Rapport de Recherche - 2025-06-30

Utilisation de méthodes d'alignement de séquences pour reconstituer des chants de baleine à bosse à partir d'extraits d'une minute

Auteur.es :

Opalin JAMES, Stéphane CHAVIN, Franck MALIGE, Giraudet PASCALE

Université de Toulon, Aix Marseille Univ, CNRS, LIS, DYNI, Marseille, France
Centre Int. d'IA en Acoustique Naturelle, Toulon, France, https://cian.univ-tln.fr

Remerciements

Je remercie tout d'abord très sincèrement mes trois encadrants Franck MALIGE, Pascale GIRAUDET et Stéphane CHAVIN pour leur confiance, pour m'avoir accompagné et épaulé tout au long de cette aventure y compris en cette fin de stage. J'ai également apprécié les échanges techniques, les conseils avisés, la disponibilité et surtout la patience et la bienveillance de mes encadrants tout au long de ce projet.

Un merci à Hervé GLOTIN pour nous avoir intégrés à son équipe du LIS. Les différents séminaires organisés par le groupe auxquels j'ai eu la chance d'assister m'ont permis de mieux comprendre les enjeux de la bioacoustique marine et d'étendre mes horizons scientifiques.

Je remercie aussi l'ensemble de mes collègues présents durant cette période pour tous les conseils, l'aide et autres longs échanges qui m'ont souvent animé en début ou fin de journée et qui ont contribué à faire de cette première expérience un magnifique stage.

Je souhaite dire un grand merci à mes amis, Noé BAUP, Elouan DERUCHE et Luca BENETTI pour leurs nombreuses aides lors de la conception des programmes python pendant ce stage. Leur savoir-faire, leur disponibilité et leur abondance de ressources ont été précieuses pour moi tout au long du stage. Sans eux, je n'aurais de toute évidence pas pu faire autant de choses ni aboutir aux nombreux résultats présentés dans ce rapport de stage.

Enfin, je tiens à remercier tous les membres et chercheurs du LIS, pour leur chaleureux accueil et la qualité de l'environnement de recherche qu'ils mettent à disposition.

TABLE DES MATIÈRES

TABLE DES FIGURES

Abréviations

ADN : Acide désoxyribonucléique

CARI'MAM : Caribbean Marine Mammals Preservation Network

CNN : Convolutional Neural Networks (Réseaux de neurones convolutifs)

IA : Intelligence artificielle

LIS : Laboratoire d'Informatique et Systèmes

Méthode itérative : Alignement itératif semi-global

Méthode progressive : Progressive Overlap Multiple Sequence Alignment

PAM : Surveillance acoustique passive (Passive Acoustic Monitoring)

YOLOv5 : You Only Look Once version 5

Introduction

Les baleines à bosse (*Megaptera novaeangliae*) produisent l'un des systèmes de communication acoustique les plus complexes du règne animal. Décrits pour la première fois en 1971, leurs chants sont constitués de séquences de sons variés qui durent de 7 à 30 minutes et sont répétées avec une précision remarquable [1]. Ces unités structurées sont principalement émises par les mâles pendant la saison de reproduction et constituent un phénomène vocal hautement élaboré [2]. Ces chants complexes leur permettent d'augmenter leurs chances de reproduction en faisant étalage de leur condition physique et jouent un rôle crucial dans la sélection sexuelle, que ce soit pour attirer les femelles ou pour médiatiser les interactions entre mâles [3], [4], [5]. Les recherches récentes ont démontré que les mâles modifient la structure de leur chant en présence d'autres chanteurs, suggérant des interactions acoustiques complexes similaires à celles observées chez les oiseaux chanteurs [4].

La structure des chants de baleines à bosse présente une organisation hiérarchique élaborée : des unités sonores fondamentales appelées vocalisations pouvant varier de 20 Hz à plus de 24 000 Hz [6] qui dure entre 1 à 3 secondes s'assemblent en phrases qui se répètent pour former des thèmes. Ces thèmes s'enchaînent ensuite dans un ordre relativement constant pour constituer un chant complet [7]. Ces chants peuvent se répéter sans interruption notable pendant plusieurs heures, formant ainsi des sessions de chant caractéristiques [1]. Les chants produits par une population de mâles sur un même lieu donné et au même moment possèdent de nombreuses similarités [7]. Ces chants présentent une évolution structurelle progressive au cours du temps, caractérisée par des modifications graduelles des unités acoustiques et de la composition des phrases qui s'orientent dans une direction évolutive constante pendant les périodes de chant [8].

Au-delà de cette évolution progressive, la transmission culturelle peut même prendre la forme de révolutions culturelles, où le chant d'une population entière est rapidement remplacé par un nouveau type de chant introduit par une population voisine [9], [10]. Le niveau et le rythme de ces changements culturels sont sans équivalent chez tout autre animal non-humain et impliquent un changement culturel à une échelle océanique [10]. Ces caractéristiques font des chants de baleines à bosse un modèle exceptionnel pour étudier l'évolution de la transmission culturelle chez les animaux.

Cependant, les contraintes techniques liées à l'enregistrement acoustique sous-marin créent un obstacle méthodologique important. Les dispositifs d'enregistrement acoustique autonomes utilisés pour la surveillance passive doivent équilibrer plusieurs facteurs limitants : autonomie énergétique, capacité de stockage, fréquence d'échantillonnage et durée de déploiement [11]. Ces limitations sont particulièrement contraignantes pour l'étude des chants de baleines à bosse, qui peuvent durer jusqu'à 30 minutes et se répéter pendant plusieurs heures. Pour surmonter ces contraintes, la plupart des projets de surveillance acoustique passive adoptent des protocoles d'échantillonnage intermittent, enregistrant par exemple une minute de son toutes les cinq à dix minutes [12]. Mais cette discontinuité induit des lacunes significatives et complexifie la distinction entre les unités des chants et celles utilisées dans la communication non chantée, tel que l'unité "D" [13], [14].

Cette fragmentation des enregistrements pose un défi majeur pour l'analyse des chants complets. Sans la continuité de la structure vocale, il devient difficile d'identifier avec précision les transitions entre thèmes, d'étudier la variabilité individuelle ou de suivre l'évolution temporelle des chants. Cette recherche vise à résoudre ce problème en développant une méthode de reconstruction des chants entiers à partir de fragments temporellement espacés.

Pour résoudre cette problématique de reconstruction, deux approches algorithmiques spécifiquement adaptées aux caractéristiques des chants de baleines à bosse, inspirées des techniques d'alignement de séquences utilisées en bio-informatique pour l'analyse de l'ADN, ont été développées. Ces méthodes tirent parti de la nature répétitive et ordonnée des unités vocales de ces mammifères marins : l'existence de structures récurrentes permet d'identifier des chevauchements entre fragments, tandis que l'organisation séquentielle hiérarchique des chants facilite la détermination de l'ordre correct des segments. La robustesse naturelle de ces unités, qui maintiennent leur structure malgré les variations individuelles et temporelles, constitue un avantage majeur pour l'approche de reconstruction automatique mise en place.

Matériels et méthodes

Le développement des méthodes de reconstruction de chants de baleines à bosse a suivi une approche méthodologique allant de l'entraînement contrôlé à l'application sur données acoustiques réelles.

1. Données d'entraînement et développement méthodologique

Avant l'application des algorithmes sur les données acoustiques réelles, une phase d'entraînement et de calibration a été réalisée sur des enregistrements complets de chants de baleines à bosse, fournis par Renata Sousa-Lima de L'UFRN (universidad federal Rio Grande do Norte). Il s'agit de trois enregistrements du même chant, collectés en septembre 2000 issus du stock A (population du Brésil), contenant respectivement 1046, 1290 et 2514 unités, pour des durées de 58, 97 et 127 minutes et correspondant aux enregistrements 1, 2 et 3 de l'article de F. Malige *et al.*, 2020 [15]. Pour simuler les conditions réelles d'échantillonnage du réseau CARI'MAM (Caribbean Marine Mammals Preservation Network), les chant fournis ont été artificiellement segmentés en fragments d'une minute, séparés par des intervalles de 4 à 5 minutes. Ainsi, ce protocole reproduit le schéma d'acquisition intermittent du réseau CARI'MAM (1 min enregistrée toutes les 6 minutes), permettant une évaluation réaliste des performances des méthodes d'alignement et une comparaison directe entre séquences consensus et originelles.

2. Méthodes d'alignement et de reconstruction des séquences consensus

Deux méthodes distinctes ont été développées pour l'alignement et l'assemblage de fragments de chants de baleines à bosse, en réponse aux contraintes spécifiques du protocole d'enregistrement discontinu.

2.1. Alignement itératif semi-global

La première méthode repose sur un alignement itératif semi-global. Cette méthode est un intermédiaire entre l'alignement global [16] (qui compare deux séquences complètes du début à la fin pour trouver la meilleure correspondance sur toute leur longueur) et l'alignement local [17] (qui identifie les zones de similarité les plus fortes entre deux séquences, indépendamment de leur position). L'alignement semi-global permet d'identifier des zones de chevauchement partiel entre fragments, en contraignant une seule extrémité de chaque séquence à s'aligner. Cette méthode est donc adaptée aux données fragmentées issues d'un échantillonnage intermittent et ayant pour but de reconstruire un chant complet.
Les fragments sont comparés deux à deux indépendamment pour limiter le temps de calcul, et regroupés à partir des paires présentant les meilleurs scores d'alignement. L'alignement multiple est ensuite construit de manière progressive, où les positions relatives sont déterminées à partir du décalage optimal identifié dans chaque paire d'alignement, en ajoutant chaque nouveau fragment au groupe existant le plus compatible. Cependant, cette méthode présente un biais dans le traitement des délétions, car les gaps ne sont pas réintroduits dans les séquences additionnelles, décalant ainsi l'alignement global.

2.2. Alignement progressif basé sur graphe

La seconde approche, s'apparente à du Progressive Overlap Multiple Sequence Alignment, et adopte une stratégie fondée sur la construction d'un graphe de chevauchement [18]. Chaque fragment est modélisé comme un nœud, et les arêtes entre nœuds représentent des chevauchements significatifs entre fragments, identifiés selon deux critères : une longueur minimale de recouvrement et un seuil de similarité basé sur la distance de Levenshtein [19] (qui mesure le nombre minimum d'opérations - insertions, suppressions ou substitutions - nécessaires pour transformer une séquence en une autre).
Les composantes connexes révèlent des groupes de fragments potentiellement issus d'un même chant. Un arbre couvrant de poids maximal est calculé pour chaque groupe de séquences en transformant les poids de similarité en valeurs négatives et en appliquant l'algorithme de Kruskal [20] pour obtenir l'arbre couvrant de poids minimum. L'algorithme de Kruskal procède en triant toutes les arêtes par poids croissant, puis en les intégrant progressivement à l'arbre en évitant la formation de cycles, jusqu'à obtenir une structure arborescente qui relie tous les nœuds. Cette approche privilégie la connectivité globale de l'ensemble plutôt qu'une optimisation locale des liaisons individuelles, permettant d'identifier des relations structurelles qui pourraient échapper à une sélection basée uniquement sur les scores de similarité les plus élevés.

Le nœud racine est sélectionné selon la centralité d'intermédiarité [21] (qui mesure la fréquence à laquelle une séquence se trouve sur les chemins reliant les autres séquences de l'arbre). Les positions relatives sont ensuite déterminées par parcours en largeur à partir de cette racine, en utilisant les informations de chevauchement pour calculer les décalages entre séquences adjacentes.

2.3. Résolution des conflits par vote majoritaire harmonisé

La phase de reconstruction du consensus fait appel à un système de vote majoritaire à trois étapes (*Annexe 1*). Dans la première étape, chaque position de la séquence consensus est déterminée par le caractère le plus fréquemment observé parmi les séquences alignées. L'étape 2 utilise un critère de similarité globale : pour chaque position ambiguë (position avec égalité lors de l'étape 1), la séquence présentant la similarité la plus élevée avec le consensus provisoire (distance de Levenshtein) détermine le caractère sélectionné. Enfin, l'étape 3 procède à une résolution aléatoire des ambiguïtés résiduelles, garantissant un consensus complet. Cette approche en cascade permet de préserver la cohérence biologique tout en assurant la robustesse du processus de reconstruction.

3. Métriques d'évaluation et optimisation paramétrique

3.1. Métrique d'évaluation

L'évaluation des reconstructions combine deux familles de métriques complémentaires, en tenant compte de la nature cyclique des chants par un alignement circulaire.

3.1.1. Distance de Levenshtein

La première métrique mise en place est la distance de Levenshtein [19]. Il s'agit d'une métrique permettant de déterminer le nombre d'opérations nécessaires (insertion, délétion et substitution) pour transformer une séquence en une autre. Cette métrique permet de détecter les erreurs d'alignement tout en étant robuste par rapport aux décalages locaux. La similarité de Levenshtein utilisée s'établit en calculant : 1 − distance de Levenshtein / longueur de la séquence de référence, fournissant un score normalisé entre 0 et 1 où 1 indique une correspondance parfaite.

3.1.2. Divergence de Jensen-Shannon

La seconde métrique mise en place pour évaluer les séquences consensus est la divergence de Jensen-Shannon [22]. Elle est calculée à partir de la distribution des k-mers (sous séquence de longueur k) présents dans les fragments alignés. Cette métrique, mesure la similarité entre les distributions de probabilité des fenêtres de taille k, capturant ainsi les motifs structuraux locaux caractéristiques des chants. L'utilisation de k-mers de longueur 4 permet de détecter la préservation ou l'altération des motifs récurrents comme les phrases, qui sont cruciaux dans l'organisation hiérarchique des unités de baleines à bosse. L'utilisation de k-mers plus courts ou plus longs ne permet pas d'identifier efficacement les motifs biologiquement pertinents les plus représentatifs des chants.

3.1.3. Score composite

Enfin une combinaison des 2 dernières métriques a été mise en place par une moyenne pondérée (50% Levenshtein et 50% Jensen-Shannon), offrant un score qui intègre à la fois la fidélité au niveau des caractères individuels et la préservation des motifs structuraux. De cette manière, ce score multiparamétrique permet une évaluation robuste qui s'avère importante pour prendre en compte la forte variabilité individuelle et temporelle que les chants peuvent subir.

3.2. Optimisation des paramètres

Les performances des deux méthodes d'alignement dépendent des combinaisons de leurs paramètres initiaux. Ainsi un programme d'optimisation de ces paramètres a été mis en place sur le jeu de données d'entraînement pour identifier la configuration la plus efficace pour chaque méthode d'alignement.
Pour la méthode itérative, les paramètres clés incluent le score de correspondance, la pénalité d'insertion/suppression, celle de substitution, et le seuil de qualité. Pour la méthode progressive, les paramètres clés comprennent cette fois-ci la longueur minimale de chevauchement, le seuil de similarité et les pondérations des scores.
Plusieurs milliers de combinaisons paramétriques ont été évaluées pour chacune des méthodes, avec les tests mis en place précédemment. Ainsi une évaluation objective des séquences consensus s'est effectuée

par rapport aux séquences originelles à travers les différentes métriques d'évaluation. Les configurations optimales ont ensuite été sélectionnées sur la base du score composite maximal, garantissant une performance équilibrée sur l'ensemble des critères d'évaluation.

4. Détermination du nombre minimal de fragments pour la reconstruction

Au-delà de l'optimisation algorithmique, la question de la quantité minimale de données nécessaire constitue un enjeu pratique fondamental pour l'application opérationnelle des méthodes développées.

4.1. Protocole d'échantillonnage simulé

Un prélèvement aléatoire de fragments de 18 caractères successifs dans la séquence de référence a été effectué sur un chant du jeu de données d'entraînement, simulant fidèlement les conditions réelles d'acquisition du réseau CARI'MAM (une minute d'enregistrement par fragment). Le nombre de fragments extraits varie de 2 à 100, avec 5 répétitions garantissant la robustesse statistique des résultats, et l'analyse de la précision des séquences consensus en fonction de la densité d'échantillonnage.

4.2. Modélisation par courbes de saturation exponentielles

Une méthode objective a été développée pour détecter les points de transition via un modèle de saturation exponentiel (*Annexe 2*) identifiant le "knee point" [23], point où la courbe de performance atteint un plateau et où l'ajout de fragments supplémentaires n'apporte plus d'amélioration significative. Le principe repose sur l'identification du point de la courbe de tendance présentant la distance perpendiculaire maximale par rapport à la droite reliant les points extrêmes de la série de données.

4.3. Évaluation comparative des méthodes

La comparaison des deux méthodes est effectuée selon plusieurs métriques dans des conditions strictement identiques et sur le même jeu de données d'entraînement où les fragments sont extraits aléatoirement. Les différentes métriques incluent la similarité de Levenshtein et les ratios de longueur entre séquences consensus et originelles, garantissant ainsi que les scores obtenus reflètent directement les performances relatives des méthodes d'alignements.

5. Données acoustiques de CARI'MAM et prétraitement

5.1. Données acoustiques et prétraitement par le LIS

Les données proviennent du réseau CARI'MAM, comprenant 19 stations hydroacoustiques déployées dans l'arc antillais [13], [24] dont 14 opérationnels. Le protocole d'acquisition suit un échantillonnage intermittent (1 minute toutes les 6 minutes), générant environ 240 fichiers audio par jour et par station.
La détection et classification des unités de baleines à bosse ont été réalisées en amont de cette étude par le LIS via le modèle YOLOv5 (You Only Look Once version 5), un modèle d'apprentissage profond ou "deep learning" appliqué aux représentations spectrographiques [13], permettant d'identifier et catégoriser 353 958 unités. Pour faciliter l'analyse computationnelle et l'application d'algorithmes d'alignement de séquences tout en préservant l'information structurelle, les 28 types d'unités identifiées ont été représentés par un symbole alphabétique unique et seuls les enregistrements sans superposition de chants ont été conservés.

5.2. Sélection et filtrage des données

Pour l'analyse, un filtrage rigoureux a été appliqué à l'ensemble des données issues du réseau CARI'MAM avec comme exigence un minimum de trois caractères vocaux dont minimum deux distincts par séquence, évitant toute confusion avec des unités de communication simples ou des artefacts de détection.
L'analyse s'est concentrée sur les 3 stations en Guadeloupe et celle des Bermudes. Les stations de Guadeloupe représentent une zone de reproduction importante où les baleines à bosse mettent bas pendant leur migration avant de remonter vers les zones d'alimentation nordiques [25]. Les Bermudes, zone plus septentrionale, constituent une zone de "stopover" [26] caractérisée par une relative stabilité des chants. Les individus n'étant pas sur les sites de reproductions en présence de leurs congénères maintiennent des structures vocales consistantes pendant leurs séjours prolongés.
Seules les données entre le 22 janvier et le 4 mai ont été conservées, correspondant au pic d'activité acoustique des baleines à bosse, période durant laquelle la production de chants est la plus intense.

Les données ont été organisées par station et période hebdomadaire pour une analyse temporelle fine, limitant l'association artificielle de fragments trop distants temporellement qui pourraient avoir subi des variations structurelles significatives. Seules les semaines avec le nombre minimal de fragments, déterminé comme seuil nécessaire pour assurer un fonctionnement fiable des méthodes d'alignement ont été retenues pour l'analyse finale.

6. Analyses statistiques

L'ensemble des tests statistiques a été réalisé sous la version de Python 3.8. Le premier test est celui de Mann-Whitney, il permet de comparer les longueurs des séquences consensus avec celles des fragments réels. Le second test appliqué est celui de Kruskal-Wallis, pour analyser les variations temporelles des longueurs entre différentes semaines et comparer les performances entre stations d'enregistrement. Le troisième test utilise le χ^2 d'indépendance pour comparer la distribution des unités entre deux thèmes considérés comme identiques. Enfin le troisième test concerne l'analyse probabiliste des phrases dans les séquences consensus. L'analyse se base sur les fréquences des unités de chaque semaine pour chaque station et les probabilités finales sont obtenues par application de la loi de Poisson (*Annexe 3*) pour déterminer la probabilité d'observer un certain nombre d'événements en connaissant le nombre moyen d'occurrences attendues.

Résultats

---

1. Optimisation des paramètres d'alignement

L'évaluation des séquences consensus obtenue à partir de 10 000 combinaisons, sur le jeu de données d'entraînement, a permis d'identifier les paramètres optimaux pour chacune des deux méthodes d'alignement développées.

Tableau I. *Scores de performance obtenus*

Métrique	Méthode itérative	Méthode progressive
Score combiné	0.6124	0.5730
Similarité Levenshtein	0.2759	0.2148
Similarité Jensen-Shannon	0.9489	0.9311

Tableau II. *Paramètres optimaux identifiés*

Paramètre	Méthode itérative	Méthode progressive
Score de correspondance	1	1
Coût/Pénalité de délétion	-8	-2
Coût/Pénalité de mutation	-3	-6
Seuil de qualité	0.1	-
Chevauchement minimal	-	5
Seuil de chevauchement	-	0.4
Seuil de similarité	-	0.3

A l'issue de ces résultats la méthode itérative semble être à privilégier avec des scores plus élevés que la méthode progressive dans toutes les métriques (score combiné de 0.6124 contre 0.5730).

2. Qualité de la reconstruction et détermination du nombre minimal de fragments pour la reconstruction

L'analyse de la qualité de la reconstruction et du nombre minimal de fragments nécessaire pour obtenir une séquence consensus de qualité acceptable réalisé sur les données d'entraînement révèle des différences significatives entre les deux méthodes développées.

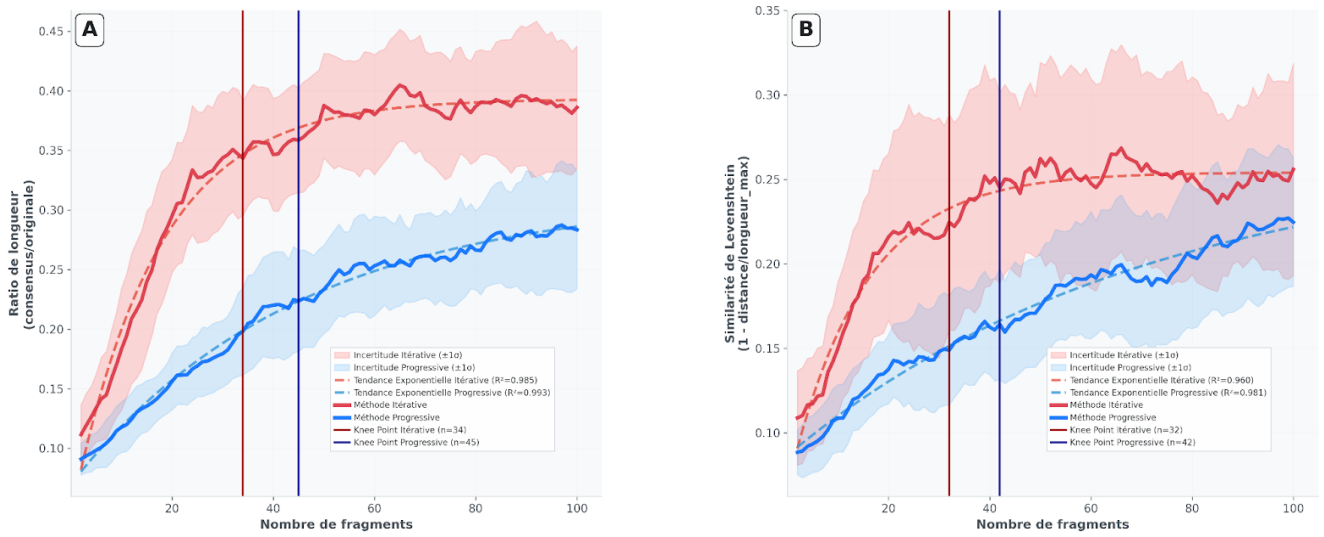


Figure 1. *(A) Évolution du ratio de longueur (consensus / originel) des séquences consensus en fonction du nombre de fragments. (B) Évolution de la similarité de Levenshtein directe en fonction du nombre de fragments pour les méthodes itérative et progressive.*

La recréation de la séquence consensus utilisant les paramètres optimaux de la méthode itérative est fonctionnelle, avec une relativement bonne préservation de l'ordre de première apparition des thèmes observés dans les chants originels (*Annexe 4*). Cependant, les reconstructions n'atteignent qu'environ 37% de la longueur des chants complets (*Figure 1A*), révélant les limites actuelles de la méthode d'assemblage. Cette diminution de la taille des séquences consensus s'explique par une réduction de la taille de chaque thème au cours du processus de reconstruction, ainsi que l'absence de répétitions de ces thèmes lorsque ceux-ci sont présents plusieurs fois dans le chant. De plus, les thèmes de faible longueurs tel que le thème 4 ne sont pas reconstruits dans les séquences consensus (*Annexe 4*).

L'analyse de la stabilité des longueurs consensus (*Figure 1A*) révèle la convergence des séquences reconstruites. La méthode itérative atteint un knee point à 34 fragments sur la régression exponentielle (R^2=0.9847), indiquant que l'amélioration du ratio de longueur se stabilise au-delà de ce seuil avec une décélération de la progression. La méthode progressive converge plus tardivement avec un knee point à 45 fragments (R^2=0.9931), montrant une accélération progressive dans l'optimisation du ratio de longueur.

L'analyse de la qualité de reconstruction (*Figure 1B*) complète cette évaluation. La méthode itérative présente un knee point à 32 fragments (R^2=0.9595), point où la similarité de Levenshtein passe d'une phase d'accélération initiale à une phase de saturation de l'amélioration. La méthode progressive nécessite 42 fragments pour atteindre son point d'inflexion optimal (R^2=0.9814), avec une accélération continue mais plus modérée de l'amélioration de la similarité.

Ces résultats démontrent que la méthode itérative est plus efficace en termes de quantité de données requises, atteignant une performance optimale avec environ 24% moins de fragments que la méthode progressive (32 contre 42 fragments pour la similarité et 34 contre 45 fragments pour le ratio de longueur). Cette différence s'explique par la nature de l'algorithme itératif qui exploite plus directement les chevauchements locaux entre fragments adjacents.

3. Application aux données acoustiques réelles du réseau CARI'MAM

L'application de la méthode itérative aux données acoustiques réelles issues du réseau CARI'MAM (35 séquences consensus et 7971 fragments réels) avec les paramètres optimaux, révèle des résultats contrastés qui soulignent la complexité de la reconstruction automatique de chants complets.

L'analyse structurelle des reconstructions (*Figure 2A*) révèle que les séquences consensus présentent une longueur moyenne de 43.17 ± 13.20 caractères (médiane = 43.0, intervalle = [18, 83]), significativement supérieure à celle des fragments réels (*Figure 2D*) (7.53 ± 4.35 caractères, médiane = 6.0). Le test de Mann-Whitney confirme cette différence (U = 280 150.5, p-value < 10^{-6}), démontrant que l'algorithme de reconstruction assemble effectivement les fragments courts en séquences consensus plus longues.

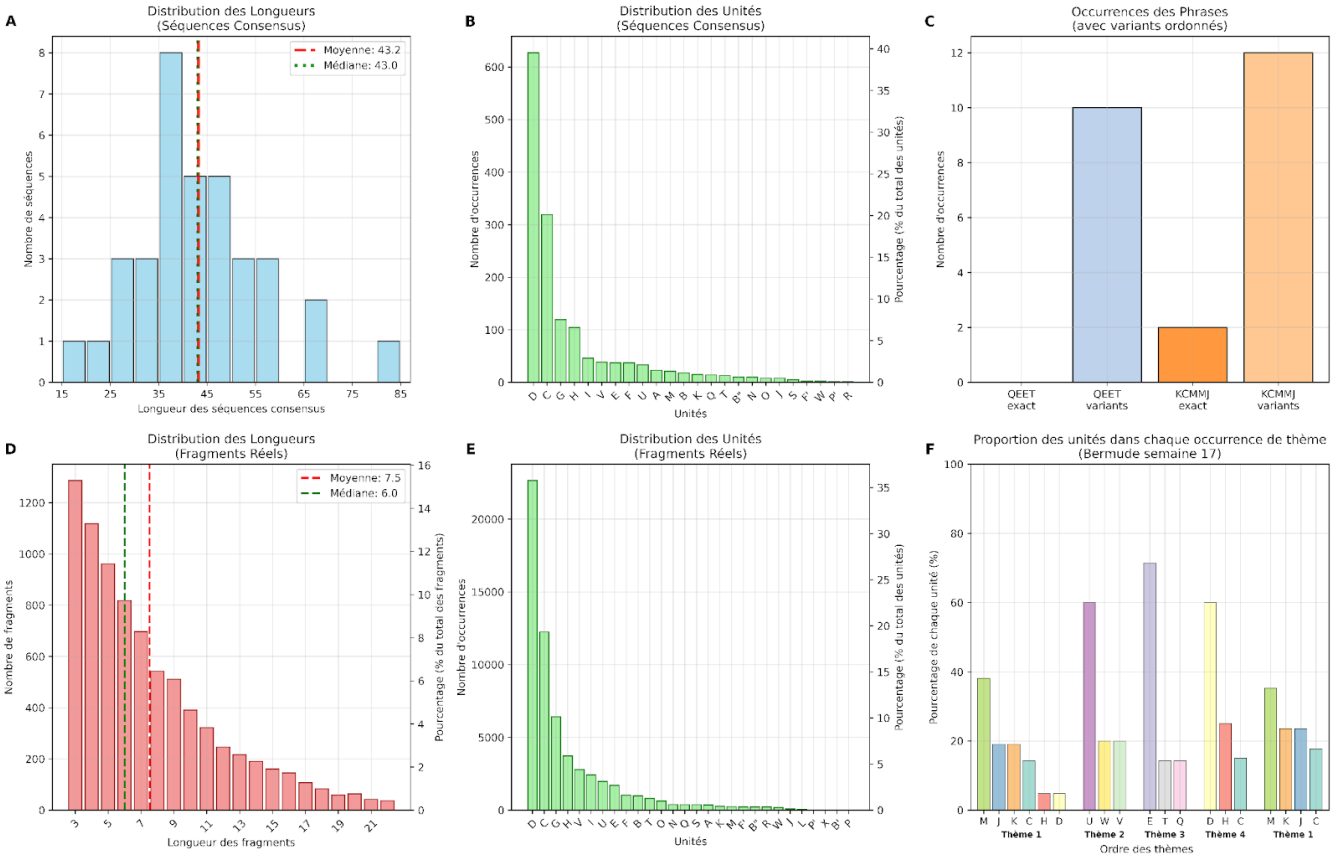


Figure 2. *Analyse comparative complète des séquences consensus et des fragments réels du réseau CARI'MAM. (A) Distribution des longueurs des séquences consensus. (B) Distribution des unités dans les séquences consensus. (C) Occurrences des phrases biologiques QEET et KCMMJ avec variants ordonnés. (D) Distribution des longueurs des fragments réels. (E) Distribution des unités dans les fragments réels. (F) Proportion des unités dans chaque occurrence de thème pour la séquence Bermude semaine 17.*

L'analyse de la composition caractéristique (*Figures 2B et 2E*) montre une dominance des unités “D” (41.5% dans les consensus, 37.5% dans les fragments) et “C” (21.1% dans les consensus, 20.3% dans les fragments), confirmant la cohérence biologique entre les données reconstruites et réelles malgré leurs différences de longueur. Qui plus est, l’ordre des fréquences est remarquablement bien conservé au-delà de ces deux types dominants entre les fragments et les séquences consensus. Par ailleurs la distribution des types d’unités varie significativement entre les différentes stations d'enregistrement (test de Kruskal-Wallis : $H = 257.03$, p-value $< 10^{-6}$), reflétant probablement les spécificités géographiques et temporelles des populations observées, notamment entre les zones de transit (Guadeloupe) et les zones de halte migratoire (Bermudes).

L'examen des motifs biologiquement pertinents (*Figure 2C*) apporte un poids mitigé à la validité des reconstructions. Les phrases QEET et KCMMJ, précédemment reconnus comme spécifiques aux chants des baleines à bosse, ne sont pas forcément identifiées dans les séquences de consensus Cependant leurs variantes le sont (10 variantes pour QEET et 12 variantes pour KCMMJ qui conservent l'ordre des unités). L’analyse probabiliste, prenant en compte les fréquences des unités de chaque semaine de chaque station montre que ces phrases ne relèvent pas de l’aléatoire. La probabilité d’obtenir au moins 10 variants de QEET est de $3{,}548\times10^{-13}$, celle d’avoir au moins 2 fois la phrase KCMMJ est de $1{,}391\times10^{-5}$ et celle pour au moins 12 variants de KCMMJ est de $4{,}663\times10^{-15}$. Ces probabilités extrêmement faibles semblent indiquer que nous retrouvons des motifs biologiques et non des formations aléatoires.

La plus grande séquence consensus reconstruite (BERMUDE semaine 17, 83 unités) a été segmentée en cinq thèmes distincts basés sur des observations inter-observateurs (*Annexe 5*). L'analyse de la proportion de chaque unités au seins de chaque thème (Figure 2F) démontre que chaque thème présente un profil caractéristique : le Thème 1 dominé par M, J et K, le Thème 2 par U, W et V, le Thème 3 par E, T et Q, et le Thème 4 par D, H et C. La comparaison statistique entre le premier et le dernier thème confirme la similarité compositionnelle ($\chi^2 = 1.89$, ddl $= 5$, p-value $\geq 8{,}65\times10^{-1}$), contrastant avec les différences significatives observées entre tous les autres thèmes. Cette concordance entre le début et la fin de la

séquence s'accorde parfaitement avec la nature cyclique des chants de baleines à bosse, suggérant que la méthode de reconstruction permet effectivement de reconstruire des chant

Discussion

1. Efficacité comparative des approches de reconstruction

La méthode itérative a démontré une supériorité, fournissant de meilleures performances, avec des séquences consensus plus grandes que la méthode progressive, restant cependant limitée à environ 37% de la taille du chant originel. De plus, la méthode itérative présente une similarité de Levenshtein plus haute (0.25 pour la méthode itérative contre 0.21 pour la méthode progressive). La comparaison du nombre de fragments minimum révèle que la méthode itérative nécessite 24% moins de fragments (34 vs 45) pour atteindre une performance optimale comparée à la méthode progressive. Cette efficacité s'explique par l'exploitation directe des chevauchements locaux entre fragments adjacents.

Par ailleurs, la méthode itérative semble également être un meilleur compromis en termes de complexité algorithmique. La complexité de calcul des deux méthodes est très faible lors d'un alignement d'un faible montant de fragments (moins de 1 seconde pour aligner 2 fragments). Néanmoins la méthode progressive devient particulièrement coûteuse pour de jeux de données d'environ 500 fragments, en raison de sa complexité $O(n^2 \times L + n^3)$ (où n représente le nombre de fragments et L la longueur moyenne des fragments) qui inclut les algorithmes de graphes NetworkX, pouvant nécessiter jusqu'à 10 minutes de temps de calcul. En comparaison, la méthode itérative avec sa complexité $O(n^2 \times L^2)$ ne prend que 2 à 3 minutes de temps de calcul pour aligner 500 fragments.

Les développements récents en bioinformatique montrent une évolution significative des approches, depuis l'amélioration continue des algorithmes d'alignement de séquences multiples face aux défis d'échelle croissante [27] jusqu'à l'automatisation complète des analyses multi-omiques par les agents IA [28], corroborant le choix d'adapter les techniques d'alignement de séquences biologiques au contexte bioacoustique. Compte tenu de ces avancées méthodologiques et des performances démontrées lors des tests appliqués, la méthode itérative est privilégiée pour la reconstruction de chants de baleines à bosse.

Bien que la méthode progressive basée sur graphe présente certains avantages théoriques en termes de flexibilité structurelle pour gérer les fragments désordonnés, ses performances inférieures et sa complexité computationnelle plus élevée la rendent moins adaptée aux applications pratiques de reconstruction de chants. Les algorithmes bio-inspirés pour l'alignement multiple de séquences continuent d'évoluer avec une diversité de méthodologies, incluant des approches multi-objectifs et hybrides, suggérant des perspectives d'amélioration pour les méthodes futures.

2. Validation des reconstructions thématiques

La réduction de la taille de chaque thème au cours du processus de reconstruction s'explique par les mécanismes d'alignement qui fusionnent les phrases d'un même thème au même endroit dans la séquence consensus, entraînant une compression de l'information thématique. La méthode d'alignement a tendance à privilégier l'alignement de structures robustes plutôt que des zones variables, créant ainsi un alignement des phrases d'un même thème au même endroit. Par ailleurs, bien que l'ordre d'apparition des thèmes soit relativement bien préservé, les thèmes de faible longueur ne sont pas reconstruits, comme le thème 4 dans les données d'entraînement, probablement dû au fait que leur détection nécessite un échantillonnage précis et plusieurs fragments comportant les zones de transition adjacentes, particulièrement difficiles à aligner en raison de leur variabilité.

Malgré les limitations identifiées, certains résultats suggèrent que les méthodes utilisées parviennent à identifier les différents thèmes des chants. L'analyse de la séquence consensus des Bermudes à la semaine 17 révèle un structure particulièrement intéressante : la séquence présente plusieurs changements de thèmes distincts et commence et finit par le même thème. Cette structure circulaire est cohérente avec la nature cyclique bien documentée des chants de baleines à bosse [1], suggérant que nous avons effectivement réussi à reconstituer les différents thèmes constitutifs du chant. L'analyse des proportions d'unités révèle des thèmes bien distincts en termes de composition (*Figure 2F*), confirmant que la méthode itérative parvient à identifier et préserver l'organisation structurelle hiérarchique des chants.

Cette observation valide partiellement l'approche méthodologique retenue et démontre le potentiel de reconstruction de structures complexes à partir de fragments discontinus. Cependant, cette validation reste limitée à quelques cas spécifiques et nécessiterait une analyse plus systématique sur l'ensemble des séquences consensus produites pour établir des conclusions généralisables.
Enfin, la conservation de l'ordre des fréquences des types d'unités entre les fragments réels et les séquences consensus, confirme la cohérence du processus d'assemblage sans pour autant garantir la validité biologique des structures reconstruites. La dominance des unités "D" (41.5%) et "C" (21.1%) soulève cependant une problématique importante : ces unités étant fréquemment utilisées dans la communication non-chantée [13], [14], leur surreprésentation suggère une contamination significative par des éléments de communication qui pourrait masquer la véritable structure des chants. Cette variabilité inter-stations dans la distribution des unités reflète probablement les spécificités géographiques et temporelles des populations observées, notamment entre les zones de reproduction (Guadeloupe) [25] et les zones de halte migratoire (Bermudes) [26].

3. Validation biologique et conservation des motifs structurels

La conservation des motifs biologiquement significatifs (QEET et KCMMJ) dans les séquences consensus constitue un indicateur crucial de la validité écologique des approches choisies. Ces phrases spécifiques ont déjà été identifiées dans des enregistrements géographiquement et temporellement distants (QEET aux Bermudes le 24 février 2021 puis à Guadeloupe Anse Bertrand le 25 février, KCMMJ aux Bermudes les 15 avril 2021 et 5 février 2022), confirmant la reproduction fidèle de structures comportementales naturelles et validant la cohérence des reconstructions avec les processus de transmission culturelle observés chez les baleines à bosse [13].
Cependant l'examen de ces phrases révèle des résultats plus nuancés concernant la fidélité des reconstructions. La phrase QEET, bien qu'absent dans sa forme exacte, apparaît sous 10 variantes conservant l'ordre des unités avec une probabilité d'occurrence extrêmement faible (probabilité : $p = 3{,}548\times10^{-13}$). La phrase KCMMJ est retrouvé 2 fois dans sa forme exacte (probabilité : $p = 1{,}391\times10^{-5}$) accompagné de 12 variantes supplémentaires (probabilité $p = 4{,}663\times10^{-15}$). Cette capacité à reproduire partiellement la phrase exacte tout en générant des variantes structurellement cohérentes indique que les approches conservent l'essence des motifs biologiques sans reproduire leur complexité intégrale. Ces limitations pourraient refléter les contraintes méthodologiques.

4. Limitations méthodologiques et biais systémiques

Plusieurs limitations méthodologiques critiques méritent d'être soulignées. Les limites de l'utilisation précise et généralisable des paysages sonores pour surveiller la biodiversité montrent que les approches analytiques existantes se comportent de manière imprévisible entre les études [29]. Cette variabilité souligne l'importance de développer des méthodes robustes et reproductibles pour l'analyse des données bioacoustiques.
L'annotation des unités en lettres, qu'elle soit humaine ou automatisée par IA, constitue un biais en amont de cette étude qui introduit des erreurs dans l'analyse. L'annotation humaine est sujette à la subjectivité et à la variabilité inter-observateur, notamment pour les unités avec des signatures acoustiques proches [30], tandis que les CNN (Convolutional Neural Networks) peuvent avoir des confusions entre des unités sous-représentées (types R/M, S/N) et des difficultés de généralisation liées aux variabilités parfois importantes des conditions d'enregistrements [13]. De plus, l'annotation des données d'entraînement et celles de CARI'MAM n'étant pas issues de la même méthode [13], [15], cela induit de potentielles erreurs de calibration des méthodes d'alignement et affectent directement la fiabilité des séquences consensus reconstruite avec les données de CARI'MAM. Cela met en évidence l'importance de mettre en place des méthodes intégrant les incertitudes de classification.
La méthode utilisée dans cette étude est dépendante de la capture de fragments spécifiques marquant les zones de transitions entre thèmes pour reconstituer un chant entier. Cela signifie qu'il faut avoir enregistré ces zones de transitions entre les différents thèmes pour une semaine d'une station donnée. Cette dépendance explique en partie les variations de qualité de reconstruction qui peuvent survenir pour les stations avec une couverture temporelle limitée. Il convient également de noter que les fragments d'entraînement utilisés sont quasiment sans imperfections et complets, tandis que les fragments issus de CARI'MAM ne présentent parfois que quelques unités et que les unités "D" et "C", surreprésentés sont

souvent associés à de la communication non chantée. Cette contamination par ces unités se caractérise par la répétition en boucle d'une même unité, caractéristique des séquences de communication plutôt que des structures thématiques organisées des chants [14]. Il est donc crucial de développer une méthode pour déterminer la nature d'une unité et ainsi limiter les biais lors de l'alignement des fragments et de la recréation des séquences consensus. Enfin les séquences consensus ne représentent qu'environ un tiers de la longueur attendue des chants complets. Ce résultat est en partie dû aux méthodes d'alignement qui ont tendance à aligner des phrases aux mêmes positions, alors que celles-ci s'avèrent souvent être répétées à plusieurs endroits dans les chants naturels.
La validation des reconstructions reste également complexe. L'approche d'alignement circulaire, ne peut capturer toutes les nuances comportementales des chants naturels. Les variations individuelles et les évolutions temporelles des répertoires vocaux représentent des sources de variabilité qui peuvent affecter la qualité des reconstructions. Les méthodes de détection automatique pour la bioacoustique présentent encore des défis considérables, particulièrement dans la phase de détection qui implique la distinction de nombreux signaux temporellement et spectralement superposés [30].

5. Perspectives d'amélioration et implications pour la surveillance acoustique passive

Plusieurs axes d'amélioration peuvent être envisagés pour optimiser ces méthodes. L'amélioration des méthodes d'alignements multiples pourrait résoudre les problèmes de désalignement identifiés dans la méthode itérative et réduire un biais fondamental dans la fiabilité des séquences consensus. L'implémentation d'approches probabilistes pour modéliser l'incertitude dans les alignements représente par exemple une solution qui pourrait augmenter la robustesse et la fiabilité des séquences consensus. Ces approches probabilistes consisteraient à attribuer des scores de confiance ou des probabilités à chaque position de l'alignement, plutôt que de considérer chaque alignement comme définitif. Cette quantification de l'incertitude permettrait de pondérer différemment les régions bien alignées de celles plus ambiguës lors de la construction de la séquence consensus finale.
Les avancées récentes en apprentissage profond, utilisant des techniques comme les CNN et les transformers, ont montré des précisions de détection impressionnantes [31]. De plus, l'intégration de ces approches facilite le traitement d'enregistrements acoustiques à long terme en générant de vastes volumes de données pré-annotées [13]. Ces méthodes d'apprentissage automatique pourraient considérablement améliorer la robustesse des approches d'alignement, notamment pour différencier la nature des unités et identifier clairement celles liées à la communication non chantée de celles utilisées dans un chant.
Ces résultats ont également des implications importantes pour l'amélioration des protocoles de surveillance acoustique passive (PAM). La surveillance acoustique passive est devenue un outil transformateur pour l'écologie appliquée, la conservation et le suivi de la biodiversité, mais son potentiel de contribution à l'écologie fondamentale reste encore sous-exploité [32]. La fragmentation temporelle des enregistrements, contrainte par les limitations énergétiques et de stockage des dispositifs autonomes, peut désormais être partiellement compensée par des approches de reconstruction algorithmique. Les protocoles d'échantillonnage intermittent (1 minute toutes les 6 minutes) adoptés par les réseaux de surveillance peuvent désormais être optimisés en tenant compte des seuils minimaux identifiés (34-45 fragments selon la méthode). Cette optimisation permettrait d'améliorer le rapport coût-efficacité des déploiements tout en maintenant une qualité de reconstruction acceptable.
L'application de ces méthodes à d'autres espèces de cétacés représente une perspective prometteuse. Les approches comparatives en bioacoustique permettent de comprendre les causes et conséquences de la variation acoustique entre taxa avec des analyses phylogénétiques comparatives [33]. Cette perspective comparative pourrait révéler des structures générales dans l'organisation des unités de mammifères marins et contribuer à une meilleure compréhension de l'évolution de la communication acoustique.
La dimension culturelle des chants de baleines offre également des opportunités d'application uniques. Les révolutions culturelles des chants de baleines à bosse, où le chant partagé par toute une population est rapidement remplacé par un nouveau type introduit par une population voisine, représentent un exemple extraordinaire de transmission culturelle à l'échelle océanique [34]. Les méthodes de reconstruction améliorées proposées par cette étude pourraient faciliter l'étude de ces phénomènes de transmission culturelle en permettant l'analyse de chants complets même à partir de données fragmentées, ouvrant ainsi de nouvelles perspectives pour comprendre les mécanismes de diffusion culturelle chez les mammifères marins.

CONCLUSION

Les méthodes employés lors de cette étude montrent des résultats contrastés mais représentant tout de même un pas en avant et contribue au progrès de la bioacoustique et à la mise en place de méthodologie pour traiter efficacement les bases de données importantes obtenus provenant des PAM ou d'autres protocoles d'échantillonnage intermittent.

Les méthodes utilisées dans cette étude démontrent qu'un assemblage des différents fragments permettent effectivement de reconstruire une partie des chants de baleines à bosse (*Megaptera novaeangliae*) lorsque les conditions préalables sont réunies, la méthode itérative démontrant de meilleure performance. Ces conditions imposent évidemment d'avoir des fragments d'une minute claire où il n'y aucun problème d'identification du chant ou de l'individu chantant par exemple et cela pour chaque partie du chant. De plus, les analyses démontrent qu'un minimum de 34 fragments sont nécessaires pour procéder à un alignement fiable avec les méthodes testées ici.

Ces méthodes représentent une avancée vers l'automatisation intégrale de l'analyse des chants des mammifères marins, offrant ainsi de nouvelles perspectives pour saisir les processus de transmission culturelle et d'évolution comportementale chez ces espèces symboliques. Cependant, plusieurs biais méthodologiques significatifs subsistent et requièrent encore la mise en place de solutions ou d'alternatives. La distinction imparfaite entre les unités de chant et de communication fait partie des points exigeant une attention particulière pour limiter les erreurs lors de la création des séquences consensus et la dissimulation de certaines structures du chants de ces cétacés.

Enfin, cette étude s'inclut dans une démarche plus large de conservation marine, où la compréhension des structures de communication pourrait permettre la mise en place de nouvelle stratégie de protections d'espèces ou d'habitats et de limiter perturbations anthropiques sur ces populations de cétacés.

Bibliographie

[1] R. S. Payne et S. Mcvay, « Humpbacks emit sounds in long, predictable pattern ranging over frequencies audible to human », *Science*, vol. 173, p. 585-597.

[2] H. E. Winn et L. K. Winn, « The song of the humpback whale Megaptera novaeangliae in the West Indies », *Mar. Biol.*, vol. 47, nº 2, p. 97-114, 1978, doi: 10.1007/BF00395631.

[3] K. Kowarski, S. Cerchio, H. Whitehead, et H. Moors-Murphy, « Where, when, and why do western North Atlantic humpback whales begin to sing? », *Bioacoustics*, vol. 31, nº 4, p. 450-469, juill. 2022, doi: 10.1080/09524622.2021.1972838.

[4] D. M. Cholewiak, S. Cerchio, J. K. Jacobsen, J. Urbán-R., et C. W. Clark, « Songbird dynamics under the sea: acoustic interactions between humpback whales suggest song mediates male interactions », *R. Soc. open sci.*, vol. 5, nº 2, p. 171298, févr. 2018, doi: 10.1098/rsos.171298.

[5] L. M. Herman, « The multiple functions of male song within the humpback whale (*Megaptera novaeangliae*) mating system: review, evaluation, and synthesis », *Biological Reviews*, vol. 92, nº 3, p. 1795-1818, août 2017, doi: 10.1111/brv.12309.

[6] W. W. L. Au, A. A. Pack, M. O. Lammers, L. M. Herman, M. H. Deakos, et K. Andrews, « Acoustic properties of humpback whale songs », *The Journal of the Acoustical Society of America*, vol. 120, nº 2, p. 1103-1110, août 2006, doi: 10.1121/1.2211547.

[7] D. M. Cholewiak, R. S. Sousa-Lima, et S. Cerchio, « Humpback whale song hierarchical structure: Historical context and discussion of current classification issues », *Marine Mammal Science*, vol. 29, nº 3, juill. 2013, doi: 10.1111/mms.12005.

[8] S. C. Tyarks, A. S. Aniceto, H. Ahonen, G. Pedersen, et U. Lindstrøm, « Changes in humpback whale song structure and complexity reveal a rapid evolution on a feeding ground in Northern Norway », *Front. Mar. Sci.*, vol. 9, p. 862794, déc. 2022, doi: 10.3389/fmars.2022.862794.

[9] M. J. Noad, D. H. Cato, M. M. Bryden, M.-N. Jenner, et K. C. S. Jenner, « Cultural revolution in whale songs », *Nature*, vol. 408, nº 6812, p. 537-537, nov. 2000, doi: 10.1038/35046199.

[10] E. C. Garland *et al.*, « Dynamic Horizontal Cultural Transmission of Humpback Whale Song at the Ocean Basin Scale », *Current Biology*, vol. 21, nº 8, p. 687-691, avr. 2011, doi: 10.1016/j.cub.2011.03.019.

[11] R. Sousa-Lima, « A Review and Inventory of Fixed Autonomous Recorders for Passive Acoustic Monitoring of Marine Mammals », *Aquat Mamm*, vol. 39, nº 1, p. 23-53, mars 2013, doi: 10.1578/AM.39.1.2013.23.

[12] S. Van Parijs *et al.*, « Management and research applications of real-time and archival passive acoustic sensors over varying temporal and spatial scales », *Mar. Ecol. Prog. Ser.*, vol. 395, p. 21-36, déc. 2009, doi: 10.3354/meps08123.

[13] S. Chavin, H. Glotin, P. Best, et Y. Ourmieres, « Spatial and Temporal Dynamics of the Humpback Whales' Vocal Repertoire : An Automatic Analysis Over Two Years of a Large Hydrophone Network ». (in submission)

[14] M. E. Fournet, A. Szabo, et D. K. Mellinger, « Repertoire and classification of non-song calls in Southeast Alaskan humpback whales (*Megaptera novaeangliae*) », *The Journal of the Acoustical Society of America*, vol. 137, nº 1, p. 1-10, janv. 2015, doi: 10.1121/1.4904504.

[15] F. Malige, D. Djokic, J. Patris, R. Sousa-Lima, et H. Glotin, « Use of recurrence plots for identification and extraction of patterns in humpback whale song recordings », *Bioacoustics*, vol. 30, nº 6, p. 680-695, nov. 2020, doi: 10.1080/09524622.2020.1845240.

[16] S. B. Needleman et C. D. Wunsch, « A general method applicable to the search for similarities in the amino acid sequence of two proteins », *Journal of Molecular Biology*, vol. 48, nº 3, p. 443-453, mars 1970, doi: 10.1016/0022-2836(70)90057-4.

[17] M. Waterman, « General methods of sequence comparison », *Bulletin of Mathematical Biology*, vol. 46, n°4, p. 473-500, 1984, doi: 10.1016/S0092-8240(84)800054-3

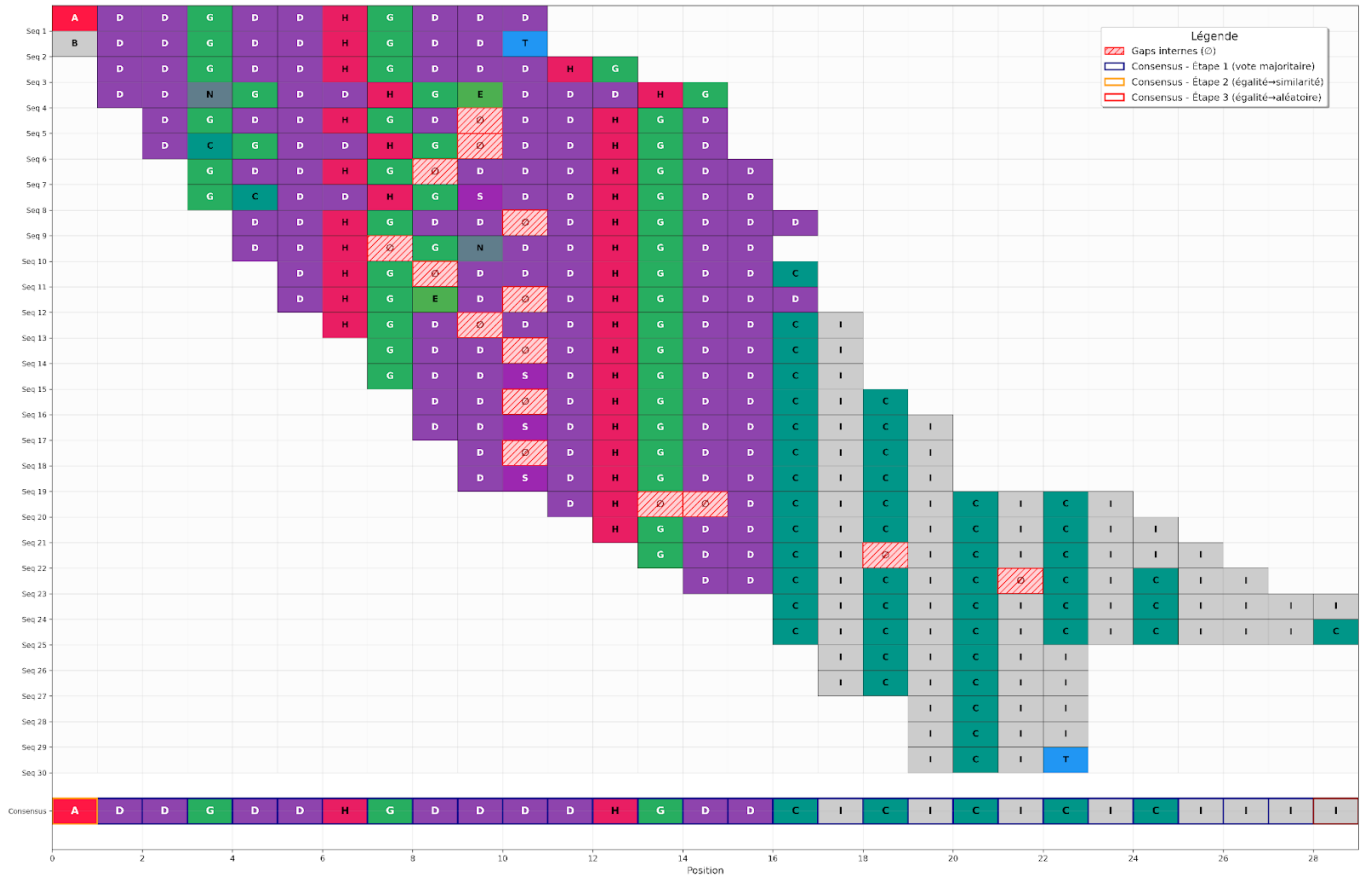
[18] C. Lee, C. Grasso, et M. F. Sharlow, « Multiple sequence alignment using partial order graphs », *Bioinformatics*, vol. 18, nº 3, p. 452-464, mars 2002, doi: 10.1093/bioinformatics/18.3.452.

[19] V. I. Levenshtein, « Binary Codes Capable of Correcting Deletions, Insertions and Reversals », *Soviet Physics Doklady*, vol. 10, p. 707, févr. 1966.

[20] J. B. Kruskal, « On the shortest spanning subtree of a graph and the traveling salesman problem », *Proc. Amer. Math. Soc.*, vol. 7, nº 1, p. 48-50, févr. 1956, doi: 10.1090/S0002-9939-1956-0078686-7.
[21] L. C. Freeman, « A Set of Measures of Centrality Based on Betweenness », *Sociometry*, vol. 40, nº 1, p. 35, mars 1977, doi: 10.2307/3033543.
[22] J. Lin, « Divergence measures based on the Shannon entropy », IEEE Trans. Inform. Theory, vol. 37, n°1, p. 145-151, janv. 1991, doi: 10.1109/18.61115
[23] V. Satopaa, J. Albrecht, D. Irwin, et B. Raghavan, « Finding a "Kneedle" in a Haystack: Detecting Knee Points in System Behavior », in *2011 31st International Conference on Distributed Computing Systems Workshops*, Minneapolis, MN, USA: IEEE, juin 2011, p. 166-171. doi: 10.1109/ICDCSW.2011.20.
[24] Hervé Glotin, Maxence Ferrari, Paul Best, Marion Poupard, Nicolas Thellier, et al.. CARIMAM REPORT BIOACOUSTIC DATA PROCESSING. [Research Report] DYNI LIS. 2021. ⟨hal-03629286⟩
[25] Nadège Gandilhon. Contribution au recensement des cétacés dans l'archipel de Guadeloupe. Océanographie. Université des Antilles (UA) - Site de Guadeloupe, FRA, 2012. Français. ⟨NNT : ⟩. ⟨tel-03910549⟩
[26] T. Grove, R. King, A. Stevenson, et L.-A. Henry, « A decade of humpback whale abundance estimates at Bermuda, an oceanic migratory stopover site », *Front. Mar. Sci.*, vol. 9, p. 971801, janv. 2023, doi: 10.3389/fmars.2022.971801.
[27] Y. Zhang, Q. Zhang, J. Zhou, et Q. Zou, « A survey on the algorithm and development of multiple sequence alignment », *Briefings in Bioinformatics*, vol. 23, nº 3, p. bbac069, mai 2022, doi: 10.1093/bib/bbac069.
[28] J. Zhou *et al.*, « An AI Agent for Fully Automated Multi-Omic Analyses », *Advanced Science*, vol. 11, nº 44, p. 2407094, nov. 2024, doi: 10.1002/advs.202407094.
[29] S. S. Sethi *et al.*, « Limits to the accurate and generalizable use of soundscapes to monitor biodiversity », *Nat Ecol Evol*, vol. 7, nº 9, p. 1373-1378, juill. 2023, doi: 10.1038/s41559-023-02148-z.
[30] R. Gibb, E. Browning, P. Glover-Kapfer, et K. E. Jones, « Emerging opportunities and challenges for passive acoustics in ecological assessment and monitoring », *Methods Ecol Evol*, vol. 10, nº 2, p. 169-185, févr. 2019, doi: 10.1111/2041-210X.13101.
[31] A. N. Allen *et al.*, « A Convolutional Neural Network for Automated Detection of Humpback Whale Song in a Diverse, Long-Term Passive Acoustic Dataset », *Front. Mar. Sci.*, vol. 8, p. 607321, mars 2021, doi: 10.3389/fmars.2021.607321.
[32] S. R. P. -J. Ross *et al.*, « Passive acoustic monitoring provides a fresh perspective on fundamental ecological questions », *Functional Ecology*, vol. 37, nº 4, p. 959-975, avr. 2023, doi: 10.1111/1365-2435.14275.
[33] K. J. Odom *et al.*, « Comparative bioacoustics: a roadmap for quantifying and comparing animal sounds across diverse taxa », *Biological Reviews*, vol. 96, nº 4, p. 1135-1159, août 2021, doi: 10.1111/brv.12695.
[34] E. C. Garland et P. K. McGregor, « Cultural Transmission, Evolution, and Revolution in Vocal Displays: Insights From Bird and Whale Song », *Front. Psychol.*, vol. 11, p. 544929, sept. 2020, doi: 10.3389/fpsyg.2020.544929.

Annexes

A.1. Exemple de création d'une séquences consensus à partir d'un alignement multiple de séquences



Résolution de conflits dans l'assemblage de séquences. Représentation matricielle du processus de vote majoritaire harmonisé montrant l'évolution du consensus à travers les trois étapes : vote majoritaire initial (étape 1), résolution par égalité-similarité (étape 2), et résolution aléatoire finale (étape 3). Les zones hachurées indiquent les gaps internes (délétion créée pour avoir un meilleur alignement), les couleurs différentes représentent les divers caractères des séquences d'entrée, et la séquence consensus finale est affichée en bas de la matrice.

A.2. Formulation mathématique du modèle de saturation exponentiel

Le modèle de saturation exponentiel utilisé pour la détection des knee points est défini par :

$$f(x) = ae^{-bx} + c$$

où :

- **x** : variable normalisée entre 0 et 1 selon $\frac{(x-x_{min})}{(x_{max}-x_{min})}$ pour améliorer la convergence numérique
- **a** : amplitude de l'amélioration initiale (paramètre libre : -∞ à +∞)
- **b** : coefficient de vitesse de convergence vers le plateau (contraint : 0.01 à 10.0)
- **c** : niveau de performance asymptotique (paramètre libre : -∞ à +∞)

A.3. Analyse probabiliste des phrases dans les séquences consensus

A.3.1. Fréquences des unités par station et semaine

Pour chaque station s et semaine w, la fréquence d'une unité u est calculée par :

$$f_{s,w}(u) = \frac{N_{s,w}(u)}{N_{s,w}^{total}}$$

où :

- $N_{s,w}(u)$ = nombre d'occurrences de l'unité "u" dans la station "s" et semaine "w"
- $N_{s,w}^{total}$ = nombre total d'unités observées dans la station "s" et semaine "w"

Cette métrique permet d'établir les probabilités de base pour chaque type d'unité en tenant compte des variations spatiales et temporelles.

A.3.2. Probabilité d'occurrence d'un motif spécifique

La probabilité d'observer un phrase donné dans une séquence de longueur L est calculée par :

$$P(phrase) = \prod_{i=1}^{n} f_{s,w}\left(u_i\right)$$

où :

- n = longueur de la phrase
- u_i = i-ème unité du de la phrase
- $f_{s,w}\left(u_i\right)$ = fréquence de l'unité u_i dans la station et semaine considérée

Cette formule suppose l'indépendance des unités successives, hypothèse simplificatrice mais nécessaire pour l'évaluation statistique.

A.3.3. Application de la loi de Poisson

Pour déterminer la probabilité d'observer au moins *k* occurrences d'une phrase, la loi de Poisson est appliquée :

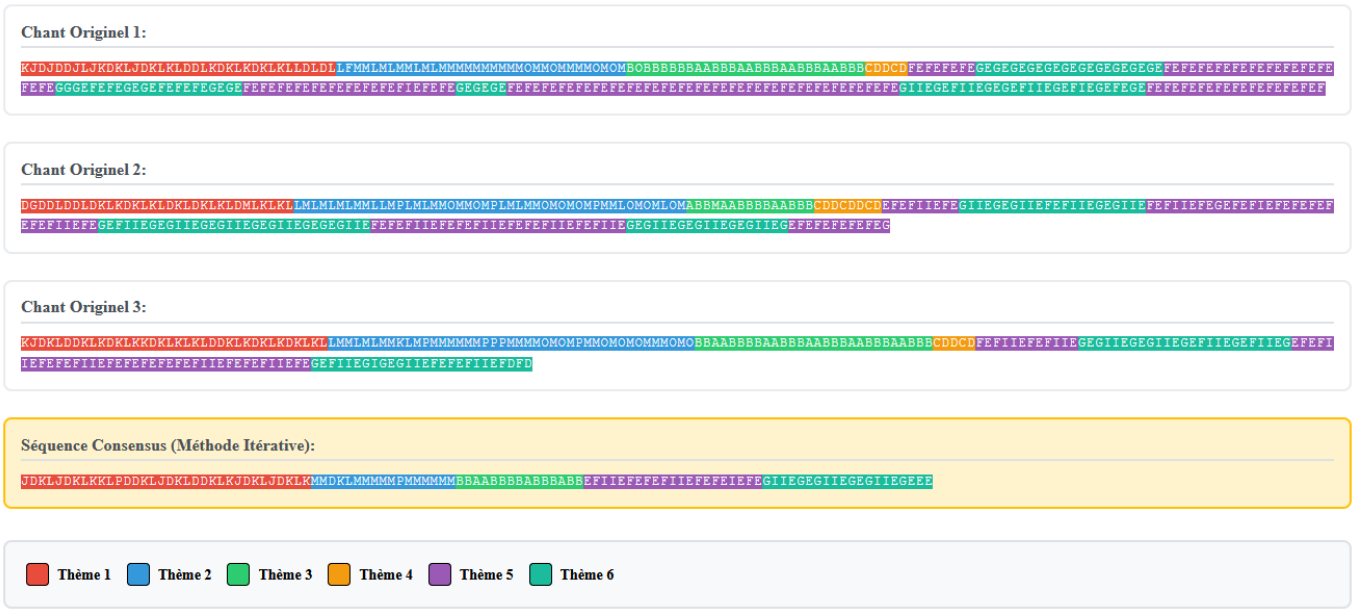
$$P(X \geq k) = 1 - \sum_{i=0}^{k-1} \frac{\lambda^i e^{-\lambda}}{i!}$$

où :

- $\lambda = (L - n + 1) \times P(phrase)$ = nombre moyen d'occurrences attendues
- L = longueur totale de la séquence consensus
- k = nombre d'occurrences observées

La loi de Poisson est utilisée car elle modélise efficacement les événements rares et indépendants dans des séquences longues, permettant d'évaluer si les motifs observés dépassent significativement le seuil du hasard.

A.4. Séquence consensus reconstruite avec les paramètres optimaux de la méthode itérative et séquences originelles



A.5. Chant reconstruit de la semaine 17 de la station des Bermude

Résumé

Les mammifères marins possèdent l'un des systèmes de communication acoustique les plus sophistiqués dans le règne animal. Les chants des baleines à bosse (*Megaptera novaeangliae*) montrent une organisation hiérarchique complexe où les unités se combinent pour former des phrases, puis des thèmes, créant des chants complets qui peuvent s'étendre sur une durée de 30 minutes et évoluer culturellement à l'échelle océanique. Néanmoins, les restrictions techniques des PAM nécessitent des protocoles d'échantillonnage discontinu qui segmentent temporellement les enregistrements, ce qui entrave l'analyse complète des structures vocales et l'examen des processus évolutifs de ces chants.
La fragmentation des données acoustiques constitue un obstacle majeur pour reconstituer l'intégralité des chants et comprendre leur organisation séquentielle. Nous démontrons ici que les techniques d'alignement de séquences tirées de la bioinformatique permettent de reconstituer partiellement les chants de ces cétacés à partir de segments d'une minute, grâce à une méthode itérative qui nécessite 24% moins de données qu'une approche progressive basée sur un graphe (34 fragments contre 45). À l'opposé des méthodologies antérieures qui se limitaient à l'analyse de segments isolés, les reconstructions proposées conservent dans une moindre mesures les phrases biologiquement pertinentes (QEET, KCMMJ) et restituent la complexité structurale en conservant partiellement l'ordre de premières apparition des différents thèmes d'un chant. L'application de cette méthode sur les données du réseau CARI'MAM prouve sa viabilité opérationnelle sur des enregistrements concrets, malgré certains biais méthodologiques résiduels.

Abstract

Marine mammals possess one of the most sophisticated acoustic communication systems in the animal kingdom. Humpback whale (*Megaptera novaeangliae*) songs exhibit a complex hierarchical organization where units combine to form phrases and then themes, creating complete songs that can span 30 minutes and evolve culturally across oceanic scales. However, the technical limitations of PAM require discontinuous sampling protocols that temporally segment recordings, hampering the comprehensive analysis of vocal structures and the examination of song evolutionary processes. The fragmentation of acoustic data poses a major obstacle to reconstructing complete songs and understanding their sequential organization. Here we demonstrate that sequence alignment techniques derived from bioinformatics can partially reconstruct the songs of these cetaceans from one-minute segments, using an iterative method that requires 24% less data than a progressive graph-based approach (34 fragments versus 45). Unlike previous methodologies that were limited to the analysis of isolated segments, the proposed reconstructions preserve to a lesser extent biologically relevant phrases (QEET, KCMMJ) and restore structural complexity by partially preserving the order of first appearance of the different themes of a song. The application of this method on data from the CARI'MAM network proves its operational viability on concrete recordings, despite some residual methodological biases.