# Classifying Humpback Whale Calls to Song and Non-song Vocalizations using Bag of Words Descriptor on Acoustic Data

Hamed Mohebbi-Kalkhoran*, Chenyang Zhu†, Matthew Schinault‡, Purnima Ratilal§
*Department of Electrical and Computer Engineering*
*Northeastern University*
Boston, MA, USA
{*mohebbikalkhoran.h, †zhu.che, ‡schinault.m}@northeastern.edu, §purnima@ece.neu.edu

***Abstract*—Humpback whale behavior, population distribution and structure can be inferred from long term underwater passive acoustic monitoring of their vocalizations. Here we develop automatic approaches for classifying humpback whale vocalizations into the two categories of song and non-song, employing machine learning techniques. The vocalization behavior of humpback whales was monitored over instantaneous vast areas of the Gulf of Maine using a large aperture coherent hydrophone array system via the passive ocean acoustic waveguide remote sensing technique over multiple diel cycles in Fall 2006. We use wavelet signal denoising and coherent array processing to enhance the signal-to-noise ratio. To build features vector for every time sequence of the beamformed signals, we employ Bag of Words approach to time-frequency features. Finally, we apply Support Vector Machine (SVM), Neural Networks, and Naive Bayes to classify the acoustic data and compare their performances. Best results are obtained using Mel Frequency Cepstrum Coefficient (MFCC) features and SVM which leads to 94% accuracy and 72.73% F1-score for humpback whale song versus non-song vocalization classification, showing effectiveness of the proposed approach for real-time classification at sea.**

***Index Terms*—bioacoustic, ocean acoustic, remote sensing, humpback whale, vocalization, SVM, Neural Networks, Bag of Words, wavelet, song, nonsong, passive acoustic, classification, beamforming, array processing**

## I. Introduction

Marine mammal vocalizations are associated with a variety of purposes such as echolocation, sexual display while mating, singing while migrating to breeding and feeding grounds, communication, as well as contact calls for coordinated movement during group feeding and other activities [1], [2]. Humpback whale vocalizations can be divided into two classes, song [3], [4] and non-song [5], [6]. Song vocalizations are sequences of calls that are structured and organized into repeatable pattern of phrases [3] with short inter-pulse intervals [5]. Humpback whale songs are regarded as breeding displays by males in mating grounds [4], and the whales have been observed to carry their tunes into feeding grounds [7]. The non-song vocalizations that include feeding cries, bow-shaped and downsweep or meow moans are suited for night time communication among humpback individuals and coordination during group feeding activities, have larger and highly variable inter-pulse intervals [1], [5], [8]. In Fig. 1, the logarithm of power spectrogram for sample song and non-song vocalizations are shown. Fig. 1 (a) shows part of a nonsong sequence while Fig. 1 (b) shows song vocalizations.

Here we classify humpback whale vocalizations recorded on a large-aperture densely-populated coherent hydrophone array system containing 160 elements during an experiment in the Gulf of Maine (GOM) in Fall 2006. These vocalizations have been previously analyzed using the Passive Ocean Acoustic Waveguide Remote Sensing (POAWRS) technique and applied to study marine mammals behavior and distributions, and their temporospatial correlation to prey species behavior and distribution in the GOM feeding ground [1], [5], [8], [9]. The large volume of underwater acoustic data recorded were previously manually labeled after semi-automatic processing and audio-visual inspection of tens of thousands of beamformed time-frequency spectrograms, which is a time-consuming process, demanding significant human interaction, and typical in analysis of large acoustic datasets [10].

Machine learning techniques can help to analyze the large amount of acoustic data efficiently and within a significantly reduced time frame. In [11], the performance of Mel Frequency Cepstrum Coefficients (MFCC), the linear prediction coding (LPC) coefficients, and Cepstral coefficients for representing humpback whale vocalizations were explored. And then K-means clustering was used to cluster units and subunits of humpback whale vocalizations into 21 and 18 clusters respectively. In [5], temporal and spatial statistics of humpback whales song and non-song calls in the Gulf of Maine were investigated. Subclassification of humpback whale downsweep moan calls into 13 sub-groups were accomplished using K-means clustering after pitch-tracking and time-frequency feature extraction. In [12], temporal-spatial and time-frequency characteristics of fin whale vocalizations in the Norwegian Sea were studied and found to comprise of 5 distinct call types. Several classifiers including SVM, Decision Tree, Logistic Regression, and CNN were later developed and tested in [13] to automatically extract the five fin whale call types. In [14], blue whale calls were classified using neural network with features derived from short-time Fourier and wavelet transforms. In [15], an automatic detection and classification of baleen

978-1-7281-4550-1/19/$31.00 ©2019 IEEE
DOI 10.1109/ICMLA.2019.00150

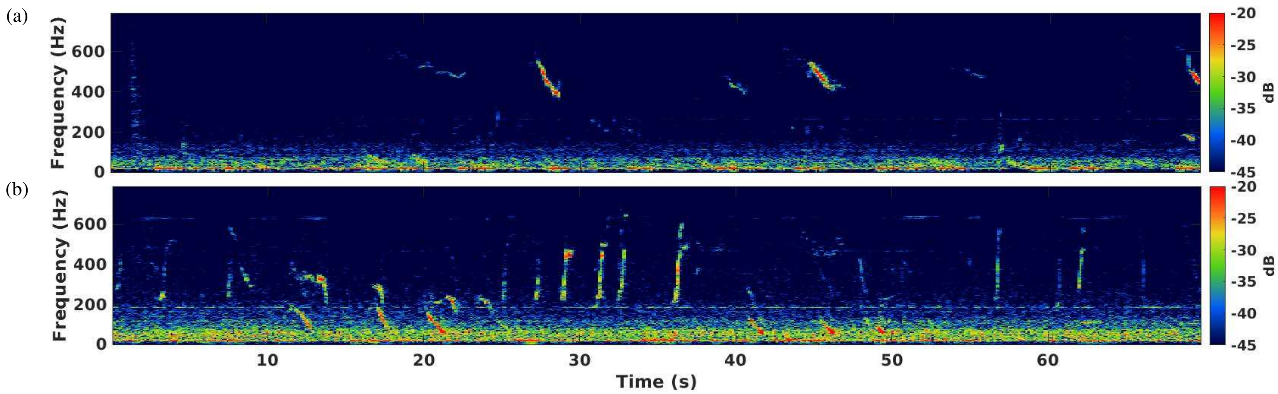


Fig. 1. Log-transformed normalized power spectrogram for (a) nonsong and (b) song calls.

whale calls system was developed using pitch tracking and quadratic discriminant function analysis. Echolocation clicks of odontocetes were classified by exploiting cepstral features and Gaussian mixture models in [16], while in [17], whale call classification was done using Convolutional Neural Networks and transfer learning on time-frequency features. Fish sounds were classified using random forest and SVM in [18].

In this study, we develop automatic methods for classifying humpback whale vocalizations into two classes, song and non-song. We use signals of 70 seconds duration as the input to classifiers. This duration corresponds to hydrophone array signal recording time frame in each file, determined from past experiments to be a suitable observation time frame for distinguishing acoustic signals from a large variety of underwater sound sources, as well as for distinguishing humpback whale song from non-song calls [5]. To summarize the statistics of features in one time period, we employ Bag of Words (BoW) method which has already been used in different applications such as speech analysis [19] and video classification [20]. Finally, we apply Support Vector Machine (SVM), Neural Networks, and Naive Bayes to classify the acoustic data and compare their performances.

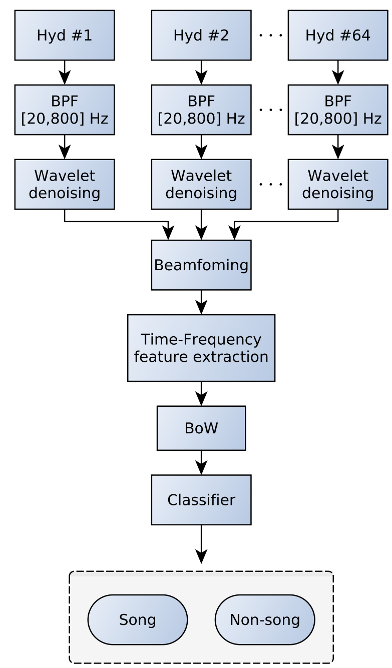


Fig. 2. General workflow of proposed approach.

## II. Materials and Methods

In this section, we describe the dataset and approaches we use for the task of classifying humpback whale song and non-song vocalizations, illustrated in Fig. 2. We explain each part of this workflow in the following subsections.

### A. Dataset

The Gulf of Maine 2006 Experiment dataset [1], [8] was acquired in Fall 2006 in this important North Atlantic marine mammal feeding ground containing large populations of spawning fish, the Atlantic herring [21], [22], [23]. Acoustic recordings of whale vocalizations were acquired using a large-aperture densely-populated coherent hydrophone array with 160 elements [24], [25] towed by a research vessel along designated tracks in Franklin Basin, north of Georges Bank. The acoustic data is sampled at 8000 Hz per element. The mid-frequency (MF) sub-aperture, consisting of 64 equally spaced hydrophones with inter-element spacing of 0.75 m was used to analyze the humpback whale calls. The water depth ranged from 180 m to 250 m at the array locations. The array tow depth was roughly 105 m and tow speed was roughly 2 m/s.

Previous analysis in [1] and [5] provided the labeled set of humpback whale vocalization signals for this experiment. There the acoustic pressure-time series measured by sensors across the coherent hydrophone array were converted to two-dimensional beam-time series by beamforming. A total of 64 beams were formed spanning 360 degree horizontal azimuth about the receiver array for data from the MF sub-aperture. Each beam-time series was converted to a beamformed spectrogram by short-time Fourier transform (sampling frequency 8000 Hz, frame 2048 samples, overlap 3/4, Hann window).

Significant sounds present in the beamformed spectrograms were automatically detected by first applying a pixel intensity threshold detector [26] followed by pixel clustering, and verified by visual inspection. Beamformed spectrogram pixels with local intensity values that stood 5.6 dB above the background were grouped using a clustering algorithm according to a nearest-neighbor criteria that determines if the pixels can be grouped into one or more significant sound signals. Humpback whale vocalization signal detections were verified and labeled as song or non-song manually by visual inspection and listening to the sounds [1], [5], [8].

Here we build the samples for song versus non-song classification task using the following procedure. Each 70 seconds duration beamformed pressure-time series signal in bearings containing humpback whale vocalizations are first gathered [5]. Samples with song labels are those that contain some song vocalizations within the 70 seconds duration beamformed pressure-time series, while non-song samples are those which do not contain any song vocalization in that duration. A total of 1193 recorded acoustic data files each containing 70 s duration of coherent hydrophone array observation are analyzed here. These files all contain humpback whale vocalizations, some of which originate from multiple and distinct bearing directions. Subsequently, after beamforming we obtain 1788 labeled samples, of which 199 are labeled as song and 1589 are labeled as non-song. This subset of data corresponds to roughly 24 hours of coherent hydrophone array observation.

### B. Preprocessing

Underwater acoustic data collected by hydrophones contain different sources of sound that include biological, geophysical and of man-made origins [27]. The ambient noise which is always present in the background in seas and oceans is often due to breaking waves, turbulence and rain [27]. Other sources of ocean sound include nearby and distant ships, offshore piling and geophysical prospecting activities, fish grunts, marine mammal vocalizations, and natural geological sounds, such as seafloor earthquakes and volcanic activities . Underwater sound signals undergo both spreading and absorption losses that are dependent on signal frequency. As a result, the recorded background sound is also frequency-dependent. In order to mitigate these effects and improve the classification performance, signal preprocessing is necessary. We first apply a bandpass filter with passband frequency range between 20 Hz and 800 Hz. Then to decrease the size of data, we down-sample the signal 4 times, so the new sampling frequency is 2000 Hz. In order to reduce the background noise, symlet4 wavelet denoising with soft thresholding for every signal channel (hydrophone) is exploited [28].

We employ delay and sum array beamforming to amplify signals in relative bearing directions containing humpback whale vocalizations while simultaneously suppressing signals from other directions. The Signal-to-Noise Ratio (SNR) gain from beamforming when using the $n$ hydrophone array compared to using only one hydrophone is $10\log_{10} n = 18$ dB, where $n = 64$ [1]. So using a large densely-populated array of hydrophones enables us to detect whale vocalizations from greater distances. The relative bearing (horizontal azimuth of source, which here is humpback whale, with respect to the hydrophone array) was previously determined by high-resolution beamforming and included in the GOM dataset. The time delay of signal arrival between two successive hydrophones is computed using the following equation:

$$t = \frac{c}{d} \sin \theta \tag{1}$$

where $c = 1500$ m/s is the speed of sound propagation in the water, $d$ is the distance between two successive hydrophones in meters, and $\theta$ is the relative bearing. After preprocessing, the spectrogram still contains noise elements. To further enhance the signal, we apply a two dimensional Gaussian filter on the two-dimensional spectrogram. The visual spectrogram results from different stages of preprocessing are illustrated in Fig. 3.

### C. Feature extraction

For the humpback whale song vs non-song vocalization classification task, since both temporal and spectral components are important, we use features that have information about both time and frequency. We exploit two types of features, power spectrogram and MFCC. For spectrogram, a hanning window of size 512 samples with 50 percent overlap is used.

In order to construct a fixed-length feature vector for every 70 s duration beamformed pressure-time series sample, we use different statistical measures. Approaches using simple statistics such as minimum, maximum and and average value of time-frequency features have been used previously in video classification [29] as well as in acoustic data classification [5], [12]. Here we exploit Bag of Words (BoW) method to build a fixed-length feature vector representation of an input time series with variable length according to the following procedure. First extract feature vector for every time-step (using MFCC or power spectrogram or other methods). We denote the $t^{th}$ time step feature vector of the $i^{th}$ sample by $f_t^{(i)}$. For the $i^{th}$ sample $S^{(i)}$, with total number of $T$ time steps we have the following:

$$S^{(i)} = [f_1^{(i)}, f_2^{(i)}, ..., f_t^{(i)}, ..., f_T^{(i)}] \tag{2}$$

Let the total number of samples in the dataset be $N$. Then the whole dataset denoted by $D$, can be represented as follows:

$$D = [S^{(1)}, S^{(2)}, ..., S^{(i)}, ..., S^{(N)}] \tag{3}$$

After constructing $D$, we perform a clustering method with total number of clusters equal to $K$ on the dataset $D$. At the end, we build the feature vector of dimension $K$ for every sample, $S^{(i)}$. We denote this feature vector by $F^{(i)}$, and the $j^{th}$ element of this feature vector by $h_j^{(i)}$. So we have:

$$F^{(i)} = [h_1^{(i)}, h_2^{(i)}, ..., h_j^{(i)}, ..., h_K^{(i)}] \tag{4}$$

Note that in contrast to $f_t^{(i)}$ which is a vector of any dimension (based on the feature extraction method, such as

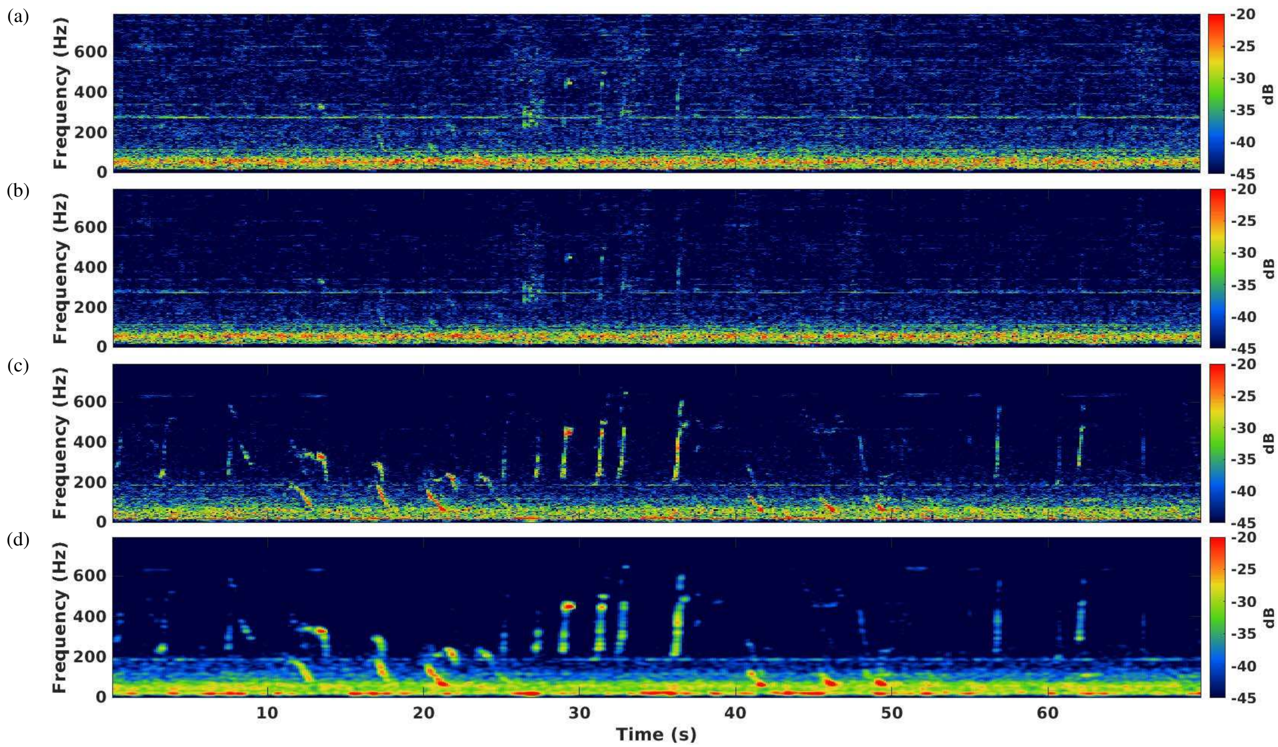


Fig. 3. Effect of different preprocessing steps on log of power spectrogram. (a) is for a sample hydrophone, (b) is for the same hydrophone after wavelet denoising, (c) is after beamforming using 64 hydrophones, and (d) is after smoothing using 2 dimensional Gaussian filter.

MFCC or spectrogram), $h_j^{(i)}$ is a scalar that signifies the number of $f_t^{(i)}$s belonging to the $j^{th}$ cluster

At the end, feature vectors are normalized by dividing each element of $F^{(i)}$ by the total number of time steps, $T$, in that sequence.

### D. Classification

With the labeled samples, we employ supervised learning approaches. We examined several classifiers including Support Vector Machine (SVM), Neural Networks (NN), and Gaussian Naive Bayes (GNB). SVM is based on finding maximum margin, uses a convex cost function, and it always reaches the global minimum of the cost function [30]. Multi-layer perceptron NN can model non-linear classification. Compared to SVM, NN parameters optimization is not convex and the solution is not guaranteed to obtain the global minimum of the cost function. Learning the parameters is done by back-propagation method. Although it is non-convex, NN can model highly non-linear and complex structures. Using effective learning algorithms for NNs and also large volume datasets, NNs have been used widely in different machine learning applications [31]. Naive Bayes (NB) classifier is based on probabilistic model and Bayes' theorem assuming that features are independent. In Gaussian Naive Bayes (GNB) classifier, it is assumed that each class follows a Gaussian distribution [32].

## III. Experiments

In this section, we explain the parameters and setup for our approach and evaluate the performance of the methods we considered for humpback whale song versus nonsong vocalization classification.

### A. Experimental Setup

We divide the dataset into training and test data, where 200 samples were selected randomly for test data, and parameters are chosen using 5-fold cross validation on training data. For wavelet denoising, the value of threshold is selected to be 0.1. For K-means clustering in BoW method, the number of clusters, $K$, is set to 50. For MFCC, we tried different values for number of Discrete Cosine Transform (DCT) coefficients, and the performance was relatively robust for different values of this parameter, so to obtain the best result we set this parameter to 13. In classification, for C-SVM, we used third degree polynomial kernel, and C=10. For NN, 2 hidden layer with 10 neurons in each layer, Rectified Linear Unit (ReLU) activation function, and weight decay of 0.001 were used. Adam optimizer [33] with learning rate of 0.01, and batch size of 200 was exploited to train the neural network.

### B. Classification results

In this section, we investigate and compare the humpback whale song versus nonsong vocalization classification performance. To evaluate the results, we calculate accuracy and

TABLE I
RESULT FOR DIFFERENT FEATURES AND CLASSIFIERS

| Classifier | PSD | MFCC | Acc(%) | AUC(%) | F1(%) |
|---|---|---|---|---|---|
| **SVM** | ✓ | × | 92.00 | 80.75 | 61.90 |
| | × | ✓ | **94.00** | 90.39 | **72.73** |
| | ✓ | ✓ | 93.00 | 85.13 | 69.57 |
| **GNB** | ✓ | × | 68.00 | 87.24 | 39.62 |
| | × | ✓ | 68.50 | 81.39 | 37.62 |
| | ✓ | ✓ | 67.50 | 80.78 | 38.10 |
| **NN** | ✓ | × | 91.00 | 91.07 | 55.00 |
| | × | ✓ | 92.50 | **94.27** | 70.59 |
| | ✓ | ✓ | **94.00** | 92.14 | 70.00 |

receiver operating characteristic (ROC) curve, and area under ROC curve (AUC) as the measures. Since our dataset is imbalanced, we also report the F1-score which provides insights on associated errors. Besides applying the classifiers on MFCC and Power Spectrogram Density (PSD) separately, we also investigate to see if there are improvements in performance by combining MFCC and PSD. We concatenated PSD and MFCC feature vectors to build the combined feature vector (concatenating the output of BoW for each feature). Table I shows the results when using PSD, MFCC, and both together as the features.

Fig. 4 shows ROC curves for different classifiers and features. As can be noted from both the numerical results and ROC curves, using MFCC as the feature usually leads to superior performance over PSD. Also in terms of AUC, NN performs better compared to SVM and Naive Bayes. Because the dataset is imbalanced, F1 score is a better measurement to evaluate the performance in this case. When considering F1-score, the best results are achieved when using MFCC features and SVM classifier. The Combination of PSD and MFCC did not show significant improvement.

## IV. CONCLUSION

In this study, we presented several machine learning approaches to address the problem of humpback whale vocalization classification into the two classes of song and non-song. We examined different preprocesing and classifiers to improve the results. Our analysis demonstrates the potential of machine learning approaches for real-time classification in field experiments applied to bio-acoustics and marine mammal vocalization analysis. Future work will involve investigating other preprocessing techniques to enhance the quality of signals and also applying the classification approaches developed and trained here to other coherent hydrophone array datasets from various undersea regions of the world to ascertain generalization of the proposed methods.

## V. ACKNOWLEDGEMENTS

This research was supported by the US National Science Foundation and the US Office of Naval Research (Ocean Acoustics Program).

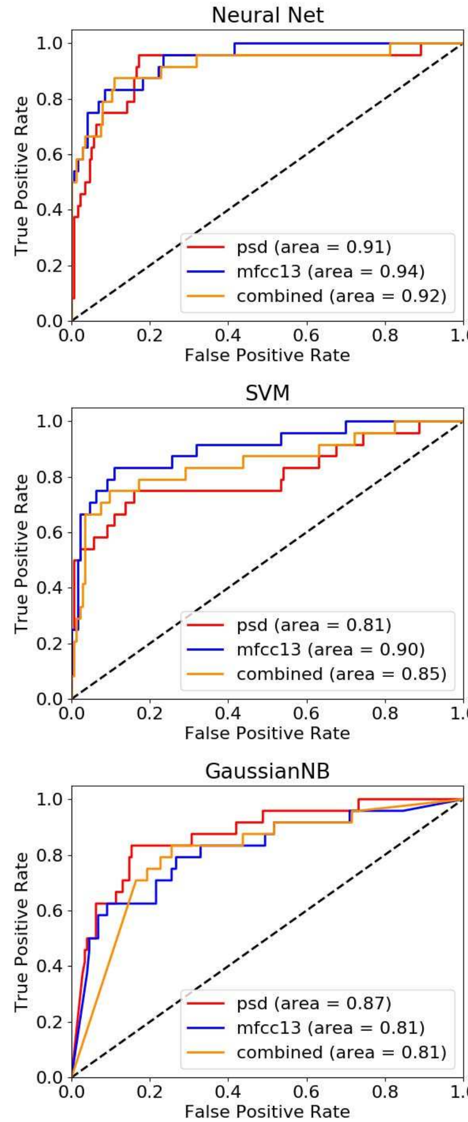


Fig. 4. ROC curves for different classifiers using MFCC and PSD features.

## REFERENCES


[1] D. Wang, H. Garcia, W. Huang, D. D. Tran, A. D. Jain, D. H. Yi, Z. Gong, J. M. Jech, O. R. Godø, N. C. Makris *et al.*, "Vast assembly of vocal marine mammals from diverse species on fish spawning ground," *Nature*, vol. 531, no. 7594, p. 366, 2016.

[2] D. Cato, R. McCauley, T. Rogers, and M. Noad, "Passive acoustics for monitoring marine animals-progress and challenges," in *Proceedings of ACOUSTICS*, vol. 2006, 2006, pp. 453–460.

[3] R. S. Payne and S. McVay, "Songs of humpback whales," *Science*, vol. 173, no. 3997, pp. 585–597, 1971.

[4] M. J. Noad, D. H. Cato, M. Bryden, M.-N. Jenner, and K. C. S. Jenner, "Cultural revolution in whale songs," *Nature*, vol. 408, no. 6812, p. 537, 2000.

[5] W. Huang, D. Wang, and P. Ratilal, "Diel and spatial dependence of humpback song and non-song vocalizations in fish spawning ground," *Remote Sensing*, vol. 8, no. 9, p. 712, 2016.

[6] R. A. Dunlop, D. H. Cato, and M. J. Noad, "Non-song acoustic communication in migrating humpback whales (megaptera novaeangliae)," *Marine Mammal Science*, vol. 24, no. 3, pp. 613–629, 2008.

[7] D. K. Mattila, L. N. Guinee, and C. A. Mayo, “Humpback whale songs on a north atlantic feeding ground,” *Journal of Mammalogy*, vol. 68, no. 4, pp. 880–883, 1987.

[8] Z. Gong, A. D. Jain, D. Tran, D. H. Yi, F. Wu, A. Zorn, P. Ratilal, and N. C. Makris, “Ecosystem scale acoustic sensing reveals humpback whale behavior synchronous with herring spawning processes and re-evaluation finds no effect of sonar on humpback song occurrence in the gulf of maine in fall 2006,” *PloS one*, vol. 9, no. 10, p. e104733, 2014.

[9] C. Zhu, H. Garcia, A. Kaplan, M. Schinault, N. Handegard, O. Godø, W. Huang, and P. Ratilal, “Detection, localization and classification of multiple mechanized ocean vessels over continental-shelf scale regions with passive ocean acoustic waveguide remote sensing,” *Remote Sensing*, vol. 10, no. 11, p. 1699, 2018.

[10] M. J. Bianco, P. Gerstoft, J. Traer, E. Ozanich, M. A. Roch, S. Gannot, C.-A. Deledalle, and W. Li, “Machine learning in acoustics: a review,” *arXiv preprint arXiv:1905.04418*, 2019.

[11] F. Pace, F. Benard, H. Glotin, O. Adam, and P. White, “Subunit definition and analysis for humpback whale call classification,” *Applied Acoustics*, vol. 71, no. 11, pp. 1107–1112, 2010.

[12] H. A. Garcia, C. Zhu, M. E. Schinault, A. I. Kaplan, N. O. Handegard, O. R. Godø, H. Ahonen, N. C. Makris, D. Wang, W. Huang *et al.*, “Temporal–spatial, spectral, and source level distributions of fin whale vocalizations in the norwegian sea observed with a coherent hydrophone array,” *ICES Journal of Marine Science*, vol. 76, no. 1, pp. 268–283, 2018.

[13] H. A. Garcia, S. Penna, J. Topple, and P. R. Makris, “Comparing neural networks with conventional classifiers for fin whale vocalizations in beamformed spectrograms of coherent hydrophone array,” *The Journal of the Acoustical Society of America*, vol. 144, no. 3, pp. 1911–1911, 2018.

[14] M. Bahoura and Y. Simard, “Blue whale calls classification using short-time fourier and wavelet packet transforms and artificial neural network,” *Digital Signal Processing*, vol. 20, no. 4, pp. 1256–1263, 2010.

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

[16] M. A. Roch, H. Klinck, S. Baumann-Pickering, D. K. Mellinger, S. Qui, M. S. Soldevilla, and J. A. Hildebrand, “Classification of echolocation clicks from odontocetes in the southern california bight,” *The Journal of the Acoustical Society of America*, vol. 129, no. 1, pp. 467–475, 2011.

[17] L. Zhang, D. Wang, C. Bao, Y. Wang, and K. Xu, “Large-scale whale-call classification by transfer learning on multi-scale waveforms and time-frequency features,” *Applied Sciences*, vol. 9, no. 5, p. 1020, 2019.

[18] M. Malfante, J. I. Mars, M. Dalla Mura, and C. Gervaise, “Automatic fish sounds classification,” *The Journal of the Acoustical Society of America*, vol. 143, no. 5, pp. 2834–2846, 2018.

[19] M. Schmitt, F. Ringeval, and B. W. Schuller, “At the border of acoustics and linguistics: Bag-of-audio-words for the recognition of emotions in speech.” in *Interspeech*, 2016, pp. 495–499.

[20] L. Ballan, M. Bertini, A. Del Bimbo, and G. Serra, “Video event classification using bag of words and string kernels,” in *International Conference on Image Analysis and Processing*. Springer, 2009, pp. 170–178.

[21] N. C. Makris, P. Ratilal, S. Jagannathan, Z. Gong, M. Andrews, I. Bertsatos, O. R. Godø, R. W. Nero, and J. M. Jech, “Critical population density triggers rapid formation of vast oceanic fish shoals,” *Science*, vol. 323, no. 5922, pp. 1734–1737, 2009.

[22] Z. Gong, M. Andrews, S. Jagannathan, R. Patel, J. M. Jech, N. C. Makris, and P. Ratilal, “Low-frequency target strength and abundance of shoaling atlantic herring (clupea harengus) in the gulf of maine during the ocean acoustic waveguide remote sensing 2006 experiment,” *The Journal of the Acoustical Society of America*, vol. 127, no. 1, pp. 104–123, 2010.

[23] W. Overholtz, J. Jech, W. Michaels, L. Jacobson, and P. Sullivan, “Empirical comparisons of survey designs in acoustic surveys of gulf of maine-georges bank atlantic herring,” *J. Northw. Atl. Fish. Sci*, vol. 36, pp. 127–144, 2006.

[24] K. Becker and J. Preston, “The onr five octave research array (fora) at penn state,” in *Oceans 2003. Celebrating the Past... Teaming Toward the Future (IEEE Cat. No. 03CH37492)*, vol. 5. IEEE, 2003, pp. 2607–2610.

[25] D. Wang and P. Ratilal, “Angular resolution enhancement provided by nonuniformly-spaced linear hydrophone arrays in ocean acoustic waveguide remote sensing,” *Remote Sensing*, vol. 9, no. 10, p. 1036, 2017.

[26] M. I. Sezan, “A peak detection algorithm and its application to histogram-based image data reduction,” *Computer vision, graphics, and image processing*, vol. 49, no. 1, pp. 36–51, 1990.

[27] G. M. Wenz, “Acoustic ambient noise in the ocean: Spectra and sources,” *The Journal of the Acoustical Society of America*, vol. 34, no. 12, pp. 1936–1956, 1962.

[28] M. S. Chavan, N. Mastorakis, M. N. Chavan, M. Gaikwad *et al.*, “Implementation of symlet wavelets to removal of gaussian additive noise from speech signal,” in *Proceedings of Recent Researches in Communications, Automation, Signal Processing, Nanotechnology, Astronomy and Nuclear Physics: 10th WSEAS International Conference on Electronics, Hardware, Wireless and Optical Communications (EHAC11), Cambridge*, 2011, p. 37.

[29] H. M. Kalkhoran and E. Fatemizadeh, “Pain level estimation in video sequences of face using incorporation of statistical features of frames,” in *2015 9th Iranian Conference on Machine Vision and Image Processing (MVIP)*. IEEE, 2015, pp. 172–175.

[30] K. P. Murphy, *Machine learning: a probabilistic perspective*. MIT press, 2012.

[31] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *nature*, vol. 521, no. 7553, p. 436, 2015.

[32] H. Zhang, “The optimality of naive bayes,” *AA*, vol. 1, no. 2, p. 3, 2004.

[33] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *arXiv preprint arXiv:1412.6980*, 2014.