Make the CPUs do the hard work - Automated acoustic feature extraction and visualization for marine ecoacoustics applications illustrated using marine mammal Passive Acoustic Monitoring datasets

Simone Cominelli[1], Nicolo' Bellin[2], Carissa Brown[1], Valeria Rossi[2], and Jack Lawson[3]

[1]Memorial University of Newfoundland
[2]Università degli Studi di Parma
[3]Fisheries and Oceans Canada Newfoundland and Labrador Region

August 25, 2022

Abstract

1. Passive Acoustic Monitoring is emerging as a solution for monitoring species and environmental change over large spatial and temporal scales. However, drawing rigorous conclusions based on acoustic recordings is challenging, as there is no consensus over which approaches and indices are best suited for characterizing marine acoustic environments. 2. We present an alternative to the use of ecoacoustic indices and describe the application of multiple machine learning techniques to the analysis of a large PAM dataset. We combine pre-trained acoustic classification models, dimensionality reduction, and random forest algorithms to demonstrate how machine-learned acoustic features capture different aspects of the marine environment. We processed two PAM databases and conducted 13 trials showing how acoustic features can be used to: i) discriminate between the vocalizations of marine mammals, beginning with high-level taxonomic groups, and extending to detecting differences between conspecifics belonging to distinct populations; ii) discriminating amongst different marine environments; and iii) detecting and monitoring anthropogenic and biological sound sources. 3. Acoustic features and their UMAP projections exhibited good performance in the classification of marine mammal vocalizations. Most of the taxonomic levels investigated here could be classified using the UMAP projections, apart from species that were underrepresented. Both anthropogenic (ships and airguns) and biological (humpback whales) sound sources could also be identified in field recordings. 4. We argue that acoustic feature extraction, visualization, and analysis allows the retention of most of the environmental information contained in PAM recordings, overcoming the limitations encountered when using ecoacoustics indices. Acoustic features are universal, permitting comparisons of results collected from multiple environments. Our approach can be used to simultaneously investigate the macro and micro characteristics of marine soundscapes, with a more objective method and with far less human effort.

INTRODUCTION

Abrupt changes in the ocean environment are increasing in frequency as climate change accelerates (Ainsworth et al., 2020), resulting in loss of key ecosystems (Sully et al., 2019), and shifts in endangered species' distributions (Plourde et al., 2019). Detecting such changes requires both historical and real-time (or near-real time) data made readily available to managers and decision-makers. Scientists and practitioners are being tasked with finding efficient solutions for monitoring environmental health and detecting incipient change (Gibb et al., 2019; Kowarski & Moors-Murphy, 2020). This challenge includes monitoring for changes in species' presence, abundance, distribution, and behaviour (Durette-Morin et al., 2019; Fleming et al., 2018; Root-Gutteridge et al., 2018), monitoring anthropogenic activity and disturbance levels (Gómez et al., 2018), monitoring the physical environment (Almeira & Guecha, 2019), and detecting harmful events (Rycyk et al., 2020), among others.

Environmental sounds provide a proxy to investigate ecological processes (Gibb et al., 2019; Rycyk et al., 2020), including exploring complex interactions between anthropogenic activity and biota (Erbe et al., 2019; Kunc et al., 2016). Sound provides useful information on environmental conditions and ecosystem health, allowing, for example, the rapid identification of disturbed coral reefs (Elise et al., 2019). In concert, numerous species (i.e., birds, mammals, fish, and invertebrates) rely on acoustic communication for foraging, mating and reproduction, habitat use and other ecological functions (Eftestøl et al., 2019; Kunc & Schmidt, 2019; Luo et al., 2015; Schmidt et al., 2014). Noise produced by anthropogenic activities (e.g., vehicles, stationary machinery, explosions) can interfere with such processes, affecting the health and reproductive success of multiple marine taxa (Kunc & Schmidt, 2019). In response to concerns about noise pollution, increasing effort is being invested in developing, testing, and implementing noise management measures in both terrestrial and marine environments. Consequently, Passive Acoustic Monitoring (PAM) has become a mainstream tool in biological monitoring (Gibb et al., 2019). PAM represents a set of techniques that are used for the systematic collection of acoustic recordings for environmental monitoring. It allows collecting large amounts of environmental information at multiple locations and over extended periods of time.

One of PAM's most common applications is in marine mammal monitoring and conservation. Marine mammals produce complex vocalizations that are species-specific (if not individually unique), and such vocalizations can be used when estimating species' distributions and habitat use (Durette-Morin et al., 2019; Kowarski & Moors-Murphy, 2020). PAM applications in marine mammal research span from the study of their vocalizations and behaviours (Madhusudhana et al., 2019; Vester et al., 2017) to assessing anthropogenic disturbance (Nguyen Hong Duc et al., 2021). PAM datasets can reach considerable sizes, particularly when recorded at high sampling rates, and projects often rely on experts to manually inspect the acoustic recordings for the identification of sounds of interest (Nguyen Hong Duc et al., 2021). For projects involving recordings collected over multiple months at different locations, conducting a manual analysis of the entire dataset can be prohibitive, and often only a relatively small portion of the acoustic recordings is subsampled for analysis.

At its core, studying acoustic environments is a signal detection and classification problem in which a large number of spatially and temporally overlapping acoustic energy sources need to be differentiated to better understand their relative contributions to the soundscape. Such an analytical process, termed acoustic scene classification (Geiger et al., 2013), is a key step in analysing environmental information collected by PAM recorders. Acoustic scenes can contain multiple overlapping sound sources, which generate complex combinations of acoustic events (Geiger et al., 2013). This definition overlaps with the ecoacoustics definition of soundscape (Farina & Gage, 2017), providing a bridge between the two fields, where a soundscape represents the total acoustic energy contained within an environment and consists of three intersecting sound sources: geological (i.e., geophony), biological (i.e., biophony), and anthropogenic (i.e., anthrophony). A goal of ecoacoustics is to understand how these sources interact and influence each other, with a particular focus on biological-anthropogenic acoustic interactions.

Automated acoustic analysis can overcome some of the limitations encountered in manual PAM analysis, allowing ecoacoustics researchers to explore full datasets (Houegnigan et al., 2017). Deep learning represents a novel set of computer-based artificial intelligence approaches which has profoundly changed biology and ecology research (Christin et al., 2019). Among the deep learning approaches, Convolutional Neural Networks (CNNs) have demonstrated high accuracy in performing image classification tasks, including the classification of spectrograms (i.e., visual representations of sound intensity across time and frequency) (Hershey et al., 2017; LeBien et al., 2020).

CNNs have been applied successfully to several ecological problems, and their use in ecology has been growing (Christin et al., 2019), such as to process camera trap images to identify species, age classes, numbers of animals, and to classify behaviour patterns (Lumini et al., 2019; Norouzzadeh et al., 2018; Tabak et al., 2019).

CNN's algorithms perform well for acoustic classification (Hershey et al., 2017), including the identification of a growing number of species vocalizations such as crickets and cicadas (Dong et al., 2018), birds and frogs (LeBien et al., 2020), fish (Mishachandar & Vairamuthu, 2021), and lately marine mammals (Usman et al., 2020). The latter include training neural networks for detecting North Atlantic right whale calls using a mix of real and synthetic data (Padovese et al., 2021), and the classification of sperm whale clicks (Bermant et al., 2019). Most CNN applications focus on species detection rather than a broader characterization of the acoustic environment. Furthermore, automated acoustic analysis algorithms often rely on supervised classification based on large datasets of known sounds (i.e., training datasets) used to train acoustic classifiers; creating training datasets is time-consuming and requires expert-driven manual classification of the acoustic data (Bittle & Duncan, 2013).

Recent developments in acoustic scene analysis demonstrate how the implementation of acoustic feature sets derived from CNNs, along with the use of dimensionality reduction, can improve our ability to understand ecoacoustics datasets while providing a common ground for analysing recordings collected across multiple environments and temporal scales (Clink & Klinck, 2020; Mishachandar & Vairamuthu, 2021; Sethi et al., 2020). Our study explores the application of acoustic scene analysis to two sets of PAM recordings containing marine mammal vocalizations (Fig 1). The first dataset, the *Watkins Marine Mammal Sound Database* (Woods Hole Oceanographic Institution and the New Bedford Whaling Museum; WMD hereafter), allowed us to test if acoustic features can be used for classifying marine mammal vocalizations according to multiple levels of taxonomic organization. The second dataset, which consisted of approximately 72 hours of recordings collected by Fisheries and Oceans Canada at two different locations within Placentia Bay (Newfoundland, Canada; PBD hereafter), allowed us to test if the acoustic features can be used to identify different sound sources, namely ships, seismic airguns, and humpback whales (*Megaptera novaengliae*).

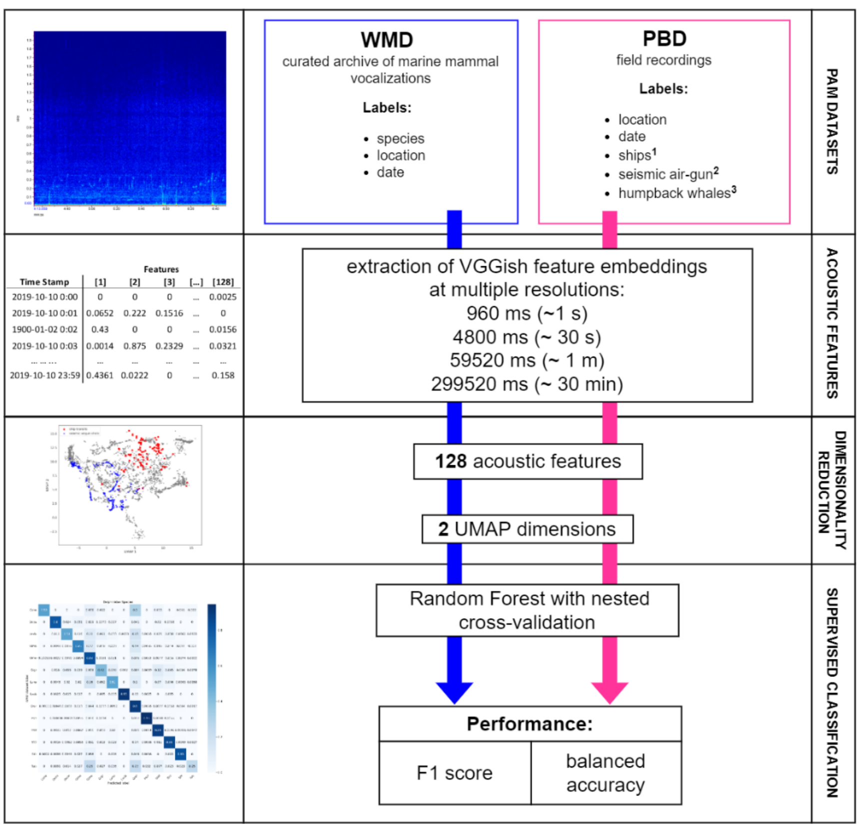


Figure 1. Analytical framework showing the different steps outlined in the Materials and Methods section. [1]labelled using ship band noise statistics; [2]labelled through visual inspection of spectrograms;[3]labelled using Google & NOAA humpback whale detector (Allen et al., 2021).

MATERIALS ANd METHODS

Data acquisition and preparation

We collected all records available in the Watkins Marine Mammal Database website listed under the "all cuts" page. We limited the analysis to 37 marine mammal species by discarding data for species with a low number of audio samples; we processed 17.1 hours of audio. For each audio file in the WMD the associated metadata included a label for the sound sources present in the recording (biological, anthropogenic, and environmental), as well as information related to the location and date of recording. We selected audio clips that contained a marine mammal as the main and only sound source present in the recording and labelled the vocalizations according to taxonomic group (*Odontocetae, Mysticetae, Otariidae,* and *Phocidae*), order, family, and species.

We created an additional label defining the population of origin for the orca (*Orcinus orca*) samples, which split them into five groups. The first three, *EN Atlantic, WN Atlantic* and *EN Pacific* , are recordings of orcas in the wild. *EN Atlantic* samples include orcas recorded in the Norwegian Sea and in a Norwegian fjord. *WN Atlanti* c samples include orcas recorded outside St. John's Harbour (Newfoundland, Canada) and orcas recorded approximately 130 km south of Martha's Vineyard (Massachusetts, U.S.). The *EN Atlantic* and *WN Atlantic* samples most likely contain a mix of two orca ecotypes (T1 and T2). *EN Pacific* samples included whales recorded in Saanich Inlet (British Columbia, Canada) and Dabob Bay (Washington, U.S.). These recordings could belong to three orca ecotypes (i.e., resident, offshore, and transient). The last two labels, *EN Atlantic - captive* and *EN Pacific - captive* , indicate recordings of captive whales Moby Doll, a resident orca captured in British Columbia, and Keiko, captured in Iceland (either a T1 or a T2 ecotype).

The Placentia Bay Database includes recordings collected by Fisheries and Oceans Canada at multiple stations within Placentia Bay (Newfoundland, Canada), from 2017 to 2020. From the PBD, we selected three days of recordings from summer 2019. The first two days (2019/08/10 and 2019/10/10) were collected by an AMAR G4 hydrophone (sensitivity: -165.02 dB re 1V/µPa at 250 Hz) deployed at 65 m of depth, approximately 13 km south of the town of Burin. The third day of recordings was collected by an AMAR G4 hydrophone (sensitivity: -164.92 dB re 1V/µPa at 250 Hz) deployed at 100 m of depth, approximately 2 km south of Red Island. Both hydrophones were set to operate following 15 min cycles, with the first 60 s sampled at 512 kHz, and the remaining 14 min sampled at 64 kHz. For the purpose of this study, we limited the analysis to the 64 kHz recordings. From the Burin deployment, we selected the 10th of August as it contained seismic airgun noise from oil and gas exploration activity happening in the Grand Banks, approximately 170 km south of the hydrophone deployment location. From the Red Island deployment, we selected the 26th of July, which contained ship transits and humpback whale vocalizations. Before proceeding with the analysis, all recordings were labelled by time stamp and location. All days contained humpback whale vocalizations.

Acoustic feature extraction

The audio files from the WMD and PBD databases were used as input for VGGish (Abu-El-Haija et al., 2016; Simonyan & Zisserman, 2014), a CNN developed and trained to perform general acoustic classification. VGGish was trained on the Youtube8M dataset, containing more than two million user-labelled audio-video files. Rather than focusing on the final output of the model (i.e., the assigned labels), here the model was used as a feature extractor (Sethi et al., 2020). VGGish converts audio input into a semantically meaningful vector consisting of 128 features. The model returns features at multiple resolution: ~1 s (960 ms); ~1 min (59'520 ms); ~5 min (299'520 ms). All of the visualizations and results pertaining to the WMD were prepared using the finest feature resolution of ~1 s. The visualizations and results pertaining to the PBD were prepared using the ~1 min features, except for the humpback whale detection test, which was conducted on the ~1 s features.

UMAP ordination and visualization

To allow for data visualization and to reduce the 128 features to two dimensions for further analysis, we applied Uniform Manifold Approximation and Projection (UMAP) to both datasets in full, and inspected the resulting plots (Figs. 2 to 7). UMAP is a non-linear dimensionality reduction algorithm based on the concept of topological data analysis which, unlike other dimensionality reduction techniques (e.g., tSNE), preserves both the local and global structure of multivariate datasets (McInnes et al., 2018).

The UMAP algorithm generates a low-dimensional representation of a multivariate dataset while maintaining the relationships between points in the global dataset structure (i.e., the 128 features extracted from VGGish). Each point in a UMAP plot in this paper represents an audio sample with duration of either ~ 1 sec or ~ 1 min. Each point in the two-dimensional UMAP space also represents a vector of 128 VGGish features. The nearer two points are in the plot space, the nearer the two points are in the 128-dimensional space, and thus the distance between two points in UMAP reflects the degree of similarity between two audio samples in our datasets. Areas with a high density of samples in UMAP space should, therefore, contain sounds with similar characteristics, and such similarity should decrease with increasing point distance. The visualizations and classification trials presented here illustrate how the two techniques (VGGish and UMAP) can be used together for marine ecoacoustics analysis.

Labelling sound sources

Sample labels were obtained with a mix of techniques: labels for the WMD records were obtained from the database metadata; for the PBD recordings, the start and end of seismic exploration was identified through manual inspection, ship presence was inferred from sound pressure levels (SPL) in the ship noise band (40-315 Hz) (Baldwin et al., 2021), and humpback whale presence was inferred using an acoustic detection model (Allen et al., 2021).

To label anthropogenic noise sources in the PBD, we first used PAMGuide (Merchant et al., 2015) to process the acoustic recordings. We computed broadband SPL (dB re 1 μPa) between 50 and 4,000 Hz (1 min resolution) as a global measure of sound pressure level in the dataset. As an indicator of ship noise, we computed the SPL between 40 and 315 Hz (i.e., ship band hereafter) at 1 min resolution. The ship band encompasses the 63, 150, and 250 Hz 1/3 octave bands (Baldwin et al., 2021), which are indicators of low-frequency ship noise levels (Merchant, et al., 2014). Samples that satisfied the following two conditions were considered as ship presences: 1) the ship band SPL was within 12 dB of the broadband SPL; 2) the 5 min mean ship band SPL was 3 dB above the global median SPL (i.e., computed on the full dataset) (Appendix S2.2). PBD samples collected near Burin on 08/10/2019 were inspected to identify the start and end of seismic airgun activity. All 1-min samples with a time stamp falling within the period of seismic exploration were marked as airgun noise present and contained multiple blasts.

Biological noise sources in the PBD recordings were processed using the humpback whale acoustic detector created by NOAA and Google (Allen et al., 2021), providing a model score for every ~1 s sample. The model returns scores ranging from 0 to 1 indicating the confidence in the predicted humpback whale presence. We used the results of this detection model to label the PBD samples according to presence of humpback whale vocalizations. We selected 0.8 as the minimum model score needed to declare a humpback present, while every sample with a score lower than 0.8 was labelled as an absence.

Label prediction performance

To predict labels from the acoustic features for both the WMD and the PBD datasets, we applied nested k-fold cross validation to a random forest model, with ten-folds in the outer loop, and five-folds in the inner loop. We selected nested cross validation as it allows model optimization of hyperparameters and performance evaluation in a single step. Models were trained either on the two UMAP dimensions, or on the full set of 128 acoustic features, depending on model performance. Model performance was evaluated using two metrics: F1 and balanced-accuracy scores, both on a scale from 0 to 1. The F1 score combines model recall and precision, favouring models with a high score in both metrics (Chinchor, 1992). Balanced accuracy is suited for measuring model performance when samples are highly imbalanced, and represents the average recall obtained for each model class (Brodersen et al., 2010). When the F1 and balanced accuracy scores indicated poor performance of the classifier, we repeated the trial using the 128 acoustic features instead of the two UMAP dimensions.

In total, we conducted 13 trials on the two databases, six on the WMD, and seven on the PBD (Table 1). The first WMD trial included building a classifier for *Mysticete* , *Odontocete* , and *Pinniped* . For the remaining five trials, we created subsets of the WMD and ran classifiers for: three *Mysticete* (*Balaenidae* ,*Balaenopteridae* , and *Eschrichtiidae*) and four*Odontocete* families (*Delphinidae* , *Monodontidae* ,*Phocoenidae* , and *Physteridae*); three*Balaenopteridae* species (minke, fin, and humpback whales), 14*Delphinidae* species (see Appendix S1 for a complete list); and three distinct orca populations. Species with less than 100 samples were removed from the analysis.

Trials on the PBD labels proceeded as follows: i) classification of hydrophone locations (i.e., Burin and Red Island); classification of anthropogenic noise sources, including ii-iii) seismic airguns and iv-v) ships; and presence of humpback whales using vi) the two UMAP dimensions and vii) the 128 acoustic features, respectively. Presences represented a very small fraction of the PBD (<0.003 %), leading to high class imbalance. We used two strategies to reduce class imbalance: we selected a subset of the PBD containing only hours with at least ten presences (this reduced the PBD dataset to 19 hours of PAM recordings); and then implemented a balanced random forest classifier (Lemaître et al., 2017) in place of the model used for the previous trials.

RESULTS

Watkins Marine Mammals Sounds Database

UMAP Visualizations

Our inspection of the UMAP 2D ordination plot of three large marine mammal taxonomic groups, *Mysticete, Odontocete* , and*Pinniped* , revealed a separation between *Mysticete* and*Odontocete* sounds (Fig. 2). However, the two groups overlapped in some areas of the plot, and *Pinniped* vocalization clustered close to the centre of the plot, scattered between the first two groups.

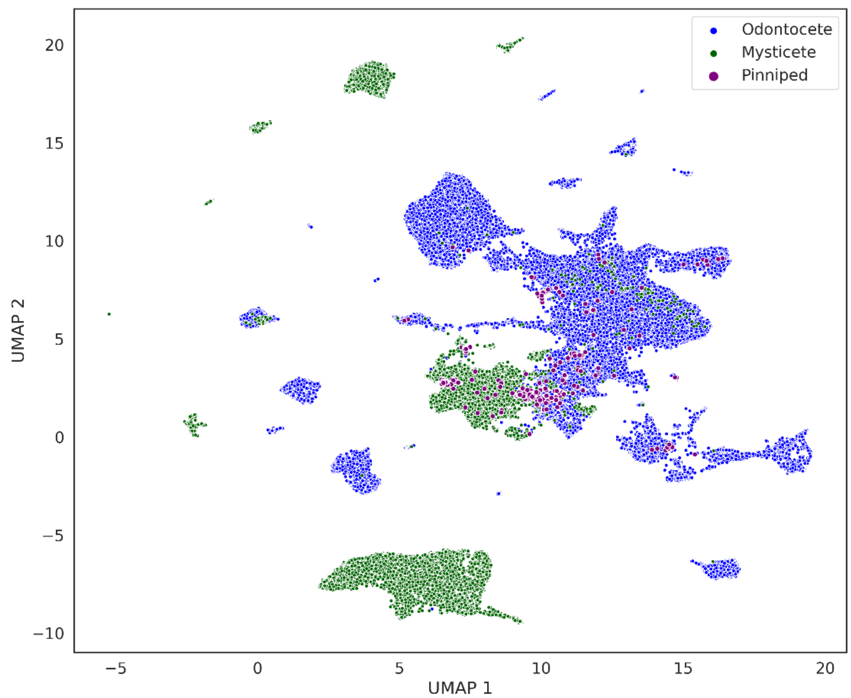


Figure 2. UMAP ordination of the WMD dataset with samples coloured according to three large taxonomic groups (Mysticete, Odontocete, and Pinniped). Pinniped sample points were plotted at double size to improve visualisation.

Within the *Mysticete* group, only three families contained enough samples to be considered for further analysis: *Balaenopteridae, Balaenidae,* and *Eschrichtiidae.* In the subsequent UMAP ordination, *Balaenidae* samples were almost completely overlapped with *Balaenopteridae* vocalizations, close to the plot centre (Fig. 3). *Eschrichtiidae* samples, the least represented label (i.e., the minority label) for the *Mysticete* , clustered in four distinct areas of the UMAP plot.

The *Odontocete* group was dominated by the *Physteridae*family, which represented the majority label for the subset, followed by*Delphinidae* and *Monodontidae* (Fig 4). *Phocoenidae*vocalisations were the minority label, and, similarly to*Eschrichtiidae,* samples belonging to this family formed small clusters scattered across the UMAP plot area.

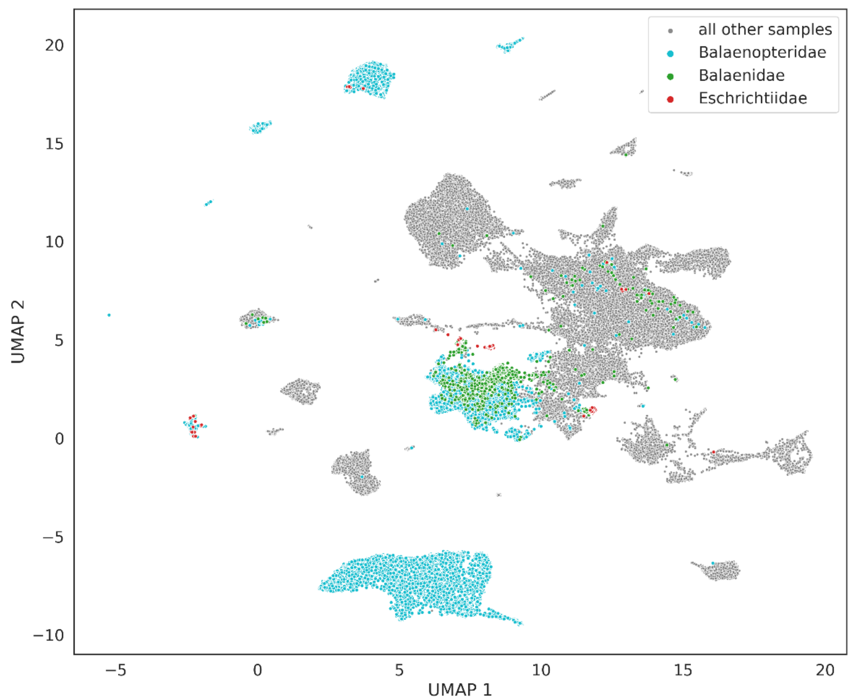


Figure 3. UMAP ordination of the WMD dataset with samples belonging to the Mysticete group coloured according to three families. All other samples (Odontocete and Pinniped) are marked in grey.

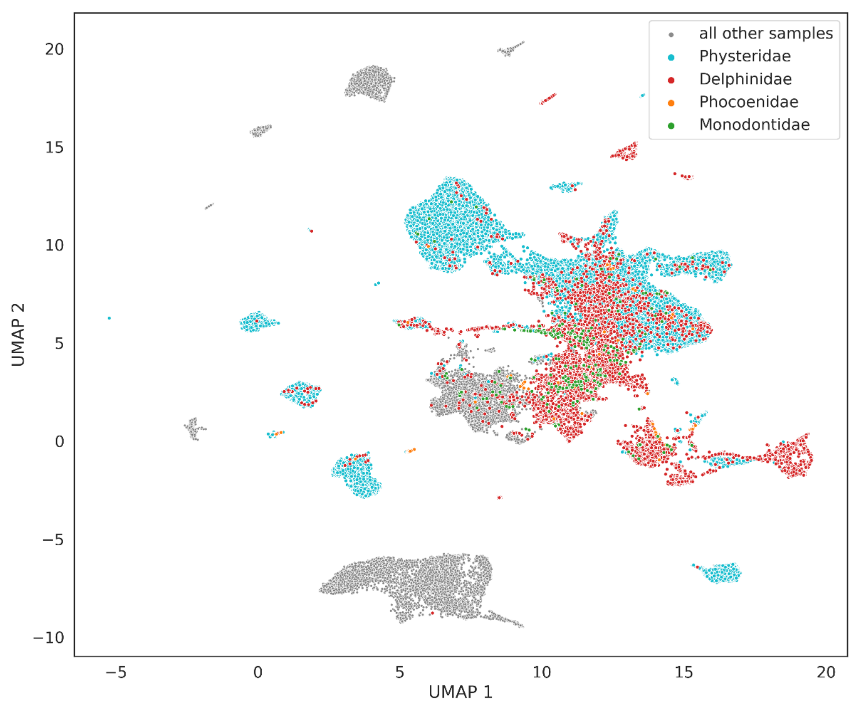


Figure 4. UMAP ordination of the WMD dataset with samples belonging to the Odontocete group coloured according to four families. All other samples (Mysticete and Pinniped) are marked in grey.

The labelled orca vocalizations (Fig. 5) showed separation between four of the five population labels, apart from *NE Pacific* orcas, the minority class of the group. *EW Atlantic* was the only label whose samples formed one large and distinct cluster. Samples from the two captive orcas (*EN Atlantic – captive* and *EN Pacific – captive*), formed two distinct clusters, while the *EW Atlantic*samples were scattered across a large area of the UMAP plot.

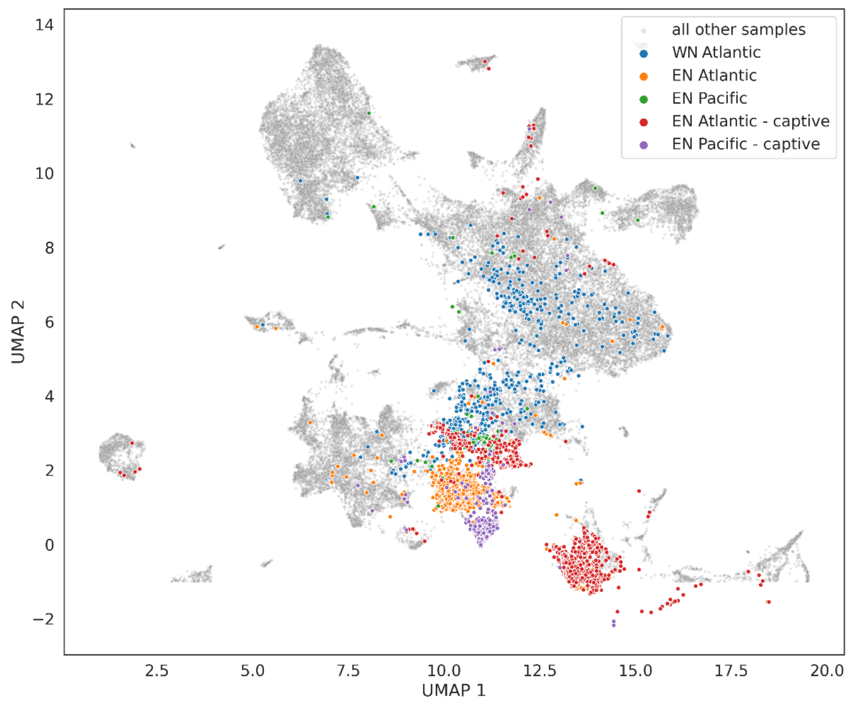


Figure 5. Detail of the WMD dataset UMAP ordination with samples belonging to Orcinus orca, coloured according to their population of origin and wild versus captive status, when recorded. All other samples are marked in grey.

UMAP label prediction performance

Model evaluation scores were above 0.7 for all of the WMD trials (Table 1), but with varying results depending on the specific label. The best classification results were obtained for *Balanopteridae* species (F1 = 0.998; balanced accuracy = 0.987), while the classifier built for*Delphinidae* species had the lowest performance (F1 = 0.829; balanced accuracy = 0.703). Classification accuracy varied across trials. For example, in the first trial, most *Mysticete* and Odontocete samples were correctly labelled, while 59% of the*Pinniped* samples were mislabelled. In the second trial, 99%, 74%, and 71% of the Balaenopteridae, *Eschrichtiidae* , and*Balaenidae* samples were correctly classified. Of the four*Odontocede* families, *Physteridae* , *Delphinidae* , and*Phocoenidae* , 99%, 90%, and 78% of the samples were correctly classified, respectively. Only 56% of the testing samples for the family *Monodontidae* were classified correctly.

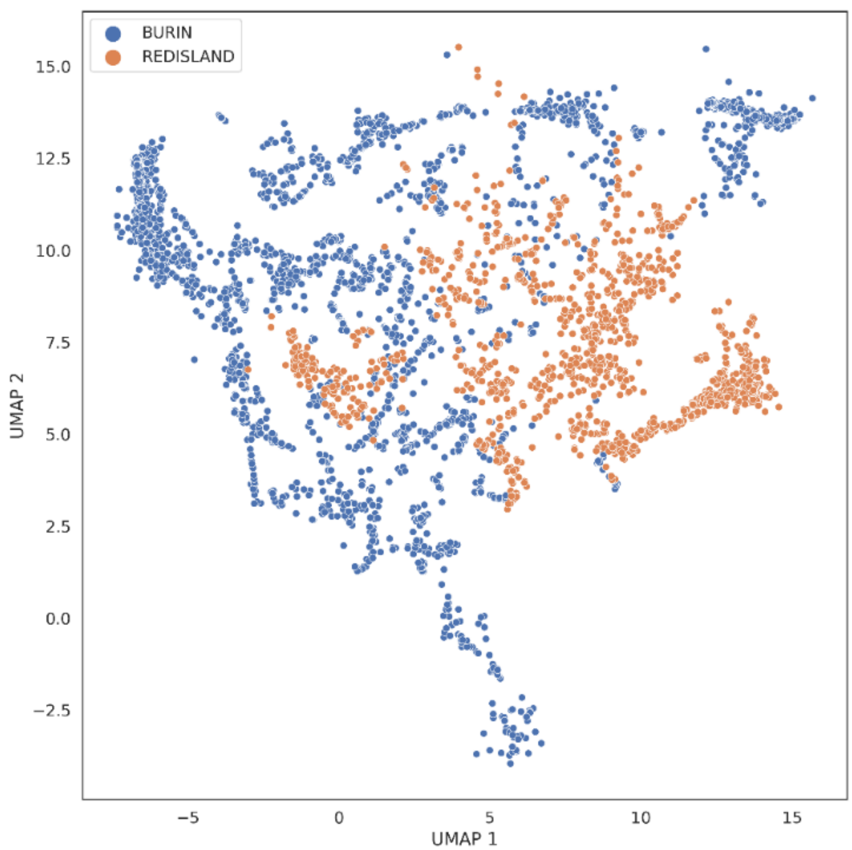
Table 1. k-fold nested cross-validation input and results. The table reports model features (X), labels (Y), and evaluation metrics (F1 score, Balanced Accuracy score). Best models, model hyperparameters, and scores per run can be found in appendix S1.

All of the three *Balaenoptera* species considered in the study were correctly classified in the vast majority of cases, with scores equal or above 98% of correct predictions. Eight of the 14*Delphinidae* species had 80% or more correct label predictions. Of the four labels tested for orcas, correct labels ranged from 87% (*WN Atlantic*) to 92% (*EN Atlantic*), except for the*EN Pacific* labels, with only 33% of the labels guessed correctly. Both model performance metrics reflected such class imbalances, with lower scores for models containing a mix of labels with low and high prediction accuracy. Balanced-accuracy scores provided a more conservative metric and were more sensitive to class imbalance than the F1 scores.

Placentia Bay Dataset

UMAP Visualizations

Our inspection of the UMAP ordination of the ˜1 min acoustic features of the two deployment locations: Burin and Red Island revealed two overlapping clusters, with samples from Burin predominantly distributed around the edges of the Red Island cluster (Fig. 6).



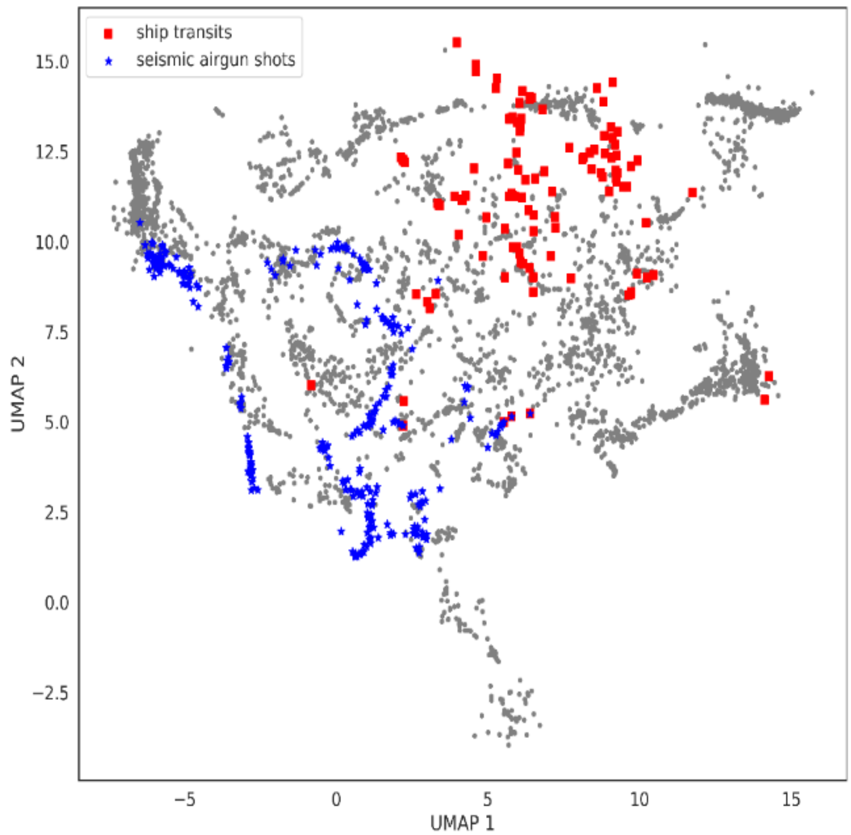


Figure 6. PBD dataset UMAP ordination at ˜1 min resolution. Samples grouped by hydrophone deployment location (left). Samples grouped by sound source (right). All other samples are shown in grey.

Samples labelled as seismic airgun noise and ship noise separated and occupied two distinct portions of the UMAP ordination plot (Fig. 6). A small number of samples from the two sources overlapped, indicating ship transits occurring during seismic exploration. However, we could not observe a clear distinction between presences and absences of the sources.

Lastly, we inspected how UMAP ordinated the ˜1 s acoustic features labelled by their chance of containing a humpback whale vocalization (Fig. 7). Detections per hour peaked at 1:00 and 13:00 and 15:00 for the Burin samples, while the Red Island samples showed a single distinct peak at 12:00 (Appendix S2.1). The ˜1 s resolution UMAP ordination showed a concentration of humpback whale detection scores (> 0.8) towards the right end of the plot, with samples densely aggregated along the second UMAP dimension. However, and similarly to the anthropogenic noise sources, we could not observe a clear separation between presences and absences.

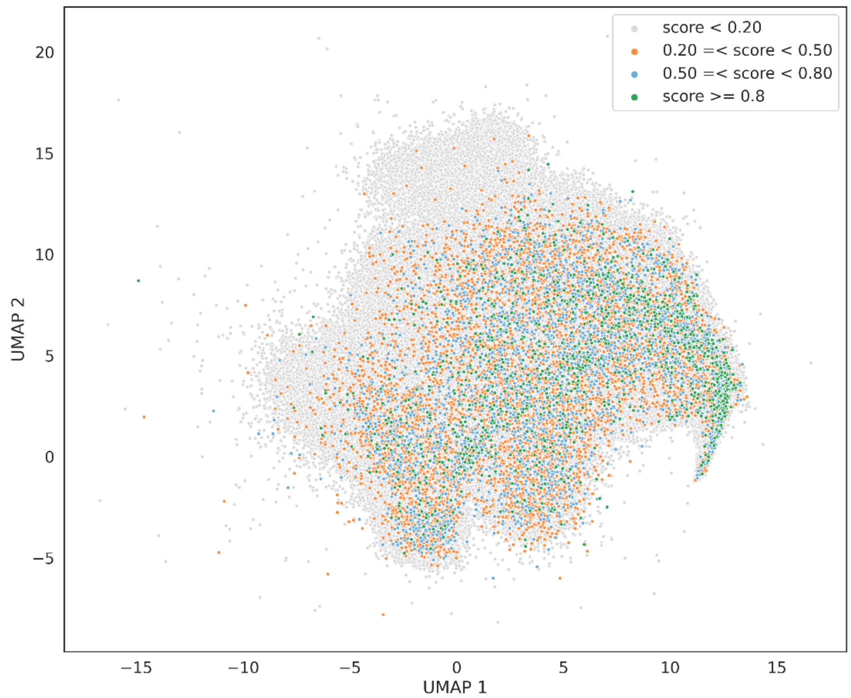


Figure 7. UMAP ordination at ~1 s resolution. Samples are coloured according to humpback whale detection probability (model scores). Scores above or equal to 0.8 were considered as presences.

Label prediction performance

Balanced accuracy scores for the 1-min UMAP dimensions were high (> 0.85) for the location label (Table 1). Of the samples labelled as 'Burin' and 'Red Island', 94% and 95% were correctly identified using the UMAP dimensions, respectively. Scores for seismic airgun presence were also high; however, model sensitivity was poor (58.3%), meaning that true positive and false negative predictions occurred with almost equal frequency. Repeating model training using the 128 acoustic features improved performance, and resulted in a drop of both false negatives and false positives. The ship presence classifier trained on the two UMAP dimensions showed a balanced accuracy score of 0.7, with only 33% of samples being correctly identified as presences. The acoustic features classifier displayed a higher balanced accuracy score (0.86), and the number of correctly predicted presences, although still low, increased to 58%.

The random forest classifiers for humpback whale presence trained on the two UMAP dimensions showed the lowest F1 and balanced accuracy score (0.59 and 0.62, respectively), resulting in a large number of mislabelled samples. Once again, repeating model fitting using the acoustic features improved model performance. Training the classifier on the 128 dimensions resulted in increased balanced accuracy score, mainly due to a dramatic increase in classifier sensitivity (93.9%) when compared to the performance of the classifier trained on UMAP dimensions (<0.001%).

Confusion matrices for the WMD and PBD cross validation runs are reported in Appendix S1.

Discussion

Managing the wellbeing of ecosystems requires identifying when and where human activities are impacting species' occurrence, movement, and behaviour. PAM is a useful approach for the detection of both large-and small-scale changes in urban and wild environments, as it allows for continuous and prolonged ecosystem monitoring. Challenges in employing PAM as a standard monitoring tool arise after data collection, when researchers and practitioners need to quickly extract useful information from large acoustic datasets, to understand when and where management actions are needed to preserve the well-being of ecosystems. The relatively new field of ecoacoustics provides the theoretical background for linking specific characteristics of the acoustic environment to biodiversity and ecosystem health. However, identifying a common analytical approach has been an obstacle to the broad application of ecoacoustics theory so far, and most studies employing ecoacoustics indices are not suited for replicability and comparison.

We addressed these problems by linking marine ecoacoustics assessment to the realms of machine learning and dimensionality reduction. We applied a deep-learning approach to characterize the biological and anthropogenic components of marine acoustic environments, and we illustrated how acoustic features derived from a pre-trained Convolutional Neural Network capture both the coarse and fine-grained structure of large PAM datasets. These methods can be applied to a broad range of marine and terrestrial systems.

Our analyses revealed several applications for inferring population- and location-specific information from acoustic datasets. When datasets are already labelled and focused on a specific taxon, such as the WMD, we found that acoustic features were particularly suited for the discrimination of marine mammal vocalizations. Understanding the evolution of vocal diversity and the role of vocalizations in the ecology of a species is one of the key objectives of bioacoustics research (Luís et al., 2021). Full acoustic repertoires are not available for most species, as building comprehensive lists of vocalizations requires considerable research effort. Here we show how a general acoustic classification model (VGGish) used as a feature extractor allows us to detect differences and similarities among marine mammal species, without requiring prior knowledge on the species' vocal repertoires. Our results for orcas are of particular interest, as they provide insights on the vocal similarities and differences between distinct populations of the same species. A large number of orca call samples labelled as *EN Pacific* were classified as *WN Atlantic* whales using the methodology in this study. Orcas show both genetic divergence and differences in call frequency that are more pronounced for sympatric ecotypes than whales found in different ocean basins (Filatova et al., 2015). Although we cannot consider the artefactual conflation of *EN Pacific* orcas with *NW Atlantic* orcas in the WMD as definitive evidence of convergence in vocal behaviour, we suggest that this aspect should be further investigated, perhaps using more recent recordings of these different orca populations.

More than 60 different ecoacoustic indices are being employed as descriptors of terrestrial soundscapes (Bradfer-Lawrence et al., 2019), making the search for indices that are successfully measuring biodiversity across widely variable environments very challenging (Minello et al., 2021). So far, ecoacoustic indices have been applied to marine environments with little success (Bohnenstiehl et al., 2018). Due to higher sound propagation efficiency, marine acoustic environments can receive acoustic energy from many sources with some that are hundreds of kilometres distant, making them more complex to study than terrestrial environments. Accordingly, the biases shown by acoustic indices measuring terrestrial species diversity (Eldridge et al., 2018; Fairbrass et al., 2017; Heath et al., 2021) are amplified when transferred to the study of marine environments (Bohnenstiehl et al., 2018; Dimoff et al., 2021; Minello et al., 2021).

Machine learned acoustic features are a promising alternative to the use of ecoacoustics indices for monitoring terrestrial biodiversity (Heath et al., 2021; Sethi et al., 2020). In this study, we show how this approach can also be extended to the study of marine soundscapes. The derived acoustic features were successful in discriminating between two different marine environments that differed in type and intensity of anthropic activity: recordings collected in Burin were dominated by distant seismic airgun pulses in the low frequency range, and the Red Island hydrophone recordings were characterized by frequent ship noise. Both sites yielded recordings of humpback whale vocalizations, and our results show that machine-learned acoustic features can be employed for detecting marine mammal sounds across different acoustic contexts. Machine-learned acoustic features respond to multiple marine sound sources, and can be employed successfully for investigating both the biological and anthropic components of marine soundscapes.

Reducing acoustic features to two UMAP dimensions, however, resulted in poorly performing classifiers for three sets of labels: airgun noise presence, ship presence, and humpback whale presence. In all three cases, repeating the analysis on a larger set of 128 features improved model performance at the cost of increased processing time. The best models used as little as two features, and as many as 64, whereas classifiers based on the full 128 features were selected as best models for all iterations of the humpback whale classifier (Appendix S1). This indicates that the number of acoustic features could be significantly reduced in some instances, thus reducing processing time and virtual memory requirements. The poor performance observed in the UMAP ship presence classifiers could be partly due to the approach adopted for labelling presences and to the fact that ship noise was almost ubiquitous in the Red Island recordings. Most samples collected at the Red Island deployment location were more than 3 dB higher than the full dataset median, but only a fraction of such samples contributed to the broadband SPL (Appendix S2.2), indicating that ship presence may have been underestimated. As an alternative, using records of vessel positions obtained from the Automatic Identification System (AIS) as an indicator of ship presence may improve model performance, at the cost of underestimating the presence of small vessels, which are rarely equipped with AIS.

Acoustic features have been shown to overcome many of the limitations of ecoacoustics indices; for example, acoustic features outperform common ecoacoustic indices in discriminating different environmental characteristics (Sethi et al., 2020). Furthermore, acoustic features are resilient to audio file compression and reduction of Nyquist frequency, and provide results that are independent from type of recorders deployed and choices relative to the temporal fragmentation of acoustic datasets (Heath et al., 2021; Sethi et al., 2020). Here, we show that acoustic features and UMAP dimensions allow for the comprehensive exploration of marine PAM datasets. Features can be used to train classification models focusing on biological and anthropogenic sound sources and allow for fine-grain comparison of marine mammal vocalizations.

Two limitations persist. VGGish, the CNN used to extract the acoustic features, is pre-trained on audio files with a sampling rate of 16 kHz, resulting in a Nyquist frequency of 8 kHz. This is sufficient to capture low frequency vocalizations but reduces its ability to discriminate high-frequency sounds. Nonetheless, we were able to correctly classify both high- and low-frequency vocalizations in the WMD examples, including *Phocoenidae* sounds, a family that includes species that can produce sounds up to 150 kHz. A second limitation is that acoustic features are not a plug and play product, as establishing links between features and relevant ecological variables requires additional analyses, while ecoacoustic indices are designed as measures of specific environmental characteristics.

By presenting a set of examples focused on marine mammals, we have demonstrated the benefits and challenges of implementing acoustic features as descriptors of marine acoustic environments. Our future research will extend feature extraction and testing to full PAM datasets spanning several years and inclusive of multiple hydrophone deployment locations. Other aspects warranting further investigation are how acoustic features perform when the objective is discriminating vocalizations of individuals belonging to the same species or population, as well as their performance in identifying samples with multiple active sound sources.

Acoustic features are abstract representations of PAM recordings which preserve the original structure and underlying relationships between the original samples, and, at the same time, are a broadly applicable set of metrices that can be used to answer ecoacoustics, ecology, and conservation questions. As such, they can help us understand how natural systems interact with, and respond to, anthropogenic pressures across multiple environments. Lastly, the universal nature of acoustic features analysis could help bridge the gap between terrestrial and marine soundscape research. This approach could deepen our understanding of natural systems by enabling multi-system environmental assessments, allowing researchers to investigate and monitor, for example, how stressor-induced changes in one system may manifest in another. And these benefits accrue from an approach that is more objective than manual analyses and requires far less human effort.

ACKNOWLEDGEMENTS

This project was funded by the Species at Risk, Oceans Protection Plan, and Marine Ecosystem Quality programmes of the Department of Fisheries and Oceans Canada, Newfoundland and Labrador Region, by Memorial University of Newfoundland and Labrador, and by the Ph.D. program in Evolutionary Biology and Ecology (University of Parma, agreement with University of Ferrara and University of Firenze). Simone Cominelli was also supported by the TD Bank Bursary for Environmental Studies. Nicolo' Bellin was also supported by the 'COMP-HUB' Initiative, funded by the 'Departments of Excellence' Project of the Italian Ministry for Education, University and Research (MIUR). We would like to express our gratitude to the curators of the Watkins Marine Mammal Sound Database as we believe that open access databases are incredibly relevant to the development of global monitoring of natural systems. We thank all the graduate students of the Northern EDGE Lab for their support, and their effort in creating a welcoming and inclusive work environment. Lastly, we would like to thank Madelyn Swackhamer, Sean Comeau, Lee Sheppard, Greg Furey, and Andrew Murphy from DFO's Marine Mammal Section for collecting the data used in this study, providing detailed information about the hydrophone deployments, and for their help and support with accessing DFO's PAM databases.

CONFLICT OF INTEREST STATEMENT

The authors declare that there are no conflicts of interest.

AUTHORS' CONTRIBUTIONS

Simone Cominelli and Nicolo' Bellin developed the concepts and methodology described here, acquired the necessary databases, ran the analysis, and prepared the first draft of the manuscript. Dr. Carissa Brown and Dr. Valeria Rossi supervised the two main authors (Simone Cominelli and Nicolo' Bellin) throughout the preparation of the manuscript, and provided space and equipment for conducting the research. Dr. Jack Lawson provided access to DFO's PAM database, provided input during the development of the methodology, and reviewed analysis results. All authors contributed critically to the drafts and final submission, and gave approval for publication.

DATA AVAILABILITY

Scripts to reproduce the images and analysis described here, as well as sample wav files, and tables containing all acoustic features and their labels are available for download as Jupiter Notebooks through Dryad:

Link_to_repository_here

REFERENCES

Abu-El-Haija, S., Kothari, N., Lee, J., Natsev, P., Toderici, G., Varadarajan, B., & Vijayanarasimhan, S. (2016). *YouTube-8M:* a large-scale video classification benchmark. Retrived from: https://doi.org//10.48550/arXiv.1609.08675

Ainsworth, T. D., Hurd, C. L., Gates, R. D., & Boyd, P. W. (2020). How do we overcome abrupt degradation of marine ecosystems and meet the challenge of heat waves and climate extremes? *Global Change Biology* , *26* (2), 343–354. https://doi.org/10.1111/gcb.14901

Allen, A. N., Harvey, M., Harrell, L., Jansen, A., Merkens, K. P., Wall, C. C., Cattiau, J., & Oleson, E. M. (2021). A convolutional neural network for automated detection of humpback whale song in a diverse, long-term passive acoustic dataset. *Frontiers in Marine Science* ,*8* . https://doi.org/10.3389/fmars.2021.607321

Almeira, J., & Guecha, S. (2019). Dominant power spectrums as a tool to establish an ecoacoustic baseline in a premontane moist forest.*Landscape and Ecological Engineering* , *15* (1), 121–130. https://doi.org/10.1007/s11355-018-0355-0

Kowarski, K.A., Wilson, C. C., Delarue, J. J. Y., Martin, S. B. (2021).. Flemish Pass acoustic monitoring: ambient characterization and marine mammal monitoring near a mobile offshore drilling unit. Technical report by JASCO Applied Scienses for Wood Environment and Infrastructure Solutions.

Bermant, P. C., Bronstein, M. M., Wood, R. J., Gero, S., & Gruber, D. F. (2019). Deep machine learning techniques for the detection and classification of sperm whale bioacoustics. *Scientific Reports* ,*9* (1), 1–10. https://doi.org/10.1038/s41598-019-48909-4

Bittle, M., & Duncan, A. (2013). A review of current marine mammal detection and classification algorithms for use in automated passive acoustic monitoring. *Proceedings of Acoustics, 2013,* 1–8. http://www.acoustics.asn.au/conference_proceedings/AAS2013/papers/p64.pdf

Bohnenstiehl, D. R., Lyon, R. P., Caretti, O. N., Ricci, S. W., & Eggleston, D. B. (2018). Investigating the utility of ecoacoustic metrics in marine soundscapes. *Journal of Ecoacoustics* ,*2* (2), 1–1. https://doi.org/10.22261/JEA.R1156L

Bradfer-Lawrence, T., Gardner, N., Bunnefeld, L., Bunnefeld, N., Willis, S. G., & Dent, D. H. (2019). Guidelines for the use of acoustic indices in environmental research. *Methods in Ecology and Evolution* ,*10* (10), 1796–1807. https://doi.org/10.1111/2041-210X.13254

Brodersen, K. H., Ong, C. S., Stephan, K. E., & Buhmann, J. M. (2010). The balanced accuracy and its posterior distribution. *20th International Conference on Pattern Recognition* , 3121–3124. https://doi.org/10.1109/ICPR.2010.764

Chinchor, N. (1992). MUC-4 evaluation metrics. *Proceedings of the 4th Conference on Message Understanding - MUC4 '92* , 22. https://doi.org/10.3115/1072064.1072067

Christin, S., Hervet, É., & Lecomte, N. (2019). Applications for deep learning in ecology. *Methods in Ecology and Evolution* , 10(10) pp. 1632–1644). https://doi.org/10.1111/2041-210X.13256

Clink, D. J., & Klinck, H. (2020). Unsupervised acoustic classification of individual gibbon females and the implications for passive acoustic monitoring. *Methods in Ecology and Evolution* ,12(2), 328-341. https://doi.org/10.1111/2041-210X.13520

Dimoff, S. A., Halliday, W. D., Pine, M. K., Tietjen, K. L., Juanes, F., & Baum, J. K. (2021). The utility of different acoustic indicators to describe biological sounds of a coral reef soundscape. *Ecological Indicators* , *124* . https://doi.org/10.1016/j.ecolind.2021.107435

Dong, X., Yan, N., & Wei, Y. (2018). Insect sound recognition based on convolutional neural network. *2018 IEEE 3rd International Conference on Image, Vision and Computing (ICIVC)* , 855–859. https://doi.org/10.1109/ICIVC.2018.8492871

Durette-Morin, D., Davies, K. T. A., Johnson, H. D., Brown, M. W., Moors-Murphy, H., Martin, B., & Taggart, C. T. (2019). Passive acoustic monitoring predicts daily variation in North Atlantic right whale presence and relative abundance in Roseway Basin, Canada. *Marine Mammal Science* , *35* (4), 1280–1303. https://doi.org/10.1111/mms.12602

Eftestøl, S., Flydal, K., Tsegaye, D., & Colman, J. E. (2019). Mining activity disturbs habitat use of reindeer in Finnmark, Northern Norway.*Polar Biology* , *42* (10), 1849–1858. https://doi.org/10.1007/s00300-019-02563-8

Eldridge, A., Guyot, P., Moscoso, P., Johnston, A., Eyre-Walker, Y., & Peck, M. (2018). Sounding out ecoacoustic metrics: Avian species richness is predicted by acoustic indices in temperate but not tropical habitats. *Ecological Indicators* , *95* (1), 939–952. https://doi.org/10.1016/j.ecolind.2018.06.012

Elise, S., Bailly, A., Urbina-Barreto, I., Mou-Tham, G., Chiroleu, F., Vigliola, L., Robbins, W. D., & Bruggemann, J. H. (2019). An optimised passive acoustic sampling scheme to discriminate among coral reefs' ecological states. *Ecological Indicators* , *107* (February), 105627. https://doi.org/10.1016/j.ecolind.2019.105627

Erbe, C., Marley, S. A., Schoeman, R. P., Smith, J. N., Trigg, L. E., & Embling, C. B. (2019). The effects of ship noise on marine mammals—A review. *Frontiers in Marine Science* , 6, *6* 06. https://doi.org/10.3389/fmars.2019.00606

Fairbrass, A. J., Rennert, P., Williams, C., Titheridge, H., & Jones, K. E. (2017). Biases of acoustic indices measuring biodiversity in urban areas. *Ecological Indicators* , *83* , 169–177. https://doi.org/10.1016/j.ecolind.2017.07.064

Farina, A., & Gage, S. H. (2017). Ecoacoustics: the ecological role of sounds. In A. Farina & S. H. Gage (Eds.), *Ecoacoustics: The ecological role of sounds, 1-366* . John Wiley & Sons, Ltd. https://doi.org/10.1002/9781119230724

Filatova, O. A., Miller, P. J. O., Yurk, H., Samarra, F. I. P., Hoyt, E., Ford, J. K. B., Matkin, C. O., & Barrett-Lennard, L. G. (2015). Killer whale call frequency is similar across the oceans, but varies across sympatric ecotypes. *The Journal of the Acoustical Society of America* , *138* (1), 251–257. https://doi.org/10.1121/1.4922704

Fleming, A. H., Yack, T., Redfern, J. V., Becker, E. A., Moore, T. J., & Barlow, J. (2018). Combining acoustic and visual detections in habitat models of Dall's porpoise. *Ecological Modelling* ,*384* , 198–208. https://doi.org/10.1016/j.ecolmodel.2018.06.014

Geiger, J. T., Schuller, B., & Rigoll, G. (2013). Large-scale audio feature extraction and SVM for acoustic scene classification.*2013 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics* , 1–4. https://doi.org/10.1109/WASPAA.2013.6701857

Gibb, R., Browning, E., Glover-Kapfer, P., & Jones, K. E. (2019). Emerging opportunities and challenges for passive acoustics in ecological assessment and monitoring. *Methods in Ecology and Evolution* , *10* (2), 169–185. https://doi.org/10.1111/2041-210X.13101

Gomez, W. E., Isaza, C. V, & Daza, J. M. (2018). Identifying disturbed habitats: A new method from acoustic indices. *Ecological Informatics* , *45* , 16–25. https://doi.org/10.1016/j.ecoinf.2018.03.001

Heath, B. E., Sethi, S. S., Orme, C. D. L., Ewers, R. M., & Picinali, L. (2021). How index selection, compression, and recording schedule impact the description of ecological soundscapes. *Ecology and Evolution* , *11* (19), 13206–13217. https://doi.org/10.1002/ece3.8042

Hershey, S., Chaudhuri, S., Ellis, D. P. W., Gemmeke, J. F., Jansen, A., Moore, R. C., Plakal, M., Platt, D., Saurous, R. A., Seybold, B., Slaney, M., Weiss, R. J., & Wilson, K. (2017). CNN architectures for large-scale audio classification. *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* , 131–135. https://doi.org/10.1109/ICASSP.2017.7952132

Houegnigan, L., Safari, P., Nadeu, C., Van Der Schaar, M., & Andre, M. (2017). A novel approach to real-time range estimation of underwater acoustic sources using supervised machine learning. *OCEANS 2017 - Aberdeen* , 1–5. https://doi.org/10.1109/OCEANSE.2017.8084914

Kowarski, K. A., & Moors-Murphy, H. (2020). A review of big data analysis methods for baleen whale passive acoustic monitoring.*Marine Mammal Science* , *37(2)* , 652–673. https://doi.org/10.1111/mms.12758

Kunc, H. P., McLaughlin, K. E., & Schmidt, R. (2016). Aquatic noise pollution: implications for individuals, populations, and ecosystems.*Proceedings of the Royal Society B: Biological Sciences* ,*283* (1836), 20160839. https://doi.org/10.1098/rspb.2016.0839

Kunc, H. P., & Schmidt, R. (2019). The effects of anthropogenic noise on animals: A meta-analysis. *Biology Letters* , *15* (11), 20190649. https://doi.org/10.1098/rsbl.2019.0649

LeBien, J., Zhong, M., Campos-Cerqueira, M., Velev, J. P., Dodhia, R., Ferres, J. L., & Aide, T. M. (2020). A pipeline for identification of bird and frog species in tropical soundscape recordings using a convolutional neural network. *Ecological Informatics* , *59* , 101113. https://doi.org/10.1016/j.ecoinf.2020.101113

Lemaitre, G., Nogueira, F., & Aridas, C. K. (2017). Imbalanced-learn: a python toolbox to tackle the curse of imbalanced datasets in machine learning. *Journal of Machine Learning Research* , *18* (1), 559–563. http://jmlr.org/papers/v18/16-365.html

Luis, A. R., Collado, L. J. M., Gospić, N. R., Gridley, T., Papale, E., & Azevedo, A. (2021). Vocal universals and geographic variations in the acoustic repertoire of the common bottlenose dolphin. *Scientific Reports* , *11* (1), 1–9. https://doi.org/10.1038/s41598-021-90710-9

Lumini, A., Nanni, L., & Maguolo, G. (2019). Deep learning for plankton and coral classification. *Applied Computing and Informatics* . https://doi.org/10.1016/j.aci.2019.11.004

Luo, J., Siemers, B. M., & Koselj, K. (2015). How anthropogenic noise affects foraging. *Global Change Biology* , *21* (9), 3278–3289. https://doi.org/10.1111/gcb.12997

Madhusudhana, S. K., Chakraborty, B., & Latha, G. (2019). Humpback whale singing activity off the Goan coast in the Eastern Arabian Sea.*Bioacoustics* , *28* (4), 329–344. https://doi.org/10.1080/09524622.2018.1458248

McInnes, L., Healy, J., & Melville, J. (2018). *UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction* . Retrived from: http://arxiv.org/abs/1802.03426

Merchant, N. D., Fristrup, K. M., Johnson, M. P., Tyack, P. L., Witt, M. J., Blondel, P., & Parks, S. E. (2015). Measuring acoustic habitats.*Methods in Ecology and Evolution* , *6* (3), 257–265. https://doi.org/10.1111/2041-210X.12330

Merchant, N. D., Pirotta, E., Barton, T. R., & Thompson, P. M. (2014). Monitoring ship noise to assess the impact of coastal developments on marine mammals. *Marine Pollution Bulletin* , *78* (1–2), 85–95. https://doi.org/10.1016/j.marpolbul.2013.10.058

Minello, M., Calado, L., & Xavier, F. C. (2021). Ecoacoustic indices in marine ecosystems: a review on recent developments, challenges, and future directions. *ICES Journal of Marine Science* , *78* (9), 3066–3074. https://doi.org/10.1093/icesjms/fsab193

Mishachandar, B., & Vairamuthu, S. (2021). Diverse ocean noise classification using deep learning. *Applied Acoustics* ,*181* , 108141. https://doi.org/10.1016/j.apacoust.2021.108141

Nguyen Hong Duc, P., Cazau, D., White, P. R., Gérard, O., Detcheverry, J., Urtizberea, F., & Adam, O. (2021). Use of ecoacoustics to characterize the marine acoustic environment off the North Atlantic French Saint-Pierre-et-Miquelon Archipelago. *Journal of Marine Science and Engineering* , *9* (2), 177. https://doi.org/10.3390/jmse9020177

Norouzzadeh, M. S., Nguyen, A., Kosmala, M., Swanson, A., Palmer, M. S., Packer, C., & Clune, J. (2018). Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning.*Proceedings of the National Academy of Sciences of the United States of America* , *115* (25), E5716–E5725. https://doi.org/10.1073/pnas.1719367115

Padovese, B., Frazao, F., Kirsebom, O. S., & Matwin, S. (2021). Data augmentation for the classification of North Atlantic right whales upcalls. *The Journal of the Acoustical Society of America* ,*149* (4), 2520–2530. https://doi.org/10.1121/10.0004258

Plourde, S., Lehoux, C., Johnson, C. L., Perrin, G., & Lesage, V. (2019). North Atlantic right whale (*Eubalaena glacialis*) and its food: (I) a spatial climatology of *Calanus* biomass and potential foraging habitats in Canadian waters. *Journal of Plankton Research* , *41* (5), 667–685. https://doi.org/10.1093/plankt/fbz024

Root-Gutteridge, H., Cusano, D. A., Shiu, Y., Nowacek, D. P., Van Parijs, S. M., & Parks, S. E. (2018). A lifetime of changing calls: North Atlantic right whales, *Eubalaena glacialis* , refine call production as they age. *Animal Behaviour* , *137* , 21–34. https://doi.org/10.1016/j.anbehav.2017.12.016

Rycyk, A. M., Tyson Moore, R. B., Wells, R. S., McHugh, K. A., Berens McCabe, E. J., & Mann, D. A. (2020). Passive Acoustic Listening Stations (PALS) show rapid onset of ecological effects of harmful algal blooms in real time. *Scientific Reports* , *10* (1), 17863. https://doi.org/10.1038/s41598-020-74647-z

Schmidt, R., Morrison, A., & Kunc, H. P. (2014). Sexy voices - no choices: Male song in noise fails to attract females. *Animal Behaviour* , *94* , 55–59. https://doi.org/10.1016/j.anbehav.2014.05.018

Sethi, S. S., Jones, N. S., Fulcher, B. D., Picinali, L., Clink, D. J., Klinck, H., Orme, C. D. L., Wrege, P. H., & Ewers, R. M. (2020). Characterizing soundscapes across diverse ecosystems using a universal acoustic feature set. *Proceedings of the National Academy of Sciences* , *117* (29), 17049–17055. https://doi.org/10.1073/pnas.2004702117

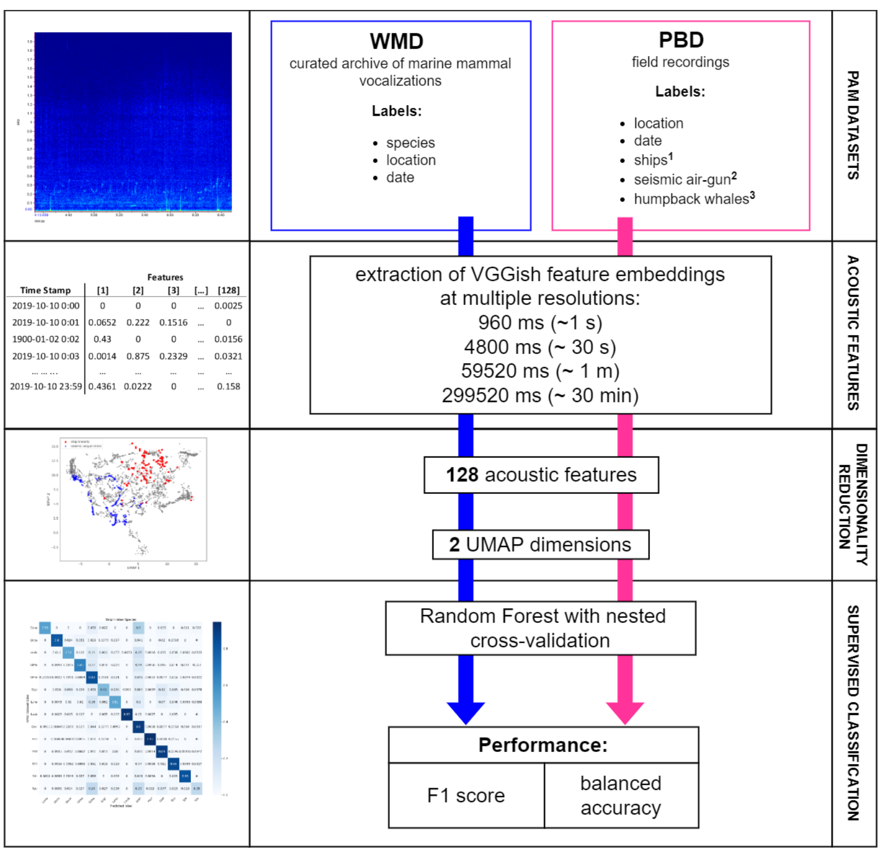
Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *Retrived from:*https://doi.org/10.48550/arXiv.1409.1556

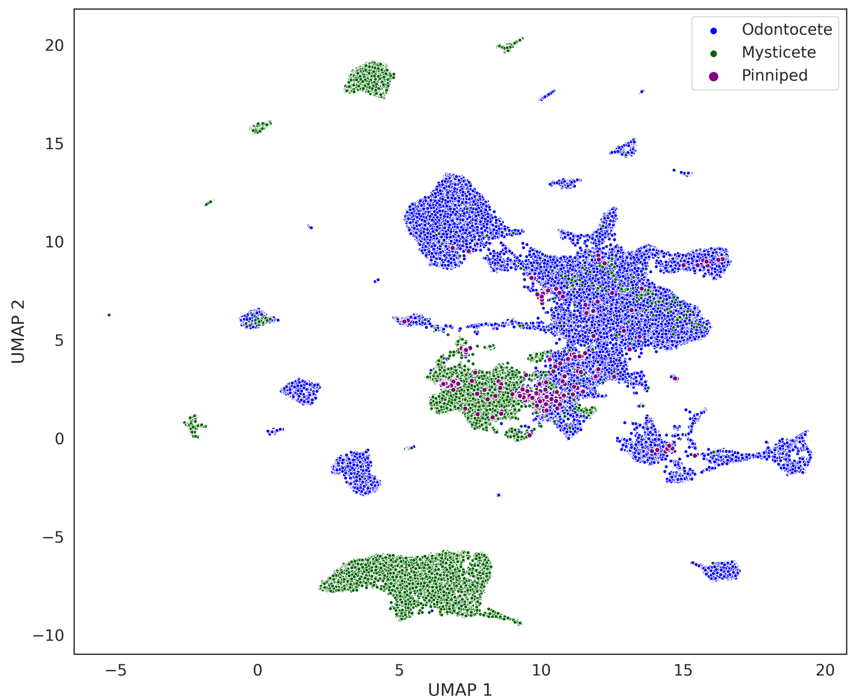
Sully, S., Burkepile, D. E., Donovan, M. K., Hodgson, G., & van Woesik, R. (2019). A global analysis of coral bleaching over the past two decades. *Nature Communications* , *10* (1), 1264. https://doi.org/10.1038/s41467-019-09238-2

Tabak, M. A., Norouzzadeh, M. S., Wolfson, D. W., Sweeney, S. J., Vercauteren, K. C., Snow, N. P., Halseth, J. M., Di Salvo, P. A., Lewis, J. S., White, M. D., Teton, B. (2019). Machine learning to classify animal species in camera trap images: Applications in ecology.*Methods in Ecology and Evolution* , *10* (4), 585–590. https://doi.org/10.1111/2041-210X.13120

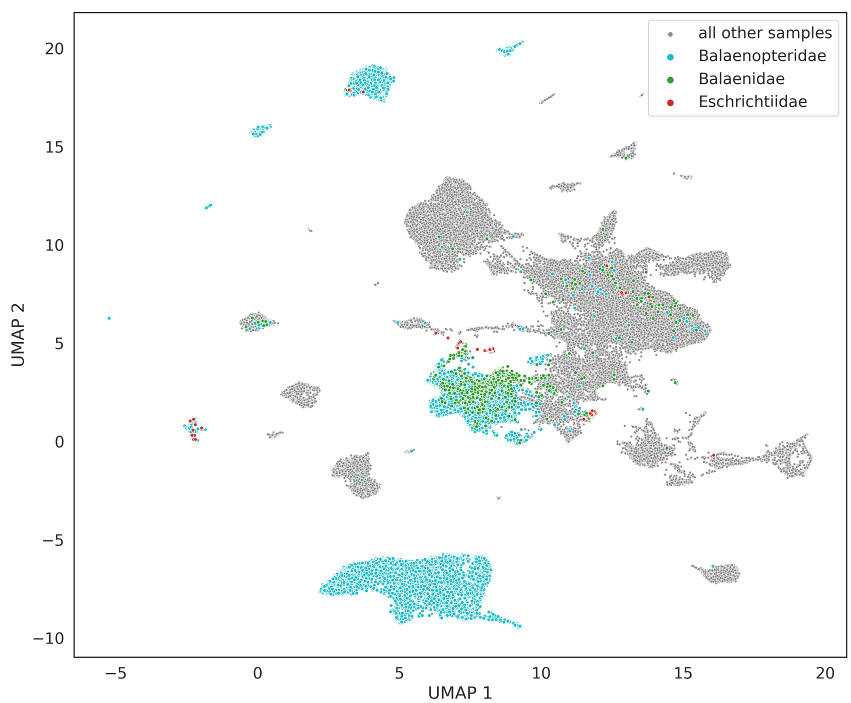
Usman, A. M., Ogundile, O. O., & Versfeld, D. J. J. (2020). Review of automatic detection and classification techniques for cetacean vocalization. *IEEE Access* , *8* , 105181–105206. https://doi.org/10.1109/ACCESS.2020.3000477

Vester, H., Hallerberg, S., Timme, M., & Hammerschmidt, K. (2017). Vocal repertoire of long-finned pilot whales (*Globicephalamelas*) in northern Norway. *The Journal of the Acoustical Society of America* , *141* (6), 4289–4299. https://doi.org/10.1121/1.4983685

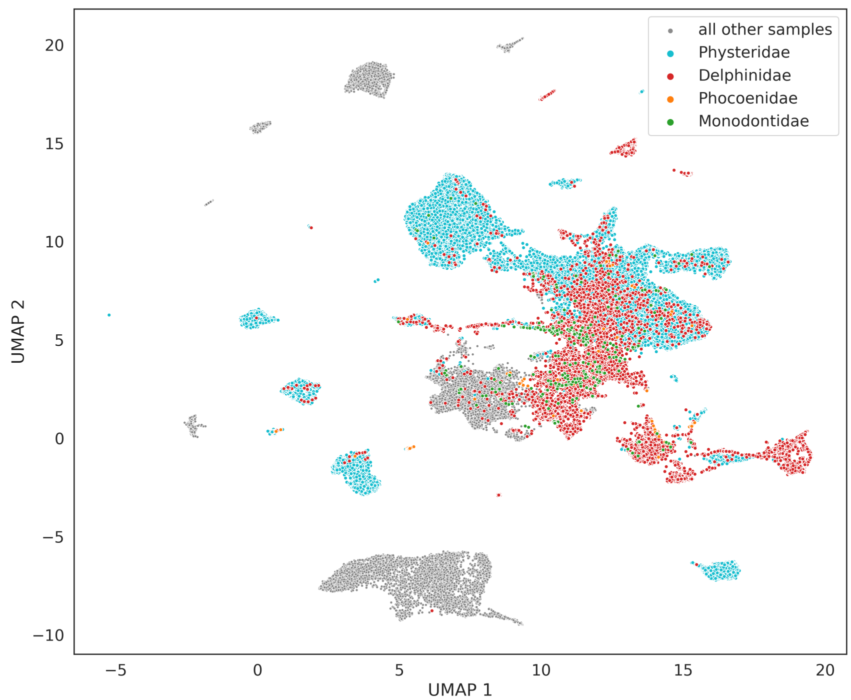


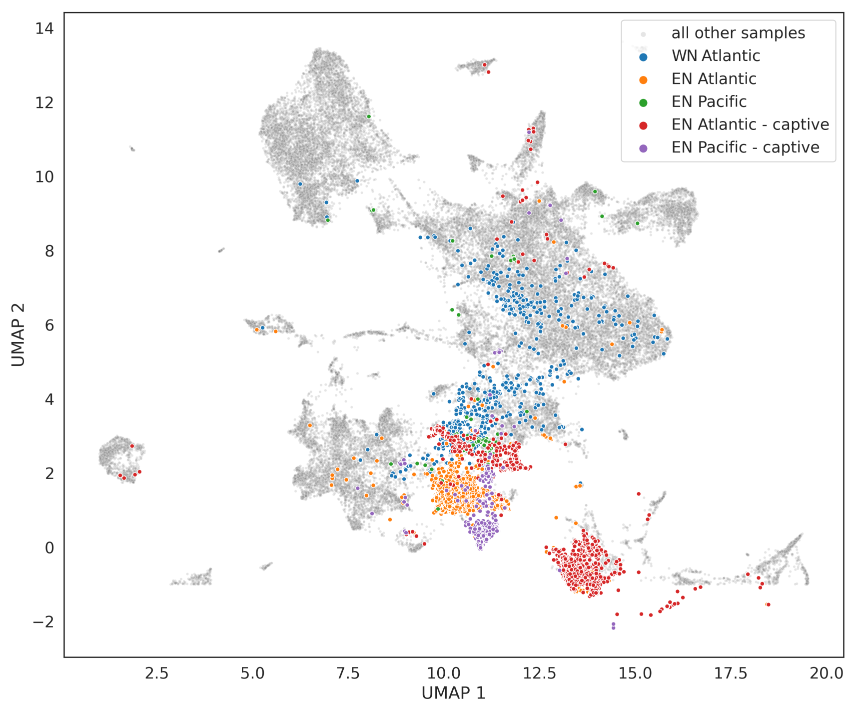


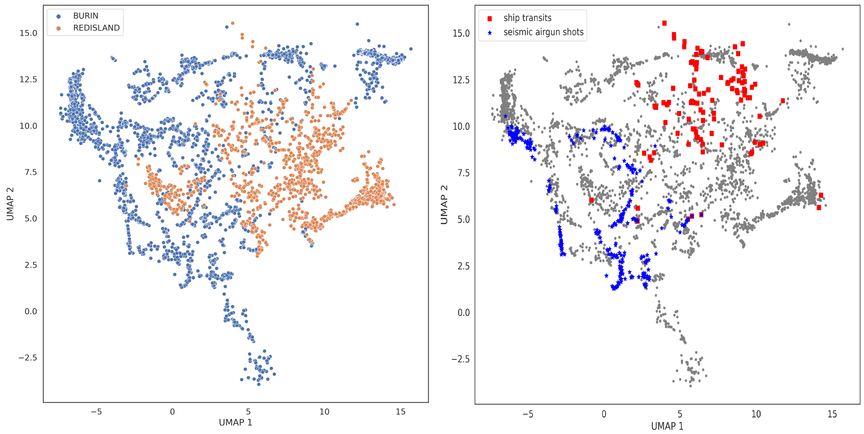
Odontocete
Mysticete
Pinniped
UMAP 2
UMAP 1

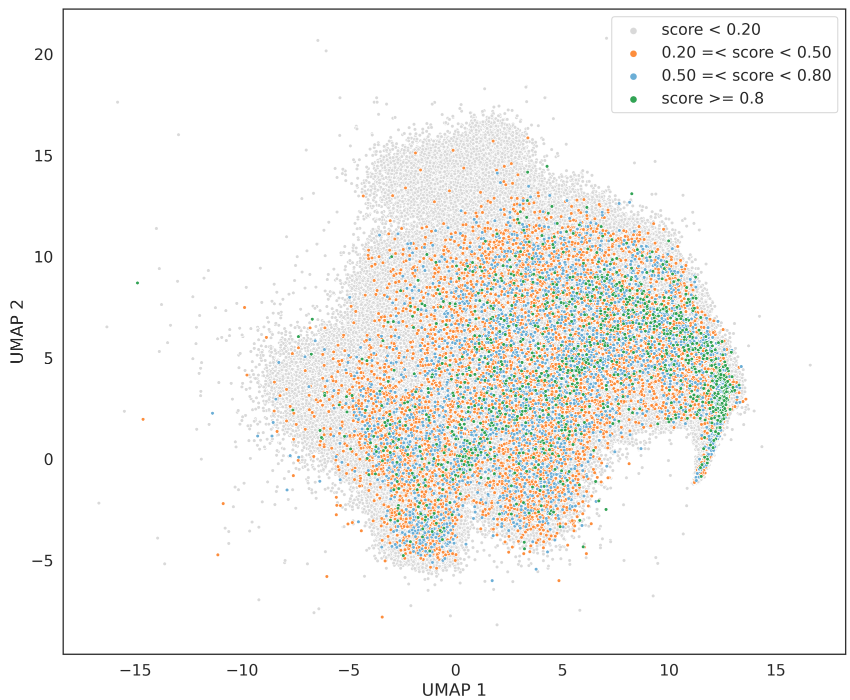


all other samples
Balaenopteridae
Balaenidae
Eschrichtiidae
UMAP 2
UMAP 1


all other samples
Physteridae
Delphinidae
Phocoenidae
Monodontidae
UMAP 1
UMAP 2


all other samples
WN Atlantic
EN Atlantic
EN Pacific
EN Atlantic - captive
EN Pacific - captive
UMAP 1
UMAP 2


BURIN
REDISLAND
UMAP 1
UMAP 2
ship transits
seismic airgun shots
UMAP 1
UMAP 2


score < 0.20
0.20 =< score < 0.50
0.50 =< score < 0.80
score >= 0.8
UMAP 1
UMAP 2