Transcription automatique des chants de baleines bleues

Léa BOUFFAUT[1*], Shyam MADHUSUDHANA[2], Valérie LABAT[1], Abdel-Ouahab BOUDRAA[1] Holger KLINCK[2]

[1]Institut de Recherche de l'École Navale, EA3634,
École Navale/Arts et Métiers ParisTech - BCRM Brest CC600, 29240 Brest Cedex 9, France

[2]Bioacoustics Research Program, Cornell Lab of Ornithology,
Cornell University, 159 Sapsucker Woods Road, Ithaca, NY 14850

lea.bouffaut@ecole-navale.fr

Résumé – L'analyse du grand volume de données généré par des enregistrements acoustiques long-terme et continus pour l'étude et le suivi des populations de baleines bleues est largement facilitée par l'extraction des signaux cibles, en particulier lorsqu'ils sont noyés dans divers bruits de fond. La méthode proposée dans ce travail permet d'aller plus loin que les algorithmes de détection et classification classiques [1] : elle permet la transcription indépendante et automatique des chants de baleines bleues présents dans un enregistrement multi-espèces. Comme preuve de concept, nous proposons une approche en trois étapes : détection des tonales, reconnaissance de formes et reconstruction. La méthode est testée sur des données réelles entachées de bruits sismiques et de chorus de mammifères marins, les résultats obtenus démontrent sa faisabilité.

Abstract – Analysis of large volumes of data resulting from continuous and long-term recordings for blue whale acoustic monitoring benefits from automated extraction of the target signals, especially when they are embedded in background noise. The method proposed in this work is going further than traditional detection and classification algorithms [1]: automated transcription of independant blue whale songs is proposed, in multi-species recordings. As a proof of concept, this approach relies on three sucessive steps: tonal signal detection, pattern recognition and signal reconstruction. The method is tested on real data with seismic noise and whale chorus and results demonstrate its feasability.

1 Introduction

L'étude des populations de baleines bleues par acoustique passive a démontré, ces dernières décennies, être une approche économique et non-intrusive permettant de couvrir de vastes zones avec un seul capteur grâce à leurs vocalises basses fréquences [2]. L'analyse du grand volume de données généré par des enregistrements long-terme et continus est largement facilitée par l'extraction des signaux cibles, en particulier lorsqu'ils sont noyés dans du bruit ambiant. Les chants de baleines bleues sont connus pour être spécifiques à chaque sous-espèce [3]. En dessous de 50 Hz, ils sont généralement décrits comme des séries régulières d'unités tonales, le plus souvent stéréotypées, ce qui en fait des candidates idéales pour la transcription automatique. Il existe deux grandes tendances pour la détection de chants de baleines [1] : les méthodes basées sur le filtrage adapté temporel ou sur spectrogramme ([4], [5]) et celles qui cherchent à classifier l'ensemble des éléments présents dans un enregistrement, pour retrouver les signaux d'intérêt [6]. La méthode proposée entre dans la deuxième catégorie. Afin d'aller plus loin que les systèmes de détection et classification traditionnels, elle s'appuie sur le fait que dans un enregistrement multi-espèces, les calls peuvent être regroupés par mesure de similarité et identifiés afin de reconstruire indépendamment chaque chant sous-jacent. Cette méthode de transcription automatique de chants de baleines bleues permet d'améliorer les analyses visuelles (et audio) des enregistrements, de faire de l'annotation automatique de données ou bien tout simplement de faciliter les analyses statistiques sur les signaux. Dans le cas où un seul type de signal tonal est présent dans l'enregistrement, elle peut être utilisée comme étape de débruitage.

* Ces travaux sont soutenus par le GdR ISIS, la Bourse de la Fondation de la Mer et l'École Navale.

2 Méthode

L'architecture proposée est détaillée dans la Figure 1. Ce travail repose sur la représentation temps-fréquence (TF) des signaux acoustiques. Tout d'abord, les signaux d'intérêt sont extraits par un détecteur de tonales dans le plan TF (§ 2.1). Puis, des attributs sont calculés à partir de la représentation TF des tonales extraites, cette étape est suivie d'une réduction de la dimension de l'espace par une analyse en composantes principales (ACP). Dans le but de grouper les tonales détectées en fonction des similitudes de leurs attributs, une étape de clustering est conduite dans ce nouvel espace (§ 2.2). Les informations de regroupement ainsi obtenues (temps, fréquence, puissance, classe) sont alors utilisées pour transcrire et reconstruire indépendamment les différents types de chants présents dans les enregistrements (§ 2.3).

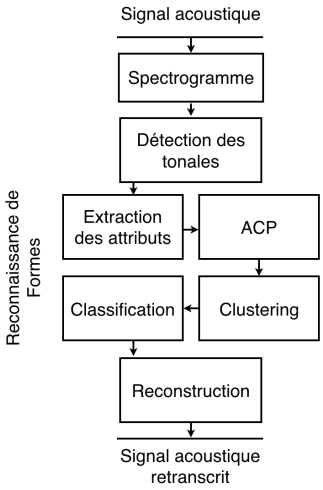

FIGURE 1: Architecture globale pour la transcription automatique des chants de baleines.

2.1 Détection de tonales

Dans une étude antérieure [7], certaines méthodes de détection de tonales populaires (telles que l'estimation de la fréquence instantanée, l'estimateur YIN, un détecteur basé sur une fonction de coût, et le détecteur de crêtes, etc.) ont été comparées en utilisant des métriques pertinentes pour quantifier (i) l'efficacité de ces détecteurs à récupérer de manière fiable les tonales et (ii) la qualité des résultats de détection. Les détecteurs ont été testés à l'aide de données couvrant un large éventail de contextes acoustiques et de valeurs de rapport signal sur bruit. Le détecteur de crêtes [8] repose sur une technique de traitement d'image qui est habituellement utilisée pour la sélection automatique de caractéristiques ou pour la segmentation d'images. Il permet d'obtenir des détections fiables et cohérentes [7] et a donc été choisi pour ce travail. En quelques mots, le détecteur de crêtes considère les régions du spectrogramme correspondant à des tonales comme des crêtes d'intensité et détermine leurs contours dans le plan TF. Tout d'abord les points TF positionnés le long de l'extrémité d'une crête sont identifiés puis, en utilisant des informations supplémentaires provenant du voisinage spectro-temporel immédiat des points précédemment identifiés, les points TF correspondant aux tonales sont connectés en utilisant une approche de filtrage bayésien. Les résultats de l'évaluation du détecteur de crêtes conduite dans [7] sont présentés en § 3.1 et comparés à un détecteur idéal.

2.2 Reconnaissance de formes

Les tonales identifiées par le détecteur de crêtes sont caractérisées à l'aide de plusieurs attributs ou caractéristiques qui sont déterminés à partir de leurs informations temps-fréquence-amplitude. L'ensemble des caractéristiques choisies vise à maximiser la capacité de discrimination. Sont mesurées sur chacune des tonales détectées les informations de :

- durée,
- fréquence centrale,
- amplitude moyenne,
- écart type de l'amplitude,
- pente minimale, maximale, moyenne et instantanée,
- ratio des fréquences des tonales présentes simultanément.

Ces attributs sont normalisés. Les composantes principales sont identifiées, les tonales détectées projetées dans ce nouvel espace sont ensuite regroupées par similarité (clustering). Le clustering est réalisé par la méthode des k-means, rassemblant dans une même classe les éléments "proches" dans la projection des données sur les 3 premières composantes principales, correspondant à 76% de la variance totale). Dans un premier temps les résultats présentés § 3.2 font lieu de preuve de concept où le nombre de clusters est déterminé arbitrairement, de manière à satisfaire le nombre déterminé visuellement.

2.3 Reconstruction du signal

Les tonales détectées regroupées dans une même classe sont utilisées pour reconstruire la forme d'onde de chants indépendants. Ceci est réalisé en multipliant point par point la transformée de Fourier à court terme (TFCT) du signal d'entrée $X(t,f) \in \mathbb{C}$ (permettant la reconstruction audio sans pertes d'informations de phase) et le masque temps fréquence (ou filtre binaire) de la $i^{\text{ème}}$ classe à reconstruire $Y_i(t,f) \in \{0,1\}$ avec $i = \{1, 2...C\}$ tel que

$$Z_i(t,f) = Y_i(t,f) \odot X(t,f), \tag{1}$$

où l'opérateur $\odot$ représente le produit de Hadamard. Une fois l'opération de filtrage réalisée, la reconstruction audio de chacune des classes est réalisée en prenant la TFCT inverse de $Z_i(t,f)$ tel que $z_i(t) = \text{iTFCT}\{Z_i(t,f)\}$, permettant de reconstruire les chants de baleine de chaque sous-espèce présente dans l'enregistrement.

3 Résultats

3.1 Évaluation du détecteur de crêtes

La qualité du détecteur de crêtes est mise en regard sur la Figure 2 du "détecteur idéal". Ces résultats sont obtenus [7] à partir de simulations Monte-Carlo réalisées en injectant des chants de baleine bleue antarctique simulés dans des enregistrements de bruits réels et en faisant varier le Rapport Signal sur Bruit (RSB) défini en adéquation avec [9]. Les scores du détecteur de crêtes sont peu dépendants du RSB. L'écart entre les fréquences attendues et mesurées est équivalent à la résolution fréquentielle du spectrogramme. Les scores de fragmentation et de couverture sont largement influencés par la nature du signal détecté : 75% de ce signal en forme de Z dans le plan TF a un aspect tonal, ce qui correspond à la valeur maximale atteinte pour la couverture. Le reste (la barre verticale du Z) correspond à un chirp linéaire descendant avec une pente très forte, ce qui explique le score de fragmentation de $\simeq 0.6$. Ces très bonnes

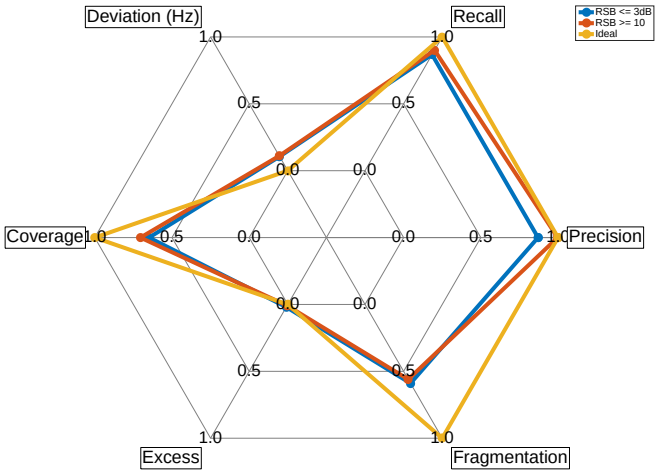

FIGURE 2: Diagramme représentant la qualité du détecteur de crêtes pour des RSB inférieurs à 3 dB et supérieurs à 10 dB, en termes de rappel, précision, fragmentation, excès de signal, couverture et l'écart entre les fréquences attendues et mesurées.

performances sur la détection de tonales ainsi que la possibilité de détecter des signaux simultanés font du détecteur de crêtes un outil idéal pour l'étape de détection de la transcription. Les performances de cette méthode sont cependant sensibles à la résolution du spectrogramme, qui doit être choisie afin optimiser la visualisation des signaux d'intérêt.

3.2 Transcription automatique

Comme preuve de concept, la méthode décrite dans la § 2 est appliquée à un enregistrement mixte de baleine bleue pygmée de Madacascar (MPBW) et de P-calls, acquis par l'hydrophone d'un sismomètre de fond de mer (Ocean Bottom Seismometer - OBS) provenant de la campagne de mesures sismiques RHUM-RUM, réalisée entre 2012 et 2013 dans l'ouest de l'océan Indien [10]. La Figure 3 (a)&(b) présente la forme d'onde et le spectrogramme à l'entrée de la méthode de transcription. Le nombre de points pour la TFCT (ainsi que pour le spectrogramme) est fixé à nfft = 512, avec un overlap de 80%, permettant une résolution fréquentielle de 0.2 Hz et temporelle de 1 s (< à la durée des signaux d'intérêt). D'après Figure 3 (c), le détecteur de crêtes permet de retrouver les unités principales de l'unité 1 de MPBW, à 13.5 et 34 Hz, celle de l'unité 2 autour de 25 Hz ainsi que le P-call à 27 Hz. Quelques tonales non attendues sont détectées autour du séisme à 400 s. Le clustering automatique permet de faire ressortir les 3 classes correspondant aux unités citées précédemment, comme le montre la couleur associée à chaque détection de Figure 3 (c). Les points ne correspondant à aucune classe sont affichés en gris. La reconstruction des différentes formes d'ondes présentée Figure 3 (d) montre qu'il est possible de reconstruire indépendamment des signaux qui se chevauchent dans le temps.

Lors du clustering automatique, certains points tombent à la frontière de deux classes comme par exemple à $\simeq$ 220 s.

4 Conclusion et perspectives

Ces résultats préliminaires démontrent qu'il est possible de débruiter et simplifier la lecture sur spectrogramme (et l'écoute) d'enregistrements multi-espèces en utilisant une méthode de transcription automatique des chants de baleines bleues.

Pour la suite de ces travaux, le nombre de clusters et leurs dimensions seront déterminés à partir d'une base de données d'entraînement annotée de 48 heures (riche de plus de 500 calls annotés par espèce) contenant l'ensemble des chants des espèces de baleines présentes dans zone d'étude. Le nombre de clusters sera alors défini comme celui qui minimisera l'erreur moyenne c'est à dire la moyenne des distances de Mahalanobis entre chaque point et le centre du cluster qui lui est attribué. Dans cette étape, permettant de fixer la définition de chaque sous-ensemble ou classe plusieurs points sont critiques : le choix des attributs doit permettre de discriminer au mieux les différentes tonales ainsi que de rapprocher les éléments similaires dans le nouvel espace, le type de distance 3D mesurée doit être représentative de la proximité des échantillons similaires observés et à l'inverse de l'espacement des différentes classe potentielles.

Pour l'application, lorsqu'une une nouvelle tonale sera détectée, elle subira la même transformation que la base de données d'entraînement et se verra attribuer la classe associée au cluster qui lui est circonscrit. Si le point n'entre dans aucun cluster prédéfini, il sera classifié comme "autre". La représentativité de la base de donnée d'apprentissage est donc déterminante dans la qualité de l'étape de clustering.

Remerciements

Ces travaux de recherche ont reçu le soutien du GdR ISIS, de la Bourse de la Fondation de la Mer et de l'École Navale.

Références

[1] M. F. Baumgartner and S. E. Mussoline, "A generalized baleen whale call detection and classification system," *Journal of the Acoustical Society of America*, vol. 129, no. 5, pp. 2889–2902, 2011.

[2] T. A. Branch, K. M. Stafford, D. M. Palacios, *et al.*, "Past and present distribution, densities and movements of blue whales balaenoptera musculus in the southern hemisphere and northern Indian Ocean," *Mammal Review*, vol. 37, no. 2, pp. 116–175, 2007.

[3] M. A. McDonald, S. L. Mesnick, and J. A. Hildebrand, "Biogeographic characterization of blue whale song worldwide : Using song to identify populations," *Journal of Cetacean Research and Management*, vol. 8, no. 1, pp. 55–65, 2006.

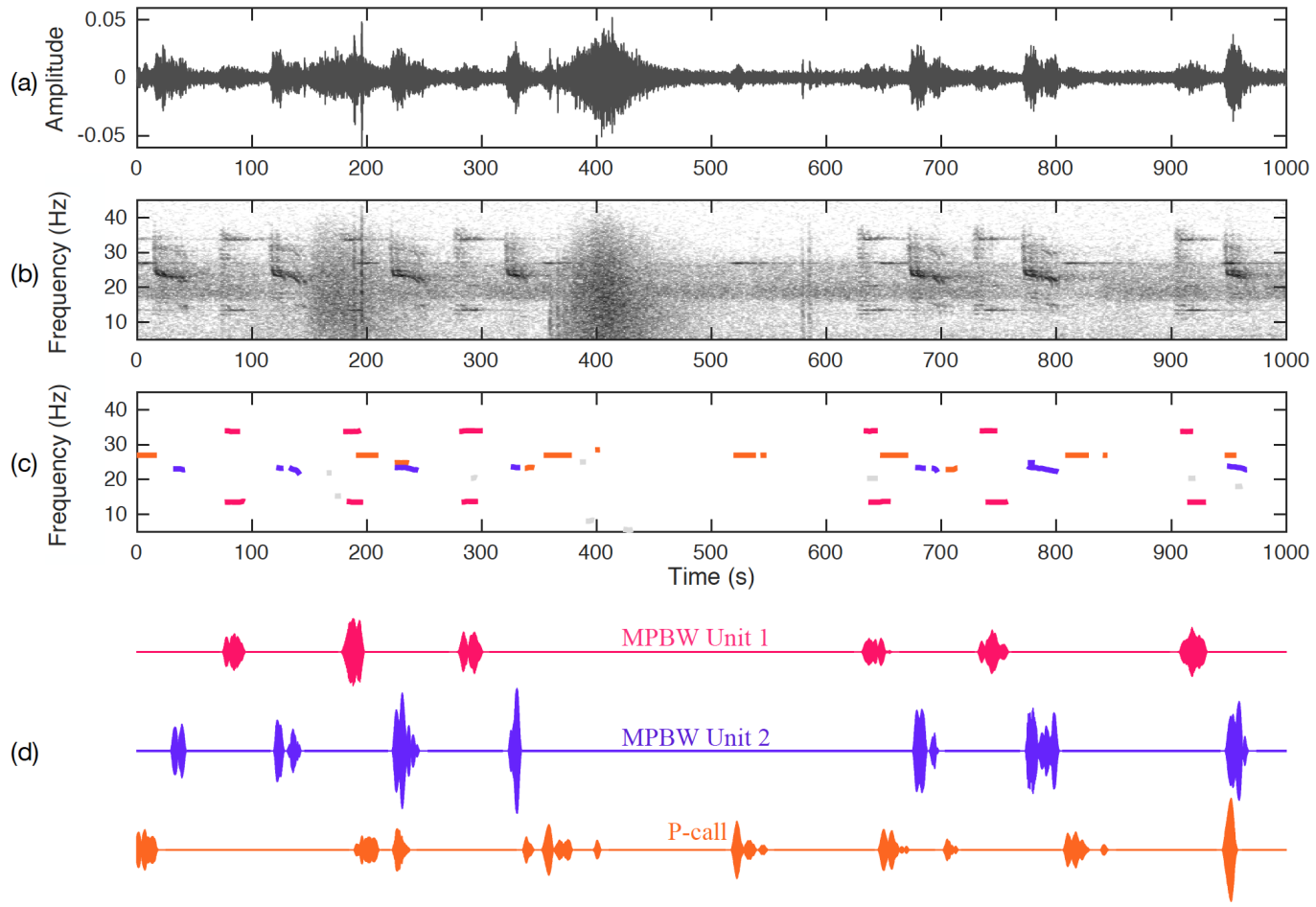

FIGURE 3: Forme d'onde (a) et spectrogramme (nfft $=$ 512, overlap $=$ 80%)(b) d'un enregistrement de calls de baleine bleue pygmée de Madagascar (MPBW) et de P-calls en présence d'un fort chorus de rorquals communs (bruit diffus "large bande" entre 17 et 27 Hz) ainsi que de deux évènements sismiques (autour de 180 s et 420 s). L'unité 1 de la MPBW est composée de deux tonales l'une à 13.5 l'autre à 34 Hz, l'unité 2 est un down-sweep autour de 24 Hz. Les P-calls signent à $\simeq$ 27 Hz. Sortie du détecteur de crêtes (c), dont la couleur correspond au type de call annoté. La forme d'onde des chants reconstruits est affichée en (d).

[4] F. Samaran, O. Adam, and C. Guinet, "Detection range modeling of blue whale calls in Southwestern Indian Ocean," *Applied Acoustics*, vol. 71, no. 11, pp. 1099–1106, 2010.

[5] L. Bouffaut, R. Dréo, V. Labat, A.-O. Boudraa, and G. Barruol, "Passive stochastic matched filter for antarctic blue whale call detection," *Journal of the Acoustical Society of America*, vol. 144, no. 2, pp. 955–965, 2018.

[6] D. Gillespie, "Detection and classification of right whale calls using an 'edge' detector operating on a smoothed spectrogram," *Canadian Acoustics*, vol. 32, no. 2, pp. 39–47, 2004.

[7] L. Bouffaut, S. Madhusudhana, V. Labat, A.-O. Boudraa, and H. Klinck, "Comparative study of tonal detectors for low frequency vocalizations of blue whales," *(under review) Journal of the Acoustical Society of America*, 2019.

[8] S. Madhusudhana, A. Gavrilov, and C. Erbe, "A generic system for the automatic extraction of narrowband signals in underwater audio," *Journal of the Acoustical Society of America*, vol. 140, no. 4, pp. 3182–3182, 2016.

[9] D. K. Mellinger and C. W. Clark, "Mobysound : A reference archive for studying automatic recognition of marine mammal sounds," *Applied Acoustics*, vol. 67, no. 11-12, pp. 1226–1242, 2006.

[10] S. C. Stähler, K. Sigloch, K. Hosseini, W. C. Crawford, G. Barruol, M. C. Schmidt-Aursch, M. Tsekhmistrenko, J.-R. Scholz, A. Mazzullo, and M. Deen, "Performance report of the RHUM-RUM ocean bottom seismometer network around La Réunion, western Indian Ocean," *Advances in Geosciences*, vol. 41, pp. 43–63, 2016.