MAT 2795: Rapport 1

Pour le 30 avril 2021

Professeur Guy Wolf

Noémie Chenail, Simon-Olivier Laperrière, Shophika Vaithyanathasarma

Table des matières

1 Introduction

L'acoustique marine, soit l'étude des sons marins, existe depuis plus d'un siècle. Son outil principal, l'hydrophone, permet de capter les sons dans l'eau. D'abord créé dans le but d'offrir un moyen de communication entre la côte et les bateaux pour réduire les naufrages contre le rivage, cet outil a grandement évolué depuis [1]. Avec de multiples applications telles que la surveillance de la taille de diverses populations et la caractérisation des sons propres à chaque espèce, l'acoustique marine permet d'offrir une meilleure représentation de l'environnement marin, ce qui permet de passer outre les difficultés propres à la représentation visuelle. Au niveau de l'écologie marine, cette technique permet une meilleure compréhension de l'habitat et des signaux sonores. Par exemple, cela a permis de comprendre la diversité des mécanismes par lesquels sont produits les sons marins allant de l'utilisation d'air pour produire des sons à l'utilisation de muscles spécialisés produisant une vibration de l'exosquelette [2].

Les sons marins ont plusieurs utilités, entre autres au niveau de la défense du territoire, de l'accouplement et de l'organisation sociale. Les sons produits par une espèce peuvent même être utilisés pour en imiter une autre en contexte de relation proie-prédateur. C'est le cas des dauphins qui imitent les sons des orques afin de ne pas se faire détecter [3]. Quant aux orques, ceux-ci encryptent leurs sons afin de ne pas être détectés par leurs proies [2]. Afin de mieux comprendre les relations entre les sons émis par les animaux marins, il convient d'utiliser des méthodes pour la comparaison et la classification des différents sons.

L'apprentissage profond apporte les outils nécessaires à la comparaison de différents échantillons sonores. En effet, dans les dernières années, l'apprentissage profond a permis de faire passer l'évaluation des signaux sonores de techniques de traitement numérique de signal à l'extraction de données issues de spectrogrammes [4]. Un spectrogramme est une représentation visuelle des ondes sonores. Différents algorithmes permettent donc d'extraire des attributs des spectrogrammes, ce qui ouvre la voie à la réduction de dimension. C'est d'ailleurs ce qui a été réalisé dans un projet de Kyle MacDonald et Manny Tam [5] visant à classifier différents sons d'oiseaux et à représenter cette classification de façon interactive. En s'inspirant des lignes directrices de ce projet, le présent projet tentera d'utiliser différents algorithmes de réduction de dimension afin d'offrir une meilleure compréhension des liens unissant les sons de plusieurs espèces marines.

2 Objectifs

Le projet tentera de classifier les enregistrements de différentes espèces marines afin d'établir s'il y a présence de regroupements intra-espèce et inter-espèces. Cela se fera par le biais de trois différentes techniques d'apprentissage de variétés appliquées sur les spectrogrammes issus des enregistrements recueillis. Les données classifiées seront aussi classifiées de façon graphique afin de représenter les résultats. De plus, les résultats obtenus avec les différentes techniques d'apprentissage de variétés seront comparés afin de déterminer la technique la plus adéquate pour regrouper les enregistrements.

3 Descriptions des données analysées

Les enregistrements utilisés proviennent du site internet du New Bedford Marine Museum [6]. Ceux-ci sont proviennent d'une soixantaine d'espèces différentes et totalisent plus de 15000 enregistrements réalisés entre 1940 et 2000 à différents endroits dans le monde par William Watkins et ses collaborateurs. La date et le lieu de l'enregistrement n'ont cependant pas été pris en compte dans nos démarches. L'accès au téléchargement de ces enregistrements est permis à des fins académiques. Ces enregistrements sonores sont des fichiers `.wav`. Nous avons utilisés les enregistrements de la section «'Best of' cuts» qui regroupe les enregistrements de meilleure qualité avec le moins de bruit. La durée de ces enregistrements varie de quelques secondes à plusieurs minutes. De plus, un prétraitement a été effectué.

4 Méthodologie

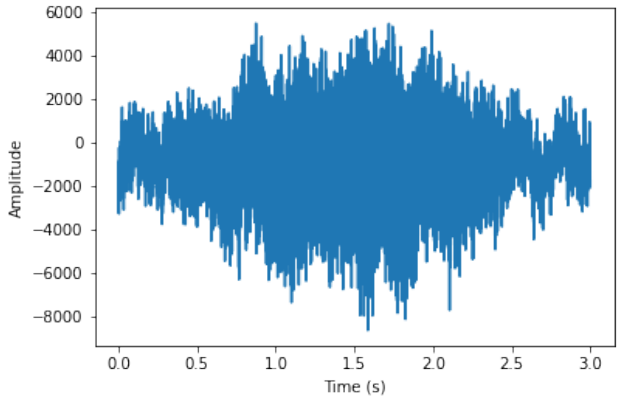
4.1 Pré-traitement audio

Nous avons tout d'abord effectué un nettoyage des données très simple. La première étape était de conserver uniquement les enregistrements dont la durée se situait entre trois et six secondes. Ce choix était plutôt arbritraire, mais nous assurait à la fois que nous aurions des enregistrements assez longs pour être analysés et assez courts pour que les sons produits par les espèces soient capturés dans la quasi-totalité de l'enregistrement. Finalement, nous avons conservé les trois premières secondes de ces enregistrements.

4.2 Transformation des .wav en spectogramme

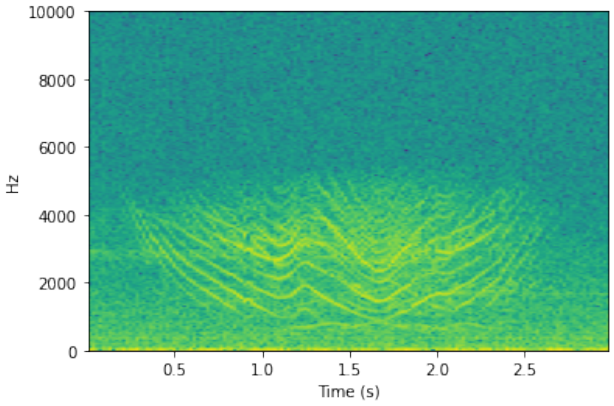
Les enregistrements sonores sélectionnés étant de type `.wav`, il fallait les transformer dans un format qui permettrait aux algorithmes de dévoiler leur structure intrinsèque. Pour ce faire, nous avons eu recours aux spectrogrammes. Afin de bien comprendre leur utilité, une brève explication du son, de sa digitalisation et de sa transformation en spectrogramme s'impose. Tout d'abord, le son est une onde acoustique qui se propage dans un milieu, tel l'air ou l'eau. La fonction correspondante aux variations de l'onde sonore en fonction du temps est ce qu'on appelle communément le signal. Il est possible de tracer le graphique correspondant à la forme d'onde du signal.

FIGURE 1 – Exemple du graphique du signal d'un enregistrement sonore d'un phoque de Ross (*Ommatophoca rossii*)



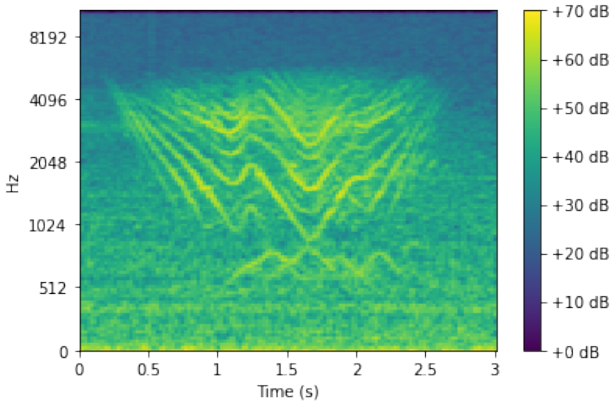
À chaque instant, la valeur du signal représente l'amplitude, qui correspond à l'intensité sonore mesurée. De plus, étant donné que le signal caractérise une onde, il est possible de calculer sa fréquence, qui correspond au nombre de cycles produits à chaque unité de temps. Cependant, cette caractéristique n'est pas explicitement indiquée dans les mesures du signal original, d'où l'intérêt d'utiliser un spectrogramme. Celui-ci est un graphique qui affiche les changements de fréquence d'un enregistrement sonore à travers le temps. L'amplitude est quant à elle représentée par une troisième «dimension», soit l'intensité lumineuse sur le graphique. Avec l'échelle de couleur utilisée dans notre travail, plus l'amplitude est grande, plus la couleur du signal sera pâle. Pour convertir nos enregistrements sonores en spectrogrammes, nous avons eu recours à la librairie `librosa`. Les détails techniques impliqués dans l'analyse spectrale effectuée par `librosa` seront omis. L'essentiel est de savoir qu'une décomposition du son est effectuée en utilisant la transformation de Fourier.

FIGURE 2 – Exemple du spectrogramme produit à partir de l'enregistrement sonore précédent



Les spectrogrammes de base offrent une représentation compacte et visuelle de l'enregistrement sonore considéré. Cependant, ils ne sont pas très représentatifs de notre façon de percevoir les sons. En effet, la capacité de l'ouïe des humains est telle que les sons sont perçus de façon logarithmique et non linéaire. Ainsi, des échelles psychoacoustiques ont été développées afin de représenter fidèlement la perception sonore des humains. Dans le cas de l'intensité sonore (le volume), il s'agit de l'échelle Décibel, tandis que pour la fréquence, il s'agit de l'échelle Mel. Nous avons donc converti nos spectrogrammes d'origine en spectogramme Mel, qui englobe à la fois les échelles décibel et Mel. Cette représentation nous permet d'extraire les caractéristiques essentielles des enregistrements.

FIGURE 3 – Exemple du spectrogramme Mel produit à partir de l'enregistrement sonore précédent. Les échelles Mel pour la fréquence et décibel pour l'intensité sonore sont utilisées.



4.3 Réduction de dimension et visualisation

Afin de visualiser les similarités entre les spectrogrammes générés lors de l'étape précédente, nous avons utilisé plusieurs méthodes de réduction de dimension. De façon générale, les méthodes de réduction de dimension consistent à réduire le nombre d'attributs d'un ensemble de données, tout en préservant le plus d'information possible. En d'autres termes, on cherche à conserver la structure et les propriétés de notre ensemble de départ. Ces méthodes peuvent être divisées en deux catégories : les méthodes de réduction de dimension linéaires et non linéaires. Lors de notre analyse des spectrogrammes, nous avons eu recours aux deux types de méthodes, mais à des fins différentes. Tous les algorithmes utilisés ont été implémentés avec la librairie `scikit-learn` dans le langage de programmation Python.

4.3.1 Utilisation de PCA

L'ensemble de données de départ, soient les sepctrogrammes obtenus après la conversion à partir des enregistrements sonores, contenait $|X| = ...$ entrées de dimension $d = ...$ Ainsi, en vertu du fléau de la dimension, l'utilisation d'algorithmes de visualisation complexes en temps n'aurait pas été pratique, voire tout simplement irréalisable. Il était donc impératif d'utiliser une méthode simple et peu coûteuse en temps. Ceci nous a permis de réduire notre ensemble de données dans un espace dimensionnel plus petit ($d = 50$), puis d'appliquer subséquemment des algorithmes de complexité plus élevée sur ce nouvel ensemble. Pour ce faire, nous avons utilisé l'algorithme PCA (analyse en composantes principales), qui est une méthode de réduction de dimension linéaire. De façon générale, PCA calcule la matrice de covariance de l'ensemble de données standardisées, puis calcule sa décomposition spectrale pour obtenir le plongement de dimension réduite. Géométriquement, ceci correspond à projeter les données sur les vecteurs qui maximisent la variance dans l'ensembles de données.

4.3.2 Utilisation de Isomap

Pour obtenir notre première représentation visuelle dans un espace 2-dimensionnel, nous avons utilisé Isomap. Isomap est une méthode de réduction de dimension non-linéaire, plus spécifiquement un algorithme d'apprentissage de variétés. Le fonctionnement d'Isomap est en quelque sorte très similaire à PCA, mais au lieu d'utiliser une matrice de covariance, Isomap utilise une matrice de dissimilarité. Plus spécifiquement, l'algorithme commence par calculer la matrice des distances

euclidiennes entre les points dans l'espace de départ. À l'aide de celle-ci, il calcule les k-plus-proches voisins de chaque points, où k est un entier déterminé préalablement. Il construit par la suite un graphe dont les noeuds sont les points de l'ensemble et le poids des arêtes sont la distance qui les séparent. Finalement, la matrice de distance géodésique entre chaque point est calculée, en utilisant le plus court chemin entre ceux-ci dans le graphe comme approximation. Le plongement est obtenu à partir de la décomposition spectrale de cette matrice.

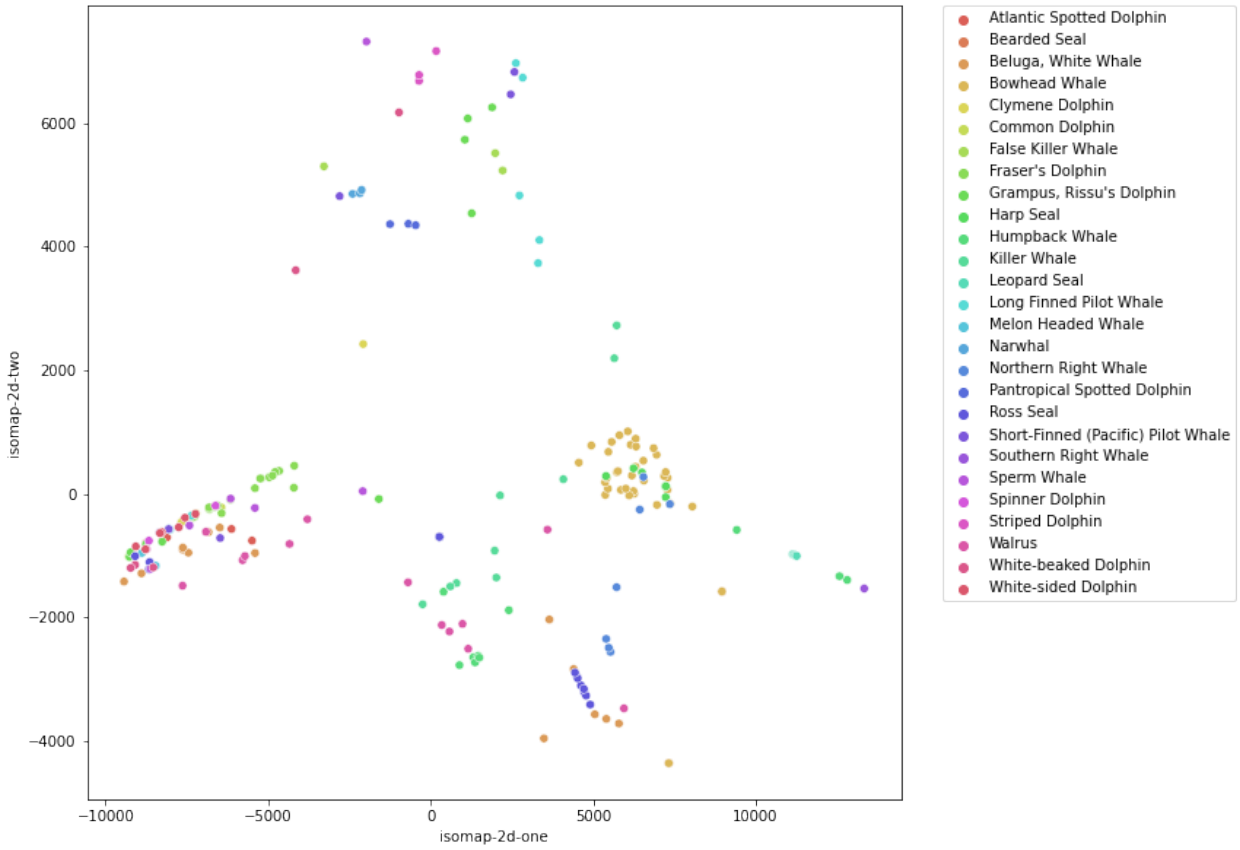
4.3.3 Utilisation de t-SNE

Pour obtenir notre seconde représentation visuelle dans un espace 2-dimensionnel, nous utilisé t-SNE (t-distributed Stochastic Neighborhood Embedding). Il s'agit d'une autre méthode d'apprentissage de variétés, beaucoup plus sophistiquée qu'Isomap. Celle-ci offre d'ailleurs trois avantages par rapport à Isomap [7]. Tout d'abord, elle permet de révéler la structure à plusieurs échelles. Puis, elle permet de déceler la présence de plusieurs regroupements différents. Finalement, elle réduit la tendance à regrouper les points au centre du plongement. Cette méthode, dont le fonctionnement dépasse la portée du cours, est basée sur la minimisation de la divergence de Kullback-Leibler entre la distribution de points dans l'espace de départ et dans l'espace en dimensions réduites [8].

5 Résultats

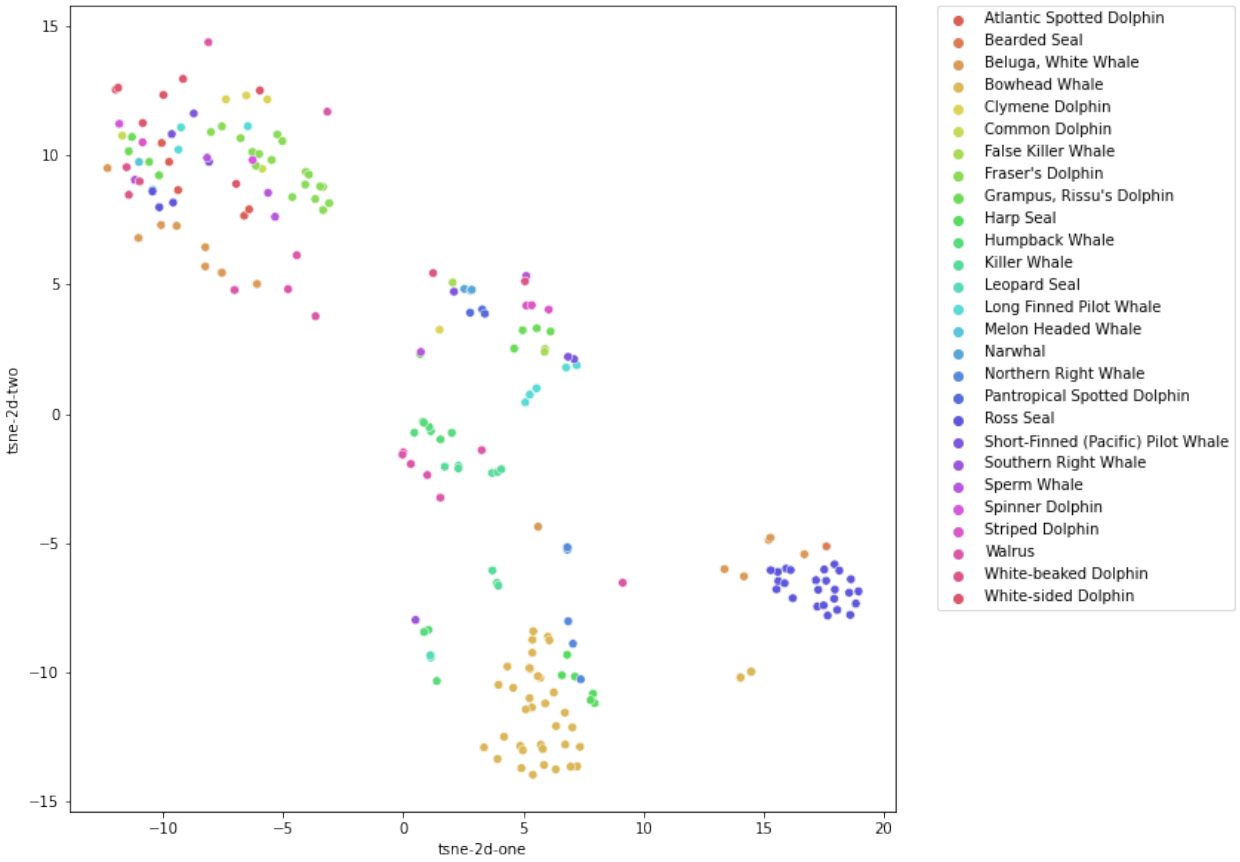
Nous avons tout d'abord tracé le nuage de points 2-dimensionels du plongement des données avec PCA et Isomap (Figure 4). À l'analyse de ce graphique, on remarque quelques rassemblements éparses, sans que la majorité des enregistrements en fasse partie. On peut environ en distinguer deux importants, soient un à gauche regroupant différentes espèces et un au centre droit regroupant les enregistrements de l'espèce des baleines boréales («Bowhead Whales»). La faible présence de regroupements indique qu'Isomap ne constitue pas une bonne façon de représenter les données.

FIGURE 4 – Nuage de points 2D obtenu par l'application successive de PCA et Isomap



À l'analyse rapide du nuage de points 2D obtenus avec PCA et t-SNE (Figure 5), il est possible de remarquer la présence de cinq regroupements de données. Par sa capacité à créer des regroupements, cette méthode est donc plus efficace dans ce cas à celle effectuée précédemment avec Isomap. Il faut toutefois noter que le regroupement constitué de plusieurs espèces en haut à gauche semble aussi être présent dans le nuage de points d'Isomap. Quant aux autres regroupements, deux sont formés d'enregistrements provenant chacun de la même espèce, soient les baleines boréales («Bowhead Whales») et les phoques de Ross («Ross Seal»). Cela n'est malheureusement pas dû à la performance de notre méthode, mais plutôt à la présence des «drop-out» dans ces enregistrements, qui correspond aux bandes noires en haut des spectrogrammes. Pour illuster ceci, une carte visuelle 2D a été produite avec Raster-Fairy pour cette transformation afin de représenter les spectrogrammes regroupés après l'application de PCA et t-SNE.

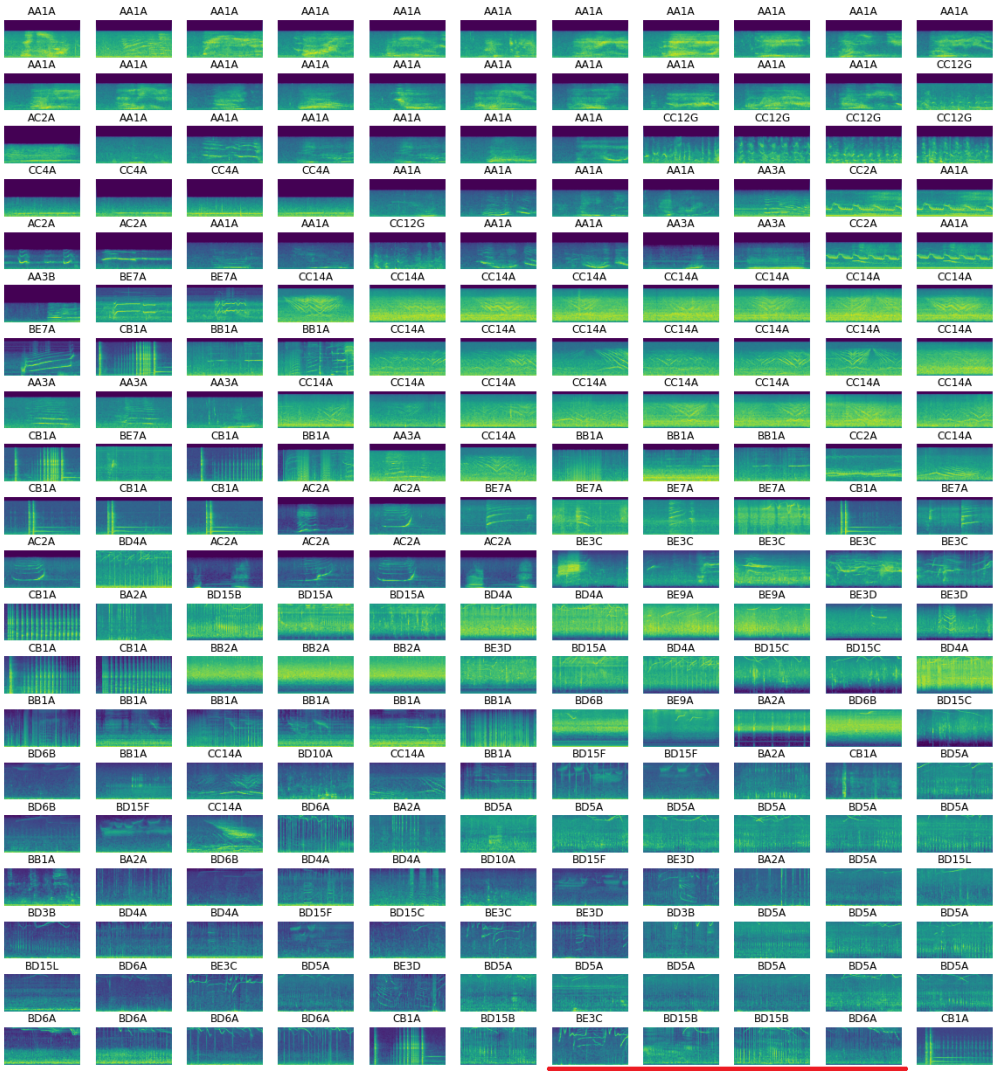
FIGURE 5 – Nuage de points 2D obtenu par l'application successive de PCA et t-SNE



Les codes indiqués en haut de chaque graphique sont liés à un nom d'espèce au sein d'un dictionnaire ayant servi à l'obtention des données de la banque de données. Les codes ont été indiqués à la place des noms d'espèce afin de ne pas surcharger la représentation par carte visuelle. Notons toutefois que le code «AA1A» correspond aux baleines boréales et que le code «CC4A» correspond aux phoques de Ross. En observant les spectrogrammes des enregistrements de ces espèces, on note que ceux-ci contiennent du «drop-out». Sur un spectrogramme, une bande noire ne veut pas nécessairement indiquer la présence de silence, mais plutôt une absence de sons captés par l'outil de mesure à cette fréquence, soit l'hydrophone dans le cas d'enregistrements sonores marins. La représentation par t-SNE regroupe donc ces enregistrements premièrement en fonction de la présence de «drop-out» et deuxièmement en fonction des fréquences perçues. Cela indique donc que les seuls regroupements interprétables dans la figure 5 sont les trois premiers, soit ceux excluant les regroupements de baleines boréales et de phoques de Ross. À l'écoute de ces enregistrements, il est en effet possible de remarquer la présence de certains motifs sonores communs. Par exemple, les

quatre avant-derniers enregistrements sonores en partant de la fin de la dernière ligne (voir le trait rouge dans la figure 6), qui correspondent respectivement aux sons produits par la baleine pilote, au dauphin clymène [1 et 2] et au dauphin à flancs blancs, possèdent tous une haute fréquence perceptible à leur écoute. L'utilisation successive de PCA et t-SNE détecte donc effectivement des patrons entre les enregistrements sonores. Cependant, il faut souligner que notre pré-traitement des fichiers audios ne comprenait pas de méthodes visant à réduire le bruit. Il est donc impossible de réfuter l'hypothèse que ces trois regroupements n'ont pas été formés en fonction de leur similitude au niveau du bruit.

FIGURE 6 – Carte visuelle 2D obtenue par l'application successive de PCA, t-SNE et la transformation Raster-Fairy (RF-transform)



6 Conclusion

En conclusion, la carte visuelle est une bonne façon de représenter le degré de similarité entre les spectrogrammes Mel produits à partir des enregistrements sonores. Celle-ci nous permet en effet de déceler des similitudes de sons produits entre les espèces marines étudiées. La grande influence du «drop-out» présent au sein de certains enregistrements sonores, ainsi que la présence de bruit, sont la principale faiblesse de notre travail. Cela ne découle pas des algorithmes de réduction de dimension utilisés, mais plutôt du pré-traitement des données. Il aurait été en effet pertinent de s'attarder aux considérations à prendre pour réduire le bruit dans nos enregistrements.
De plus, notre démarche ne permet pas de représenter les résultats de façon interactive. En effet, nous désirions au départ créer une carte visuelle permettant d'associer chaque son à son spectrogramme dans un tableau des données regroupées comme l'avait fait Kyle Macdonald et al. [5]. Cela aurait aussi contribuer à valider les regroupements produits par nos algorithmes. Nous avons choisi de ne pas réaliser cette tâche en raison des contraintes de temps.
Il est aussi important de rappeler que le projet est basé sur une perspective humaine de la perception des sons. En effet, autant dans le choix du spectrogramme Mel comme outil de comparaison des données que dans la non utilisation de l'audiogramme qui aurait permis d'analyser comment les sons sont perçus par les différents animaux, la généralisation de nos résultats au contexte communicationnel des animaux marins est affaiblie.
Finalement, il aurait été pertinent d'explorer d'autres algorithmes de réduction de dimension afin de possiblement trouver la façon optimale de représenter nos résultats. Cela aurait aussi pû nous permettre de valider la formation de certains regroupements et d'étoffer la partie de comparaison des représentations présentes dans nos objectifs.

7 Contribution des membres de l'équipe

Simon-Olivier a dirigé la section de programmation du projet. Il a travaillé à la compréhension de la démarche de Kyle Macdonald et al. Il a aussi été responsable de l'interface visuelle présentée.
Noémie s'est chargée des références biologiques concernant le sujet afin d'offrir une perspective plus proche des données lors de la démarche. Elle a aussi été active dans le téléchargement des données du site internet vers le Collab et dans la recherche concernant les algorithmes à utiliser.
Shophika a assisté Simon-Olivier dans l'implémentation des différents algorithmes, en plus de diriger la rédaction du rapport.

Références

[1] University of Rhode Island, Discovery of Sound in the Sea, 2020, https://dosits.org/people-and-sound/history-of-underwater-acoustics/the-first-practical-uses-of-underwater-acoustics-the-early-1900s/.

[2] John C. Montgomery, Craig A. Radford, Marine bioacoustics, Current Biology, Volume 27, Issue 11, 2017, Pages R502-R507, ISSN 0960-9822, https://doi.org/10.1016/j.cub.2017.01.041.

[3] Doris Elín Urrutia, Pilot Whales Show Possible Orca-Mimicking Repertoire, Scientific American, March 1st, 2021, https://www.scientificamerican.com/article/pilot-whales-show-possible-orca-mimicking-repertoire/.

[4] Ketan Doshi,Audio Deep Learning Made Simple (Part 1) : State-of-the-Art Techniques, Towards Data Science, February 11th, 2021, https://towardsdatascience.com/audio-deep-learning-made-simple-part-1-state-of-the-art-techniques-da1d3dff2504.

[5] Experiments with Google, Bird Sounds, May 2017, https://experiments.withgoogle.com/bird-sounds.

[6] Watkins Marine Mammal Sound Database, 2021, https://whoicf2.whoi.edu/science/B/whalesounds/index.cfm.

[7] Scikit learn, 2.2 Manifold Learning, 2007-2020, https://scikit-learn.org/stable/modules/manifold.html.

[8] Wikipedia, t-distributed stochastic neighbor embedding, 2021, https://en.wikipedia.org/wiki/T-distributed_stochastic_neighbor_embedding