

Computer-aided classification of bowhead whale call categories for mitigation monitoring

Delphine Mathias
Scripps Institution of Oceanography
Marine Physical Laboratory
9500 Gilman Drive
La Jolla, CA, 92093-0238
Email: delphine.mathias@gmail.com

Aaron Thode
Scripps Institution of Oceanography
Marine Physical Laboratory
9500 Gilman Drive
La Jolla, CA, 92093-0238

Susanna B. Blackwell
and Charles Greene Jr.
Greeneridge Sciences
1411 Firestone Road
Goleta, CA, 93117

Abstract—Since 2001 Directional Autonomous Seafloor Acoustic Recorders (DASARs) have been used to localize and record bowhead whale (*Balaena mysticetus*) calls during their annual migration. In 2007 DASARs were deployed at 35 locations over a 280 km swath in the Beaufort Sea, during seismic exploration activities (Fig. 1), in order to monitor potential changes in the animals' location and/or acoustic activity during the seismic activities.

The large amount of acoustic data generated (about 50 days per DASAR) motivated the development of computer-aided methods to assist in detecting and classifying bowhead whale calls.

Bowhead whale calls can be classified in various ways. Here, we divide calls into six categories: (1) upsweeps, (2) downsweeps, (3) constant calls, (4) u-shaped and (5) n-shaped undulated calls, and (6) complex calls, a catch-all category that covers both frequency-modulated calls with multiple inflections, and amplitude-modulated calls such as warbles, growls, and other such sounds. In addition, walrus and bearded seal calls can produce similar call features in a spectrogram, yielding a total of eight classification categories. The frequency range, duration, and fine structure of individual calls vary considerably even within each category, creating difficulties when using simple matched-filtering or spectrogram correlation methods.

A manually reviewed test dataset was assembled, containing examples from each call category, arranged by signal-to-noise ratio (SNR) in 5 dB bins, ranging from 5 to 40 dB. The dataset was then used to test several methods for extracting relevant parameters from the signal for subsequent classification. Contour tracing methods that estimate frequency bandwidth, inflection points, and duration were examined, as well as other boundary descriptors that utilize standard image segmentation techniques. An optimization procedure was then used to determine appropriate decision boundaries for optimum statistical classifiers.

I. INTRODUCTION

The bowhead whale, *Balaena mysticetus*, is the only baleen whale endemic to the Arctic. The Bering-Chukchi-Beaufort stock, often called the western Arctic stock, annually migrates between the northern Bering Sea (where they primarily feed in the summer) and the Canadian Beaufort Sea during the spring, and returns to the Bering Sea in the fall. [2] During their westward migration across the Alaskan Beaufort and Chukchi seas between August and October, whales encounter oil industry exploration and extraction activities as they swim over the continental shelf in shallow depth (between 20m and

50m) and relatively close to the shore (20-60 km offshore) [3] [4].

The acoustic repertoire of bowhead whales is quite complex and includes frequency-modulated (FM) calls, amplitude-modulated (AM) calls, pulsed calls, songs and sequences of repeated identical calls [5] [6]. The biological function of these calls remains unknown.

Both matched-filter [7] and spectrogram correlation [8] techniques have been used to detect and classify bowhead calls. The high variance in call duration, frequency range, and FM modulation structure of bowhead calls limits the practicality of these analysis techniques, which assume little variability in call parameters between calls. Tracing frequency contours in the spectrogram domain has been studied for both dolphins [9] and blue whales [10], but both suffer from difficulties in tracing steep FM upsweeps and downsweeps, and connecting momentary "breaks" in calls.

Since 2001, Directional Autonomous Seafloor Acoustic Recorders (DASARs) [1] have been deployed in the Alaskan Beaufort Sea to record bowhead whale calls and monitor potential changes in the animals' locations and/or acoustic activity during the seismic activities. This paper discusses a computer-aided method, based on morphological and optimization processes, that assists in detecting and classifying the large amount of bowhead whale calls.

II. METHODS

A. Data Sets

In 2006, four DASARs were deployed in pairs at two sites in the Beaufort Sea between late August and the beginning of October near Deadhorse, Alaska. The DASARs were placed on the seafloor, at depths of 18 m and 37 m. The DASARs recorded at 1kHz sampling rate between September 12 and October 1. One year later, between August and October 2007, DASARs were deployed at 35 locations between Harrison Bay and Kaktovic over a 280 km swath, divided into five sites, each site arranged as a set of equilateral triangles with 7 km sides. (Fig. 1). In the results that follow, most call samples were culled from the 2006 data set, with supplements from the 2007 data for rarer types of calls.

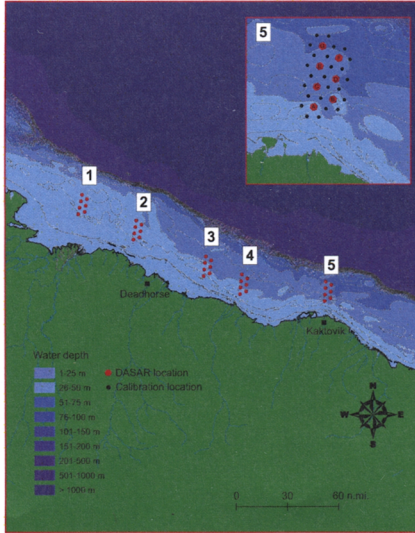


Fig. 1. Location of the 35 DASARs deployed in 2007 in the Beaufort Sea, grouped by site.

B. Manual Analysis

A group of analysts manually analyzed the 2006 dataset by listening to recordings and examining spectrograms. A call detected by more than one DASAR was logged as a single call. Calls were assigned into 10 categories :

- (1) Upcall, an FM (frequency modulated) up-sweep;
- (2) Downcall, an FM down-sweep;
- (3) Constant call, small FM tolerated, but generally no change;
- (4.1) Inflected call, FM undulation down, then up ("U"-shaped);
- (4.2) Inflected call, FM undulation up, then down ("n"-shaped);
- (5) High call, an FM sweep with most energy higher than 200 Hz;
- (6) Complex call, a mixture of FM and/or AM (amplitude modulation), often with some broadband energy;
- (7) Slap: either breach or fluke slap, or sharp report;
- (8) Seal call;
- (9) Ping;
- (10) Other;

A test dataset was then assembled, containing examples of the most common call types (i.e. call types 1, 2, 3, 4.1, and 4.2), and arranged by signal to noise ratio (SNR) in 5 dB bins, ranging from 5 dB to 40 dB (Fig. 2). The SNR was estimated by computing the root-mean-square (rms) level of the call over the frequency band visible on the spectrogram, and then dividing by the rms level of background noise measured one second before the call, and filtered to cover the same bandwidth as the call. A second manual review by a single person was performed to ensure consistency of the original

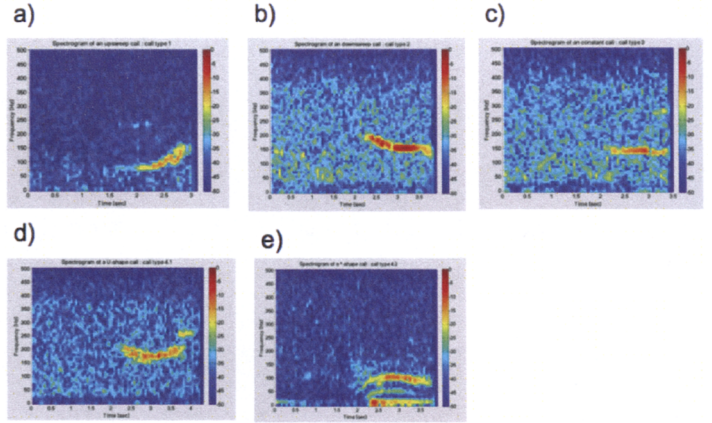


Fig. 2. Spectrograms of calls of type 1 "upsweep" (a), type 2 "downsweep" (b), type 3 "constant" (c), type 4.1 "U-shaped" (d), and type 4.2 "n-shaped" (e).

manual analysis.

The time series for each test call begins 1s before the call begins, to permit background noise levels to be measured via spectrogram equalization techniques (aka "clutter map constant false-alarm rate" methods [11]).

A test dataset containing 1000 calls (265 upsweep, 271 downsweep, 202 constant, 161 u-shaped, and 61 n-shaped) were collected. Half of the test dataset was used as a "training" dataset while the second half was used to test the performance of the classification method (i.e. as a "validating" dataset). Fig.2 shows spectrograms of calls of type 1,2,3,4.1, and 4.2.

C. Morphological processing and contour extraction

The first stage of the classification process consists in constructing spectrograms of each call and extracting the call contours. The following sequence was applied to every call sequence of the test dataset :

- (1) A spectrogram representing the power spectral density as a function of time and frequency is plotted (Fig. 3a).
- (2) The spectrogram is equalized using the median value of the background noise spectrum, using the first second of the time series.
- (3) The image is converted into a grayscale, and then binary image, (Fig. 3b), using Otsu's method to estimate the threshold of the on/off pixels. Otsu's method is a global thresholding method that minimizes the variances in intensity within the "on" and "off" class, thus maximizing the variance between the classes [12].
- (4) Pixels corresponding to frequencies below 50 Hz and above 350 Hz are set to 0 (background value). This is done to diminish the risk of tracing background noises or seal calls (Fig. 3c).
- (5) A morphological processing method called "opening" [13] is used to clear out small time-bandwidth product objects (Fig. 3d).
- (6) A morphological dilation of the image is used to fill in

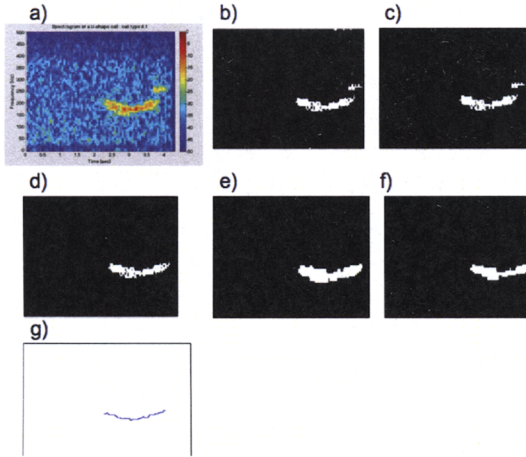


Fig. 3. (a) : original spectrogram of a U-shape call ; (b) : after conversion to a binary image ; (c): pixels corresponding to frequencies below 50 Hz and above 350 Hz are set to 0; (d): after opening, (e) after dilatation, (f) after erosion, (g) final line contour.

gaps of the contour (Fig. 3e).

(7) A morphological erosion mitigates the effects of dilation by removing perimeter pixels from larger image objects (Fig. 3f).

(8) The largest time-bandwidth product object in the image is selected.

(9) For every time bin (vertical slice) of the image, the median frequency of the 'on' pixels is estimated, reducing the object to a line (Fig. 3g).

Finally, a polynomial fit is performed on the curve to extract parameters describing the contour slope and curvature. Each contour line is divided into N sections (where $N = 1$ or 3), and the coefficients of a polynomial $P(t)$ of degree D ($D=1$ or 2) that fits each section of the contour are computed. The slope and the 2nd order derivative of each section fit is then stored.

The slope and 2nd order derivatives are derived using these formulas :

$$P(t) = a_0 + a_1 * t + a_2 * t^2$$

$$\text{slope} = a_1 + 2 * a_2 * < t > [\text{Hz/sec}]$$

$$\text{curvature} = 2 * a_2 [\text{Hz/sec}^2]$$

where $< t >$ is the average time of the curve fit, or $T_{duration}/2$, with $T_{duration}$ the duration of the call portion subjected to the polynomial fit.

The following figures display the normalized slope, i.e. the slope divided by the maximum slope encountered across all call types.

D. Classification of contours under call types

In this work we explore the efficacy of three different types of polynomial fit. The first, and simplest method, is to estimate a mean slope to the contour (e.g. fit a straight line). The second

method is to split each call into three sections, and estimate the slope for each section. For call types 1-3, the average of these slopes is taken to be the evaluation parameter. For call types 4.1 and 4.2, thresholds for slope are set for each section. The final method is to perform a second-order fit to the entire contour, yielding two parameters per call: the first and second-order derivatives. We show preliminary results for this final method.

E. Optimization procedure

The optimization computation uses the Matlab Genetic Algorithm and Direct Search Toolbox. The direct search algorithm searches the parameter space, looking for a point where the value of an objective function is lower than the value at the current parameter values. The parameter space here is a set of slope thresholds that are used to divide upslopes from downslopes and constant calls.

The objective function uses here is defined as follows:

$$\sum (\text{weight}_i * (\max([\text{bad_classification_fraction}_i - \text{goal}; 0]))^2)$$

where i represents a call type, $\text{bad_classification_fraction}_i$ is the number of misclassified type i calls divided by the total number of type i calls tested, and goal represents a minimum performance level (missed fraction) desired by the system. The quantity weight_i represents the weighting attributed to each call type. For instance if $\text{weight} = [1 \ 1 \ 10 \ 1]$, it would mean that it is 10 times more important to classify well call types 4.1 compared to other call types. In the preliminary results shown here each weight value has been set to unity, and the goal has been set to 20%. Eventually the weight vector will be set to the probability of occurrence of each call type [6].

The optimization procedure was applied to half the dataset, chosen randomly from the complete data set (the training dataset). Then the optimized parameters are then used to classify the second half of the test dataset (the validating dataset).

III. RESULTS

A. Morphological processes and contour tracing

Using the detections procedure described in Sec. II.C, a certain fraction of calls failed to be detected. 9.8% of the upsweep calls (call type 1), 15.5% of the downsweep calls (call type 2), 13.3% of the constant calls (call type 3), 7.6% of the U-shape calls (call type 4.2) and 9.8% of the "n"-shape calls (call type 4.2) were missed.

B. Optimization results

1) Classification of the "simple" call types : upsweep calls, downsweep calls, constant calls based on mean value:

Fig. 4 shows the distribution of mean slope values for call types 1,2,3, derived from a first-order polynomial fit . Only 2 optimization parameters are needed in this case—the threshold values separating a downsweep from a constant value, and a constant value from an upsweep. The initial parameters (before optimization) are shown as dashed vertical lines in Figure 4.

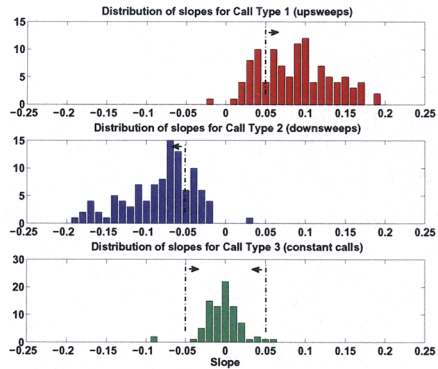


Fig. 4. Distribution of average slope values for N=1 for call types 1,2, and 3. Initial threshold guesses are drawn with vertical black lines.

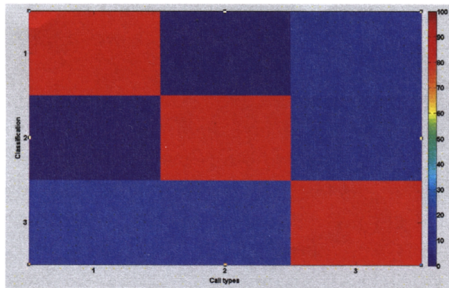


Fig. 5. Optimization: Classification of the call types 1,2,3 using mean slope, N=1.

Fig. 5 shows the classification results for the training dataset and Fig. 6 show the classification results for the validating dataset. The "simple" classification procedure achieves the goal of 80% good classification percentage for the three call types considered. When the optimized parameters are applied to the validating dataset, the 80% good classification percentage for the three call types is still attained. Note that the upsweeps (call type 1) and downsweeps (call type 2) bad classifications mostly fall in the constant call class, as expected. The constant call bad classifications mostly fall in the downsweep class. Our general experience suggests that human analysts are inconsistent in applying the "constant" label to calls.

2) Classification of all call types based three-part segmentation of contours: N=3:

Fig. 7 shows the distribution of mean slope values for N=3 for call types 1,2,3, where each row represents a different section of the call. Fig. 8 shows the distribution of slopes for the three sections of types 4.1 and 4.2, which are expected to vary for each section. Eleven thresholds are needed to distinguish the five call types; four for type 4.1, four for type 4.2, one for type 1, one for type 2, and one for type 3. All thresholds are shown as vertical black lines in the figures.

A test call is first subjected to the type 4.1 and 4.2 threshold

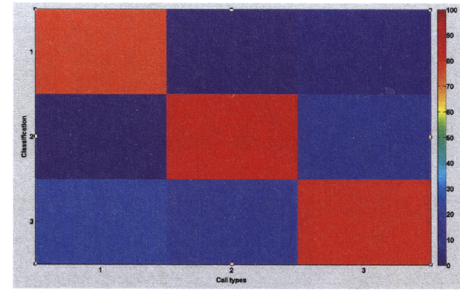


Fig. 6. Validation: Classification of the call types 1,2,3 using mean slope, N=1.

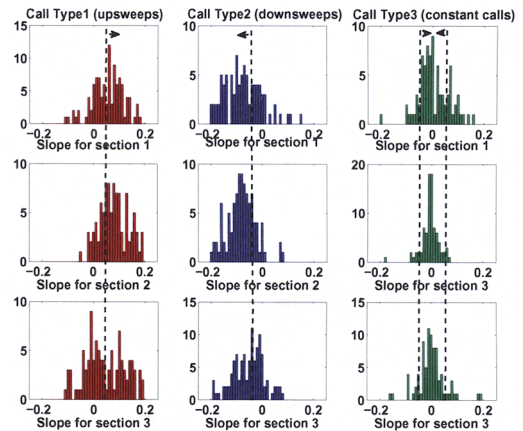


Fig. 7. Distribution of slope values for N=3 for call types 1,2,3. The three optimization thresholds are drawn with vertical black lines.

tests indicated in Figure 8. If it fails both tests, it is then subjected to the less stringent threshold tests shown in Figure 7. In principle a call can fail all tests and thus be classified as "other".

Fig. 9 shows the classification results for the training dataset, and Fig. 10 show the classification results for the validating dataset. When adding call types 4.1 and 4.2 in the optimization procedure, the overall good classification percentage decreases to 70%. When applying the optimized parameters to the validating dataset, the good classification percentage decreases for call types 1 and 4.2, remains stable for call type 2 and increases for call types 3 and 4. The average good classification percentage is now 63%.

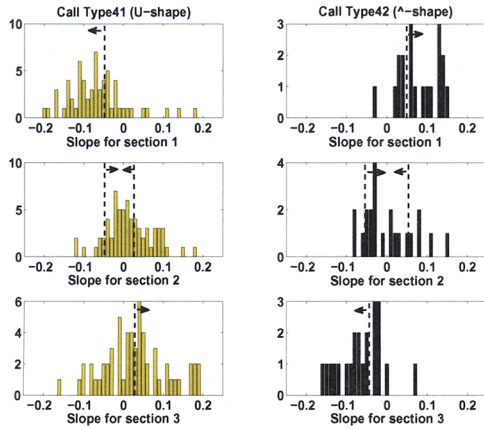


Fig. 8. Distribution of slope values for $N=3$ for call types 4.1 and 4.2. The eight optimization thresholds are drawn with vertical black lines.

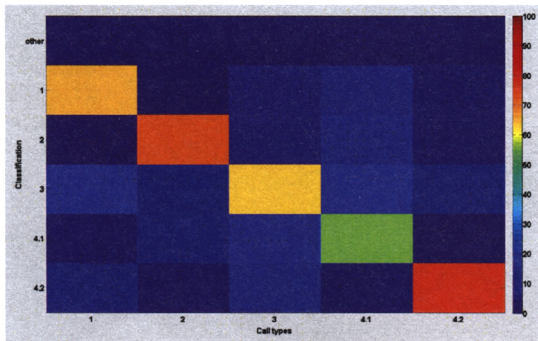


Fig. 9. Optimization: Classification of all call types using $N=3$.

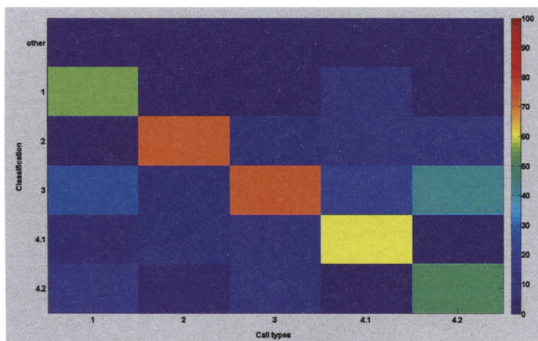


Fig. 10. Validation: Classification of all call types using $N=3$.

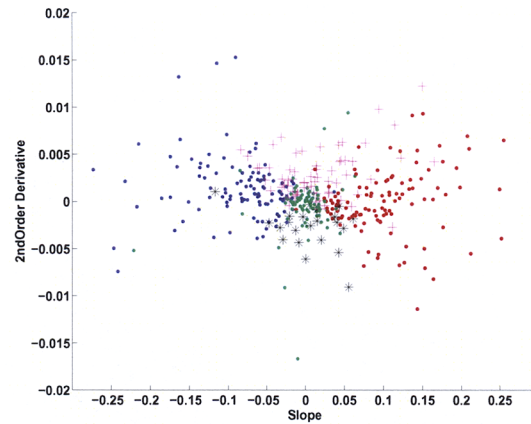


Fig. 11. 2nd order derivative versus slope for $N=1$ for call types 1 (red), 2 (blue), 3 (green), 4.1 (magenta), 4.2 (black).

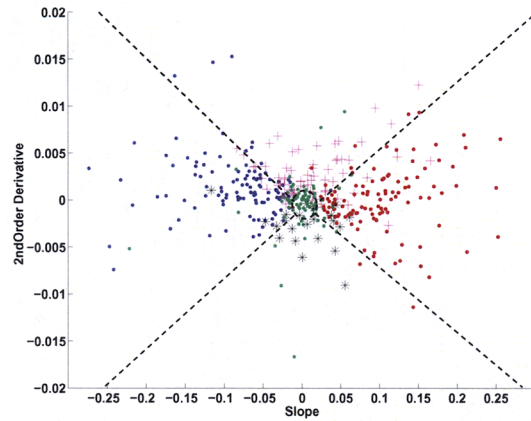


Fig. 12. 2nd order derivative versus slope for $N=1$ for call types 1 (red), 2 (blue), 3 (green), 4.1 (magenta), 4.2 (black). Optimization thresholds are drawn with dashed black lines

3) Classification of all call types based on slopes and 2nd order derivative values - $N=1$:

The final polynomial fit applies a second-order polynomial to the entire contour and estimates the slope and second-order derivative from the expressions in Section II.C. Figure 11 shows the distribution of all call types in this two-dimensional parameter space, while Figure 12 shows an initial attempt to divide the parameter space into the appropriate call types. One can draw a circle around the origin to define the "constant" type 3 call type, and then 4 lines can be drawn to separate call type classes. One can see that the U and n-shaped calls tend to have a larger second-order derivative for a given slope than the simpler call types, permitting this distinction. Optimization is a work in progress.

IV. DISCUSSION

The morphological processes used in this study miss only 11% of the test data, even though the test dataset included calls with SNR as low as 5 dB. So far we have been able to classify the call types 1,2,3 with a good classification fraction of 80%. However, when we add the more complex call types (4.1 and 4.2) in the optimization procedure, the good classification percentage falls to 63%. Three possibilities to improve those results emerge from this analysis.

First, due to the variability within each call type, the manual classification can be very subjective. Indeed, the results suggests that human analysts are inconsistent in applying the "constant" label to calls. For instance, the combination of a constant section ended by a small upswEEP tail could be classified as a constant call by one manual analyst and as a upswEEP call by somebody else. So one step would be to review the incorrect classifications and determine whether the manual classifications are appropriate.

Second, the test dataset contains different numbers of calls for each call type. Changing the weight in the optimization fit function according to the frequency of occurrence of each call type could help improve the classification results. In particular, call types 4.1 and 4.2 are giving the lowest classification percentage, but are also the call types the least represented in field data.

Finally, as mentioned in Sec. III.B.3, adding the 2nd order derivative in the optimization parameters could also help separating calls types 4.1 and 4.2 from call types 1,2 and 3. Figure 11 supports this conclusion, so work continues on optimizing this approach.

ACKNOWLEDGMENT

The authors would like to thank Greeneridge Sciences and Shell Exploration and Production Company (SEPCO) for support.

REFERENCES

- [1] C.R. Greene, Jr., M.Wm. McLennan, R.G. Norman, T.L. McDonald, R.S. Jakubczak, W.J. Richardson, *Directional frequency and recording (DIFAR) sensors in seafloor recorders to locate calling bowhead whales during their fall migration*. J. Acoust. Soc. Am. 116 (2), 2004.
- [2] S.E. Moore and R.R. Reeves, *Distribution and movement*. In: Burns, J.J., Montague, J.J., and Cowles, C.J., eds. The bowhead whale. Special Publication No. 2, Society for Marine Mammalogy. Lawrence, Kansas: Allen Press. 313-386, 1993.
- [3] S.E. Moore, J.T. Clarke, and D.K. Ljungblad, *Bowhead whale (Balaena mysticetus) spatial and temporal distribution in the central Beaufort Sea during late summer and early fall 1979-86*. Report of the International Whaling Commission 39:283-290, 1989b.
- [4] S.D. Treacy, *Aerial surveys of endangered whales in the Beaufort Sea, fall 2001*. OCS Study MMS 2002-061. Minerals Management Service, Alaska OCS Region, Anchorage, Alaska, USA, 2002.
- [5] W.J. Richardson, C.R. Greene, Jr., C.I. Malme, D.H. Thomson, *Marine Mammals and Noise*. Academic Press, San Diego, 1995.
- [6] S.B. Blackwell, W.J. Richardson, C.R. Greene, Jr. and B. Streever, *Bowhead Whale (Balaena mysticetus) Migration and Calling Behaviour in the Alaskan Beaufort Sea, Autumn 2001-04: An Acoustic Localization Study*. Artic, vol. 60, no. 3, p. 255-270, 2007.
- [7] I.R. Urazghildiev and C.W. Clark, *Acoustic detection of North Atlantic whale contact calls using the generalized likelihood ratio test*. J. Acoust. Soc. Am. 120 (4), 2006.
- [8] D.K. Mellinger and C.W. Clark, *Recognizing transient low-frequency whale sounds by spectrogram correlation*. J. Acoust. Soc. Am. 107 (6), 2000.
- [9] S. Datta and C. Sturtivant, *Dolphin whistle classification for determining group identities*. Signal Processing (82), p.251-258, 2002.
- [10] S.K. Madhusudhana, E.M. Oleson, M.S. Soldevilla, M.A. Roch, J.A. Hildebrand, *Frequency based Algorithm for Robust Contour Extraction of Blue Whale B and D calls*. IEEE Oceans, 2008.
- [11] N. Levanon, *Radar Principles*. A Wiley-Intersciences publication, 1988.
- [12] N. Otsu, *A Threshold Selection Method from Gray-Level Histograms*. IEEE Transactions on Systems, Man, and Cybernetics, Vol. 9, No. 1, pp. 62-66, 1979.
- [13] R.C. Gonzalez and R.E. Woods, *Digital Image Processing*. Pentice Hall, New Jersey, 2001.