms. Let us call x_1 and x_2 the signals of our two channels, $.^*$ the conjugation operator, and $\mathfrak{F}$ the Fourier transform. We have

$$x_1 * x_2 = \mathfrak{F}^{-1}\left(\mathfrak{F}(x_1 * x_2)\right) = \mathfrak{F}^{-1}\left(\mathfrak{F}(x_1) \times \mathfrak{F}(x_2)^*\right). \tag{3.4}$$

This trick stays valid for discrete signals, as a discrete signal is a continuous signal convolved with a Dirac comb. FFTs can thus be used to speed up the cross correlation. In practice, FFT correlation and standard correlation differ, because the TDOA computation is done on a chunk of the signal. The FFT will act as if the chunk is a period of a periodic signal, whereas the standard correlation will either pad this chunk with zeros, or used the needed neighbouring samples.

3.2.2 Generalised cross correlation

The standard cross correlation is however prone to error in low SNR, or reverberated environments. One way to improve the cross correlation is to use the generalised cross correlation (GCC). GCC is an ensemble of techniques that add a weight Ψ to equation (3.4), which becomes

$$x_1 *_\Psi x_2 = \mathfrak{F}^{-1}\left(\Psi\mathfrak{F}(x_1) \times \mathfrak{F}(x_2)^*\right). \tag{3.5}$$

This weight Ψ can be seen as a way to filter out any noise, echoes or artefacts that could deteriorate the cross correlation. The simplest Ψ, apart from the identity for which the GCC equates to a standard cross correlation, is a bandpass filter. The frequency bands which are supposed to contain the signal will have a weight of 1, whereas the others will have a weight of 0.

GCC weights may be dynamic, meaning that they will adapt to what they estimate the signal and the noise to be, in order to give more weight to a region with high SNR. Let us denote the spectrum of a signal in upper case (e.g. the spectrum of x will be X). The channel signals are still named x, the source is named s, and the noises are named n.

ROTH and SCOT

The first two filters used to weight frequency bands using an approximation of the SNR are named ROTH [112] and SCOT [113]. SCOT, or Smoothed Coherence Factor, is a declination of the ROTH weight which allows the noises to be different for the two channels.

$$\Psi_{\text{ROTH}} = \frac{1}{\|X_1\|^2} \quad \text{and} \quad \Psi_{\text{SCOT}} = \frac{1}{\sqrt{\|X_1\|^2 \|X_2\|^2}}. \tag{3.6}$$

PHAT

The next method is called PHAT [114] and is used with signals recorded in high reverberation environments. It normalizes the spectrum in order to use only the phase to compute the correlation,

$$\Psi_{\text{PHAT}} = \frac{1}{\|X_1 X_2^*\|}. \tag{3.7}$$

However, PHAT does not work well in a low SNR environment.

Eckart

The Eckart filter [114] was made to maximize the deflection criterion, which is defined by the ratio of the change in mean correlation output due to signal present compared to the standard deviation of correlation output due to noise alone,

$$\Psi_{\text{ECKART}} = \frac{\|S\|^2}{\|N_1\|^2 \|N_2\|^2}. \tag{3.8}$$

With the hypothesis of an uncorrelated noise, the following simplification can be made,

$$\begin{aligned} x_1 &= s + n_1, \\ \|X_1\|^2 &= X_1 X_1^* = \|S\|^2 + \|N_1\|^2 + S N_1^* = \|S\|^2 + \|N_1\|^2, \\ \|X_1(\omega) X_2^*(\omega)\| &= \|S(\omega) S^*(\omega) \exp^{i*\omega*t_0} + N_1(\omega) S^*(\omega) + S(\omega) N_2^*(\omega) + N_1(\omega) N_2^*(\omega)\| \\ &= \|S(\omega)\|^2, \end{aligned} \tag{3.9}$$

where ω is the angular frequency. The Eckart filler then becomes

$$\Psi_{\text{ECKART}} = \frac{\|X_1 X_2^*\|}{(|X_1\|^2 - \|X_1 X_2^*\|)(|X_2\|^2 - \|X_1 X_2^*\|)}. \tag{3.10}$$

Unlike PHAT, the Eckart filter will have a weight of zero if there is no signal. Eckart also has the advantages of SCOT, such as the preference toward high SNR frequency bands.

3.2.3 steered response power

The steered response power (SRP) is an ensemble of techniques that find the TDOA by optimising the power of a beamformer [115, 116, 117, 118], such as the delay and sum beamformer. This is a simple beamformer that will translate in time each channel, and then compute the mean of these delayed signals. The mean of uncorrelated signals will tend to zero by increasing the number of channels, whereas the mean of multiple copies of the same signal will be this signal, which is what happens when the delay and sum beamformer uses the TDOA that has delay parameters. For this beamformer, the power will be

$$\mathbf{P}(\mathbf{x}, \tau_{21}, \tau_{31}, ..., \tau_{n1}) = \int_{-\infty}^{\infty} \left(\sum_{i=1}^{n} x_i(t - \tau_{i1}) \right)^2 dt \tag{3.11}$$

$$= \int_{-\infty}^{\infty} \sum_{i=1}^{n} \sum_{j=1}^{n} x_i(t - \tau_{i1}) x_j(t - \tau_{j1}) dt \tag{3.12}$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} \int_{-\infty}^{\infty} x_i(t - \tau_{i1}) x_j(t - \tau_{j1}) dt \tag{3.13}$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} (x_i \circledast x_j)(\tau_{i1} - \tau_{j1}), \tag{3.14}$$

where $\circledast$ denotes cross-correlation.

The power of the beamformer is the sum of the cross correlation of all the possible pairs of hydrophones, including self pairs containing twice the same hydrophones. Since $(x_i \circledast x_i)(\tau_{i1} - \tau_{j1}) = (x_i \circledast x_i)(0)$ is a constant and we have $(x_i \circledast x_j)(\tau_{i1} - \tau_{j1}) = (x_j \circledast x_i)(\tau_{j1} - \tau_{i1})$, the SRP optimization can be reduced to optimizing

$$\mathbf{SRP}(\mathbf{x}, \tau_{21}, \tau_{31}, ..., \tau_{n1}) = \sum_{i=2}^{n} \sum_{j=1}^{i-1} (x_i \circledast x_j)(\tau_{i1} - \tau_{j1}). \tag{3.15}$$

3.2.4 Geometric SRP

We propose an approach similar to SRP, but in our case, we process multiplications instead of additions. We thus improve the TDoA computation by combining multiple cross-channel information. For N channels, the method starts by computing the cross-correlation c_{ij} between all the $\binom{N}{2}$ possible pairs of signals (x_i, x_j). Then, the method subtracts the minimum of c_{ij} from it. With N channels, there are only $N-1$ independent TDOA (for instance $\{\tau_{i1}\}_{i \in [\![2,N]\!]}$) because of the relation $\tau_{ij} = \tau_{ik} + \tau_{kj}$. In our setup there are in fact three independent TDOA. If the sound is recorded on all channels the maximum of each cross-correlation c_{ij} is at τ_{ij}. However, this does not hold if the recordings are noisy. That is why both methods try to maximize the values of all the cross correlation regarding a set of independent TDOA $\{\tau_{i1}\}_{i\in[\![2,N]\!]}$.

In short, our method finds the TDOA by computing

$$\tau_{21}, \tau_{31}, ..., \tau_{n1} = \arg \max_{\hat{\tau}_{21},\hat{\tau}_{31},...,\hat{\tau}_{n1}} \prod_{i=2}^{n} \prod_{j=1}^{i-1} (x_i \circledast x_j)(\hat{\tau}_{i1} - \hat{\tau}_{j1}) - \min\left(x_i \circledast x_j\right), \quad (3.16)$$

where $\circledast$ denotes cross-correlation.

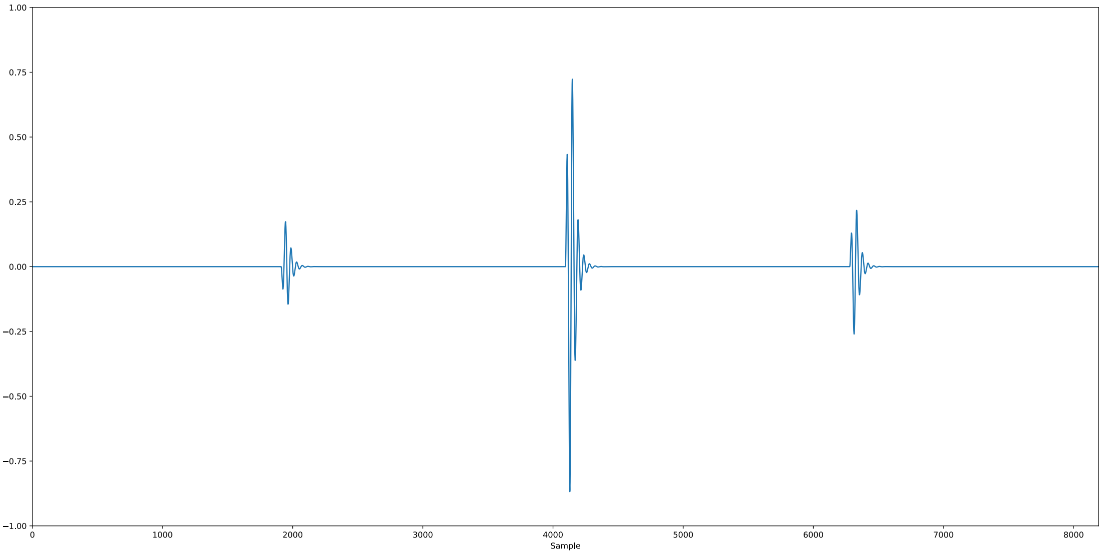


Figure 3.2: Synthetic click used for the SNR - TDoA simulation illustrated in Figure 3.3.

In Figure 3.3 we show the results of both methods to compute τ_{21} on a synthetic signal (Figure 3.2) in three or four channels with various SNR dB levels of white noise. We also compared them to τ_{21} obtained with the cross-correlation between x_1 and x_2. This figure shows the absolute error in bins of 2048 examples for each level of noise. The TDoA search was done between -128 and +128 samples. The source signal was synthesized using an adaptation of a dolphin click generation model [119], which is used here to generate the pulse P1 of the sperm whale click,

$$U(t) = U_0 \text{ ReLU}\left(\frac{a^{\text{ReLU}(t/T-m)} - a^{t/T}}{1 - a^m}\right), \quad (3.17)$$

where ReLU is the Rectified Linear Unit function, m is the number of periods to reach the maximum of amplitude, a is the reflectance factor, and T is the period of the carrier

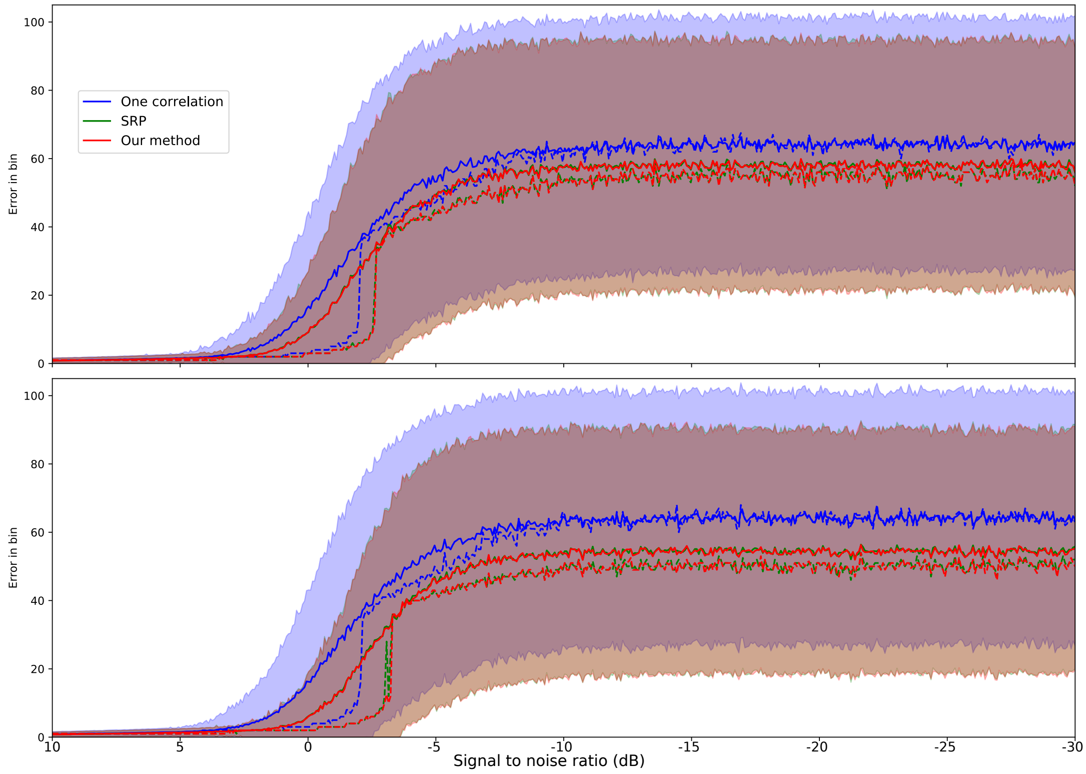


Figure 3.3: TDoA error with 3 channels (*top*) and TDoA error with 4 channels (*bottom*) computed on simulated clicks at various SNR.
The full lines are the mean error, the dashed lines are the median of the error, and the filled region corresponds to one standard deviation across the mean.

waveform. The dolphin-specific parameters, $m = 5$ and $a = 0.84$, have been changed to $m = 1$ and $a = 0.25$ to approximate sperm whale pulses. A carrier frequency of 12.5 kHz was used. The denominator has also been changed from $1 - a$ to $1 - a^m$ in order to have a maximum magnitude of U_0. The equation (3.17) describes the envelope of one pulse. To better mimic sperm whale clicks [120], two other pulses of relative amplitudes -0.2 and 0.3 were added with an Inter Pulse Interval (IPI) of 4 ms.

3.2.5 Increasing the resolution

The TDOA obtained with the previous methods, whether it was with FFT correlation or standard cross correlation, will not be more precise than the sampling period. This is simply due to the TDOA being expressed as the offset in number of samples when these methods are used on discrete signals. It is however possible to interpolate the objective function as a function of the independent TDOA. An alternative method is to use a low pass filter on the TDOA estimation along time can also be done. However it requires having a way of clustering each click by track, otherwise, this will mix the

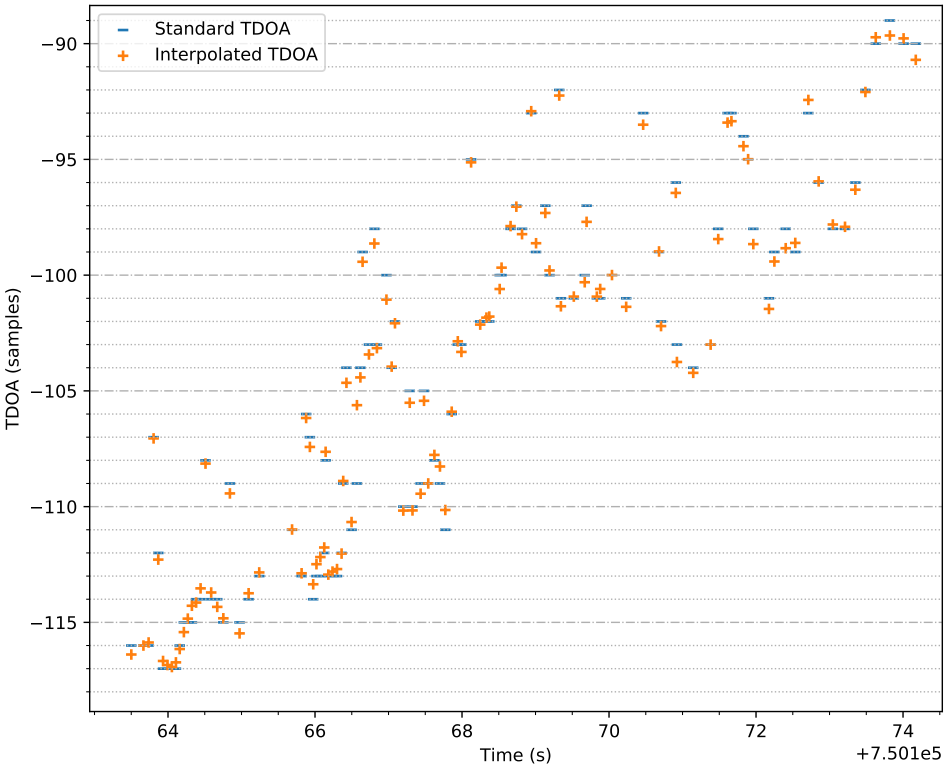


Figure 3.4: Result of interpolating the objective function.

unrelated TDOA of multiple animals together. Another issue with the filtering of the TDOA compared to the interpolation on the objective function is that it will not recover subtle movements hidden in a monotone track. An example of interpolating the objective function is shown in Figure 3.4, where the blue lines are the standard TDOA, meaning the TDOA limited by the sampling rate, and the orange pluses are the TDOA obtained by interpolating the objective function. On this figure, whereas the standard TDOA shows various plateaus, the interpolated TDOA managed to reconstruct a continuous curve. Since the TDOA of each click are estimated independently, this recovery of the curve can only be obtained by an estimation of the TDOA, as an erroneous estimation would produce random noises.

Figures 3.5 and 3.6 are TDOA curves made from 192 000 kHz 4 channel recording made near the Estagnol beach (east of Hyères). The antenna was placed on the sea floor at 5 m depth. Unlike Figure 3.4, which was done on transients (sperm whale clicks), these two figures were done on stationary sound signals. Figure 3.5 showcases a passage where a swimmer swam above the antenna, while holding a small speaker which was emitting

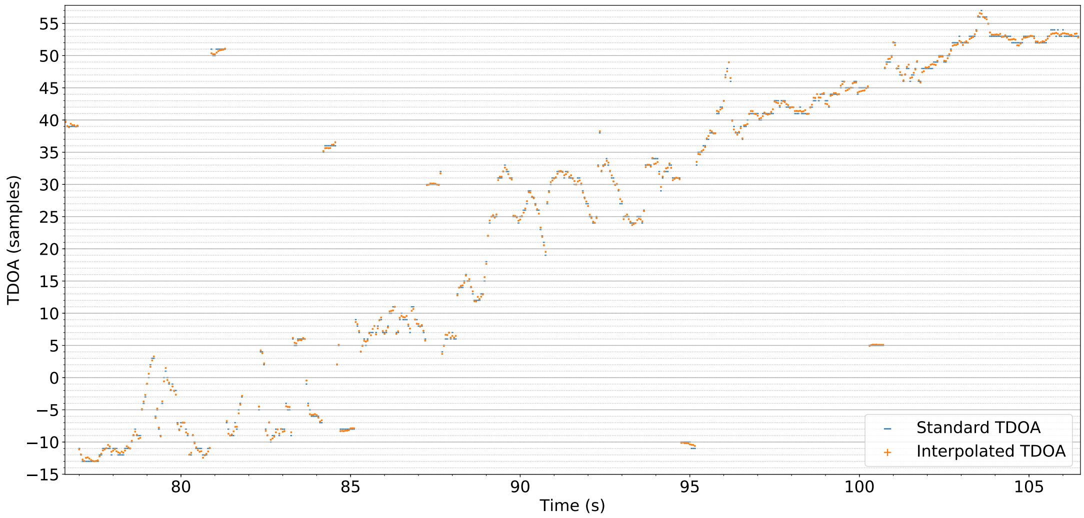


Figure 3.5: Result of interpolating the objective function for swimmer with speaker emitting killer whale vocalisations.

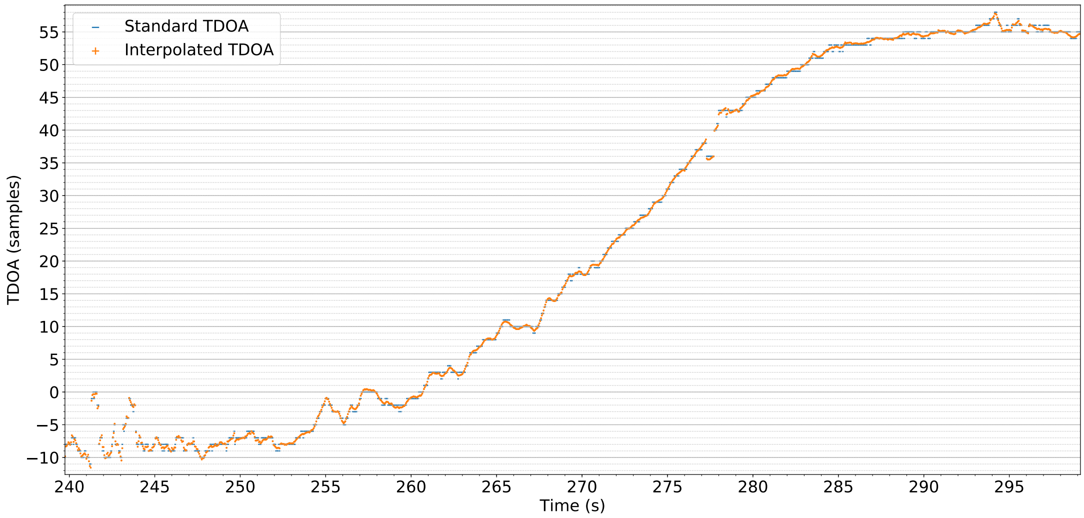


Figure 3.6: Result of interpolating the objective function for a boat signal.

killer whale vocalisations in order to test the antenna before its future deployment in Vancouver. Unfortunately the speaker was not loud enough, so the vocalizations blurred into the background noise during the other tests that were performed further away from the antenna. Figure 3.6 shows the boat TDOA while it was moving to recover the diver, and later the antenna. The discontinuity around 278 s was due to a loud click, which saturated some channels.

3.3 Ellipsoid TDOA Method

Even if 4 hydrophones are theoretically enough to retrieve the distance, the computation is not robust and any small measurement error can lead to large variations in the distance estimation. The sound speed and the inter hydrophone distances are crucial parameters that will influence the distance estimation. Using the ellipsoid TDOA method, the position of each hydrophone can be retrieved from the ellipsoid distribution generated by the measurements. This method will not produce a distance between each hydrophone in terms of length, as it will instead produce the maximum TDOA possible between each hydrophone. The distance between the hydrophones could then be obtained by multiplying by the speed of sound. By defining the distance between hydrophones this way, the estimation of the distance will then depend linearly on the speed of sound, removing the need for a precise value for the speed of sound.

The ellipsoid TDOA method was first proposed by [121] on a 2D version of the problem. The 3D approach was later addressed by [122].

3.3.1 An intuition of the underlying mechanism

Before explaining the 3D case, three specific examples of the 2D case can help to understand why the TDOA form an ellipsoid, which in the 2D case is an ellipse. These are illustrated in Figure 3.7. Let us name the three hydrophones H_0, H_1 and H_2, and the TDOA between two hydrophones i and j, τ_{ij}. Let us suppose that the distance d between H_0 and H_1 is the same as the distance between H_0 and H_2. If H_1 is at the same position as H_2, then τ_{10} is equal to τ_{20}. If H_1 is at the opposite of H_2 regarding H_0, then τ_{10} is equal to $-\tau_{20}$. The last simple case is when H_1H_0 is perpendicular to H_2H_0. If a plane wave comes from an angle α regarding $\overrightarrow{H_0H_1}$ then it will have an angle $\frac{\pi}{2} - \alpha$ regarding $\overrightarrow{H_0H_2}$, and thus $\tau_{10} = d\cos\alpha$ and $\tau_{20} = d\sin\alpha$, creating a circle.

3.3.2 The full 3D case for a four-hydrophone antenna

For this method to work, two hypotheses need to be made. The first one is that the sound speed needs to be constant in a convex hull around the antenna. This will not be verified for long recording sessions, where the water temperature or the salinity might change, hence a change in the speed of sound. The second hypothesis that needs to be made is the far field hypothesis. As can be seen in Figure 3.8, the TDOA joint distribution no longer forms an ellipse when the source is getting closer to the antenna.

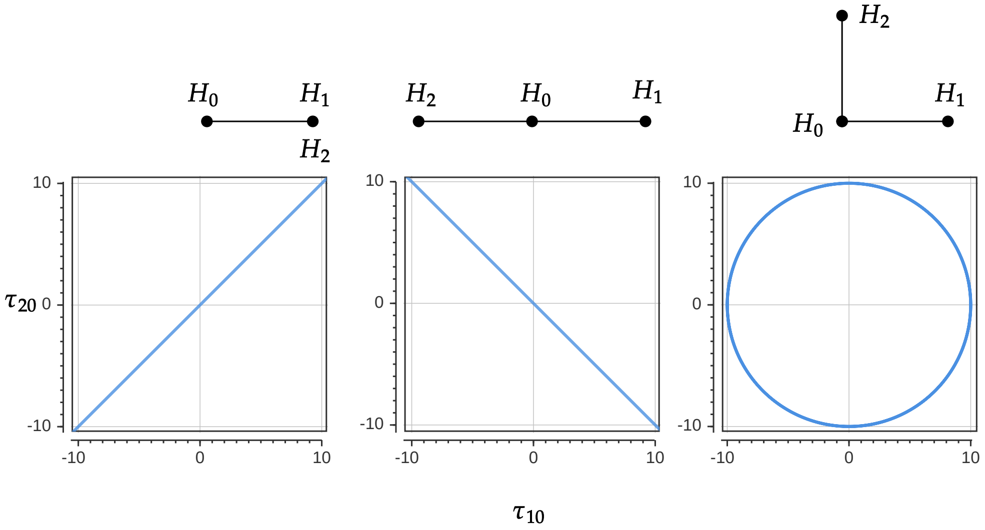


Figure 3.7: The three examples of a 2D antenna and their TDOA distribution.

In Figure 3.9, the difference of ellipse distance between the far field ellipse and the near field ellipse for each angle depending on the position of the source is displayed. Since the distances are normalised, a smaller antenna of the order of 1 m will only have an error of 0.01 for target at 10 m from its center, while a larger antenna of the order of 5 m will only achieve the same error for target at 50 m. The second hypothesis will thus be verified for closer targets when using smaller antenna. However, this also means that the distance of the target will be harder to estimate, since estimating the distance is equivalent to measuring this error. Finally the TDOA used for this method needs to be obtained by an algorithm that does not use the antenna shape as a constraint, as it will provide spurious TDOAs since the antenna shape constraint will force the estimated TDOA to be found on the ellipse of this given shape.

As before, let us call the four hydrophones H_0 to H_3, the TDOA between H_i and H_j, τ_{ij}, the distance between H_i and H_j, d_{ij}, and the angle between $\overrightarrow{H_0H_i}$ and $\overrightarrow{H_0H_j}$, α_{ij}. Let the coordinates of hydrophone H_i be (x_i, y_i, z_i) in a base with H_0 as its origin, the Z axis co-linear to $\overrightarrow{H_0H_1}$, and H_2 in the XZ plane, with x_2 being positive. Finally let us have the subscript of the source named s, and express its position in spherical coordinates (ρ, ϕ, ψ), as shown in Figure 3.10, with a ρ close to infinity. We thus have,

$$\tau_{10} = \frac{z_1}{c}\cos\psi, \tag{3.18}$$

$$\tau_{20} = \frac{x_2}{c}\cos\phi\sin\psi + \frac{z_2}{c}\cos\psi, \tag{3.19}$$

$$\tau_{30} = \frac{x_3}{c}\cos\phi\sin\psi + \frac{y_3}{c}\sin\phi\sin\psi + \frac{z_3}{c}\cos\psi. \tag{3.20}$$

If y_3 is negative, we can replace ϕ by $-\phi$. Then (3.18) and (3.19) will stay the same

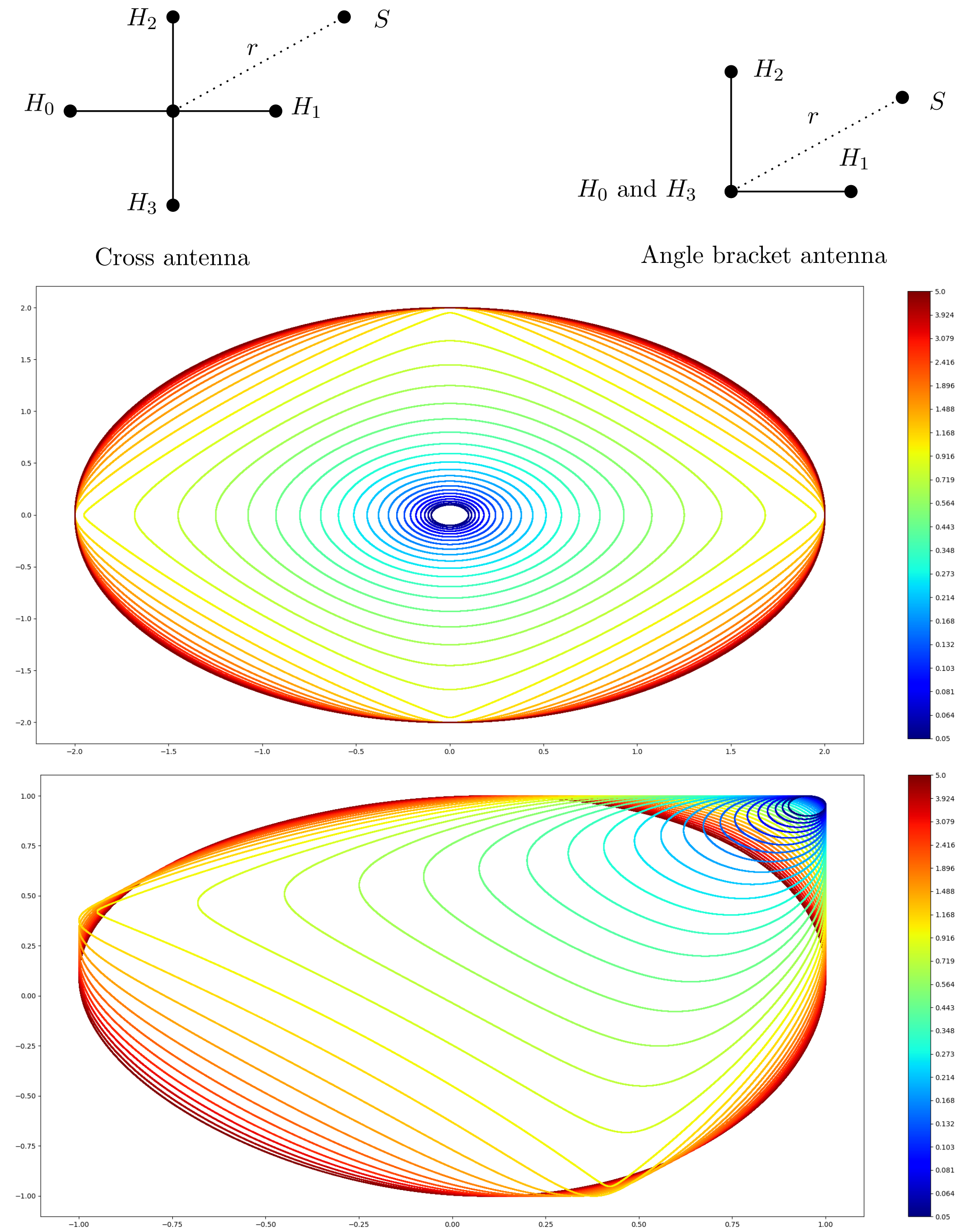


Figure 3.8: Joint TDOA distribution (τ_{10}, τ_{32}) for the cross antenna (top) and the angle bracket antenna (bottom) for multiple concentric circles of source emission. The color represents the radius of these circles. Except for the source, the distance between two dots is one unit.

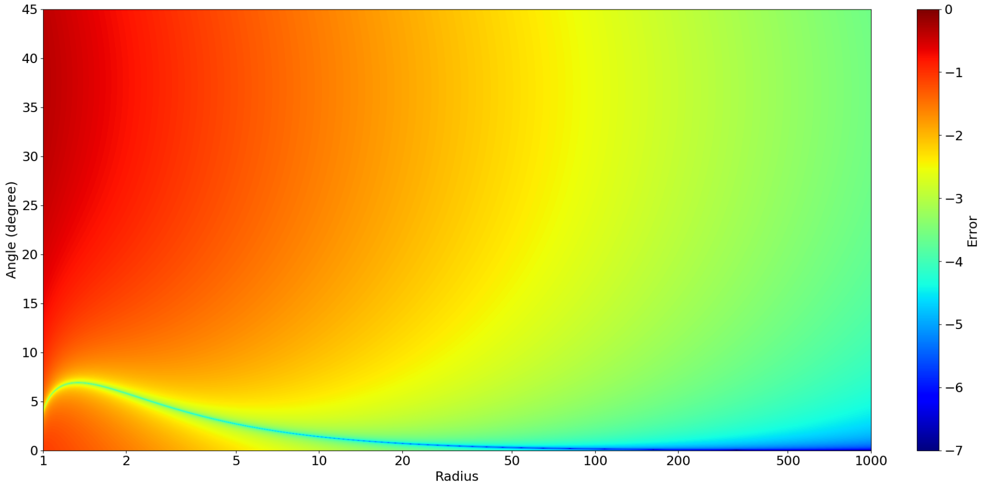


Figure 3.9: Error ($\log_{10}$ of the distance between the near-field generated TDOA pair and the far-field TDOA ellipse) for the cross antenna, depending on the distance r of the source and its angle with the closest hydrophone.

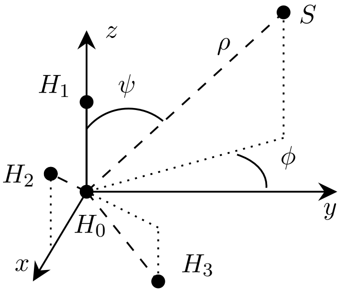


Figure 3.10: Spherical coordinate system.

while (3.20) becomes

$$\tau_{30} = \frac{x_3}{c}\cos\phi\sin\psi - \frac{y_3}{c}\sin\phi\sin\psi + \frac{z_3}{c}\cos\psi. \tag{3.21}$$

Let us assume that the joint distribution of all possible TDOAs from a far field source forms an ellipsoid, then for any TDOA tuple $(\tau_{10}, \tau_{20}, \tau_{30})$ we have

$$\begin{pmatrix}\tau_{10} & \tau_{20} & \tau_{30}\end{pmatrix} \mathbf{M} \begin{pmatrix}\tau_{10} \\ \tau_{20} \\ \tau_{30}\end{pmatrix} = 1, \tag{3.22}$$

where $\mathbf{M}$ is a positive definite matrix. If $\mathbf{P}$ is the matrix $\frac{1}{c}\begin{pmatrix} z_1 & 0 & 0 \\ z_2 & x_2 & 0 \\ z_3 & x_3 & |y_3| \end{pmatrix}$ and $\mathbf{x}$ is the vector $\begin{pmatrix} \cos\psi \\ \cos\phi\sin\psi \\ \sin\phi\sin\psi \end{pmatrix}$, then (3.22) can be rewritten as

$$\mathbf{x}^{\mathsf{T}}\mathbf{P}^{\mathsf{T}}\mathbf{M}\mathbf{P}\mathbf{x} = 1, \tag{3.23}$$

which is true if and only if $\mathbf{P}^{\mathsf{T}}\mathbf{M}\mathbf{P}$ is the identity $\mathbf{I}$ and $\mathbf{P}$ is invertible. $\mathbf{P}$ will not be invertible only in the case where the antenna is a planar antenna. Thus, we have,

$$\begin{aligned}
\mathbf{P}^{\mathsf{T}}\mathbf{M}\mathbf{P} &= \mathbf{I} && (3.24)\\
\mathbf{M} &= \left(\mathbf{P}\mathbf{P}^{\mathsf{T}}\right)^{-1} && (3.25)\\
&= c^2 \begin{pmatrix} z_1^2 & z_1 z_2 & z_1 z_3 \\ z_1 z_2 & x_2^2 + z_2^2 & x_2 x_3 + z_2 z_3 \\ z_1 z_3 & x_2 x_3 + z_2 z_3 & x_3^2 + y_3^2 + z_3^2 \end{pmatrix}^{-1} && (3.26)\\
&= c^2 \begin{pmatrix} d_{10}^2 & d_{10}d_{20}\cos\alpha_{12} & d_{10}d_{30}\cos\alpha_{13} \\ d_{10}d_{20}\cos\alpha_{12} & d_{20}^2 & d_{20}d_{30}\cos\alpha_{23} \\ d_{10}d_{30}\cos\alpha_{13} & d_{20}d_{30}\cos\alpha_{23} & d_{30}^2 \end{pmatrix}^{-1} && (3.27)\\
&= \frac{c^2}{\cos^2\alpha_{12} + \cos^2\alpha_{13} + \cos^2\alpha_{23} - 2\cos\alpha_{12}\cos\alpha_{13}\cos\alpha_{23} - 1}. && (3.28)\\
&\quad \begin{pmatrix} -\frac{\sin^2\alpha_{23}}{d_{10}^2} & \frac{\cos\alpha_{12} - \cos\alpha_{13}\cos\alpha_{23}}{d_{10}d_{20}} & \frac{\cos\alpha_{13} - \cos\alpha_{12}\cos\alpha_{23}}{d_{10}d_{30}} \\ \frac{\cos\alpha_{12} - \cos\alpha_{13}\cos\alpha_{23}}{d_{10}d_{20}} & -\frac{\sin^2\alpha_{13}}{d_{20}^2} & \frac{\cos\alpha_{23} - \cos\alpha_{12}\cos\alpha_{13}}{d_{20}d_{30}} \\ \frac{\cos\alpha_{13} - \cos\alpha_{12}\cos\alpha_{23}}{d_{10}d_{30}} & \frac{\cos\alpha_{23} - \cos\alpha_{12}\cos\alpha_{13}}{d_{20}d_{30}} & -\frac{\sin^2\alpha_{12}}{d_{30}^2} \end{pmatrix}.
\end{aligned}$$

$\mathbf{M}$ is indeed a symmetric positive definite matrix, and we obtain its inverse with its Cholesky decomposition, with $\mathbf{P}$ having strictly positive diagonal entries in the non degenerate case.

With this method, it is thus possible to obtain the shape of the antenna using either (3.27) or (3.25). Alternatively, if the shape of the antenna is known, and not used in the TDOA estimation, then it can be used to filtered of TDOA that do not lie on the ellipsoid distribution.

3.4 Synchronising audio with video signal

In the situation when multiple media coming from various sources have to be combined, the first step is to synchronise them. For example, in the Sarano protocol of Section 6.1, the audio recorded by the JASON recorder and the video recorded by the mounted GoPro need to be synchronized in order to plot the click on the video in Section 6.3. Even when each of the media contain the same type of information (e.g. audio data), simple methods such as cross correlation might not suffice as some issues might arise.

3.4.1 Synchronising JASON audio with GoPro audio

In the case of the Sarano protocol, Section 6.1, the GoPro was placed inside a case to allow a better positioning and, longer and deeper dives, which led to a distorted and directional recording. The GoPro audio signal is degraded due the lossy compression, which is allowable when dealing with speech or signals meant to be heard by humans, for which the loss of information is optimized.

Another issue that appeared was time dilation between each signal. Even though the dilation was only around 1.5 %, this means that files perfectly synchronised at their start will have an offset of 1 second after one minute. One second is long enough to have a JASON card recorded click synchronized with a GoPro recorded click, when in reality they correspond to two distinct instances. Thankfully the time dilation can be assumed constant for the time scale of a file.

With the issue of time dilation, a cross correlation on the whole file cannot be used, as a single offset time between the two file does not exist. The result of the cross correlation will be composed of all the segregated maxima belonging to actual matching events, and other maxima belonging to combinations of events. Using multiple cross correlations, each only having one small part of the first signal could be a solution. Each cross correlation will only contain one true maximum, and combination maxima. In theory the true maximum should be higher than the other maxima. In practice, the two signals are too different for it to work, especially when the allowed time offset is large (score of seconds). Since the dilation is assume to be constant within each file, a visualisation of each local correlation in a spectrogram manner will have each true maxima aligned, with the line formed having the dilation as its slope, as it can be seen in Figure 3.12.

This figure was made using windows of 4 096 samples at 48 000 kHz. Note that bicubic interpolation was used when the image was created, which in conjunction with the absolute value, allows the line to be seen in this plot, since it would be otherwise too fine for the 192 000 sample width of the local cross correlation. Note that this example was chosen because the signal made of creaks is dense enough in terms of clicks so that

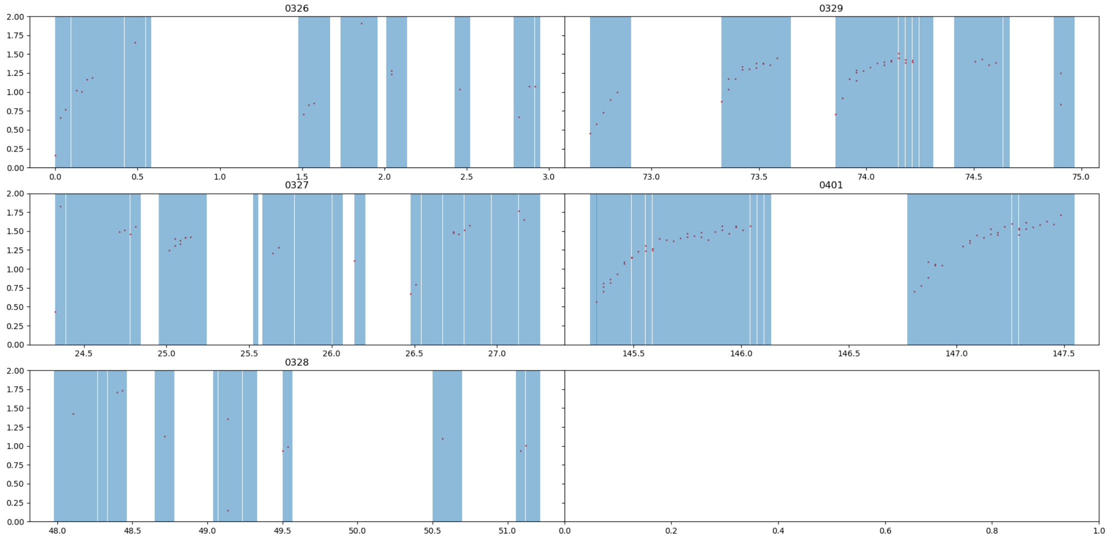


Figure 3.11: Time dilation of the JASON card in percent
The blue intervals indicate the presence of an audio file (i.e. the card was on)

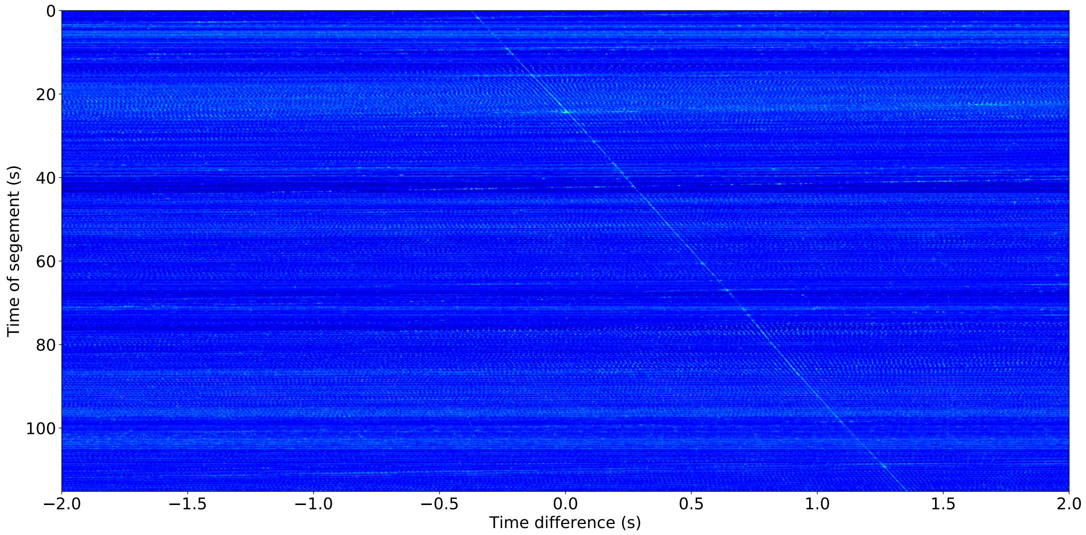


Figure 3.12: Absolute value local cross correlation between the GoPro and JASON recordings

the line is present for every time step.

The solution that was adopted here was to find the dilation and the offset at the same time. However, even with the right dilation factor, using a cross correlation does not work to find the right offset. Thus we first investigated a method to find the right offset, without any dilation. The way we found was to cross correlate the log amplitude spectrogram of each signal. This prevents the large change in energy due to the directivity, yet maintaining more information than a simple binarization. Computing the maximum of this cross correlation for various dilation factors will also give a maximum at the right dilation factor.

Instead of dilating the signal to test various dilation factors, we interpolated directly the log amplitude spectrogram in order to speed up computation time.

3.5 Conclusion

With the basis of signal processing techniques that was introduced in this chapter, any database can be processed to be later analysed. A click detector will first detect the presence of sperm whale clicks, which can then be used to estimate the TDOA, or used directly in the analysis to extract various features (IPI, spectrum, ...). This chapter also presented the ellipse method used to either calibrate the distance between the hydrophone, or to filter out spurious clicks (see Section 5.5.2). Finally, when a video channel is present, a synchronisation method was presented to synchronize the video signal with the audio signal from another device.

Chapter 4

Far Field Recording - BOMBYX

The previous chapter introduced the signal processing techniques. With them, our databases can be processed and analysed. Such databases are presented in the following chapter starting with two far field databases, BOMBYX and Sphyrna, followed by a near field one. These databases each have their specificities, which allow for different pieces of knowledge to be acquired.

BOMBYX is a sonobuoy installed in the Mediterranean sea [123], near the island of Porquerolles (42° 56 N and 6°19 E), south of Hyères, in the south-east of France, as shown in Figure 4.1. It is positioned at 27 meters of depth, and records at 50 kHz with two hydrophones spaced by 1.83 m. BOMBYX is facing south, meaning that the evolution of the TDOA allows to know if a group of sperm whales goes from east to west or west to east. Since BOMBYX is fully emerged and at 27 meters of depth, it reduces the surface noise generated and recorded.

The sound card was made by OSEAN, and the two hydrophones are Neptune D/70– see Figure 4.4. The recording protocol changed along the years (from fully continuous recording to 5 minutes every 20 minutes, varying between 16 and 24 bits encodings), with recording sessions lasting around 3 months. A team of divers is then sent to change the batteries and collect the hard drive containing the recordings.

With an estimation of 400 sperm whales in the Mediterranean sea [124], BOMBYX serves as a monitoring station to survey this endangered species. This is also why BOMBYX is situated within the Pelagos sanctuary which aims to protect marine mammals.

4.1 Estimation of the sperm whale trajectory

As shown in Section 6.2.1, BOMBYX can only estimate the azimuth of DOA, since it only has two hydrophones. Unlike the Sarano antenna of Section 6.1 which is a moving antenna, BOMBYX is a fixed buoy, meaning that other elements could be used to obtain a better position of the source, such as the surface echoes, or the interaction of the whole environment, by assuming that the click of an individual stays constant at the source [125] (which is far from reality in the case of the sperm whales.). However these techniques are harder to implement since they require a good estimation of the TDOA

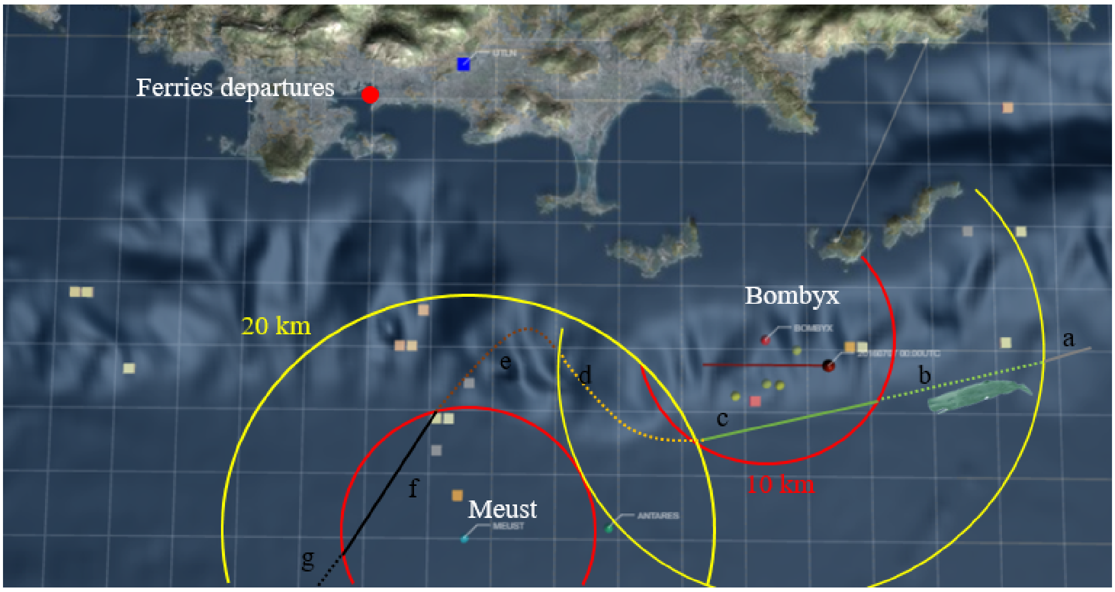


Figure 4.1: Map of Hyères where the Bombyx sonobuoy is placed.

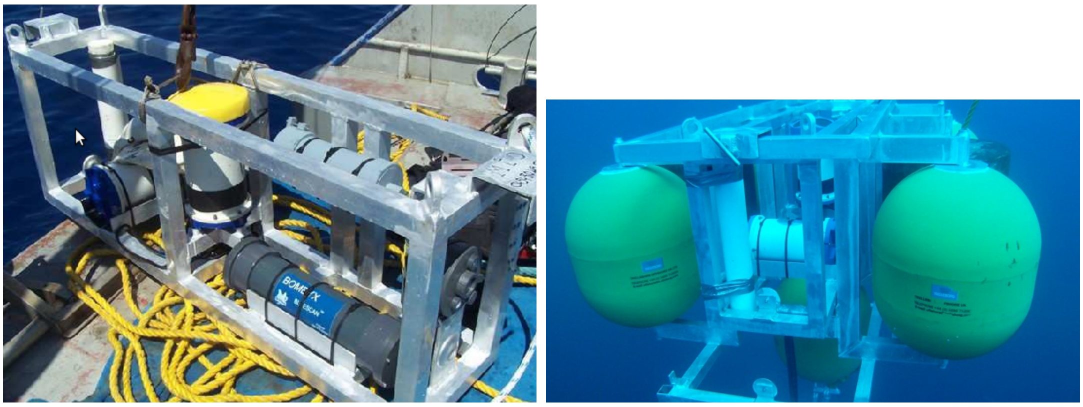


Figure 4.2: BOMBYX before mooring

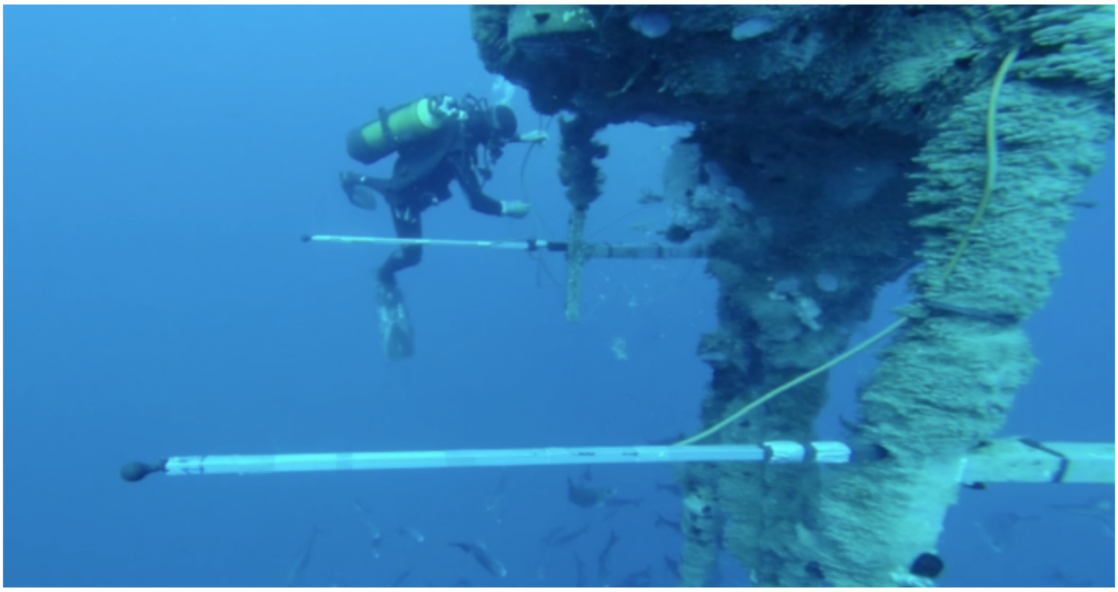


Figure 4.3: Underwater picture of BOMBYX.

Figure 4.4: Close up view of BOMBYX's sound card and hydrophones.

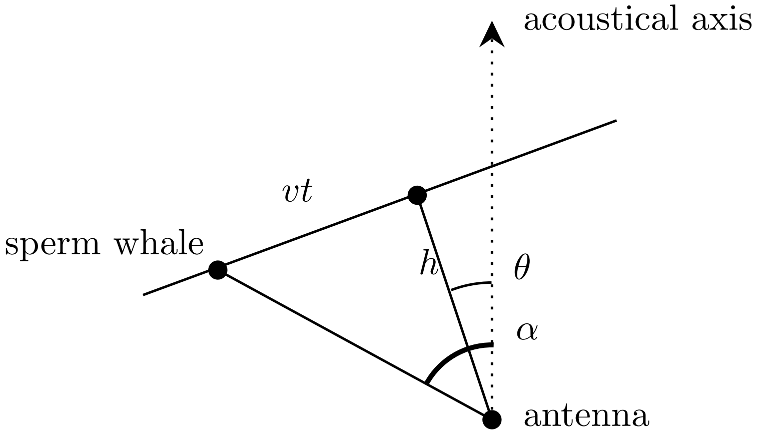


Figure 4.5: Model of a rectilinear moving whale

between the click and its surface echoes, or a sound propagation simulation in the whole environment. Even if whales were to pass right through BOMBYX, the amount of time they will be under 20 meters of distance (10 times the size of the antenna – see Figure 3.9) will be negligible compared to the rest of the recording interval, meaning that the far field hypothesis can thus be assumed to be always true. By assuming that the sperm whales have a rectilinear uniform motion, more information than just the azimuth can be obtained as shown in [123] for some some specific angles and relations between speed and range of the animal, that we extend below for all angles.

Figure 4.5 is a representation of this model where a sperm whale is moving at a speed of v. In this model, $t = 0$ is when the sperm whale is closest to the antenna, with a distance of h. The closest point, the current position of the sperm whale and the antenna form a right triangle.

In order to compute the TDOA that will be produced at the antenna, the angle α between the whale and the acoustical axis (bisector of the two sensors) needs to be found. The tangent of the angle $\alpha - \theta$ of the aforementioned triangle defines a relation with α and the other parameter of this model. Thus by expanding the tangent and regrouping the terms in $\tan\alpha$, the relation becomes

$$\tan(\alpha - \theta) = \frac{vt}{h}, \tag{4.1}$$

$$\frac{\tan\alpha - \tan\theta}{1 + \tan\alpha\tan\theta} = \frac{vt}{h}, \tag{4.2}$$

$$\tan\alpha = \frac{\frac{vt}{h} + \tan\theta}{1 - \frac{vt}{h}\tan\theta}, \tag{4.3}$$

$$\alpha = \arctan\frac{\frac{vt}{h} + \tan\theta}{1 - \frac{vt}{h}\tan\theta} + n\pi, \tag{4.4}$$

where $n \in \mathbb{Z}$ accounts for the π periodicity of the tangent function to keep the angle α continuous. Since the discontinuity occurs when $1 - \frac{vt}{h}\tan\theta$ changes sign, n can be defined

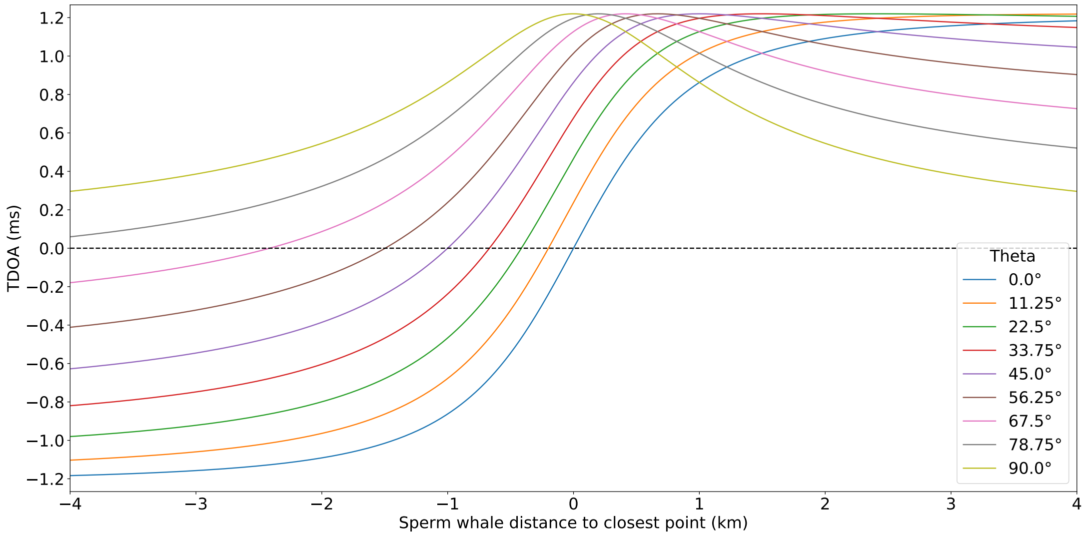


Figure 4.6: TDOA for a rectilinear moving whale depending on the angle θ

as $n = \frac{1-\mathrm{sgn}\left(1-\frac{vt}{h}\tan\theta\right)}{2}$, where $\mathrm{sgn}(x) = \begin{cases} -1 & \text{if} \quad x < 0 \\ 0 & \text{otherwise} \end{cases}$. Since $\tau_{21} = \frac{d_{21}}{c}\sin\alpha$, where d_{21} is the distance between the hydrophones, c is the speed of sound, and $\sin(\arctan(x) + n\pi) = (-1)^n \frac{x}{1+x^2}$, the TDOA τ_{21} can be obtained as

$$\begin{aligned} \tau_{21} &= \frac{d_{21}}{c} \frac{\frac{\frac{vt}{h}+\tan\theta}{\left|1-\frac{vt}{h}\tan\theta\right|}}{\sqrt{1+\left(\frac{\frac{vt}{h}+\tan\theta}{1-\frac{vt}{h}\tan\theta}\right)^2}} && (4.5) \\ &= \frac{d_{21}}{c} \frac{\frac{vt}{h}+\tan\theta}{\sqrt{1+\tan^2\theta+\frac{v^2t^2}{h^2}+\frac{v^2t^2}{h^2}\tan^2\theta}}. && (4.6) \end{aligned}$$

As a thought experiment, if we assume that an antenna has a 4.9 km detection radius, then for a whale that will not be closer than 1 km (i.e. $h = 1$ km), the maximal value for vt will be 4 km, no matter the value of θ. Figure 4.6 shows the TDOA that will be obtained in this situation. Whether the click is close to 0°, 45° or 90° without fitting of a curve, a human operator can label the general direction of the sperm whale.

Using the slope at $\tau_{21} = 0$, one can recover most of the information. First the TDOA is differentiated with respect to time,

$$\frac{d\tau_{21}}{dt} = \frac{d_{21}v}{ch} \frac{(1+\tan^2\theta)(1-\frac{vt}{h}\tan\theta)}{((1+\frac{v^2t^2}{h^2})(1+\tan^2\theta))^{\frac{3}{2}}}. \tag{4.7}$$

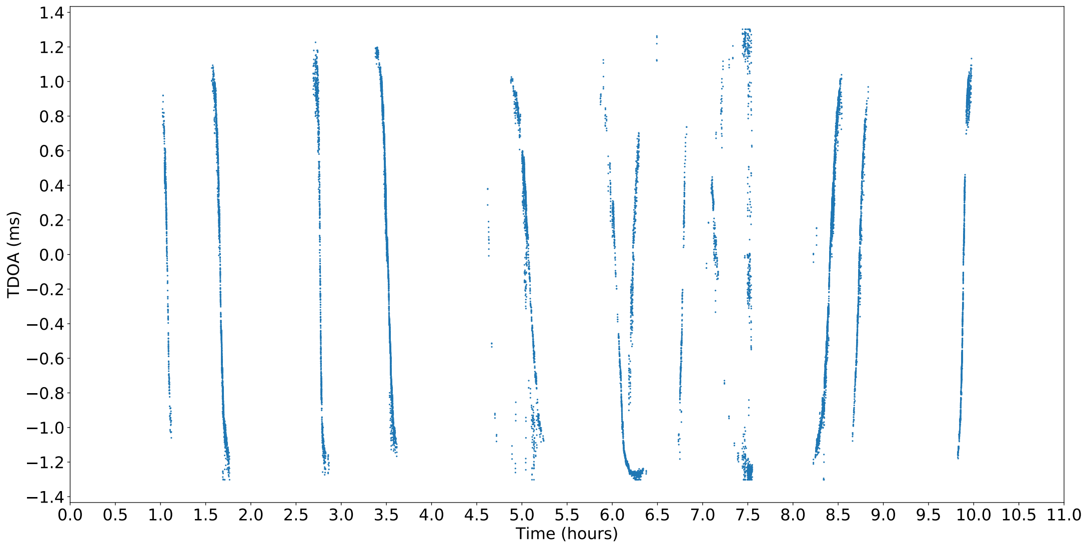


Figure 4.7: TDOA of boats passing in front of BOMBYX.

Since $\tau_{21} = 0$ when $\frac{vt}{h} + \tan\theta = 0$, the slope at $\tau_{21} = 0$ can be obtained as

$$\frac{d\tau_{21}}{dt} = \frac{d_{21}}{c}\frac{v}{h(1+\tan^2\theta)}. \tag{4.8}$$

In order to use these results, further hypotheses need to be made to assume the values of two unknowns. One can assume that the angle θ is always similar, for example in the case where sperm whales always pass trough the same corridor. Then if the closest distance is obtained using another means, such as the time of arrival of the surface echoes, one can obtain the travel speed. Otherwise, one can assume the travel speed to obtain said distance. Finally if the travel speed is assumed and the sperm whales are assumed to always pass trough one point (which could be different than the closest point), then the angle θ can be recovered.

4.2 Application to real data

In reality, only boats will have such an ideal movement. For example, Figure 4.7 and Figure 4.8 show TDOA curves from a recording session in December 2018. This session was taken as an illustration since it was the only session with a continuous recording, whereas other sessions only have a 25% duty cycle (e.g. 5 minutes of recording, followed by 15 minutes of sleep). As shown in Figure 4.7, their curves are monotone, compared to the curve of a sperm whale in Figure 4.8, which contains many oscillations.

Not only the sperm whale curve contains oscillations due to the sperm whale going back and forth in a hunting site, but it also contains a silent interval, where the whale may be moving and breathing at the surface without emitting any clicks. However,

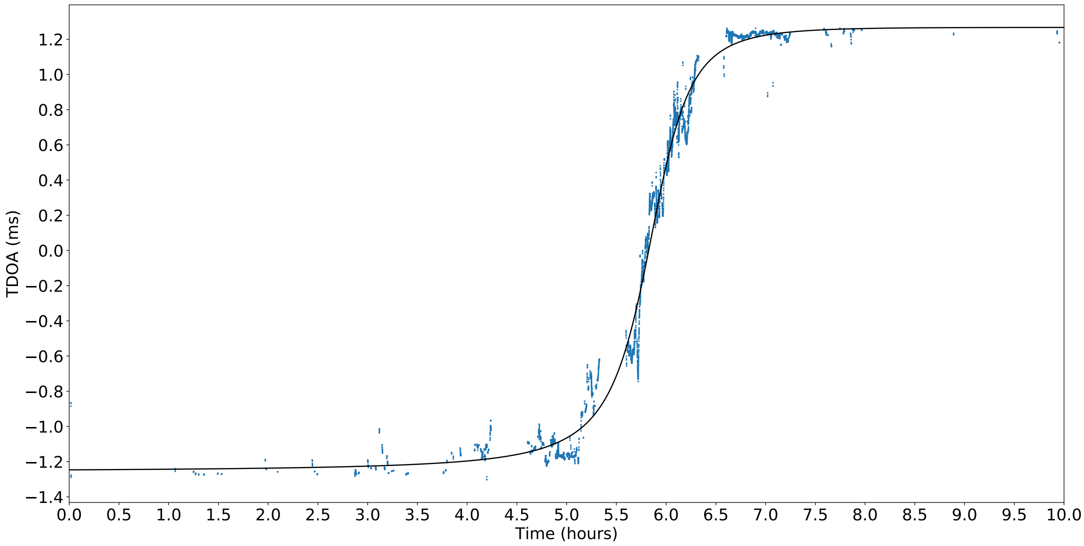


Figure 4.8: TDOA of a sperm whale swimming near BOMBYX
. Blue dots: estimated TDOA. Black line: fitting of the uniform rectilinear model.

on average sperm whales move approximately at a constant surface speed, allowing the fitting of the model. In the example of Figure 4.8, the model returns an angle of 6.07°, and a ratio v/h of $2.28h^{-1}$. This correspond to a sperm whale going from east to west following the underwater cliff, as seen in Figure 4.1. While sperm whales have a cruising speed that vary around 10 km/h, this value cannot be taken as the speed in order model. This cruising speed suppose that the sperm whale will continuously swim toward the west. However the click that were measured indicate that the sperm whale was diving. While the sperm whale will not surface at the same position it started its dive, meaning that its dives have a mean horizontal speed, the global mean speed must be slower than 10 km/h. Since the sperm whale is detected during 10 hours, its mean surface speed can be assumed to be 1 km/h, meaning that the distance h is 2.28 km, which correspond to the distance between BOMBYX and the bottom of the cliff.

4.3 Conclusion

The first of three databases that was presented in this chapter is the BOMBYX database, which has the advantage of being a fixed, merged database. Its long term session allows for various background noise conditions (weather conditions, amount of boat noise, etc.). This increase in background noise variability allows to test the generalisation of various techniques, and makes it easier to differentiate clicks from background noise.

Chapter 5

Far Field Recording - Sphyrna

Unlike the previous chapter which presented a fixed bi-hydrophone antenna, with a low frequency to record during months, the database that will be presented here is a five-hydrophone mobile antenna.

Passive acoustic monitoring (PAM) provides a viable option to assess the population status of cetaceans, model their behavior or yet to prevent ship collisions. Existing techniques use different methods to record and localize the echolocation clicks of cetaceans. Static hydrophone arrays, using underwater buoys, provide stable recordings and observations of foraging [126, 127, 123] but can only monitor a fixed location, making it difficult to track animals over longer distances. A bottom-mounted array with small aperture (2m) and high sampling rate has been deployed for 3D localization of multiple whales [128, 129]. More flexibility can be achieved by mounting hydrophones on a vessel. In another experiment [130], a wide-aperture towed array of two hydrophones was used for tracking in 2D dive profiles of sperm whales, taking advantage of surface-reflected paths, but not allowing 3D localization. In [131], a first attempt of a moving deep 6-hydrophone array resulted only in range estimation of the cetacean. In [132], we designed a high sampling rate sound card and robust passive acoustic algorithm for a small non-uniform array of 5 hydrophones mounted directly under an autonomous surface vehicle (ASV) called Sphyrna. This ASV was built by Sea Proven in France and is shown in Figure 5.1, and in a demo[1].

This setup is challenging due to the small aperture and the recording of sound close to the surface, above the thermocline. The recordings are made at 16 bits@600 kHz with our available high-resolution sound card JASON (Figure 5.3), designed in our scientific platform SMIoT [133]. We propose a method for efficient filtering of the transients and other noises generated in these extreme conditions, in order to monitor deep foraging cetaceans like the sperm whale (*Physeter macrocephalus*, Pm) which spends more than 70% of its time down to 1 km. In the following we will illustrate our algorithms to process the echolocation pulses of sperm whales that it uses for orientation and prey localization.

A moving near-surface hydrophone array was also used in [134, 135] to track cetaceans precisely in 3D over time. The analysis of their movement can also then help to deduce

---

[1] `http://sphyrna-odyssey.com`

Figure 5.1: The ASV Sphyrna, 17m long. It is a proa (South Pacific design), consisting of two unequal length parallel hulls. It is extremely stable, even at high wind force.

their behaviour, helping in the creation of a conservation program.

5.1 Method

To estimate the 3D localization of the animal, we compute the Time Delay of Arrival (TDOA) on signal chunks resulting from an automatic sperm whale click detector. TDOA estimation is cross-correlation based. The clicks are extracted using a simple spike detector on an enhanced signal (or noise-reduced bandpass-filtered signal).

The cross-correlation based method we apply to compute TDOA also yields values linked to the energy of the click and its coherence between all pairs of channels (see Section 3.2.4). However, after this phase, echoes from the surface and from objects in the water column still exist. Thus, we cluster the clicks by tracks in the time-TDOA space with the DBSCAN algorithm (Density-Based Spatial Clustering of Applications with Noise [136]). DBSCAN is a density-based clustering non-parametric algorithm: given a set of points in some space, it groups points that are closely packed together (points with many nearby neighbors), marking as outliers points that lie alone in low-density regions (whose nearest neighbors are too far away). This clustering procedure contributes to discarding spurious clicks (e.g. false positives from the spike detector) and groups together tracks of clicks facilitating the analysis of the cetacean behavior. We finally estimate 3D positions of the whale from the TDOAs using a nonlinear solver. The tracks are stabilized according to the yaw and roll from the ship's Motion Processing Unit (MPU).

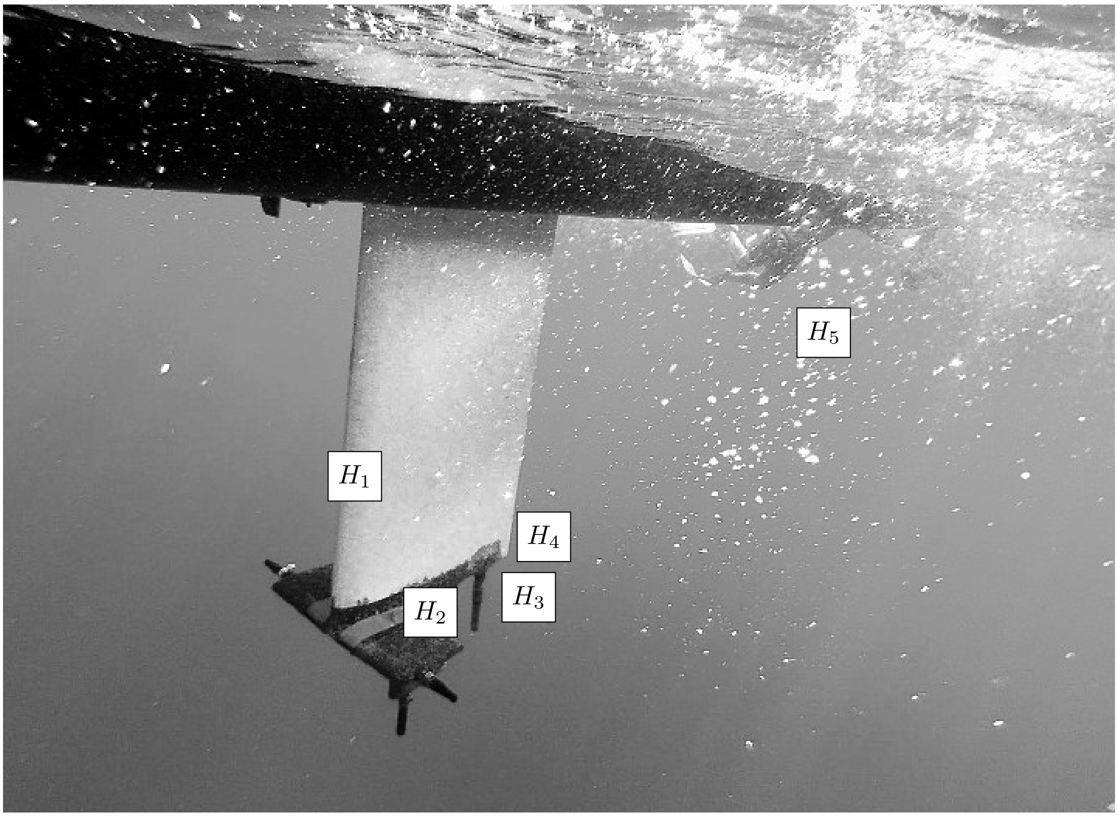


Figure 5.2: Layout of the 4+1 hydrophone array. The inter-hydrophone distances under the keel are 35, 59, 59, 63, 63, 70 cm. Hydrophone H_5 is placed at the stern, 7 m away.

5.2 TDOA computation

As shown in Figure 5.4, the process used to estimate the TDOA combines multiple methods that were described in Section 3. The Teager-Kaiser approach was used to detect the click, which was then used to estimate the TDOA using the Geometric SRP. Since it has a high algorithmic complexity, especially in our case with five hydrophones, one of which being seven meters afar, the TDOA was evaluated in two steps, both using the Geometric SRP. First, only the four hydrophones of the keel were used to evaluate the first three independent TDOAs, as if there were only four channels. Then all channels are considered, but with the previously estimated TDOAs fixed to their estimated values, leaving only one independent TDOA to evaluate. With this two-step method, two objective functions are used, thus providing two values linked to the energy of the click, and to how well the channels are correlated together.

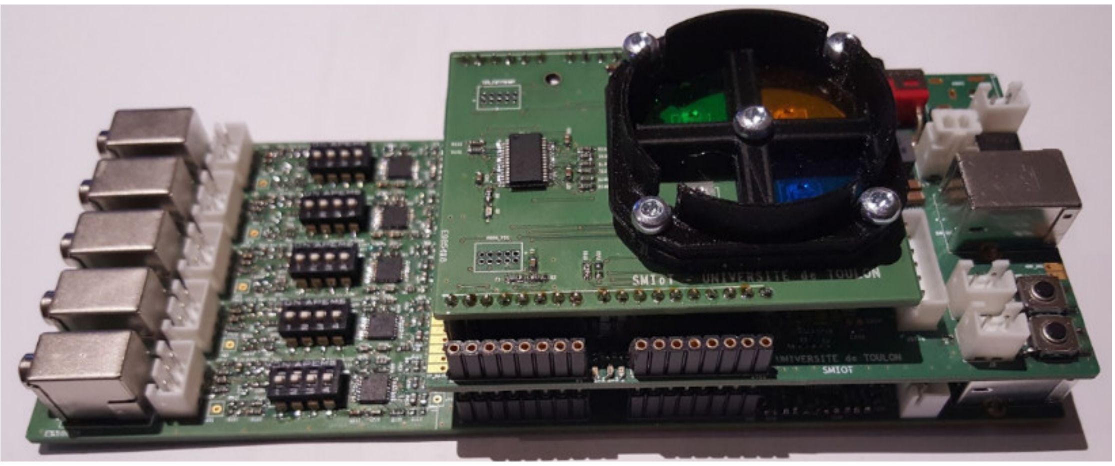


Figure 5.3: The JASON sound card (from `http://smiot.univ-tln.fr`), up to 5 x 2 MHz sampling rate at 16 bits resolution, placed into the drone (the built in luxmeter was not used in this work).

5.3 Filtering

5.3.1 Filtering on TDOA

Figure 5.5 shows the four independent TDOAs $\tau_{21}, \tau_{31}, \tau_{41}, \tau_{51}$ during a 3-hour recording session where the sperm whale we observed did three dives. Horizontal tracks of points in this figure were caused by artifacts. In order to distill the actual clicks from the echoes and false positives, we apply DBSCAN clustering on the time-TDOA space and keep clusters linked to the main track. DBSCAN is a clustering algorithm based on the number of neighbours within a ball of diameter `epsilon`, and then linking neighbouring points that have enough neighbours themselves. This clustering method is thus great for non convex clusters, such as TDOA tracks. The metric used was the Euclidean distance. Since the 2018 Sphyrna recording session contains only one sperm whale during three hours, the DBSCAN parameters were manually optimised such that the remaining clusters could be chosen by hand (see Section 5.5.1 for DBSCAN parameter optimisation). The optimal value for `epsilon` is set according to the computed number of clusters in the `epsilon` function. We then determined the clusters for five `epsilon` values that were in the middle of the plateau depicting the evolution in number of clusters. Finally we chose the values of `epsilon` that clustered the track with a minimal number of clusters (4 main clusters for the three dives and two small clusters at the end of the last track) eliminating most of the spurious and echo clicks. This gives a better result than filtering by keeping points that belong to the Gaussian with the highest energy (points that were in the upper right quadrant of (-29, -22) of Figure 5.6).

The only echoes that we could not eliminate with DBSCAN were the surface echoes (for example in Figure 5.5, TDOAs between H2 and H1, the yellow points that are above 0.00022 s), since their tracks originate from the same point as the main track, which is

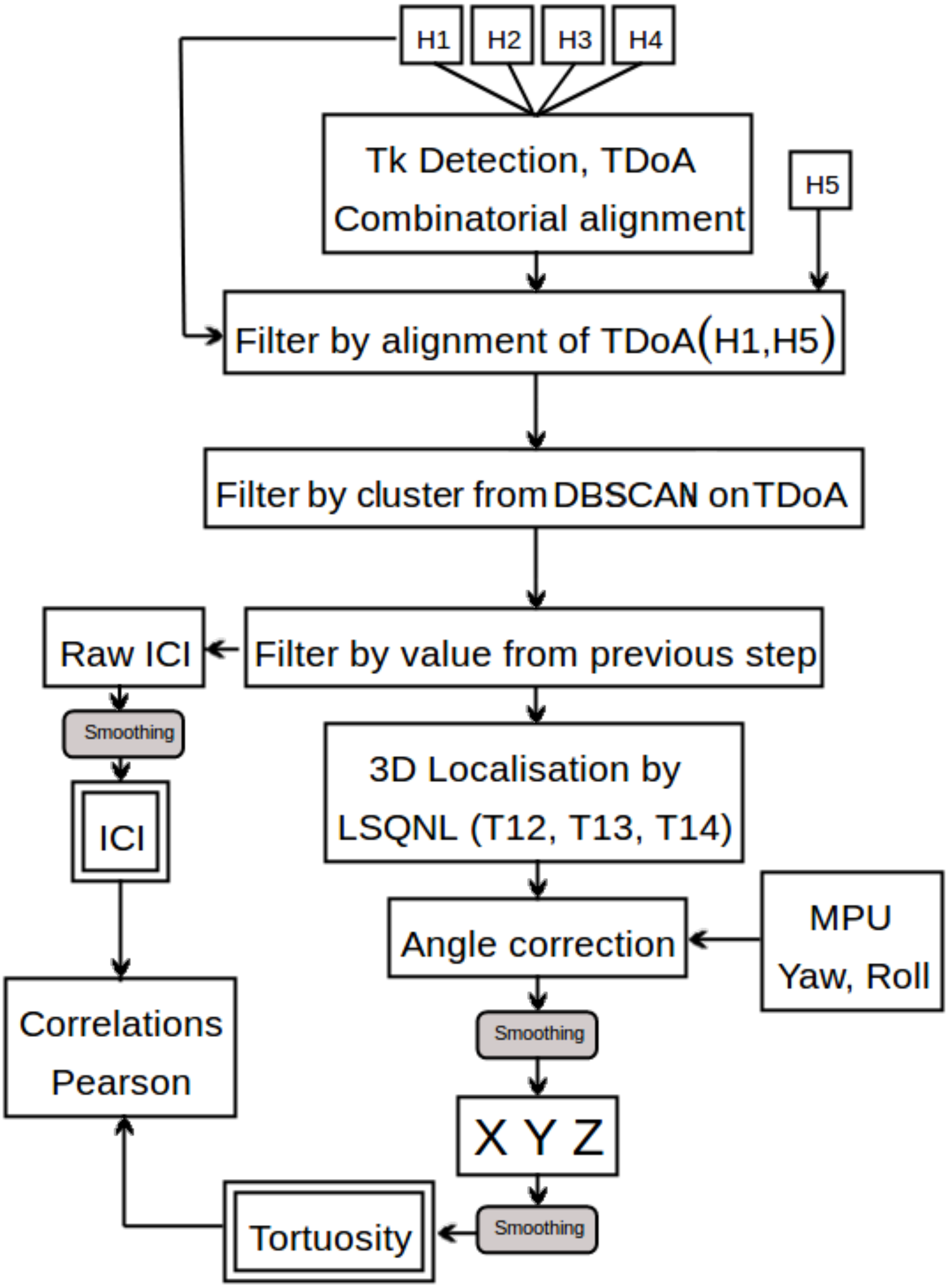


Figure 5.4: Flowchart of our method.

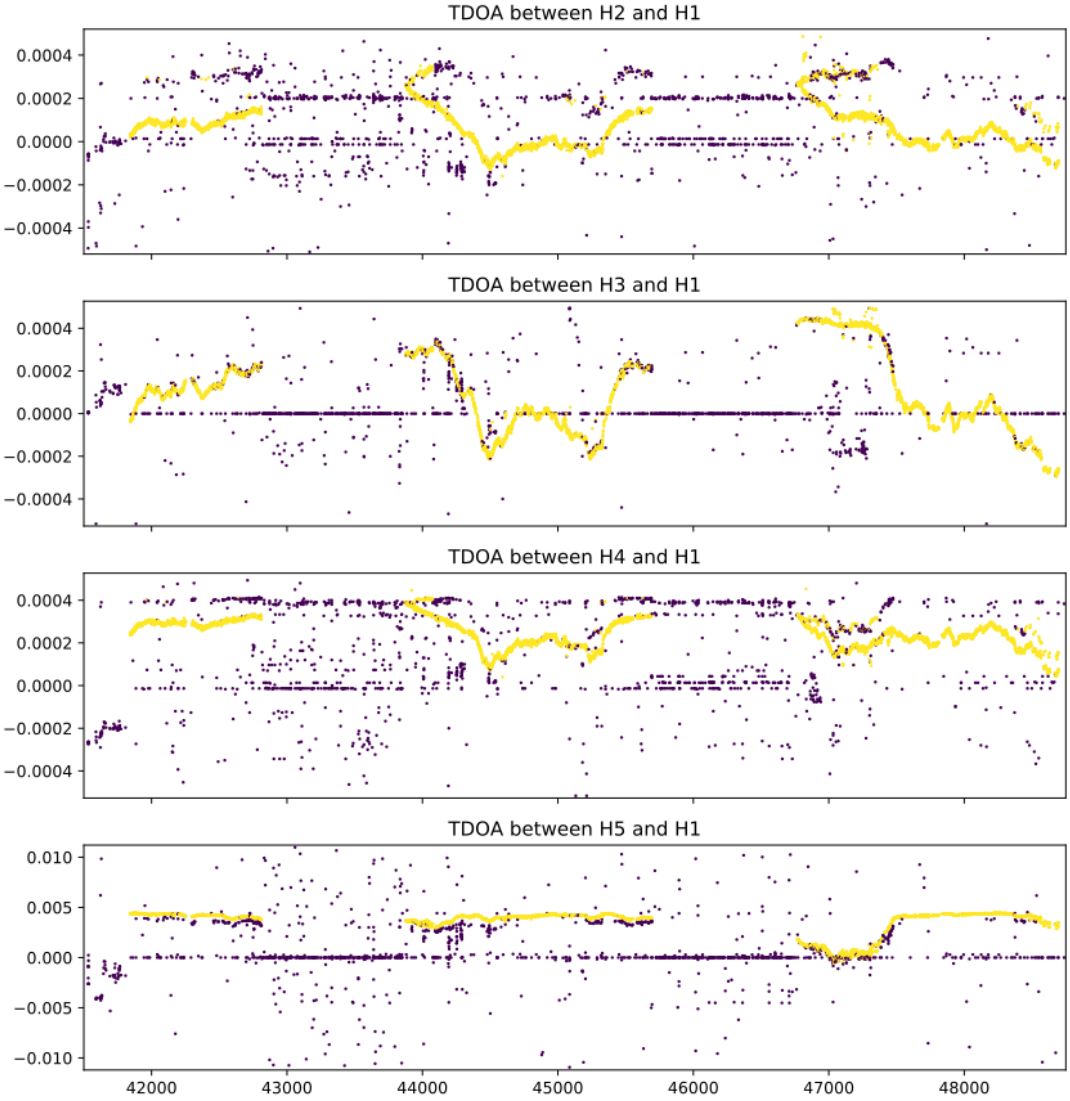


Figure 5.5: TDOA during 3h of recording. The yellow points are the points kept after the DBSCAN clustering.

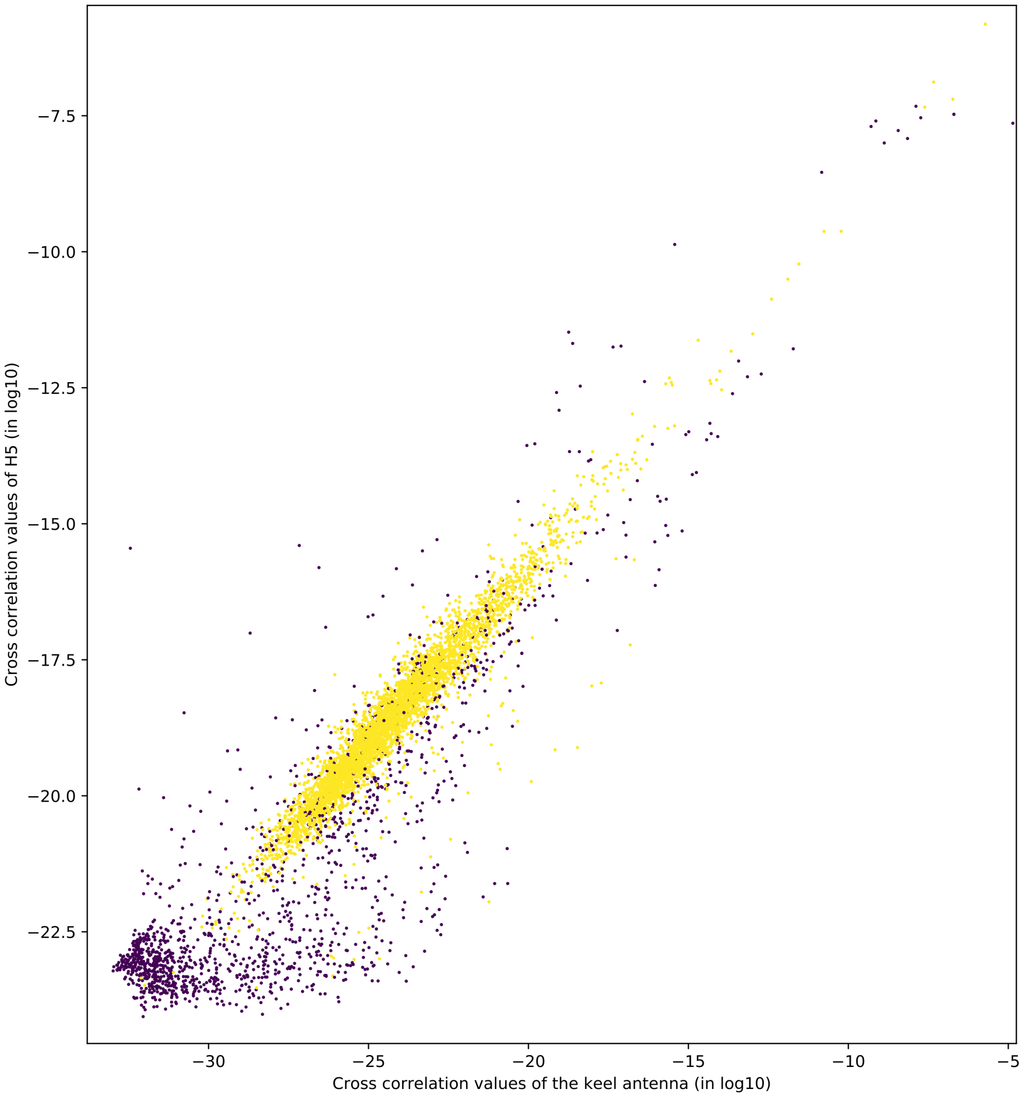


Figure 5.6: Scatter plot of the two cross correlation values. Same 3h of recording as in Figure 5.5.

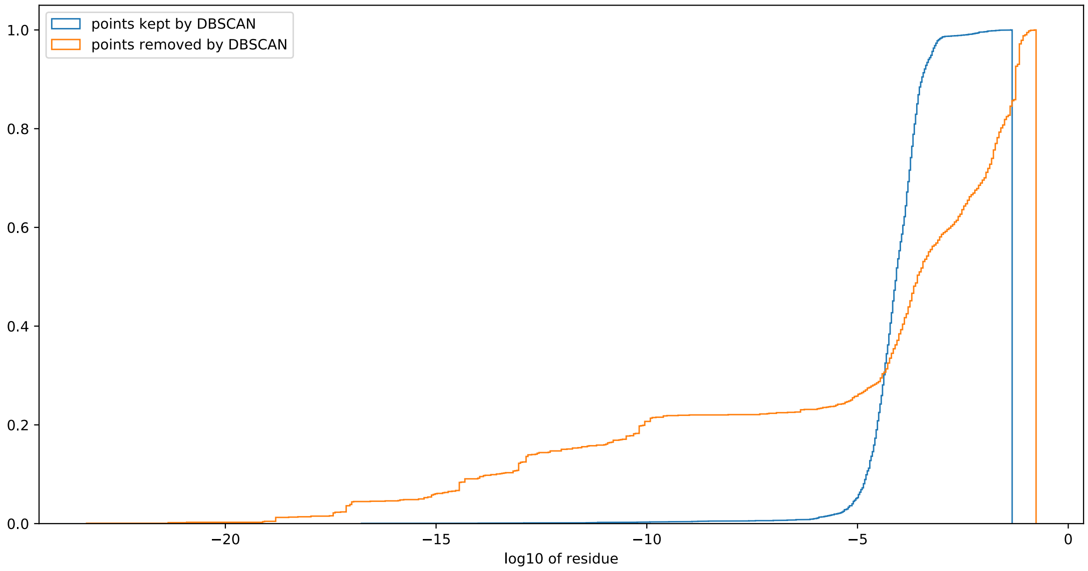


Figure 5.7: Cumulative density histogram of the residue of the localization algorithm.

the start of the dive.

5.3.2 Source localization and post filtering

Once the list of clicks has been finalized, we converted their TDOA into Cartesian coordinates using the keel antenna as base. These are then mapped to GPS positions using the yaw, roll and pitch obtained from the MPU. Surface echoes are eliminated at this stage since they lead to positions above the surface. We then remove the remaining noise with a moving median filter with a window size of 20 seconds, whose results are shown in Figure 5.10.

We used the solver detailed in ([110]) to compute the 3D coordinates from the TDOAs. The histograms of the residue of the solutions are given in Figure 5.7, on the set of points kept using DBSCAN, and those discarded using DBSCAN. They show that the residues of retained clicks lie in a small range, while the others are more spread out, from smaller residues (artifacts) to larger residues (non coherent TDOAs). Thus DBSCAN filters spurious transient detections.

5.4 Tracking results

In this section we compare our results to the 3D track obtained previously without the proposed filtering [137]. Figure 5.9 shows the improvements: the spurious clicks have disappeared, so did some of the tracks of echoes.

In order to compare how the added DBSCAN filtering improved the track we obtained, we plotted the density histogram of standard deviation over 40 ms windows.

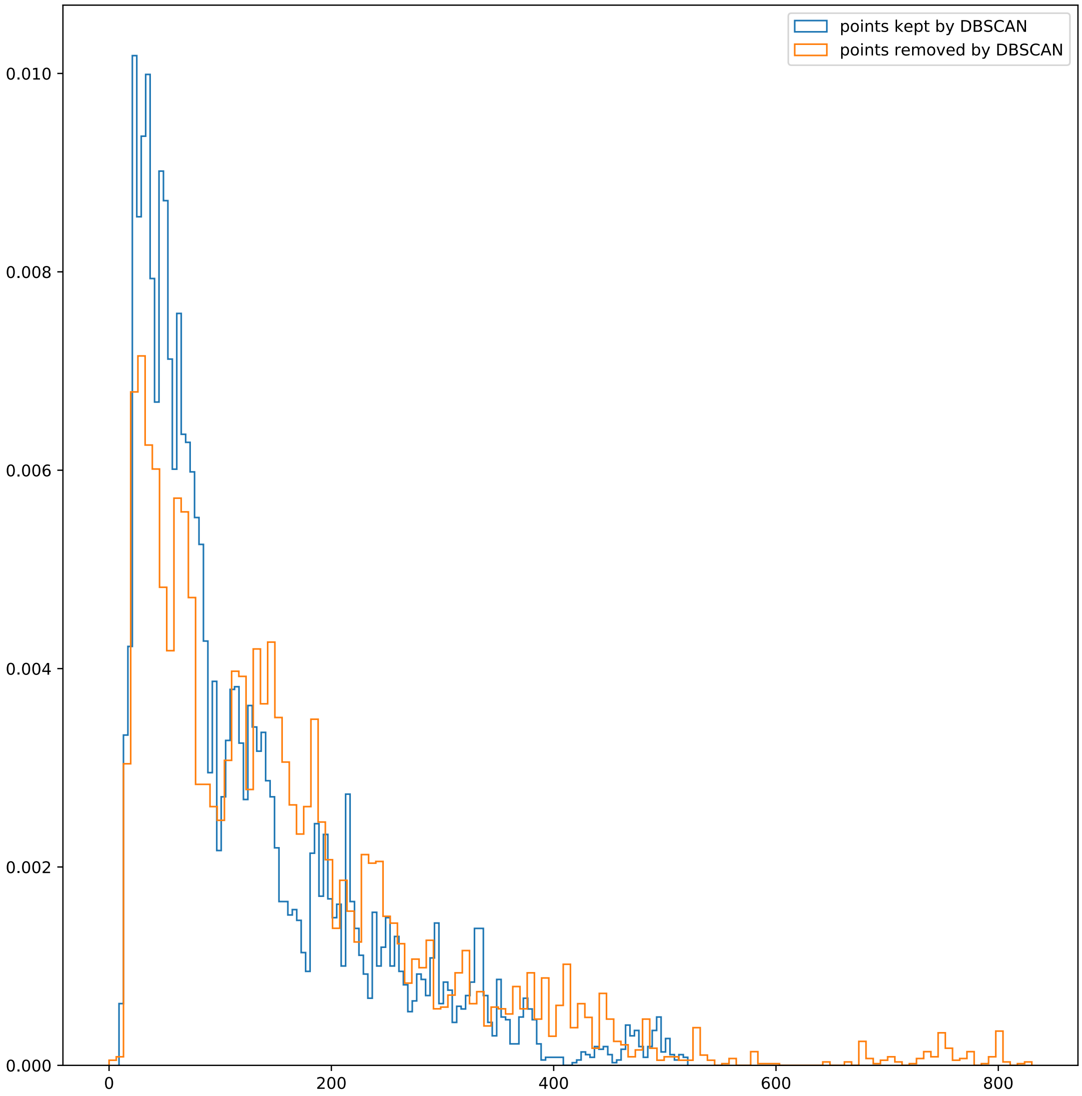


Figure 5.8: Moving standard deviations in a 40 sec window for the set of points kept by DBSCAN (blue) versus for points removed by DBSCAN (red).

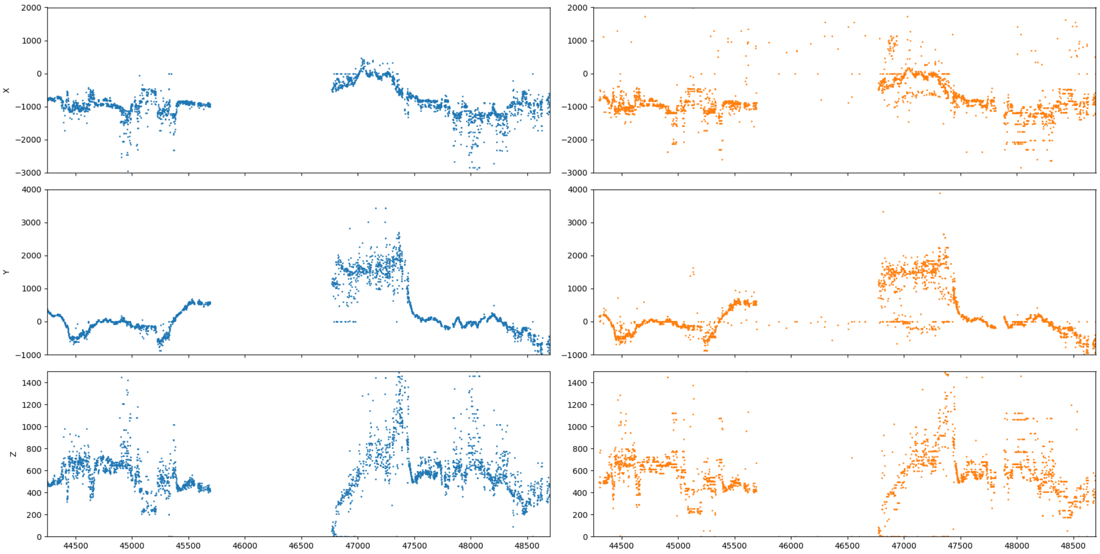


Figure 5.9: Raw coordinates of the Sperm Whale during the last two dives, in the keel antenna reference space. The blue points are from the DBSCAN filtering. The orange points are from the baseline [137].

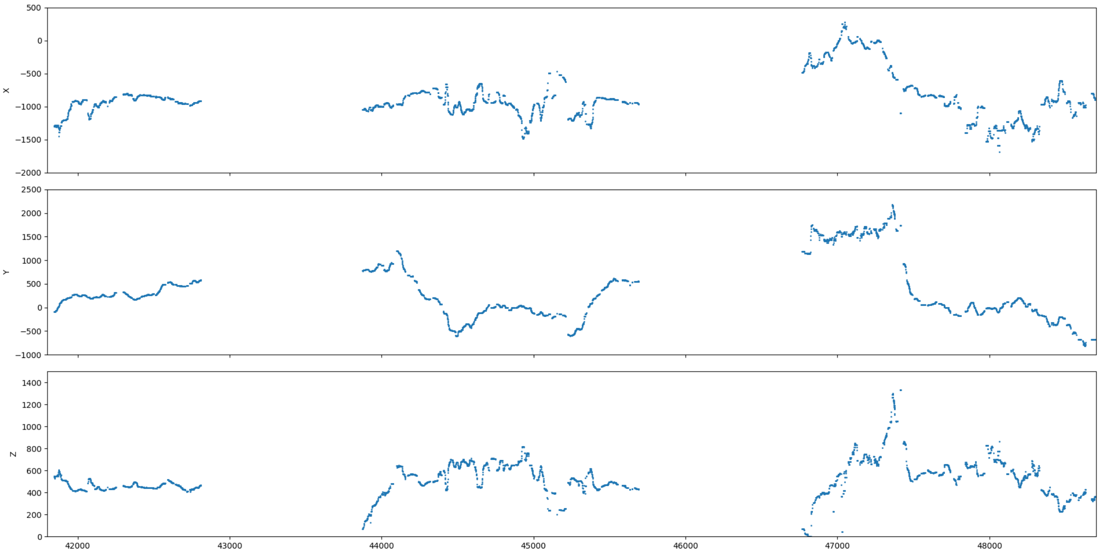


Figure 5.10: The computed track after DBSCAN of the three recorded dives, time in seconds. X is Northing (m), Y Easting (m), Z Depth (m). Z axis is directed downward.

Figure 5.8 shows that our method allows to remove all the positions with a standard deviation above 550 m, and increases by a factor 4/3 the proportion of positions with a standard deviation below 100 m.

5.5 Sphyrna 2019-2020

The 2019-2020 expedition reused the drone of 2018 with the addition of another drone. The second drone (Sphyrna 70 or S70) measures 21 m and was equipped with a 5 hydrophone antenna, disposed in the same shape as the hydrophones of the S55 (the drone of 2018). As in 2018, the legislation imposed that a manned boat follow the two ASVs. During the recording session a four hydrophone antenna was also installed at the rear of the catamaran. On the fifth channel of each card the signal from a filter PPS signal of a GPS card was added to the signal of the fifth hydrophone. PPS is a signal with a raising edge repeating exactly each second. The original PPS signal emitted by the GPS card was a rectangular signal, which we filter using an analogical high pass to only modify one sample of the original hydrophone signal. The addition of the PPS allows to synchronise each antenna, and to measure any drift that might appear in the recording frequency. The 2019-2020 mission started in September 2019 and ended in April 2020, with most of the recording effort done in January, February, and April. Note that both drones were not always present due to mechanical issues. The use of three antennas allows to cover a wider perimeter during the search for sperm whales, but also allows to combine the position estimation of the various vessel, similarly to [138, 139, 140, 141].

5.5.1 Finding DBSCAN parameters

A DBSCAN clustering can be described by three parameters: the distance metric, the `min sample`, and the `epsilon`. In the DBSCAN algorithm, a point that is distant of less than `epsilon` from another is said to be directly reachable from that other point. Points that have at least `min sample` points (including itself) that are directly reachable from them are called core points. All core points that are joined by a chain of core points that are directly reachable from each other are in the same cluster. Finally, points that are directly reachable from a core point are added to its cluster. Remaining points are labelled as noise. When DBSCAN is used to cluster tracks of clicks, `min sample` should be set to 3, since as core points, the points of the track should at least directly reach itself, the previous point and the next point. The two points at the edge will still be added to the cluster, since they are directly reachable from core points. The distance chosen should treat the TDOA equally if they have the same order of magnitude. In our case the distance chosen is a Euclidean norm with an extra parameter to scale the time. This means that the TDOA difference can be greater if the time difference between two clicks is short than if it was longer. This goes against the fact that the longer the time passes between two clicks, the further the sperm whale can move from its previous position, thus increasing the TDOA difference. However, in our case the maximum TDOA difference that can be allowed between two points of a track is of the order of the TDOA estimation error. Since the tracks of clicks are from whales far away, the `epsilon` should be chosen around the square root of the number of TDOA used plus one, if the distance is expressed in samples. This value can be increase if the TDOA estimation error is large, or if the TDOA difference is greater than one sample for each TDOA between two clicks. Finally in order to choose the value of the scaling parameter that will scale the time to be similar to the click , multiple processes are possible. A first solution is to consider the largest ICI that will make two clicks part of the same track. This value should be high enough so that clicks are very seldom not part of a track, but not too high, since it will increase the number of clustered spurious clicks. Tracks of clicks split into two clusters can be merged after the filtering of noise. With this method, the scaling parameter value will be such that the length of this largest considered ICI has the same value as `epsilon` after scaling, or slightly smaller if a variation of TDOA value is accepted for this ICI. Another method is to label multiple (to increase variability) small time frames. The scaling parameter can then be chosen to be the value that maximizes the number of clicks in a cluster, while minimizing the number of spurious clicks outside of clusters. A final method is to plot the evolution of the nearest neighbour distance regarding these parameters for a large number of clicks. The extreme values of these parameters will lead to either a time difference that is negligible compared to the TDOA, or the TDOA that are negligible compared to the time differences. In the first case, most clicks should find another click with similar TDOA, meaning that most nearest neighbours should be small. In the second case, the nearest neighbour will be the previous or next detection, with all the distance growing proportionally to the scaling parameters. The true clicks have the next/previous point in the the track as a point with similar TDOA which will be its nearest neighbour for a large enough scaling factor, while the spurious clicks need to look far in time to find such a point. Thus the spurious click will transit from the first regime to the next earlier than the true click will. The value of the scaling parameter will thus be set before the true click distribution starts growing over the `epsilon` value. The greater the number of independent TDOAs, the harder it will be for spurious clicks with ”random” TDOA to find another click with similar TDOA close to them in time, making it easier to differentiate between the two distributions. Note that while these methods seem to indicate that the two distributions are already filtered at the moment of choosing the scaling parameter value, the DBSCAN that follows will add the requirement of a second neighbour fulfilling these conditions, thus increasing the number of points filtered. Note that this method is close to using the evolution of the ratio of points filtered out by DBSCAN.

5.5.2 Additional ellipse filtering

As explained in Section 3.3, the tuple of independent TDOAs are distributed on an ellipsoid. Since the TDOA estimation method used here does not use the shape of the antenna, it does not force the TDOA to reside on this ellipse. Hence the TDOA for spurious clicks that are due to other mechanism than the wave propagation in sea water should not reside on this ellipsoid. Electric spikes, for example, is a kind of spurious click that will have all of its TDOA close to zero, since it is an artifact that appears on all

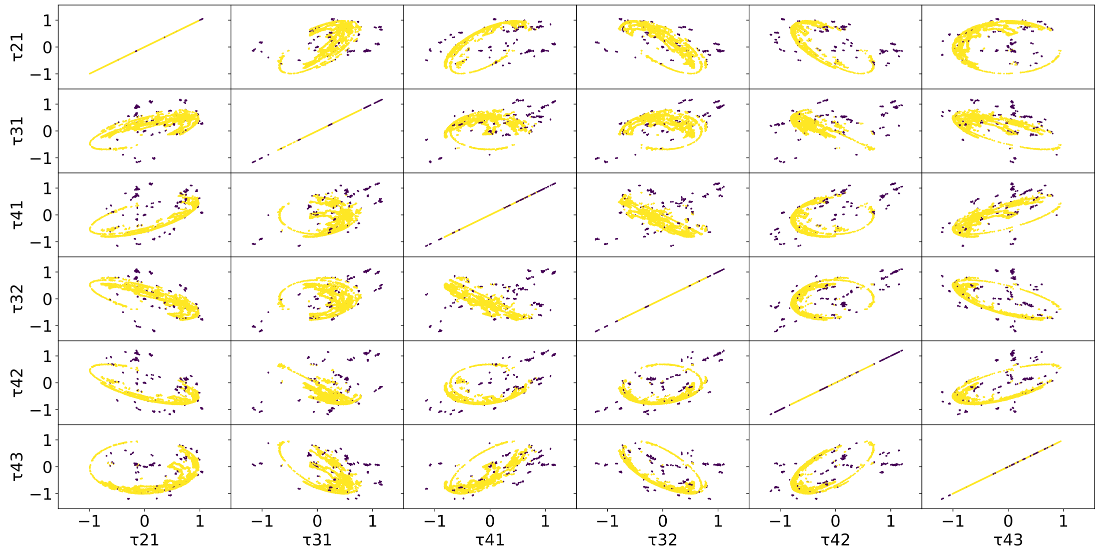


Figure 5.11: Joined TDOA distribution from the 4 hydrophone antenna
The yellow dots are the samples kept after the ellipse filtering. All the axes are in ms.

the channels simultaneously. Another example is clicks due to vibration going through the antenna body. Even though hydrophones are acoustically isolated from the antenna body, sound may still reach the hydrophone by passing through the water at the end of the holding tube. The speed of sound being faster in the antenna body than in the water, the TDOA generated this way should not reside on the same ellipse distribution. Thus, TDOAs can be filtered by using their distance to the ellipse, which can be done by using the ellipsoid matrix to define a norm, (basically the square root of (3.22)). With this norm, only TDOA on the ellipse will have a norm of 1, allowing to discard any TDOA whose norm is outside of an interval around 1.

This method can be done before the DBSCAN filtering if the measured inter hydrophone distance and speed of sound are available. Otherwise, the DBSCAN filtering needs to be done first in order to remove enough spurious clicks to allow the fitting of an ellipsoid, which can be used to remove the remaining clicks. An example can be seen in Figure 5.11.

Since this four-hydrophone antenna is made of two orthogonal horizontal axes of 1.57 m length, H_1H_2 and H_3H_4. These two axes are vertically spaced by 0.61 m, with both axes centers being aligned vertically. As explained in Section 3.3.1, when two axes are orthogonal with an equal distance between hydrophones, the TDOA distribution will form a circle as in (τ_{21}, τ_{43}). However, since the sperm whales are not only in the 2D plane formed by these two axes, points will also be present inside the circle. Other tuples, such as (τ_{21}, τ_{42}) demonstrate the case where the two axes are almost collinear, leading to an ellipse with a small minor axis, and a major axis having a slope close to the ratio of the distance between the hydrophone pairs considered.

Even if each ellipse could be entirely filled, which is equivalent to saying that the 3D ellipsoid possesses points everywhere on its surface, real recordings will never provide such results. Sperm whales stay underwater which means that a surface antenna should not have points on half of its ellipsoid, except surface echo points. On a surface antenna, surface echoes are within the window of analysis of the direct click, which means that the TDOA estimation can only produce TDOA links to the surface echoes instead of the actual click if the energy of the echoes is larger than the energy of the direct click. Note that the surface echo is in opposition of phase compared to the direct click (see Section 10.3.1), meaning that their cross correlation with the direct click will have a negative sign. During the recording, the drones/boat are drifting on the surface with no direct control of their orientation. Yet, the orientation will still be quite constant as it will be imposed by the current and the wind. Thus for a small sperm whale group, the azimuth of the clicks will be restrained to a small angular interval. When the drones/boat are moving, we respect a 1 km distance with the sperm whales, which prevents the possibility of clicks coming from under the antennas. This explains why pairs of horizontal axes such as (H_1H_2, H_3H_4) still have most of their points close to the 2D case.

5.6 Tracking results for January 14th, 2020

Throughout the months of January and February 2020, multiple recordings of sperm whales were made by the Sphyrna drones. On the 14th of January, near Nice in France, a group of around 7 sperm whales were present near the drones, and sighted at the surface. The group stayed close to the drone in the afternoon between 14h and 17h. Following the pipeline that was described earlier, the clicks where first detected using the TK filter approach, then their TDOA were estimated, and they where filtered using DBSCAN filter and the ellipsoid method [142]. Finally, from the TDOA, an estimation of the position of the source was computed. The tracks of clicks that resulted from this analysis can be seen in the results in [142] and in the following Figure 5.12.

As in [143], we observe a spatial distribution of the whales, starting at 100 m. Figure 5.13 from [142] shows the histogram of the mean of all distances between the tracks for each time step, while Figure 5.14 shows the histogram of the minimum distances between the tracks for each time step, showing this 100 meters limit. In our case, the whales are spread with a usual distance of 500 meters. This corresponds to the predicted range of 470 m , obtained in Section 2.3.1 for 180 dB. Since sperm whales vary their sound level, this range will also vary. However, this could show that the sperm whales might use each others sonar to increase their individual foraging capability.

5.7 Conclusion

With the Sphyrna dataset, made of 5 channels at 384 kHz, we can extract high definition clicks. The five channels allowed for more variability within each click (small change in background noise and different interaction with the surface echoes). The estimation of tracks helps to remove more spurious clicks, making it a useful large database which can

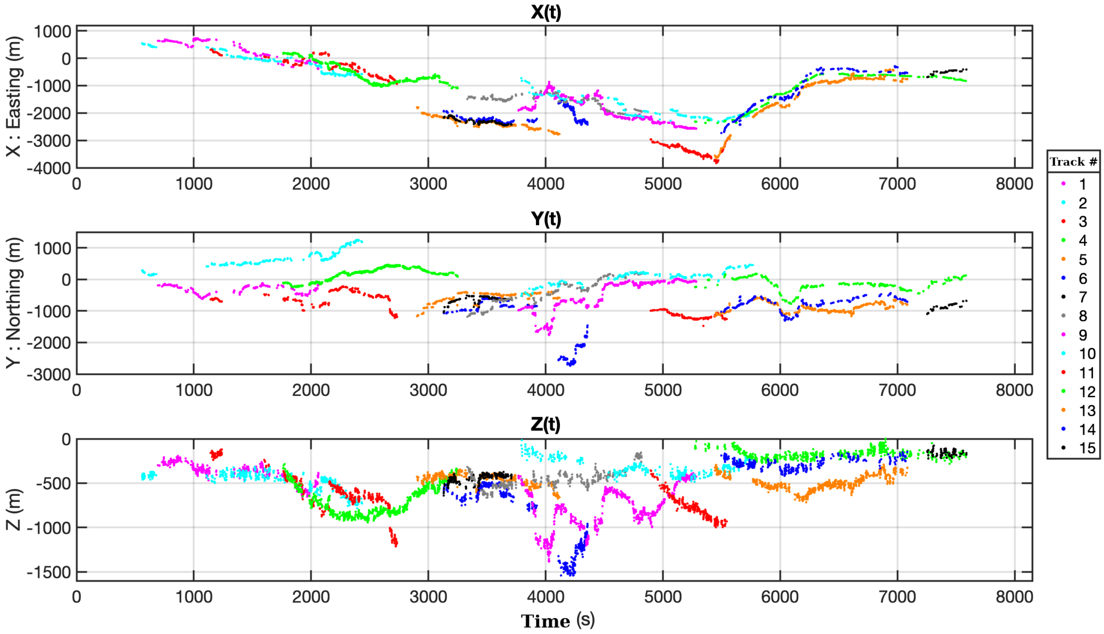


Figure 5.12: Tracking results for the 14 of January
Tracks with the same color (e.g. 1 and 9) can be distinguished by the large temporal gap between them.

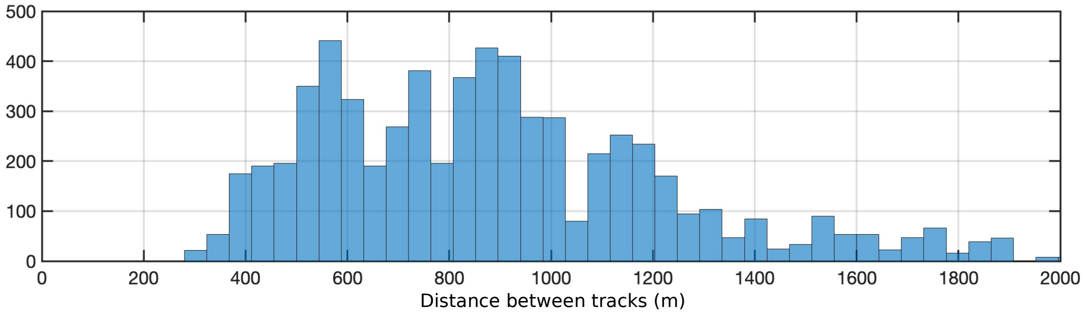


Figure 5.13: Histogram of the mean of all distances between the tracks for each time step.

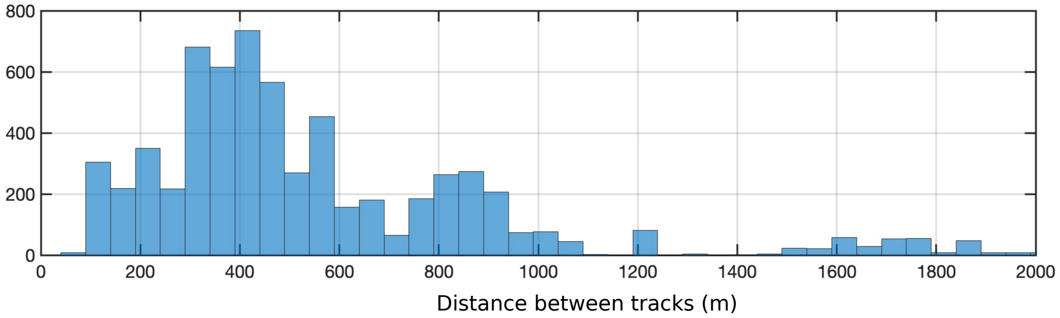


Figure 5.14: Histogram of the minimum distances between the tracks for each time step.

be used in deep learning for detection or classification tasks. The track also helps to understand the interaction between a sperm whale, the topography and the other sperm whales present in the area.

Chapter 6

Near Field Recording

After having described two databases that record sperm whale clicks far from the source, we present here a database where sperm whales were recorded at close range. We also propose a set of methods to use TDOA in order to recover the source position and the addition of video treatments in order to identify the source.

Under the authority of the Marine Megafauna Conservation Organisation directed by H. Vitry and, as part of the global program Maubydick, a team led by F. Sarano has been conducting a longitudinal study on the same group of 27 sperm whales off the west coast of the island of Mauritius, since 2013. The main goal is to understand the relationship between individuals inside the family group and the dynamics of the Mauritian sperm whale population. The main originality is that, since 2017, the data protocol is reinforced by a collaboration with H. Glotin by the use of a high sampling rate hydrophone array, Blue JASON, of SMIoT and LIS DYNI, that can record their most intimate acoustic behaviour while minimizing their disturbance.

6.1 Material

During the past years, François Sarano (see Figure 6.1) and his team have been periodically returning to the Mauritius island in order to record local sperm whales. Each year the recording protocol has been evolving to improve the data collection. Since 2017, on the initiative of H. Glotin, V. and F. Sarano have been using a GoPro Hero 4 mounted on a stereophonic acoustic antenna of our design, based on our JASON SMIoT Toulon ultra high velocity DAQ designed for these extreme recordings. Our protocol has evolved each year, with the access to additional high quality hydrophones, and the learning from past mistakes–see Table 6.1.

The hydrophones are from Cetacean Research. The DAQ is the Qualilife sound card [133]. It is able to record at a sampling rate up to 2 MHz at 16 bits per channel, up to 5 channels.

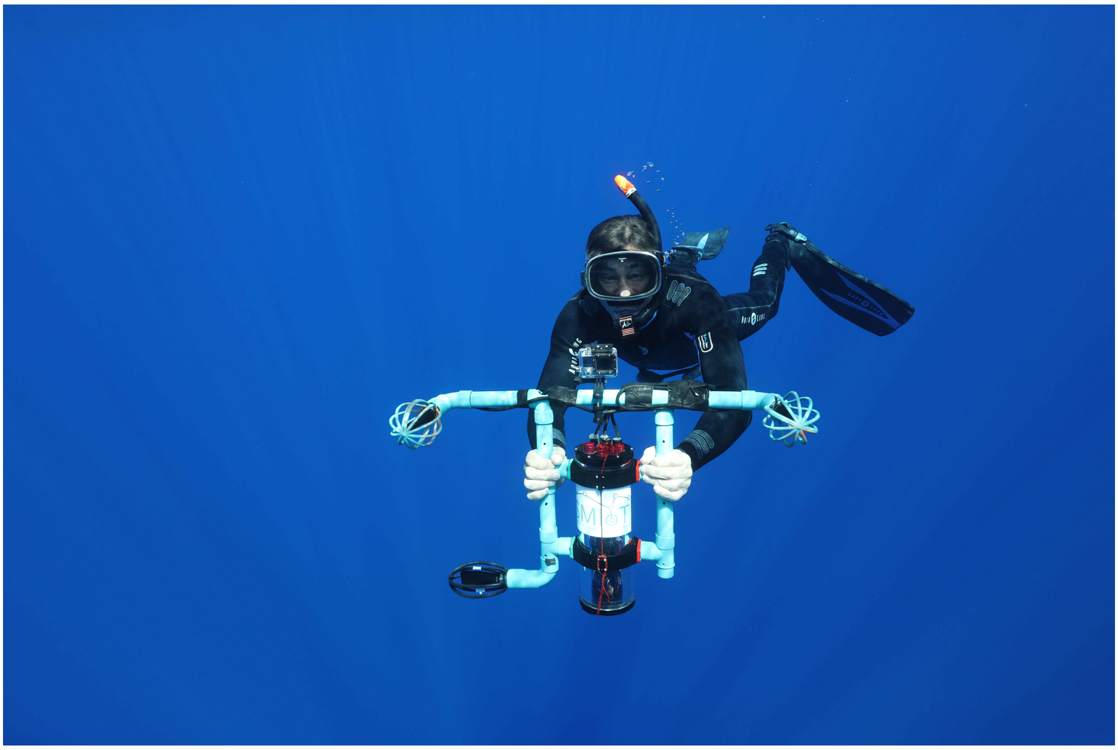


Figure 6.1: François Sarano holding the 2018 antenna (Image: F. Guerin).

6.2 From TDOA to DOA

As explained in Section 3.2, the TDOA can be used to retrieve information about the signal source position. However some hypotheses need to be considered. First of all the source is considered to be a point source, meaning that one point in space with no volume or surface is emitting the signal. Secondly, the speed of sound is considered to be homogeneous. This means that sound will have a spherical propagation from the point source, and that rays of sound are straight lines. Since N hyrdophones will only produce $N-1$ independent TDOA (see Section 3.2), only $N-1$ characteristics about the source position can be recovered. With two hydrophones, an antenna will recover the azimuth (if the hydrophones are horizontal). With three, it will additionally recover the elevation, while four hydrophones are needed to obtain the distance. However, even with four hydrophones, small errors in the value of the speed of sound or the distance between the hydrophones can be enough to prevent any useful estimation of the distance. Thus, antennas with more than four hydrophones can be useful to over constrain the problem, and reduce these errors. For an antenna made of only two or three hydrophones, a third hypothesis is added. Since the distance cannot be estimated, the source is considered to be at infinity, meaning that the sound propagation can be seen as a plane wave. If the homogeneous speed of sound hypothesis is not considered to be true, then another hypothesis can be made. This hypothesis supposes that the speed of sound is only

Year	Number of hydrophones	Model of hydrophone	Sampling rate
2017	2	C55, C57	600 kHz
2018	3	C57, C57, C75	600 kHz
2019	4	C57, C75, C57, C75	300 kHz
2020	2*5	?	300 kHz

Table 6.1: Evolution of measurement capabilities for annual campaigns off the Mauritius coast.

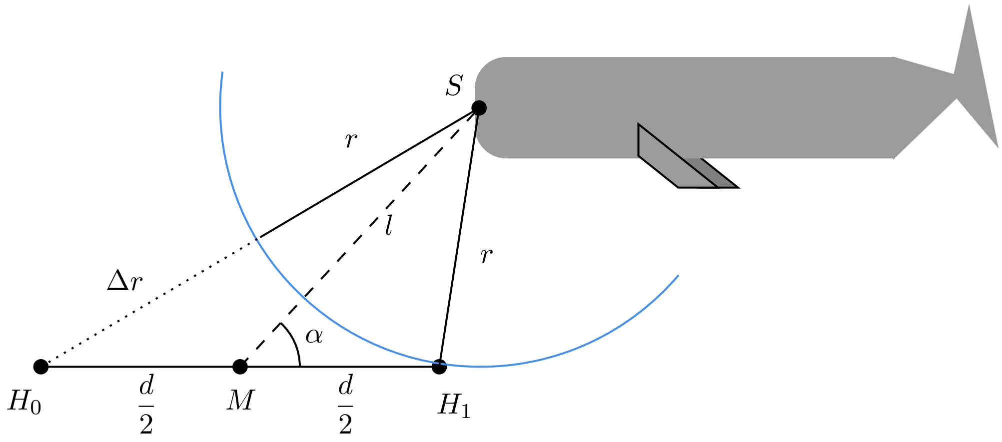


Figure 6.2: Spherical propagation of a sperm whale click

.

homogeneous around the antenna.

6.2.1 Two hydrophones

When an antenna is only composed of two hydrophones, the 3D problem can be reduced to a 2D problem using the axial symmetry of the antenna. As show in Figure 6.2, a sperm whale emits from a point S and the click will reach the hydrophones after travelling two paths of different lengths. This difference of Δr is linked to the TDOA, τ_{10}, since

$$\tau_{10} = \frac{SH_0}{c} - \frac{SH_1}{c} = \frac{r + \Delta r}{c} - \frac{r}{c} = \frac{\Delta r}{c}. \tag{6.1}$$

From the Law of Cosines, on the triangle SMH_1, the length r can be obtained,

$$r^2 = l^2 + \frac{d^2}{4} - ld\cos\alpha. \tag{6.2}$$

Similarly, for the triangle H_0MS, the edges are linked by the relation,

$$(r + \Delta r)^2 = l^2 + \frac{d^2}{4} - ld\cos(\pi - \alpha) = l^2 + \frac{d^2}{4} + ld\cos\alpha. \tag{6.3}$$

These two relations can then be combined to remove r,

$$2r\Delta r + \Delta r^2 = 4l\frac{d}{2}\cos\alpha \tag{6.4}$$

$$2\sqrt{l^2 + \frac{d^2}{4} - ld\cos\alpha}\Delta r + \Delta r^2 = 2ld\cos\alpha. \tag{6.5}$$

Solving this polynomial in Δr and keeping the non negative term gives

$$\Delta r = l\left(\sqrt{1 + \frac{d^2}{4l^2} + \frac{d}{l}\cos\alpha} - \sqrt{1 + \frac{d^2}{4l^2} - \frac{d}{l}\cos\alpha}\right). \tag{6.6}$$

The series expansion at $\frac{d}{l} = 0$ is

$$\Delta r = d\cos\alpha - \frac{d^3}{l^2}\cos\alpha\sin^2\alpha + \mathrm{o}\left(\frac{d^4}{l^3}\right) = d\cos\alpha\left(1 - \frac{d^2}{l^2}\sin^2\alpha\right) + \mathrm{o}\left(\frac{d^4}{l^3}\right). \tag{6.7}$$

The first order term $d\cos\alpha$ is independent of the distance l, hence allowing an estimation of the angle α without knowledge of the distance. When the source is at infinity (meaning that l tends to ∞), then only the first order term remains,

$$\Delta r = d\cos\alpha. \tag{6.8}$$

In practice, $l = 5d$ is enough to use this approximation. One of the main consequences of (6.8), is that the antenna will have less precision when the sperm whale is close to the line (H_0H_1), meaning that α is either close to 0 or π, which is equivalent to the absolute value of the TDOA being close to its maximal value. Figure 6.3 shows the TDOA probability density function when the the angle α is taken from a uniform distribution. The TDOA here is normalised by the maximum TDOA, $\frac{d}{c}$.

6.2.2 Three hydrophones

An antenna with three hydrophones will have its sensors defining a plane. To simplify the equations, let us define a coordinate axis whose origin is at the center of $[H_0, H_1]$, its x-axis proportional to $\overrightarrow{H_0H_1}$, and the y-axis belongs to the plane $H_0H_1H_2$. In this coordinate system, the hydrophones H_0, H_1 and H_2 are at $\begin{pmatrix}-d_0\\0\\0\end{pmatrix}$, $\begin{pmatrix}d_0\\0\\0\end{pmatrix}$, $\begin{pmatrix}d_1\\d_2\\0\end{pmatrix}$ respectively. A click emitted at a point S located at $\begin{pmatrix}\rho\sin\alpha\sin\beta\\\rho\cos\beta\\\rho\cos\alpha\sin\beta\end{pmatrix}$ will generate the following TDOA,

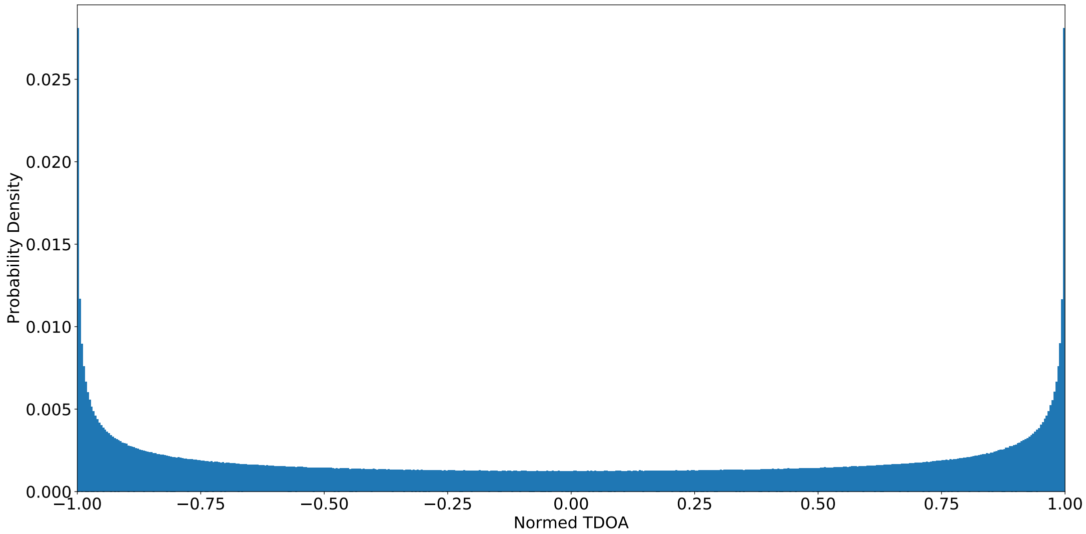


Figure 6.3: TDOA probability density function.

$$
\begin{aligned}
\tau_{10} &= \frac{SH_0 - SH_1}{c} \\
&= \frac{\sqrt{(\rho\sin\alpha\sin\beta + d_0)^2 + \rho^2\cos^2\beta + \rho^2\cos^2\alpha\sin^2\beta}}{c} \\
&\qquad - \frac{\sqrt{(\rho\sin\alpha\sin\beta - d_0)^2 + \rho^2\cos^2\beta + \rho^2\cos^2\alpha\sin^2\beta}}{c}, \qquad (6.9) \\
\tau_{20} &= \frac{SH_0 - SH_2}{c} \\
&= \frac{\sqrt{(\rho\sin\alpha\sin\beta + d_0)^2 + \rho^2\cos^2\beta + \rho^2\cos^2\alpha\sin^2\beta}}{c} \\
&\qquad - \frac{\sqrt{(\rho\sin\alpha\sin\beta - d_1)^2 + (\rho\cos\beta - d_2)^2 + \rho^2\cos^2\alpha\sin^2\beta}}{c}. \qquad (6.10)
\end{aligned}
$$

Similarly to the two-hydrophone problem, one can use the series expansion at $\frac{d_0}{\rho} = 0$ to obtain

$$
\tau_{10} = 2\frac{d_0}{c}\sin\alpha\sin\beta + \frac{d_0^3}{c\rho^2}\sin\alpha\sin\beta(\sin^2\alpha\sin^2\beta - 1) + \mathrm{o}(\frac{d_0^4}{\rho^3}), \qquad (6.11)
$$

$$
\begin{aligned}
\tau_{20} &= \frac{d_0 + d_1}{c}\sin\alpha\sin\beta + \frac{d_2}{c}\cos\beta \\
&- \frac{-(d_0^2 - d_1^2)\sin^2\alpha\sin^2\beta - d_1 d_2\sin\alpha\sin(2\beta) - d_2^2\cos^2(\beta) - d_0^2 + d_1^2 + d_2^2}{c\rho} + \mathrm{o}(\frac{d_0^2}{\rho}).
\end{aligned} \qquad (6.12)
$$

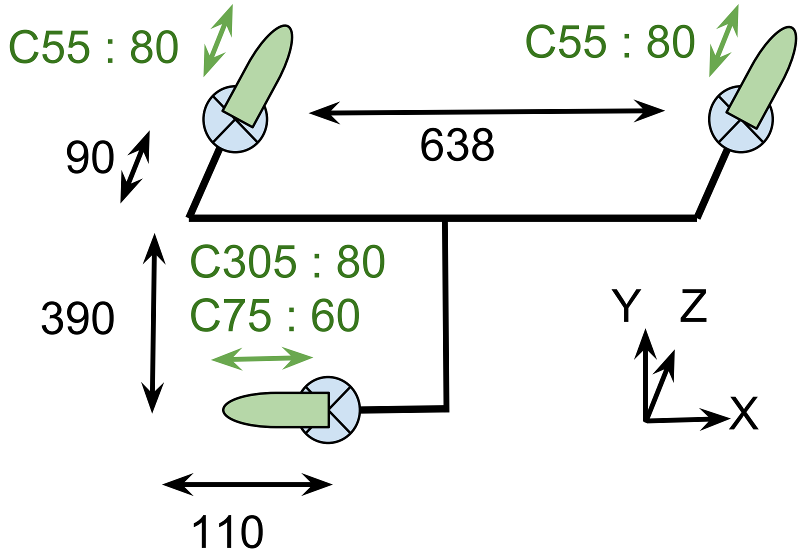


Figure 6.4: Blueprint of the 2018 antenna.

Once again, when the distance to the source ρ is large compared to the parameter d_0, only the first order term remains, which is also independent of ρ,

$$\tau_{10} = 2\frac{d_0}{c}\sin\alpha\sin\beta, \tag{6.13}$$

$$\tau_{20} = \frac{d_0+d_1}{c}\sin\alpha\sin\beta + \frac{d_2}{c}\cos\beta. \tag{6.14}$$

One may remark that the first terms of (6.13) and (6.14) are similar, and that the correlation between τ_{10} and τ_{20} is proportional to the scalar product $< \overrightarrow{H_0H_1} | \overrightarrow{H_0H_2} >$ (see Section 3.3 for details). Indeed when $d_1 = -d_0$, $\overrightarrow{H_0H_1}$ and $\overrightarrow{H_0H_2}$ are orthogonal, meaning that their scalar product is null. Thus when designing an antenna, one may want to use this orthogonality, if the only TDOAs used are τ_{10} and τ_{20}. However, note that other methods might prefer the hydrophones to form an equilateral triangle, such that the information is equally spread between τ_{10}, τ_{20} and τ_{21}.

Finally, to obtain the DOA from the TDOA, (6.13) and (6.14) need to be inverted, which results in the following relations

$$I = \sqrt{d_0^2(\tau_{10}-2\tau_{20})^2 - 2d_0d_1\tau_{10}(\tau_{10}-2\tau_{20}) - d_1^2\tau_{10}^2 + 4\frac{d_0^2d_2^2}{c^2}}, \tag{6.15}$$

$$\alpha = -\arcsin\left(\frac{d_2\tau_{10}}{I}\right), \tag{6.16}$$

$$\beta = 2\arctan\left(\frac{2d_0\tau_{20} - (d_1+d_0)\tau_{10} - 2\frac{d_2d_0}{c})}{I}\right). \tag{6.17}$$

6.2.3 Four hydrophones or more

For the problem of n hydrophones ($n \geq 4$), let us denote $\begin{pmatrix} x_i \\ y_i \\ z_i \end{pmatrix}$ the cartesian coordinates of the hydrophone H_i, t_i the click's TOA for the same hydrophone, and let us use the subscript $_S$ for the source. We compute the difference

$$\begin{aligned} d_{0S}^2 - d_{iS}^2 &= (x_0 - x_S)^2 + (y_0 - y_S)^2 + (z_0 - z_S)^2 - (x_i - x_S)^2 - (y_i - y_S)^2 - (z_i - z_S)^2 \\ &= x_0^2 - x_i^2 + y_0^2 - y_i^2 + z_0^2 - z_i^2 + 2\left(x_S(x_0 - x_n) + y_S(y_0 - y_n) + z_S(z_0 - z_n)\right). \end{aligned} \tag{6.18}$$

The left term of this equation can be developed using the TDOA τ_{0i} since

$$\begin{aligned} d_{0S}^2 - d_{iS}^2 &= d_{0S}^2 - c^2(t_i - t_s)^2 \\ &= d_{0S}^2 - c^2(t_i - t_0 - (t_s - t_0))^2 \\ &= d_{0S}^2 - (d_{0S} + c\tau_{0i})^2 \\ &= -2d_{0S}c\tau_{0i} - c^2\tau_{0i}^2. \end{aligned} \tag{6.19}$$

By grouping the terms that depend on the source position, but not the TDOA, on the left side, and the others on the right side, the equations for all the hydrophones can be summarized as the following matrix equation,

$$\begin{aligned} &\begin{pmatrix} x_0 - x_1 & y_0 - y_1 & z_0 - z_1 & -c\tau_{01} \\ x_0 - x_2 & y_0 - y_2 & z_0 - z_2 & -c\tau_{02} \\ x_0 - x_3 & y_0 - y_3 & z_0 - z_3 & -c\tau_{03} \\ \dots & \dots & \dots & \dots \\ x_0 - x_n & y_0 - y_n & z_0 - z_n & -c\tau_{0n} \end{pmatrix} \begin{pmatrix} x_S \\ y_S \\ z_S \\ d_{0S} \end{pmatrix} \\ &\qquad = \frac{1}{2} \begin{pmatrix} x_0^2 - x_1^2 + y_0^2 - y_1^2 + z_0^2 - z_1^2 - c^2\tau_{01}^2 \\ x_0^2 - x_2^2 + y_0^2 - y_2^2 + z_0^2 - z_2^2 - c^2\tau_{02}^2 \\ x_0^2 - x_3^2 + y_0^2 - y_3^2 + z_0^2 - z_1^2 - c^2\tau_{03}^2 \\ \dots \\ x_0^2 - x_n^2 + y_0^2 - y_n^2 + z_0^2 - z_n^2 - c^2\tau_{0n}^2 \end{pmatrix}. \end{aligned} \tag{6.20}$$

If the first hydrophone is placed at the origin, then the equation can be simplified as

$$\begin{pmatrix} x_1 & y_1 & z_1 & c\tau_{01} \\ x_2 & y_2 & z_2 & c\tau_{02} \\ x_3 & y_3 & z_3 & c\tau_{03} \\ \dots & \dots & \dots & \dots \\ x_n & y_n & z_n & c\tau_{0n} \end{pmatrix} \begin{pmatrix} x_S \\ y_S \\ z_S \\ d_{0S} \end{pmatrix} = \frac{1}{2} \begin{pmatrix} x_1^2 + y_1^2 + z_1^2 + c^2\tau_{01}^2 \\ x_2^2 + y_2^2 + z_2^2 + c^2\tau_{02}^2 \\ x_3^2 + y_3^2 + z_1^2 + c^2\tau_{03}^2 \\ \dots \\ x_n^2 + y_n^2 + z_n^2 + c^2\tau_{0n}^2 \end{pmatrix}. \tag{6.21}$$

This equation is of the form $A\overrightarrow{x} = \overrightarrow{b}$, which is solved by using a least square regression, i.e. minimizing $\|A\overrightarrow{x} - \overrightarrow{b}\|$. For four hydrophones, the problem is under-determined since the rank of A is at most 3. This can be solved by adding the constraint that $d_{0S}^2 = (x_0 - x_S)^2 + (y_0 - y_S)^2 + (z_0 - z_S)^2$.

Alternatively, one can fit the TOA model, which is

$$t_i = t_S + \frac{\sqrt{(x_i - x_S)^2 + (y_i - y_S)^2 + (z_i - z_S)^2}}{c}. \tag{6.22}$$

This is a nonlinear regression problem where (t_S, x_S, y_S, z_S) needs to be found to fit the n data points (t_i, x_i, y_i, z_i). Note that in this version of the problem, there is one more variable t_S to find, but there is also one more equation with the hydrophone H_0 (using $i = 0$ does not give $0 = 0$ as for previous methods).

In order to have good starting values, one can estimate the DOA using the three-hydrophone estimation, and combine it with the mean expected distance at which the sperm whale is supposed to be. Otherwise the previous least square regression can be used.

6.3 Plotting clicks on the video

Once the DOAs have been obtained, they can be used in conjunction with the GoPro video to identify the sperm whale that emitted the click. If the distance has been estimated, it can be used to obtain the DOA relative to the GoPro, instead of the origin of the base in which the DOA have been estimated (usually placed on H_0). If the distance was not estimated, then one will have to assume that the distance is large enough, such that the DOA relative to the GoPro stays the same.

6.3.1 Fisheye effect

The GoPro uses a fisheye lens. It allows to film a higher angle of view, at the cost of distortion. Straight lines will thus appear curved with a fisheye lens, unlike how they would have been if a convex lens (regular camera lens) was used. As described in Figure 6.5, a convex lens will focus rays from a faraway object on plane space from the lens of exactly its focal length f. On this plane, the focus point will be at a distance d from the optical axis, if the object rays make an angle θ with the optical axis.

Thus d follows the relation

$$d = f \tan(\theta). \tag{6.23}$$

Fisheye lenses do not follow this relation. There are multiple technologies of fisheye. In the case of the GoPro we used, the relation is

$$d = f \tan\left(\frac{\theta}{2}\right). \tag{6.24}$$

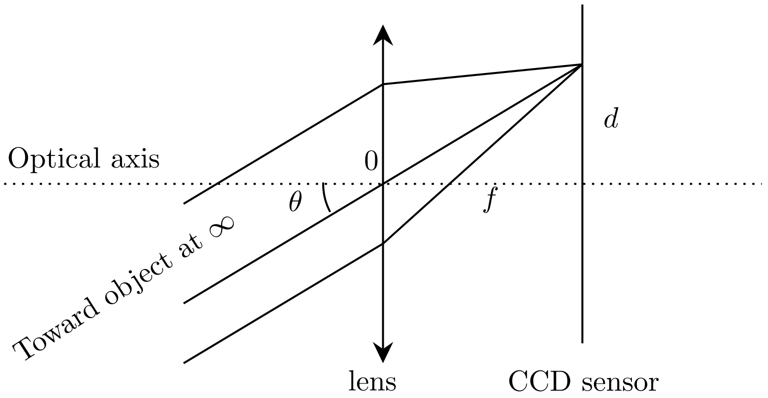


Figure 6.5: Ray diagram of a convex lens.

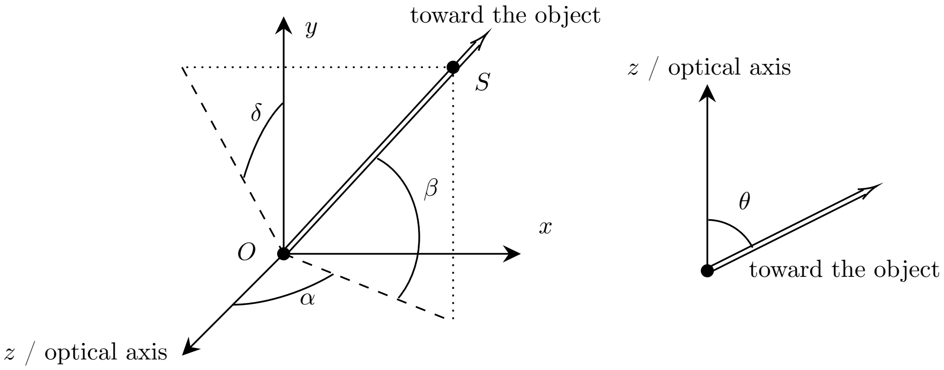


Figure 6.6: Spherical coordinate system for GoPro image analysis.

6.3.2 DOA to pixel coordinate

Usually the DOA is expressed in a spherical coordinate system, which would use (α, β) as shown in Figure 6.6. However the fisheye lens relation requires another spherical coordinate system (θ, δ), since θ is the parameter involved in the fisheye distortion.

By assuming that the point S is on the plane $z = f$, the vector $\overrightarrow{OS}$ equates to

$$\overrightarrow{OS} = \begin{pmatrix} f \tan\alpha \\ f \frac{\tan\beta}{\cos\alpha} \\ f \end{pmatrix} = f \begin{pmatrix} \tan\alpha \\ \frac{\tan\beta}{\cos\alpha} \\ 1 \end{pmatrix}. \tag{6.25}$$

Since the GoPro has limited field of view (FOV), this hypothesis covers all the possible pairs (α, β), where $\cos\alpha > 0$. Then

$$\cos\theta = \frac{f}{\|\overrightarrow{OS}\|} = \frac{f}{f\sqrt{(\tan^2\alpha + \frac{\tan^2\beta}{\cos^2\alpha} + 1}} = \frac{\cos^2\alpha}{\sqrt{\tan^2\alpha\cos^2\alpha + \tan^2\beta + \cos^2\alpha}} \tag{6.26}$$

$$= \frac{\cos^2\alpha}{\sqrt{\sin^2\alpha + \tan^2\beta + \cos^2\alpha}} = \frac{\cos^2\alpha}{\sqrt{1 + \tan^2\beta}}. \tag{6.27}$$

Note that this can be easily expanded to other values of α, for example by considering the plane $z = -f$, which would give the same equation in the end. We also find,

$$\cos\delta = \frac{f\tan\alpha}{f\sqrt{\tan^2\alpha + \frac{\tan^2\beta}{\cos^2\alpha}}} = \frac{\sin(\alpha)}{\sqrt{\sin(\alpha)^2 + \tan(\beta)^2}}, \tag{6.28}$$

$$\sin\delta = \frac{f\frac{\tan\beta}{\cos\alpha}}{f\sqrt{\tan^2\alpha + \frac{\tan^2\beta}{\cos^2\alpha}}} = \frac{\tan(\beta)}{\sqrt{\sin(\alpha)^2 + \tan(\beta)^2}}. \tag{6.29}$$

GoPro does not give the value of f, but the the value of the horizontal and vertical field of view, respectively fov_h and fov_v. If the width and the height of the output image pixel are denoted w and h, then

$$x = \frac{\tan(\frac{\theta}{2})}{\tan(\frac{\mathrm{fov}_h}{4})} \frac{w}{2} \cos\delta + \frac{w}{2}, \tag{6.30}$$

$$y = \frac{\tan(\frac{\theta}{2})}{\tan(\frac{\mathrm{fov}_v}{4})} \frac{h}{2} \sin\delta + \frac{h}{2}, \tag{6.31}$$

which when combined with the previous equation give

$$\theta = \arccos\left(\frac{\cos(\alpha)}{\sqrt{1+\tan(\beta)^2}}\right), \tag{6.32}$$

$$x = \frac{\tan(\frac{\theta}{2})}{\tan(\frac{\text{fov}_h}{4})}\frac{\sin(\alpha)}{\sqrt{\sin(\alpha)^2+\tan(\beta)^2}}\frac{w}{2}+\frac{w}{2}, \tag{6.33}$$

$$y = \frac{\tan(\frac{\theta}{2})}{\tan(\frac{\text{fov}_v}{4})}\frac{\tan(\beta)}{\sqrt{\sin(\alpha)^2+\tan(\beta)^2}}\frac{h}{2}+\frac{h}{2}. \tag{6.34}$$

However (6.33) and (6.34) use the characteristic of the fisheye lens when used in air. Thus, Snell's law needs to be used in order to convert the θ_{water} obtained using (6.32) and the DOA into a θ_{air} that can be use in (6.33) and (6.34),

$$\theta_{\text{air}} = \arcsin(1.345\sin(\theta_{\text{water}})). \tag{6.35}$$

Since there are a finite number of pixels, it is better to convert each pixel coordinate into the corresponding angle. This way, confidence intervals around a DOA will also be plotted with the fisheye distortion, instead of being a circle with a fixed pixel diameter. A similar fact can be said regarding the plotting of DOA containing only the azimuth. In order to obtain the Pixel/angle map, the conversion formula needs to be inverted. We find,

$$x' = \frac{(x-\frac{w}{2})\tan(\frac{\text{fov}_h}{4})}{\frac{h}{2}}, \tag{6.36}$$

$$y' = \frac{(y-\frac{h}{2})\tan(\frac{\text{fov}_v}{4})}{\frac{h}{2}}, \tag{6.37}$$

$$\theta = \arcsin\left(\frac{1}{1.345}\sin(2\arctan(\sqrt{x'^2+y'^2}))\right), \tag{6.38}$$

$$\delta = \text{arctan2}(y', x'), \tag{6.39}$$

$$\alpha = \arctan(\tan(\theta)\cos(\delta)), \tag{6.40}$$

$$\beta = \arctan(\tan(\theta)\sin(\delta)\cos(\alpha)). \tag{6.41}$$

Note that minus signs might be needed depending on the orientation of the pixel axes.

Figure 6.7 shows an example of a GoPro picture of a graph paper. The lines of the paper are distorted, as well as the blackboard border. In Figure 6.8, the pixel coordinates have been stretched using the previously described equations ((6.36) to (6.41) with the omission of the indices of water) to obtain their corresponding angle (θ), and then coverting back to the pixel space using (6.23).

Figure 6.9 shows an example of a frame from a GoPro video of sperm whale where the DOA of the clicks have been plotted. Each point (DOA of a click) stays for 7 frames (starting from the frames for which the corresponding click is earmarked) on the video to make them easier to see. However the antenna does not have the means to measure its

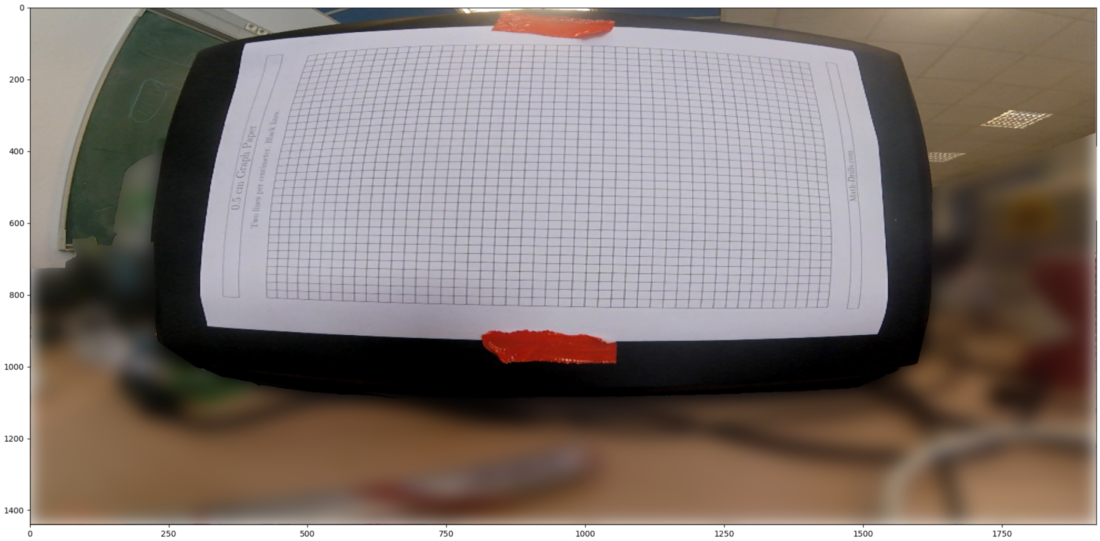


Figure 6.7: GoPro picture of a graph paper

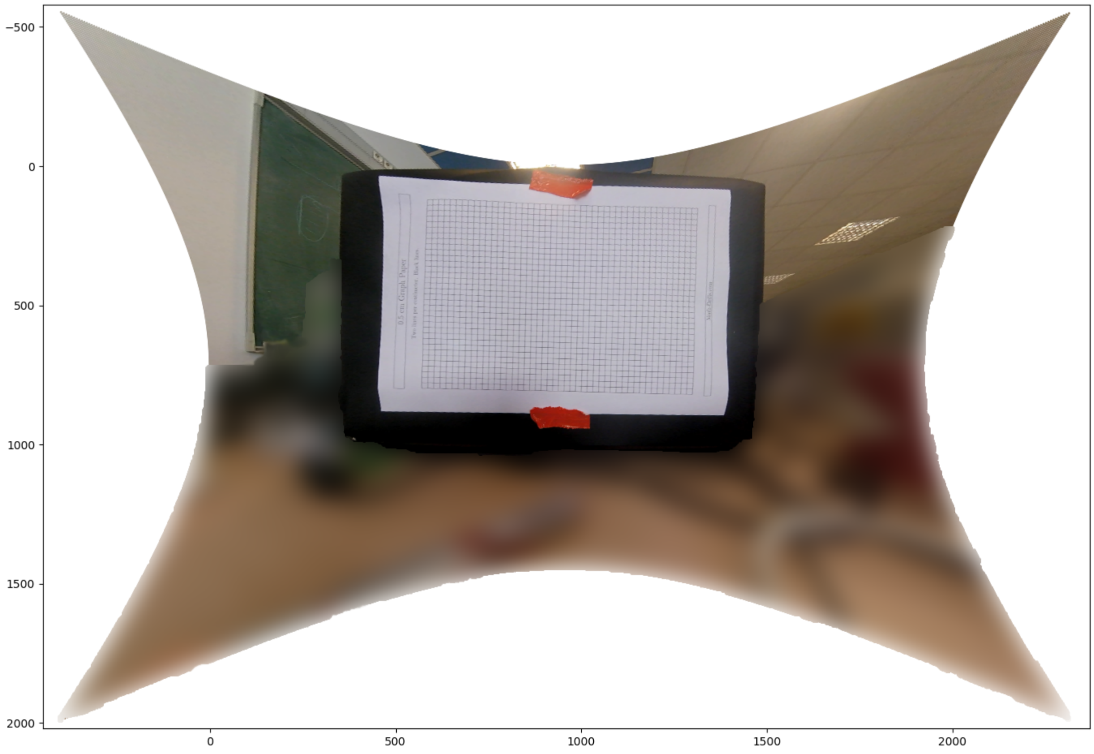


Figure 6.8: Removal of the fish eye effect in Figure 6.7

(Image credit = F. Sarano) (Process = M. Ferrari, H. Glotin, R.Marxer, M. Asch @ DGA LAMFA DYNI LIS Toulon Univ France)

Figure 6.9: Frame from a GoPro video with DOA of click (The red dots)
The black shape inside the white rectangle represent the FOV of the GoPro in the (α, β) space, both scaled to $(-\frac{\pi}{2}, \frac{\pi}{2})$. DOAs of the current frame are also plotted inside this space.

rotation in space, which means that every oscillation (which is strong due to waves) will shift the scene. Seven frames is already long enough for a point to give the impression that it is located where it should not be, when it was in fact in the right place in the first frames in which it was displayed. The black shape inside the white rectangle represents the FOV of the GoPro in the (α, β) space, both scaled to $(-\frac{\pi}{2}, \frac{\pi}{2})$. The DOAs of the current frame are also plotted inside this space. As a sperm whale swims out of the FOV of the GOPRO, this secondary plot can show to a viewer that the clicks heard are not plotted on the video since they come from outside of the GoPro FOV.

6.4 Application of Image Segmentation Techniques for Detecting Individuals

The previous section showed how to find the source position of a click, and with it, how to use this information along with a video in order to associate a click with an individual. However labeling each click manually, by looking at the video to know which sperm shale is behind each dot plotted can be quiet tedious. A less tedious alternative is to annotate each whale once per video, and let the computer annotate the click. In order to have the sperm whale label spread across all the frame of the video, one can use an instance segmentation tool. For a frame it will attribute a color to each sperm whale. The color consistency between frames can be achieved by comparing the segmentation, hence attributing each sperm whale a color, thus a label. Even if the color changes in the video, the amount of annotation work will still be greatly reduced.

Image segmentation of sperm whales is not an easy task. Unlike pictures of dogs sitting in grass, which have dissimilar colors, pictures of sperm whales extracted from MP4 video have all their colors contained in a small range, and with and underwater visibility around 20 m and divers trying to stay at a distance from the animals, sperm whales are often seamlessly blended with the background.

For this task we decided to use DeepLabV3+ [144] which gives state of the art performance. The other advantage of this choice was this network's ability to achieve satisfying results on a small training set of 10 images. Instead of spending time creating a dataset of thousands of annotated images, a loop can be made where the network is trained on the small dataset, then the trained network is used to annotate the whole dataset of unannotated images. Dozens of images with clear failure cases (e.g. whole or large part of an animal classified as the background, or scuba divers classified as a sperm whale) are then chosen to be annotated. The network is then retrained on this augmented dataset. After a few iterations of this process, the whole dataset is used during training, including the network's annotated images. Preliminary results are shown in Figure 6.10.

A test set of images from recording sessions made during days that differ from the days used for training, was also used to monitor the network performance.

The main failure cases seen during this iterative process were due to the presence of colorful fish/algae. Since the presence of non shade of blue colors was only seen in a picture taken outside of water, the network reaction was to label a large zone around the fish as background, even if sperm whales were present behind them. The other failure

Figure 6.10: Image (left) with its semantic segmentation (middle) and its instance segmentation (right)

.

cases were due to the networking struggling to fully label flukes, with most of the time flukes being disconnected from the rest of the body.

6.5 Conclusion

With this final out of three databases, closeup recorded clicks are observed. While being recorded with a high frequency similar to the Sphyrna database, they differ by the new context in which they were recorded. Indeed, the Sphyrna database is mostly composed of echolocation clicks, while the Sarano database mostly contains socializing clicks, which are composed of codas and creaks. The presence of this social aspect is also magnified by the video signal, the use of which allows clicks to be associated with their emitters. This association is what allows us to link behaviour that occurs after an exchange of coda between one or more individuals.

Part II

Machine Learning for Bioacoustic Signals

In this part, made up of two chapters we will apply novel machine learning techniques to the analysis of the recorded sperm whale calls.

Machine learning is becoming an indispensable tool, and in [145] one can find an excellent survey of the recent advances and transformative potential of machine learning (ML), including deep learning, in the field of underwater acoustics.

Chapter 7

Overview of Machine Learning Techniques for Bioacoustics

7.1 Click Classification

Passive acoustic monitoring is today a common approach for biodiversity monitoring. Its efficiency relies on a large dataset, and thus reliable automatic detection of species. This paper deals with a particular type of emission, transients from odontocetes, which are short-duration wide-band impulses. We will present a case study, the CARI'MAM project, and describe how a reference dataset could be built for such monitoring. Then we propose a novel approach for click classification based on an End-to-End CNN model.

The CARI'MAM project aims to create a network of Marine Protected Area Managers spread across the whole Caribbean sea for the conservation of marine mammals. In order to survey the distribution of marine mammals, a mono-hydrophone system was to be deployed this spring during 40 days in 20 different locations, but the deployment have been delayed . The amount of data collected will be too large to analyse manually. To prepare for this analysis, we created a first dataset made of clicks from the various species present in the Caribbean. The proposed dataset contains 10 out of the 30 species that the CARI'MAM project aims to study. This first corpus will allow us to test the different techniques of semi- or fully automated analysis as well as train preliminary deep learning models to solve the classification task. This dataset is also distributed as a benchmark for click classification in the DOCC10 (Dyni Odontocete Click Classification) challenge.[1]

To build a dataset large enough to train neural networks we gathered data from different sources: i) the 2018 DLCDE challenge[2], created by HILDEBRAND John, GLOTIN Hervé, FRASIER kait, ROCH Marie et al. [146], and ii) sperm whale clicks from the 2018 Sphyrna Odyssey expedition [132]. These existing sets contain long sequences of audio with rough annotations of the temporal regions with clicks. Our goal is to produce a set with individual clicks associated to a particular species. In this work we present

---

[1] `https://challengedata.ens.fr/participants/challenges/32/`

[2] `http://sabiod.univ-tln.fr/DCLDE/challenge.html`

our methodology to extract the clicks and label them with the species identity. We also present a preliminary analysis of the resulting corpus, a data split useful for benchmarking and a baseline deep learning model to classify the clicks. Even though our method to extract clicks and labels may induce some label noise, this is a situation encountered in a real scenario, thus increasing the ecological validity of the dataset. Furthermore this permits exploring the use of techniques specifically dealing with these issues, such as negative learning [147, 148]. We thus decided to increase the number of samples, at the cost of a possible increase of mislabeling.

7.2 Construction of the DOCC10 dataset

7.2.1 2018 DCLDE challenge

The high-frequency dataset from the 2018 DCLDE challenge [146] consists of marked encounters with echolocation clicks of species commonly found along the US Atlantic Coast and in the Gulf of Mexico:

- Mesoplodon europaeus - Gervais' beaked whale
- Ziphius cavirostris - Cuvier's beaked whale
- Mesoplodon bidens - Sowerby's beaked whale
- Lagenorhynchus acutus - Atlantic white-sided dolphin
- Grampus griseus - Risso's dolphin
- Globicephala macrorhynchus - Short-finned pilot whale
- Stenella sp. - Stenellid dolphins
- Delphinid type A
- Delphinid type B

The goal for the DCLDE dataset is to identify the times at which echolocating individuals of a particular species approached the area covered by the sensors. Analysts examined the data in search of echolocation clicks and approximated the start and end times of acoustic encounters. Any period that was separated from another by five minutes or more was marked as a separate encounter. Whistle activity was not considered. Consequently, while the use of whistle information during echolocation activity is appropriate, reporting a species based on whistles in the absence of echolocation activity would be considered a false positive for this classification task.

Data were recorded at different locations in the Western North Atlantic and Gulf of Mexico as shown in Figure 7.1. In the accompanying Table 7.1, we list the coordinates and depths of the various sites. These data were collected between 2011 and 2015, and the time period for each recording can be inferred directly from the data.

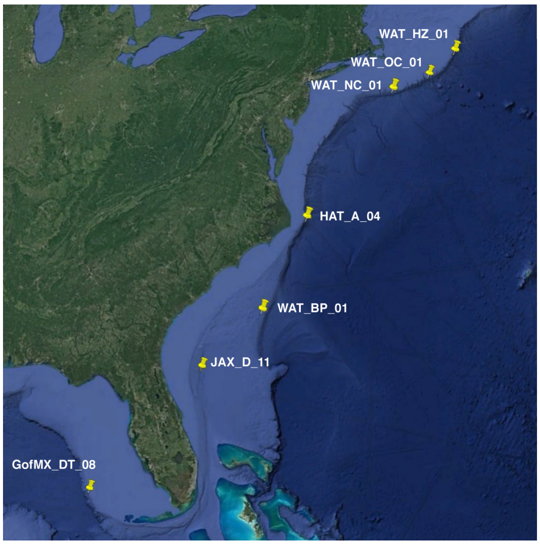


Figure 7.1: Recording locations of the 2018 DCLDE challenge

7.2.2 Enhancing the weak labels of DCLDE 2018

For each of the 9 species contained in the DCLDE dataset, the labels are lapse of time indicating the presence of the corresponding species. The longest interval between two clicks in a segment can last up to 5 minutes. We consider these labels as weak, in the sense that they do not reflect precisely the timestamp of each click. We are interested in the detection and classification of the individual clicks, therefore annotations at a much finer temporal scale are required. These are the labels that we will refer to as *strong*.

In order to extract strong labels from the weak DCLDE 2018 weak labels, we first retain only energy components in the frequency ranges of the clicks by applying a bandpass filter. After this filtering step, we use a Teager-Kaiser (TK) filter [109, 110] combined with a local maximum extractor having a half window length of 0.02 s, to obtain the position of all these clicks. Since most of the maxima will not be actual clicks but background noise, a median filter is used on the logarithms of these maxima to evaluate the background noise level. Any maxima above the noise level plus 0.5 dB are kept. Windows of 8192 samples are then extracted around these clicks.

We then proceed to label these maxima with the labels from the DCLDE challenge. If a click is in the interval of two or more weak labels, we assign it all of the corresponding labels. We also extract multiple acoustic features to curate the new DOCC10 dataset from mislabeled clicks. One must note that in the DCLDE data the clicks of all present species are not labeled. There may be segments labeled as containing a single species that contain clicks from other species that are not part of the DCLDE label set, such as sperm whales. We decided to use the spectral centroid as the feature to perform the final filtering, since it is the feature with which the outliers are better distinguishable from actual clicks. The spectral centroid is the weighted mean of the frequency, using the Fourier transform amplitude as weights.

Project	Site	Deployment	Preamp	Lat N	Long W	Depth
WAT	HZ	1	734	41-03.7	66-21.1	850
WAT	OC	1	707	40-15.8	67-59.2	1100
WAT	NC	1	740	39-49.9	69-58.9	980
HAT	A	4	685	35-20.8	74-50.9	840
WAT	BP	1	810	32-06.4	77-05.7	945
JAX	D	11	681	30-09.0	79-46.2	800
GofMX	DT	8	638	25-32.3	84-37.9	1200

Table 7.1: DCLDE recording meta data

The spectral centroid is however not useful to classify clicks on its own, as most of the DCLDE species will have clicks with similar spectral centroids, mainly in the range 30 kHz - 40 kHz. Thus it cannot be used to chose one label for clicks that have multiple labels.

7.2.3 Sphyrna Odyssey expedition data

Many applications, such as the one faced in the CARI'MAM project, require the detection of species not available in the DCLDE 2018 dataset. We consider the possibility of mixing data from different recording experiments into the corpus. In our case we use data obtained from the 2018 Sphyrna Odyssey expedition. This set contains clicks from sperm whales, *Physeter Macrocephalus*. All the clicks are from a single sperm whale 3 hour encounter.

The clicks were recorded at 300 kHz by a Cetacean Research C57 hydrophone and JASON sound card from SMIoT UTLN. The sperm whale clicks were detected using a detection process similar to the one used to create strong labels from the DCLDE dataset. We cross-correlated the signal with one period of a 12.5 kHz sine wave which acts as a band-pass filter (bandwidth of echolocation clicks is 10–15 kHz [88]). We then apply a Teager-Kaiser filter [109, 110] and extract the local maxima in 20 ms windows (twice the largest inter-pulse interval of 10 ms [120]). For each 1 minute audio file we compute the mean and standard deviation of the maxima values in decibels (dB), and only keep samples over three times the standard deviation [111]. To incorporate them in DOCC10, we down-sampled the signal at 200 kHz to match the sampling rate of the DCLDE datset.

Since the data added contain a single unseen species, we are introducing a bias of high correlation between recording configuration, environment and the species label. However this can be seen as a usual approach to composing bioacoustics datasets for

Label	Scientific name	Common name
Gg	Grampus griseus	Risso's dolphin
Gma	Globicephala macrorhynchus	Short-finned pilot whale
La	Lagenorhynchus acutus	Atlantic white-sided dolphin
Mb	Mesoplodon bidens	Sowerby's beaked whale
Me	Mesoplodon europaeus	Gervais' beaked whale
Pm	Physeter macrocephalus	Sperm whale
Ssp	Stenella sp.	Stenellid dolphins
UDA		Delphinid type A
UDB		Delphinid type B
Zc	Ziphius cavirostris	Cuvier's beaked whale

Table 7.2: Class labels

machine learning and will evidence the issues with such a method in the benchmark.

7.2.4 DOCC10 challenge

The new DOCC10 dataset consists of clicks centered in a window of 8192 samples. This was motivated by the possibility of analysing clicks in a window of 4096 samples while being able to offset this shorter window. The combination of DCLDE and Sphyrna Odyssey brought this new dataset to a total count of $134,080$, that we split into a training set of $113,120$ clicks and a test set of $20,960$ clicks for the DOCC10 challenge, which produces an approximately 85-15 split. The test set is balanced with 2096 clicks per class. For the challenge, the test set was split into a private test set (90%) and a public test set (10%). This split was done randomly, so that the classes are no longer perfectly balanced. The training set is also perfectly balanced with $11,312$ clicks per class. The class names are detailed in Table 7.2. Figures 7.2 and 7.3 show example clicks contained in the DOCC10 dataset for each class except for the sperm whale.

This challenge is distributed by DYNI LIS UTLN on sabiod.fr and MADICS CNRS (`http://sabiod.fr/pub/docc10`) and similarly in the DATA challenge of ENS (`https://challengedata.ens.fr/challenges/32`).

7.3 Baseline

A large part of machine learning research is done on image classification [149, 150, 151]. When working on sounds, the usage of spectrograms or Mel-frequency cepstral coefficients (MFCC) allows one to convert these 1D signals into images, and use the state of the art techniques such as ResNet [152]. Even if this trick is largely used in signal processing, it has the disadvantage of having a number of parameters that need to be tuned beforehand, such as the stride, the window size for the FFT, which will affect the time/frequency resolution. Not only choosing the right representation for each specific task is not obvious, but choosing the wrong parameters for these hand-crafted features might decrease the performance.

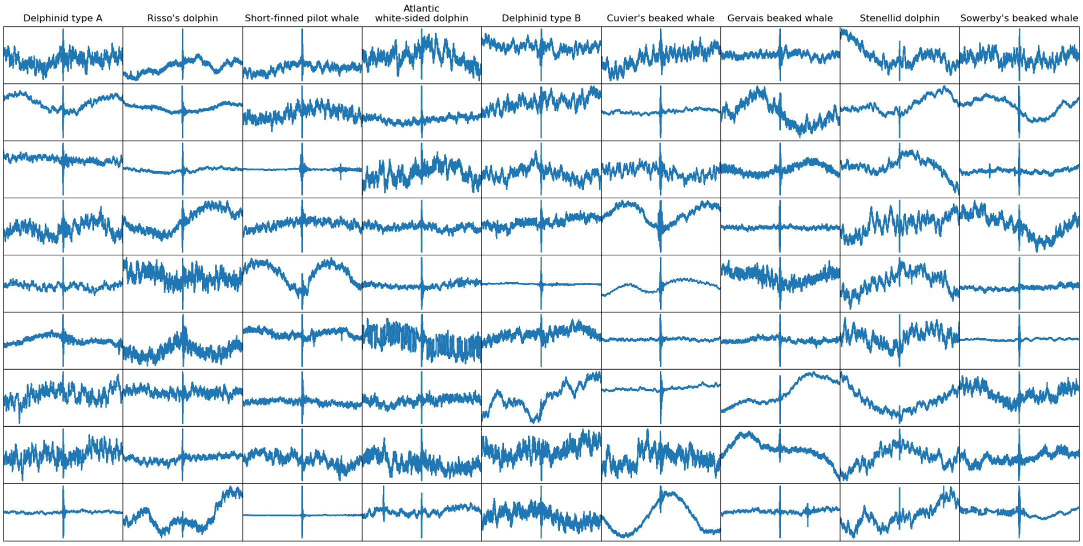


Figure 7.2: Examples of DCLDE test instances for each class (4096 samples long)

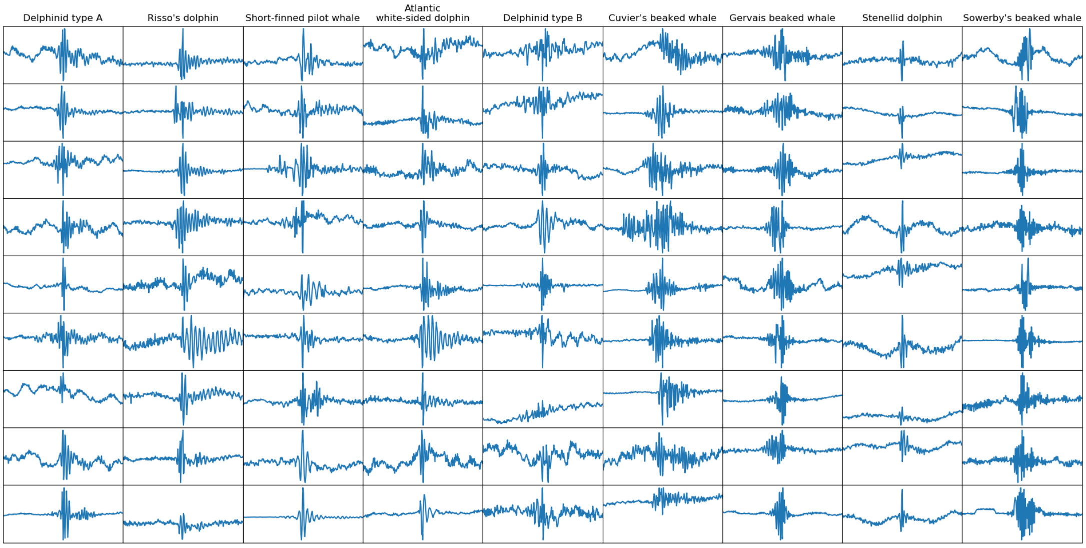


Figure 7.3: Zoom on the same examples of DCLDE test instances for each class (256 samples long)

In bioacoustics, *bulbul* and *sparrow* [153] are two architectures using the STFT magnitude spectrograms that were made for the Bird audio detection challenge[3] and are nowadays used as the state of the art since *bulbul* won the challenge [154, 155].

The first test we did with this architecture did not work, which is to be expected

---

[3] `http://machine-listening.eecs.qmul.ac.uk/bird-audio-detection-challenge`

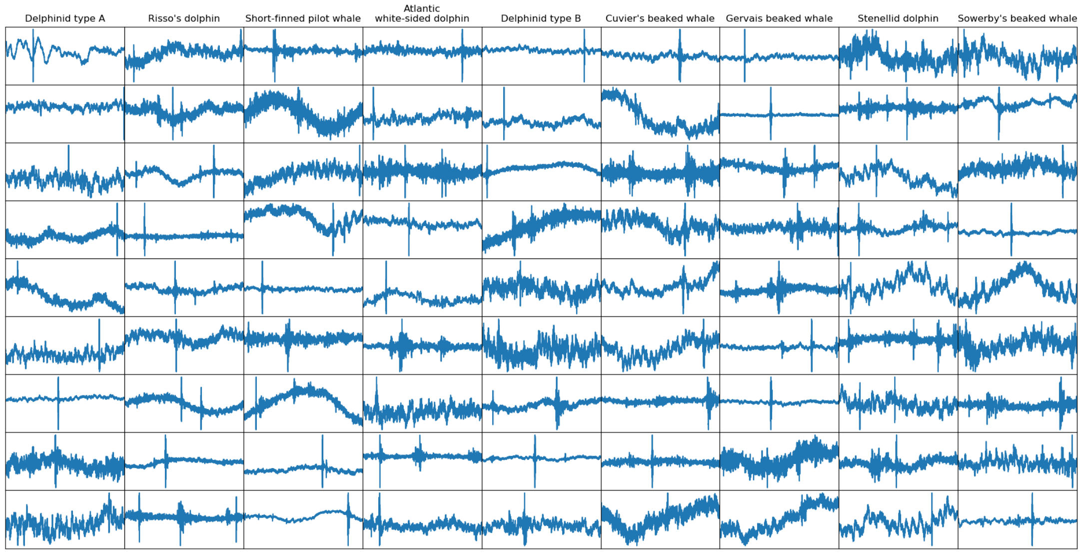


Figure 7.4: Examples of DCLDE test instances for each class (4096 samples long)

since clicks are far from the long signals of bird vocalisations. Instead of using 2D spectrograms, which are better for the analysis of chirps or stationary signals, we decided to learn directly from the raw signal, starting with convolution layers similarly to what is done in the study of ECG signals [156, 157]. The advantage of a convolution layer over a dense layer is that it will force the learned filter to be invariant to a translation of the signal [158]. The multiple filters of a convolution layer will output multiple features per time step, which can be considered as a new dimension with one feature. Two-dimensional convolution can thus be used on this 2D signal, reducing the amount of parameters per layer amongst the other advantages of convolutions [159]. This can be done after the first layer, or after multiple 1D convolution layers [160, 161]. The operation can then be repeated to perform a 3D convolution. For convenience, we call this increase of dimension followed by a convolution, UpDim. This operation could then be repeated to increase the number of dimensions to 4D and more. However, usual deep learning libraries such as Tensorflow or PyTorch do not support convolution on tensors with more than 3 dimension (5 if the batch and feature dimension are taken into account).

7.3.1 Topology of the baseline

We apply the new operator UpDim in a CNN of 12 layers using the raw audio as an input. Windows of 4096 bins are extracted randomly from the 8192-wide samples, and random pink, white and transient noises are added to it, each having an independent amplitude distribution that is log-uniform (to be uniform in dB scale). The result is then normalised and given to the first layer of the CNN. Figure 7.4 shows samples of this process, which are the inputs of the CNN for its training. The topology of the model is given in Table 7.3. The first layers of this model are alternates of convolution and

Layer name	Input size	Kernel	Strides	In features	Out features
Conv-1D	$N * 4096$	5	2	1	16
Conv-2D	$N * 2048 * 16$	5*3	2*1	1	16
Conv-3D	$N * 512 * 16 * 16$	5*3*3	4*1*1	1	16
Conv-3D	$N * 128 * 16 * 16 * 16$	5*3*3	2*1*1	16	32
Conv-3D	$N * 64 * 16 * 16 * 32$	5*3*3	2*2*2	32	64
MaxPool	$N * 32 * 8 * 8 * 64$	5*3*3	4*2*2	64	64
Conv-3D	$N * 8 * 4 * 4 * 64$	5*3*3	2*2*2	64	64
Conv-3D	$N * 4 * 2 * 2 * 64$	5*3*3	2*2*2	64	64
Reshape	$N * 2 * 1 * 1 * 64$				
OneByOne	$N * 2 * 1 * 64$	1*1	1*1	64	64
Max					
OneByOne	$N * 1 * 1 * 64$	1*1	1*1	64	64
OneByOne	$N * 1 * 1 * 64$	1*1	1*1	64	11
Flatten	$N * 1 * 1 * 11$				

Table 7.3: Topology of baseline model

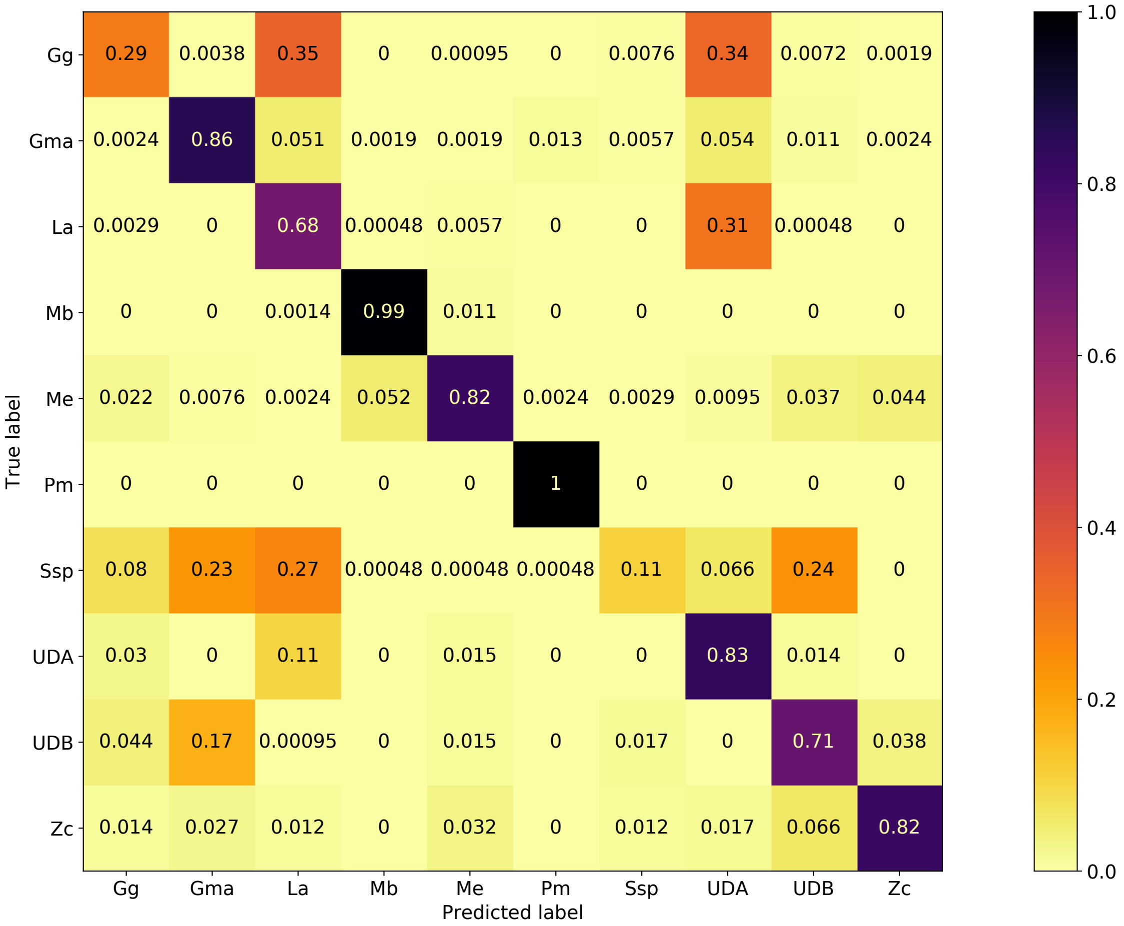
Note that the 11th class was trained to detect noise and was discarded for DOCC10 prediction. Dimensions are given in NHWC order.

increase in dimension using our proposed UpDim operator.

The activation between each layer is a leaky ReLu with an alpha of 0.01. The loss is the cross entropy with softmax. An L2 loss on the weights is added as regularization, with a factor of 0.0005. The model was trained with Adam [162] with a learning rate of 0.0005, during 16 epochs, with mini batches of 32 samples.

7.4 Results

As this baseline was originally built for the CARI'MAM project, it was trained with an additional class, the noise class, which was trained with the artificial noise cited earlier. Hence the network topology has 11 classes instead of the 10 of the dataset. For the evalutation of the full DOCC10 test set, the logit of the noise class was dropped before the softmax. The confusion matrix shown in Figure 7.5 is thus obtained by the prediction without the noise logit. Note that the confusion matrix on a test set which includes noise sample is the same as the one shown in this paper, with all the noise



	Gg	Gma	La	Mb	Me	Pm	Ssp	UDA	UDB	Zc
Gg	0.29	0.0038	0.35	0	0.00095	0	0.0076	0.34	0.0072	0.0019
Gma	0.0024	0.86	0.051	0.0019	0.0019	0.013	0.0057	0.054	0.011	0.0024
La	0.0029	0	0.68	0.00048	0.0057	0	0	0.31	0.00048	0
Mb	0	0	0.0014	0.99	0.011	0	0	0	0	0
Me	0.022	0.0076	0.0024	0.052	0.82	0.0024	0.0029	0.0095	0.037	0.044
Pm	0	0	0	0	0	1	0	0	0	0
Ssp	0.08	0.23	0.27	0.00048	0.00048	0.00048	0.11	0.066	0.24	0
UDA	0.03	0	0.11	0	0.015	0	0	0.83	0.014	0
UDB	0.044	0.17	0.00095	0	0.015	0	0.017	0	0.71	0.038
Zc	0.014	0.027	0.012	0	0.032	0	0.012	0.017	0.066	0.82

Figure 7.5: Baseline confusion matrix on the test set

sample being classified as noise, and one Stenellid dolphin being classified as noise. The baseline obtains a MAP (mean Average Precision) of 77.12% and an accuracy of 71.13% on the full test set. On the public portion of the test set, the MAP is 77.68% and the accuracy is 70.52%.

7.4.1 First challenger results

Since the release of the DOCC10 challenge in early 2020, 26 challengers have participated. The current top 10 scores are reported in Table 7.4. The full up-to-date leaderboard can be found on the Challenge Data website (`https://challengedata.ens.fr/participants/challenges/32/ranking/public`). The top two scores were obtained by the same team, who used a semi-supervised approach on the test set, hence the score gap with the other participants.

For this challenge, we decided to try a modified version of a resnet that uses the UpDim principle as shown in Table 7.5. The activation functions used are leaky ReLu

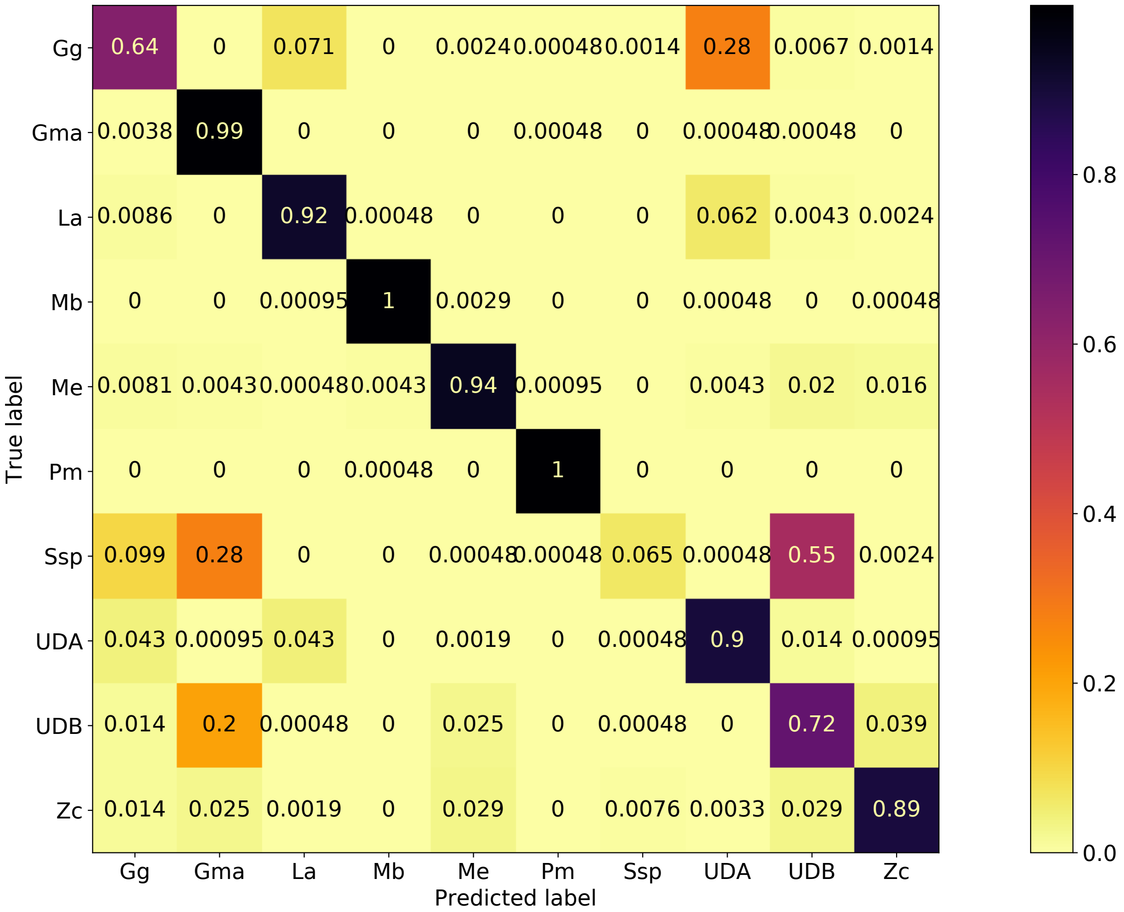
Ranking	Date	User(s)	Public score
1	March 22, 2020	alain.dr	0.8702
2	March 23, 2020	Judy35 & alcodias_data & levilain	0.8659
3	March 28, 2020	TBF	0.8034
4	April 21, 2020	jvasso & RaphaelGin	0.8015
5	March 28, 2020	mclergue	0.7963
6	Feb. 24, 2020	BastienD	0.7953
7	March 19, 2020	trollinou	0.7867
8	March 17, 2020	nattochaduke	0.7858
9	March 3, 2020	BastienD & morhan	0.7772
10	March 18, 2020	LeGrosTroll	0.7677

Table 7.4: Top 10 scores as of May 1, 2020

with an alpha of 0.001. Batchnorm was also used after each convolution layer except the ones of the skip connections.The loss is the cross entropy with softmax. An L2 loss on the weights is added as regularization, with a factor of 0.05. The model was also trained with Adam using beta's of $(0.8, 0.999)$ and a epsilon of 0.0001, with a learning rate of 0.0002. These parameters were not optimised. A Mixup data augmentation [163, 164, 165] using an alpha of 0.2 was also used. The confusion matrix of this experiment with an accuracy of 80.62% can be seen in Figure 7.6.

7.5 DOCC7

An alternate version of the DOCC10 dataset, called DOCC7, has been generated. It has the same samples, but restricted to only 7 species, which are Gg, Gma, La, Mb, Me, Pm, and Zc. The reason for the removal of UDA and UDB is more straightforward. When the DCLDE dataset was made, they used clustering methods to detect the various species. These two labels were then given to dolphin species that could not be identified. We decided to leave them in the DOCC10 challenge since they still represent clicks that belong to groups of dolphins, even if they do not represent only one species, unlike the other labels. These clusters are also useful to train a classifier that would be used after a click detector, and prevent it to classify these dolphin clicks as another species. However, trained network (with various architectures from various labs) have shown that, unlike the seven other classes in DOCC7, the trained networks had lower accuracy on the UDA and UDB labels. We believe that the networks prediction might not be wrong, meaning that these classes have a higher label noise. Finally, the Ssp were also removed for two reasons. Firstly, Stenella is a genus and not a species unlike the other remaining classes.



	Gg	Gma	La	Mb	Me	Pm	Ssp	UDA	UDB	Zc
Gg	0.64	0	0.071	0	0.0024	0.00048	0.0014	0.28	0.0067	0.0014
Gma	0.0038	0.99	0	0	0	0.00048	0	0.00048	0.00048	0
La	0.0086	0	0.92	0.00048	0	0	0	0.062	0.0043	0.0024
Mb	0	0	0.00095	1	0.0029	0	0	0.00048	0	0.00048
Me	0.0081	0.0043	0.00048	0.0043	0.94	0.00095	0	0.0043	0.02	0.016
Pm	0	0	0	0.00048	0	1	0	0	0	0
Ssp	0.099	0.28	0	0	0.00048	0.00048	0.065	0.00048	0.55	0.0024
UDA	0.043	0.00095	0.043	0	0.0019	0	0.00048	0.9	0.014	0.00095
UDB	0.014	0.2	0.00048	0	0.025	0	0.00048	0	0.72	0.039
Zc	0.014	0.025	0.0019	0	0.029	0	0.0076	0.0033	0.029	0.89

Figure 7.6: Confusion matrix on the test set

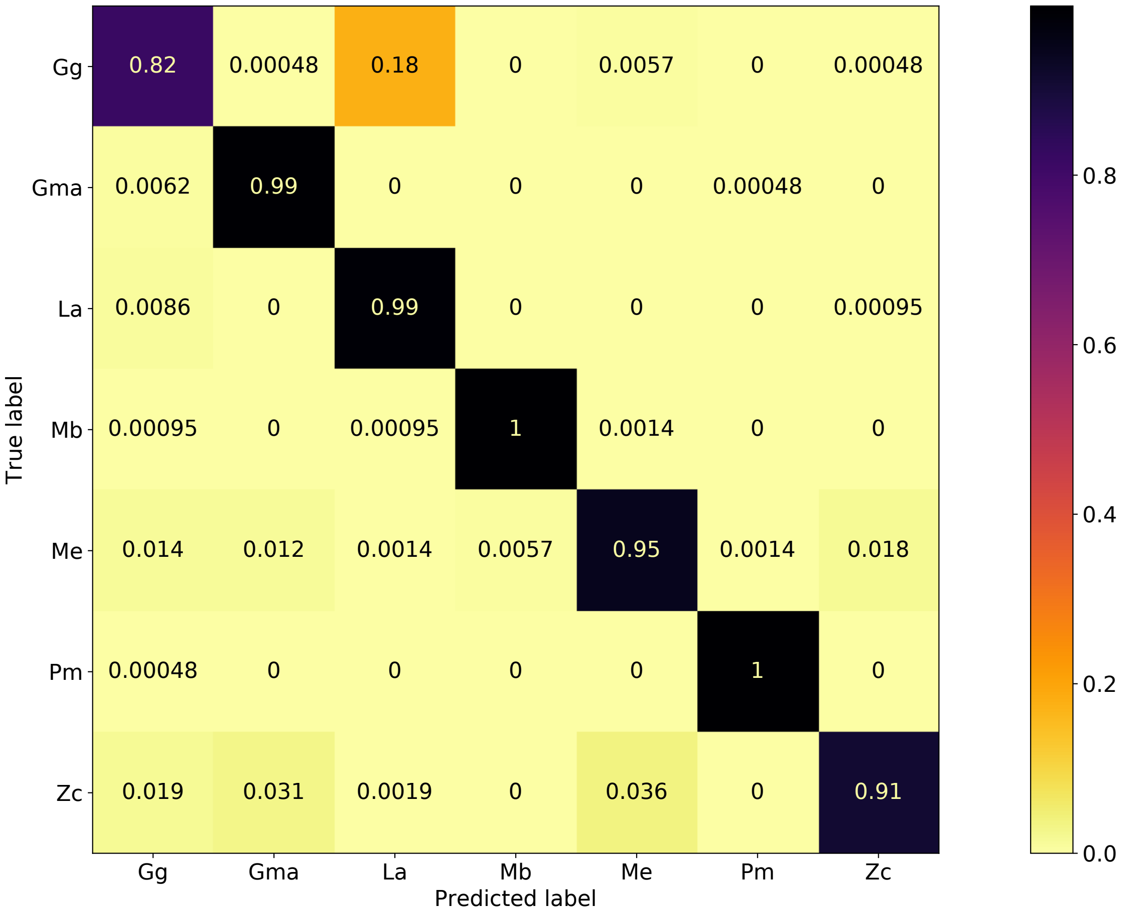


Figure 7.7: Confusion matrix on the test set of DOCC7

Secondly, there seems to be a large covariate shift between the training and test sets for this class. A number of reasons could explain this difference between the training and test set, such as different species, different groups, different types of clicks, or mislabeling. As seen in Figure 7.6, these three classes represent the majority of the confusions. The modified resnet of Table 7.5 was also tried on DOCC7, and it obtains an accuracy of 95.09%.

However, this smaller version of the test dataset will not be released until the end of the challenge, to prevent any challenger from gaining information on the test set.

7.6 Conclusion

We created a new DOCC10 dataset with strong labels for marine mammal transient classification. It has a total of 134,080 clicks for 10 species. Except for part of the test reserved for the scoring of the DOCC10 challenge that has been opened with this dataset, the dataset is publicly available. We also proposed a new neural network model to classify these marine mammal transients. With the new recording from the Sphyrna Odyssey 2019-2020 mission, containing other species, we plan to release an augmented version of the DOCC10 dataset with more classes, such as Tursiops, or *Globicephala Macrorhynchus*. We also plan to include records from the CARI'MAM project, composed of 20 recording stations spread over the Caribbean islands. The CARI'MAM project targeted around 30 species. This augmented dataset will probably be released in late 2020 or early 2021. The increase of variety in the acoustic environment and recording devices should allow networks trained on it to be more robust to unseen background noise and other details linked to these changes.

Layer name	Input size	Kernel	Strides	Out features
Conv-1D	$N * 4096 * 1$	3	1	32
Conv-1D	$N * 4096 * 32$	3	2	32
Skip	$N * 4096 * 1$	1	2	32
Conv-1D	$N * 2048 * 32$	3	2	64
Conv-1D	$N * 1024 * 64$	3	2	128
Skip	$N * 2048 * 32$	1	4	128
Conv-2D	$N * 1024 * 128 * 1$	3*3	1*1	32
Conv-2D	$N * 1024 * 128 * 32$	3*3	2*2	32
Skip	$N * 1024 * 128 * 1$	1*1	2*2	32
Conv-2D	$N * 512 * 64 * 32$	3*3	2*2	64
Conv-2D	$N * 256 * 32 * 64$	3*3	2*2	128
Skip	$N * 512 * 64 * 32$	1*1	4*4	128
Conv-3D	$N * 128 * 16 * 128 * 1$	3*3*3	1*2*1	32
Conv-3D	$N * 128 * 8 * 128 * 32$	3*3*3	2*2*2	64
Skip	$N * 128 * 8 * 128 * 1$	1*1*1	2*4*2	64
Conv-3D	$N * 64 * 4 * 64 * 64$	3*3*3	2*2*2	128
Conv-3D	$N * 32 * 2 * 32 * 128$	3*3*3	2*2*2	256
Skip	$N * 64 * 8 * 64 * 64$	1*1*1	4*4*4	256
Softmax	$N * 16 * 1 * 16 * 256$	16*1*1		
MaxPool	$N * 16 * 1 * 16 * 256$	16*1*1		
Flatten	$N * 1 * 1 * 16 * 256$			
Dense	$N * 4096$			1024
Dense	$N * 1024$			512
Dense	$N * 512$			10

Table 7.5: Topology of UpDimV2 model

Dimensions are given in NHWC order. Horizontal lines separate each residual block.

Chapter 8

Application of Autoencoders to Click Analysis

8.1 An Introduction to Autoencoders

In the bioacoustic domain, there are many unknown or badly known signals. Usual approaches based on neural networks (NN) for sound signals are made for human speech, and use knowledge on how this signal is created, such as existing phonemes or even language structure, to improve performance. Some research has been done on raw audio signals [166, 167]. NN on raw waveforms have also shown to obtain better denoising abilities [168, 169], especially compared to other methods that just copy the original phase [170]. Without this knowledge of a dictionary of signals to classify, one might try unsupervised learning, such as an autoencoder. In this paper we study stereo datasets of transient noises. We show that by adding a second branch in the decoder, it is possible to achieve a better untangling of the embeddings. This study was motivated by the analysis of an unlabeled sperm whales' click dataset.

8.2 The Two Models

In this paper, we compare two Autoencoders. Their goal is to reconstruct the noisy stereo inputs they are given. Both share the same encoder which is composed of 7 convolution layers and 3 dense layers. The two models then share a part of the decoder which is made of 3 dense layers and 3 transpose convolution layers, with the only difference being that the first model decoder will reconstruct the channel whereas the second model decoder will reconstruct only one signal which will be tiled to two channels. For the first model an alternative decoder has been made with two additional dense layers in order to have a fairer comparison with the second model. The second model has an extra branch (Table 8.1) coming from the embedding that will predict a filter. This filter will then be applied to the tiled signal using a cross convolution. The noisy signal is fed raw to the network, with no normalization, in order to let the network learn the loudness of the signal, or

Layer type, Activation	Input shape	Kernel, stride	Filters
Encoder			
Convolution layer, tanh	2*1024*1	1*11, 1*4	64
Convolution layer, leaky relu 5%	2*256*64*1	1*1*64, 1*1*1	1
Convolution layer, tanh	2*256*64	1*11, 1*4	128
Convolution layer, leaky relu 5%	2*64*128*1	1*1*128, 1*1*1	1
Convolution layer, tanh	2*64*128	1*11, 1*4	128
Convolution layer, leaky relu 5%	2*16*128*1	1*1*128, 1*1*1	1
Convolution layer	2*16*128	2*1, 1*1	256
Dense layer leaky relu 5%	4096		2048
Dense layer leaky relu 5%	2048		512
Dense layer	512		128
Decoder			
Dense layer leaky relu 5%	128		1024
Dense layer leaky relu 5%	1024		2048
Dense layer leaky relu 5%	2048		2048
Transpose convolution, leaky relu 5%	2*128*8	1*5, 1*2	8
Transpose convolution	2*256*8	1*5, 1*2	8
Transpose convolution	2*512*8	1*5, 1*2	1
Extra branch			
Dense layer, leaky relu 5%	128		1024
Dense layer, leaky relu 5%%	1024		2048
Softmax	2*1024	1*1024	1024

Table 8.1: Model architecture

Note that for the decoder of the Extra branch autoencoder, the shape only has one channel instead of two.

the SNR. Both networks have the same 2 losses. The main loss is the reconstruction loss which is the mean squared error loss between the input noisy signals and the output of the autoencoder. The regulation losses are L2 losses on the weights. The optimizer used is ADAM [162].

8.3 Proof of concept

The 3 datasets that we introduce here all have different characteristics, which can be used to compute the various capabilities of the network to cluster and separate the data. The metric used was the error (for the wanted clustering parameters) between the point and each of its 40 nearest neighbours in the embedding space. This choice was motivated by the study of 3D t-SNE on the embedding [171], which showed that embeddings were composed of manifolds inter-twinned together, with most of the time only one manifold per family. All the datasets have the center of each pulse as a feature, which can also be viewed as one center and a Time Difference Of Arrival (TDOA).

8.3.1 Description of the first synthetic dataset

The first dataset is made of a sinusoid with a Gaussian envelope. It is spread into five families. Each child of a family is created with a carrier frequency which is sampled from a Gaussian distribution around the frequency of the family. The support size and the amplitude of each child is also randomly chosen. With all these parameters defined, each transient is created with a random offset from the center, with a different offset for the second channel, to create a TDOA. Then a white noise is added to the signal with a random level. This mean that this dataset has a distribution of SNR, but also a distribution of signal loudness.

8.3.2 Description of the second synthetic dataset

The second synthetic dataset is made from an adaptation of a model of dolphin click generation [119], which is used here to generate the pulse P1 of the sperm whale click.

$$U(t) = U_0 \, \mathrm{ReLU}\left(\frac{a^{\mathrm{ReLU}(t/T-m)} - a^{t/T}}{1-a^m}\right), \tag{8.1}$$

where m is the number of periods to reach the maximum of amplitude, a is the reflectance factor, and T is the period of the carrier waveform. The parameters that were $m = 5$ and $a = 0.84$ for the dolphins have been changed to $m = 1$ and $a = 0.25$ for sperm whales. The denominator has also been changed from $1 - a$ to $1 - a^m$ in order to have a maximum magnitude of U_0.

A sperm whale click is composed of a first pulse P0, leaking from its monkey lips dorsal bursae (MLDB) complex, the main pulses P1, and other echo pulses following the same path as P1 (Figure 8.1) . The other echo pulses all have the same inter pulse interval (IPI), while the IPI of P0P1 can have a ±10% length depending on the orientation of the animal with regard to the recorder. The pulse P0 is also in opposition of phase, since it was not affected by the reflection from the back of the head of the sperm whale.

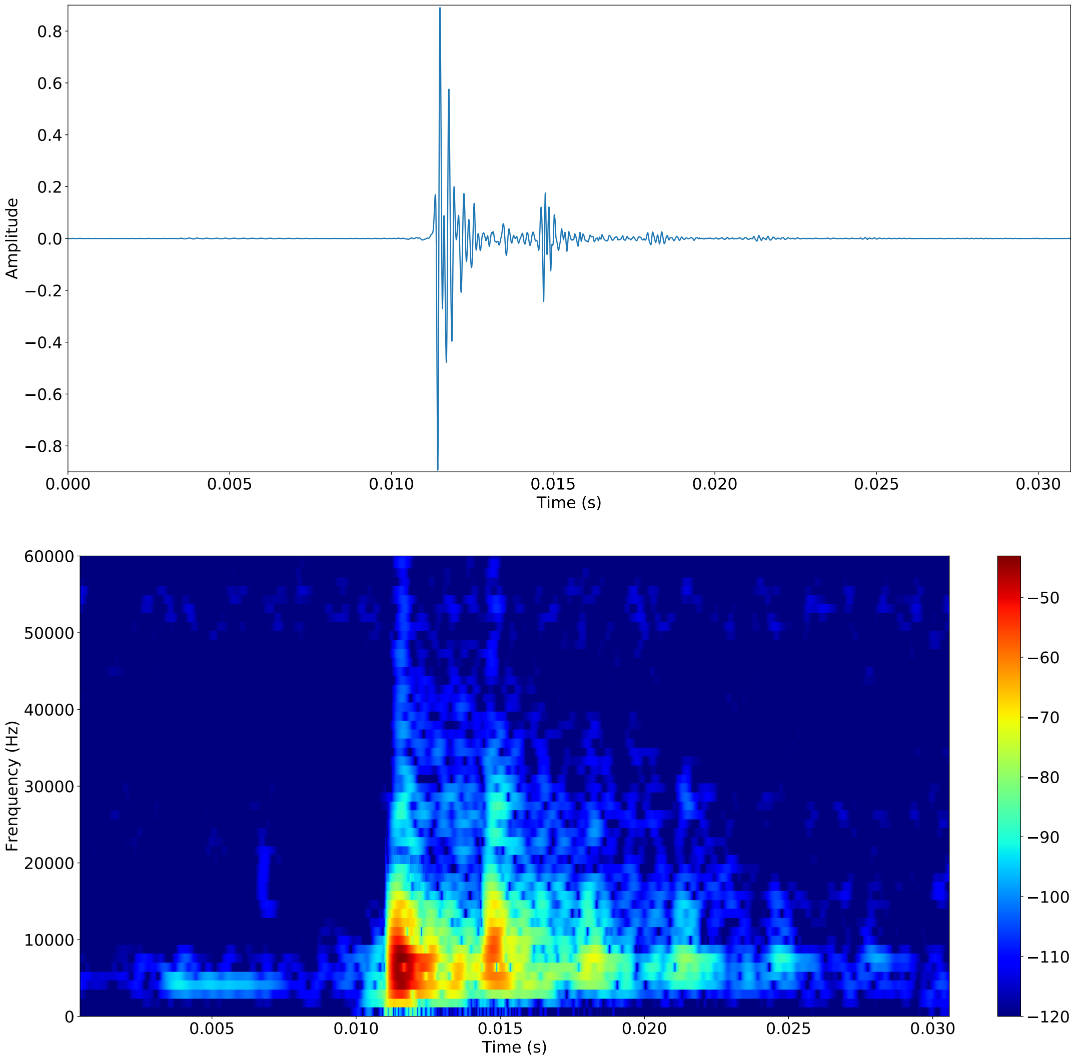


Figure 8.1: Example of a real click with its spectrogram

The pulses present between the labeled pulses are pulses leaking from the back of the sperm whale head and are called false pulses

This dataset is thus based on these facts, and has 9 families, bred from 3 main values of IPI and 3 main carrier frequencies. The IPI values are 3, 5 and 7 ms, and the center frequency values are from the set $\{3.25 \text{ kHz}, 6.5 \text{ kHz}, 13 \text{ kHz}\}$. The time is 1024 samples from -0.01 to 0.01, endpoints included. The amplitude of P2 is around 35% of the amplitude of P1, while P0 is only 20%. In the same manner as the previous dataset, each click is off-centered, and TDOA is introduced between the channels.

8.3.3 Description of the third synthetic dataset

The third synthetic dataset was created with the aim of showing what would happen when the discrimination criterion for the families is not a temporal or frequency criterion, but a variation in the spectrum. All families have the same fundamental frequency at 3.25 kHz (with a slight variation on it for each child) and the second and fourth harmonic. The signal has the same envelope as the second dataset (8.1), but only has one pulse. The five families shown in Figure 8.2 are $[1, 2, 3], [3, 2, 1], [3, 1, 2], [1, 3, 2], [2, 3, 1]$ where each digit is the respective relative amplitude of the harmonics 1, 2 and 4. The pulses are off centered and TDOA shifted the same way as the previous datasets were.

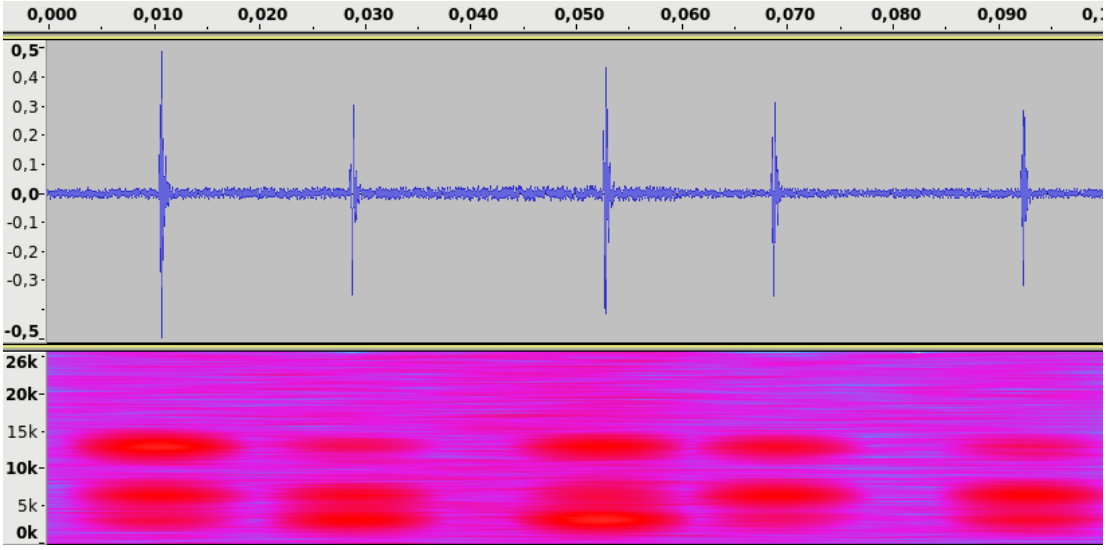


Figure 8.2: One member of each of the five families
The pulses present between the labeled pulses are pulses leaking from the back of the sperm whale head and are called false pulses

8.4 Results

In Section 8.2, we explained that the main loss was an MSE between the noisy inputs and the outputs. Since the two datasets presented here are synthetic, we also have the known TDOA and known clean signal as a metric to observe how well the networks were learning these parameters. The clean metric is a MSE between the clean signal and the output, while the TDOA metric is a MSE between the TDOA and the position of the maximum of the cross correlation of the outputs.

The clean metric helps to quantify the fact that even if the task of the autoencoders is to reproduce the inputs, they are unable to reproduce the random noise and are learning the clean version of the signals. Since the outputs are denoised versions of the signal, the TDOA metric was also improving during the learning.

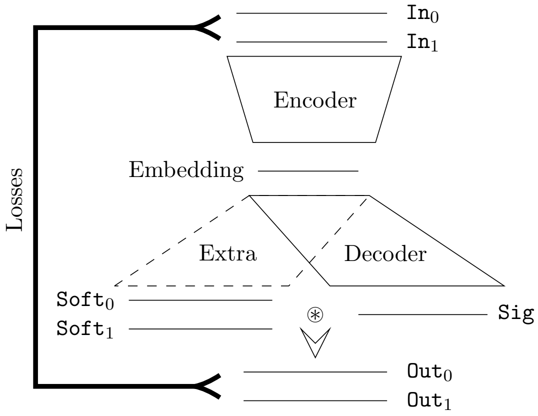


Figure 8.3: Autoencoder with the extra branch

Assuming we use the notation of Figure 8.3, what happens to the `Soft` signal and the `Sig` signal for the extra branch autoencoder is interesting. The only restriction on these two signals is that the `Soft` signal is positive and must sum to 1. Other than that, the network is free to encode each feature on either one of the signals, or mix it in both. This will happen most of the time, but we had some iterations where the network was learning the signal with `Sig`, while `Soft` was learning the position. In future studies, we will investigate what changes can be made to improve this behaviour. For example, reducing `Sig` to a short support, or having the last layer of the extra branch being a convolution. Another behaviour that can be noted is that the network will use zero padding of the convolution as a perk. For example, when the noise makes the two input clean signals a bit too different, the network will reconstruct a pulse at each extremity of `Sig`, and the shifting of the `Soft` signals will make each `Out` only have one of them at the right position.

After the training, we computed the l_1 loss between the ground truth TDOA and the predicted TDOA. As it is shown in Figure 8.4, we compared the TDOA with three oracle versions of generalized cross correlation with Eckart weight. The first version is the usual weight, the second has a mean noise amplitude throughout frequency bins, and the last has its weight clipped to 1 (best clipping possible). We also added the results of networks trained with only one channel, and then applied separately on both channels to denoise the signal and compute the TDOA. The cumulative histogram (Figure 8.5) for the third dataset has secondary peaks that are located on periods of the signal, while the histogram for the second dataset has a second local maximum that is located on the IPI.

The previous results have shown the possibilities of the autoencoders on the TDOA estimation. As it was said in the introduction, the autoencoders are also used to cluster the data with regards to their physical features.

The aim of the first dataset was to introduce five simple families of clicks, which were determined by frequency and by support width to see if the two networks were able to cluster them. With this simple dataset both networks are able to disentangle the

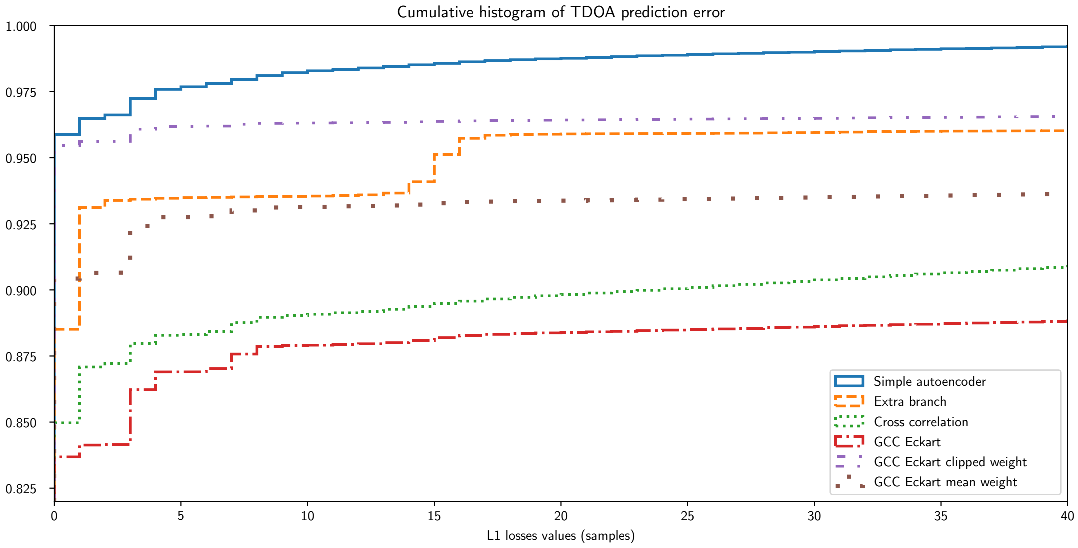


Figure 8.4: Cumulative histogram of the error between the predicted TDOA and the ground truth for the first 100 bins

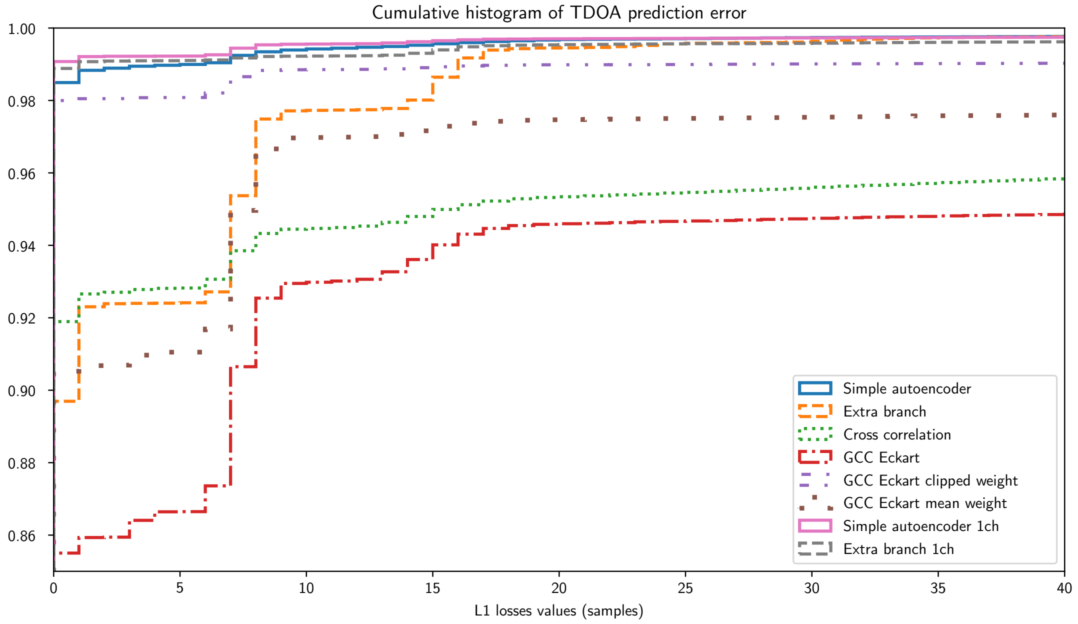


Figure 8.5: Cumulative histogram of the error between the predicted TDOA and the ground truth for the first 40 samples

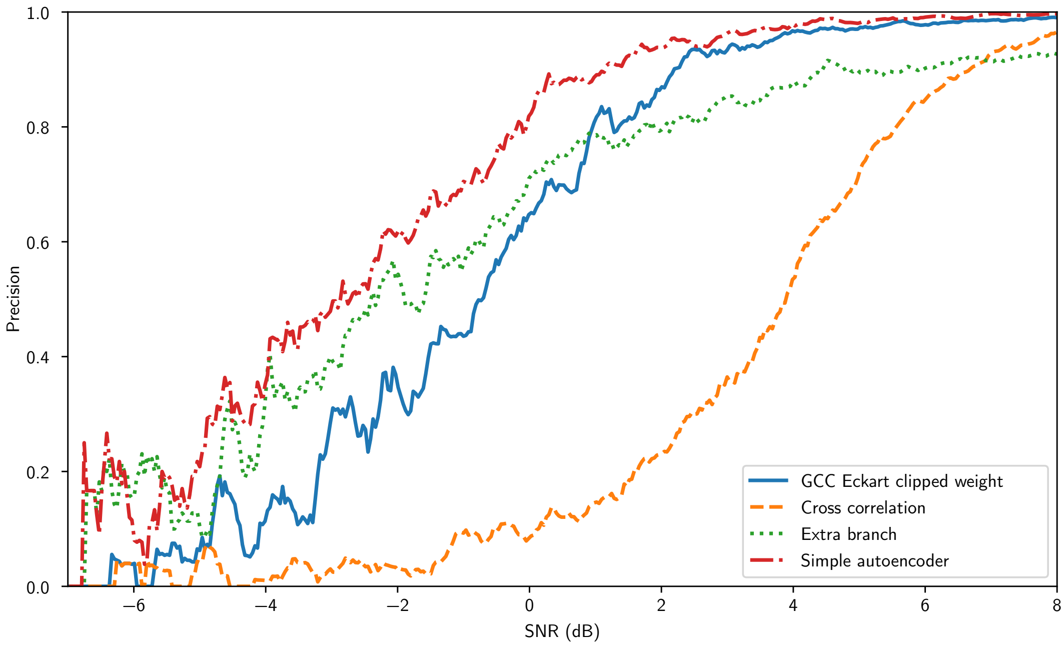


Figure 8.6: Precision of the predicted TDOA within an error smaller than 3 samples. Remarkably, the autoencoder enhances GCC by 3dB at precision 0.5.

Dataset	AE	5	25	Med.	Mean	75	95
1 Freq	SA	-0,068	0,111	0,243	0,263	0,359	0,789
	EB	**-0,146**	**0,00**	**0,133**	**0,163**	**0,272**	**0,674**
	RN	0.756	0.82	0.888	0.903	0.99	1.07
2 Freq	SA	2,06	2,28	2,81	2,74	3,10	3,43
	EB	**2,05**	**2,27**	**2,53**	**2,36**	**2,76**	**3,07**
	RN	3.47	3.55	3.64	3.64	3.72	3.81
2 IPI	SA	-2,90	-2,83	-2,73	-2,74	-2,67	-2,60
	EB	**-2,91**	-2,83	**-2,74**	**-2,75**	-2,67	-2,60
	RN	-2.90	-2.82	-2.73	-2.74	-2.66	-2.60
3	SA	0,55	0,65	0,725	0,711	0,755	0,85
	EB	**0,45**	**0,55**	**0,625**	**0,618**	**0,7**	**0,8**
	RN	0.70	0.75	0.80	0.80	0.85	0.90

Table 8.2: Statistics on the mean difference between each example and its 40 nearest neighbours in the embedding space. The frequency and IPI stats on the log10 distribution of errors. For the third datatset, an error of one was attributed to neighbours that belong to other families. SA : Simple Autoencoder model, EB : Extra Branch model, RN : random neighbours

embedding by family, but even in this case, the extra branch network is able to disentangle the embedding more. With the second dataset, the goal was to see if the networks were able to distinguish between the two families with the same center frequency and the same pulse, but with various echoes. This dataset is kind of a disadvantage for the extra branch, because the pulse P0 has the opposite phase, which the network cannot reproduce with the shifting of the positive `Soft` signals. As the Table 8.2 shows, the extra branch network is able to disentangle the frequencies, however it performs as badly as the other network on the IPI. This can be explained by the fact that the other pulses were not always reconstructed by the network, since they were most of the time hidden in the noise. The aim of the last dataset was to see if the networks were able to distinguish between clicks with the same frequency content, but with variations on the amplitude. The extra branch model did once again better, even thought it may have done even better if the families were less entangled (the Gaussian variation on each harmonic amplitude is large enough to mix some families together)

8.5 Test on real data

We tested our method on two real datasets. We used a month of data from the sonobouy Bombyx [123], which represent 500 000 clicks. Bombyx is located in the Mediterranean sea, near the island of Porquerolles, at 27 meters depth. It records at 50 kHz on two hydrophones distant of 1.8 meters. The second dataset we used is made of clicks that came from an antenna that we fixed under the Sphyrna [137], a surface drone. This Sphyrna dataset is small (6 000 clicks on 4 channels). To stay coherent with the previous datasets, the signals have been downsampled to 50 kHz. The clicks were fed to the network in windows of 1024 samples. Both datasets have been filtered with an 8-th order high-pass Butterworth filter at 2 kHz in order to remove low frequency noise that was magnitudes higher compared to the signal. Without this filtering, the networks will mainly learn how to reconstruct this low frequency noise. We also found that a normalization on the energy of each signal allowed the networks to learn with the same number of epochs as when they learn on the synthetic datasets. Otherwise they take at least twice the number of epochs. However, we were not able to quantify the results as we do not have any precise ground truth. Even though we have TDOA for the Sphyrna, the signals contained multiple sources and the network sometimes predicts where other sources should be (using interpolation). We will inspect these results more thoroughly in future works.

8.6 Conclusion

We have shown, for synthetic data, that a simple autoencoder can be used as an unsupervised tool to compute the TDOA on noisy data, and that by adding an extra branch, we are able to better disentangle the embedding in order to cluster the data.

Part III

Wave Propagation and Simulation for Bioacoustics

This part, consisting of three chapters, sets up the models and numerical methods that are required for simulating the generation and propagation of sounds in the complex medium made up of the head of the sperm whale and its surrounding environment. Particular attention will be paid to an accurate representation of the anatomy and morphology, as well as the different types of waves, acoustic and elastic, that play a role. Though we restrict our treatment here to linear theory, there are nonlinear phenomena which could be taken into account in future work.

Note that the propagation model will be needed for the final part of the thesis, where we aim to couple models and data in an inverse problem approach.

Chapter 9

State of the Art in Wave Propagation

This part of this manuscript is dedicated to simulation of the sperm whale sonar. The theory of acoustic is thus explain in order to have the multiple equation that will be used in the simulation, but also used to interpret the results.

The vast majority of the acoustic studies will use results or equations related to the acoustic wave equation,

$$\frac{\partial^2 p}{\partial t^2} - \nabla \cdot (c(x)\nabla p) = 0, \tag{9.1}$$

where p is the acoustic pressure, $c = c(x)$, is the sound speed at point $x \in \Omega$, and Ω is the domain of propagation that we want to study. This equation must be completed by initial and boundary conditions. and is the basis of linear acoustics. If c is constant, we can write

$$\nabla^2 p = \frac{1}{c^2}\frac{\partial^2 p}{\partial t^2}, \tag{9.2}$$

However, this is only an approximation, and in numerous applications nonlinear terms becomes non negligible. This area of acoustics is called nonlinear acoustics, and includes the sound generation of musical instruments, voices, or shock waves—see Appendix A.

9.1 Linear wave propagation

To derive the equations that govern linear acoustics, we use the two main equations of continuum mechanics, the conservation of mass and Newton's second law,

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \boldsymbol{v}) = 0, \tag{9.3}$$

$$\frac{\mathrm{D}\boldsymbol{v}}{\mathrm{D}t} = \rho \boldsymbol{g} + \sum_{i=1}^{3}\sum_{j=1}^{3} \frac{\partial \sigma_{ij}}{\partial x_j}\boldsymbol{e}_i, \tag{9.4}$$

where ρ is the local mass density, $\boldsymbol{v}$ is the particle velocity, $\boldsymbol{g}$ is the gravitational acceleration, $\boldsymbol{e}_i$ are the unit vectors of a Cartesian basis in Galilean reference frame, σ_{ij} are Cartesian components of the stress tensor and $\frac{\mathrm{D}}{\mathrm{D}t}$ is the Stokes' total time derivative operator [172],

$$\frac{\mathrm{D}}{\mathrm{D}t} = \frac{\partial}{\partial t} + \boldsymbol{v} \cdot \nabla. \tag{9.5}$$

9.1.1 Fluid

The second starting point that is needed for fluids is the convective form of the Cauchy momentum equation.

$$\frac{\mathrm{D}u}{\mathrm{D}t} = -\nabla \boldsymbol{q} + \sum_{i=1}^{3} \sum_{j=1}^{3} \sigma_{ij} \frac{\partial v_i}{\partial x_j}, \tag{9.6}$$

which can incorporate the entropy since

$$T\mathrm{d}s = \mathrm{d}u + p\mathrm{d}\rho^{-1}, \tag{9.7}$$

$$\mathrm{d}\rho = \frac{1}{c^2}\mathrm{d}p - \frac{\alpha\rho T}{c_p}\mathrm{d}s, \tag{9.8}$$

where c_p is the specific heat at constant pressure and α is the coefficient of thermal expansion. To link the stress to the viscosity of a fluid, the properties of the material need to be considered. Here the hypothesis is made that the fluids considered are Newtonian fluids. Then

$$\sigma_{ij} = \sigma_n \delta_i^j + \mu\phi_{ij}, \tag{9.9}$$

$$\phi_{ij} = \frac{\partial v_i}{\partial x_j} + \frac{\partial v_j}{\partial x_i} - 2/3 \nabla \cdot \boldsymbol{v} \delta_i^j, \tag{9.10}$$

$$\sigma_n = -p + \zeta \nabla \cdot \boldsymbol{v}, \tag{9.11}$$

where ϕ_{ij} is rate of shear, ζ is the bulk viscosity, μ is the shear viscosity and σ_n is the average normal stress.

Finally, the theory of heat [173] says that the flux vector $\boldsymbol{q}$ should be the opposite of the temperature gradient times the thermal conduction κ,

$$\boldsymbol{q} = -\kappa \nabla T. \tag{9.12}$$

All the previous equations can then be combined to obtain the Navier-Stokes-Fourier equations

$$\rho \frac{\mathrm{D}\boldsymbol{v}}{\mathrm{D}t} = -\nabla p + \nabla(\zeta \nabla \cdot \boldsymbol{v}) + \rho \boldsymbol{g} + \sum_{i=1}^{3} \sum_{j=1}^{3} \frac{\partial \mu \phi_{ij}^2}{x_j} \boldsymbol{e}_i, \tag{9.13}$$

$$\rho T \frac{\mathrm{D}s}{\mathrm{D}t} = \zeta(\nabla \cdot \boldsymbol{v})^2 + \nabla \cdot (\kappa \nabla T) + \frac{1}{2}\mu \sum_{i=1}^{3} \sum_{j=1}^{3} \phi_{ij}^2. \tag{9.14}$$

In most applications, both viscosity and thermal conductivity can be neglected, which allows us to simplify the Navier-Stokes-Fourier equations and obtain the Euler equation,

$$\rho \frac{\mathrm{D}\boldsymbol{v}}{\mathrm{D}t} = -\nabla p + \rho \boldsymbol{g}. \tag{9.15}$$

Since $\frac{\mathrm{D}s}{\mathrm{D}t} = 0$ with the new hypothesis, (9.3) and (9.8) can be used to obtain

$$\rho \frac{\mathrm{D}p}{\mathrm{D}t} = +\rho c^2 \nabla \cdot \boldsymbol{v}, \tag{9.16}$$

which will lead to the wave equation.

9.1.2 Elastic solids

If we consider a solid to be an ideal elastic solid [174], and make the hypothesis that the displacement of the particles are small enough to neglect the convective time derivative, then (9.4) becomes

$$\rho \frac{\partial^2 \xi_i}{\partial t^2} = \sum_{j=1}^{3} \frac{\partial \sigma_{ij}}{\partial x_j}, \tag{9.17}$$

where ξ_i is the displacement of the particle in the $\boldsymbol{e}_i$ direction relative to its nominal position. With the former hypothesis, the stress σ_{ij} can be computed using the strains ε_{ij} with Hooke's law [175],

$$\varepsilon_{ij} = \frac{1}{2}\left(\frac{\partial \xi_i}{\partial x_j} + \frac{\partial \xi_j}{\partial x_i}\right), \tag{9.18}$$

$$\sigma_{ij} = \sum_{k=1}^{3}\sum_{l=1}^{3} c_{ijkl}\varepsilon_{kl}, \tag{9.19}$$

where c_{ijkl} are the components of the stiffness tensor. Note that only 21 components of the stiffness tensor are independent due to major and minor symmetries [176]. The Voigt notation [177] allows us to reduce the number of indices by taking advantage of the lower number of independent stiffness components. We will write

$$\begin{pmatrix}\sigma_{11}\\ \sigma_{22}\\ \sigma_{33}\\ \sigma_{23}\\ \sigma_{13}\\ \sigma_{12}\end{pmatrix} = \begin{pmatrix}\sigma_{1}\\ \sigma_{2}\\ \sigma_{3}\\ \sigma_{4}\\ \sigma_{5}\\ \sigma_{6}\end{pmatrix}, \qquad \begin{pmatrix}\varepsilon_{11}\\ \varepsilon_{22}\\ \varepsilon_{33}\\ \varepsilon_{23}\\ \varepsilon_{13}\\ \varepsilon_{12}\end{pmatrix} = \begin{pmatrix}\varepsilon_{1}\\ \varepsilon_{2}\\ \varepsilon_{3}\\ \varepsilon_{4}\\ \varepsilon_{5}\\ \varepsilon_{6}\end{pmatrix}, \qquad \sigma_i = C_{ij}\varepsilon_j. \tag{9.20}$$

Isotropic solids

Isotropic solids are solids that have identical proprieties in any direction. With this much symmetry, the number of independent components is reduced to 2. Now (9.19) can thus be rewritten as

$$\sigma_{ij} = 2\mu\varepsilon_{ij} + \delta_i^j \lambda \sum_{k=1}^{3} \varepsilon_{kk}, \tag{9.21}$$

where δ_i^j is the Kronecker delta, and (λ, μ) are the Lamé parameters.

9.1.3 Wave equation

For a compressible fluid, if there is no ambient velocity, and the impact of the gravity is neglected, (9.15) and (9.16) become

$$\rho \frac{\partial \boldsymbol{v}}{\partial t} = -\nabla p, \tag{9.22}$$

$$\frac{\partial p}{\partial t} + \rho c^2 \nabla \cdot \boldsymbol{v} = 0. \tag{9.23}$$

These two equations can then be combined to obtain

$$\nabla \cdot \left(\frac{1}{\rho} \nabla p \right) - \frac{1}{\rho c^2} \frac{\partial^2 p}{\partial t^2} = 0. \tag{9.24}$$

A constant density will allow us to simplify this to

$$\nabla^2 p - \frac{1}{c^2} \frac{\partial^2 p}{\partial t^2} = 0. \tag{9.25}$$

For isotropic elastic solids, by taking the same hypothesis of a constant density, and also considering that the Lamé coefficients are independent of the position, we can combine the equations (9.17), (9.18), and (9.21) to obtain

$$\frac{\partial^2 \boldsymbol{\xi}}{\partial t^2} = \frac{\lambda + 2\mu}{\rho} \nabla(\nabla \cdot \boldsymbol{\xi}) - \frac{\mu}{\rho} \nabla \times (\nabla \times \boldsymbol{\xi}). \tag{9.26}$$

By using the Helmholtz decomposition [178], the vector field $\boldsymbol{\xi}$ can be decomposed into two fields, an irrotational (curl-free) vector field, and a solenoidal (divergence-free) vector field by using the scalar potential Φ and the vector potential $\mathbf{A}$.

$$\boldsymbol{\xi} = \nabla\Phi + \nabla \times \mathbf{A} \tag{9.27}$$

This allows us to decompose (9.26) into two equations,

$$\nabla^2 \Phi - \frac{1}{c_1^2} \frac{\partial^2 \Phi}{\partial t^2} = 0, \tag{9.28}$$

the equation of the longitudinal wave propagation and

$$\nabla^2 \mathbf{A} - \frac{1}{c_2^2} \frac{\partial^2 \mathbf{A}}{\partial t^2} = 0, \tag{9.29}$$

the equation of the shear wave propagation, where $c_1 = \sqrt{\frac{\lambda+2\mu}{\rho}}$ and $c_2 = \sqrt{\frac{\mu}{\rho}}$. Equations (9.25), (9.28) are (9.29) the linear wave equations and are often noted as

$$\Box_c p = 0, \tag{9.30}$$

where $\Box$ is the d'Alembert operator, which is defined by $\Box_c = \nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}$. Using a unique notation for this equation adds to the fact that similarly to the diffusion equation, it can then be used in multiple fields. For example, the wave equation is also used in electromagnetism (see Section 10.1).

9.1.4 Plane waves

A simple solution to the wave equation is the plane wave. This is a wave whose features are independent along the directions that are normal to the wave propagation. By choosing the Cartesian base to have its x-axis aligned with this normal, (9.25) becomes

$$\frac{\partial^2 p}{\partial x^2} - \frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} = 0, \tag{9.31}$$

$$\left(\frac{\partial}{\partial x} - \frac{1}{c}\frac{\partial}{\partial t}\right)\left(\frac{\partial}{\partial x} + \frac{1}{c}\frac{\partial}{\partial t}\right) p = 0. \tag{9.32}$$

This is solve by nullifying either one of the factors, thus the general solution is the sum of two functions, f zeroing the left factor and g taking care of the right one. We obtain

$$p(x,t) = f(x+ct) + g(x-ct), \tag{9.33}$$

where f represents the backward propagation of a plane wave, and g the forward propagation, both at a speed of c. For a plane wave travelling in the direction of a unit vector $\boldsymbol{n}$, for any point $\boldsymbol{x}$ we thus have

$$p = g(\boldsymbol{x}\cdot\boldsymbol{n} - ct). \tag{9.34}$$

9.1.5 Acoustic impedance and intensity

The acoustic impedance Z represent the relation between the pressure and acoustic flow, and is similar to the electrical impedance. For a plane wave, the acoustic impedance is

$$Z = \frac{p}{v} = \rho c. \tag{9.35}$$

The sound intensity is defined by

$$I = pv. \tag{9.36}$$

It is a value that measures the power per unit area of a sound wave.

The effective pressure p_e is defined as the root mean square of the pressure,

$$p_e = \sqrt{< p^2 >_\tau} = \sqrt{\frac{1}{\tau}\int_0^\tau p^2\,dt}, \tag{9.37}$$

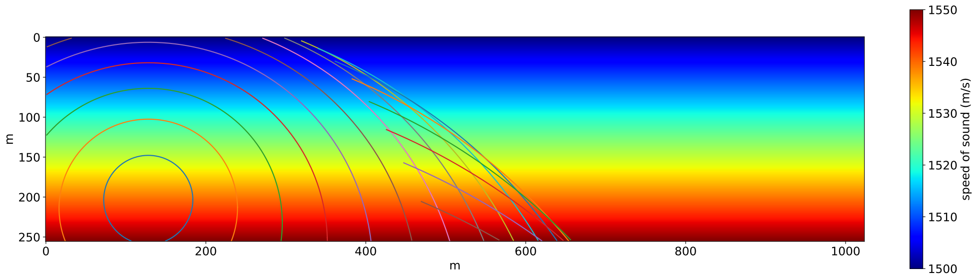


Figure 9.1: Example of multiple wavefront

where τ is the length of the signal. The effective velocity v_e is defined in the same way. With these definitions, the average sound intensity can be obtained,

$$I =< pv >_{\tau}= \frac{1}{\tau}\int_0^{\tau} pvdt = \frac{1}{\tau}\int_0^{\tau} \frac{p^2}{\rho c}dt = \frac{p_e^2}{\rho c} = p_e v_e. \tag{9.38}$$

9.1.6 Wavefront

A wavefront is a surface on which each point shares the same waveform characteristics. The theory of plane waves can be applied on each infinitesimal part of the surface of the wavefront, showing that the wavefront will locally move toward its normal at a speed of c. An example of a multiple wavefront is shown in Figure 9.1.

9.2 Nonlinear wave propagation

Please see Appendix A for details of the nonlinear theory. Although this theory is not used directly in the thesis, it is nonetheless very important and will enable a much finer analysis of the complex sound generation and propagation in the head of the sperm whale.

9.3 Conclusion

This chapter defined the basis of linear acoustic theory, with the key equations of wave propagation in solid and fluid, and the first order equations linking the stress and the velocity that were used to obtain them. These equations are key, since they are the foundation needed to build any kind of wave propagation simulation, and are a requirement to interpret the results produced by these models or further analyses of the signals obtained in real recordings. Moreover this chapter also introduces other useful concepts such as the acoustic impedance, or the notion of a plane wave.

Chapter 10

Numerical methods

With the equations presented in the previous chapter, a numerical method is needed in order to build a simulation. This chapter presents an overview of the existing numerical methods and their characteristics.

For simple geometries and homogeneous materials, analytical solutions to the acoustic propagation model can be found, which provide insight on what is happening and the mechanisms involved. However in most real world applications with irregular geometries and inhomogeneous media, analytical solutions cannot be found. One thus has to resort to the use of numerical methods to solve the model. Even though numerical methods give only an approximation of the solution, they have the advantage that they can easily adapt to any parameter changes (geometry, materials, etc.). Many numerical methods exist, each answering to different needs. One will thus choose the appropriate method depending on the needs (accuracy, types of material) and available resources, such as time, or CPU, GPU and memory usage.

10.1 Finite difference approximation

Finite difference methods (FDM) are a type of numerical method used to solve differential equations. FDM are usually chosen as a first approach, since they are easy to program and are able to give an estimation of the solution with a relatively short computation time [179]. Unlike other more precise methods, such as finite element methods, FDM are easier to implement. This allows the user to implement one's own program, which leads to an easier implementation of any additional custom code needed for data analysis, or adding effects that were not modelled by the differential equations.

For a differentiable function f, let us denote its derivative f'. Then Taylor's theorem implies that

$$\forall a \in \mathbb{R} \quad f(a+h) = f(a) + f'(a)h + \mathcal{O}(h^2). \tag{10.1}$$

The finite difference aproximation will thus replace the derivative f' in the differential equations with either the forward difference $\Delta_h(f)$, the backward difference $\nabla_h(f)$, or the central difference $\delta_h(f)$, divided by the step size h. These are defined by

$$\frac{\Delta_h(f)(a)}{h} = \frac{f(a+h)-f(a)}{h} = f'(a) + \mathcal{O}(h), \tag{10.2}$$

$$\frac{\nabla_h(f)(a)}{h} = \frac{f(a)-f(a-h)}{h} = f'(a) + \mathcal{O}(h), \tag{10.3}$$

$$\frac{\delta_h(f)(a)}{h} = \frac{f(a+\frac{h}{2})-f(a-\frac{h}{2})}{h} = f'(a) + \mathcal{O}(h^2). \tag{10.4}$$

Higher order differences also exist, defined recursively by $\forall n \in \mathbb{N}^*, \quad \Delta_h^n(f) = \Delta(\Delta_h^{n-1}(f))$, with $\nabla_h^n(f)$ and $\delta_h^n(f)$ defined in a similar manner. Using the Binomial theorem, this can be expanded into

$$\Delta_h^n(f)(a) = \sum_{k=0}^{n}(-1)^k\binom{n}{k}f(a+(n-k)h), \tag{10.5}$$

$$\nabla_h^n(f)(a) = \sum_{k=0}^{n}(-1)^k\binom{n}{k}f(a-kh), \tag{10.6}$$

$$\delta_h^n(f)(a) = \sum_{k=0}^{n}(-1)^k\binom{n}{k}f(a+(\frac{n}{2}-k)h). \tag{10.7}$$

The n^{th} order derivatives $f^{(n)}$ can also be approximated using the n^{th} order difference,

$$f^{(n)} = \frac{\delta_h^n(f)}{h^n} + \mathcal{O}(h^{n+1}) = \frac{\Delta_h^n(f)}{h^n} + \mathcal{O}(h^n) = \frac{\nabla_h^n(f)}{h^n} + \mathcal{O}(h^n). \tag{10.8}$$

10.2 Finite element approximations

Another method to numerically solve partial differential equation is the finite element method (FEM). FEM will solve the problem by using a mesh of finite elements, which are small subdivisions of the global domain Ω. Once the subdivision is done, the partial differential equations are locally approximated for each element. Once approximated, the partial differential equations become ordinary differential equations for transient problems, and are algebraic equations for steady state problems. The problem is solved at the mesh nodes, and basis functions (also called shape functions) are used to interpolate between the values obtained at the nodes.

Another way to explain the FEM, is that it will try to approximate the solution u with $\hat{u}$, which is the weighted sum over the finite elements of the local solutions u_i, using the basis functions N_i as weights,

$$u(\boldsymbol{x},t) \approx \hat{u}(\boldsymbol{x},t) = \sum_{i=1}^{n} N_i(\boldsymbol{x})u_i(t), \tag{10.9}$$

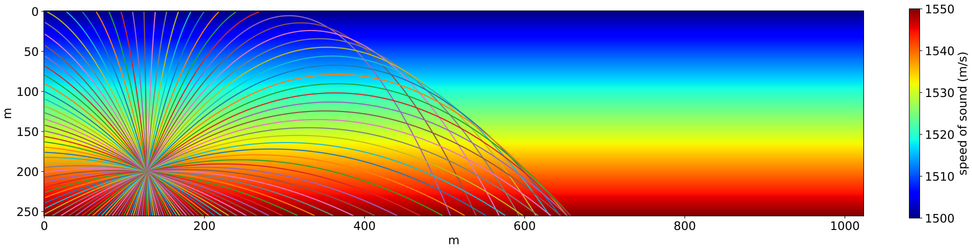


Figure 10.1: Example of multiple rays

with N_i having a non zero value only around its node. For example, in a 1D domain split into finite elements at the nodes x_i, the shape function N_i could be defined by

$$N_i = \begin{cases} \frac{x-x_i}{x_i-x_{i-1}} & \text{if} \quad x \in [x_{i-1}, x_i], \\ \frac{x_{i+1}-x}{x_{i+1}-x_i} & \text{if} \quad x \in [x_i, x_{i+1}], \\ 0 & \text{otherwise.} \end{cases} \tag{10.10}$$

10.2.1 Spectral element method

The spectral element method [180] is a special case of FEM, where the basis functions are high degree piecewise polynomials. Since the spectral element method uses high degree piecewise polynomials basis functions, it can achieve a high order of accuracy.

10.3 Ray tracing

In an analogy to light, acoustics can be modelled using rays. In [181], a ray is defined as the curve whose tangent at each point is in the direction of the velocity of the waveform that passes through the point. An example is shown in Figure 10.1.

In a medium moving at a velocity $\boldsymbol{v}$, if the wave velocity is $c\boldsymbol{n}$, where $\boldsymbol{n}$ is the normal to the waveform and c the speed of sound at the studied position $\boldsymbol{x}$, the velocity $\boldsymbol{v}_{\text{ray}}$ of the point will be

$$\boldsymbol{v}_{\text{ray}} = \boldsymbol{v} + c\boldsymbol{n}. \tag{10.11}$$

For any point $\boldsymbol{x}$ in space, let us define the function τ which returns the first time at which the wavefront passes through $\boldsymbol{x}$ (any further passage can still be studied with the same following process by restricting the function to some time interval). Let us also call s the gradient of τ so that $s = \nabla\tau(\boldsymbol{x})$. Then

$$\boldsymbol{s} = \frac{\boldsymbol{n}}{c + \boldsymbol{v} \cdot \boldsymbol{n}}, \tag{10.12}$$

$$\boldsymbol{n} = \frac{c\boldsymbol{s}}{1 - \boldsymbol{v} \cdot \boldsymbol{s}}. \tag{10.13}$$

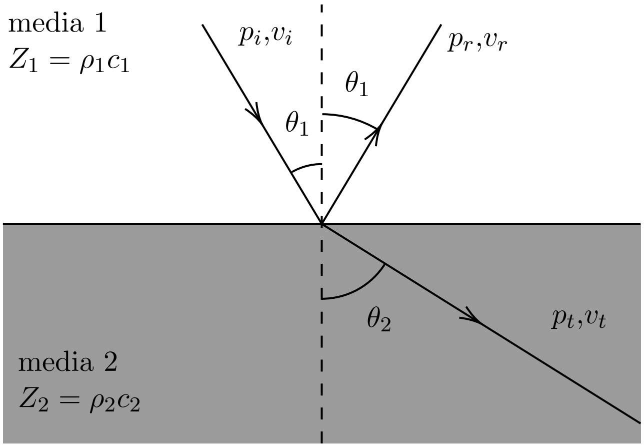


Figure 10.2: Ray of sound splitting into reflected and transmitted sound at the interface of two media.

Differentiating, we obtain the ray equations,

$$\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} = \frac{c^2\boldsymbol{s}}{1-\boldsymbol{v}\cdot\boldsymbol{s}} + \boldsymbol{v}, \tag{10.14}$$

$$\frac{\mathrm{d}\boldsymbol{s}}{\mathrm{d}t} = -\frac{1-\boldsymbol{v}\cdot\boldsymbol{s}}{c}\nabla c - (\boldsymbol{s}\cdot\nabla)\boldsymbol{v} - \boldsymbol{s}\times(\nabla\times\boldsymbol{v}). \tag{10.15}$$

In the simpler case when there is no ambient velocity $\boldsymbol{v}$, the equations can be simplified to

$$\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} = c^2\boldsymbol{s}, \tag{10.16}$$

$$\frac{\mathrm{d}\boldsymbol{s}}{\mathrm{d}t} = -\frac{1}{c}\nabla c. \tag{10.17}$$

If c is also constant, then $\boldsymbol{s}$ will be constant, leading to a constant normal $\boldsymbol{n}$, meaning that the rays are lines.

10.3.1 Interfaces and diffraction

Similarly to light, a ray of sound will follow Snell's law at the interfaces between two media, as shown in Figure 10.2.

We have

$$\frac{1}{c_1}\sin\theta_1 = \frac{1}{c_2}\sin\theta_2. \tag{10.18}$$

For a normal incident wave ($\theta_1 = 0$), the pressure must be continuous at the boundary, meaning that the sum of the incident wave pressure p_i and the reflected wave pressure p_r must equate the value of the transmitted wave pressure p_t. The same continuity condition holds for the velocity, thus

$$p_i + p_r = p_t, \tag{10.19}$$
$$v_i - v_r = v_t. \tag{10.20}$$

By using the acoustic impedance, the latter equation can be rewritten as $\frac{p_i}{Z_1} - \frac{p_r}{Z_1} = \frac{p_t}{Z_2}$. This pair of equations leads to relations between the reflection and transmission coefficients and the acoustic impedances,

$$r = \frac{p_r}{p_i} = \frac{Z_2 - Z_1}{Z_2 + Z_1}, \tag{10.21}$$
$$t = \frac{p_t}{p_i} = \frac{2Z_2}{Z_2 + Z_1}. \tag{10.22}$$

Equation (10.21) is useful to derive global behaviour of waves interacting with media interfaces. The first simple case is when the two impedances are matching (i.e. $Z_1 = Z_2$). In that case there will be no reflection, with everything being transferred from one medium to the other. The other case will be when one impedance is far greater than the other. In this case, the wave will be totally reflected, with little or no energy going to the other medium. Note that r will be close to 1 if Z_1 is small compared to Z_2, whereas r will tend to -1 if Z_1 is large compared to Z_2, meaning that the reflected wave will be in opposition of phase.

In the case of the sperm whale, the complete acoustic path is composed of media with an acoustic impedance close to that of water. On the other hand, the acoustic impedance of the air of the distal and frontal sac is small compared to the impedance of the spermaceti oil, explaining why they are often referred to as acoustic mirrors.

Note that by using (10.18), one can deduce the value of the transmission and reflection coefficients for any angle of incidence θ_1,

$$r = \frac{p_r}{p_i} = \frac{Z_2 \cos\theta_1 - Z_1 \cos\theta_2}{Z_2 \cos\theta_1 + Z_1 \cos\theta_2}, \tag{10.23}$$
$$t = \frac{p_t}{p_i} = \frac{2Z_2 \cos\theta_1}{Z_2 \cos\theta_1 + Z_1 \cos\theta_2}. \tag{10.24}$$

In an analogy to the study of light, the intensity ratio can be deduced, which in the case of a normal incidence gives

$$R = \frac{I_r}{I_i} = \left(\frac{Z_2 - Z_1}{Z_2 + Z_1}\right)^2, \tag{10.25}$$
$$T = \frac{I_t}{I_i} = \frac{4Z_2Z2}{(Z_2 + Z_1)^2}, \tag{10.26}$$
$$T + R = 1. \tag{10.27}$$

Regarding the diffraction, [182] developed the geometrical theory of diffraction which allows rays with discontinuous direction.

10.4 Numerical stability

The Courant-Friedrichs-Lewy (CFL) criterion [183] is a criterion that determines if a numerical simulation is stable, meaning that the values of the solution will remain bounded. Stability does not mean accuracy. In fact, using a finer mesh to increase accuracy may invalidate the CFL criteria.

To satisfy the CFL criteria, the Courant number C needs to be lower than a limit $C_{\max}$

$$C = \Delta t \sum_{i=1}^{n} \frac{v_i}{\Delta x_i} \leq C_{\max}, \tag{10.28}$$

where v_i is the magnitude of the velocity in the i^{th} dimension, Δt is the time step size and Δx_i is the length interval in the i^{th} dimension. $C_{\max}$ depends on the discretization method used.

The main outcome of this numerical stability condition is that the time step size needs to be proportionally smaller than the space step size. There are two reasons to decrease the space step size. The first one is to obtain a solution with a higher spatial resolution. The second one is the need to simulate higher frequencies, since the space step size determines the lowest wavelength that can be approximated. One will thus usually determine a space step size to validate accuracy criteria, then chose a time step size to validate the CFL criterion.

The complexity of a 3D numerical solver is $\mathcal{O}(N_T N_n)$ where N_T is the number of time steps and N_n the number of nodes, since a step size of the solver is usually linearly dependent on the number of nodes N_n. The number of nodes is itself proportional to $\frac{V_{\text{sim}}}{\Delta x^3}$ where V is the volume simulated. Since $N_T = \frac{T_{\text{sim}}}{\Delta t}$, with T the total time simulated, the complexity becomes $\mathcal{O}(\frac{V_{\text{sim}} T_{sim}}{\Delta t \Delta x^3})$, or $\mathcal{O}(\frac{1}{\Delta t \Delta x^3})$ if both V_{sim} and T_{sim} are not considered as parameters.

Finally the CFL condition simplifies the complexity to $\mathcal{O}(\frac{1}{\Delta x^4})$ or $\mathcal{O}(f_m^4)$ with f_m the maximum frequency of the simulation. This means that for the same problem, refining the mesh by a factor 2 will increase the number of operation by a factor of $2^4 = 16$. Note that the spatial complexity (memory consumption) is proportional to the number of nodes N_n, thus a complexity of $\mathcal{O}(\frac{1}{\Delta x^3})$, implies that the same refinement will also increase the memory consumption by a factor of 8.

Thankfully, the update of each node can be done independently of the other nodes. Each time step still needs to be done sequentially, but the computation time of each single time step will gain from parallelizing the problem as predicted by Amdahl's law [184]. This is why most implementations of the aforementioned numerical methods are done on HPC (high performance computer) systems or GPUs. Note that if the number of nodes N_n already surpasses the available processing power, then all the nodes will not be processed simultaneously, and despite a speedup provided by a partial parallelization, the total computation time will still grow following the complexity of $\mathcal{O}(\frac{1}{\Delta x^4})$.

10.5 Conclusion

As presented in this chapter, multiple numerical methods exist to approximate a solution of a problem which has no (easily obtainable) analytic solution. All these methods have their pros and cons, but are usually compared and chosen based on the two opposing criteria that are the accuracy, and the computation time, with sometimes an added criterion for the easiness of implementing additional custom code. The CFL criterion was also introduced, with its implication on the temporal complexity.

Chapter 11

Sound Propagation in the Head of Sperm Whales

Amongst the previously described methods, the finite difference method was chosen to build the simulation for its easiness of implementation, and the speed of the simulation, allowing for more time steps for a given spatial discretization. In this chapter, this method is used in the time domain in order to model the wave propagation inside the sperm whale's head.

Since the 90's [185], scientists have been modeling the propagation of vocalized sound waves in marine mammals' heads. The ability to model wave propagation in marine mammals allows a better understanding of the interaction between all the organs responsible for the sound creation, or the molding of the sound wave, to achieve the highly directive beam pattern of such species [186, 187]. To the best of our knowledge these types of simulations have not been performed on the biosonar of sperm whales.

Most of these simulations are based on anatomical data derived from computed tomography (CT) scans. This information enables the construction of the model geometry, and to obtain the mechanical parameters for each material and their location (up to the CT scan resolution). However, most of the employed scans were performed on postmortem individuals. In [188], data was compared between dead and live specimens and their effects on the simulations. Dead specimens are prone to introducing artifacts in the model, such as air-filled blood vessels, but will not suffer from scanning errors due to the movement of a living specimen. However these deviations are likely not to change the mechanical parameters of the various tissues, and thus the Hounsfield unit that the CT-scan will measure, has been shown [189] to be well correlated to the density and speed of sound.

11.1 Model

Unlike other small marine mammals, sperm whales cannot be CT-scanned by normal means due to their size and weight. The only tomography data available have been performed on postmortem neonate sperm whales [90, 190]. However, those models cannot be simply scaled up since some anatomical elements do not match those of adult individuals obtained from dissections, such as the one shown in [96, 191]. In order to shape our model we have used dissection data. We model each organ using Computer Assisted Design (CAD) software based on the slices from [96]. Since single blueprints did not match each other exactly, we had to scale some of them, or take the mean shape. While the shapes used to model the junk are basic and might not model some focusing effects, its complex shape might have evolved to serve a structural aspect [192], with less influence on the acoustic effect.

For our numerical experiments we had to chose the mechanical parameters for each of the simulated media (skin, bones, *spermaceti*, water, etc.). While FDTD and our model are able to cope with anisotropic coefficients, for the sake of simplicity, in this first approach we have made an isotropic assumption. We have combined the measurements of [193] (assuming a temperature of 30 °C and atmospheric pressure), [96] and the measurements done on the *Kogia breviceps* in [194]. For the parameters not found in the literature, we have used values from the human body, based on the observation that the other parameter values are shared between the species (*Physeter macrocephalus, Kogia breviceps, Ziphius cavirostris, Homo sapiens sapiens*). The little variation introduced by the values borrowed from the other species will not have a significant impact on the results, since even a change of the order of 5% to 10% has little effect on the resultant beam [186]. The most important factor for the position of the various focal points is the geometry of the organs.

11.2 Finite difference in time domain

As stated in its name, Finite Difference Time Domain (FDTD) is a finite difference method. This method was developed by Kane S. Yee for solving Maxwell's equations on a transitory electromagnetic field [195]. This method was latter adapted to acoustics in elastic media [196, 197].

We will formulate below a discrete FDTD form of the following initial-boundary value problem. In the fluid regions, Ω_f, we will solve the acoustic wave equation system,

$$\rho\frac{\partial \boldsymbol{v}}{\partial t} = -\nabla p \quad \text{in } \Omega_\mathrm{f}\times[0,T], \tag{9.22}$$

$$\frac{\partial p}{\partial t} = -\rho c^2 \nabla\cdot\boldsymbol{v} \quad \text{in } \Omega_\mathrm{f}\times[0,T]. \tag{9.23}$$

In the solid regions, Ω_s, we will solve the elastic wave equation system,

$$\rho\frac{\partial v_i}{\partial t} = \sum_{j=1}^{3}\frac{\partial \sigma_{ij}}{\partial x_j} \quad \text{in } \Omega_\mathrm{s}\times[0,T], \tag{9.17}$$

$$\frac{\partial \sigma_{ij}}{\partial t} = \frac{1}{2}\sum_{k=1}^{3}\sum_{l=1}^{3} c_{ijkl}\left(\frac{\partial v_k}{\partial x_l}+\frac{\partial v_l}{\partial x_k}\right) \quad \text{in } \Omega_\mathrm{s}\times[0,T]. \tag{11.1}$$

The acoustic source will be simulated as a forcing term, $f(x,t)$, on the right-hand side of the pressure equation (9.23), or (11.1), depending on whether it is located in the fluid or solid regions, respectively. The exact form of the forcing function will be described below. To complete this system, we add the following boundary conditions:

- On the exterior, fluid boundary, an absorbing boundary condition on p.
- On the interior boundaries, between different materials, interface conditions that are described below.

Finally, the initial conditions are set equal to zero for p, v and σ since a forcing function is used.

11.2.1 FDTD applied to acoustics equation

The FDTD method uses a central difference approximation to approximate the space and time partial derivatives. In a grid using a spacing Δx_i in the direction $\boldsymbol{e}_i$, the fluid equations (9.22) and (9.23) become with a time step size Δt,

$$\rho\frac{v_i(t+\frac{\Delta t}{2},\boldsymbol{x})-v_i(t-\frac{\Delta t}{2},\boldsymbol{x})}{\Delta t} = -\frac{p(t,\boldsymbol{x}+\frac{\Delta x_i}{2}\boldsymbol{e}_i)-p(t,\boldsymbol{x}-\frac{\Delta x_i}{2}\boldsymbol{e}_i)}{\Delta x_i}, \tag{11.2}$$

$$\frac{p(t+\frac{\Delta t}{2},\boldsymbol{x})-p(t-\frac{\Delta t}{2},\boldsymbol{x})}{\Delta t} = -\rho c^2\sum_{i=1}^{3}\frac{v_i(t,\boldsymbol{x}+\frac{\Delta x_i}{2}\boldsymbol{e}_i)-v_i(t,\boldsymbol{x}-\frac{\Delta x_i}{2}\boldsymbol{e}_i)}{\Delta x_i}. \tag{11.3}$$

Similarly, by using the velocity $v_i=\frac{\partial\xi_i}{\partial t}$, the elastic solid equations (9.17), (9.18), and (9.19) become

$$\rho\frac{v_i(t+\frac{\Delta t}{2},\boldsymbol{x})-v_i(t-\frac{\Delta t}{2},\boldsymbol{x})}{\Delta t} = \sum_{j=1}^{3}\frac{\sigma_{ij}(t,\boldsymbol{x}+\frac{\Delta x_j}{2}\boldsymbol{e}_j)-\sigma_{ij}(t,\boldsymbol{x}-\frac{\Delta x_j}{2}\boldsymbol{e}_j)}{\Delta x_j}, \tag{11.4}$$

$$\begin{aligned}\frac{\sigma_{ij}(t+\frac{\Delta t}{2},\boldsymbol{x})-\sigma_{ij}(t-\frac{\Delta t}{2},\boldsymbol{x})}{\Delta t} &= \frac{1}{2}\sum_{k=1}^{3}\sum_{l=1}^{3}c_{ijkl}\Bigg(\frac{v_k(t,\boldsymbol{x}+\frac{\Delta x_l}{2}\boldsymbol{e}_l)-v_k(t,\boldsymbol{x}-\frac{\Delta x_l}{2}\boldsymbol{e}_l)}{\Delta x_l}\\ &\quad+\frac{v_l(t,\boldsymbol{x}+\frac{\Delta x_k}{2}\boldsymbol{e}_k)-v_l(t,\boldsymbol{x}-\frac{\Delta x_k}{2}\boldsymbol{e}_k)}{\Delta x_k}\Bigg).\end{aligned} \tag{11.5}$$

As shown in both fluid and solid equations, The FDTD method will create two grids, a stress/pressure grid and a velocity grid, with the latter shifted one-half a space step size from the former. In the solid case, a third grid is needed for the cross stress $(\sigma_{23},\sigma_{13},\sigma_{12})$, as seen in Figure 11.1 which is tiled into Figure 11.2.

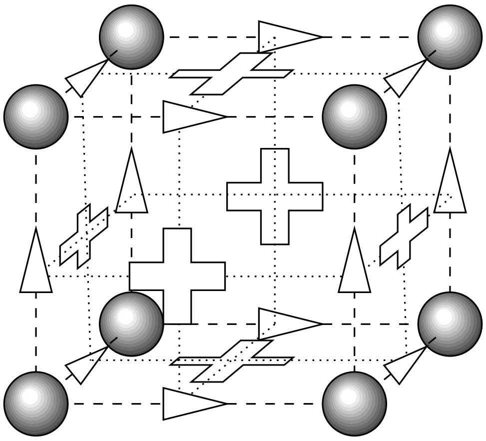


Figure 11.1: Part of the FDTD grid
Sphere: normal stress. Triangle: velocity. Cross: cross-stress.

For an isotropic elastic solid, (11.5) can be simplified to

$$\begin{aligned}\frac{\sigma_{ij}(t+\frac{\Delta t}{2},\boldsymbol{x})-\sigma_{ij}(t-\frac{\Delta t}{2},\boldsymbol{x})}{\Delta t} = & \quad \mu\frac{v_i(t,\boldsymbol{x}+\frac{\Delta x_j}{2}\boldsymbol{e}_j)-v_i(t,\boldsymbol{x}-\frac{\Delta x_j}{2}\boldsymbol{e}_j)}{\Delta x_j} \\ & + \quad \mu\frac{v_j(t,\boldsymbol{x}+\frac{\Delta x_i}{2}\boldsymbol{e}_i)-v_j(t,\boldsymbol{x}-\frac{\Delta x_i}{2}\boldsymbol{e}_i)}{\Delta x_i} \\ & + \lambda\delta_i^j\frac{v_i(t,\boldsymbol{x}+\frac{\Delta x_i}{2}\boldsymbol{e}_i)-v_i(t,\boldsymbol{x}-\frac{\Delta x_i}{2}\boldsymbol{e}_i)}{\Delta x_i},\end{aligned} \tag{11.6}$$

where δ_i^j is the Kronecker's delta.

Instead of using the pair of equations (11.2) and (11.3), fluids can be modelled using the isotropic elastic solid equations (11.4) and (11.6) by letting $\mu_{\text{fluid}} = 0$ and $\lambda_{\text{fluid}} = \rho_{\text{fluid}}c^2_{\text{fluid}}$. In this case the pressure will be equal to minus the average normal stress (9.11). Thus to compute the pressure at a node, the average of the stress on this node should be taken. One could also take the value of only one of the stress components on this node, as they should all have the same value since they are updated by the same amount, except for some numerical erosion, or a different initialisation which should not be done since they model fluid nodes.

Having the pressure node modelled this way, will increase the memory consumption and also increase the computation time, but allow to have a simpler program when both fluid and solid materials are present. The other way to model a volume containing both fluid and solid is to ignore the shear wave modes and model the solids as equivalent fluids [198].

After the initialisation, the grid will be updated in a leapfrog manner, alternating between the stress update and the velocity update as shown in Figure 11.3. One iteration of the algorithm is one time step size, meaning one update of the velocity and one update of the pressure. When updating elements on the border of the grid, the elements needed

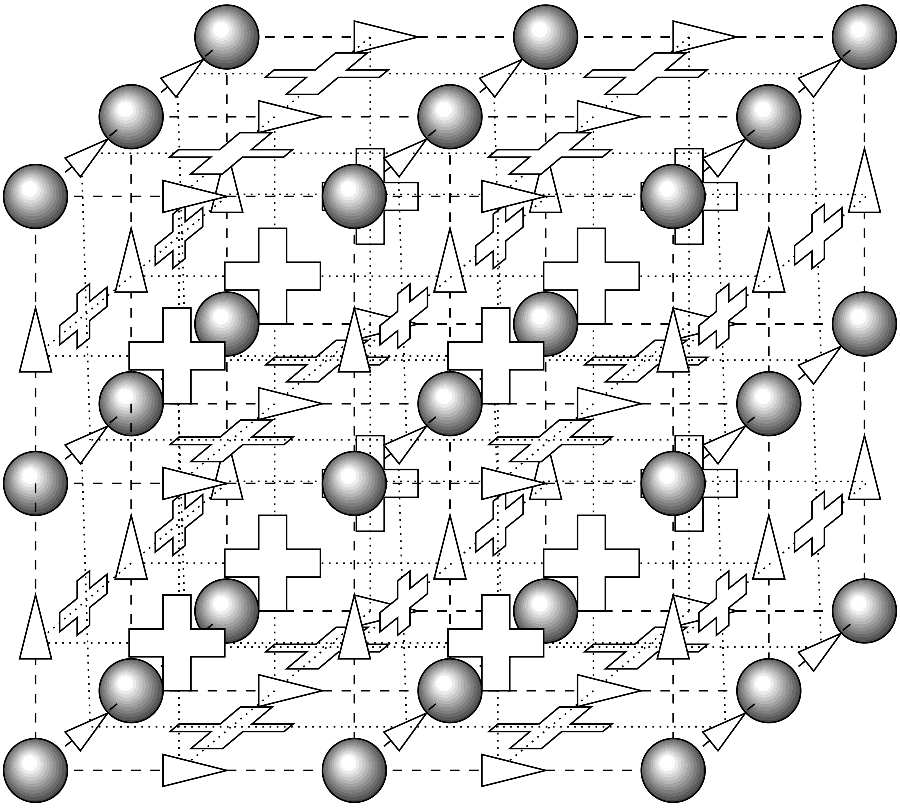


Figure 11.2: Part of the FDTD grid (tiled once more in every direction compared to Figure 11.1).

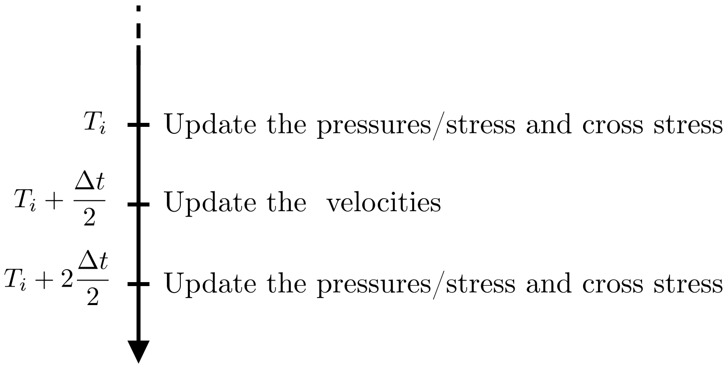


Figure 11.3: Update scheme of the FDTD at iteration T_i.

for the update that are outside of the grid will be considered as zero. Depending on the type of node chosen (either pressure or velocity) to be on the outer layer, the border of the simulation will act as an infinitely stiff wall or as void. In both cases, any incoming wave will be reflected.

11.2.2 Absorbing boundary condition

As usual for this kind of simulation, where we have a finite-sized object of interest (the sperm whale, or its head) evolving in an infinite domain (the ocean), the aim is to simulate the target body inside an infinite medium. To do this, we must use a finite simulation domain. The standard way of getting rid of reflections from the border of the simulation, and thus simulating a infinite medium while treating only a finite box, consists either of having multiple dampening layers near the border, or having special equations for the border that will make them 'invisible' to waves. All of those methods are always an approximation and will still produce some reflections in certain cases. For FDTD, multiple Absorbing Boundary Conditions (ABC) have been developed, starting with [199]. Since ABCs only deals with the nodes near the surface of the simulation, the computation time is usually multiple orders of magnitude smaller compared to the update of the computation time needed to update the nodes inside the whole volume of the simulation.

Initially, we have used the ABC from [200], with angles of 2.86 ° (0.05 rad) and 65 °. The computation time of the boundary update is negligible compared to the stress and speed update (two orders of magnitude in our case), and we could have increased the number of angles of incidence with perfect absorption without any perceptible decrease in performance. However we consider this number of absorption angles to be enough.

In [200] it is shown that for a wave with an incidence angle θ (the angle between the normal of the boundary and the normal of the wave), the boundary condition

$$\left(\cos\theta\frac{\partial}{\partial t} - c\frac{\partial}{\partial x}\right)p = 0 \tag{11.7}$$

will totally absorb the reflection. Thus for multiple absorption angles (α_i), the boundary condition

$$\left(\prod_{i=1}^{n}\left(\cos\alpha_i\frac{\partial}{\partial t}-c\frac{\partial}{\partial x}\right)\right)p=0 \tag{11.8}$$

can be used. This boundary condition can then be discretized using the box scheme operator that was proposed and developed in [201, 202],

$$\left(\prod_{i=1}^{n}\left(\cos\alpha_i\frac{I-Z^{-1}}{\Delta t}\frac{I+K}{2}-c\frac{I-Z^{-1}}{2}\frac{I+K}{\Delta x}\right)\right)p=0, \tag{11.9}$$

where I is the identity, K is the space shift operator respectively to the normal $\boldsymbol{n}$ of the boundary, and Z is the time shift operator, meaning that

$$K\left(p(\boldsymbol{x},t)\right)=p(\boldsymbol{x}+\Delta x\boldsymbol{n},t), \tag{11.10}$$

$$Z^{-1}\left(p(\boldsymbol{x},t)\right)=p(\boldsymbol{x},t+\Delta t). \tag{11.11}$$

The reflection coefficient r defined as the ratio between the amplitude of the reflection and outgoing wave can be calculated as follows, when using (11.8) as a boundary condition,

$$r(\theta)=-\prod_{i=1}^{n}\frac{\cos\alpha_i-\cos\theta}{\cos\alpha_i+\cos\theta}. \tag{11.12}$$

This means that for this boundary condition, a wave with a tangential incidence ($\theta = 90°$) will always have a reflection coefficient of 1, no matter the number of absorption angles. However these kinds of waves do not matter as they do not enter the rest of the simulation. The choice of the absorption angles should be done by trying to keep the reflection coefficient below a desired threshold, knowing that the angle close to 90° will not satisfy this threshold, unless choosing other absorption angles close to them. An absorption angle near 0° should be used to prevent the main modes of the simulation box. Using an absorption angle near 0° instead of exactly 0° allows to have a wave with an incident angle near 90° matter less, as they stay close to the border and will be absorbed by another wall of the simulation as a wave with a normal incidence.

Figure 11.4 shows the behaviour of the reflection for an angle α_1 of 2.86° and various values of α_2. As the absorption angle tends to 90°, $(\cos\alpha_i-\cos\theta)/(\cos\alpha_i+\cos\theta)$ will tend to 1 (the reflection coefficient made of one absorption angle of 90° is a constant 1) for incident angles θ smaller than the absorption angle. One might choose to use a smaller absorption angle in order to have a greater impact on other incident angles than just around the chosen angle.

C-PML

An alternative to the Higdon ABCs are the so-called perfectly matched layers (PML) that are shown schematically in Figure 11.5. They were introduced by [203] for Maxwell's equations. PML have the propriety of having a theoretically zero reflection coefficient,

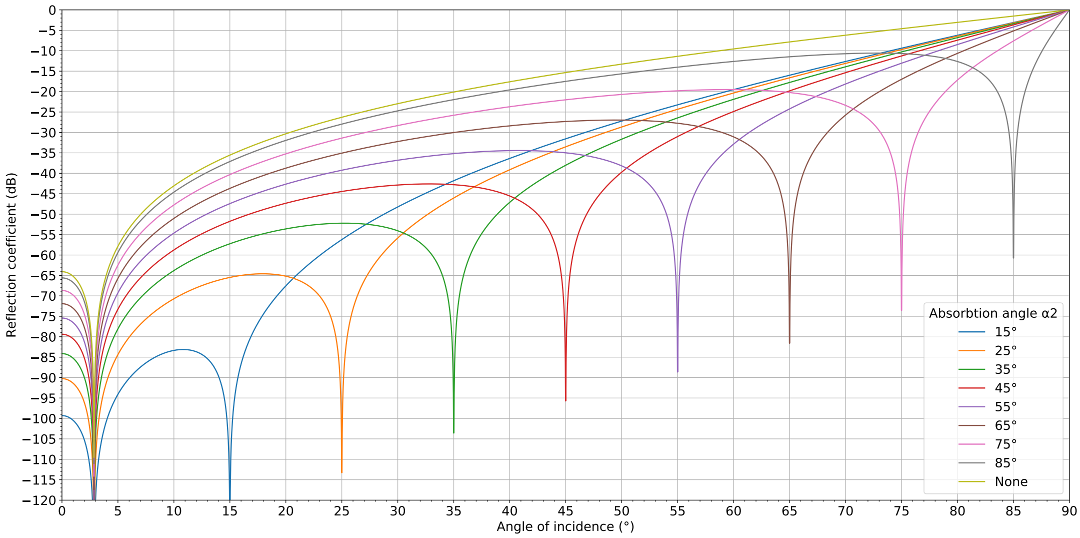


Figure 11.4: Reflection coefficient level for multiple absorption angle α_2.

for any incidence angle and any frequency. While this zero reflection coefficient does not hold once the discretization is done, this technique is widely applied as it is more efficient than other classical techniques. In [204], a comparison can be seen between Higdon ABC and the PML, in which the PML do perform better.

In [203] a reformulation based on a complex coordinate stretching technique was developed [205, 206], leading to a more efficient implementation of the PML, with

$$\tilde{x}_i = \int_0^{x_i} s_i(x_i')dx_i', \tag{11.13}$$

where s_i are complex-coordinate stretching functions. In the PML region, the spatial derivatives need to be replaced by a stretched version,

$$\frac{\partial}{\partial \tilde{x}_i} = \frac{1}{S_i(x_i,\omega)}\frac{\partial}{\partial x_i}, \tag{11.14}$$

$$S_i(x_i,\omega) = 1 + \frac{d_i(x_i)}{j\omega}, \tag{11.15}$$

where ω is the angular frequency and $j = \sqrt{-1}$. Here, $d_i(x_i)$ is the damping profile of the PML region, thus equal to 0 outside the border. In the time domain, (11.14) becomes

$$\frac{\partial}{\partial \tilde{x}_i} = s_i(x_i,t) *_t \frac{\partial}{\partial x_i}, \tag{11.16}$$

where $*_t$ is the temporal correlation, hence the name convolutional PML, or C-PML, and s_i is the inverse Fourier transform of S_i. As said before, the discrete version of the PML does not absorb perfectly all incoming waves, especially for incoming waves that

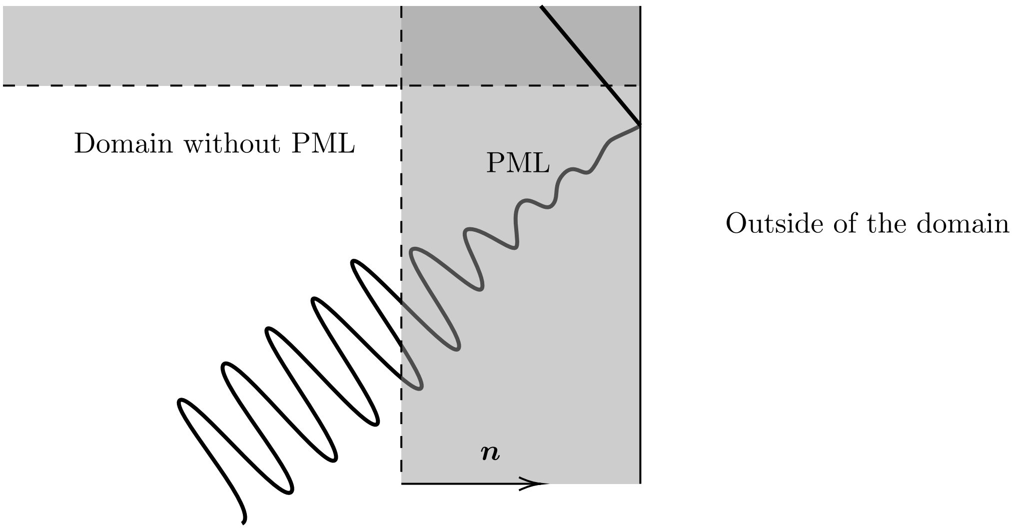


Figure 11.5: Slice of the domain showing the C-PML region with its normal $\boldsymbol{n}$.

are almost orthogonal with the normal of the border–the so-called grazing angles. This was solved by changing the scaling factor S_i [206, 207, 208] which becomes,

$$S_i(x_i,\omega) = \kappa_i(x_i) + \frac{d_i(x_i)}{\beta_i(x_i) + j\omega} \quad s_i(x_i,t) \qquad = \frac{1}{\kappa_i(x_i)}\delta(t) + \zeta_i(x_i,t), \tag{11.17}$$

where δ is the Dirac distribution, κ_i and β_i are profile parameters and $\zeta_i(x_i,t)$ is

$$\zeta_i(x_i,t) = \frac{d_i(x_i)}{\kappa_i(x_i)^2} e^{-(\frac{d_i(x_i)}{\kappa_i(x_i)} + \beta_i(x_i))t} H(t), \tag{11.18}$$

where H is the Heaviside distribution. The new spatial derivative which is now

$$\frac{\partial}{\partial \tilde{x}_i} = \frac{1}{\kappa_i(x_i)} \frac{\partial}{\partial x_i} + \zeta_i(t) *_t \frac{\partial}{\partial x_i} \tag{11.19}$$

needs a temporal convolution, which is not computationally friendly, since for each time step a sum over all the previous time steps is needed. Fortunately, it is possible to use a memory variable as in [209] to replace the computation of the correlation. Let us name $\psi_{x_i}(t)$ the memory variable at position x_i in the direction $\boldsymbol{e}_i$ at time t. Then (11.19) becomes

$$\frac{\partial}{\partial \tilde{x}_i} = \frac{1}{\kappa_i(x_i)} \frac{\partial}{\partial x_i} + \psi_{x_i}(t). \tag{11.20}$$

In the discrete case, using the superscript ^{n} to indicate time step n, the memory variable is updated using the following scheme,

$$\psi_{x_i}^n = b_i \psi_i^{n-1} + a_i(\frac{\partial}{\partial x_i})^{n-\frac{1}{2}}, \tag{11.21}$$

$$a_i(x_i) = \frac{d_i(xv)}{\kappa_i(x_i)(d_i(x_i) + \kappa_i(x_i)\beta_i(x_i)}(b_i(x_i) - 1), \tag{11.22}$$

$$b_i(x_i) = e^{-(\frac{d_i(x_i)}{\kappa_i(x_i)} + \beta_i(x_i))\Delta t}. \tag{11.23}$$

As in [208], $\kappa_i(x_i) = 1$ is used, thus simplifying the equations, which when applied to (11.4) and (11.6) become

$$\rho \frac{\partial v_i}{\partial t} = \sum_{j=1}^{3} \frac{\partial \sigma_{ij}}{\partial \tilde{x}_j} = \sum_{j=1}^{3} \left(\frac{\partial \sigma_{ij}}{\partial x_j} + \psi_{x_j,\sigma_{ij}}(t) \right), \tag{11.24}$$

$$\begin{aligned}\frac{\partial \sigma_{ij}}{\partial t} &= \mu \left(\frac{\partial v_i}{\partial \tilde{x}_j} + \frac{\partial v_j}{\partial \tilde{x}_i} \right) + \delta_i^j \lambda \sum_{k=1}^{3} \frac{\partial v_k}{\partial \tilde{x}_k} \\ &= \mu \left(\frac{\partial v_i}{\partial x_j} + \psi_{x_j,v_i}(t) + \frac{\partial v_j}{\partial x_i} + \psi_{x_i,v_j}(t) \right) + \delta_i^j \lambda \sum_{k=1}^{3} \left(\frac{\partial v_k}{\partial x_k} + \psi_{x_k,v_k}(t) \right),\end{aligned} \tag{11.25}$$

where $\psi_{x_i,u}(t)$ is $\psi_{x_i}(t)$ applied to the variable u according to (11.21) if the local PML boundary has $\boldsymbol{e}_i$ as its normal, otherwise $\psi_{x_i,u}(t) = 0$. Henceforth, in the corner where the PML domains of two faces overlap, the PML terms $\psi_{x_i,u}(t)$ of the two or more faces are added together. Note that the first terms of these equations are the update terms without the C-PML. Thus the update can be done in two steps, within each step of the leapfrog scheme that was described previously (see Figure 11.3). The first part is the usual update of the values that need to be updated at this time step, followed by the update of the memory variable which is then added to the previously updated value. Separating the update into two, allows us to keep the code for the standard update separated from the code for the C-PML update, which is both useful for easily changing the type of boundary treatment used, but also simplifies the cost of the update as the C-PML is only applied on a small region of the model. By using ˜ to indicate the variable updated with the standard scheme, the update equations become

$$v_i(t + \frac{\Delta t}{2}, \boldsymbol{x}) = \tilde{v}_i(t + \frac{\Delta t}{2}, \boldsymbol{x}) + \frac{\Delta t}{\rho} \sum_{j=1}^{3} \psi_{x_j,\sigma_{ij}}(t), \tag{11.26}$$

$$\sigma_{ij}(t + \frac{\Delta t}{2}, \boldsymbol{x}) = \tilde{\sigma}_{ij}(t + \frac{\Delta t}{2}, \boldsymbol{x}) + \mu \Delta t(\psi_{x_j,v_i}(t) + \psi_{x_i,v_j}(t)) + \lambda \Delta t \delta_i^j \sum_{k=1}^{3} \psi_{x_k,v_k}(t), \tag{11.27}$$

and the memory variables

$$\psi_{x_i,v_j}(t) = b_i \psi_{x_i,v_j}(t-\Delta t) + a_i \frac{v_j(t-\frac{\Delta t}{2}, \boldsymbol{x} + \frac{\Delta x_k}{2}\boldsymbol{e}_k) - v_j(t-\frac{\Delta t}{2}, \boldsymbol{x} - \frac{\Delta x_k}{2}\boldsymbol{e}_k)}{\Delta x_k}, \tag{11.28}$$

$$\psi_{x_i,\sigma_{ij}}(t) = b_i \psi_{x_i,\sigma_{ij}}(t-\Delta t) + a_i \frac{\sigma_{ij}(t-\frac{\Delta t}{2}, \boldsymbol{x} + \frac{\Delta x_i}{2}\boldsymbol{e}_i) - \sigma_{ij}(t-\frac{\Delta t}{2}, \boldsymbol{x} - \frac{\Delta x_i}{2}\boldsymbol{e}_i)}{\Delta x_i}. \tag{11.29}$$

Figure 11.6 shows a comparison of three simulations with different treatments of the boundary. The figure shows the evolution of the amplitude of the stress component σ_x along the x-axis at $y = 45$ and $z = 10$. The simulations were done in a 200*200*100 box, with the source placed at $(45, 45, 10)$. The simulation material is water. The source is the same as the one that will be used in the sperm whale simulation (11.47), which is a sinusoidal wave at 10 kHz multiply by half a period of a sinusoidal wave at 5 kHz to serve as an envelope. The parameters for the C-PML are taken similarly to those in [208], which in this case are,

$$N = 2, \tag{11.30}$$

$$L = 5, \tag{11.31}$$

$$c_p = 1500, \tag{11.32}$$

$$d_0 = -(N+1)c_p \frac{\log(0.01)}{2L}, \tag{11.33}$$

$$d(x) = d_0 \left(\frac{x}{L}\right)^N, \tag{11.34}$$

$$\beta_{\max} = \pi 10000, \tag{11.35}$$

$$\beta(x) = \beta_{\max}\left(1 - \frac{x}{L}\right). \tag{11.36}$$

Here c_p is the sound speed of water, L is the thickness of the PML region and x is the distance to the start of the considered PML boundary.

Since these simulations are homogeneously made in water, the speed of sound is constant throughout the simulation. For the direct signal coming from the source, and advancing though the line used in Figure 11.6, the wavefront will form a straight line in this figure with a slope of plus or minus the speed of sound in water. In the case without any boundary treatment, a parabolic shape starting at 0.25 ms can be seen. This shape is the reflection from the border $z = 0$. It is a parabola since the distance the wave needs to travel to reach position x on this line is made up of the difference between x and 45 (position of the source) for the $x$axis and the difference between the source and the $z = 0$ border plus the difference between the $z = 0$ border and the recorded line for the z-axis, leading to the equation $t = \frac{1}{c_p}\sqrt{(x-45)^2 + 20^2}$. Globally, reflection with any number of faces will form a parabolic curve in this figure, with the particular case of the two straight lines starting at approximately 0.4 ms and 1.2 ms which are the reflections on the $x = 0$ and $x = 200$ borders. We clearly observe the improvements due to the use of the C-PML.

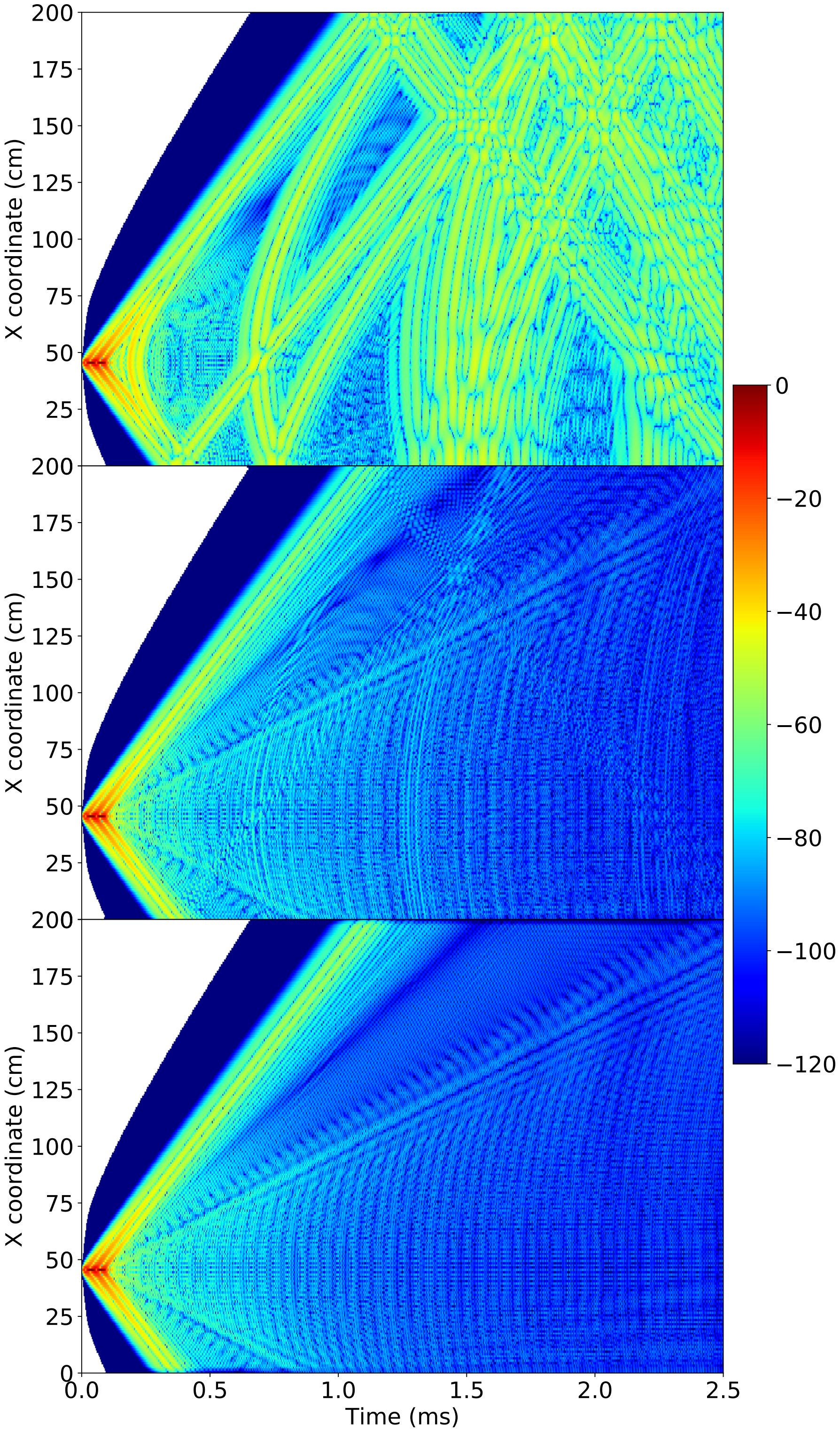


Figure 11.6: Comparison between different treatment of the limits of the simulation Top: nothing, middle: second order Higdon ABC, bottom : 5 layer C-PML. The color scale is in dB.

11.2.3 Grid initialisation

Before the first update of the velocities, the grids need to be initialised. If they are not initialised at rest (both velocities and stress nodes are set to zero), they can be initialised as the stress/velocity distribution matching a wave that needs to be studied. If the gird is initialised at rest, the model will stay at rest. To have a useful simulation, an outside source thus needs to be added. An outside source can be added in two manners. The first one is to manually set the value of the stress velocity field in the source volume amid the source activation period, discarding the update for those nodes. The second option is to include the source directly into the equation. While the former option allows a more precise control over the source nodes, and allows to simulate node values that would be totally controlled by the source mechanism, it prevents any incoming wave to pass through the source field. We will thus use the second option.

11.2.4 Material grids

As shown in the FDTD equations, physical characteristics of the materials are required for the simulation. For inhomogeneous materials, each node of both the stress and velocity grids needs to have access to the local values of these characteristics. This can be seen as a three-material grid, having respectively the shape of the two stress grids and the velocity grid, where each node contains the mechanical characteristics needed, such as the density. When the distribution of the mechanical parameters are described by an analytical function (these functions being constant functions for homogeneous media), the values on these grids can be calculated on the fly. When sufficient memory is available, it is preferable to store these values multiplied by the other constants (e.g. time step size, space step size, etc.) that will be used in the equations, as it will reduce the computation time of each iteration of the simulation.

The simulation of multiple materials requires the simulation of the boundary between these materials. The boundaries geometries are approximated and fitted to the nodes as seen in Figure 11.7. As shown in [210], these boundary mechanical values can be calculated to have a simulation generating a solution close to the analytical solution.

On a boundary between two materials (red and blue) whose normal is in the $\boldsymbol{e}_i$ vector, one can use the forward or backward difference to obtain equations that are defined in only one material, as shown in Figure 11.8. Let us use the subscript $_b$ and $_r$ to denote the blue and red material respectively, and let j and k be the indices of the two other directions (see [210] for a non uniform grid using different space step size per material). We write

$$\rho_r \frac{\partial v_i}{\partial t} = \frac{\Delta_{\Delta x_i/2}(\sigma_{ii})(t,\boldsymbol{x})}{\Delta x_i/2} + \frac{\delta_{\Delta x_j}(\sigma_{ij})(t,\boldsymbol{x})}{\Delta x_j} + \frac{\delta_{\Delta x_k}(\sigma_{ik})(t,\boldsymbol{x})}{\Delta x_k}, \tag{11.37}$$

$$\rho_b \frac{\partial v_i}{\partial t} = \frac{\nabla_{\Delta x_i/2}(\sigma_{ii})(t,\boldsymbol{x})}{\Delta x_i/2} + \frac{\delta_{\Delta x_j}(\sigma_{ij})(t,\boldsymbol{x})}{\Delta x_j} + \frac{\delta_{\Delta x_k}(\sigma_{ik})(t,\boldsymbol{x})}{\Delta x_k}. \tag{11.38}$$

These equations can then be combined to remove the stress node $\sigma_{ii}(\boldsymbol{x})$, which allows to obtain an equation using only nodes that were already present in our staggered grids,

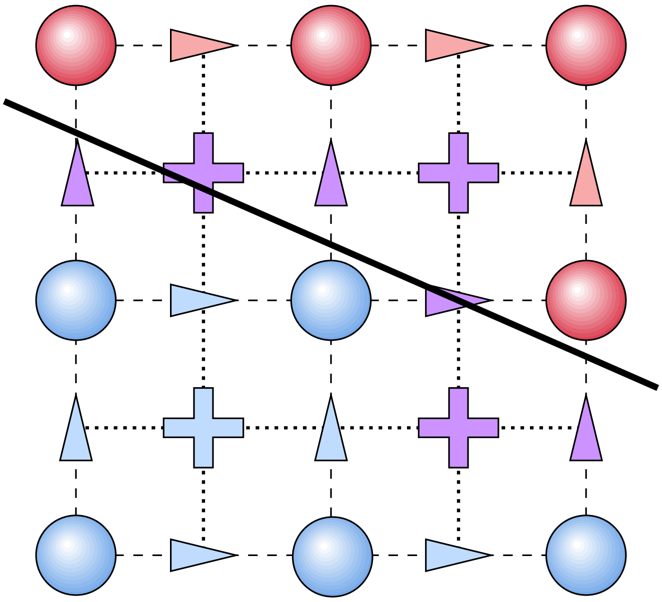


Figure 11.7: Material approximation in a 2D grid
The full line is the actual material boundary between the blue and red materials. The background is the discretized materials. The blue and red nodes use the materials characteristics, while the purple nodes will used a mixture of the two materials.

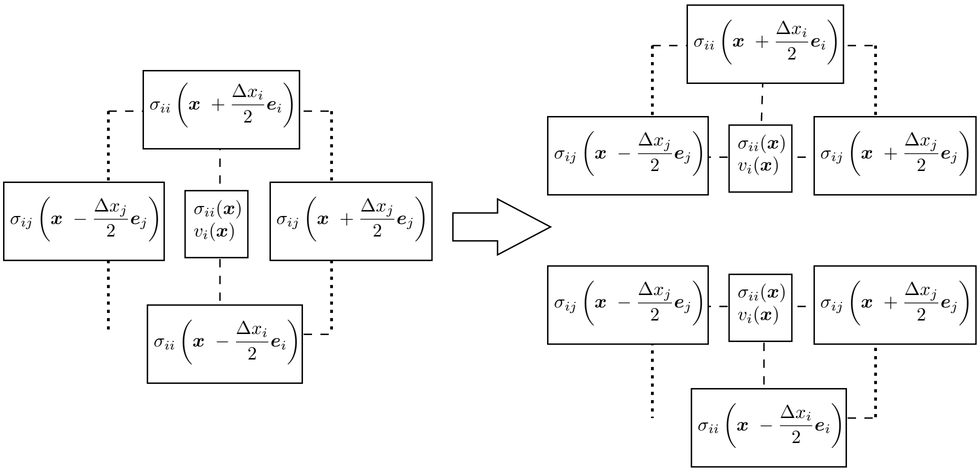


Figure 11.8: Separation of a two-medium velocity node problem into two one-medium subproblems.

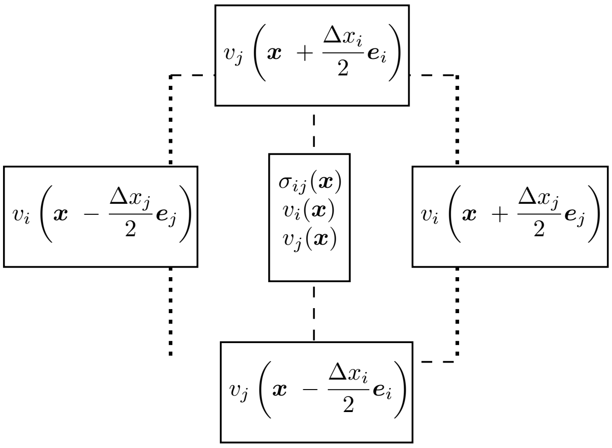


Figure 11.9: Four media layout around a cross stress node.

$$\frac{\rho_r+\rho_b}{2}\frac{\partial v_i}{\partial t}=\frac{\delta_{\Delta x_i}(\sigma_{ii})(t,\boldsymbol{x})}{\Delta x_i}+\frac{\delta_{\Delta x_j}(\sigma_{ij})(t,\boldsymbol{x})}{\Delta x_j}+\frac{\delta_{\Delta x_k}(\sigma_{ik})(t,\boldsymbol{x})}{\Delta x_k}. \tag{11.39}$$

Here, a comparison with (11.4) allows to identify a density $\bar{\rho}$ that can be used to simulate the boundary without changing the equation used,

$$\bar{\rho}=\frac{\rho_r+\rho_b}{2}. \tag{11.40}$$

For the cross stress, the same process can be applied to derive the Lamé coefficient μ. However, since up to four different media can surround the cross stress, the equation (11.6) needs to be rewritten four times using only forward and backward differences such that they are applied on only one medium. Thus, with the subscript $({}_r,{}_g,{}_b,{}_y)$ for the four media red, green, blue and yellow respectively, in a configuration shown by Figure 11.9. We have

$$\frac{1}{\mu_r}\frac{\partial\sigma_{ij}}{\partial t}=\frac{\nabla_{\Delta x_j/2}(v_i)(t,\boldsymbol{x})}{\Delta x_j/2}+\frac{\Delta_{\Delta x_i/2}(v_j)(t,\boldsymbol{x})}{\Delta x_i/2}, \tag{11.41}$$

$$\frac{1}{\mu_g}\frac{\partial\sigma_{ij}}{\partial t}=\frac{\Delta_{\Delta x_j/2}(v_i)(t,\boldsymbol{x})}{\Delta x_j/2}+\frac{\Delta_{\Delta x_i/2}(v_j)(t,\boldsymbol{x})}{\Delta x_i/2}, \tag{11.42}$$

$$\frac{1}{\mu_b}\frac{\partial\sigma_{ij}}{\partial t}=\frac{\nabla_{\Delta x_j/2}(v_i)(t,\boldsymbol{x})}{\Delta x_j/2}+\frac{\nabla_{\Delta x_i/2}(v_j)(t,\boldsymbol{x})}{\Delta x_i/2}, \tag{11.43}$$

$$\frac{1}{\mu_y}\frac{\partial\sigma_{ij}}{\partial t}=\frac{\Delta_{\Delta x_j/2}(v_i)(t,\boldsymbol{x})}{\Delta x_j/2}+\frac{\nabla_{\Delta x_i/2}(v_j)(t,\boldsymbol{x})}{\Delta x_i/2}. \tag{11.44}$$

We can sum the 4 medium-dependent equations to remove the two speed values $v_i(\boldsymbol{x})$ and $v_j(\boldsymbol{x})$,

$$\left(\frac{1}{\mu_r}+\frac{1}{\mu_g}+\frac{1}{\mu_b}+\frac{1}{\mu_y}\right)\frac{\partial\sigma_{ij}}{\partial t}=4\frac{\delta_{\Delta x_j}(v_i)(t,\boldsymbol{x})}{\Delta x_j}+4\frac{\delta_{\Delta x_i}(v_j)(t,\boldsymbol{x})}{\Delta x_i}. \tag{11.45}$$

Material	λ (GPa)	μ (GPa)	ρ (kg.m^{-3})	Calculated from	Reference
Water	2.250	0	1000	speed of sound 1500m/s	
Bone	9.808	6.538	1600	human bone	
Blubber	1.946	1.297	939	human with p-speed of 1730 m/s and Poisson ratio of 0.3, but with actual sperm whale blubber density	[96]
Spermaceti oil	1.702	0	832	speed of sound 1405m/s at 30°C	[193]
Air	0.1521	0	1.293	speed of sound	
Muscle	1.846	1.231	1075	human with p-speed of 1575m/s and Poisson ratio of 0.3	

Table 11.1: Mechanical parameters used in the simulation.

The parameter identification with (11.6) defines the boundary Lamé parameter value $\bar{\mu}$ as

$$\bar{\mu} = \frac{4}{\frac{1}{\mu_r} + \frac{1}{\mu_g} + \frac{1}{\mu_b} + \frac{1}{\mu_y}}. \tag{11.46}$$

11.3 Results

We simulated a sperm whale head in a $520 * 240 * 220$ cm^3 volume, with 1 cm resolution, and the materials were averaged following [210]. The simulation was implemented using PyTorch (a Deep Learning Python library) and run on an NVidia Titan X GPU. The implementation performs at 4.6 iterations per second. Thus for a simulation of 20 ms with a time step of 1 μs, the computation time is 1 h 12. The simulation starts at rest. We then add to pressure nodes located next to the *museau de singe* in the spermaceti, the difference of a 10 kHz sinusoidal wave during one period, defined by

$$\sigma_i(t) = \sin\left(2\pi \frac{t}{0.0001\,s}\right) * \sin\left(\pi \frac{t}{0.0001\,s}\right) \chi_{[0,0.0001\,s]}(t) \tag{11.47}$$

where $\chi_{[0,0.0001\,s]}$ is the characteristic function of the interval $[0, 0.0001\,s]$.

Figure 11.11 shows a recorded sound wave of a sperm whale click and the simulated pressure at the *museau de singe*. In both the recorded and simulated sounds we observe three pulses of a sperm whale click. In the simulated case these correspond to P0, P1 and P2. In the simulation we measure an offset of 6662 bins (or μs) between each of

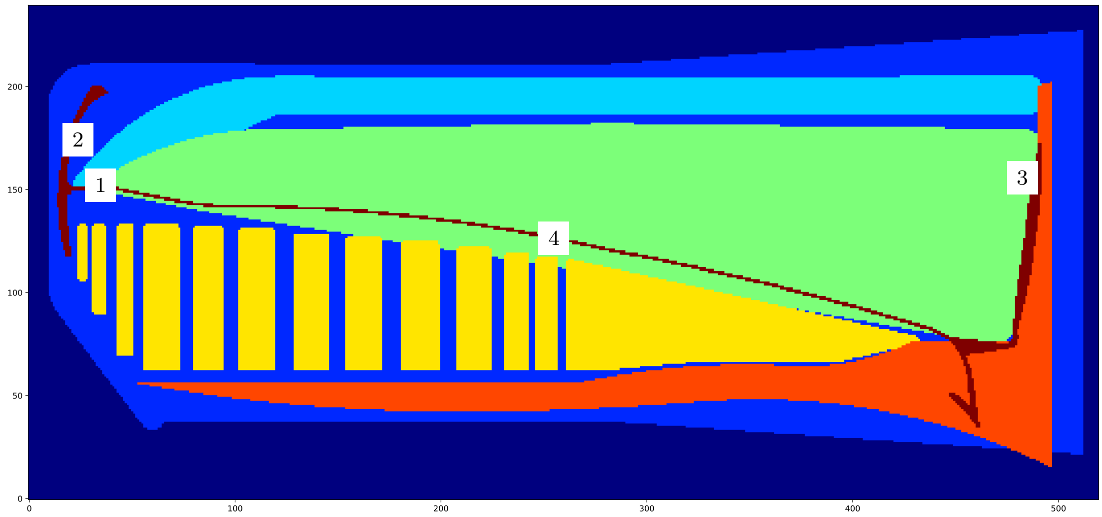


Figure 11.10: Material in the sagittal plane. Colors: deep blue=water, blue=blubber and skin, cyan=muscle, yellow=junk, green=spermaceti, orange=bone, dark red=air. Numbers: 1=museau de singe, 2=distal sac, 3=frontal sac, 4=right nasal passage.

these pulses. These intervals are known as the inter pulse interval (IPI) and have often been used to estimate the total body length of the sperm whale ([96], [95], and [98]).

While the proposed model still fails to reproduce individual pulse wave shapes, such as those found in recorded vocalisations, it does produce a signal with a valid IPI. By using the three different methods cited above to estimate the body size from the IPI, we obtain sizes of 14.97 m, 14.47 m and 14.12 m respectively, which match the length of the actual sperm whale that the model is based on (14.2 m). This result mainly depends on three parameters: the bulk modulus, the density, and the length of the *spermaceti*. Yet, it is still a comforting proof that this part of the model is working. In Chapters 12 and 13 we will study a parameter identification inverse problem that will allow an improved approximation of these parameters.

In Figure 11.12, we can see the evolution of the simulation, with the sound wave propagating from the *museau de singe* to the frontal sac, then being reflected by it, and going back to the *museau de singe* to be reflected by the distal sac.

11.4 Conclusion

The FDTD method, along with ABC and methodical treatment of boundaries between various media, was introduced in this chapter. This method was then applied to a 3D model of a sperm whale head, in order to simulate the wave propagation inside said head from the emission of the first pulse by the sperm whale, to the moment when the acoustic wave reaches the water.

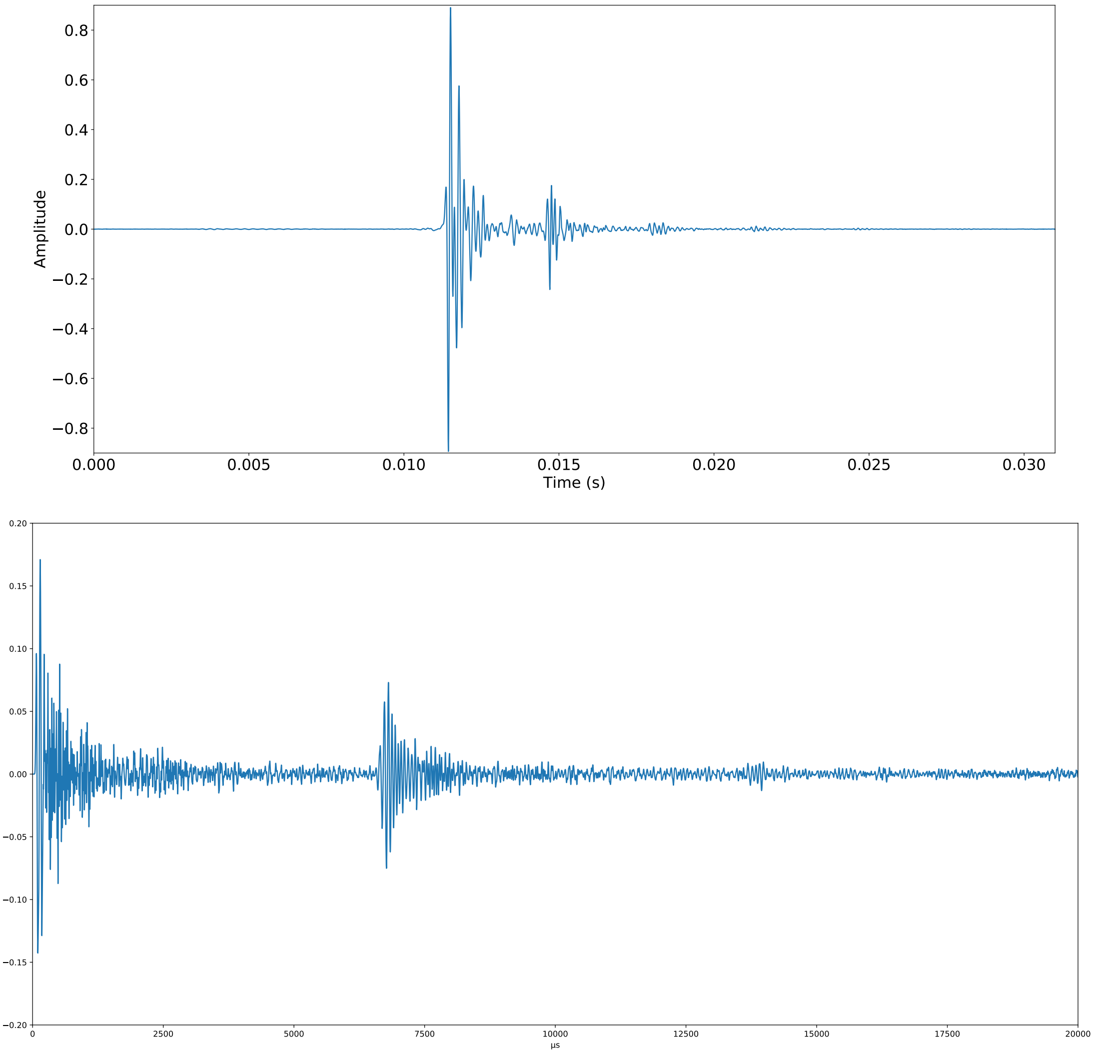


Figure 11.11: Top: Recording of sperm whale. Bottom: Simulated pressure at the excitation point.

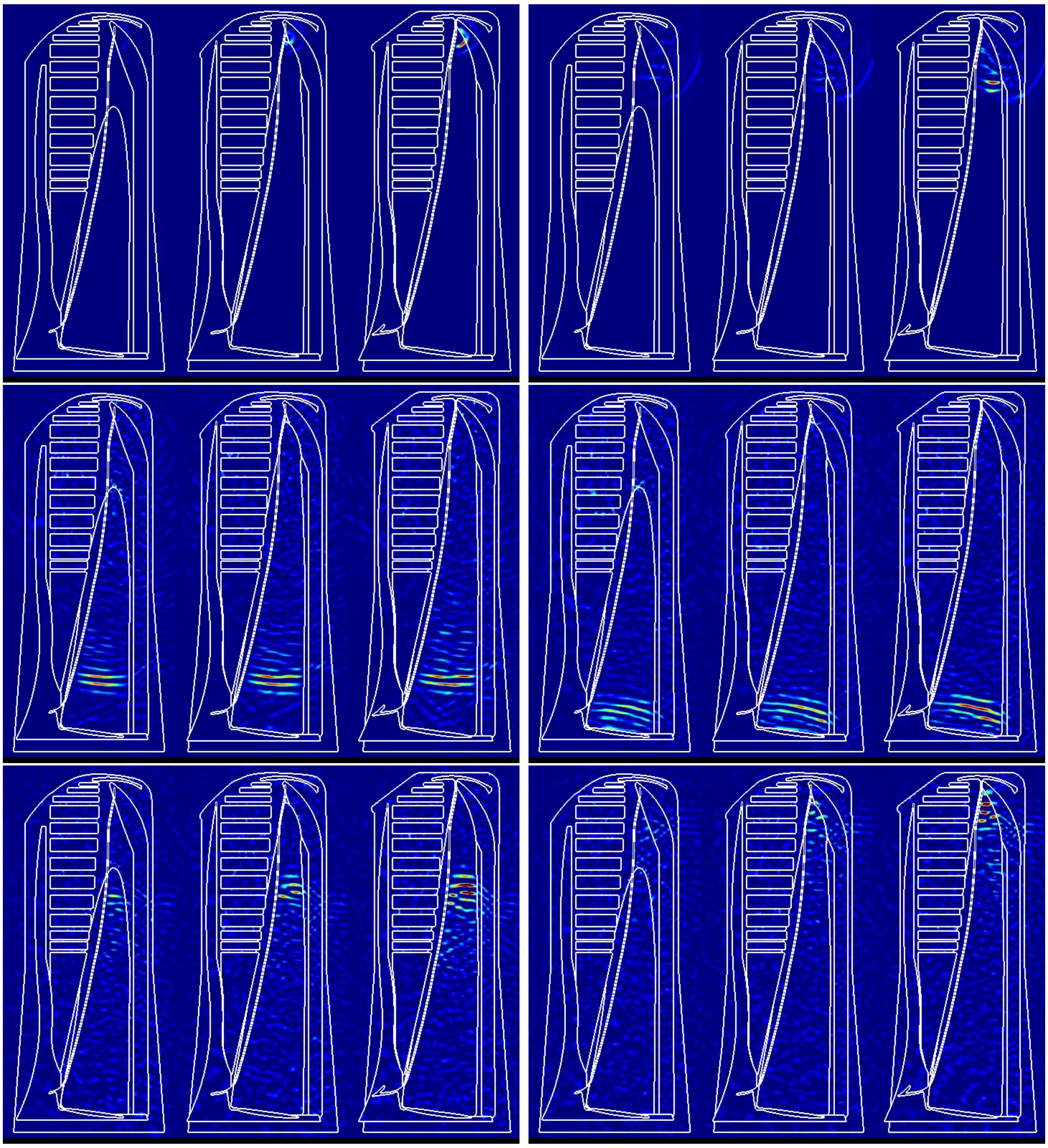


Figure 11.12: Multiple frames of the simulation, with the stress component (normalized) being plotted. Each picture is made of three slices of the 3D volume. The right one is the sagittal plane, the middle one is a plane 10 cm on the left of the sagittal plane, and the left one has an offset of 20 cm regarding the sagittal plane. Time steps shown are (top-left to bottom-right): 0.06μs, 0.6μs, 2.8μs, 3.5μs, 5.8μs and 6.6μs

Part IV

Toward Coupling of Simulations and Measurements

This final part of the thesis lays the theoretical and algorithmic groundwork for solving parameter identification inverse problems, that combine measurement data and propagation models. As was seen in all the previous parts, we can never have a completely accurate model of the acoustics. So, the objective of the inverse approach is to use the recorded data to *improve* the models, and then these improved models can be used in their turn to better analyze, identify and predict the sounds themselves. In particular, we could then study the effects of externalities on the behavior and safety of the cetaceans.

This part is not complete, due to lack of time, but is the subject of ongoing and future research efforts.

Chapter 12

The Inverse Problem

In this chapter we present the theoretical basis for coupling the measurements with the propagation models. The approach we use is based on an adjoint equation and is closely related to data assimilation [211]. If there is noise in the measurements, and we would like to obtain a complete characterisation of the uncertainty, then a Bayesian approach can be used. Other couplings are possible between machine learning and PDE models. In particular, we could use the model to generate realizations for the ML algorithm. Another possibility is to insert the differential equations into the neural net, forming a so-called "neural differential equation" method that couples the two. This will be the object of longer term research projects.

12.1 Parameter Estimation and Inverse Problems

Direct and inverse problems are depicted in Figure 12.1.

Our aim here is to couple data and models in order to improve the model. We seek to estimate the unknown, or badly known parameters of the model so that the model output is as close as possible to the observation data and the updated model can thus be used as a better predictive tool. This coupling can be done in several ways:

1. Adjoint methods, continuous or discrete.
2. Bayesian estimation.
3. Linear or nonlinear optimisation.
4. Machine learning approaches.

We will present in detail the first method, then briefly touch on the remaining three.

12.1.1 Adjoint Method

Let $\mathbf{u}$ be the state of a *dynamical system* whose behaviour depends on model parameters $\mathbf{m}$ and is described by a differential operator equation

$$\mathbf{L}(\mathbf{u}, \mathbf{m}) = \mathbf{f},$$

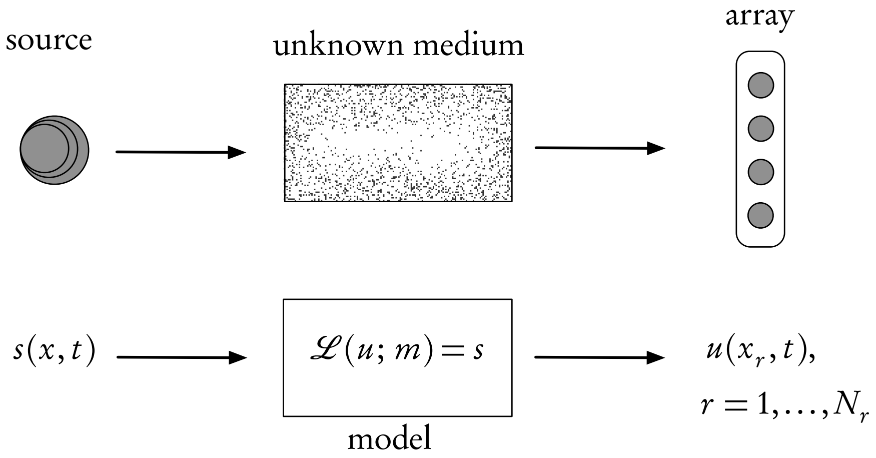


Figure 12.1: Ingredients of an inverse problem: the physical reality (top) and the direct mathematical model (bottom). The inverse problem uses the difference between the model-predicted observations, u, (calculated at the receiver array points x_r) and the real observations measured on the array, in order to find the unknown model parameters, m, or the source s (or both).

where $\mathbf{f}$ represents external forces. Define a *cost function* $J(\mathbf{m})$ as an energy functional or, more commonly, as a misfit functional that quantifies the error (L^2-distance[1]) between the observation and the model prediction $\mathbf{u}(\mathbf{x},t;\mathbf{m})$. For example,

$$J(\mathbf{m}) = \int_0^T \int_\Omega \Big(\mathbf{u}(\mathbf{x},t;\mathbf{m}) - \mathbf{u}^{\mathrm{obs}}(\mathbf{x},t)\Big)^2 \,\mathrm{d}\mathbf{x}\,\mathrm{d}t,$$

where $\mathbf{x} \in \Omega \subset \mathbb{R}^n, \quad n = 1,2,3,\ 0 \leq t \leq T$. Our objective is to choose the model parameters $\mathbf{m}$ as a function of the observed output $\mathbf{u}^{\mathrm{obs}}$, such that the cost function $J(\mathbf{m})$ is minimized.

The minimisation is most frequently performed by a gradient-based method, the simplest of which is steepest gradient, though usually some variant of a quasi-Newton approach is used [212]. If we can obtain an expression for the gradient, then the minimisation is greatly facilitated. This is the objective of the adjoint method that provides an *explicit* formula for the gradient of $J(\mathbf{m})$.

Suppose we are given a pde,

$$F(u;m) = 0, \tag{12.1}$$

where u is the state vector, m is the parameter vector (for simplicity, we drop the bold face notation here) and F represents a partial differential operator with right-hand side (source), boundary and initial conditions, such as the wave equatiuon, for example. Note that the components of m can appear as coefficients in the equation, the source,

---

[1] The L^2-space is a Hilbert space of (measurable) functions that are square-integrable (in Lebesgue sense).

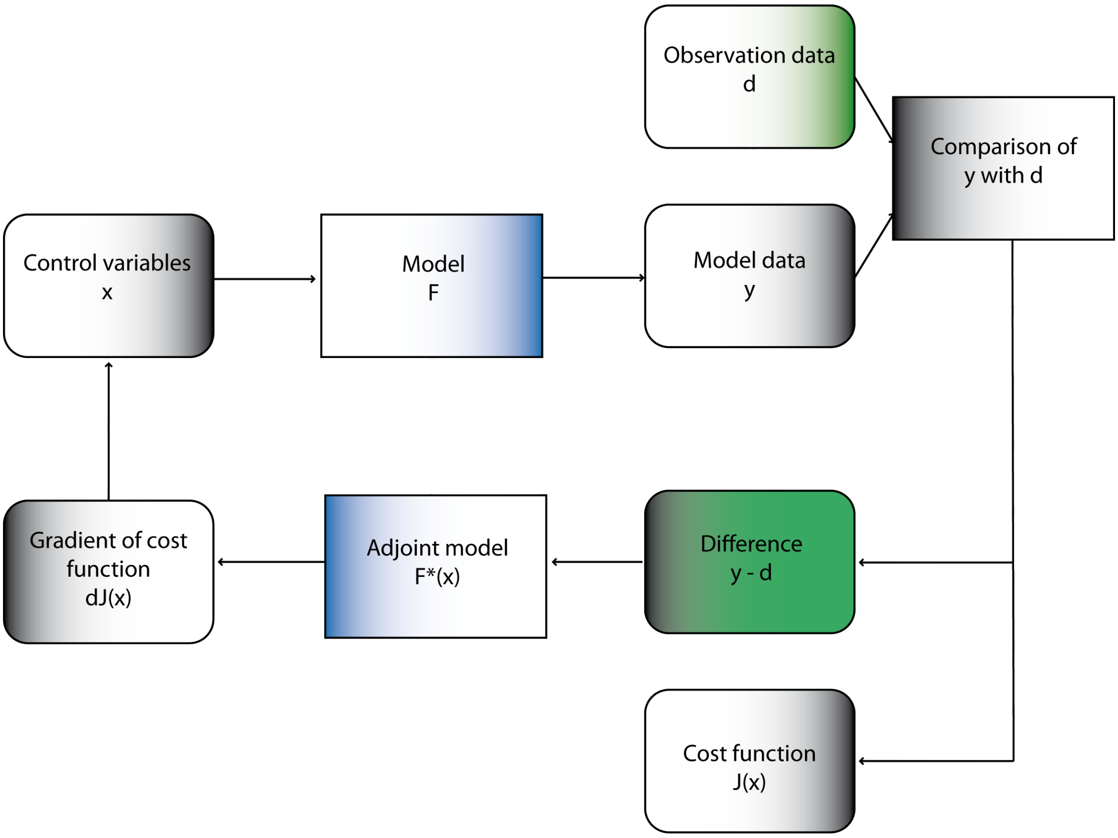


Figure 12.2: Data flow for an adjoint-based inversion.

or as components of the boundary/initial conditions. To solve the parameter estimation problem, we are given a cost function $J(u)$. The optimization problem is then,

$$\begin{cases} \text{minimize}_m & J\left(u(m), m\right) \\ \text{subject to} & F(u; m) = 0, \end{cases} \tag{12.2}$$

where J can depend on both u and on m explicitly in the presence of eventual regularization terms. The gradient of J with respect to m (also known as the *sensitivity*) is then given by

$$\nabla_m J = \frac{\partial J}{\partial u}\frac{\partial u}{\partial m} + \frac{\partial J}{\partial m}.$$

The derivatives of J with respect to u and m are readily computed from the expression for J, but the derivative of u with respect to m requires a potentially huge number of evaluations, corresponding to the product of the dimensions of u and m that can both be very large. The adjoint method is a way to avoid calculating all of these derivatives. We use the fact that $F(u; m) = 0$ everywhere, implies that the total derivative of F with respect to m is equal to zero everywhere too. Differentiating the PDE (12.1), we can thus write

$$\frac{\partial F}{\partial u}\frac{\partial u}{\partial m} + \nabla_m F = 0.$$

This can be solved to give,

$$\frac{\partial u}{\partial m} = -\left(\frac{\partial F}{\partial u}\right)^{-1} \nabla_m F$$

assuming that the inverse of F_u exists. Substituting in the expression for the gradient of J, we obtain

$$\begin{aligned} \nabla_m J &= -\frac{\partial J}{\partial u}\left(\frac{\partial F}{\partial u}\right)^{-1} \nabla_m F + \frac{\partial J}{\partial m}, \\ &= \lambda \nabla_m F + \frac{\partial J}{\partial m} \end{aligned} \tag{12.3}$$

where λ is the solution of the *adjoint equation*

$$\left(\frac{\partial F}{\partial u}\right)^T \lambda = -\frac{\partial J}{\partial u}. \tag{12.4}$$

In summary, we have a three-step procedure:

1. Solve (12.4) for the adjoint state, λ.
2. Compute the gradient of the cost function J from (12.3), where we suppose that $\partial J/\partial m$ and $\nabla_m F$ are known (or easily computed).
3. Solve the optimisation problem (12.2) to estimate the parameters m that minimise the mismatch between model and observations.

The complete data flow is shown in Figure 12.2.

12.1.2 Continuous and Discrete Adjoints

There are basically two approaches for forming and computing the adjoint state:

1. A continuous approach based on integration by parts of the (continuous) partial differential equation. This is sometimes called "adjoint then discretize (AtD)."
2. A discrete approach based on taking the adjoint of the discrete approximation of the partial differential equation. This is sometimes called "discretize then adjoint (DtA)."

A third approach, used for example in weather prediction codes, is automatic differentiation of the numerical code itself—see [213], [214], [215]. The use of this method is justified in cases where the system equations are very complex, and usually composed of coupled systems for which analytical and discrete adjoints are not feasible to derive. We will not use this approach here.

The continuous approach is described below, in Section 12.2, where it is applied to the wave equation. Here, we will briefly present the discrete approach which can be particularly useful in cases where the continuous adjoint is difficult to derive analytically.

Suppose now that we have a solution vector $\mathbf{x}$ of a *discretized* partial differential equation, or of any other set of n equations. Assume that $\mathbf{x}$ depends as usual on a parameter vector, $\mathbf{m}$, made up of p components - these are sometimes called control variables, design parameters or decision parameters. If we want to optimize these values for a given cost function, $J(\mathbf{x},\mathbf{m})$, we need to compute, as for the continuous case, the gradient $\mathrm{d}J/\mathrm{d}\mathbf{m}$. As we have seen above, this should be possible with an adjoint method at a cost that is independent of p and comparable to the cost of a single solution for $\mathbf{x}$. In the finite-dimensional case, this implies the inversion of a linear system, usually $\mathcal{O}(n^3)$ operations. This efficiency, especially for large values of p, is what makes the solution of the inverse problem tractable - if it were not for this, many problems would be simply impossible to solve within reasonable resource limits.

We will first consider systems of linear algebraic equations, then we can readily generalize to nonlinear systems of algebraic equations and to initial-value problems for linear systems of ordinary differential equations.

Linear systems

Let $\mathbf{x}$ be the solution of the $(n \times n)$ linear system

$$\mathbf{A}\mathbf{x} = \mathbf{b} \tag{12.5}$$

and suppose that $\mathbf{x}$ depends on the parameters $\mathbf{m}$ through $\mathbf{A}(\mathbf{m})$ and $\mathbf{b}(\mathbf{m})$. Define a cost function $J = J(\mathbf{x},\mathbf{m})$ that depends on $\mathbf{m}$ through $\mathbf{x}$. In order to evaluate the gradient of J with respect to $\mathbf{m}$ directly, we need to compute by the chain rule

$$\frac{dJ}{d\mathbf{m}} = \frac{\partial J}{\partial \mathbf{m}} + \frac{\partial J}{\partial \mathbf{x}}\frac{\partial \mathbf{x}}{\partial \mathbf{m}} = J_{\mathbf{m}} + J_{\mathbf{x}}\mathbf{x}_{\mathbf{m}}, \tag{12.6}$$

where $J_{\mathbf{m}}$ is a $(p \times 1)$ column vector, $J_{\mathbf{x}}$ is an $(1 \times n)$ row vector and $\mathbf{x}_{\mathbf{m}}$ is an $(n \times p)$ matrix. For a given function, J, the derivatives with respect to $\mathbf{x}$ and $\mathbf{m}$ are assumed to be easily computable. However, it is clearly much more difficult to differentiate $\mathbf{x}$ with respect to $\mathbf{m}$. Let us try and do this directly. We can differentiate, term-by-term, equation (12.5) with respect to the parameter m_i and solve for $\mathbf{x}_{m_i}$ from (applying the chain rule)

$$\mathbf{x}_{m_i} = \mathbf{A}^{-1}(\mathbf{b}_{m_i} - \mathbf{A}_{m_i}\mathbf{x}).$$

This must be done p times, and rapidly becomes unfeasible for large n and p. Recall that p can be of the order of 10^6 in practical data assimilation problems.

The adjoint method, which reduces this to a *single* solve, relies on the trick of adding zero in an astute way. We can do this, as was done above in the continuous case, by introducing a "Lagrange multiplier". Since the residual vector $\mathbf{r}(\mathbf{x}, \mathbf{m}) = \mathbf{A}\mathbf{x} - \mathbf{b}$ vanishes for the true solution $\mathbf{x}$, we can replace the function J by the augmented function

$$\hat{J} = J - \boldsymbol{\lambda}^T \mathbf{r} \tag{12.7}$$

where we are free to choose $\boldsymbol{\lambda}$ at our convenience and we will use this liberty in order to make the difficult-to-compute term in (12.6), $\mathbf{x}_{\mathbf{m}}$, disappear. So let us take the expression for the gradient (12.6) and evaluate it at $\mathbf{r} = 0$,

$$\begin{aligned} \left.\frac{dJ}{d\mathbf{m}}\right|_{\mathbf{r}=0} &= \left.\frac{d\hat{J}}{d\mathbf{m}}\right|_{\mathbf{r}=0}, \\ &= J_{\mathbf{m}} - \boldsymbol{\lambda}^T \mathbf{r}_{\mathbf{m}} + \left(J_{\mathbf{x}} - \boldsymbol{\lambda}^T \mathbf{r}_{\mathbf{x}}\right) \mathbf{x}_{\mathbf{m}}. \end{aligned} \tag{12.8}$$

Then in order to "kill" the troublesome $\mathbf{x}_{\mathbf{m}}$ term, we must require that $\left(J_{\mathbf{x}} - \boldsymbol{\lambda}^T \mathbf{r}_{\mathbf{x}}\right)$ vanishes, that implies

$$\mathbf{r}_x^T \boldsymbol{\lambda} = J_{\mathbf{x}}^T.$$

But $\mathbf{r}_{\mathbf{x}} = \mathbf{A}$ and hence $\boldsymbol{\lambda}$ must satisfy the *adjoint equation*

$$\mathbf{A}^T \boldsymbol{\lambda} = J_{\mathbf{x}}^T, \tag{12.9}$$

which is a single $(n \times n)$ linear system. Equation (12.9) is of identical complexity as the original system (12.5) since the adjoint matrix $\mathbf{A}^T$ has the same condition number, sparsity and preconditioner as $\mathbf{A}$ *i.e.* if we have a numerical scheme (and hence a computer code) for solving the direct system, we will use precisely the same one for the adjoint.

With $\boldsymbol{\lambda}$ known now, we can compute the gradient of J from (12.8) as follows,

$$\begin{aligned} \left.\frac{dJ}{d\mathbf{m}}\right|_{\mathbf{r}=0} &= J_{\mathbf{m}} - \boldsymbol{\lambda}^T \mathbf{r}_{\mathbf{m}} + 0 \\ &= J_{\mathbf{m}} - \boldsymbol{\lambda}^T (\mathbf{A}_{\mathbf{m}}\mathbf{x} - \mathbf{b}_{\mathbf{m}}). \end{aligned}$$

Once again, we assume that when $\mathbf{A}(\mathbf{m})$ and $\mathbf{b}(\mathbf{m})$ are explicitly known, this permits an easy calculation of the derivatives with respect to $\mathbf{m}$. If this is not the case, we must resort to automatic differentiation (AD) in order to compute these derivatives. The AD approach will be presented below, after we have discussed nonlinear and initial-value problems.

Initial-value problems

We consider an initial-value problem for a linear, time-independent, homogeneous system of ordinary differential equations, in matrix-vector form,

$$\dot{\boldsymbol{x}} = \mathbf{B}\boldsymbol{x}$$

with $\boldsymbol{x}(0) = \boldsymbol{b}$. We know that the solution is given by

$$\boldsymbol{x}(t) = \mathrm{e}^{\mathbf{B}t}\boldsymbol{b},$$

but this can be rewritten as a linear system, $\mathbf{A}\boldsymbol{x} = \boldsymbol{b}$ where $\mathbf{A} = \mathrm{e}^{-\mathbf{B}t}$. Now we simply can use our results from above. Suppose we want to minimize $J(\mathbf{x}, \mathbf{m})$ based on the solution $\boldsymbol{x}$ at time t. As before, we can compute the adjoint vector $\boldsymbol{\lambda}$ using (12.9),

$$\mathrm{e}^{-\mathbf{B}^T t}\boldsymbol{\lambda} = J_{\mathbf{x}}^T,$$

but this is equivalent to the adjoint ODE,

$$\dot{\boldsymbol{\lambda}} = \mathbf{B}^T\boldsymbol{\lambda},$$

with $\boldsymbol{\lambda}(0) = J_{\mathbf{x}}^T$. This is exactly what we would expect: solving for the adjoint state vector, $\boldsymbol{\lambda}$, is a problem of the same complexity and type as that of finding the state vector $\boldsymbol{x}$. Clearly we are not obliged to use matrix exponentials for the solution, but can choose among: Runge-Kutta formulas, forward Euler, Crank-Nicolson, etc. [216]. What about the important issue of stability? The eigenvalues of $\mathbf{B}$ and $\mathbf{B}^T$ are complex conjugates and thus the stability of one (spectral radius less than one) implies the stability of the other. Finally, using

$$\left.\frac{dJ}{d\mathbf{m}}\right|_{\mathbf{r}=0} = J_{\mathbf{m}} - \boldsymbol{\lambda}^T\mathbf{r}_{\mathbf{m}}, \tag{12.10}$$

we obtain the gradient of the cost function in the time-dependent case,

$$\begin{aligned}\frac{dJ}{d\mathbf{m}} &= J_{\mathbf{m}} - \boldsymbol{\lambda}^T(\mathbf{A}_{\mathbf{m}}\mathbf{x} - \mathbf{b}_{\mathbf{m}}) \\ &= J_{\mathbf{m}} + \int_0^t \boldsymbol{\lambda}^T(t-t')\mathbf{B}_{\mathbf{m}}\mathbf{x}(t')\,dt' + \boldsymbol{\lambda}^T\mathbf{b}_{\mathbf{m}},\end{aligned}$$

where we have differentiated the expression for $\mathbf{A}$. We observe that this computation of the gradient via the adjoint requires that we save in memory $\mathbf{x}(t')$ for all times $0 \le t' \le t$ in order to be able to compute the gradient. This is a well-known issue in adjoint approaches for time-dependent problems and can be dealt with in three ways (that are problem-, or more precisely, dimension-dependent):

1. Store everything in memory, if feasible.
2. If not, use some kind of checkpointing [217], which means that we divide the time interval into a number of subintervals, and store consecutively sub-interval by sub-interval.
3. Re-solve “simultaneously” forward and adjoint, and at the same time compute the integral. i.e. at each time step of the adjoint solution process, recompute the direct solution up to this time.

12.1.3 Linear and Nonlinear Optimization

If we do not choose to use an adjoint approach, then the optimisation problem where we seek the parameters to minimise the model-observations mismatch, can be solved directly by classical optimisation techniques. These problems arise from the discretisation of the underlying PDE model.

A linear inverse problem is one that can be written as a linear system of equations,

$$G\mathbf{m} = \mathbf{d}, \tag{12.11}$$

where $\mathbf{m} \in \mathbb{R}^p$ and $\mathbf{d} \in \mathbb{R}^n$ are vectors, and $G \in \mathbb{R}^{n\times p}$ is an $(n \times p)$ dimensional matrix. To invert this equation for $\mathbf{m}$, and solve the inverse problem, we need to "simply" invert G. But this is not a trivial linear algebra problem because of the dimensions and rank of G that make the problem (12.11), in general, undetermined or inconsistent. We need to find the best solution possible in this case, one that will reduce to the classical solution,

$$\mathbf{m} = G^{-1}\mathbf{d},$$

when G is a square, invertible (or non singular) matrix.

The most frequent situation is the least squares solution, which is optimal in the sense that it minimizes the squares of the errors between the model predictions and the data. The errors, or residuals, are defined as

$$\mathbf{r} = G\mathbf{m} - \mathbf{d},$$

and their sum of squares can be expressed as a discrete L_2-norm of $\mathbf{r}$.

In the most commonly encountered case, there will be more equations than unknowns, more rows than columns, in the matrix A. In this overdetermined case, there is a very important theorem.

Theorem 12.1.1 *The optimal least-squares solution of any system, $Ax = b$, is the vector $\bar{x}$ given by*

$$\bar{x} = A^{\dagger}b,$$

where the pseudoinverse, $A^{\dagger}$, is defined as

$$A^{\dagger} = \left(A^T A\right)^{-1} A^T.$$

A more general result for the existence of a unique least squares solution is provided by the following theorem.

Theorem 12.1.2 *For $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$, let $\epsilon = \epsilon(x) = Ax - b$. The general least-squares problem is to find the vector x that minimizes the residual sum of squares,*

$$\sum_{i=1}^{m} \epsilon_i^2 = \epsilon^T \epsilon = (Ax - b)^T (Ax - b).$$

Any vector that provides a minimal value is called a least-squares solution. The set of all least-squares solutions is precisely the set of solutions of the normal equations, $A^TAx = A^Tb$.

There exists a unique least-squares solution, given by $x = \left(A^TA\right)^{-1}A^Tb$, *if and only if* $\operatorname{rank}(A) = n$. *If the linear system* $Ax = b$ *is consistant, then the solution of* $Ax = b$ *is the same as the least-squares solution.*

Note that when A is invertible (implying that its transpose is invertible as well), we indeed obtain

$$A^\dagger = \left(A^TA\right)^{-1}A^T = A^{-1}\left(A^T\right)^{-1}A^T = A^{-1}.$$

However, if A is rank deficient, then the inverse of A^TA does not exist and we have to resort to regularization techniques, or to pseudoinverses based on singular value decompositions.

Nonlinear Case We suppose that the discrete inverse problem is now nonlinear,

$$G(\mathbf{m}) = \mathbf{d},$$

where G is some nonlinear operator, or function, describing the relation between $\mathbf{m}$ and $\mathbf{d}$ and we seek $\mathbf{m}$ from known measurements $\mathbf{d}$. There is no general theory for the solution of the nonlinear inverse problem, but iterative methods, based on underlying linear strategies, are usually quite effective for their solution.

The methods commonly employed are precisely methods that are used in nonlinear optimization, namely Newton's method and certain of its variants that are specific for least-squares problems–in this case, Gauss-Newton and Levenberg-Marquardt methods. These are also the methods employed for nonlinear regression.

Let the objective function,

$$\phi(x) = \frac{1}{2}\sum_{j=1}^{p} r_j^2(x),$$

where each $r_j(x) = y_j - f(t_j, x)$ is the residual function, with observations (t_i, y_i), and maps $\mathbb{R}^n$ into $\mathbb{R}$. This type of objective function appears in most parameter identification problems where the discrepancy (or residual) measures the mismatch between modelled and measured system behavior. The minimization clearly finds the best match between model and observations.

The special structure of ϕ enables us to formulate adapted algorithms for its minimization. Writing the residual function as a vector function,

$$\phi(x) = \frac{1}{2}\|r(x)\|_2^2 = \frac{1}{2}\mathbf{r}^T\mathbf{r},$$

we can express the derivatives (or variation) of $\phi(x)$ with respect to x as the Jacobian matrix of first partial derivatives

$$\begin{aligned} J(x) &= \left[\frac{\partial r_j}{\partial x_i} \right]_{j=1,\ldots,m,\, i=1,\ldots,n} \\ &= \begin{bmatrix} \nabla r_1(x)^T \\ \vdots \\ \nabla r_m(x)^T \end{bmatrix} \\ &= \begin{bmatrix} \frac{\partial r_1}{\partial x_1} & \cdots & \frac{\partial r_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial r_m}{\partial x_1} & \cdots & \frac{\partial r_m}{\partial x_n} \end{bmatrix}. \end{aligned}$$

The gradient and Hessian of the objective function, ϕ, can then be written in terms of J and r as

$$\begin{aligned} \nabla \phi(x) &= J^T(x) r(x), \\ \nabla^2 \phi(x) &= J^T(x) J(x) + \sum_{j=1}^{m} r_j(x) \nabla^2 r_j(x). \end{aligned}$$

In many applications, the first derivatives can be obtained relatively easily, by various approximation methods. It is the second derivatives, present in the Hessian, that are delicate to compute accurately. But, if the second term in the expression for the Hessian is small—which it will be when r is small, or when the residuals are almost affine—we can compute everything with only the knowledge of J. In fact, the term J^TJ is often larger than the second term, and then this approximation can be exploited to formulate nice algorithms. These are based on Newton or quasi-Newton approaches.

12.1.4 Machine Learning and Neural Networks

There is much potential, and recent interest, in using machine learning methods to solve inverse problems. This entails some kind of coupling between the model, the observations and the statistical learning approaches. There are several promising avenues to explore here:

1. Use the model to generate additional members of the training set for a neural network.
2. From the observations (assuming that we have a large volume) deduce the parameters of the model, or the model itself, by machine learning techniques.
3. Incorporate the differential equations directly into the neural network using a neural differential equation approach.

We will not consider here the first point. Regarding point number 2, a number of recent works have been inspired by the approach of [218] who proposed a new feed-forward deep network, called PDE-Net, that fulfills two objectives simultaneously: to accurately predict dynamics of complex systems and to uncover the underlying hidden PDE models. A recent contribution in [219] combines symbolic calculus and a neural network generator.

In the paper, [220], the problem of learning complex spatio-temporal dynamics with neural networks is also treated, but using an adjoint. The dynamical system is expressed as

$$\frac{\mathrm{d}x}{\mathrm{d}t} = f_\theta(x),$$

with (partial) observations

$$y = h(x).$$

The adjoint approach is used to express the parameter estimation problem to estimate f_θ, as was see above. Then a residual net (ResNet) architecture is used to learn the evolution operator, f_θ. The training algorithm, though purely data-driven, is then very similar to the one shown above in Figure 12.2.

The basic method, used in these cases, is a neural differential equation approach - see `https://github.com/SciML/DiffEqFlux.jl`. With this package one can explore various ways to integrate the two methodologies (NN and ODE):

- Neural networks can be defined where the “activations” are nonlinear functions described by differential equations.
- Neural networks can be defined where some layers are ODE solves.
- ODEs can be defined where some terms are neural networks.
- Cost functions on ODEs can define neural networks.

The basic principle derives from the fact that a recurrent neural network,

$$x_{k+1} = x_k + \mathrm{DNN}(x_k, k; p),$$

is equivalent to a forward Euler method, with time-step equal to one, applied to the neural ODE,

$$x' = \mathrm{DNN}(x, t; p),$$

where p represents the parameters and DNN is a deep neural network architecture. This can easily be extended to neural PDEs. We just use the method of lines to convert the PDE into a system of ODES and then apply the neural ODE approach.

This can be generalized to more complicated finite difference stencils (eg. for higher order derivatives), and to other ODE solvers, such a Runge-Kutta methods. For this, a new package, PDE-NetgGen, that combines Keras and sympy, a symbolic mathematics package for the stencil evaluations, has been developed in [219].

In a series of very recent papers, Ying’s group at Stanford University has developed a similar approach that is based on a wavelet basis for approximating integral operators [221]. This approach has been successfully applied to a number of parameter estimation and inverse problems, and in particular to scattering problems for the wave equation.

12.2 Application to the Wave Equation: Adjoint Method

12.2.1 Wave propagation

Traditionally, seismo-acoustic wave propagation for underwater acoustics is modelled in a bounded computational domain Ω in $R^3 \times (0, T)$ that represents a layered medium, underlying a seawater layer with varying bathymetry. The sea surface is traction- (pressure-) free and the lateral boundaries (sides and bottom) are suitably absorbing. Continuity conditions are imposed on all interfaces between layers. There is a source and there are points in the domain where the signals are measured (receivers). Mathematically, this situation can be described by an initial-boundary-value-problem for the seismo-acoustic, linear wave equation,

$$\rho \boldsymbol{u}_{tt} - (\lambda + \mu)\nabla(\nabla \cdot \boldsymbol{u}) - \mu \nabla^2 \boldsymbol{u} \quad = \quad \boldsymbol{f}, \tag{12.12}$$

where $\boldsymbol{u} = (u_1, u_2, u_3)^T$ is the displacement in the x-, y- and z-direction, λ and μ are the Lamé coefficients, ρ is the medium density and $\boldsymbol{f}$ is an initial impulse that represents the acoustic source. The relations between these coefficients and the wave speeds are

$$c_p^2 = \frac{\lambda + 2\mu}{\rho} \qquad \text{and} \qquad c_s^2 = \frac{\mu}{\rho},$$

where c_p is the pressure (or primary) wave speed and c_s is the shear (or secondary) wave speed. Thus an acoustic layer is obtained in the model by simply setting $\mu = 0$ locally.

Note that the elastodynamic system (12.12) can also be expressed in displacement-stress form, component-by-component, as

$$\begin{aligned} \rho \frac{\partial^2 u_i}{\partial t^2} &= \frac{\partial \sigma_{ij}}{\partial x_j} + f_i\,, \\ \sigma_{ij} &= \lambda \delta_{ij} \epsilon_{kk} + 2\mu \epsilon_{ij}\,, \\ \epsilon_{ij} &= \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right)\,, \end{aligned} \tag{12.13}$$

or

$$\begin{aligned} \rho u_{i,tt} &= \sigma_{ij,j} + f_i \,\text{in}\, \boxtimes \times (0, \mathrm{T}), \\ \sigma_{ij} &= \lambda \delta_{ij} \epsilon_{kk} + 2\mu \epsilon_{ij} = c_{ijkl}\epsilon_{kl}\,, \\ \epsilon_{ij} &= \frac{1}{2}\left(u_{i,j} + u_{j,i}\right)\,, \end{aligned} \tag{12.14}$$

where $i, j, k, l = 1, \ldots, 3$, we have used Einstein's summation convention for the indices, and δ is the Kronecker delta function. Finally, this can be written as a single vector equation,

$$\rho \ddot{\mathbf{u}} - \nabla \cdot \boldsymbol{\sigma} = f. \tag{12.15}$$

The advantage of the system (12.12) or (12.13) is that it intrinsically models *all* the different types of waves that can arise in layered media - compressional and shear waves in the bulk, and Love, Stoneley and Raleigh waves along the interfaces. These equations must be completed with physically relevant *boundary and initial conditions.* On the surface of the water we usually specify a zero pressure condition. Between layers, continuity conditions on the normal components of u and the stresses must be satisfied. At the bottom-most level we give a suitable absorbing condition. On the lateral boundaries, suitable absorbing/radiating conditions need to be specified. The initial condition is usually a Ricker wavelet with the desired frequency content, located at the source position in the water layer - other initial conditions are possible. The hydrophones (or other measurement devices) are “simulated” by simply recording the solution at given points.

We recognize that this is directly applicable to our wave propagation problem of Chapter 11 inside and surrounding the sperm whale’s head.

12.2.2 Adjoint of the Wave Equation

As a preliminary step, we will derive the adjoint of the scalar wave equation. Then in Section 12.2.3 below, we will obtain an adapted version for the sperm whale head, that in particular will enable us to reuse the simulation code of Chapter 11 for solving the inverse problem.

The initial boundary value problem for the (scalar) wave equation is,

$$\begin{aligned} \frac{\partial^2 u}{\partial t^2} - \nabla\cdot(\nu\nabla u) = 0, \quad & x\in(0,L), && t>0,\\ u(x,0)=u_0(x), \quad & \tfrac{\partial u}{\partial t}(x,0)=0, && t=0,\\ u(0,t)=0, \quad & u(L,t)=\eta(t), && t>0. \end{aligned}$$

A variety of different inverse problems can be formulated for this system:

- *internal* control: $\nu(x)$ - this is the parameter identification problem, also known as tomography;
- *initial* control: $\xi(x)=u_0(x)$ - this is a source detection IP or DA problem;
- *boundary* control: $\eta(t)=u(L,t)$ - this is the “classical” boundary control problem, also a parameter identification IP.

As above, we can define the cost function,

$$J[\nu,\xi,\eta]=\frac{1}{LT}\int_0^T\int_0^L\left(u-u^{\mathrm{obs}}\right)^2\,\mathrm{d}x\,\mathrm{d}t,$$

which is now a space-time multiple integral, and its related Lagrangian,

$$J^*=\frac{1}{LT}\int_0^T\int_0^L\left(u-u^{\mathrm{obs}}\right)^2\,\mathrm{d}x\,\mathrm{d}t+\frac{1}{LT}\int_0^T\int_0^L p\left[u_{tt}-(\nu u_x)_x\right]\,\mathrm{d}x\,\mathrm{d}t.$$

Now take the variation of J^*,

$$\begin{aligned}\delta J^* &= \frac{1}{LT}\int_0^T\int_0^L 2(u-u^{\mathrm{obs}})\delta u\,\mathrm{d}x\,\mathrm{d}t + \frac{1}{LT}\int_0^T\int_0^L \delta p\,\overbrace{\left[u_{tt}-(\nu u_x)_x\right]}^{=0}\,\mathrm{d}x\,\mathrm{d}t \\ &+\frac{1}{LT}\int_0^T\int_0^L p\left[\delta u_{tt}-(\delta\nu\, u_x+\nu\delta u_x)_x\right]\,\mathrm{d}x\,\mathrm{d}t\end{aligned}$$

and perform integration by parts to obtain,

$$\delta J^* = \frac{1}{LT}\int_0^T\int_0^L \delta\nu\, u_x p_x\,\mathrm{d}x\,\mathrm{d}t + \frac{1}{LT}\int_0^L p_t\ \left.\delta u\right|_{t=0}\mathrm{d}x\ + \frac{1}{LT}\int_0^T p\ \left.\delta\eta\right|_{x=L}\mathrm{d}t, \tag{12.16}$$

where we have defined the adjoint equation as

$$\begin{aligned}&\frac{\partial^2 p}{\partial t^2} - \nabla\cdot(\nu\nabla p) = 2(u-u^{\mathrm{obs}}), \quad x\in(0,L), \quad t>0,\\ &p(0,t)=0, \quad p(L,t)=0,\\ &p(x,T)=0, \quad p_t(x,T)=0.\end{aligned}$$

As before, this equation is of the same type as the original wave equation, but must be solved backwards in time. Now we use the definition of the variation of u with respect to the parameters $\mathbf{m}$ in the direction $\delta\mathbf{m}$ (the Gâteaux differential),

$$\delta\mathbf{u} = \nabla_m\mathbf{u}\,\delta\mathbf{m}$$

to write the variation of J as

$$\begin{aligned}\delta J &= \nabla_m J\,\delta\mathbf{m}\\ &= \nabla_u J\,\delta\mathbf{u}\\ &= \left\langle\nabla_u J_1\,\delta\mathbf{u}\right\rangle, \end{aligned}\tag{12.17}$$

where in the second line we have used the chain rule together with the above definition of $\delta\mathbf{u}$, and in the third line $\langle\cdot\rangle$ denotes the space-time integral. Here we have passed the "derivative" under the integral sign and J_1 is the integrand. Finally, based on the result (12.17), we can pick off in (12.16) each of the three desired terms of the gradient,

$$\begin{aligned}\nabla_{\nu(x)}J^* &= \frac{1}{T}\int_0^T u_x p_x\,\mathrm{d}t,\\ \nabla_{\left.u\right|_{t=0}}J^* &= \left.p_t\right|_{t=0},\\ \nabla_{\left.\eta\right|_{x=L}}J^* &= \left.p\right|_{x=L}.\end{aligned}$$

At the expense of a single (backward) solution of the adjoint equation, we obtain explicit expressions for the gradient of the cost function with respect to each of the three control variables. This is quite remarkable and completely avoids "brute force" or exhaustive minimization, though, as mentioned earlier, we only have the guarantee to find a local minimum. However, if we have a good starting guess that is usually obtained from historical or other "physical" knowledge of the system, we are sure to arrive at a good (or at least, better) minimum.

An alternative approach for seismo-acoustic inversion, based on Fréchet kernels, is implemented in the SPECFEM framework [222]. Details can be found in Appendix B.

12.2.3 Tangent linear model

In our context, where we seek to model the sound generation and propagation inside the sperm whale's head, we use (see Section 11.2.1):

- A system of two first-order partial differential equations.
- Piecewise constant material properties in each anatomical subdomain of the head, following the data of Table 11.1.

To derive the adjoint in this specific case, we need to use a different approach based on a tangent linear model. The tangent linear model (TLM) is used to obtain the the gradient of the cost function J by perturbing it with a small perturbation. Unlike the Lagrangian method, it can be used with constant-valued parameters.

Let us consider the following matrix definitions

$$D_1 = \begin{pmatrix} \frac{\partial}{\partial x_1} & 0 & 0 & 0 & \frac{\partial}{\partial x_3} & \frac{\partial}{\partial x_2} \\ 0 & \frac{\partial}{\partial x_2} & 0 & \frac{\partial}{\partial x_3} & 0 & \frac{\partial}{\partial x_1} \\ 0 & 0 & \frac{\partial}{\partial x_3} & \frac{\partial}{\partial x_2} & \frac{\partial}{\partial x_1} & 0 \end{pmatrix}, \quad D_2 = \frac{1}{2}\begin{pmatrix} 2\frac{\partial}{\partial x_1} & 0 & 0 \\ 0 & 2\frac{\partial}{\partial x_2} & 0 \\ 0 & 0 & 2\frac{\partial}{\partial x_1} \\ 0 & \frac{\partial}{\partial x_3} & \frac{\partial}{\partial x_2} \\ \frac{\partial}{\partial x_3} & 0 & \frac{\partial}{\partial x_1} \\ \frac{\partial}{\partial x_2} & \frac{\partial}{\partial x_1} & 0 \end{pmatrix},$$

$$N_1 = \begin{pmatrix} n_1 & 0 & 0 & 0 & n_3 & n_2 \\ 0 & n_2 & 0 & n_3 & 0 & n_1 \\ 0 & 0 & n_3 & n_2 & n_1 & 0 \end{pmatrix}, \qquad N_2 = \frac{1}{2}\begin{pmatrix} 2n_1 & 0 & 0 \\ 0 & 2n_2 & 0 \\ 0 & 0 & 2n_3 \\ 0 & n_3 & n_2 \\ n_3 & 0 & n_1 \\ n_2 & n_1 & 0 \end{pmatrix},$$

$$D_3 = \begin{pmatrix} 0_{6,6} & D_2 \\ D_1 & 0_{3,3} \end{pmatrix}, \qquad N_3 = \begin{pmatrix} 0_{6,6} & N_2 \\ N_1 & 0_{3,3} \end{pmatrix},$$

$$P_1 = \begin{pmatrix} C & 0_{6,3} \\ 0_{3,6} & I_3 \end{pmatrix}, \qquad P_2 = \begin{pmatrix} I_6 & 0_{6,3} \\ 0_{3,6} & \rho I_3 \end{pmatrix},$$

$$R = \begin{pmatrix} I_6 \\ 0_{3,6} \end{pmatrix}$$

where $\boldsymbol{n} = \begin{pmatrix} n_1 \\ n_2 \\ n_3 \end{pmatrix}$ is the local normal of the surface considered, I_n is the identity matrix and $0_{m,n}$ is the null matrix. With these matrices, the solid wave equations (9.17), (9.18), (9.19) can be rewritten using Voigt notation (see (9.20)), to obtain a matrix relation between the stress $\boldsymbol{\sigma}$, the strain $\boldsymbol{\varepsilon}$, and the particle speed $\boldsymbol{v}$, which is the temporal partial derivative of the particle displacement $\boldsymbol{\xi}$,

$$\rho\frac{\partial \boldsymbol{v}}{\partial t} = D_1\boldsymbol{\sigma}, \tag{12.18}$$

$$\frac{\partial \boldsymbol{\varepsilon}}{\partial t} = D_2\boldsymbol{v}, \tag{12.19}$$

$$\boldsymbol{\sigma} = C\boldsymbol{\varepsilon} + \boldsymbol{f}(\boldsymbol{x},t) \tag{12.20}$$

where $\boldsymbol{f}(\boldsymbol{x},t)$ is an added source term.

If we define the augmented unknown vector as $\boldsymbol{u} = \begin{pmatrix}\boldsymbol{\sigma}\\ \boldsymbol{v}\end{pmatrix}$, then we can rewrite the wave equation as

$$P_2\frac{\partial \boldsymbol{u}}{\partial t} = P_1D_3u + R\boldsymbol{f}(\boldsymbol{x},t). \tag{12.21}$$

On the boundary Γ of the simulation domain Ω, the vector $\boldsymbol{u}$ is

$$\boldsymbol{u}_\Gamma = 0 \tag{12.22}$$

since both $\boldsymbol{v}$ and $\boldsymbol{\sigma}$ are at rest outside the model, and on its boundary. Since the model starts at rest, we also have

$$\boldsymbol{u}(t=0) = 0 \tag{12.23}$$

Using this unknown, we now define the cost function to be minimized,

$$J(C,\rho) = \frac{1}{2}\int_0^T\int_\Omega \|\boldsymbol{u} - \boldsymbol{u}^{obs}\|_2^2 dV dt \tag{12.24}$$

and we can compute the perturbation difference,

$$J(C+\alpha\delta C, \rho+\alpha\delta\rho) - J(C,\rho) = \frac{1}{2}\int_0^T\int_\Omega \|\tilde{\boldsymbol{u}} - \boldsymbol{u}^{obs}\|_2^2 - \|\boldsymbol{u} - \boldsymbol{u}^{obs}\|_2^2 dV dt \tag{12.25}$$

$$= \frac{1}{2}\int_0^T\int_\Omega < \tilde{\boldsymbol{u}} + \boldsymbol{u} - 2\boldsymbol{u}^{obs} | \tilde{\boldsymbol{u}} - \boldsymbol{u} > dV dt. \tag{12.26}$$

Now, the variation of the cost function is defined by

$$\delta J_{C,\rho}(\delta C, \delta\rho) = \int_0^T\int_\Omega < \boldsymbol{u} - \boldsymbol{u}^{obs} | \hat{\boldsymbol{u}} > dV dt, \tag{12.27}$$

and $\tilde{\boldsymbol{u}}$ is the solution of the perturbed system,

$$\begin{cases} (P_2 + \alpha\delta P_2)\frac{\partial \tilde{\boldsymbol{u}}}{\partial t} - (P_1 + \alpha\delta P_1)D_3\tilde{\boldsymbol{u}} = R\boldsymbol{f}(\boldsymbol{x},t), \\ \tilde{\boldsymbol{u}}_\Gamma = 0, \\ \tilde{\boldsymbol{u}}(t=0) = 0. \end{cases} \tag{12.28}$$

To obtain the TLM, we must now subtract and divide by α, then take the limit as $\alpha \to 0$,

$$P_2 \frac{\partial \frac{\tilde{\boldsymbol{u}}-\boldsymbol{u}}{\alpha}}{\partial t} + \delta P_2 \frac{\partial \tilde{\boldsymbol{u}}}{\partial t} - P D_3 \frac{\tilde{\boldsymbol{u}}-\boldsymbol{u}}{\alpha} - \delta P_1 D_3 \tilde{\boldsymbol{u}} = 0, \tag{12.29}$$

$$P_2 \frac{\partial \hat{\boldsymbol{u}}}{\partial t} + \delta P_2 \frac{\partial \boldsymbol{u}}{\partial t} - P_1 D_3 \hat{\boldsymbol{u}} - \delta P_1 D_3 \boldsymbol{u} = 0. \tag{12.30}$$

along with the boundary condition

$$\begin{cases} \hat{\boldsymbol{u}}_\Gamma = 0, \\ \hat{\boldsymbol{u}}(t=0) = 0. \end{cases} \tag{12.31}$$

The scalar product of (12.30) with a vector $\boldsymbol{\eta}$ is taken and then integrated over the simulation time and the domain Ω which results in

$$\int_0^T \int_\Omega < P_2 \frac{\partial \hat{\boldsymbol{u}}}{\partial t} | \boldsymbol{\eta} > dV dt - \int_0^T \int_\Omega < P_1 D_3 \hat{\boldsymbol{u}} | \boldsymbol{\eta} > dV dt = \int_0^T \int_\Omega < \delta P_2 \frac{\partial \boldsymbol{u}}{\partial t} - \delta P_1 D_3 \boldsymbol{u} | \boldsymbol{\eta} > dV dt. \tag{12.32}$$

The goal is now to transfer the derivatives of $\tilde{\boldsymbol{u}}$ onto the newly introduced vector $\boldsymbol{\eta}$. In order to do so, integration by parts can be used which on the first term is

$$\int_0^T \int_\Omega < P_2 \frac{\partial \hat{\boldsymbol{u}}}{\partial t} | \boldsymbol{\eta} > dV dt = \left[\int_\Omega < \hat{\boldsymbol{u}} | P_2^* \boldsymbol{\eta} > dV \right]_0^T - \int_0^T \int_\Omega < \hat{\boldsymbol{u}} | P_2^* \frac{\partial \boldsymbol{\eta}}{\partial t} > dV dt \tag{12.33}$$

Here the first term on the right side can be simplified since $\hat{\boldsymbol{u}}(t=0) = 0$. Integration by parts on the second term of (12.32), using Gauss' theorem, is

$$\int_0^T \int_\Omega < P_1 D_3 \hat{\boldsymbol{u}} | \boldsymbol{\eta} > dV dt = \int_0^T \int_\Gamma < \hat{\boldsymbol{u}} | N_3^* P_1^* \boldsymbol{\eta} > dS dt - \int_0^T \int_\Omega < D_3^* P_1^* \boldsymbol{\eta} | \hat{\boldsymbol{u}} > dV dt, \tag{12.34}$$

which can be simplified due to the Dirichlet boundary condition on $\hat{\boldsymbol{u}}$, which when integrated back into (12.32) leads to the equation

$$\int_\Omega < \hat{\boldsymbol{u}} | P_2^* \boldsymbol{\eta} > (t=T) dV - \int_0^T \int_\Omega < \hat{\boldsymbol{u}} | P_2^* \frac{\partial \boldsymbol{\eta}}{\partial t} > dV dt + \int_0^T \int_\Omega < D_3^* P_1^* \boldsymbol{\eta} | \hat{\boldsymbol{u}} > dV dt = \int_0^T \int_\Omega < \delta P_2 \frac{\partial \boldsymbol{u}}{\partial t} - \delta P_1 D_3 \boldsymbol{u} | \boldsymbol{\eta} > dV dt. \tag{12.35}$$

The adjoint model can now be defined to get rid of the remaining terms in $\hat{\boldsymbol{u}}$, which in this case is defined as

$$\begin{cases} -P_2^* \frac{\partial \boldsymbol{\eta}}{\partial t} + D_3^* P_1^* \boldsymbol{\eta} = \boldsymbol{u} - \boldsymbol{u}^{obs}, \\ \boldsymbol{\eta}(t=T) = 0. \end{cases} \tag{12.36}$$

Since the initial conditions are at $t = T$, this is a backward model in time. With this model, the previous equation becomes

$$\int_0^T \int_\Omega < \boldsymbol{u} - \boldsymbol{u}^{obs} | \hat{\boldsymbol{u}} > dVdt = \int_0^T \int_\Omega < \delta P_2 \frac{\partial \boldsymbol{u}}{\partial t} - \delta P_1 D_3 \boldsymbol{u} | \boldsymbol{\eta} > dVdt, \tag{12.37}$$

$$\delta J_{C,\rho}(\delta C, \delta\rho) = \int_0^T \int_\Omega < \delta P_2 \frac{\partial \boldsymbol{u}}{\partial t} - \delta P_1 D_3 \boldsymbol{u} | \boldsymbol{\eta} > dVdt, \tag{12.38}$$

$$\nabla_C J(C) = \int_0^T \int_\Omega < C D_2 \boldsymbol{v} | R^* \boldsymbol{\eta} > dVdt, \tag{12.39}$$

$$\nabla_\rho J(\rho) = \int_0^T \int_\Omega \sum_{i=1}^{3} \frac{\partial v_i}{\partial t} | \eta_{i+6} dVdt. \tag{12.40}$$

Isotropic solids

In the case of an isotropic solid, the stiffness tensor C is

$$C = \begin{pmatrix} 2\mu+\lambda & \lambda & \lambda & 0 & 0 & 0 \\ \lambda & 2\mu+\lambda & \lambda & 0 & 0 & 0 \\ \lambda & \lambda & 2\mu+\lambda & 0 & 0 & 0 \\ 0 & 0 & 0 & 2\mu & 0 & 0 \\ 0 & 0 & 0 & 0 & 2\mu & 0 \\ 0 & 0 & 0 & 0 & 0 & 2\mu \end{pmatrix}, \tag{12.41}$$

which means that the gradient of J can be expressed in terms of λ and μ as

$$\nabla_\lambda J(\lambda) = \int_0^T \int_\Omega < 4\lambda \sum_{i=1}^{3} \sum_{j=3}^{3} \frac{\partial v_j}{\partial x_j} \eta_i \boldsymbol{v} | R^* \boldsymbol{\eta} > dVdt, \tag{12.42}$$

$$\nabla_\mu J(\mu) = \int_0^T \int_\Omega < 2\mu I_6 D_2 \boldsymbol{v} | R^* \boldsymbol{\eta} > dVdt. \tag{12.43}$$

Absorbing Boundary Conditions

The TLM can also be applied to the equation applied on the boundary. The Higdon ABCs are

$$\left(\prod_{i=1}^{n} \left(\cos\alpha_i \frac{\partial}{\partial t} - c \frac{\partial}{\partial x} \right) \right) p = 0 \tag{11.8}$$

and the initial conditions are

$$\begin{cases} \frac{\partial^k p}{\partial t^k}(t=0) = 0, \forall k \in [\![0,n]\!], \\ \frac{\partial^k p}{\partial x^k}(t=0) = 0, \forall k \in [\![0,n]\!], \end{cases} \tag{12.44}$$

since the simulation starts at rest. The Higdon ABCs are defined using coordinates related to the normal $\boldsymbol{n}$ of the surface Γ of the boundary. To make this clearer (11.8) can be rewritten as

$$\left(\prod_{i=1}^{n}\left(\cos\alpha_i\frac{\partial}{\partial t}-c<\boldsymbol{n}|\nabla>\right)\right)p=0. \tag{12.45}$$

Similarly to the TLM applied on the wave propagation equation, a cost function is defined as follows,

$$J(P)=\frac{1}{2}\int_0^T\int_\Omega(p-p^{\text{obs}})^2dV\,dt, \tag{12.46}$$

which after the same process of applying a small perturbation and taking the limit, the variation of the cost J with respect to the parameters of the model P is

$$\delta J_P(\delta P)=\int_0^T\int_\Omega(p-p^{\text{obs}})(\hat{p})dV\,dt. \tag{12.47}$$

Since none of the parameters in P intervene in (12.45), the equation obtained in $\hat{p}$ is the same,

$$\left(\prod_{i=1}^{n}\left(\cos\alpha_i\frac{\partial}{\partial t}-c<\boldsymbol{n}|\nabla>\right)\right)\hat{p}=0. \tag{12.48}$$

Once again, a variable η is introduced in the aim to transfer the derivatives from p to η in the equation

$$\int_0^T\int_\Gamma\eta\prod_{i=1}^{n}\left(\cos\alpha_i\frac{\partial}{\partial t}-c<\boldsymbol{n}|\nabla>\right)\hat{p}\,dS\,dt=0, \tag{12.49}$$

for which the integration by parts is (see below for proof)

$$\begin{aligned}
&\int_0^T \int_\Gamma \eta \prod_{i=1}^{n} \left(\cos\alpha_i \frac{\partial}{\partial t} - c < \boldsymbol{n}|\nabla > \right) \hat{p}\, dS dt = \\
&\qquad \left[\int_\Gamma \sum_{k=1}^{n} (-1)^{k+1} \prod_{i=1}^{k-1} \left(\cos\alpha_i \frac{\partial}{\partial t} - c < \boldsymbol{n}|\nabla > \right) \eta \right. \\
&\qquad\quad \left. \cos\alpha_k \prod_{i=k+1}^{n} \left(\cos\alpha_i \frac{\partial}{\partial t} - c < \boldsymbol{n}|\nabla > \right) \hat{p} dS \right]_0^T \\
&\qquad + \int_0^T \int_L \sum_{k=1}^{n} (-1)^{k} \prod_{i=1}^{k-1} \left(\cos\alpha_i \frac{\partial}{\partial t} - c < \boldsymbol{n}|\nabla > \right) \eta \\
&\qquad\quad c \prod_{i=k+1}^{n} \left(\cos\alpha_i \frac{\partial}{\partial t} - c < \boldsymbol{n}|\nabla > \right) \hat{p}\, dl dt \\
&\qquad\quad + (-1)^n \int_0^T \int_\Gamma \hat{p} \prod_{i=1}^{n} \left(\cos\alpha_i \frac{\partial}{\partial t} - c < \boldsymbol{n}|\nabla > \right) \eta\, dS dt, \qquad (12.50)
\end{aligned}$$

with the convention that an empty product is equal to 1, and L is the boundary of the surface Γ. As before, the goal is to remove the remaining terms in $\hat{p}$. The first term in the time bracket vanishes at $t = 0$, since $\hat{p}$ and its derivative are null. To simplify at $t = T$, the same condition will be applied on η at that time, creating the initial condition for the backward model as before. The second terms is null since the surface Γ is a closed surface. We obtain the adjoint system

$$\begin{cases} \prod_{i=1}^{n} \left(\cos\alpha_i \frac{\partial}{\partial t} - c < \boldsymbol{n}|\nabla > \right) \eta = p - p^{\text{obs}}, \\ \frac{\partial^k \eta}{\partial t^k}(t = T) = 0, \forall k \in [\![0, n]\!], \\ \frac{\partial^k \eta}{\partial x^k}(t = T) = 0, \forall k \in [\![0, n]\!]. \end{cases} \qquad (12.51)$$

Proof For $n = 1$, (12.45) only has one term, for which the integration by parts is

$$\begin{aligned}
&\int_0^T \int_\Gamma \eta \left(\cos\alpha_1 \frac{\partial}{\partial t} - c < \boldsymbol{n}|\nabla > \right) \hat{p}\, dS dt = \\
&\quad \left[\int_\Gamma \eta \cos\alpha_k \hat{p} dS \right]_0^T - \int_0^T \int_L \eta\, c p\, dl dt - \int_0^T \int_\Gamma \hat{p} \left(\cos\alpha_1 \frac{\partial}{\partial t} - c < \boldsymbol{n}|\nabla > \right) \eta\, dS dt.
\end{aligned} \qquad (12.52)$$

This shows that (12.50) is true for $n = 1$. For $n \in \mathbb{N}$, let us assume (12.50) to be true. For $n+1$, the term in α_{n+1} can be separated from the product of derivatives as follows,

$$\int_0^T \int_\Gamma \eta \prod_{i=1}^{n+1} \left(\cos\alpha_i \frac{\partial}{\partial t} - c < \boldsymbol{n}|\nabla > \right) \hat{p}\, dSdt = \\ \int_0^T \int_\Gamma \eta \prod_{i=1}^{n} \left(\cos\alpha_i \frac{\partial}{\partial t} - c < \boldsymbol{n}|\nabla > \right) \left(\cos\alpha_{n+1} \frac{\partial}{\partial t} - c < \boldsymbol{n}|\nabla > \right) \hat{p}\, dSdt, \quad (12.53)$$

which then allows us to use the hypothesis in n to develop the product terms, thus giving

$$\int_0^T \int_\Gamma \eta \prod_{i=1}^{n+1} \left(\cos\alpha_i \frac{\partial}{\partial t} - c < \boldsymbol{n}|\nabla > \right) \hat{p}\, dSdt = \\ \left[\int_\Gamma \sum_{k=1}^{n} (-1)^{k+1} \prod_{i=1}^{k-1} \left(\cos\alpha_i \frac{\partial}{\partial t} - c < \boldsymbol{n}|\nabla > \right) \eta \cos\alpha_k \prod_{i=k+1}^{n+1} \left(\cos\alpha_i \frac{\partial}{\partial t} - c < \boldsymbol{n}|\nabla > \right) \hat{p} dS \right]_0^T \\ + \int_0^T \int_L \sum_{k=1}^{n} (-1)^{k} \prod_{i=1}^{k-1} \left(\cos\alpha_i \frac{\partial}{\partial t} - c < \boldsymbol{n}|\nabla > \right) \eta\, c \prod_{i=k+1}^{n+1} \left(\cos\alpha_i \frac{\partial}{\partial t} - c < \boldsymbol{n}|\nabla > \right) \hat{p}\, dldt \\ + (-1)^n \int_0^T \int_\Gamma \left(\cos\alpha_{n+1} \frac{\partial}{\partial t} - c < \boldsymbol{n}|\nabla > \right) \hat{p} \prod_{i=1}^{n} \left(\cos\alpha_i \frac{\partial}{\partial t} - c < \boldsymbol{n}|\nabla > \right) \eta\, dSdt. \\ (12.54)$$

The integration by parts of the last term is

$$(-1)^n \int_0^T \int_\Gamma \left(\cos\alpha_{n+1} \frac{\partial}{\partial t} - c < \boldsymbol{n}|\nabla > \right) \hat{p} \prod_{i=1}^{n} \left(\cos\alpha_i \frac{\partial}{\partial t} - c < \boldsymbol{n}|\nabla > \right) \eta\, dSdt = \\ \left[\int_\Gamma (-1)^{n+2} \prod_{i=1}^{n} \left(\cos\alpha_i \frac{\partial}{\partial t} - c < \boldsymbol{n}|\nabla > \right) \eta \cos\alpha_{n+1} \hat{p} dS \right]_0^T \\ + \int_0^T \int_L (-1)^{n+1} \prod_{i=1}^{n} \left(\cos\alpha_i \frac{\partial}{\partial t} - c < \boldsymbol{n}|\nabla > \right) \eta\, c\hat{p}\, dldt \\ + (-1)^{n+1} \int_0^T \int_\Gamma \hat{p} \prod_{i=1}^{n+1} \left(\cos\alpha_i \frac{\partial}{\partial t} - c < \boldsymbol{n}|\nabla > \right) \eta\, dSdt, \\ (12.55)$$

thus finally giving

$$\begin{aligned}
&\int_0^T \int_\Gamma \eta \prod_{i=1}^{n+1} \left(\cos\alpha_i \frac{\partial}{\partial t} - c < \boldsymbol{n} | \nabla > \right) \hat{p}\, dSdt = \\
&\left[\int_\Gamma \sum_{k=1}^{n+1} (-1)^{k+1} \prod_{i=1}^{k-1} \left(\cos\alpha_i \frac{\partial}{\partial t} - c < \boldsymbol{n} | \nabla > \right) \eta \cos\alpha_k \prod_{i=k+1}^{n+1} \left(\cos\alpha_i \frac{\partial}{\partial t} - c < \boldsymbol{n} | \nabla > \right) \hat{p} dS \right]_0^T \\
&\quad + \int_0^T \int_L \sum_{k=1}^{n+1} (-1)^{k} \prod_{i=1}^{k-1} \left(\cos\alpha_i \frac{\partial}{\partial t} - c < \boldsymbol{n} | \nabla > \right) \eta\, c \prod_{i=k+1}^{n+1} \left(\cos\alpha_i \frac{\partial}{\partial t} - c < \boldsymbol{n} | \nabla > \right) \hat{p}\, dldt \\
&\qquad + (-1)^{n+1} \int_0^T \int_\Gamma \hat{p} \prod_{i=1}^{n+1} \left(\cos\alpha_i \frac{\partial}{\partial t} - c < \boldsymbol{n} | \nabla > \right) \eta\, dSdt.
\end{aligned} \tag{12.56}$$

Since (12.50) is true for $n = 1$, and if it is true for n, it is true for $n+1$, then by induction (12.50) is true $\forall n \in \mathbb{N}^*$.

Adjoint for C-PML In [223], using a Lagrange multiplier technique, the authors derive a complete adjoint system for the C-PML boundary conditions in the wave equation.

12.3 Conclusion

We have shown that the adjoint approach can provide explicit expressions for the gradient of a cost function that we want to minimize in order to find optimal material parameters. These optimal values guarantee the best possible fit between model simulations and measured data. The solution of the inverse problem of parameter identification can thus be found by solving for the adjoint state, computing the gradient, then using this gradient to minimize the mismatch cost function.

This theoretical approach will be applied in the next Chapter to wave propagation in and around the sperm whale's head. Then, using data from the measurement campaigns that were described in the first Part of this thesis, we should be able to construct an efficient simulation tool of the whale's biosonar system.

Chapter 13

Feasibility and Implementation of Coupling

In this chapter we describe how to apply the results from our propagation model and inverse problem formulations, based first on synthetic and then on field measurements, to the sperm whale biosonar acoustic system.

13.1 Introduction

As explained in the previous chapter, we can use parameter estimation methods to solve the inverse problem that consists of determining (or estimating) the material properties (or the properties of the source) so that the output of the model simulations reproduces the measured signals as closely as possible. This can be done in three ways:

1. Adjoint method.
2. Bayesian inference.
3. Neural networks.

The second and third approaches will not be dealt with here, but are an important element in our future research plans, where we would like to combine them in some intelligent way. We will concentrate here on the adjoint method, which itself can later be used as a basis for a neural differential equation approach.

Before attempting to invert real data, it is customary to begin with twin experiments [211], where we use the propagation model to generate synthetic signals, and then use these synthetics for solving the inverse problem. This approach enables an initial calibration of the inversion method, which is vital to obtain before attempting inversions of field data.

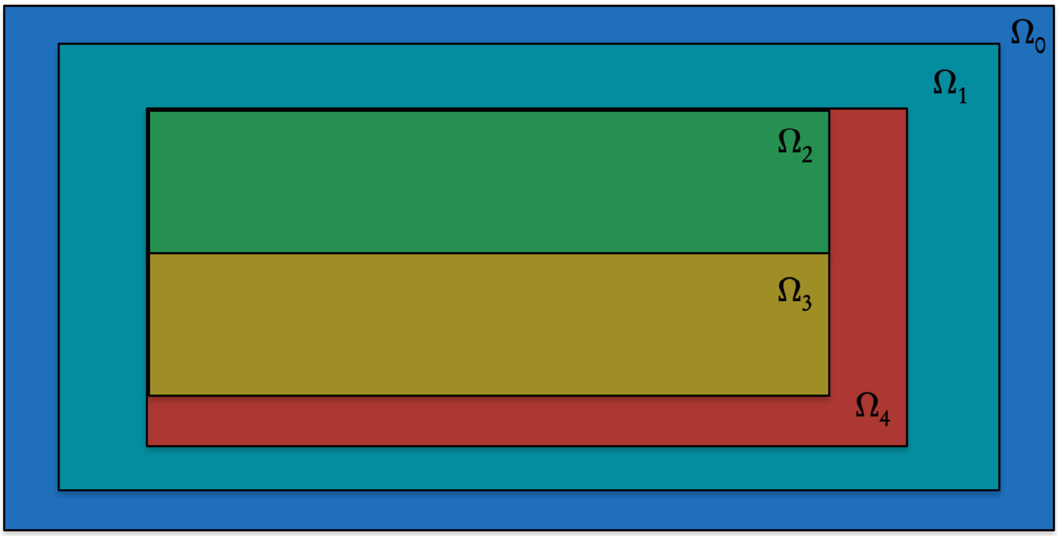


Figure 13.1: Simple model of a sperm whale's head.

13.2 Twin Experiments

Twin experiments, or synthetic runs, are a basic and indispensable tool for all inverse problems. In order to evaluate the performance of an inversion chain we begin with the following methodology.

1. Fix all parameters and unknowns and define a reference trajectory, obtained from a run of the direct model - call this the "truth".
2. Derive a set of (synthetic) measurements, or background data, from this "true" run.
3. Optionally, perturb these observations with random noise in order to generate a more realistic observed state.
4. Run the inverse problem algorithm, starting from an initial guess (different from the "true" initial state used above), using the synthetic observations.
5. Evaluate the performance using a suitable loss function, modify the model/algorithm/observations and cycle back to Step 1.

13.3 Inverse Problem Formulation

We set up a model problem, based on a simplified representation of the sperm whale's head (see Figure 11.10), that is shown in Figure 13.1. We define a domain $\Omega = \bigcup_{i=0}^{4} \Omega_i$, where

- Ω_0 represents the surrounding sea water,
- Ω_1 represents the blubber and skin,
- Ω_2 represents the junk,
- Ω_3 represents the spermaceti, and
- Ω_4 represents the bone,

in the whale's head. We formulate the direct problem of elastic wave propagation in the domain Ω and apply absorbing boundary conditions on Γ, its outer border. We suppose that we have measurements/recordings available in the water layer Ω_0 and we will use these to solve the following inverse problem of parameter estimation: for given measurements of the acoustic field, find the material properties that minimize the cost function J defined by

$$J(\theta) = \frac{1}{2}\int_0^T \int_\Omega e^2(x,t;\theta)\,\mathrm{d}x\,\mathrm{d}t + \alpha \int_0^T \int_\Omega \left|\nabla\theta\right|^2 \,\mathrm{d}x\,\mathrm{d}t,$$

where e is the mismatch between simulated and measured data, and α is a regularization coefficient.

More specific inverse problems can be formulated from the general setting. In particular:

1. Measurements are partially known, either in a restricted spatial domain, or on sub intervals of time.
2. Some material parameters are known, and we seek a subset consisting of the unknown (or badly known) materials.
3. Different recordings, possibly from different individuals, can be accumulated in a composite cost function.

13.4 Implementation of Coupling

Our initial objective will be to implement a complete inversion chain for the simplified, twin experiment described in the previous sections. Then we will describe the steps required for extending the methodology to real, measured data. The actual implementations in the synthetic and real cases, are work in progress.

13.4.1 Synthetic Inversion

As detailed above, we want to start by performing an inversion on an "academic" case, where we suppose in particular that data is available at all points on the boundary of the domain. We then relax the conditions, by:

1. Taking sparse measurements, at a few grid points only.
2. Increasing the noise level of the synthetic signals.

By doing this, we can test and tune the inversion algorithm, thus preparing the system for dealing with real data. These simulations are in progress, and due to a lack of time, will be reported later.

13.4.2 Towards Inversion of Real Data

The major difficulties that we encounter when dealing with real data, are due to the limited number of points where we have measurements available and to the noisy nature of the recorded signals. The measurement points are the hydrophone positions in the sea campaigns and they cannot capture all the points surrounding the head. As a result, the cost function can be evaluated at relatively few positions and this results in a very ill-posed inverse problem. The question is then, how to regularize this problem in order to ensure convergence to meaningful parameter estimations? Numerous regularization options exist:

1. Test and choose among a range of different regularization norms.
2. Use a composite (sum) cost function, possibly from distinct measurements, individuals and campaigns.
3. Perform second-order, sensitivity analysis to find the best (most sensitive) parameters to estimate.
4. Use machine learning techniques to obtain a surrogate model.
5. Use Bayesian estimation techniques that take into account the stochastic, or probabilistic nature of the signals.

This requires a vast modelling and simulation effort, and is the subject of ongoing and future research in the framework of the ANR-funded AI chair, ADSIL, led by H. Glotin.

13.5 Conclusion

Although the actual numerical computations have not been performed yet, due to lack of time, all the necessary ingredients have been prepared in this chapter and the two preceding ones in order to construct an efficient simulation tool for the complete biosonar system of the sperm whale.

Chapter 14

Conclusions and Outlook

The complete chain of analysis developed throughout this manuscript is aimed at improving our knowledge of the mechanisms that govern the sperm whale sonar. With sperm whales, classical experiments that are done on bats or dolphins to fully qualify their sonar in every aspect cannot be done, since sperm whales are wild and large species. This means that there are no measurements of their nervous response, testing their hearing capability, precise measurement of their beam pattern, or any other experiment that requires the animal to wear large equipment, or be trained to have a specific behaviour for an experiment. Hence, the need for a simulation of the sperm whale sonar.

As described in Chapter 1, the need was split into three goals:

- The making of the simulation.
- The coupling of the measurement with the simulation.
- The use of machine learning to classify and improve measurements.

The results of these goals that were explored by this thesis will be now summarized, and then some promising research outlooks will be presented.

Data acquisition and processing

Three databases were presented, each one having specific characteristics:

- BOMBYX, a database made from two hydrophones, at 50 kHz, during long term sessions, presented in Chapter 4.
- Sphyrna, a database made from five hydrophones, at 384 kHz, recorded on ASV, presented in Chapter 5.
- Sarano, a database made from two to four hydrophones, at 300 kHz, recorded near the sperm whale, presented in Chapter 6.

These databases were analysed using the signal processing techniques introduced in Chapter 3, following the pipeline of detection, TDOA estimation, and filtering.

Detection The detection step that was proposed and used for the three databases is also made up of successive steps, starting with a band-pass filter, followed by a TK filter, and ending with the use of a threshold. An alternative detection process is also mentioned, using the neural network architecture of Chapter 7 to classify ten species of marine mammals based on raw recordings of their clicks.

TDOA estimation The geometric SRP method was presented to estimate the TDOA, and was compared to the standard SRP. Both methods have the same complexity and have the same advantages when compared to a single pair of channel TDOA estimations, when used on a system with the same amount of noise and signal level on all channels. However, they differ in the fact that SRP uses addition, while geometric SRP uses multiplication. If a click is not present on a channel, SRP will still produce good TDOA for the other channels, while the geometric SRP will produce random TDOA with a lower correlation value. When all the TDOAs are needed to estimate the position of the source, these two elements make it easier to filter out these clicks. Since BOMBYX has only two hydrophones, the geometric SRP was not used for its analysis, but it was used for the two other databases.

Filtering Two methods of post filtering were presented. First, the ellipse method was used to recover the shape of the antenna, but also use to filter out spurious clicks not generated from signals coming from the water, thus with TDOA far from the ellipsoid. The other filtering method was the DBSCAN clustering presented in Section 5.5.1. From the idea that a point inside a track with its two neighbours can be seen as a cluster of three points, with the track being a succession of connected clusters, DBSCAN is the clustering method that corresponds to this exact definition. With BOMBYX only having two channels, clicks with random TDOA have a higher chance to be considered as a part of one cluster, and not be filtered out. However, this method was still employed for all three databases as it showed good results even on BOMBYX.

Machine learning

In Chapter 7 we introduce a new challenge of *Odontoceti* click detection, named DOCC10. It answers the need for a large database of marine mammal transients in the intent of having enough training examples to train a neural network, or other machine learning method, to be able to classify each click according to its species. The presentation of DOCC10 was introduced along with another database named DOCC7, a version of DOCC10 where three classes have been removed, as they intersect other classes, or are too different between the train and test sets.

A neural network architecture was also proposed alongside the DOCC10 database. As of today, it achieved a state of the art score on DOCC10 when only the train set is used during training. This model, which uses directly the raw signal as an input, can also be used as a detector.

Coupling and simulation

With the theory of wave propagation and a finite difference in time domain method, a simulation was built to model the sperm whale sonar. To correctly calibrate the parameters of this simulation, a coupling method using a tangent linear model was conceived. The use of this method, which improves the parameters of the simulation using measurements as a target, was not applied to real data in this thesis, due to time constraints. Yet the simulation was still run by itself, showing that at least some mechanisms that govern the sonar of the sperm whale were recovered, namely, the multiple pulse structure of sperm whale clicks and the correct IPI value were measured.

Outlook

As explained in Chapter 13, the implementations of coupling in the synthetic and the real cases are ongoing work, and did not make it into this manuscript due to time constraints. The implementation in the real case will tune the parameters of the model, improving the synthetic click generated. At this point, the mechanisms that appear in the model can be studied to hypothesize the actual mechanisms that govern the sonar. However, the model could still be improved at this point. Indeed the numerical method that was chosen was the FDTD due to its simplicity and iteration speed, with the downfall of having a lesser accuracy. With the improved parameters, the model and coupling method could be implemented with a more accurate numerical method, such as the spectral finite element method. The ultimate objective is to combine this propagation modelling with the machine learning techniques used for signal analysis. This is ongoing work.

As stated in Chapter 12, machine learning can also be used to estimate the parameters of a model using observation as an input. This usage alongside with the possibility of incorporating the differential equations directly into the neural network are two key aspects of neural networks that could help solving inverse problems, making them research axes with great potential.

Another improvement point is the use of nonlinear wave theory. The use of linear wave theory is useful as a first approach, and makes it easier to interpret the results generated by the model. Yet, we are well aware that the louder clicks emitted by sperm whales have a short shock distance, meaning that nonlinear effects cannot be neglected, even inside its head.

The database created at the end of Chapter 6 linked clicks to the sperm whale that emitted them. This database have the advantages of exploring the features that are shared between clicks from the same sperm whale, which if they exist and are discriminatory enough (thus excluding the IPI) would lead to a classifier able to distinguish sperm whales by only using acoustics. Moreover, social interaction between multiple individuals can be studied, since the video channel allows us to see the behaviour that resulted from various coda exchanges, while the video database, allows us to know which individual emitted which coda.

These three research outlooks show that the amount of work remaining is not small, but will surely lead to interesting results.

Chapter 15

List of Publications

Conference Proceedings or Report

[1] M. Ferrari, M. Poupard, P. Giraudet, R. Marxer, J.-M. Prévot, T. Soriano, and H. Glotin, “Efficient artifacts filter by density-based clustering in long term 3d whale passive acoustic monitoring with five hydrophones fixed under an autonomous surface vehicle,” in *OCEANS 2019-Marseille*, IEEE, 2019, pp. 1–7.

[2] M. Ferrari, H. Glotin, R. Marxer, V. Barchasz, V. Sarano, V. Gies, M. Asch, and F. Sarano, “High-frequency near-field physeter macrocephalus monitoring by stereo-autoencoder and 3d model of sonar organ,” in *OCEANS 2019-Marseille*, IEEE, 2019, pp. 1–4.

[3] M. Ferrari, R. Marxer, M. Asch, and H. Glotin, “Wave propagation in the biosonar organ of sperm whales using a finite difference time domain method,” in *VIHAR*, 2019.

[4] M. Ferrari, H. Glotin, R. Marxer, and M. Asch, “Docc10: Open access dataset of marine mammal transient studies and end-to-end cnn classification,” in *IJCNN*, 2020.

[5] H. Glotin, P. Spong, H. Symonds, V. Roger, R. Balestriero, M. Ferrari, M. Poupard, J. Towers, S. Veirs, R. Marxer, *et al.*, “Deep learning for ethoacoustical mapping: Application to a single cachalot long term recording on joint observatories in vancouver island,” *The Journal of the Acoustical Society of America*, vol. 144, no. 3, pp. 1776–1777, 2018.

[6] V. Roger, M. Ferrari, R. Marxer, F. Chamroukhi, and H. Glotin, “Towards the topology of autoencoder of calls versus clicks of marine mammal,” *The Journal of the Acoustical Society of America*, vol. 144, no. 3, pp. 1777–1778, 2018.

[7] M. Poupard, M. Ferrari, J. Schluter, R. Marxer, P. Giraudet, V. Barchasz, V. Gies, G. Pavan, and H. Glotin, “Real-time passive acoustic 3d tracking of deep diving cetacean by small non-uniform mobile surface antenna,” in *ICASSP 2019-*

2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2019, pp. 8251–8255.

[8] M. Poupard, M. Ferrari, J. Schluter, P. Astruch, B. Schohn, B. Rouanet, A. Goujard, A. Lyonnet, P. Giraudet, V. Barchasz, *et al.*, "Passive acoustics to monitor flagship species near boat traffic in the unesco world heritage natural reserve of scandola," in *Input Academy: International Conference on Innovation in Urban and regional planning*, 2019.

[9] P. Best, M. Ferrari, M. Poupard, S. Paris, R. Marxer, H. Symonds, P. Spong, and H. Glotin, "Deep learning and domain transfer for orca vocalization detection," in *International joint conference on neural networks*, 2020.

[10] H. Glotin, N. Thellier, P. Best, M. Poupard, M. Ferrari, S. Viera, G. Donzé, M. Campana, J. Chevallier, P. Giraudet, F. Malige, J. Patris, J. Prévot, V. Giés, N. Prévot d' Alvise, V. Barchasz, S. Marzetti, F. Sarano, S. Gaillard, and F. de Varenne, "Sphyrna-odyssey : Découvertes de chasses abyssales de cachalots en alliance et des effets du confinement covid19," 220, CNRS LIS Univ Toulon Ed. 2020. [Online]. Available: `https://sabiod.univ-tln.fr/pub/SPHYRNA/SO_Premier_Rapport.pdf`.

Appendix A

Nonlinear Acoustics

Nonlinear acoustics might seem to be a rare phenomenon, yet it was the first form of acoustics observed [224, 225, 226]. Nonlinear acoustics is mostly seen in sound generation, while sound propagation is usually linear, especially in mediums such as water or air when the signal amplitude is not too high to let the nonlinear effect be observed within the distance of the observable environment.

A.1 Nonlinear wave propagation in solids

As presented earlier, linear wave propagation is based on Hooke's law, which describes the linear elastic theory using second order elastic constants c_{ijkl}. Nonlinear effects start to appear in the equation when third order elastic constants are introduced, in the strain-stress relation. This relation is developed in the nonlinear elastic theory [227] to include those additional nonlinear term. These third order elastic constants c_{ijklmn} will affect the speeds of sound, causing what is called the acoustoelastic effect, which was first studied by [228].

By assuming that the elastic potential energy E can be approximated by a Taylor series, and that it is null at the equilibrium, E is obtain by

$$E(\varepsilon) = \frac{1}{2!}\sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{k=1}^{3}\sum_{l=1}^{3} c_{ijkl}\varepsilon_{ij}\varepsilon_{kl} + \frac{1}{3!}\sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{k=1}^{3}\sum_{l=1}^{3}\sum_{m=1}^{3}\sum_{n=1}^{3} c_{ijklmn}\varepsilon_{ij}\varepsilon_{kl}\varepsilon_{mn} + \ldots \tag{A.1}$$

which can be used to define the constants c_{ijkl} and c_{ijklmn}. Note that similarly to the second order elastic constants, the Voigt notation can be used for the third order elastic constants (i.e. C_{ijk}). In the case of cubic crystals, only 6 independent constants are needed (C_{111}, C_{112}, C_{144}, C_{166}, C_{123}, C_{456}) [229]. In [230] they demonstrate that for a cubic crystal, in the direction [100], [110] and [111], only purely longitudinal waves exist, and follow the equation

$$\rho_0 \frac{\partial^2 u}{\partial t^2} = K_2 \frac{\partial^2 u}{\partial x^2} + (3K_2 + K_3)\frac{\partial u}{\partial x}\frac{\partial^2 u}{\partial x^2} = K_2 \frac{\partial^2 u}{\partial x^2}\left(1 + \left(3 + \frac{K_3}{K_2}\right)\frac{\partial u}{\partial x}\right), \tag{A.2}$$

Direction	K_2	K_3
[100]	C_{11}	C_{111}
[110]	$\frac{C_{11}+C_{12}+2C_{44}}{2}$	$\frac{C_{111}+3C_{112}+12C_{166}}{4}$
[111]	$\frac{C_{11}+2C_{12}+4C_{44}}{3}$	$\frac{C_{111}+C_{112}+12C_{144}+24C_{166}+2C_{123}+16C_{456}}{9}$

Table A.1: K_2 and K_3 value in a cubic crystal

where K_2 and K_3 are the coefficients given in Table A.1

From (A.2), the nonlinear coefficient β for a cubic crystal can be defined as

$$\beta = -\left(3 + \frac{K_3}{K_2}\right). \tag{A.3}$$

The minus sign is introduced to ensure that β is positive. In case that the nonlinear term can be neglected, the remaining linear term gives $\rho_0 \frac{\partial^2 u}{\partial t^2} = K_2 \frac{\partial^2 u}{\partial x^2}$ leading to the relation $K_2 = \rho_0 c_0$.

A.2 Nonlinear wave propagation in fluids

To solve the state of a fluid at a point, 6 equations are needed to evaluate the 6 state variable p, ρ, T and $\boldsymbol{v}$. The 6 equations needed are the conservation of mass (9.3), the conservation of momentum (9.13), the conservation of energy (9.14), and a state equation,

$$p = p(\rho, s), \tag{A.4}$$

which can be developed in a Taylor series. In the isentropic case

$$p = p_0 + (\rho - \rho_0)\left(\frac{\partial p}{\partial \rho}\right)_{s,\rho=\rho_0} + \frac{(\rho-\rho_0)^2}{2}\left(\frac{\partial^2 p}{\partial \rho^2}\right)_{s,\rho=\rho_0} + \mathcal{O}\left((\rho-\rho_0)^3\right) \tag{A.5}$$

Here and going on, the following common notation is adopted. The subscript $_0$ denotes the quantities at rest (in a medium without perturbation) while the superscript $'$ denotes a small perturbation around that state (e.g. $p' = p - p_0$). The superscript $'$ is often dropped (e.g. p' is simply denoted by p) and will be dropped starting from (A.19). Note that since the value at rest represents a valid state, the equations can be simplified by cancelling out the terms describing that state.

Let us define the coefficients A and B [231] as

$$A = \rho_0 \left(\frac{\partial p}{\partial \rho}\right)_{s,\rho=\rho_0} = \rho_0 c_0^2, \tag{A.6}$$

$$B = \rho_0^2 \left(\frac{\partial^2 p}{\partial \rho^2}\right)_{s,\rho=\rho_0}, \tag{A.7}$$

$$c_0 = \sqrt{\left(\frac{\partial p}{\partial \rho}\right)_{s,\rho=\rho_0}}, \tag{A.8}$$

where c_0 is the isentropic sound velocity. The Taylor development thus becomes

$$p - p_0 = A\frac{\rho - \rho_0}{\rho_0} + B\frac{(\rho - \rho_0)^2}{2\rho_0^2} + \mathcal{O}\left((\rho - \rho_0)^3\right). \tag{A.9}$$

The ratio $\frac{B}{A} = \frac{\rho_0}{c_0^2}\left(\frac{\partial^2 p}{\partial \rho^2}\right)_{s,\rho=\rho_0}$ is called the nonlinearity parameter. However, this ratio cannot be measured using this definition in part due to the error induced by the second order derivative,

$$\begin{aligned}\left(\frac{\partial^2 p}{\partial \rho^2}\right)_{s,\rho=\rho_0} &= \left(\frac{\partial c^2}{\partial \rho}\right)_{s,\rho=\rho_0}\\ &=2c_0\left(\frac{\partial c}{\partial \rho}\right)_{s,\rho=\rho_0}\\ &=2c_0\left(\frac{\partial c}{\partial p}\right)_{s,p=p_0}\left(\frac{\partial p}{\partial \rho}\right)_{s,\rho=\rho_0}\\ &=2c_0^3\left(\frac{\partial c}{\partial p}\right)_{s,p=p_0}.\end{aligned} \tag{A.10}$$

This leads to a definition of the nonlinearity parameter that can be obtained with a measure of the variation of the sound speed due to an isentropic pressure variation,

$$\frac{B}{A} = 2\rho_0 c_0 \left(\frac{\partial c}{\partial p}\right)_{s,p=p_0} \tag{A.11}$$

and

$$\beta = 1 + \frac{B}{2A}. \tag{A.12}$$

A.2.1 Dimensionless quantities

As for the wave equation section, the ambient velocity will be considered as negligible.

The usual solution for a thermoviscous fluid is the sum of three modes [232, 233, 234]. The first one is the acoustic mode $\boldsymbol{v}_p$, which is studied below. The second one is the entropy mode $\boldsymbol{v}_s$, which is used to study conduction and heat transfer. Finally the last one is the vorticity mode $\boldsymbol{v}_\times$. We have

$$\boldsymbol{v}' = \boldsymbol{v}_p + \boldsymbol{v}_s + \boldsymbol{v}_\times. \tag{A.13}$$

In this decomposition, only the vorticity mode has a non zero curl, but its divergence is zero (similar to the Helmholtz decomposition (9.27)).

The hypothesis can be made that these modes are only coupled in the boundary conditions, if the two following conditions are met:

$$w \ll \frac{\rho c^2}{\frac{4}{3}\mu + \zeta}, \tag{A.14}$$

$$w \ll \frac{\rho c^2 c_p}{\kappa}, \tag{A.15}$$

where $\omega = 2\pi f$ is the angular frequency. This means that the hypothesis is met for frequencies below 1 GHz in the air, and below 1 THz in water.

The acoustic mach number is

$$\epsilon = \frac{v}{c_0} \tag{A.16}$$

and

$$\eta = \frac{\mu\omega}{\rho_0 c_0^2} = \frac{1}{R_e} \tag{A.17}$$

which measures the ratio between the amplitude of the viscous stress and the amplitude of pressure variation for a plane wave. It is the inverse of R_e the acoustic Reynolds number. The Prandtl number is

$$P_r = \frac{\mu c_p}{\kappa}. \tag{A.18}$$

The difference between the acoustic mode and the entropy mode, is that the acoustic modes have no entropy perturbation, while the entropy modes have zero pressure perturbation.

A.2.2 Second order equation

By assuming that the conditions (A.14) and (A.15) are met, and that the fluid is studied outside of the boundary layers, then the acoustics mode can be studied independently of the other two modes. To shorten the notation, the acoustic mode will now be noted $\boldsymbol{v}$ instead of $\boldsymbol{v}_p$.The conservation of mass (9.3) can be approximated at the first order in ϵ, by assuming that ϵ and η have the same order of magnitude, to become

$$\frac{\partial \rho}{\partial t} + \rho_0 \nabla \cdot \boldsymbol{v} = -\rho \nabla \cdot \boldsymbol{v} - \boldsymbol{v} \cdot \nabla \rho. \tag{A.19}$$

Its two right terms can be expanded using (9.22) and (9.23)

$$-\rho\nabla\cdot\boldsymbol{v} = -\left(\frac{p}{c_0^2}\right)\left(-\frac{1}{\rho_0}\frac{\partial\rho}{\partial t}\right) = -\left(\frac{p}{c_0^2}\right)\left(-\frac{1}{\rho_0 c_0^2}\frac{\partial p}{\partial t}\right) = \frac{p}{\rho_0 c_0^4}\frac{\partial p}{\partial t}, \tag{A.20}$$

$$-\boldsymbol{v}\cdot\nabla\rho = -\boldsymbol{v}\cdot\frac{\nabla p}{c_0^2} = -\frac{1}{c_0^2}\boldsymbol{v}\cdot\left(-\rho_0\frac{\partial\boldsymbol{v}}{\partial t}\right) = \frac{\rho_0}{2c_0^2}\frac{\partial\boldsymbol{v}}{\partial t} = \frac{1}{c_0^2}\frac{\partial}{\partial t}\left(\frac{1}{2}\rho_0\boldsymbol{v}^2\right). \tag{A.21}$$

Thus leading to the simplification

$$\begin{aligned}\frac{\partial\rho}{\partial t} + \rho_0\nabla\cdot\boldsymbol{v} &= \frac{p}{\rho_0 c_0^4}\frac{\partial p}{\partial t} + \frac{1}{c_0^2}\frac{\partial}{\partial t}\left(\frac{1}{2}\rho_0\boldsymbol{v}^2\right)\\ &= \frac{1}{\rho_0 c_0^4}\frac{\partial p^2}{\partial t} - \frac{1}{2}\frac{1}{\rho_0 c_0^4}\frac{\partial p^2}{\partial t} + \frac{1}{c_0^2}\frac{\partial}{\partial t}\left(\frac{1}{2}\rho_0\boldsymbol{v}^2\right)\\ &= \frac{1}{\rho_0 c_0^4}\frac{\partial p^2}{\partial t} + \frac{1}{c_0^2}\frac{\partial L}{\partial t},\end{aligned} \tag{A.22}$$

where L is the Lagrangian density, defined as

$$L = \frac{1}{2}\left(\rho_0 v^2 - \frac{p^2}{\rho_0 c_0^2}\right). \tag{A.23}$$

In the same manner, the momentum equation (9.13) becomes

$$\rho_0\frac{\partial\boldsymbol{v}}{\partial t} + \nabla p = -\rho_0\boldsymbol{v}\cdot\nabla\boldsymbol{v} + \mu\nabla^2\boldsymbol{v} + \left(\zeta + \frac{1}{3}\mu\right)\nabla(\nabla\cdot\boldsymbol{v}) - \rho\frac{\partial v}{\partial t}. \tag{A.24}$$

To further develop this equation the following identities can be used

$$\nabla(\nabla\cdot\boldsymbol{v}) = \nabla^2\boldsymbol{v} + \nabla\times\nabla\times\boldsymbol{v}, \tag{A.25}$$

$$\boldsymbol{v}\cdot\nabla\boldsymbol{v} = \frac{1}{2}\nabla\boldsymbol{v}^2 - \boldsymbol{v}\times\nabla\times\boldsymbol{v} \tag{A.26}$$

Since the curl is null for the acoustic mode ($\nabla\times\boldsymbol{v} = 0$), (A.24) becomes

$$\rho_0\frac{\partial\boldsymbol{v}}{\partial t} + \nabla p = -\frac{1}{2}\rho_0\nabla\boldsymbol{v}^2 + \left(\zeta + \frac{4}{3}\mu\right)\nabla(\nabla\cdot\boldsymbol{v}) - \rho\frac{\partial v}{\partial t}. \tag{A.27}$$

The two most right terms can then be treated

$$\left(\zeta + \frac{4}{3}\mu\right)\nabla(\nabla\cdot\boldsymbol{v}) = \left(\zeta + \frac{4}{3}\mu\right)\nabla\left(-\frac{1}{\rho_0}\frac{\partial\rho}{\partial t}\right) = -\frac{1}{\rho_0 c_0^2}\left(\zeta + \frac{4}{3}\mu\right)\nabla\frac{\partial p}{\partial t}, \tag{A.28}$$

$$-\rho\frac{\partial\boldsymbol{v}}{\partial t} = -\frac{p}{c_0^2}\left(-\frac{1}{\rho_0}\nabla\right) = \frac{1}{2\rho_0 c_0^2}\nabla p^2 = \nabla\left(\frac{p^2}{2\rho_0 c_0^2}\right). \tag{A.29}$$

Thus leading to a linearized version of the conservation of momentum equation

$$\rho_0\frac{\partial\boldsymbol{v}}{\partial t} + \nabla p = -\frac{1}{\rho_0 c_0^2}\left(\zeta + \frac{4}{3}\mu\right)\nabla\frac{\partial p}{\partial t} - \nabla L. \tag{A.30}$$

The next equation that needs to be considered is the density state equation

$$\rho = \frac{p}{c_0^2} - \frac{1}{\rho_0 c_0^4}\frac{B}{2A}p^2 - \frac{1}{c_0^2}\left(\frac{\partial p}{\partial s}\right)_\rho s \tag{A.31}$$

At the first order, the temperature state equation is

$$T = \frac{T_0 \alpha}{\rho_0 c_p} p \tag{A.32}$$

which allows us to integrate the first order approximation of the energy conservation (9.14) regarding time

$$\rho_0 T_0 \frac{\partial s}{\partial t} = \kappa \nabla^2 T = \frac{\kappa}{c_0^2}\frac{\partial^2 T}{\partial t^2}, \tag{A.33}$$

$$s = \frac{\kappa}{c_0^2 \rho_0 T_0}\frac{\partial T}{\partial t}. \tag{A.34}$$

The temperature can then be replaced by the pressure since

$$\left(\frac{\partial p}{\partial s}\right)_\rho \frac{\partial T}{\partial t} = \left(\frac{\partial p}{\partial s}\right)_\rho \left(\frac{\partial T}{\partial \rho}\right)_s \frac{\partial \rho}{\partial t} = \frac{1}{c_0^2}\left(\frac{\partial p}{\partial s}\right)_\rho \left(\frac{\partial T}{\partial \rho}\right)_s \frac{\partial p}{\partial t}, \tag{A.35}$$

and by using

$$\left(\frac{\partial p}{\partial s}\right)_\rho = \rho^2 \left(\frac{\partial T}{\partial \rho}\right)_s, \tag{A.36}$$

$$\left(\frac{\partial T}{\partial \rho}\right)_s^2 = (c_p - c_v)\frac{T_0 c_0^2}{c_p c_v \rho_0}, \tag{A.37}$$

the density state equation become

$$\rho = \frac{p}{c_0^2} - \frac{1}{\rho_0 c_0^4}\frac{B}{2A}p^2 - \frac{\kappa}{\rho_0 c_0^4}\left(\frac{1}{c_v} - \frac{1}{c_p}\right)\frac{\partial p}{\partial t}. \tag{A.38}$$

Finally, to obtain the second order equation, the conservation of mass (A.22) is derived and the divergence of the conservation of momentum (A.30) are taken

$$\frac{\partial^2 \rho}{\partial t^2} + \rho_0 \nabla \cdot \frac{\partial \boldsymbol{v}}{\partial t} = \frac{1}{\rho_0 c_0^4}\frac{\partial^2 p^2}{\partial t^2} + \frac{1}{c_0^2}\frac{\partial^2 L}{\partial t^2}, \tag{A.39}$$

$$\rho_0 \nabla \cdot \frac{\partial \boldsymbol{v}}{\partial t} + \nabla^2 p = -\frac{1}{\rho_0 c_0^2}\left(\zeta + \frac{4}{3}\mu\right)\nabla^2 \frac{\partial p}{\partial t} - \nabla^2 L, \tag{A.40}$$

and are subtracted together and simplified using the state equation (A.38),

$$\nabla^2 p - \frac{\partial^2 \rho}{\partial t^2} = -\frac{1}{\rho_0 c_0^2}\left(\zeta + \frac{4}{3}\mu\right)\nabla^2 \frac{\partial p}{\partial t} - \nabla^2 L - \frac{1}{\rho_0 c_0^4}\frac{\partial^2 p^2}{\partial t^2} - \frac{1}{c_0^2}\frac{\partial^2 L}{\partial t^2} \tag{A.41}$$

$$\begin{aligned} \nabla^2 p - \frac{1}{c_0^2}\frac{\partial^2 p}{\partial t^2} + \frac{1}{\rho_0 c_0^4}\frac{B}{2A}\frac{\partial^2 p^2}{\partial t^2} + \frac{\kappa}{\rho_0 c_0^4}\left(\frac{1}{c_v} - \frac{1}{c_p}\right)\frac{\partial^3 p}{\partial t^3} = \\ -\frac{1}{\rho_0 c_0^2}\left(\zeta + \frac{4}{3}\mu\right)\nabla^2 \frac{\partial p}{\partial t} - \nabla^2 L - \frac{1}{\rho_0 c_0^4}\frac{\partial^2 p^2}{\partial t^2} - \frac{1}{c_0^2}\frac{\partial^2 L}{\partial t^2} \end{aligned} \tag{A.42}$$

$$\begin{aligned} \nabla^2 p - \frac{1}{c_0^2}\frac{\partial^2 p}{\partial t^2} + \left(\frac{\kappa}{\rho_0 c_0^4}\left(\frac{1}{c_v} - \frac{1}{c_p}\right) + \frac{1}{\rho_0 c_0^2}\left(\zeta + \frac{4}{3}\mu\right)\right)\frac{\partial^3 p}{\partial t^3} = \\ -\frac{\beta}{\rho_0 c_0^4}\frac{\partial^2 p^2}{\partial t^2} - \left(\frac{1}{c_0^2}\frac{\partial^2 L}{\partial t^2} + \nabla^2 L\right). \end{aligned} \tag{A.43}$$

The last equation can be clarified by introducing the sound diffusivity δ defined as

$$\delta = \frac{1}{\rho_0}\left(\frac{4}{3}\mu + \zeta\right) + \frac{\kappa}{\rho_0}\left(\frac{1}{c_v} - \frac{1}{c_p}\right) = 2\delta_{cl}, \tag{A.44}$$

where δ_{cl} characterizes the dissipation in the acoustic mode. We then obtain

$$\nabla^2 p - \frac{1}{c_0^2}\frac{\partial^2 p}{\partial t^2} + \frac{\delta}{c_0^4}\frac{\partial^3 p}{\partial t^3} = -\frac{\beta}{\rho_0 c_0^4}\frac{\partial^2 p}{\partial t^2} - \left(\nabla^2 + \frac{1}{c_0^2}\frac{\partial^2}{\partial t^2}\right)L. \tag{A.45}$$

This equation was first published by [235] and is used to derive the mainly used nonlinear acoustic equations.

A.2.3 Kuznetsov's equation

Equation (A.45) also has an alternate form know as the Kuznetsov's equation. It is obtained by using the velocity potential ϕ defined by

$$\boldsymbol{v} = -\nabla\phi. \tag{A.46}$$

The Lagrangian density (A.23) can be rewritten using the velocity potential,

$$L = \frac{\rho_0}{2}\left((\nabla\phi)^2 - \frac{1}{c_0^2}\left(\frac{\partial\phi}{\partial t}\right)^2\right), \tag{A.47}$$

which can be further developed using the trick

$$\frac{\partial^2 \phi^2}{\partial t^2} = \frac{\partial}{\partial t}\left(2\phi\frac{\partial\phi}{\partial t}\right) = \left(\frac{\partial\phi}{\partial t}\right)^2 + 2\phi\frac{\partial^2\phi}{\partial t^2}, \tag{A.48}$$

$$L = \frac{\rho_0}{4}\left(\nabla^2\phi^2 - \frac{1}{c_0^2}\frac{\partial^2\phi}{\partial t^2}\right) - \frac{1}{2}\rho_0\phi\left(\nabla^2\phi^2 - \frac{1}{c_0^2}\frac{\partial^2\phi}{\partial t^2}\right). \tag{A.49}$$

Meanwhile, the first order equations (9.22) and (9.25) thus become

$$-\rho_0 \frac{\partial \nabla \phi}{\partial t} + \nabla p = 0 \tag{A.50}$$

$$\nabla \left(-\rho_0 \frac{\partial \phi}{\partial t} + p \right) = 0 \tag{A.51}$$

Which after integrating in time gives

$$p = \rho_0 \frac{\partial \phi}{\partial t}, \tag{A.52}$$

$$\nabla^2 \phi^2 - \frac{1}{c_0^2} \frac{\partial^2 \phi}{\partial t^2} = 0. \tag{A.53}$$

These first order equation thus lead to the approximation

$$L \approx \frac{\rho_0}{4} \left(\nabla^2 \phi^2 - \frac{1}{c_0^2} \frac{\partial^2 \phi}{\partial t^2} \right). \tag{A.54}$$

By inserting the state equation (A.38) into the mass conservation (A.22), the following equation is obtained

$$\frac{\partial p}{\partial t} + \rho_0 c_0^2 \nabla \cdot \boldsymbol{v} - \frac{\kappa}{\rho_0 c_0^2} \left(\frac{1}{c_v} - \frac{1}{c_p} \right) \frac{\partial^2 p}{\partial t^2} = \frac{\beta}{\rho c_0^2} \frac{\partial p^2}{\partial t} + \frac{\partial L}{\partial t}, \tag{A.55}$$

in which the the velocity potential can be used

$$\frac{\partial p}{\partial t} - \rho_0 c_0^2 \nabla^2 \phi - \frac{\kappa}{c_0^2} \left(\frac{1}{c_v} - \frac{1}{c_p} \right) \frac{\partial^3 \phi}{\partial t^3} = \frac{\rho \beta}{c_0^2} \frac{\partial}{\partial t} \left(\frac{\partial \phi}{\partial t} \right)^2 + \frac{\partial L}{\partial t}, \tag{A.56}$$

$$\frac{\partial \nabla p}{\partial t} - \rho_0 c_0^2 \nabla (\nabla^2 \phi) - \frac{\kappa}{c_0^2} \left(\frac{1}{c_v} - \frac{1}{c_p} \right) \nabla \frac{\partial^3 \phi}{\partial t^3} = \frac{\rho \beta}{c_0^2} \nabla \frac{\partial}{\partial t} \left(\frac{\partial \phi}{\partial t} \right)^2 + \nabla \frac{\partial L}{\partial t}. \tag{A.57}$$

Here ∇p can be obtained by using (A.30), rewritten using the velocity potential ϕ,

$$\nabla p = \rho_0 \frac{\partial \nabla \phi}{\partial t} - (\zeta + \frac{4}{3}\mu) \nabla^2 (\nabla \phi) - \nabla L. \tag{A.58}$$

Thus

$$\begin{aligned} \rho_0 \nabla \left(\frac{\partial^2 \phi}{\partial t^2} - c_0^2 \nabla^2 \phi - \frac{1}{\rho_0 c_0^2} \left(\zeta + \frac{4}{3}\mu + \frac{\kappa}{c_v} - \frac{\kappa}{c_p} \right) \frac{\partial^3 \phi}{\partial t^3} \right) \\ = \rho_0 \nabla \left(\frac{\beta}{c_0^2} \frac{\partial}{\partial t} \left(\frac{\partial \phi}{\partial t} \right)^2 + \frac{2}{\rho_0} \frac{\partial L}{\partial t} \right). \end{aligned} \tag{A.59}$$

By using (A.47), the right term can then be simplified since

$$\frac{\beta}{c_0^2}\frac{\partial}{\partial t}\left(\frac{\partial\phi}{\partial t}\right)^2+\frac{2}{\rho_0}\frac{\partial L}{\partial t}=\frac{1}{c_0^2}\frac{\partial}{\partial t}\left(\beta\left(\frac{\partial\phi}{\partial t}\right)^2+\frac{2}{\rho_0}\frac{\rho_0}{2}\left((\nabla\phi)^2-\left(\frac{\partial\phi}{\partial t}\right)^2\right)\right) \tag{A.60}$$

$$=\frac{1}{c_0^2}\frac{\partial}{\partial t}\left((\beta-1)\left(\frac{\partial\phi}{\partial t}\right)^2+c_0^2(\nabla\phi)^2\right) \tag{A.61}$$

$$=\frac{1}{c_0^2}\frac{\partial}{\partial t}\left(\frac{B}{2A}\left(\frac{\partial\phi}{\partial t}\right)^2+c_0^2(\nabla\phi)^2\right). \tag{A.62}$$

With this simplification and the integration of the terms, the equation becomes

$$\frac{\partial^2\phi}{\partial t^2}-c_0^2\nabla^2\phi-\frac{\delta}{c_0^2}\frac{\partial^3\phi}{\partial t^3}=\frac{\partial}{\partial t}\left(\frac{B}{2A}\frac{1}{c_0^2}\left(\frac{\partial\phi}{\partial t}\right)^2+(\nabla\phi)^2\right), \tag{A.63}$$

which is know as Kuznetsov's equation.

A.2.4 Westervelt's equation

By assuming the $L=0$, which is true for plane waves for which $p=\rho_0c_0v$, (A.45) reduces to

$$\nabla^2p-\frac{1}{c_0^2}\frac{\partial^2p}{\partial t^2}+\frac{\delta}{c_0^4}\frac{\partial^3p}{\partial t^3}=-\frac{\beta}{\rho_0c_0^4}\frac{\partial^2p}{\partial t^2}, \tag{A.64}$$

which is know as Westervelt's equation

A.2.5 Burgers' equation

Burgers' equation describes a nonlinear thermo-viscous fluid wave in 1D [236]. In 1D, the condition for Westervelt's equation are valid. In this situation, the Westervelt's equation becomes

$$\frac{\partial^2p}{\partial z^2}-\frac{1}{c_0^2}\frac{\partial^2p}{\partial t^2}+\frac{\delta}{c_0^4}\frac{\partial^3p}{\partial t^3}=-\frac{\beta}{\rho_0c_0^4}\frac{\partial^2p}{\partial t^2}. \tag{A.65}$$

Let us introduce two variables: $\tau=t-\frac{z}{c_0}$ the retarded time, and $z'=\epsilon z$ the slow space scale. Then

$$\frac{\partial}{\partial z}=\frac{\partial z'}{\partial z}\frac{\partial}{\partial z'}+\frac{\partial\tau}{\partial z}\frac{\partial}{\partial\tau}=\epsilon\frac{\partial}{\partial z'}-\frac{1}{c_0}\frac{\partial}{\partial\tau}, \tag{A.66}$$

$$\frac{\partial^2}{\partial z^2}=\left(\epsilon\frac{\partial}{\partial z'}-\frac{1}{c_0}\frac{\partial}{\partial\tau}\right)^2=\epsilon^2\frac{\partial^2}{\partial z'^2}-\frac{2\epsilon}{c_0}\frac{\partial^2}{\partial\tau\partial z'}+\frac{1}{c_0^2}\frac{\partial^2}{\partial\tau^2}, \tag{A.67}$$

$$\frac{\partial}{\partial t}=\frac{\partial z'}{\partial t}\frac{\partial}{\partial z'}+\frac{\partial\tau}{\partial t}\frac{\partial}{\partial\tau}=\frac{\partial}{\partial\tau}. \tag{A.68}$$

By applying this change of variables and only keeping the terms in $\wr(\epsilon)$,

$$\epsilon^2 \frac{\partial^2 p}{\partial z1^2} - \frac{2\epsilon}{c_0} \frac{\partial^2 p}{\partial \tau \partial z'} + \frac{1}{c_0^2} \frac{\partial^2 p}{\partial \tau^2} - \frac{1}{c_0^2} \frac{\partial^2 p}{\partial \tau^2} + \frac{\delta}{c_0^4} \frac{\partial^3 p}{\partial \tau^3} = -\frac{\beta}{\rho_0 c_0^4} \frac{\partial^2 p}{\partial \tau^2}, \tag{A.69}$$

$$-\frac{2\epsilon}{c_0} \frac{\partial^2 p}{\partial \tau \partial z'} + \frac{\delta}{c_0^4} \frac{\partial^3 p}{\partial \tau^3} = -\frac{\beta}{\rho_0 c_0^4} \frac{\partial^2 p}{\partial \tau^2}, \tag{A.70}$$

which can be integrated with respect to τ,

$$\epsilon \frac{\partial p}{\partial z'} - \frac{\beta}{2\rho_0 c_0^3} p \frac{\partial p}{\partial \tau} = \frac{\delta}{2c_0^3} \frac{\partial^2 p}{\partial \tau^2}. \tag{A.71}$$

Finally, by expressing the equation using z instead of the slow scale z'

$$\frac{\partial p}{\partial z} = \frac{\beta}{2\rho_0 c_0^3} p \frac{\partial p}{\partial \tau} + \frac{\delta}{2c_0^3} \frac{\partial^2 p}{\partial \tau^2}. \tag{A.72}$$

The right part of the equation is made of two terms. The first one represents the attenuation, while the second one manages the nonlinearities.

For a source pressure $p(0,t) = p_0 F(t)$, the Burgers' equation has a general solution. It is obtained by using the Hopf-Cole change of variables [237, 238],

$$p = \frac{\rho_0 \delta}{\beta} \frac{\partial \ln(\zeta)}{\partial \tau}, \tag{A.73}$$

where ζ is not the bulk viscosity but a dimensionless variable. The Burgers' equation thus reduces to

$$\frac{\partial \zeta}{\partial z} = \frac{\delta}{2c_0^2} \frac{\partial^2 \zeta}{\partial \tau^2}, \tag{A.74}$$

which is a diffusion equation, for which the solution is known. We have

$$\begin{cases} \zeta(z,\tau) = \sqrt{\frac{c_0^3}{2\pi z \delta}} \int_{-\infty}^{\infty} \zeta(0,\tau') e^{-\frac{c_0^3(\tau-\tau')^2}{2z\delta}} d\tau', \\ \zeta(0,\tau') = e^{\frac{\beta p_0}{\rho_0 \delta} \int_{-\infty}^{\tau'} F(\tau'') d\tau''} \end{cases}, \tag{A.75}$$

$$\begin{cases} p(z,\tau) = \frac{\rho_0 c_0^3}{\beta z} \frac{\int_{-\infty}^{\infty} \zeta(0,\tau')(\tau-\tau') e^{-\frac{c_0^3(\tau-\tau')^2}{2z\delta}} d\tau'}{\int_{-\infty}^{\infty} \zeta(0,\tau') e^{-\frac{c_0^3(\tau-\tau')^2}{2z\delta}} d\tau'}, \\ p(0,\tau') = p_0 F(t). \end{cases} \tag{A.76}$$

An alternative version of the Burgers' equation exists that models other mechanisms of attenuation [239]. Some examples are given in Table A.2 using the generalized Burgers' equation,

$$\frac{\partial p}{\partial z} = \frac{\beta}{2\rho_0 c_0^3} p \frac{\partial p}{\partial \tau} + \nu(p). \tag{A.77}$$

$\nu(p)$	Attenuation mechanism
$\alpha\frac{\partial^2 p}{\partial\tau^2}$	Nonlinear attenuation
$\gamma\frac{\partial^3 p}{\partial\tau^3}$	Pure dispersion
$\frac{1}{2c_0}\left(\frac{c_\infty^2}{c_0^2}-1\right)\frac{\partial}{\partial\tau}\int_{-\infty}^{\tau} e^{-\frac{\tau-\tau'}{\tau_R}}\frac{\partial p(z,\tau')}{\partial\tau'}d\tau'$	Relaxation
$-b\sqrt{\frac{2}{\pi}}\int_{-\infty}^{\tau}\frac{\partial p(z,\tau')}{\partial\tau'}\frac{d\tau'}{\sqrt{\tau-\tau'}}$	Rigid cylinder
$\alpha\frac{\partial^\gamma}{\partial\tau^\gamma}\int_{-\infty}^{\tau}\frac{\partial p(z,\tau')}{\partial\tau'}\frac{d\tau'}{(\tau-\tau')^\beta}$	viscoelastic medium

Table A.2: Various attenuation mechanisms in the generalized Burgers' equation

A.2.6 KZK's equation

KZK's equation is an extension of the Burgers' equation in 3D [235, 240]. It is meant to take into account the diffraction effect from a directional sound beam, in addition to the nonlinearity and absorption. Here the hypothesis is made that the sound beam direction is the z-axis, and as for the Burgers' equation the variables τ and z' are defined, in addition to the variables $x' = \sqrt{\epsilon}x$ and $y' = \sqrt{\epsilon}y$. The Laplacian ∇^2 thus becomes

$$\nabla^2 = \epsilon\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}\right)+\epsilon^2\frac{\partial^2}{\partial z'^2}-\frac{2\epsilon}{c_0}\frac{\partial^2}{\partial\tau\partial z'}+\frac{1}{c_0^2}\frac{\partial^2}{\partial\tau^2}, \tag{A.78}$$

changing the Westervelt's equation into

$$\epsilon\left(\frac{\partial^2 p}{\partial x^2}+\frac{\partial^2 p}{\partial y^2}\right)+\epsilon^2\frac{\partial^2 p}{\partial z'^2}-\frac{2\epsilon}{c_0}\frac{\partial^2 p}{\partial\tau\partial z'}+\frac{1}{c_0^2}\frac{\partial^2 p}{\partial\tau^2}-\frac{1}{c_0^2}\frac{\partial^2 p}{\partial\tau^2}+\frac{\delta}{c_0^4}\frac{\partial^3 p}{\partial\tau^3}=-\frac{\beta}{\rho_0 c_0^4}\frac{\partial^2 p}{\partial\tau^2}, \tag{A.79}$$

$$\epsilon\left(\frac{\partial^2 p}{\partial x^2}+\frac{\partial^2 p}{\partial y^2}\right)-\frac{2\epsilon}{c_0}\frac{\partial^2 p}{\partial\tau\partial z'}+\frac{\delta}{c_0^4}\frac{\partial^3 p}{\partial\tau^3}=-\frac{\beta}{\rho_0 c_0^4}\frac{\partial^2 p}{\partial\tau^2}. \tag{A.80}$$

By noting $\nabla_\perp = \begin{pmatrix}\frac{\partial}{\partial z}\\ \frac{\partial}{\partial z}\\ 0\end{pmatrix}$, integrating in τ, and going back to (x, y, z, τ),

$$\frac{\partial p}{\partial z\partial} = \frac{c_0}{2}\nabla_\perp^2 p+\frac{\beta}{2\rho_0 c_0^3}\frac{\partial p}{\partial\tau}+\frac{\delta}{2c_0^3}\frac{\partial^2 p}{\partial\tau^2}. \tag{A.81}$$

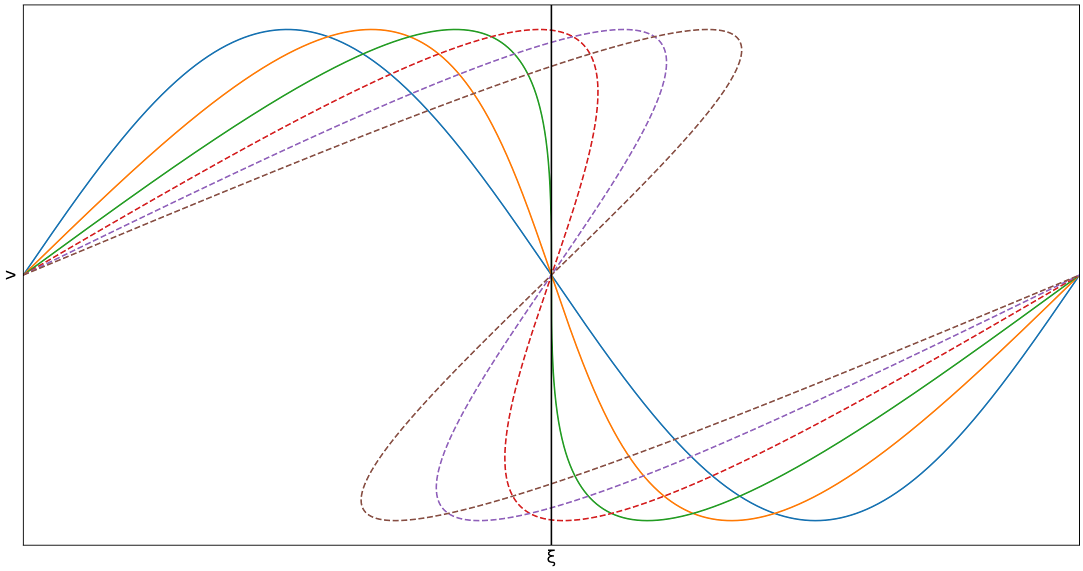


Figure A.1: Evolution of a sinusoidal wave

As for the Burgers' equation, the two right terms representing nonlinearity and absorption are present, with the addition of a third new term $\frac{c_0}{2}\nabla_\perp^2 p$ representing the diffraction.

A.3 Shock waves

[241] [242]

$$\frac{\partial v}{\partial t} + (v + c)\frac{\partial v}{\partial z} = 0 \tag{A.82}$$

$$v = f\left(t - \frac{z}{c + v}\right) \tag{A.83}$$

$$v = f\left(t - \frac{z}{c_0 + \beta v}\right) \tag{A.84}$$

In Figure A.1 we show how a sinusoidal wave evolves and develops a shock.

$$\bar{z} = \frac{(c_0 + \beta v)^2}{\beta \frac{df}{dz}} \tag{A.85}$$

For a sinusoidal wave source, $v = v_0 \sin(\omega t)$ at $z = 0$, the implicit solution is

$$v = v_0 \sin\left(\omega\left(t - \frac{z}{c_0 + \beta v}\right)\right), \tag{A.86}$$

which can be rewritten using the retarded time τ as

$$v = v_0 \sin\left(\omega\left(\tau + \frac{z}{c_0}\frac{\frac{\beta v}{c_0}}{1+\frac{\beta v}{c_0}}\right)\right). \tag{A.87}$$

If $\beta\frac{v}{c_0} \ll 1$ then, by defining $\sigma = \frac{\omega z v_0 \beta}{c_0^2} = \frac{\omega z p_0 \beta}{\rho_0 c_0^3} = \frac{z}{\bar{z}}$,

$$v = v_0 \sin\left(\omega\left(\tau + \frac{z\beta v}{c_0^2}\right)\right) = v_0 \sin\left(\omega\tau + \sigma\frac{v}{v_0}\right), \tag{A.88}$$

$$p = p_0 \sin\left(\omega\tau + \sigma\frac{p}{p_0}\right), \tag{A.89}$$

$$\bar{z} = \frac{c_0^2}{\omega v_0 \beta}. \tag{A.90}$$

In [243] there is an explicit solution using a Fourier series,

$$p = p_0 \sum_{n=1}^{\infty} \frac{2J_n(n\sigma)}{n\sigma} \sin(n\omega\tau), \tag{A.91}$$

where J_n are the Bessel functions of the first kind. This solution is only valid until the shock distance $\bar{z}$ is reached. For small $\sigma \ll 1$, the pressure p can be considered to only have the two first harmonics p_1 and p_2,

$$\begin{cases} p_1 = p_0\frac{2J_1(\sigma)}{\sigma}\sin(\omega\tau) = p_0\sin(\omega\tau), \\ p_2 = p_0\frac{2J_2(2\sigma)}{2\sigma}\sin(2\omega\tau) = z\frac{\omega p_0^2\beta}{\rho_0 c_0^3}\sin(2\omega\tau). \end{cases} \tag{A.92}$$

This shows that the second harmonic is linearly proportionally to the distance and the nonlinear coefficient β, while it has a quadratic relation with the source pressure.

In a medium with loss, the Burgers' equation needs to be used. With the same source $p(0,t) = p_0 \sin(\omega t)$, the solution obtained using (A.76) is

$$p(\sigma,\tau) = p_0 \frac{4\Gamma^{-1}\sum_{n=1}^{\infty}(-1)^{n+1} I_n\left(\frac{\Gamma}{2}\right) e^{-n^2\frac{\sigma}{\Gamma}} \sin(n\omega\tau)}{I_0\left(\frac{\Gamma}{2}\right) + 2\sum_{n=1}^{\infty}(-1)^{n+1} I_n\left(\frac{\Gamma}{2}\right) e^{-n^2\frac{\sigma}{\Gamma}} \cos(n\omega\tau)}, \tag{A.93}$$

where $I_n(x) = \mathrm{i}^{-n} J_n(\mathrm{i}x)$ are the modified Bessel functions, $\Gamma = \frac{\beta\epsilon\omega}{c_0\alpha} = \frac{l_a}{\bar{z}}$ is the Gol'dberg number, $\alpha = \frac{\delta\omega^2}{c_o^3} = \frac{1}{l_a}$ is the damping constant for linear wave and l_a is the attenuation length. The Gol'dberg number measures the importance of the nonlinearity compared to the damping. A large $\Gamma > 1$ means that the nonlinearity effect will prevail, whereas a small $\Gamma < 1$ means that the damping prevents the nonlinearity effect from appearing.

A.4 Bubbly liquids

Bubbles in liquid will increase the nonlinear property of a material. Compared to the liquid, bubbles are made of compressible gas, and will also oscillate following the nonlinear equation [244]. This interaction between bubble and nonlinear acoustics leads to many effects, with some of them, such as acoustic cavitation [245, 246], having useful applications in the industry. Other notable effects are ultrasound self focusing [247], sound amplification [248, 249], sub harmonic generation [250], or difference frequency generation [251, 252].

Appendix B

Adjoint Inversion with Fréchet Kernels

For the full wavefield inversions, in the SPECFEM framework [222], we use a continuous adjoint approach that we formulate, following [253, 254, 255, 256], in terms of Fréchet kernels. The continuous adjoint, when it is feasible to obtain, is far less complex to program and to maintain than a discrete adjoint generated by automatic or semi-automatic differentiators, such as Tapenade [213] or YAO [214], for example.

The kernels are obtained as follows. Suppose that $\boldsymbol{u}(\boldsymbol{x}, t)$ is a displacement (or pressure) field that is related to a set of model parameters, $\boldsymbol{m}(\boldsymbol{x})$, and a force field, $\boldsymbol{f}(\boldsymbol{x}, t)$, by a wave equation (such as (12.15))

$$L(\boldsymbol{u}, \boldsymbol{m}) = \boldsymbol{f}(\boldsymbol{x}, t),$$

which is completed by initial and boundary conditions. Suppose that we have a cost function in the form of a time integral,

$$E(u) = \int_{t_0}^{t_1} e(\boldsymbol{u}(\boldsymbol{x}^r, t))\, \mathrm{d}t,$$

where the receiver positions are given by $\boldsymbol{x} = \boldsymbol{x}^r$ and e is a mismatch term describing the (squared) error between the model solution and the physical measurements or observations. Additional terms can be added to E to provide regularization of the inverse problem. The Fréchet kernel (or sensitivity kernel) is the volumetric density of E with respect to the model parameters $\boldsymbol{m}$. In its most general form, it is given by

$$\delta_m E \doteq \nabla_m E \, \delta \boldsymbol{m} = \int_T \boldsymbol{u}^\dagger \cdot \partial_m L(\boldsymbol{u}, \boldsymbol{m})\, \mathrm{d}t,$$

where the subscript m denotes variation or partial differentiation with respect to each of the model parameters, $T = [t_0, t_1]$ is the time interval and $\boldsymbol{u}^\dagger$ is the adjoint field, solution of the adjoint (wave) equation,

$$\rho \ddot{\boldsymbol{u}}^\dagger - \nabla \cdot \boldsymbol{\sigma} = -\nabla_u E, \tag{B.1}$$

subject to zero terminal conditions and adjoint boundary conditions. This equation is to be solved backwards in time. In the case of (12.12), $\boldsymbol{m} = (\rho, \lambda, \mu)$, and the kernels with respect to the individual parameters ρ, μ and λ can be easily shown to be

$$\begin{aligned}
\delta_\rho E &= -\int_T \partial_t \boldsymbol{u}^\dagger \cdot \partial_t \boldsymbol{u} \, \mathrm{d}t, \\
\delta_\mu E &= \int_T \left(\nabla \cdot \boldsymbol{u}^\dagger\right) : (\nabla \cdot \boldsymbol{u}) + \left(\nabla \cdot \boldsymbol{u}^\dagger\right) : (\nabla \cdot \boldsymbol{u})^T \, \mathrm{d}t, \qquad \text{(B.2)} \\
\delta_\lambda E &= \int_T \left(\nabla \cdot \boldsymbol{u}^\dagger\right) (\nabla \cdot \boldsymbol{u}) \, \mathrm{d}t.
\end{aligned}$$

The kernels with respect to the wave speeds c_p and c_s can be computed from the above three expressions. The code SPECFEM provides these kernels.

The solution of the inverse problem for finding the model parameters from knowledge of the measurements at the receivers, consists of using these kernels in a minimization of the cost function, usually accomplished by a variant of a gradient method,

$$m^{(k+1)} = m^{(k)} - \alpha_k \nabla_m E(m^{(k)}).$$

Here k represents the iteration number and α_k the (optimal) step in the minimisation direction (such as steepest descent). We must still choose a suitable cost function, which might include regularization terms, and construct an optimization loop in order to compute the minimum. Here one customarily uses a quasi-Newton algorithm, but there is some interest in investigating full Newton methods, though these require Hessians that can be computed by taking the second variation of the cost function, following the same steps as above.

References

[1] B. Møhl, M. Wahlberg, P. T. Madsen, A. Heerfordt, and A. Lund, "The monopulsed nature of sperm whale clicks," *The Journal of the Acoustical Society of America*, vol. 114, no. 2, pp. 1143–1154, 2003.

[2] K. S. Norris and G. W. Harvey, "A theory for the function of the spermaceti organ of the sperm whale (physeter catodon l.)," 1972.

[3] P. Dudley, "Ii. an essay upon the natural history of whales, with a particular account of the ambergris found in the sperma ceti whale. in a letter to the publisher, from the honourable paul dudley, esq; fr s," *Philosophical Transactions of the Royal Society of London*, vol. 33, no. 387, pp. 256–269, 1724.

[4] A. Starbuck, "History of the american whale fishery fromits earliest inception to the year 1876," *US Com-mission of Fish and Fisheries, Part IV, Report of the Com-mission for*, vol. 76, 1875.

[5] J. N. Tønnessen and A. O. Johnsen, *The history of modern whaling.* Univ of California Press, 1982.

[6] H. Whitehead, "Estimates of the current global population size and historical trajectory for sperm whales," *Marine Ecology Progress Series*, vol. 242, pp. 295–304, 2002.

[7] R. M. Rolland, S. E. Parks, K. E. Hunt, M. Castellote, P. J. Corkeron, D. P. Nowacek, S. K. Wasser, and S. D. Kraus, "Evidence that ship noise increases stress in right whales," *Proceedings of the Royal Society B: Biological Sciences*, vol. 279, no. 1737, pp. 2363–2368, 2012.

[8] A. Farcas, P. M. Thompson, and N. D. Merchant, "Underwater noise modelling for environmental impact assessment," *Environmental Impact Assessment Review*, vol. 57, pp. 114–122, 2016.

[9] J. Gordon and A. Moscrop, "Underwater noise pollution and its significance for whales and dolphins," *The conservation of whales and dolphins: science and practice. John Wiley & Sons, Chichester, United Kingdom*, pp. 281–319, 1996.

[10] P. Madsen, B. Møhl, B. Nielsen, and M. Wahlberg, "Male sperm whale behaviour during exposures to distant seismic survey pulses," *Aquatic mammals*, vol. 28, no. 3, pp. 231–240, 2002.

[11] B. R. Mate, K. M. Stafford, and D. K. Ljungblad, “A change in sperm whale (physeter macroephalus) distribution correlated to seismic surveys in the gulf of mexico,” *The Journal of the Acoustical Society of America*, vol. 96, no. 5, pp. 3268–3269, 1994.

[12] R. Bassila, T. Bertet, L. Somaglino, M.-E. Bouhier, H. Glotin, P. Best, V. Baron, C. Noel, P. Cristini, F. Fayet, *et al.*, “Abysound, an end to end system for noise impact measurement of deep sea mining production tools,” in *OCEANS 2019-Marseille*, IEEE, 2019, pp. 1–9.

[13] N. D. Merchant, R. C. Faulkner, and R. Martinez, “Marine noise budgets in practice,” *Conservation Letters*, vol. 11, no. 3, e12420, 2018.

[14] T. J. O’SHEA, “Environmental contaminants and marine mammals,” *Biology of marine mammals*, pp. 485–563, 1999.

[15] D. Viale, N. Verneau, and Y. Tison, “Stomach obstruction in a sperm whale beached on the lavezzi islands: Macropollution in the mediterranean.,” *Journal de recherche oceanographique. Paris*, vol. 16, no. 3, pp. 100–102, 1992.

[16] L. David, S. Alleaume, and C. Guinet, “Evaluation of the potential of collision between fin whales and maritime traffic in the north-western mediterranean sea in summer, and mitigation solutions,” *Journal of Marine Animals and Their Ecology Vol*, vol. 4, no. 1, 2011.

[17] A. Gannier and G. Marty, “Sperm whales ability to avoid approaching vessels is affected by sound reception in stratified waters,” *Marine Pollution Bulletin*, vol. 95, no. 1, pp. 283–288, 2015.

[18] S. D. Kraus, M. W. Brown, H. Caswell, C. W. Clark, M. Fujiwara, P. K. Hamilton, R. D. Kenney, A. R. Knowlton, S. Landry, C. A. Mayo, *et al.*, “North atlantic right whales in crisis,” *Science*, vol. 309, no. 5734, pp. 561–562, 2005.

[19] S. Panigada, G. Pesante, M. Zanardelli, F. Capoulade, A. Gannier, and M. T. Weinrich, “Mediterranean fin whales at risk from fatal ship strikes,” *Marine Pollution Bulletin*, vol. 52, no. 10, pp. 1287–1298, 2006.

[20] V. Trygonis, E. Gerstein, J. Moir, and S. McCulloch, “Vocalization characteristics of north atlantic right whale surface active groups in the calving habitat, southeastern united states,” *The Journal of the Acoustical Society of America*, vol. 134, no. 6, pp. 4518–4531, 2013.

[21] A. Mallawaarachchi, S. Ong, M. Chitre, and E. Taylor, “Spectrogram denoising and automated extraction of the fundamental frequency variation of dolphin whistles,” *The Journal of the Acoustical Society of America*, vol. 124, no. 2, pp. 1159–1170, 2008.

[22] O. Mac Aodha, R. Gibb, K. E. Barlow, E. Browning, M. Firman, R. Freeman, B. Harder, L. Kinsey, G. R. Mead, S. E. Newson, *et al.*, “Bat detective—deep learning tools for bat acoustic signal detection,” *PLoS computational biology*, vol. 14, no. 3, e1005995, 2018.

[23] W. Luo, W. Yang, and Y. Zhang, “Convolutional neural network for detecting odontocete echolocation clicks,” *The Journal of the Acoustical Society of America*, vol. 145, no. 1, EL7–EL12, 2019.

[24] W. Yang, W. Luo, and Y. Zhang, “Classification of odontocete echolocation clicks using convolutional neural network,” *The Journal of the Acoustical Society of America*, vol. 147, no. 1, pp. 49–55, 2020.

[25] M. Leavey and J. Sweeney, “The influence of long-range transport of air pollutants on summer visibility at dublin,” *International Journal of Climatology*, vol. 10, no. 2, pp. 191–201, 1990.

[26] J. A. S. Kinney, S. M. Luria, and D. O. Weitzman, “Visibility of colors underwater,” *JOSA*, vol. 57, no. 6, pp. 802–809, 1967.

[27] C. Levenson, “Source level and bistatic target strength of the sperm whale (physeter catodon) measured from an oceanographic aircraft,” *The Journal of the Acoustical Society of America*, vol. 55, no. 5, pp. 1100–1103, 1974.

[28] G. S. Schorr, E. A. Falcone, D. J. Moretti, and R. D. Andrews, “First long-term behavioral records from cuvier’ s beaked whales (ziphius cavirostris) reveal record-breaking dives,” *PloS one*, vol. 9, no. 3, 2014.

[29] E. C. Kirk, “Visual influences on primate encephalization,” *Journal of Human Evolution*, vol. 51, no. 1, pp. 76–90, 2006.

[30] D. R. Ketten, “The marine mammal ear: Specializations for aquatic audition and echolocation,” in *The evolutionary biology of hearing*, Springer, 1992, pp. 717–750.

[31] D. Wartzok and D. R. Ketten, “Marine mammal sensory systems,” *Biology of marine mammals*, vol. 1, p. 117, 1999.

[32] G. Fleischer, *Evolutionary principles of the mammalian middle ear*. Springer Science & Business Media, 2013, vol. 55.

[33] S. Hemilä, S. Nummela, and T. Reuter, “A model of the odontocete middle ear,” *Hearing Research*, vol. 133, no. 1-2, pp. 82–97, 1999.

[34] W. W. Au, P. W. Moore, and D. A. Pawloski, “Detection of complex echoes in noise by an echolocating dolphin,” *The Journal of the Acoustical Society of America*, vol. 83, no. 2, pp. 662–668, 1988.

[35] N. A. Dubrovskiy, “On the two auditory subsystems in dolphins,” in *Sensory abilities of cetaceans*, Springer, 1990, pp. 233–254.

[36] P. Moore, R. Hall, W. Friedl, and P. Nachtigall, “The critical interval in dolphin echolocation: What is it?” *The Journal of the Acoustical Society of America*, vol. 76, no. 1, pp. 314–317, 1984.

[37] A. Y. Supin and V. Popov, “Recovery cycles of the dolphin’ s brain stem responses to paired acoustic stimuli,” in *Dokl. Akad. Nauk SSSR*, vol. 283, 1985, pp. 740–743.

[38] V. Popov and A. Y. Supin, “Auditory brain stem responses in characterization of dolphin hearing,” *Journal of Comparative Physiology A*, vol. 166, no. 3, pp. 385–393, 1990.

[39] V. Popov and A. Supin, “Electrophysiological studies of hearing in some cetaceans and a manatee,” in *Sensory abilities of cetaceans*, Springer, 1990, pp. 405–415.

[40] A. Y. Supin and V. V. Popov, “Temporal resolution in the dolphin' s auditory system revealed by double-click evoked potential study,” *The Journal of the Acoustical Society of America*, vol. 97, no. 4, pp. 2586–2593, 1995.

[41] A. Y. Supin and V. Popov, “Frequency tuning and temporal resolution in dolphins,” *Sensory systems of aquatic mammals*, pp. 95–110, 1995.

[42] V. Popov, A. Y. Supin, and V. Klishin, “Paradoxical lateral suppression in the dolphin's auditory system: Weak sounds suppress response to strong sounds,” *Neuroscience letters*, vol. 234, no. 1, pp. 51–54, 1997.

[43] I. Glezer, P. Hof, P. Morgane, A. Fridman, T. Isakova, D. Joseph, A. Nair, P. Parhar, A. Thengampallil, S. Thomas, *et al.*, “Chemical neuroanatomy of the inferior colliculus in brains of echolocating and nonecholocating mammals: Immunocytochemical study,” *Echolocation in Bats and Dolphins. University of Chicago Press, Chicago, Il*, pp. 161–172, 2004.

[44] S. Gero, D. Engelhaupt, L. Rendell, and H. Whitehead, “Who cares? between-group variation in alloparental caregiving in sperm whales,” *Behavioral Ecology*, vol. 20, no. 4, pp. 838–843, 2009.

[45] F. Sarano, *Le retour de Moby Dick: ou ce que les cachalots nous enseignent sur les océans et les hommes.* Éditions Actes Sud, 2017.

[46] H. Whitehead, R. Antunes, S. Gero, S. N. Wong, D. Engelhaupt, and L. Rendell, “Multilevel societies of female sperm whales (physeter macrocephalus) in the atlantic and pacific: Why are they so different?” *International Journal of Primatology*, vol. 33, no. 5, pp. 1142–1164, 2012.

[47] T. Lyrholm, O. Leimar, B. Johanneson, and U. Gyllensten, “Sex–biased dispersal in sperm whales: Contrasting mitochondrial and nuclear genetic structure of global populations,” *Proceedings of the Royal Society of London. Series B: Biological Sciences*, vol. 266, no. 1417, pp. 347–354, 1999.

[48] H. Whitehead, *Sperm whales: social evolution in the ocean.* University of Chicago press, 2003.

[49] R. H. Backus and W. E. Schevill, “Physeter clicks,” *Whales, dolphins, and porpoises*, pp. 510–527, 1966.

[50] W. A. Watkins and W. E. Schevill, “Sperm whale codas,” *The Journal of the Acoustical Society of America*, vol. 62, no. 6, pp. 1485–1490, 1977.

[51] L. E. Rendell and H. Whitehead, “Vocal clans in sperm whales (physeter macrocephalus),” *Proceedings of the Royal Society of London. Series B: Biological Sciences*, vol. 270, no. 1512, pp. 225–231, 2003.

[52] M. Cantor, L. G. Shoemaker, R. B. Cabral, C. O. Flores, M. Varga, and H. Whitehead, “Multilevel animal societies can emerge from cultural transmission,” *Nature communications*, vol. 6, no. 1, pp. 1–10, 2015.

[53] G. Pouchet and H. Beauregard, *Recherches sur le cachalot.* 1889.

[54] G. Neuweiler, “Auditory adaptations for prey capture in echolocating bats,” *Physiological Reviews*, vol. 70, no. 3, pp. 615–641, 1990. eprint: `http://physrev.physiology.org/content/70/3/615.full.pdf`. [Online]. Available: `http://physrev.physiology.org/content/70/3/615`.

[55] W. W. L. Au, J. L. Pawloski, P. E. Nachtigall, M. Blonz, and R. C. Gisner, “Echolocation signals and transmission beam pattern of a false killer whale (pseudorca crassidens),” *The Journal of the Acoustical Society of America*, vol. 98, no. 1, pp. 51–59, 1995. DOI: `10.1121/1.413643`. eprint: `http://dx.doi.org/10.1121/1.413643`. [Online]. Available: `http://dx.doi.org/10.1121/1.413643`.

[56] W. W. Au, J. Thomas, C. Moss, and M. Vater, “A comparason of the sonar capabilities of bats and dolphins,” pp. xiii–xxvii, 2004.

[57] R. J. Urick, *Principles of Underwater Sound*, 3rd Edition. New York, NY (USA) McGraw-Hill Book, 1983.

[58] S. A. Kick and J. A. Simmons, “Automatic gain control in the bat’s sonar receiver and the neuroethology of echolocation,” *Journal of Neuroscience*, vol. 4, no. 11, pp. 2725–2737, 1984.

[59] R. Kastelein, W. Au, H. Rippe, and N. Schooneman, “Target detection by an echolocating harbor porpoise (phocoena phocoena),” *The Journal of the Acoustical Society of America*, vol. 105, no. 4, pp. 2493–2498, 1999.

[60] W. Au, *The Sonar of Dolphins.* Springer-Verlag, 1993, ISBN: 9780387978352. [Online]. Available: `https://books.google.fr/books?id=yg0XAQAAIAAJ`.

[61] N. Troest and B. Møhl, “The detection of phantom targets in noise by serotine bats; negative evidence for the coherent receiver,” *J. Comp. Physiol. A*, vol. 159, no. 4, pp. 559–567, Oct. 1986.

[62] B. Møhl, “Detection by a pipistrelle bat of normal and reversed replica of its sonar pulses,” *Acta Acustica united with Acustica*, vol. 61, no. 1, pp. 75–82, 1986.

[63] A. E. Murchison, “Detection range and range resolution of echolocating bottlenose porpoise (tursiops truncatus),” in *Animal sonar systems*, Springer, 1980, pp. 43–70.

[64] J. A. Simmons, “The resolution of target range by echolocating bats,” *The Journal of the Acoustical Society of America*, vol. 54, no. 1, pp. 157–173, 1973.

[65] E. Š. Ajrapet’ânc and A. I. Konstantinov, *Eholokaciâ v prirode: Echolocation in nature.* Nauka, 1974.

[66] J. Simmons, W. Lavender, B. Lavender, C. Doroshow, S. Kiefer, R. Livingston, A. Scallet, and D. Crowley, “Target structure and echo spectral discrimination by echolocating bats,” *Science*, vol. 186, no. 4169, pp. 1130–1132, 1974.

[67] W. W. Au and D. A. Pawloski, "Cylinder wall thickness difference discrimination by an echolocating atlantic bottlenose dolphin," *Journal of Comparative Physiology A*, vol. 170, no. 1, pp. 41–47, 1992.

[68] B. D. Lawrence and J. A. Simmons, "Measurements of atmospheric attenuation at ultrasonic frequencies and the significance for echolocation by bats," *The Journal of the Acoustical Society of America*, vol. 71, no. 3, pp. 585–590, 1982.

[69] J. Simmons, S. Kick, B. Lawrence, C. Hale, C. Bard, and B. Escudie, "Acuity of horizontal angle discrimination by the echolocating bat, eptesicus fuscus," *Journal of comparative physiology*, vol. 153, no. 3, pp. 321–330, 1983.

[70] D. L. Renaud and A. N. Popper, "Sound localization by the bottlenose porpoise tursiops truncatus," *Journal of Experimental Biology*, vol. 63, no. 3, pp. 569–585, 1975.

[71] C. S. Johnson, "Relation between absolute threshold and duration-of-tone pulses in the bottlenosed porpoise," *The Journal of the Acoustical Society of America*, vol. 43, no. 4, pp. 757–763, 1968. DOI: `10.1121/1.1910893`. eprint: `http://dx.doi.org/10.1121/1.1910893`. [Online]. Available: `http://dx.doi.org/10.1121/1.1910893`.

[72] A. Surlykke and O. Bojesen, "Integration time for short broad band clicks in echolocating FM-bats (Eptesicus fuscus)," *J. Comp. Physiol. A*, vol. 178, no. 2, pp. 235–241, Feb. 1996.

[73] K. Seger, M. Guerra, and A. Thode, "Laguna san ignacio ecosystem science program acoustic research 2013 winter season report," 2013.

[74] F. Fisher and V. Simmons, "Sound absorption in sea water," *The Journal of the Acoustical Society of America*, vol. 62, no. 3, pp. 558–564, 1977.

[75] R. Francois and G. Garrison, "Sound absorption based on ocean measurements: Part i: Pure water and magnesium sulfate contributions," *The Journal of the Acoustical Society of America*, vol. 72, no. 3, pp. 896–907, 1982.

[76] ——, "Sound absorption based on ocean measurements. part ii: Boric acid contribution and equation for total absorption," *The Journal of the Acoustical Society of America*, vol. 72, no. 6, pp. 1879–1890, 1982.

[77] G. Schuller, K. Beuter, and H. .-.-U. Schnitzler, "Response to frequency shifted artificial echoes in the batrhinolophus ferrumequinum," *Journal of comparative physiology*, vol. 89, no. 3, pp. 275–286, Sep. 1974, ISSN: 1432-1351. DOI: `10.1007/BF00696191`. [Online]. Available: `https://doi.org/10.1007/BF00696191`.

[78] H.-U. Schnitzler and W. H. O' Dell, *Performance of Airborne Animal Sonar Systems: I. Microchiroptera*. Jan. 1980, pp. 109–181, ISBN: 978-1-4684-7256-1.

[79] E. Gould, "Studies of maternal-infant communication and development of vocalizations in the bats myotis and eptesicus," *Communications in Behavioral Biology*, vol. 5, pp. 263–313, 1971.

[80] S. Matsumura, "Mother-infant communication in a horseshoe bat (rhinolophus ferrumequinum nippon): Development of vocalization," *Journal of Mammalogy*, vol. 60, no. 1, pp. 76–84, 1979.

[81] C. F. Moss, "Ontogeny of vocal signals in the big brown bat, eptesicus fuscus," in *Animal sonar*, Springer, 1988, pp. 115–120.

[82] C. F. Moss, D. Redish, C. Gounden, and T. H. Kunz, "Ontogeny of vocal signals in the little brown bat, Myotis lucifugus," *Anim Behav*, vol. 54, no. 1, pp. 131–141, Jul. 1997.

[83] T. W. Cranford, W. G. V. Bonn, M. S. Chaplin, J. A. Carr, T. A. Kamolnick, D. A. Carder, and S. H. Ridgway, "Visualizing dolphin sonar signal generation using high-speed video endoscopy," *The Journal of the Acoustical Society of America*, vol. 102, no. 5, pp. 3123–3123, 1997. DOI: `10.1121/1.420593`. eprint: `http://dx.doi.org/10.1121/1.420593`. [Online]. Available: `http://dx.doi.org/10.1121/1.420593`.

[84] T. Cranford, W. Elsberry, D. Blackwood, J. A. Carr, T. Kamolnick, M. Todd, W. G. van Bonn, D. A. Carder, S. Ridgway, D. M. Bozlinski, and E. C. Decker, "Two independent sonar signal generators in the bottlenose dolphin: Physiologic evidence and implications," vol. 108, pp. 2613–2614, Nov. 2000.

[85] N. A. Dubrovsky and L. R. Giro, "A plausible mechanism of acoustic click production in the dolphin," *The Journal of the Acoustical Society of America*, vol. 105, no. 2, pp. 1263–1263, 1999. DOI: `10.1121/1.426044`. eprint: `http://dx.doi.org/10.1121/1.426044`. [Online]. Available: `http://dx.doi.org/10.1121/1.426044`.

[86] М. А. Исакович, *Общая акустика: учебное пособие*. Наука. Гл. ред. физ.-мат. лит., 1973.

[87] W. M. Zimmer, P. T. Madsen, V. Teloni, M. P. Johnson, and P. L. Tyack, "Off-axis effects on the multipulse structure of sperm whale usual clicks with implications for sound production," *The Journal of the Acoustical Society of America*, vol. 118, no. 5, pp. 3337–3345, 2005.

[88] P.-T. Madsen, R. Payne, N. Kristiansen, M. Wahlberg, I. Kerr, and B. Møhl, "Sperm whale sound production studied with ultrasound time/depth-recording tags," *J. of Exp. Biology*, vol. 205, no. 13, pp. 1899–1906, 2002.

[89] S. Huggenberger, M. André, and H. H. A. Oelschläger, "An acoustic valve within the nose of sperm whales physeter macrocephalus," *Mammal Review*, vol. 44, no. 2, pp. 81–87, Apr. 2014, ISSN: 1365-2907. DOI: `10.1111/mam.12017`. [Online]. Available: `https://doi.org/10.1111/mam.12017`.

[90] T. W. Cranford, "The sperm whale's nose: Sexual selection on a grand scale?1," *Marine Mammal Science*, vol. 15, no. 4, pp. 1133–1157, Oct. 1999, ISSN: 1748-7692. DOI: `10.1111/j.1748-7692.1999.tb00882.x`. [Online]. Available: `https://doi.org/10.1111/j.1748-7692.1999.tb00882.x`.

[91] V. Teloni, W. M. ZIMMER, M. Wahlberg, and P. T. MADSEN, “127 consistent acoustic size estimation of sperm whales using clicks recorded from unknown aspects,” *J. Cetacean Res. Manage*, vol. 9, no. 2, pp. 127–136, 2007.

[92] F. Caruso, V. Sciacca, G. Bellia, E. De Domenico, G. Larosa, E. Papale, C. Pellegrino, S. Pulvirenti, G. Riccobene, F. Simeone, *et al.*, “Size distribution of sperm whales acoustically identified during long term deep-sea monitoring in the ionian sea,” *PLoS One*, vol. 10, no. 12, 2015.

[93] M. Nishiwaki, S. Ohsumi, and Y. Maeda, “Change of form in the sperm whale accompanied with growth,” *Scientific Reports of the Whales Research Institute, Tokyo*, vol. 17, pp. l–17, 1963.

[94] B. Møhl, E. Larsen, and M. Amundin, “ ‘sperm whale size determination: Outlines of an acoustic approach,” *FAO Fisheries*, vol. 3, no. 5, pp. 327–331, 1981.

[95] J. C. Gordon, “Evaluation of a method for determining the length of sperm whales (physeter catodon) from their vocalizations,” *Journal of Zoology*, vol. 224, no. 2, pp. 301–314, 1991.

[96] M. R. Clarke, “Structure and proportions of the spermaceti organ in the sperm whale,” *Journal of the Marine Biological Association of the United Kingdom*, vol. 58, no. 1, pp. 1–17, 1978.

[97] M. Q. Rhinelander and S. M. Dawson, “Measuring sperm whales from their clicks: Stability of interpulse intervals and validation that they indicate whale length,” *The Journal of the Acoustical Society of America*, vol. 115, no. 4, pp. 1826–1831, 2004.

[98] A. Growcott, B. Miller, P. Sirguey, E. Slooten, and S. Dawson, “Measuring body length of male sperm whales from their clicks: The relationship between interpulse intervals and photogrammetrically measured lengths,” *The Journal of the Acoustical Society of America*, vol. 130, no. 1, pp. 568–573, 2011.

[99] B. P. Bogert, “The quefrency alanysis of time series for echoes; cepstrum, pseudoautocovariance, cross-cepstrum and saphe cracking,” *Time series analysis*, pp. 209–243, 1963.

[100] R. Antunes, L. Rendell, and J. Gordon, “Measuring inter-pulse intervals in sperm whale clicks: Consistency of automatic estimation methods,” *The Journal of the Acoustical Society of America*, vol. 127, no. 5, pp. 3239–3247, 2010.

[101] R. Abeille, Y. Doh, P. Giraudet, H. Glotin, J.-M. Prevot, and 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,” *Journal of the Scientific Reports of Port-Cros National Park*, vol. 28, 2014.

[102] B. Møhl, “Sound transmission in the nose of the sperm whale physeter catodon. a post mortem study,” *Journal of Comparative Physiology A*, vol. 187, no. 5, pp. 335–340, 2001.

[103] A. Goodson, J. Flint, and T. Cranford, "The harbor porpoise (phocoena phocoena)—modeling the sonar transmission mechanism," *Echolocation in Bats and Dolphins (ed. JA Thomas, CF Moss and M. Vater)*, pp. 64–85, 2004.

[104] Z. Song, Y. Zhang, S. W. Thornton, S. Li, and J. Dong, "The influence of air-filled structures on wave propagation and beam formation of a pygmy sperm whale (kogia breviceps) in horizontal and vertical planes," *The Journal of the Acoustical Society of America*, vol. 142, no. 4, pp. 2443–2453, 2017.

[105] W. A. Watkins, "Acoustics and the behavior of sperm whales," in *Animal sonar systems*, Springer, 1980, pp. 283–290.

[106] B. Møhl, M. Wahlberg, P. T. Madsen, L. A. Miller, and A. Surlykke, "Sperm whale clicks: Directionality and source level revisited," *The journal of the Acoustical Society of America*, vol. 107, no. 1, pp. 638–648, 2000.

[107] W. M. Zimmer, P. L. Tyack, M. P. Johnson, and P. T. Madsen, "Three-dimensional beam pattern of regular sperm whale clicks confirms bent-horn hypothesis," *The Journal of the Acoustical Society of America*, vol. 117, no. 3, pp. 1473–1485, 2005.

[108] E.-M. Nosal and L. N. Frazer, "Sperm whale three-dimensional track, swim orientation, beam pattern, and click levels observed on bottom-mounted hydrophones," *The Journal of the Acoustical Society of America*, vol. 122, no. 4, pp. 1969–1978, 2007.

[109] V. Kandia and Y. Stylianou, "Detection of sperm whale clicks based on the Teager–Kaiser energy operator," *Applied Acoustics*, vol. 67, pp. 1144–1163, 2006.

[110] H. Glotin, F. Caudal, and P. Giraudet, "Whale cocktail party: Real-time multiple tracking and signal analyses," *Canadian acoustics*, vol. 36, no. 1, pp. 139–145, 2008.

[111] F. Pukelsheim, "The three sigma rule," *The American Statistician*, vol. 48, no. 2, pp. 88–91, 1994.

[112] P. R. Roth, "Effective measurements using digital signal analysis," *IEEE spectrum*, vol. 8, no. 4, pp. 62–70, 1971.

[113] G. C. Carter, A. H. Nuttall, and P. G. Cable, "The smoothed coherence transform," *Proceedings of the IEEE*, vol. 61, no. 10, pp. 1497–1498, 1973.

[114] C. Knapp and G. Carter, "The generalized correlation method for estimation of time delay," *IEEE transactions on acoustics, speech, and signal processing*, vol. 24, no. 4, pp. 320–327, 1976.

[115] W. J. Bangs and P. M. Schultheiss, "Space-time processing for optimal parameter estimation," in *Signal processing*, Academic San Diego, Calif., 1973, pp. 577–590.

[116] G. C. Carter, "Variance bounds for passively locating an acoustic source with a symmetric line array," *The Journal of the Acoustical Society of America*, vol. 62, no. 4, pp. 922–926, 1977.

[117] W. Hahn and S. Tretter, "Optimum processing for delay-vector estimation in passive signal arrays," *IEEE Transactions on Information Theory*, vol. 19, no. 5, pp. 608–614, 1973.

[118] J. DiBiase, H. Silverman, and M. Brandstein, "Robust localization in reverberant rooms," in *Microphone Arrays*, Springer, 2001, pp. 157–180.

[119] N. Dubrovsky and L. Giro, "Modeling of the click-production mechanism in dolphins," in *Echolocation in Bats and Dolphins.* University of Chicago Press, 2004, p. 68.

[120] R. Abeille, Y. Doh, P. Giraudet, H. Glotin, J.-M. Prevot, and 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," *Journal of the Scientific Reports of Port-Cros National Park*, vol. 28, 2014.

[121] S. Thrun, "Affine structure from sound," in *Proceedings of Conference on Neural Information Processing Systems (NIPS)*, Cambridge, MA: MIT Press, 2005.

[122] Y. Kuang, E. Ask, S. Burgess, and K. Åström, "Understanding toa and tdoa network calibration using far field approximation as initial estimate.," in *ICPRAM (2)*, 2012, pp. 590–596.

[123] H. Glotin, P. Giraudet, J. Ricard, F. Malige, P. Patris, V. Roger, J.-M. Prévot, M. Poupard, O. Philippe, and P. Cosentino, "Projet vamos : Observations de mammifères marins par la bouée stéreophonique bombyx," 53, PELAGOS int. cetacean sanctuary, France, Italie, Monaco Ed. 2017. [Online]. Available: `https://www.sanctuaire-pelagos.org/fr/tous-les-telechargements/etudes-scientifiques-studi-scientifici-studies/etudes-francaises/789-14-037-vamos`.

[124] L. Rendell, S. Simião, J. Brotons, S. Airoldi, D. Fasano, and A. Gannier, "Abundance and movements of sperm whales in the western mediterranean basin," *Aquatic Conservation: Marine and Freshwater Ecosystems*, vol. 24, no. S1, pp. 31–40, 2014.

[125] J. Patris, D. Komatitsch, M. Sepulveda, M. Santos, H. Glotin, F. Malige, S. Buchan, and M. Asch, "Mono-hydrophone localization of baleen whales: A study of propagation using a spectral element method applied in northern chile," in *OCEANS 2019-Marseille*, IEEE, 2019, pp. 1–9.

[126] Y. Simard and N. Roy, "Detection and localization of blue and fin whales from large-aperture autonomous hydrophone arrays: A case study from the St. Lawrence estuary," *Canadian Acoustics*, vol. 36.1, no. 1, pp. 104–110, 2008.

[127] N. Roy, Y. Simard, and C. Gervaise, "3D tracking of foraging belugas from their clicks: Experiment from a coastal hydrophone array," *Applied Acoustics*, vol. 71, no. 11, pp. 1050–1056, 2010.

[128] F. Bénard-Caudal, P. Giraudet, and H. Glotin, "Whale 3D monitoring using astrophysic NEMO ONDE two meters wide platform with state optimal filtering by rao-blackwell monte carlo data association," *App. Acoustics*, vol. 71, no. 11, pp. 994–999, 2010.

[129] P. Favali, F. Chierici, G. Marinaro, G. Giovanetti, *et al.*, "NEMO-SN1 abyssal cabled observatory in the western ionian sea," *J. Oceanic Eng.*, vol. 38, no. 2, pp. 358–374, 2013.

[130] A. Thode, "Tracking sperm whale (p. macrocephalus) dive profiles using a towed passive acoustic array," *J. of the Acoustical Society of America*, vol. 116, no. 1, pp. 245–253, 2004.

[131] W. Zimmer, "Range estimation of cetaceans with compact volumetric arrays," *J. of Acoustical Society of America*, vol. 134, no. 3, pp. 2610–2618, 2013.

[132] M. Ferrari, M. Poupard, P. Giraudet, R. Marxer, J.-M. Prévot, T. Soriano, and H. Glotin, "Efficient artifacts filter by density-based clustering in long term 3d whale passive acoustic monitoring with five hydrophones fixed under an autonomous surface vehicle," in *OCEANS 2019-Marseille*, IEEE, 2019, pp. 1–7.

[133] M. Fourniol, V. Gies, V. Barchasz, E. Kussener, H. Barthelemy, R. Vauché, and H. Glotin, "Low-power wake-up system based on frequency analysis for environmental internet of things," in *Int. Conf. on Mechatronic, Embedded Systems, App.*, IEEE, 2018, pp. 1–6.

[134] W. A. Watkins and W. E. Schevill, "Sound source location by arrival-times on a non-rigid three-dimensional hydrophone array," in *Deep Sea Research and Oceanographic Abstracts*, Elsevier, vol. 19, 1972, pp. 691–706.

[135] J. Macaulay, J. Gordon, D. Gillespie, C. Malinka, and S. Northridge, "Passive acoustic methods for fine-scale tracking of harbour porpoises in tidal rapids," *The Journal of the Acoustical Society of America*, vol. 141, no. 2, pp. 1120–1132, 2017.

[136] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *KDD'96 Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, AAAI Press, 1996, pp. 226–231.

[137] M. Poupard, M. Ferrari, J. Schluter, R. Marxer, P. Giraudet, V. Barchasz, V. Gies, G. Pavan, and H. Glotin, "Real-time passive acoustic 3d tracking of deep diving cetacean by small non-uniform mobile surface antenna," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2019, pp. 8251–8255.

[138] R. Hirotsu, M. Yanagisawa, T. Ura, M. Sakata, H. Sugimatsu, J. Kojima, and R. Bahl, "Localization of sperm whales in a group using clicks received at two separated short baseline arrays," *The Journal of the Acoustical Society of America*, vol. 127, no. 1, pp. 133–147, 2010.

[139] R. Bahl, T. Nakatani, T. Ura, and M. Sakata, “Automatic classification of diving sperm whales by analysis of click time delay using two hydrophones,” in *Oceans’ 04 MTS/IEEE Techno-Ocean’04 (IEEE Cat. No. 04CH37600)*, IEEE, vol. 4, 2004, pp. 2316–2320.

[140] 浦環, 坂田雅雄, 小島淳一, *et al.*, “2 組のハイドロフォンアレイを使ったマッコウクジラの潜水行動の推定,” 生産研究, vol. 56, no. 6, pp. 471–474, 2004.

[141] 浦環, 坂田雅雄, 能勢義昭, 福地鐵雄, 小島淳一, 浦純也, 杉松治美, and 中谷武志, “2 組のハイドロフォンアレイを用いたマッコウクジラの追跡観測実験,” 生産研究, vol. 56, no. 2, pp. 157–160, 2004.

[142] H. Glotin, N. Thellier, P. Best, M. Poupard, M. Ferrari, S. Viera, G. Donzé, M. Campana, J. Chevallier, P. Giraudet, F. Malige, J. Patris, J. Prévot, V. Giés, N. Prévot d’ Alvise, V. Barchasz, S. Marzetti, F. Sarano, S. Gaillard, and F. de Varenne, “Sphyrna-odyssey : Découvertes de chasses abyssales de cachalots en alliance et des effets du confinement covid19,” 220, CNRS LIS Univ Toulon Ed. 2020. [Online]. Available: https://sabiod.univ-tln.fr/pub/SPHYRNA/SO_Premier_Rapport.pdf.

[143] W. A. Watkins and W. E. Schevill, “Spatial distribution of physeter catodon (sperm whales) underwater,” *Deep Sea Research*, vol. 24, no. 7, pp. 693–699, 1977.

[144] L. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, “Encoder-decoder with atrous separable convolution for semantic image segmentation,” *CoRR*, vol. abs/1802.02611, 2018. arXiv: 1802.02611. [Online]. Available: http://arxiv.org/abs/1802.02611.

[145] M. Bianco, P. Gerstoft, J. Traer, E. Ozanich, M. A. Roch, S. Gannot, and C.-A. Deledalle, “Machine learning in acoustics: Theory and applications,” *The Journal of the Acoustical Society of America*, vol. 146, no. 5, pp. 3590–3628, 2019.

[146] H. Glotin, J. Hildebrand, K. Dunleavy, and M. Roch, “Dclde challenge,” 134, Sorbonne Université, Université de Toulon Ed. 2018. [Online]. Available: http://sabiod.univ-tln.fr/DCLDE/files/DCLDE2018_OfficialProgram.pdf.

[147] Y. Kim, J. Yim, J. Yun, and J. Kim, “Nlnl: Negative learning for noisy labels,” in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, pp. 101–110.

[148] E. Fonseca, M. Plakal, D. P. Ellis, F. Font, X. Favory, and X. Serra, “Learning sound event classifiers from web audio with noisy labels,” in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2019, pp. 21–25.

[149] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, “Backpropagation applied to handwritten zip code recognition,” *Neural computation*, vol. 1, no. 4, pp. 541–551, 1989.

[150] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “Imagenet: A large-scale hierarchical image database,” in *2009 IEEE conference on computer vision and pattern recognition*, Ieee, 2009, pp. 248–255.

[151] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[152] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[153] T. Grill and J. Schlüter, “Two convolutional neural networks for bird detection in audio signals,” in *2017 25th European Signal Processing Conference (EUSIPCO)*, IEEE, 2017, pp. 1764–1768.

[154] A. J. Fairbrass, M. Firman, C. Williams, G. J. Brostow, H. Titheridge, and K. E. Jones, “Citynet—deep learning tools for urban ecoacoustic assessment,” *Methods in Ecology and Evolution*, vol. 10, no. 2, pp. 186–197, 2019.

[155] M. Poupard, P. Best, J. Schlüter, H. Symonds, P. Spong, and H. Glotin, “Large-scale unsupervised clustering of orca vocalizations: A model for describing orca communication systems,” PeerJ Preprints, Tech. Rep., 2019.

[156] K. Fukumori, H. T. T. Nguyen, N. Yoshida, and T. Tanaka, “Fully data-driven convolutional filters with deep learning models for epileptic spike detection,” in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2019, pp. 2772–2776.

[157] S. Kiranyaz, T. Ince, O. Abdeljaber, O. Avci, and M. Gabbouj, “1-d convolutional neural networks for signal processing applications,” in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2019, pp. 8360–8364.

[158] Y. LeCun, Y. Bengio, *et al.*, “Convolutional networks for images, speech, and time series,” *The handbook of brain theory and neural networks*, vol. 3361, no. 10, p. 1995, 1995.

[159] J. Jacobsen, E. Oyallon, S. Mallat, and A. W. Smeulders, “Multiscale hierarchical convolutional networks,” *arXiv preprint arXiv:1703.04140*, 2017.

[160] J. J. Huang and J. J. A. Leanos, “Aclnet: Efficient end-to-end audio classification cnn,” *arXiv preprint arXiv:1811.06669*, 2018.

[161] D. Amodei, S. Ananthanarayanan, R. Anubhai, J. Bai, E. Battenberg, C. Case, J. Casper, B. Catanzaro, Q. Cheng, G. Chen, *et al.*, “Deep speech 2: End-to-end speech recognition in english and mandarin,” in *International conference on machine learning*, 2016, pp. 173–182.

[162] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *arXiv preprint arXiv:1412.6980*, 2014.

[163] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, "Mixup: Beyond empirical risk minimization," *arXiv preprint arXiv:1710.09412*, 2017.

[164] S. Thulasidasan, G. Chennupati, J. A. Bilmes, T. Bhattacharya, and S. Michalak, "On mixup training: Improved calibration and predictive uncertainty for deep neural networks," in *Advances in Neural Information Processing Systems*, 2019, pp. 13 888–13 899.

[165] L. Carratino, M. Cissé, R. Jenatton, and J.-P. Vert, "On mixup regularization," *arXiv preprint arXiv:2006.06049*, 2020.

[166] Y. Aytar, C. Vondrick, and A. Torralba, "Soundnet: Learning sound representations from unlabeled video," in *Advances in Neural Information Processing Systems 29*, D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett, Eds., Curran Associates, Inc., 2016, pp. 892–900. [Online]. Available: `http://papers.nips.cc/paper/6146-soundnet-learning-sound-representations-from-unlabeled-video.pdf`.

[167] H. Lee, P. Pham, Y. Largman, and A. Y. Ng, "Unsupervised feature learning for audio classification using convolutional deep belief networks," in *Advances in Neural Information Processing Systems 22*, Y. Bengio, D. Schuurmans, J. D. Lafferty, C. K. I. Williams, and A. Culotta, Eds., Curran Associates, Inc., 2009, pp. 1096–1104. [Online]. Available: `http://papers.nips.cc/paper/3674-unsupervised-feature-learning-for-audio-classification-using-convolutional-deep-belief-networks.pdf`.

[168] S.-W. Fu, Y. Tsao, X. Lu, and H. Kawai, "Raw waveform-based speech enhancement by fully convolutional networks," in *2017 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, IEEE, 2017, pp. 006–012.

[169] X. Feng, Y. Zhang, and J. Glass, "Speech feature denoising and dereverberation via deep autoencoders for noisy reverberant speech recognition," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2014, pp. 1759–1763. DOI: `10.1109/ICASSP.2014.6853900`.

[170] K. Paliwal, K. Wójcicki, and B. Shannon, "The importance of phase in speech enhancement," *Speech Commun.*, vol. 53, no. 4, pp. 465–494, Apr. 2011, ISSN: 0167-6393. DOI: `10.1016/j.specom.2010.12.003`. [Online]. Available: `http://dx.doi.org/10.1016/j.specom.2010.12.003`.

[171] D. Ulyanov, *Multicore-tsne*, `https://github.com/DmitryUlyanov/Multicore-TSNE`, 2016.

[172] G. G. Stokes, "On the theories of the internal friction of fluids in motion, and of the equilibrium and motion of elastic solids," *Transactions of the Cambridge Philosophical Society*, vol. 8, 1880.

[173] J. B. J. Baron Fourier, *The analytical theory of heat*. The University Press, 1878.

[174] A. E. H. Love, *A treatise on the mathematical theory of elasticity*. Cambridge university press, 2013.

[175] R. Hooke, *Lectures de potentia restitutiva, or of spring explaining the power of springing bodies*, 6. John Martyn, 2016.

[176] G. L. Belen' kii, E. Y. Salaev, and R. A. Suleimanov, "Deformation effects in layer crystals," *Usp. Fiz. Nauk*, vol. 155, no. 5, pp. 89–127, 1988. DOI: `10.3367/UFNr.0155.198805c.0089`. [Online]. Available: `https://ufn.ru/ru/articles/1988/5/c/`.

[177] W. Voigt *et al.*, *Lehrbuch der kristallphysik*. Teubner Leipzig, 1928, vol. 962.

[178] R. Aris, *Vectors, tensors and the basic equations of fluid mechanics*. Courier Corporation, 2012.

[179] "Finite difference methods," in *Numerical Treatment of Partial Differential Equations: Translated and revised by Martin Stynes*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 23–124, ISBN: 978-3-540-71584-9. DOI: `10.1007/978-3-540-71584-9_2`. [Online]. Available: `https://doi.org/10.1007/978-3-540-71584-9_2`.

[180] A. T. Patera, "A spectral element method for fluid dynamics: Laminar flow in a channel expansion," *Journal of computational Physics*, vol. 54, no. 3, pp. 468–488, 1984.

[181] E. Milne, "Vi. sound waves in the atmosphere," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 42, no. 247, pp. 96–114, 1921.

[182] J. B. Keller, "Geometrical theory of diffraction," *Josa*, vol. 52, no. 2, pp. 116–130, 1962.

[183] R. Courant, K. Friedrichs, and H. Lewy, "Über die partiellen differenzengleichungen der mathematischen physik," *Mathematische annalen*, vol. 100, no. 1, pp. 32–74, 1928.

[184] G. M. Amdahl, "Validity of the single processor approach to achieving large scale computing capabilities," in *Proceedings of the April 18-20, 1967, spring joint computer conference*, 1967, pp. 483–485.

[185] J. L. Aroyan, T. W. Cranford, J. Kent, and K. S. Norris, "Computer modeling of acoustic beam formation in d elphinusdelphis," *The Journal of the Acoustical Society of America*, vol. 92, no. 5, pp. 2539–2545, 1992.

[186] T. W. Cranford, P. Krysl, and J. A. Hildebrand, "Acoustic pathways revealed: Simulated sound transmission and reception in cuvier's beaked whale (ziphius cavirostris)," *Bioinspiration & Biomimetics*, vol. 3, no. 1, p. 016 001, 2008.

[187] C. Wei, Y. Zhang, and W. W. Au, "Simulation of ultrasound beam formation of baiji (lipotes vexillifer) with a finite element model," *The Journal of the Acoustical Society of America*, vol. 136, no. 1, pp. 423–429, 2014.

[188] T. W. Cranford, V. Trijoulet, C. R. Smith, and P. Krysl, "Validation of a vibroacoustic finite element model using bottlenose dolphin simulations: The dolphin biosonar beam is focused in stages," *Bioacoustics*, vol. 23, no. 2, pp. 161–194, 2014.

[189] M. S. Soldevilla, M. F. McKenna, S. M. Wiggins, R. E. Shadwick, T. W. Cranford, and J. A. Hildebrand, "Cuvier's beaked whale (ziphius cavirostris) head tissues: Physical properties and ct imaging," *Journal of experimental biology*, vol. 208, no. 12, pp. 2319–2332, 2005.

[190] S. Huggenberger, M. Andre, and H. H. Oelschläger, "The nose of the sperm whale: Overviews of functional design, structural homologies and evolution," *Journal of the Marine Biological Association of the United Kingdom*, vol. 96, no. 4, pp. 783–806, 2016.

[191] M. R. Clarke, "The head of the sperm whale," *Scientific American*, vol. 240, no. 1, pp. 128–141, 1979.

[192] O. Panagiotopoulou, P. Spyridis, H. M. Abraha, D. R. Carrier, and T. C. Pataky, "Architecture of the sperm whale forehead facilitates ramming combat," *PeerJ*, vol. 4, e1895, 2016.

[193] J. C. Goold, J. D. Bennell, and S. E. Jones, "Sound velocity measurements in spermaceti oil under the combined influences of temperature and pressure," *Deep Sea Research Part I: Oceanographic Research Papers*, vol. 43, no. 7, pp. 961–969, 1996.

[194] Z. Song, X. Xu, J. Dong, L. Xing, M. Zhang, X. Liu, Y. Zhang, S. Li, and P. Berggren, "Acoustic property reconstruction of a pygmy sperm whale (kogia breviceps) forehead based on computed tomography imaging," *The Journal of the Acoustical Society of America*, vol. 138, no. 5, pp. 3129–3137, 2015.

[195] K. Yee, "Numerical solution of initial boundary value problems involving maxwell's equations in isotropic media," *IEEE Transactions on antennas and propagation*, vol. 14, no. 3, pp. 302–307, 1966.

[196] R. Madariaga, "Dynamics of an expanding circular fault," *Bulletin of the Seismological Society of America*, vol. 66, no. 3, pp. 639–666, 1976.

[197] J. Virieux, "P-sv wave propagation in heterogeneous media: Velocity-stress finite-difference method," *Geophysics*, vol. 51, no. 4, pp. 889–901, 1986.

[198] J. L. Aroyan, M. A. McDonald, S. C. Webb, J. A. Hildebrand, D. Clark, J. T. Laitman, and J. S. Reidenberg, "Acoustic models of sound production and propagation," in *Hearing by whales and dolphins*, Springer, 2000, pp. 409–469.

[199] G. Mur, "Absorbing boundary conditions for the finite-difference approximation of the time-domain electromagnetic-field equations," *IEEE transactions on Electromagnetic Compatibility*, no. 4, pp. 377–382, 1981.

[200] R. L. Higdon, “Absorbing boundary conditions for difference approximations to the multidimensional wave equation,” *Mathematics of computation*, vol. 47, no. 176, pp. 437–459, 1986.

[201] H. B. Keller, “Accurate difference methods for linear ordinary differential systems subject to linear constraints,” *SIAM Journal on Numerical Analysis*, vol. 6, no. 1, pp. 8–30, 1969.

[202] M. Lentini and V. Pereyra, “An adaptive finite difference solver for nonlinear two-point boundary problems with mild boundary layers,” *SIAM Journal on Numerical Analysis*, vol. 14, no. 1, pp. 91–111, 1977.

[203] J.-P. Berenger *et al.*, “A perfectly matched layer for the absorption of electromagnetic waves,” *Journal of computational physics*, vol. 114, no. 2, pp. 185–200, 1994.

[204] F. Collino and C. Tsogka, “Application of the perfectly matched absorbing layer model to the linear elastodynamic problem in anisotropic heterogeneous media,” *Geophysics*, vol. 66, no. 1, pp. 294–307, 2001.

[205] W. C. Chew and W. H. Weedon, “A 3d perfectly matched medium from modified maxwell’s equations with stretched coordinates,” *Microwave and optical technology letters*, vol. 7, no. 13, pp. 599–604, 1994.

[206] J. A. Roden and S. D. Gedney, “Convolution pml (cpml): An efficient fdtd implementation of the cfs–pml for arbitrary media,” *Microwave and optical technology letters*, vol. 27, no. 5, pp. 334–339, 2000.

[207] M. Kuzuoglu and R. Mittra, “Frequency dependence of the constitutive parameters of causal perfectly matched anisotropic absorbers,” *IEEE Microwave and Guided wave letters*, vol. 6, no. 12, pp. 447–449, 1996.

[208] D. Komatitsch and R. Martin, “An unsplit convolutional perfectly matched layer improved at grazing incidence for the seismic wave equation,” *Geophysics*, vol. 72, no. 5, SM155–SM167, 2007.

[209] R. J. Luebbers and F. Hunsberger, “Fdtd for nth-order dispersive media,” *IEEE transactions on Antennas and Propagation*, vol. 40, no. 11, pp. 1297–1301, 1992.

[210] M. Toyoda, D. Takahashi, and Y. Kawai, “Averaged material parameters and boundary conditions for the vibroacoustic finite-difference time-domain method with a nonuniform mesh,” *Acoustical Science and Technology*, vol. 33, no. 4, pp. 273–276, 2012.

[211] M. Asch, M. Bocquet, and M. Nodet, *Data Assimilation: Methods, Algorithms, and Applications*. SIAM, 2016.

[212] J. Nocedal and S. J. Wright, *Numerical Optimization*. Spriger, 2006.

[213] L. Hascoët and V. Pascual, “TAPENADE 2.1 user’s guide,” INRIA, Sophia Antipolis, Rapport technique 300, 2004. [Online]. Available: `http://www.inria.fr/rrrt/rt-0300.html`.

[214] L. Nardi, C. Sorror, F. Badran, and S. Thiria, “Lecture notes in computer science,” in *Lecture Notes in Computer Science*, O. Gervasi, D. Taniar, B. Murgante, A. Lagana, Y. Mun, and M. Gavrilova, Eds. Springer Berlin / Heidelberg, 2009, vol. 5593, ch. YAO: A Software for Variational Data Assimilation Using Numerical Models, pp. 621–636. [Online]. Available: `http://dx.doi.org/10.1007/978-3-642-02457-3_53`.

[215] J.-P. Hermand, M. Meyer, M. Asch, and M. Berrada, “Adjoint-based acoustic inversion for the physical characterization of a shallow water environment,” *The Journal of the Acoustical Society of America*, vol. 119, no. 6, pp. 3860–3871, 2006.

[216] A. Quarteroni, F. Saleri, and G. Paola, *Scientific Computing with MATLAB and Octave*, Fourth. Springer-Verlag Berlin Heidelberg, 2014.

[217] A. Griewank and A. Walther, “Revolve: An implementation of checkpointing for the reverse or adjoint mode of computational differentiation.,” *ACM Trans. Math. Software*, vol. 26, pp. 19–45, 2000.

[218] Z. Long, Y. Lu, and B. Dong, “Pde-net 2.0: Learning pdes from data with a numeric-symbolic hybrid deep network,” *Journal of Computational Physics*, vol. 399, p. 108 925, 2019, ISSN: 0021-9991. DOI: `https://doi.org/10.1016/j.jcp.2019.108925`. [Online]. Available: `http://www.sciencedirect.com/science/article/pii/S0021999119306308`.

[219] R. Lguensat, J. Le Sommer, S. Metref, E. Cosme, and R. Fablet, “Learning generalized quasi-geostrophic models using deep neural numerical models,” in *Neurisp workshop on Machine Learning and the Physical Sciences*, Vancouver, Canada, Dec. 2019.

[220] I. Ayed, E. de Bézenac, A. Pajot, J. Brajard, and P. Gallinari, “Learning dynamical systems from partial observations,” *CoRR*, vol. abs/1902.11136, 2019. arXiv: `1902.11136`. [Online]. Available: `http://arxiv.org/abs/1902.11136`.

[221] Y. Fan, C. Orozco Bohorquez, and L. Ying, “Bcr-net: A neural network based on the nonstandard wavelet form,” *Journal of Computational Physics*, vol. 384, pp. 1–15, 2019, ISSN: 0021-9991. DOI: `https://doi.org/10.1016/j.jcp.2019.02.002`. [Online]. Available: `http://www.sciencedirect.com/science/article/pii/S0021999119300762`.

[222] J. Tromp, D. Komatitsch, and Q. Liu, “Spectral-element and adjoint methods in seismology.,” *Communications in Computational Physics*, vol. 3, pp. 1–32, 2008.

[223] Z. Xie, D. Komatitsch, R. Martin, and R. Matzen, “Improved forward wave propagation and adjoint-based sensitivity kernel calculations using a numerically stable finite-element pml,” *Geophysical Journal International*, vol. 198, pp. 1714–1747, 2014.

[224] G. A. Sorge, *Vorgemach der musicalischen Composition, oder: Ausführliche, ordentliche und zur heutigen Praxis hinlänglichen Anweisung zum General-Baß*. 1745, vol. 1.

[225] ——, *Vorgemach der musicalischen Composition, oder: Ausführliche, ordentliche und vor heutige Praxin hinlängliche Anweisung zum General-Baß...: Der dritte Theil von dem Vorgemach der musicalischen Composition: in welchem Der mannichfaltige Gebrauch der Dissonantzen... gezeigt wird...* Selbstverl., 1747, vol. 3.

[226] M. F. Hamilton, D. T. Blackstock, *et al.*, *Nonlinear acoustics.* Academic press San Diego, 1998, vol. 237.

[227] R. W. Ogden, *Non-linear elastic deformations.* Courier Corporation, 1997.

[228] L. Brillouin, "Les tensions de radiation; leur interprétation en mécanique classique et en relativité," 1925.

[229] K. Brugger, "Pure modes for elastic waves in crystals," *Journal of Applied Physics*, vol. 36, no. 3, pp. 759–768, 1965.

[230] M. Breazeale, "Third-order elastic constants of cubic crystals," in. Jan. 2001, pp. 489–510, ISBN: 9780124457607. DOI: `10.1016/B978-012445760-7/50023-X`.

[231] R. T. Beyer, "Parameter of nonlinearity in fluids," *The Journal of the Acoustical Society of America*, vol. 32, no. 6, pp. 719–721, 1960.

[232] L. Cremer, "Über die akustische grenzschicht vor starren wänden," *Archiv der elektrischen Übertragung*, vol. 2, pp. 136–139, 1948.

[233] L. S. Kovasznay, "Turbulence in supersonic flow," *Journal of the Aeronautical Sciences*, vol. 20, no. 10, pp. 657–674, 1953.

[234] G. Kirchhoff, "Ueber den einfluss der wärmeleitung in einem gase auf die schallbewegung," *Annalen der Physik*, vol. 210, no. 6, pp. 177–193, 1868.

[235] S. I. Aanonsen, T. Barkve, J. N. Tjo/tta, and S. Tjo/tta, "Distortion and harmonic generation in the nearfield of a finite amplitude sound beam," *The Journal of the Acoustical Society of America*, vol. 75, no. 3, pp. 749–768, 1984.

[236] D. T. Blackstock, "Generalized burgers equation for plane waves," *The Journal of the Acoustical Society of America*, vol. 77, no. 6, pp. 2050–2053, 1985.

[237] E. Hopf, "The partial differential equation $u_t + uu_x = \mu_{xx}$," *Communications on Pure and Applied mathematics*, vol. 3, no. 3, pp. 201–230, 1950.

[238] J. D. Cole, "On a quasi-linear parabolic equation occurring in aerodynamics," *Quarterly of applied mathematics*, vol. 9, no. 3, pp. 225–236, 1951.

[239] S. Makarov and M. Ochmann, "Nonlinear and thermoviscous phenomena in acoustics, part ii," *Acta Acustica united with Acustica*, vol. 83, no. 2, pp. 197–222, 1997.

[240] Y.-S. Lee and M. F. Hamilton, "Time-domain modeling of pulsed finite-amplitude sound beams," *The Journal of the Acoustical Society of America*, vol. 97, no. 2, pp. 906–917, 1995.

[241] B. Riemann *et al.*, *Über die Fortpflanzung ebener Luftwellen von endlicher Schwingungsweite.* Verlag der Dieterichschen Buchhandlung, 1860.

[242] R. S. Earnshaw and M. Sheffield, "On the mathematical theory of sound," in *Classic Papers in Shock Compression Science*, Springer, 1998, pp. 83–108.

[243] E. F. Ghiron, *Anomalie nella propagazione di onde acustiche di grande ampiezza.* Industrie Grafiche Italiane Stucchi, 1935.

[244] Z. Feng and L. Leal, "Nonlinear bubble dynamics," *Annual review of fluid mechanics*, vol. 29, no. 1, pp. 201–243, 1997.

[245] I. Akhatov, U. Parlitz, and W. Lauterborn, "Pattern formation in acoustic cavitation," *The Journal of the Acoustical Society of America*, vol. 96, no. 6, pp. 3627–3635, 1994.

[246] P. Ciuti, G. Iernetti, and M. Sagoo, "Optical visualization of non-linear acoustic propagation in cavitating liquids," *Ultrasonics*, vol. 18, no. 3, pp. 111–114, 1980.

[247] G. Askar'yan, "Self-focusing of powerful sound during the production of bubbles," *Soviet Journal of Experimental and Theoretical Physics Letters*, vol. 13, p. 283, 1971.

[248] I. S. Akhatov and V. Baikov, "Propagation of sound perturbations in heterogeneous gas-liquid systems," *Journal of engineering physics*, vol. 50, no. 3, pp. 276–280, 1986.

[249] I. S. Akhatov, V. Baikov, and R. Baikov, "Propagation of nonlinear waves in gas-liquid media with a gas content variable in space," *Fluid Dynamics*, vol. 21, no. 1, pp. 161–164, 1986.

[250] E. Zabolotskaya, "Nonlinear waves in liquid with gas bubbles," *Trudi IOFAN*, vol. 18, p. 121, 1989.

[251] O. A. Druzhinin, L. A. Ostrovsky, and A. Prosperetti, "Low-frequency acoustic wave generation in a resonant bubble layer," *The Journal of the Acoustical Society of America*, vol. 100, no. 6, pp. 3570–3580, 1996.

[252] L. A. Ostrovsky, A. M. Sutin, I. A. Soustova, A. I. Matveyev, and A. I. Potapov, "Nonlinear, low-frequency sound generation in a bubble layer: Theory and laboratory experiment," *The Journal of the Acoustical Society of America*, vol. 104, no. 2, pp. 722–726, 1998.

[253] A. Fichtner, H.-P. Bunge, and H. Igel, "The adjoint method in seismology: I. Theory," *Physics of the Earth and Planetary Interiors*, vol. 157, pp. 86–104, 2006, ISSN: 0031-9201. DOI: `10.1016/j.pepi.2006.03.016`. [Online]. Available: `http://www.sciencedirect.com/science/article/pii/S0031920106001051`.

[254] ——, "The adjoint method in seismology II. Applications: Traveltimes and sensitivity functionals," *Physics of the Earth and Planetary Interiors*, vol. 157, pp. 105–123, 2006, ISSN: 0031-9201. DOI: `10.1016/j.pepi.2006.03.018`. [Online]. Available: `http://www.sciencedirect.com/science/article/pii/S0031920106001038`.

[255] A. Fichtner and J. Trampert, “Resolution analysis in full waveform inversion,” *Geophysical Journal International*, vol. 187, no. 3, pp. 1604–1624, 2011, ISSN: 1365-246X. [Online]. Available: `http://dx.doi.org/10.1111/j.1365-246X.2011.05218.x`.

[256] D. Peter, D. Komatitsch, Y. Luo, R. Martin, N. Le Goff, E. Casarotti, P. Le Loher, F. Magnoni, Q. Liu, C. Blitz, T. Nissen-Meyer, P. Basini, and J. Tromp, “Forward and adjoint simulations of seismic wave propagation on fully unstructured hexahedral meshes,” *Geophysical Journal International*, vol. 186, no. 2, pp. 721–739, 2011, ISSN: 1365-246X. DOI: `10.1111/j.1365-246X.2011.05044.x`. [Online]. Available: `http://dx.doi.org/10.1111/j.1365-246X.2011.05044.x`.

Abstract

he sperm whale, *Physeter macrocephalus*, possesses the largest biosonar in nature. Made of multiple oil sacs, the sperm whale sonar is tailored to function from the sea surface down to a depth of 2 kilometers, emitting clicks as loud as 236 dB, and is multipurpose, as it produces clicks for either echolocation or socializing. However, the liquid wax that composes its sonar, made the sperm whales the target of whaling until 1986, when the remaining population was far too small to remain commercially viable, especially with the arrival of similar products from the petrochemical industry. The sperm whale population still faces some human threats, with the ingestion of plastic and collision with boats continuing to take a toll on their numbers. Studying sperm whales thus will have outcomes in multiple fields, in conservation, ethology, as well as in bioacoustics. Understanding the mechanism that governs the sperm whale sonar will help to study these other fields, as it is a key element in the sperm whale life. Aiming for this goal, this thesis analyzes three databases with distinct characteristics, obtaining the trajectory of sperm whale dives. Clicks were also linked with the sperm whale that emitted them over multiple years of recording of the same population. An efficient End-to-End deep learning classifier was trained to classify biosonar waveforms. A simulation of wave propagation through the sperm whale head was also developed to better understand the complex mechanism of this sonar. Finally, a coupling method was developed to improve the parameters of the simulation using the recorded clicks from the aforementioned databases.
Keyword: Sperm whale, Sonar, Simulation

Résumé

Le cachalot, *Physeter macrocephalus*, possède le plus grand biosonar de la nature. Composé de plusieurs poches d'huile, le sonar du cachalot est conçu pour fonctionner de la surface de la mer jusqu'à une profondeur de 2 kilomètres, émettant des clics pouvant aller jusqu'à 236 dB, et est polyvalent, car il produit des clics pour l'écholocation ou la socialisation. Cependant, la cire liquide qui compose le sonar a fait des cachalots la cible de la chasse jusqu'en 1986, lorsque la population restante était beaucoup trop petite pour rester commercialement viable, en particulier avec l'arrivée de produits similaires développés par l'industrie pétrochimique. La population de cachalots est toujours confrontée à certaines menaces humaines, comme l'ingestion de plastique et la collision avec des bateaux qui continuent de faire des ravages sur leur nombres. L'étude des cachalots donne ainsi des résultats dans de multiples domaines, en conservation, en éthologie, ainsi qu'en bioacoustique. Comprendre le mécanisme qui régit le sonar du cachalot aidera à étudier ces autres domaines, car il s'agit d'un élément clé de la vie du cachalot. Dans ce but, cette thèse analyse trois bases de données aux caractéristiques distinctes, obtenant la trajectoire des plongées de cachalots. Les clics enregistrés ont été également reliés au cachalot qui les avait émis, et ce sur plusieurs années d'enregistrement fait sur la même population. Un modèle original End-to-End par Deep Learning est construit pour classer efficacement les formes d'onde de biosonars. Une simulation de propagation des ondes à travers la tête du cachalot a également été développée pour mieux comprendre le mécanisme complexe de ce sonar. Enfin, une méthode de couplage a été développée pour améliorer les paramètres de la simulation en utilisant les clics enregistrés des bases de données précédemment citées. Un résumé de la thèse en français est présent au début du manuscrit.
Mots clé : Cachalot, Sonar, Simulation