Underwater Acoustic Monitoring Using MFCC for Fuzzy C-Means Clustering, Naive-Bayes and Hidden Markov Model-Based Classifiers

Cristian Alzate¹, Samira Ortiz¹ and Vidya Manian¹
¹Department of Electrical & Computer Engineering, University of Puerto Rico at Mayaguez
Mayaguez, Puerto Rico

Abstract - The whale sounds help researchers in population assessments and to follow the migratory path of whales. Acoustics is the best way to study and observe cetaceans since it is automatic and non-invasive. A technique capable of differentiating between whale songs, other marine sounds and man-made sounds would be very useful for the scientific community. This paper presents a system that can classify signals collected from humpback whales vocalization data through acoustic monitoring from Puerto Rico coastal marine habitats. The fuzzy c-means algorithm is used to cluster, in an unsupervised manner, sounds from the same marine source. The Naive-Bayes classifier and a Hidden Markov Model are used to classify marine sounds such as whale songs, man-made sounds and natural background noise from hydrophone recordings under water. The MFCC (Mel-Frequency Cepstral Coefficients) approach is used to extract the most important characteristic of the signals and a vector quantization and clustering approach is used to obtain the states for training the HMM. Satisfactory results are obtained through HMM classifier which outperforms the Naive-Bayes classifier for whale songs and marine sounds.

Keywords: Classifier, Clustering, Hidden Markov Model, Humpback whales, Naïve-Bayes

1 Introduction

A humpback whale song consists of units of sound which combine to form what is known as a phrase. An unbroken sequence of similar phrases is a theme. Themes are sung in a specific order; and several distinct themes combine to form a song. Once completed, the whale will return to the beginning and start again. Whales will sing for hours or days at a time. Whales in different areas sing in different dialects. And, the songs change gradually over time. [1][2]

Although there is not a sure explanation about the purpose of the songs, they allow researchers to follow the migratory path of whales throughout this season and also to estimate population density within an area. The whale watching season in Puerto Rico is between December and March and reaches its peak in February. The rare and impressive Humpback whales visit Puerto Rico every year on their migratory route to the North Atlantic where they mate during the summer months. Hence, in an effort to monitor marine ecological habitats in the coasts of the island of Puerto Rico and gain new knowledge and further the understanding of harbor acoustic monitoring, we set out to implement different techniques to extract songs features, cluster whales individuals based on their singing patterns and classify underwater sounds to differentiate whale songs from other marine sounds, background and man-made sounds.

With growing concern about environmental problems, cetaceans and their habitats are receiving significant attention and studies have been developed for location and cetacean behavior analysis [3] [4]. However there is not much work in classification of acoustic signals from marine habitats especially those around Puerto Rico. There has been some analysis of West Indian mammals’ vocalization [5] [6] and marine mammal behavioral response in southern California [7]. This paper presents the results of clustering and classification of signals recorded from acoustic monitoring of humpback whales in Rincon and Boqueron bay area in Puerto Rico. Due to highly song nature and underwater channel distortions, it has been a challenge the task of classification.

The acoustic signals considered in this work are whale songs, background noise, swimming noise, kayak, boat and jestki sounds. We test different techniques to achieve the best results in classification and make a remarkable contribution in the study of cetaceans. A clustering method and two different classification methods are implemented; for clustering the fuzzy c-means algorithm is used and for classification, the Naive-Bayes classifier and Hidden Markov Model. The feature extraction method is the Mel-Frequency Cepstral Coefficients (MFCC). The paper is organized as follows. In section two, we first explain data acquisition and preprocessing methods, later we present the feature extraction, data segmentation to later explain the clustering and classifying methods. Section three presents the experimental results and conclusions are shown in section 4.

2 Methodology

2.1 Data Acquisition and preprocessing

Sound under water has important characteristics; it travels rapidly over long distances due to the fact that seawater is approximately 800 times denser than air [8]. The attenuation in sounds is related to frequency. High frequencies are attenuated faster than low frequencies and, given the same
environmental conditions; the louder the sound the further it can travel.

The active vocalization range of humpback whales has been estimated to be 0.02-8.2 kHz at 15 to 160km. Hydrophone recordings of whale songs have been used in this work. Since the highest frequency component of humpback whales is 8.2 kHz, the sampling rate of the hydrophone was set to be at least 22 kHz.

Once the sounds are recorded, the next step is the signal preprocessing, at this stage, it is extremely important to reduce the background noise so that the sounds can be processed optimally enhancing features and improving classification. In the case of recording marine species like whales, multiple sources of noise can be identified since the underwater sounds is generated by a variety of natural sources such as rain, marine life and breaking waves. It is also generated by a variety of man-made sources, such as military sonars, ships, boats, fishing and so on. The frequency range of ambient noise underwater is 50-100 kHz. This noise is due to spray and bubbles associated with breaking waves [9][10]. It increases with increasing wind speed.

The vocalizations emitted by humpback whales can be defined as broad-band transient signals. Transient signals can be defined as having a short duration and finite energy, whereas their power is zero. For this reason, the correlation function in the analysis of these kinds of sounds is determined using their energy [11]. Fig. 2(a) shows the spectrogram (Short Time Fourier Transform) of an original whale sound, the spectrogram after removal of noise due to ocean waves (Fig. 2(b)) and after complete noise removal (Fig. 2 (c)). From the spectrogram, we can see that the principal frequency components are between 0 and 5 KHz. approximately. The noise from the breaking waves was reduced by creating a sound profile of the source and designing a filter to remove it.

Fig. 1. Marine acoustic data acquisition scheme

Fig. 2. (a) Spectrogram of whale sound, (b) Spectrogram without noise from waves, (c) Spectrogram with full noise duction.
2.2 Feature Extraction using Mel-Frequency Cepstral Coefficients (MFCC)

The Mel-Frequency Cepstral Coefficients is a good parametric representation of acoustic signals. The coefficients are a result of the cosine transform of the real logarithm of the short-term energy spectrum expressed on a Mel-frequency scale. The MFCC calculated for three samples from a whale audio source is shown in Fig. 3. It can be seen that the coefficients are similar as it is for the same individual whale.

![Fig. 3. MFCC applied to different samples of the same whale.](image)

2.3 Fuzzy C-means (FCM)

FCM is a data clustering technique where each data point belongs to a cluster to some degree that is specified by a membership grade the cluster centers, which are updated iteratively. It was proposed by Dunn [12] in 1973 and eventually modified by Bezdek [13] in 1981.

Let \( x_i \) represent the \( N \) data points of \( M \)-dimension to be clustered, where \( i = 1,2,...,N \). And \( C \) the number of clusters to be made. The level of cluster fuzziness \( f \) higher than one is chosen and the membership matrix \( U \) of size \( N \times C \times M \) is initialized at random, such \( U_{ijm} \in [0,1] \) and \( \sum_{j=1}^{C} U_{ijm} = 1.0 \), for each \( i \) and a fixed value of \( m \). The cluster centers \( C_{Cjm} \) for \( j^{th} \) cluster and its \( m^{th} \) dimension is determined by using the expression [14]

\[
C_{Cjm} = \frac{\sum_{i=1}^{N} U_{ijm} x_{im}}{\sum_{i=1}^{N} U_{ijm}^f} \quad (1)
\]

Then the Euclidean distance between \( i^{th} \) data point and \( j^{th} \) cluster center with respect to \( m^{th} \) dimension is defined by:

\[
D_{ijm} = \| (x_{im} - C_{Cjm}) \| \quad (2)
\]

Then, the fuzzy membership matrix \( U \) is updated according to \( D_{ijm} \). While \( D_{ijm} > 0 \)

\[
U_{ijm} = \frac{1}{\sum_{j=1}^{C} (U_{ijm})^{2-f}} \quad (3)
\]

When \( D_{ijm} = 0 \), \( C_{Cjm} \) will have the full membership \( U_{ijm} = 0 \). This procedure is done until \( U \leq \epsilon \), where \( \epsilon \) is a pre-specified termination criterion [14].

2.4 Classification

2.4.1 Naive-Bayes Classifier

The Naive Bayes classifier is a classification method based on Bayes Theorem [15]. \( C_j \) denotes the class of vector \( X \) as belonging to the \( j \)-th class, \( j = 1,2,...,J \) out of \( J \) possible classes. Let \( P(X|C_j) \) denote the probability of the sample vector belonging in the \( j \)-th class given the individual characteristics \( X_1, X_2, ..., X_p \). Furthermore, let \( P(X_1, X_2, ..., X_p|C_j) \) denote the probability of a sample with individual characteristics \( X_1, X_2, ..., X_p \) belonging to the \( j \)-th class and \( P(C_j) \) denote the unconditional (i.e. without regard to individual characteristics) prior probability of belonging to the \( j \)-th class. For a total of \( J \) classes, Bayes theorem gives us the following probability rule for calculating the case-specific probability of a sample vector falling in the \( j \)-th class:

\[
P(C_j|X_1, X_2, ..., X_p) = \frac{P(X_1, X_2, ..., X_p|C_j)P(C_j)}{\text{Denom}} \quad (4)
\]

Where:

\[
\text{Denom} = P(X_1, X_2, ..., X_p|C_1)P(C_1) + \cdots + P(X_1, X_2, ..., X_p|C_J)P(C_J) \quad (5)
\]

All the MFCC of each individual is saved in a data structure named “Codebook”. Each MFCC has a dimension of 20x16. The Naive-Bayes algorithm is trained with the entire codebook.

2.4.2 Hidden Markov model

Hidden Markov Model is used for modeling stochastic processes and sequences in various applications, like natural language modeling, handwriting recognition and voice signal processing [16]. Also it is used to solve problem that has states at time \( t \) that are influenced directly by a state at time (\( t-1 \)). Hidden Markov Model has a number of parameters, whose values are set so as to characterize well the training patterns for the known category. Later, a test pattern is classified by the model that has the highest posterior probability. For \( n \) events \( A_1, ..., A_n \), the chain rule of probability [17] is given by:
Where,

\[ P(A_1 \cap A_2 \ldots \cap A_n) = P(A_1) \prod_{i=2}^{n} P(A_i|A^{i-1}) \]  

(6)

If the event \( A_i \) pertain to the event \( \{X_i = a_i\} \) then,

\[ P(X^n = a^n) = P(X_1 = a_1) \prod_{i=2}^{n} P(X_i = a_i|X^{i-1} = a^{i-1}) \]  

(8)

The formula above show the causal decomposition of the joint distribution of \( x^n \) into a product of conditional distributions of the present value of \( X \) given the past values. \( X^n \) is a Markov Process if the future value \( X^{i+1} \) is independent of the past value \( X_{i-1} \) given the present \( X_i \):

\[ P(X_i = a_i|X^{i-1} = a^{i-1}) = P(X_i = a_i|X_{i-1} = a_{i-1}) \]  

(9)

3 Results

The algorithms were trained using two samples of each of the acoustic sources and were tested using 15 to 20 other audio samples that were not used for training. Audio files are 5 seconds length. They belong to different acoustic sources of whales, jet skis, kayaks, human voices and boats. Features were extracted from the MFCC coefficients of the data. One set of coefficients were calculated for each audio sample.

For clustering, FCM is determining the cluster centers, calculating the Euclidean distances and updating the fuzzy membership matrix until the centroid distances are less than a threshold, which is the minimum distance between centroids and our termination criterion. Once this threshold is computed with just the two samples of each of the acoustic sources, the algorithm is implemented in the testing data set. Fig 6 shows the clustering results after the FCM is implemented to group these sounds, the algorithm precisely detects the number of sources.

For the naive-Bayes classifier, after one set of features extracted from the MFCC coefficients of the data, two samples of each audio file were stored in a codebook to train the network and the others coefficients were used to test the classifier. The algorithm for naive Bayes classifier is described in Fig. 7 (a).

![Fig. 4. Clustering using FCM algorithm](image)

For the naive-Bayes classifier, after one set of features extracted from the MFCC coefficients of the data, two samples of each audio file were stored in a codebook to train the network and the others coefficients were used to test the classifier. The algorithm for naive Bayes classifier is described in Fig. 7 (a).

Tables 1 and 2 show the results for this classifier. In the case of whale songs the accuracy of the algorithm was 72.7%, which corresponds to 30 errors in 110 trials. For the marine sounds, the accuracy was 87.6% which corresponds to 13 errors from 105 trials.

For the classifier based on Markov Model, a segmentation process is used to create a sequence for each sound sample. Each file was segmented into 64 parts, and then the MFCC is calculated for each part and stored into a data structure called codebook. Then, a clustering algorithm based on the Euclidean distance between the codebooks and the coefficients is applied to find similar coefficients and create groups which will represent the hidden states of the HMM. After running the clustering algorithm, a hidden state is assigned to each segment, which results in a sequence for each audio file. Finally, to create the training data set, all the sequences for the same sound source are grouped. Fig. 7 (b) shows the general algorithm. The classification results are shown in Tables 1 and 2 along with the results of Bayes classification for comparison. This classifier results in 95.2% accuracy for classification of whale songs and 95.7 for the other unknown audio files.

![Fig. 5. Algorithm for Naive Bayes (a) and HMM classifier (b).](image)

The HMM is more accurate for whale songs and marine sounds classification. The disadvantage of this classifier is the training time; the segmentation and the creation of sequences are time consuming steps. Table 3 shows the computational run time for both classifier algorithms using Matlab in a computer whose technical specifications are: 24 GB of RAM memory and 2 Intel Xeon processors with a 2.67 GHz clock each. The naive-Bayes classifier is less expensive in computational time compared to HMM classifier, and
depends on the number of sound sources. The computational
time for classification of marine sounds (6 sources) is greater
than that of whale songs (5 sources). The HMM classification
timing cost depends only on the number of files, hence the
computational time for classifying whale songs (110 files) is
greater than that of marine sounds (105 files).

Table 1. Overall accuracy (OA) classification for Whale songs

<table>
<thead>
<tr>
<th>Whales</th>
<th>Naive-Bayes (%)</th>
<th>HMM (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>66.6</td>
<td>89</td>
</tr>
<tr>
<td>B</td>
<td>50</td>
<td>95</td>
</tr>
<tr>
<td>C</td>
<td>50</td>
<td>97.5</td>
</tr>
<tr>
<td>D</td>
<td>100</td>
<td>99.5</td>
</tr>
<tr>
<td>E</td>
<td>97</td>
<td>95</td>
</tr>
<tr>
<td>OA</td>
<td>72.7</td>
<td>95.2</td>
</tr>
</tbody>
</table>

Table 2. Overall accuracy (OA) classifications for Marine Sounds.

<table>
<thead>
<tr>
<th>Marine Sound</th>
<th>Naive-Bayes (%)</th>
<th>HMM (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boat</td>
<td>1</td>
<td>88</td>
</tr>
<tr>
<td>Jestki</td>
<td>66.7</td>
<td>92.3</td>
</tr>
<tr>
<td>Kayak</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Noise</td>
<td>100</td>
<td>99</td>
</tr>
<tr>
<td>Swimming</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Voice</td>
<td>67</td>
<td>95</td>
</tr>
<tr>
<td>OA</td>
<td>87.5</td>
<td>95.7</td>
</tr>
</tbody>
</table>

Table 3. Timing results for the Classifiers.

<table>
<thead>
<tr>
<th>Sound</th>
<th>Time (in seconds)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Naive-Bayes</td>
</tr>
<tr>
<td>Whale songs</td>
<td>2.42</td>
</tr>
<tr>
<td>Marine</td>
<td>4.61</td>
</tr>
</tbody>
</table>

4 Conclusions

In this paper, we implemented a series of algorithms to study
and analyze the humpback whales behavior through their
songs. The series of algorithms implemented for noise
removal, feature extraction, clustering and classification
allowed successfully monitoring marine life in the island of
Puerto Rico and other coastal areas. Classifiers allowed
identifying songs from the same whale, showing an important
fact about whales; each one of them has an individual
signature. Besides, different sound sources could be
successfully clustered and classified.

The feature extraction using Mel-Frequency Cepstral
Coefficients proved to be effective for whale songs and
marine sounds. By using the MFCC for fuzzy c-means
clustering, the number of acoustic sources could be
satisfactorily obtained. HMM classifier performed better than
the naive Bayes classifier in spite of the complexity involved
in finding the sequences and states. The HMM classifier
performs well once a suitable segmentation of the sequences
is obtained and state transitions are set up correctly. Training
the naive-Bayes Classifier is comparatively easier and is less
complex compared to the HMM in time and memory, it works
well with marine sounds and gives an accuracy of 87.5% for
the whale songs, hence the techniques implemented can be
used for whales and other marine mammal surveys and
census.

Acknowledgment

We would like to thank Ms. Mithriel M. MacKay, Director of
Research and Education Marine and Coastal Ecology
Research Center San German, Puerto Rico, and Mr. Jose
Rivera, National Marine Fisheries Service, NOAA for
providing us the humpback whale songs. We would also like
to thank Mr. Samuel Rosario for his technical support in the
data collection and the Laboratory for Applied Remote Sensing and Image Processing, University of Puerto Rico,
Mayaguez for providing the software platforms for data
processing. Finally, we would like to say that the views
presented in this paper are the opinions of the authors and do
not represent any official position of any of the US federal
agencies that partially supported this work.
This work was supported by DHS grant CIMES 2008-ST-
061ML0002.

5 References

[1] Songs of Humpback Whales, roger S. Payne and
Scott McAvay 13 August 1971, Volume 173, pp 587-597
Signature in Humpback Whales,” OCEANS 2008 - MTS/IEEE Kobe Techno-Ocean , vol.,no., pp.1,8, 8-
11 April 2008


