

COMPARISON OF TWO METHODS FOR DETECTION OF NORTH ATLANTIC RIGHT WHALE UPCALLS

Mahdi Esfahanian, Hanqi Zhuang, Nurgun Erdol, and Edmund Gerstein

Dept. of Computer and Electrical Eng. and Computer Science, Florida Atlantic University, Boca Raton, FL
Emails: {mesfahan, zhuang, erdol, egerste1}@fau.edu

ABSTRACT

In this paper, a study is carried out for detecting North Atlantic Right Whale upcalls with measurements from passive acoustic monitoring devices. Preprocessed spectrograms of upcalls are subjected to two different tasks, one of which is based on extraction of time-frequency features from upcall contours, and the other that employs a Local Binary Pattern operator to extract salient texture features of the upcalls. Then several classifiers are used to evaluate the effectiveness of both the contour-based and texture-based features for upcall detection. Detection results reveal that popular classifiers such as Linear Discriminant Analysis, Support Vector Machine, and TreeBagger can achieve high detection rates. Furthermore, using LBP features for call detection shows improved accuracy of about 3% to 4% over time-frequency features when an identical classifier is used.

Index Terms— North Atlantic Right Whale, Local Binary Patterns, Spectral Denoising, Upcall Detection.

1. INTRODUCTION

North Atlantic Right Whale (NARW) is one of the critically endangered whales as the decline in its population is not compensated with its low birth rate [1]. It is recorded that some 300-500 individuals remain off the east coast of North America [2]. Therefore, it is important to be able to detect the presence of such animals in high-risk areas in order that mitigation measures protecting them from such damages as collision with ships may be activated. Passive acoustic methods have been shown to be the most effective mechanisms for determining whale presence in critical habitats [4]. Upcalls are narrow-band frequency-modulated chirps in the 50-250Hz frequency band produced by NARW for long-range communication [3].

The detection of NARW upcalls has attracted researchers in the field of bioacoustics since these species are highly endangered and automatic detection systems have to be developed in order to find right whale calls amidst other marine mammal vocalizations. Mellinger [5] compared the performance of spectrogram correlation and neural network methods. The former uses an optimization program to find the synthetic kernel that best correlates to a sample space of 20 right whale upcalls. The latter method trains weights of a NN via backpropagation on 9/10 of the test dataset. The neural network performed better,

achieving an error rate of less than 6%. Munger *et.al* [6] also used spectrogram cross-correlation with a synthetic kernel [7] for automatically detecting right whales using the software program *Ishmael* [8]. Despite the high number of false detections and missed individual calls, Munger's spectrogram cross-correlation helped a human analyst identify segments of data that contained right whale calls with high probability. Gillespie [9] constructed a two-stage detector where the vocalization outlines are extracted from a smoothed spectrogram using an edge detection method. In the second stage, parameters measured from time-frequency contours are fed into a classifier to determine the sounds associated with right whales. The problem was also addressed by Urazghildiiev *et.al* [10] who used a generalized likelihood ratio test (GLRT) detector of polynomial-phase signals with unknown amplitude and polynomial coefficients observed in the presence of locally stationary Gaussian noise. The closed form representation for a minimal sufficient statistic was derived and a realizable detection scheme was developed. The performance was shown to be superior to other detection techniques. Urazghildiiev and Clark [11] designed an automatic detector for a passive acoustic NARW monitoring system that determines the time of signals' occurrence but a human operator makes the final decision after spectrogram inspection of the marked areas.

In this paper, we propose two new texture-based techniques for NARW upcall detection. The first method performs elaborate pre-processing in order to isolate a spectrogram contour associated with an upcall and drives time-frequency parameters from the contour to use feature vectors for classification. The second approach applies the Local Binary Pattern (LBP) operator on a region of interest in the spectrogram to capture important texture features. Finally, both types of features are fed to classifiers and detection results are evaluated.

2. PROPOSED FEATURE EXTRACTION ALGORITHMS

Both procedures for NARW upcall detection proposed in this section consist of several steps to extract either contour features or texture features for the purpose of upcall detection.

2.1. Contour-based approach

Spectrograms were obtained of NARW upcalls recorded in two-second clips at a sampling rate of 2000 Hz. The data were segmented into 80% of overlapping frames of 128 ms duration (frequency resolution of 7.8 Hz) and Hanning windowed with no zero padding. In the first step, the spectrogram is normalized to give:

$$S_N(t, f_i) = \frac{S(t, f_i) - \mu_i}{\sigma_i} \quad \text{for } i = 1, \dots, N \quad (1)$$

where $S_N(t, f)$ and $S(t, f)$ represent the normalized and original spectrograms, respectively. Also μ_i and σ_i are the mean and standard deviation calculated for each frequency band f_i , respectively. Such normalization emphasizes long-lasting narrowband noises made by ships, wind, and electrical machineries enhancing the short-duration sounds such as upcalls. To reduce background noise and avoid extreme values, the spectrogram is equalized by hard-limiting the spectral upper and lower bounds of amplitudes as:

$$S_H(t, f) = \max \left[S_{\text{floor}}, \min \left(S_{\text{ceiling}}, S_N(t, f) \right) \right] - S_{\text{floor}} \quad (2)$$

where S_{floor} and S_{ceiling} are the new lower and upper bounds of the normalized spectrogram. In another word, the equalization algorithm renders the spectrogram values in the range between S_{floor} and S_{ceiling} .

After spectrogram normalization and equalization, important regions within the spectrogram have to be analyzed in order to find areas associated with NARW upcalls. Equalized spectrogram is subsequently converted to a binary image, a process aimed at finding continuous segments. Considering the fact that some upcalls might be very faint inside the background noise in the spectrogram, it is worthwhile to point out that a well-chosen low threshold has to be set in such a way that no target objects are missed during the binarization process. The binary image obtained from the equalized spectrogram of Fig. 1 is given in top image of Fig. 2.

The binary image shows many spurious contours remaining as a result of the tradeoff of choosing a low threshold. Ranging from tiny to large, these contours are irrelevant clutter and must be separated from the object of interest corresponding to the upcall contour. Toward that end, an 8-connected neighborhood and Moore-Neighbor tracing technique using Jacob's stopping criteria are applied on the spectrogram to locate individual continuous objects and trace their exterior boundaries as shown in the lower image of Fig. 2. Then a set of properties such as perimeter (pixels), area (pixels), height (Hz) as a measure of frequency range, width (sec) measures the time duration, is extracted from each object in the image for further processing.

The parameters are used to make an initial decision to discard an object or keep it for further analysis. The thresholds for each object are chosen to minimize the number of missed objects associated with an upcall. If no objects are detected, the spectrogram contains no upcall regions and is labeled as non upcall. If there is one object detected, it is considered as a potential upcall and passed to the second detection phase as depicted in Fig. 3.

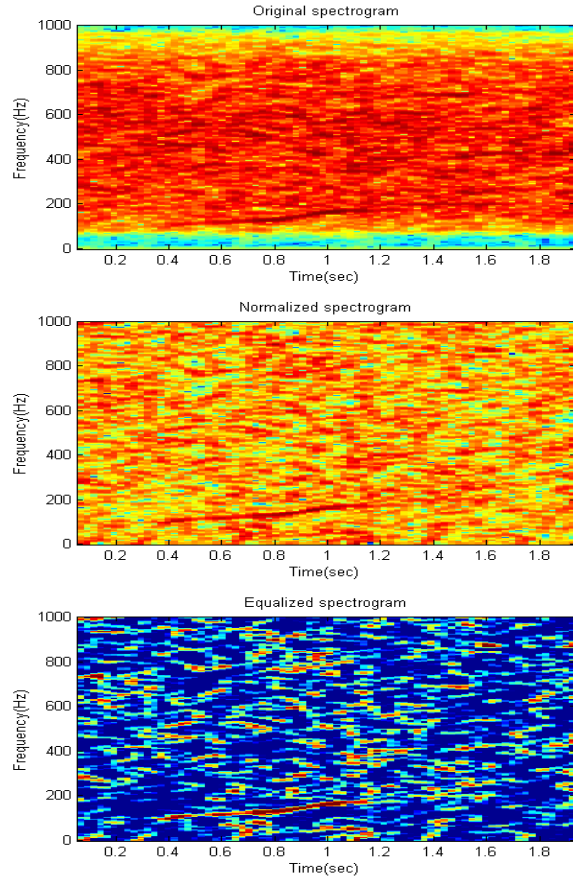


Fig. 1. Original, normalized and equalized spectrograms

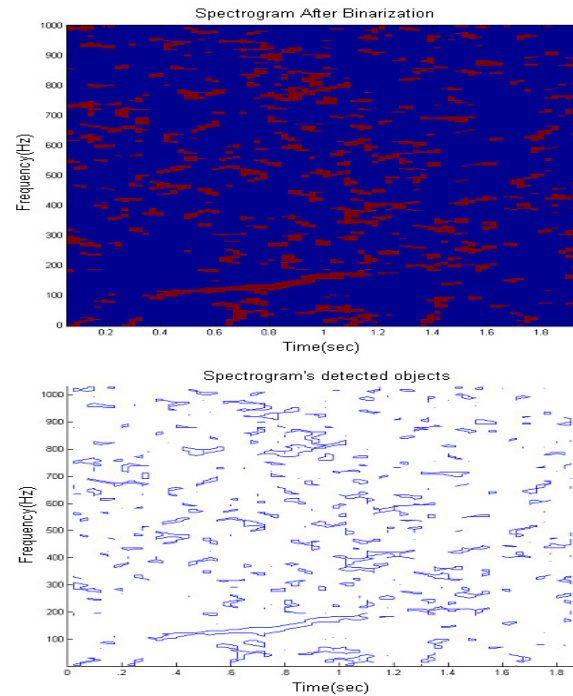


Fig. 2. Spectrogram after binarization and object detection

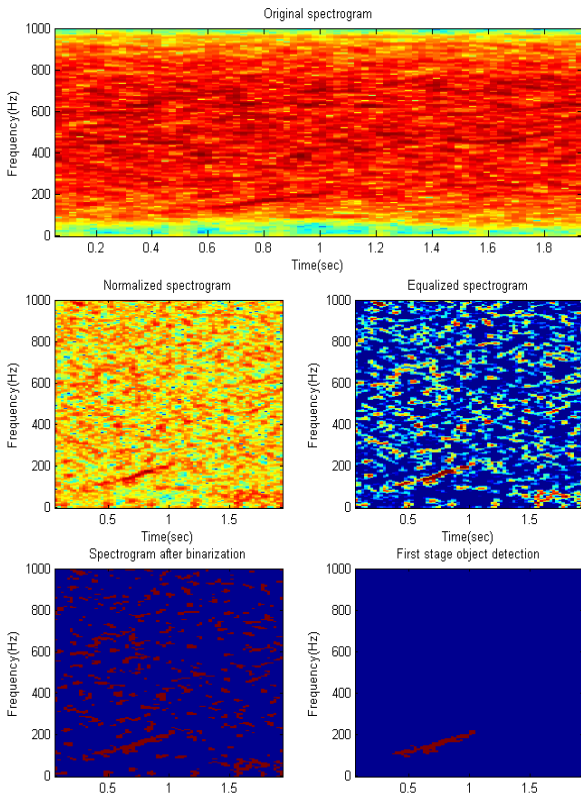


Fig. 3. The outputs of preprocessing steps and first stage detector leading to potential upcall decision

The first stage of detection categorizes the audio signal segments in which no objects are found into the "non-upcall" class. Any signal segment which does not belong to the 'non-upcall' class is fed to the second stage to determine if there is an upcall in the segment. At this point, a feature vector has to be computed for all objects considered as potential upcalls. For this purpose, a set of features named "TFP-2 features" are extracted from detected objects. These features are: minimum frequency (Hz), maximum frequency (Hz), frequency band (Hz), perimeter (pixels), area (pixels), orientation (degree) and time duration (sec).

2.2. Texture based approach

Most of the upcalls in our data set have been observed to occur within the frequency range of 80 Hz and 320Hz, respectively. Hence, the first step of preprocessing in this approach involves a band-pass filter that limits the frequency range of NARW calls. The next step which is similar to that of the first method is to run normalization and equalization algorithms on the spectrogram in order to enhance the upcalls and remove clutter.

A 3x3 median filter is applied to the resulting image to smooth the spectrogram and enhance the contour edges. Fig. 4 shows the spectrogram before and after median filtering. This is followed by hard-thresholding of the pixels at 70% of the maximum intensity of the spectrogram. Since the pixels along the upcall contour are expected to have high intensities, this approach is capable of

finding locations in the spectrogram where the probability of an upcall presence is very high. Then, all objects in the spectrogram are detected by the technique described in Section 2.1 which keeps only those objects with intensity larger than a threshold value for feature extraction as illustrated in the bottom spectrogram of Fig. 4.

LBP is a feature extraction method which is capable of describing texture patterns in the image [12]. To detect an upcall in the image, the LBP operator scans the entire spectrogram using a circular 8-point neighborhood of radius 1 yielding to the LBP image as depicted in the top image of Fig. 5. It is evident from the image that the upcall in the bottom image of Fig. 4 is preserved after the LBP operation. Feature vectors are subsequently derived from the LBP histogram, as depicted in the bottom image of Fig. 5, following the method described in [12].

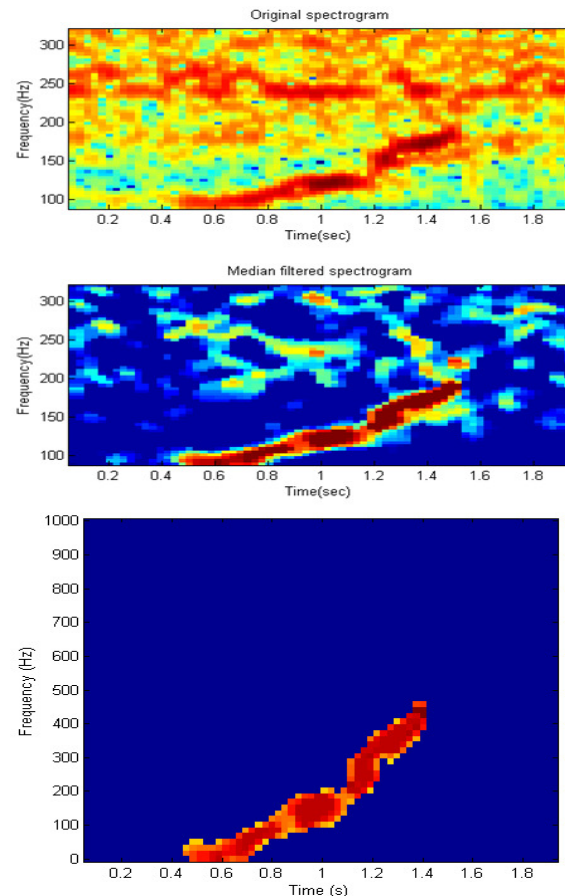


Fig. 4. Original spectrogram (top), output of median filter (middle), and output of high-intensity region selection (bottom)

3. DETECTION RESULTS

The effectiveness of the features described in the previous section is evaluated for NARW upcall detection with various popular classifiers. In the training phase, 4000 right whale audio segments are utilized, in which there are 1265 NARW upcalls and 2735 non-upcalls. In addition, the classification methods for upcall detection are applied

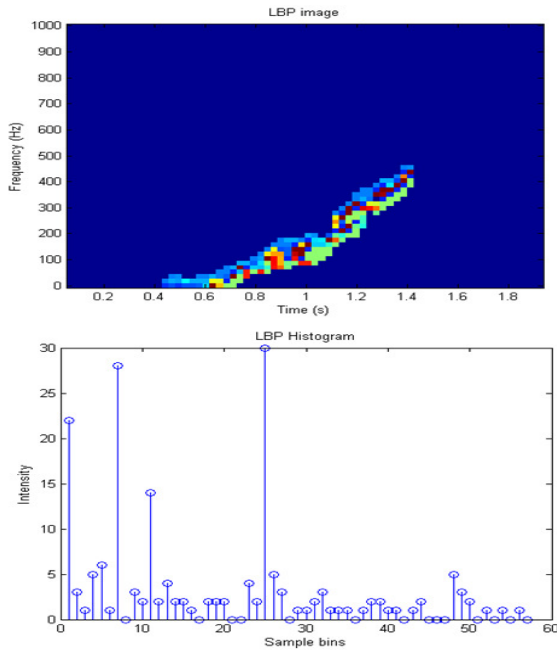


Fig. 5. LBP image (top), LBP histogram (bottom)

on 3000 NARW audio segments in which there are 699 upcalls and 2301 non-upcalls. In this section, three different detection rates are used to analyze detection results:

$$\text{Overall detection rate} = \frac{\text{number of correctly classified calls}}{\text{total number of calls}} \quad (3)$$

$$\text{Upcall detection rate} = \frac{\text{number of correctly classified upcalls}}{\text{total number of upcalls}} \quad (4)$$

$$\text{Non-upcall detection rate} = \frac{\text{number of correctly classified non-upcalls}}{\text{total number of non-upcalls}} \quad (5)$$

	Upcall detection rate (%)	Non-upcall detection rate (%)	Overall detection rate (%)
LDA	80.11	95.21	91.7
QDA	67.38	92.78	86.87
KNN	63.23	95.74	88
Decision Tree	31.18	98.52	83.97
Linear SVM	70.81	97.08	90.97
TreeBagger	76.25	95.83	91.27

Table 1. Detection results using TFP-2 features

Table 1 shows the detection results obtained from various classifiers using TFP-2 features. It is observed that the highest rate of correct detection is achieved by Linear Discriminant Analysis (LDA) 80% corresponding to 560 upcalls followed by TreeBagger (533 upcalls) and Linear SVM (495 upcalls). Although Decision Tree performance is very poor in detection of NARW upcalls followed by KNN but the best non-upcall detection rate is achieved by Decision Tree 98.5% (2267 non-upcalls) and Linear SVM 97% (2234 non-upcalls). The last column also reveals that LDA, Treebagger, and Linear SVM with 91.7%, 91.27%,

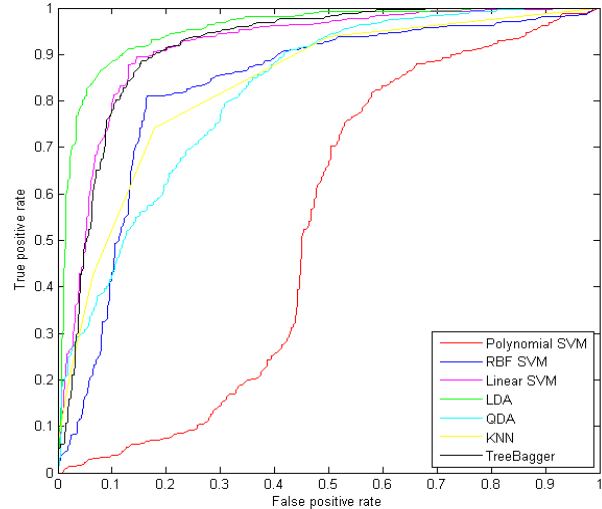


Fig. 6. ROC plot of different classifiers using TFP-2 features

and 90.97%, respectively, demonstrate better performance in terms of overall detection accuracy. To summarize the entire classification results, the Receiver Operating Characteristics (ROC) curves which are the plot of true positive rate (correctly classified upcall) against false positive rate (non-upcall classified as upcall) are shown for all scenarios in Fig. 6. The closer the ROC curve follows the vertical axis and then the top border of the Fig., the more accurate the classifier is. The above conclusion is also proven since LDA, Treebagger, and Linear SVM curves, respectively, tend to achieve high true positive and low false positive. Therefore, the area under these ROC curves is greater than others.

	Upcall detection rate (%)	Non-upcall detection rate (%)	Overall detection rate (%)
LDA	72.96	97.82	92.03
QDA	78.40	94.44	90.70
KNN	78.97	95.35	91.53
Decision Tree	57.79	90.65	83
Linear SVM	90.41	93.44	92.73
TreeBagger	89.98	93.48	92.67

Table 2. Detection results using LBP features

Detection results are given in Table 2 using LBP features. In terms of best upcall detection rate, Linear SVM with 90.41% accuracy (632 upcalls) outperforms the other classifiers followed by TreeBagger with 89.98% accuracy (629 upcalls). On the other hand, LDA is capable of obtaining 98% accuracy in non-upcall detection corresponding to 2251 upcalls. It is also interesting that although KNN is a very simple classifier but it can achieve a high non-upcall detection rate (95%) while keeping the upcall detection rate acceptably high (79%). The best detection performance is achieved by Linear SVM, TreeBagger, and

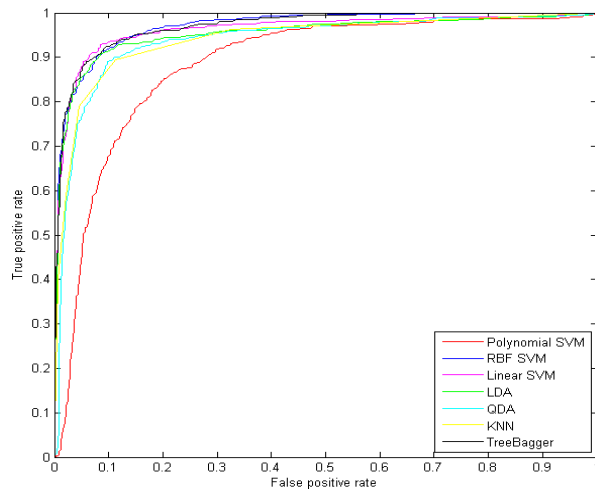


Fig. 7. ROC plot of different classifiers using LBP features

LDA with accuracy of around 92-93%. For the overall comparison of classifiers tested with LBP features, their ROC curves are plotted in Fig. 7 confirming the above claim since these three classifiers have the largest area under the curve. Comparing ROC curves in Fig. 6 and Fig. 7 reveals that classifiers with LBP features have gained about 3% to 4% accuracy improvement over TFP-2 features with an identical classifier.

4. CONCLUSION

Features of NARW upcalls were extracted using both contour and texture based algorithms. Various classifiers such as LDA, SVM, and Tree bagger are paired with these feature extractors and their detection results are analyzed. Considering TFP-2 features, the detector that acquired the highest accuracy ~ 91% is TreeBagger since this approach creates an ensemble of decision trees where every tree is grown on an independently drawn bootstrap replica of input data. On the other hand, LDA is observed to detect the highest number of upcalls and linear SVM demonstrates the least false negative rate. With only LBP features, LDA, linear SVM, and Tree Bagger are amongst the high-ranked detectors with accuracies close to 93%. It seems that LBP features exhibit linear characteristics that discriminant analysis and SVM with linear kernels can well distinguish upcalls in the dataset. The largest percent of upcall detection (true positive) belong to linear SVM and on the other hand, LDA produces the least number of non-upcall misclassifications. An important observation is that switching from TFP-2 features to LBP features produces considerably high detection rate with all classifiers tested which once again indicate highly-informative property of the LBP features.

ACKNOWLEDGMENT

The authors would like to acknowledge the financial support from a FAU Seed Grant.

REFERENCES

- [1] S.D. Kraus, M.W. Brown, H. Caswell, C.W. Clark, M. Fujiwara, P.K. Hamilton, R.D. Kenney, A.R. Knowlton, S. Landry, C.A. Mayo, W.A. McLellan, M.J. Moore, D.P. Nowacek, D.A. Pabst, A.J. Read, and R.M. Rolland, "North Atlantic Right Whale in Crisis," *Science*, vol. 309, no. 5737, pp. 561-562, 2005.
- [2] R.R. Reeves, B.D. Smith, E.A. Crespo, and G. Notarbartolo di Sciari, "Dolphins, Whales and Porpoises: 2002-2010 Conservation Action Plan for the World's Cetaceans," Gland/Cambridge, Switzerland/U.K.: IUCN/SSC Cetacean Specialist Group, 4th chapter, 2003.
- [3] C.W. Clark, "The Acoustic Repertoire of the Southern Right Whale, a Quantitative Analysis," *Anim. Behav.*, vol. 30, pp. 1060-1071, 1982.
- [4] C.W. Clark, M. W. Brown, and P. Corkeron, "Visual and Acoustic Surveys for North Atlantic Right Whales, *Eubalaena Glacialis*, In Cape Cod Bay, Massachusetts, 2001-2005: Management Implications," *Mar. Mamm. Sci.* vol. 26, pp. 837-854, 2010.
- [5] D.K. Mellinger, "A Comparison of Methods for Detecting Right Whale Calls," *Can. Acoust.*, vol. 32, no. 2, pp. 55-65, 2004.
- [6] L.M. Munger, D.K. Mellinger, S.M. Wiggins, S.E. Moore, and J.A. Hildebrand, "Performance of Spectrogram Cross-Correlation in Detecting Right Whale Calls in Long-Term Recordings from the Bering Sea," *Can. Acoust.*, vol. 33, pp. 25-34, 2005.
- [7] D.K. Mellinger, and C.W. Clark, "Recognizing Transient Low-Frequency Whale Sounds by Spectrogram Correlation," *J. Acoust. Soc. Am.*, vol. 107, no. 6, pp. 3518-3529, 2000.
- [8] D.K. Mellinger, "ISHMAEL 1.0 User's Guide," NOAA Technical Memorandum OAR PMEL-120, available from NOAA/PMEL, 7600 Sand Point Way, NE, Seattle, WA 98115-6349, 2001.
- [9] D. Gillespie, "Detection and Classification of Right Whale Calls Using Edge Detector Operating on a Smoothed Spectrogram," *Can. Acoust.*, vol. 32, no. 2, pp. 39-47, 2004.
- [10] I.R. Urazghildiev, and C.W. Clark, "Acoustic Detection of North Atlantic Right Whale Contact Calls Using the Generalized Likelihood Ratio Test," *J. Acoust. Soc. Am.*, vol. 120, pp. 1956-1963, 2006.
- [11] I.R. Urazghildiev, and C.W. Clark, "Acoustic Detection of North Atlantic Right Whale Contact Calls Using Spectrogram-Based Statistics," *J. Acoust. Soc. Am.*, vol. 122, no. 2, pp. 769-776, 2007.
- [12] M. Esfahanian, H. Zhuang, and N. Erdol, "Local Binary Patterns for Classification of Dolphin Whistles," *J. Acoust. Soc. Am.*, vol. 134, no. 1, 2013.