Acoustics'08 Paris
June 29-July 4, 2008
www.acoustics08-paris.org

Detection and classification of call types in the vocalizations of north-east pacific blue whales

Nicolas Josso[a], Cornel Ioana[a] and Jack McLaughlin[b]

[a]GIPSA-lab, dep. DIS, 961, rue de la Houille Blanche, 38402 St Martin d'Hères, France
[b]University of Washington, 1013 NE 40th St, Seattle, WA 98105-6698, USA
cornel.ioana@gipsa-lab.inpg.fr

Characterization of marine mammal vocalizations is of great help for understanding underwater issues such as underwater communication, sonar detection and localization, marine mammal monitoring, ect. The vocalizations of the North-East Pacific (NEPAC) blue whales are known to be made of at least three different call types: the A call, the B call and the C call. This study aims at the development of a *wholly automatic* process of detection and classification for the two most common call types of the NEPAC population which are the A call and the B call. We created one template for the A call and one for the B call in order to extract features with matchfiltering operations. Features are then analyzed and we show that a simple Gaussian Mixture Model classifier can be used to accurately track and identify the call types in 24 hours long records. The proposed methodology is applied to real data sets recorded by seismic sensors gathered thanks to the Keck Foundation.

1 Introduction

The blue whale is the largest mammal and perhaps the largest animal ever to inhabit the Earth. Because of its size, the sounds it emits are of extraordinarily low frequency – in the tens of Hertz, although the powerful vocalizations can produce harmonics up to hundreds of Hertz. The largest known concentration, consisting of about 2,000 individuals, is the North-East Pacific population. It ranges from Alaska to Costa Rica but is most commonly seen from California in summer.

Characterizing the typical time-frequency content of blue whale call types using automatic means is akin to the task of automatic speaker recognition in which the "average" spectral content of an individual human speaker's speech is mathematically modelled and then subsequently tested with speech from unknown speakers. We adapted APL-UW's existing speaker recognition technology to the problem of detection and classification of call types for the North-East Pacific blue whale vocalizations.

According to previous studies [4][5], the different call types of the NEPAC population are consistent. This study shows it is possible to develop a *wholly automatic process* of detection and classification for the most common call types of the NEPAC population, the A calls and the B calls. Time-frequency tools and analyses were used to create features computed by a simple Gaussian mixture Model classifier to track and identify vocalise features.

The paper is organized as follows. Section 2 presents the real data used in our study as well as the pre-processing methods applied before classification. Section 3 describes the detection and classification methodology. The classification results are presented in Section 4. We close in Section 5 with Conclusions.

2 Data and pre-processing

2.1 Data collection

The data used in this paper have been provided by the Keck Endeavour Seismic Network via Dr. William Wilcock of the School of Oceanography of UW. These data were recorded by underwater sensors measuring speed variation and used for seismic purposes. These data represent 575 hours of recordings with a sampling frequency, *Fs*, of 128.66Hz.

2.2 Pre-processing and sorting of calls

With this great amount of data, an automatic way to proceed to the call detection was necessary. For this purpose, we employed a software package called Ishmael [11] which is used by most marine mammal biologists. Calls were detected automatically using the spectrogram correlation of Ishmael. Spectrogram settings were a 256–point Hanning window shifted by 64 samples.

Most of NEPAC vocalizations are made of A calls followed by B calls but a B call can be followed by another B call (figure 1.a). We wanted to classify A calls and B calls so the classifier was trained with two sets of data: the first one is composed of short files (25 seconds) containing only A calls with high SNR and the second one of short files (25 seconds also) containing only B calls. We decided to sort the development testing and the testing data into two different groups which are A-B calls and B-B calls. This organisation of the data allows us to check the classification and the detection part of this study in a realistic way by working with almost all the different call combinations possible. The repartition of the NEPAC blue whale vocalization can be summarized in this table:

	Training	Dvpt Testing	Testing
B call	900	0	0
A call	396	0	0
AB structure	0	744	771
BB structure	0	157	136

Table 1 Repartition of the NEPAC blue whale vocalization

3 Detection and classification

3.1 Feature extraction

For classifying blue whales calls, we used as features correlations between the spectrogram of the analyzed file and the spectrogram of a template of a vocalization. A template of the spectrogram of an A call and the template of the spectrogram of a B call were made (figure 1). The features are the result of the correlation between the spectrogram of the file and the spectrogram of a template which corresponds to a part of a NEPAC vocalization.

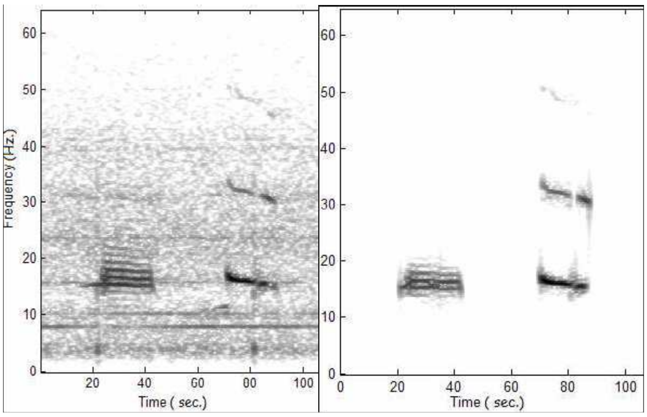

Fig. 1 . Spectrograms of the original vocalization (left side) and of the reconstructed call (right side)

So, there are two features for each analyzed file; one feature is the correlation with the template of the A call and the other feature is the correlation with the template of the B call (figure 2). The lengths of the templates of A and B calls are the same (19 seconds). A shifting window process has been used to enable the tracking of the time of occurrence for a detected call. The correlation is calculated with a window of the length of the templates, and then this window is shifted. It was observed that an overlapping of 80% appeared to be a good compromise between rapidity and accuracy.

When a record is studied, correlations are computed in the time-frequency domain to obtain new features. It could be possible to proceed to a first detection and a first classification using these features which leads to a good detection and a good classification of B calls. Though A calls would not be well detected or classified because the result of a correlation between an A call and the template of the A call is often more sensitive to the noise. That is why the features are then compared to models created by a Gaussian Mixture Model (GMM) classifier in order to enable the detection and the classification. The scheme below summarizes the process used for the detection and the classification:

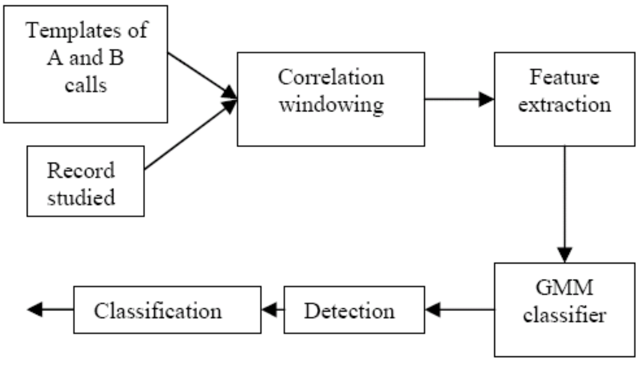

Fig. 2 . General organization of detection-classification methodology

3.2 Gaussian mixture model classifier

Characterising the typical time–frequency content of blue whale population groups using automatic means is akin to the task of automatic speaker recognition in which the "average" spectral content of an individual human speaker's speech is mathematically modelled and then subsequently tested with speech from unknown speakers. That is why we decided to use a GMM classifier in which the GMM is simply a sum of weighted Gaussians which can be described by:

$$p(\vec{x}/\lambda)=\sum_{i=1}^{M} p_i b_i(\vec{x}), \qquad (1)$$

where x is an N-dimensional random vector (the feature vector), b_i are the M component densities, and p_i are the component weights. Each component is an N-dimensional Gaussian of the form

$$b_i(\vec{x})=\frac{1}{(2\pi)^{\frac{N}{2}}(\Sigma_i)^{\frac{1}{2}}}\exp\left(\frac{1}{2}(\vec{x}-\overrightarrow{\mu_i})^T\Sigma_i^{-1}(\vec{x}-\overrightarrow{\mu_i})\right), \quad (2)$$

where Σ_i and μ_i are the mean and covariance of the component. Weights are scaled such that they sum to one making the GMM a proper probability density function. Calculation of the probability of a set of test vectors given a GMM is straightforward. The call type of a set of test vectors is determined by calculating the probability of the test vectors given models for all call types and selecting the call type whose model scores the highest probability.

3.3 Call detection method

Some files analysed may not contain any call and others may contain a great number of calls showing the necessity of a threshold. In order to reduce the false alarm ratio, a threshold is used in the usual way: if one or both of the likelihoods are over the threshold, a vocalization will be detected. Files different from the training set of data and containing only A calls and B calls have been created (101 A calls and 305 B calls) to find the threshold. These files are longer than the files used to train the classifier and the features obtained with these files were used to plot receiver operating characteristic curves (ROC curves) In order to plot the ROC curves, the likelihoods obtained were sorted into two groups according to the method shown in the following scheme.

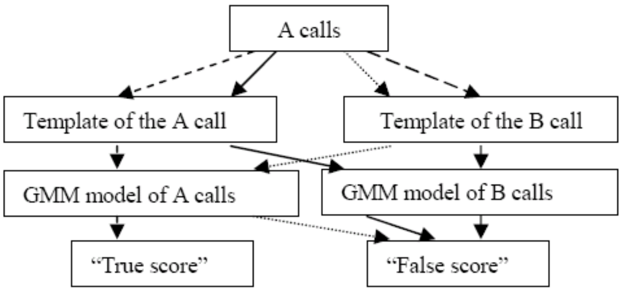

Fig. 3 . Methodology of ROC evaluation

In this scheme, the first step is the correlation between the call studied and the template of a call. The features obtained are compared to the GMM models. Features and likelihoods coming from files containing B calls were sorted the same way. The first threshold used corresponds to the equal error rate on the ROC curves. The equal error rate corresponds to the point where the miss probability and the false alarm probability are equal. In figure 4, the equal error point is 27.5%:

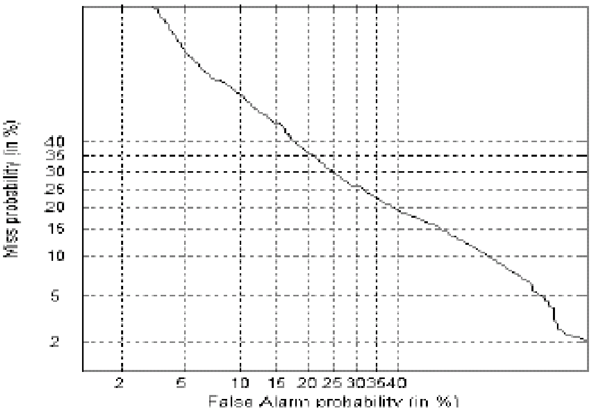

Fig. 4 . ROC curve used to find the threshold corresponding to the equal error rate

This corresponds to a particular value of threshold. New features are computed for every new window and windows are spaced by 2.9835 seconds for an overlapping of 80%. If one or both of the features have a value bigger than the threshold for a certain time, a call is detected and will be classified in the next stage.

3.4 Classification

Once a call detected, it is classified between the two groups considered: A calls or B calls. If the likelihood of a feature scored against the model of A calls is higher than the likelihood of this feature scored against the model of B calls, the call is sorted as an A call. B calls are identified in the same way. A and B calls are often at least 20 seconds long so they are detected in more than only one window. The average length of A and B calls are well known [4], so as soon as a likelihood is higher than the threshold, a maximum is tracked over both time and the likelihoods to be an A call or a B call. This maximum is computed on a number of windows corresponding to the average length of A and B calls.

Windows are numerated from the beginning of the recording and the number of the windows corresponding to the detection of a call is kept. This process was tested with the development data with different thresholds. An optimized threshold close to the threshold calculated with the ROC curves was then kept. The figure 5 illustrates the progression of the curves of likelihood for both of the features as a function of the number of the windows and the spectrogram of the call studied.

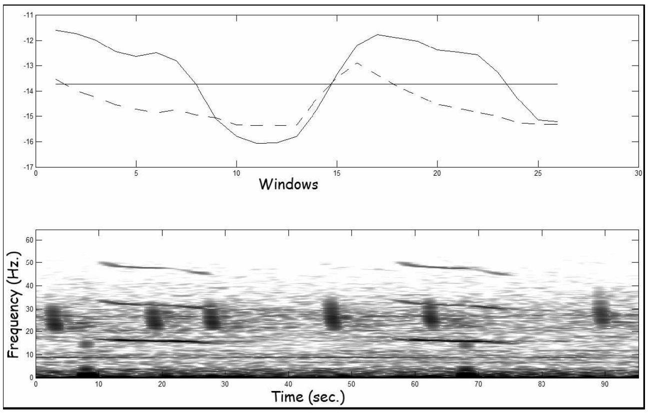

Fig. 5 . Likelihood curves of A and B calls versus spectrogram

4 Results

The automatic detection and classification was run on all the development data which are made of files containing A-B calls and B-B calls. The result of the classification is considered correct only if both the parts of the call are well detected and then well classified. The results can be summarized in this array:

	Number of correct results	Percentage of correct results
A – B calls	692	93%
B – B calls	152	96.8%

Table 2. Results of the classification and the detection for the development data

As the results match well with the development data, we tested it on the testing data without any modification and this led to the following results:

	Number of correct results	Number of correct results
A – B calls	709	92%
B – B calls	129	94.8%

Table 3. Results of the classification and the detection for the testing data

These results show that the classification and the detection work quite well even if a few calls are not detected or classified correctly. This work also enables us to record the time of occurrence of the calls when they are detected on files. We decided to work on long files that have more than one or two calls to check the automatic time detection. A 24 hours long file containing an important number of NEPAC vocalizations was divided into 15 smaller files of 4370 seconds to simplify the interpretation of results. 129 A calls and 184 B calls were detected with an average accuracy of 2.56 seconds for the A calls and 1.15 seconds for the B calls. The standard deviation of the time of detection is 4.26 seconds for the A calls and 6.67 seconds for the B calls.

Another way to test a classification is to calculate the Word Error Rate (which is often called the WER) and can be computed as:

$$WER = \frac{S + D + I}{N}, \qquad (3)$$

where S is the number of substitutions, D is the number of the deletions, I is the number of the insertions and N is the number of words in the reference. The automatic processes of detection and classification have been applied to 6 files of 4370 seconds long which leads to the detection of 326 calls. Among the calls detected, 313 were real calls with 129 A calls and 184 B calls and there were 13 insertions. The number of substitutions were 10 and 2 calls were not detected at all.

$$N = 326;\ S = 10;\ D = 2;\ I = 13.$$

$$WER = 0.076.$$

5 Conclusion

This study shows it is possible to detect and then classify the different types of call that are part of blue whale's vocalizations. This process is *wholly automatic* and could aid the study of blue whale's vocalization. The automatic detection is quite accurate but it can be improved by increasing the overlapping. The identification of A calls and B calls is good and around 93% of the calls of the testing data are well sorted. It should also enable a real time detection and classification while recording underwater sounds if the correct overlapping is chosen.

According to a previous study by the same authors, the same process should be able to detect and then classify the vocalizations of four populations of blue whales living around the world: the NEPAC, South-east Pacific, Atlantic and Antarctic populations. Finally, it indicates that a sea mammal speech recognition system could be possible with blue whales.

Acknowledgments

Without the help of many individuals, particularly with regard to acquiring data, this work would not have been possible. The authors would like to thank Dr. William Wilcock and his student David Englund for providing us with the NE Pacific blue whale recordings from seismic sensors, and the Keck Foundation for their support in gathering these recordings. Dr. Kathleen Stafford was an invaluable source of general knowledge concerning blue whales. A great deal of data for this work came from the NOAA PMEL website. We would like to thank the NOAA/PMEL Acoustics Project Office, the Woods Hole Oceanographic Institution and the Lamont-Doherty Earth Observatory for making this possible. Finally, we would like to thank Marc Stewart and Mike Boyd of the Applied Physics Lab - University of Washington for their help with seaGlider data.

References

[1]http;//Newport.pmel.noaa.gov/whales/bluecal.html

[2]http://www.birds.cornell.edu/SoudsBlueWhale.html

[3]Clark C. W. and Altman N.S. (2006) "Acoustic Detection of Blue Whale (Balaenoptera musculus) and Fin Whale (B. physalus) Sounds During a SURTASS LFA Exercise"

[4] David K. Mellinger, Christopher W. Clark (2003) "Blue Whale (Balaenoptera musculus) sounds from the North Atlantic".

[5]Julie A. Rivers (1997) "Blue whale, *Balaenoptera musculus*, vocalizations from the waters off central California"

[6]Kathleen M. Stafford, Sharon L. Nieukirk, Christopher G. Fox (1999) "Low-frequency whale sounds recorded on hydrophones moored in the eastern tropical Pacific"

[7]Kathleen M. Stafford, Sharon L. Nieukirk, Christopher G. Fox (1999) "An acoustic link between blue whales in the eastern tropical Pacific and the northeast Pacific"

[8]Kathleen M. Stafford and Sue E. Moore (2005) "Atypical calling by a blue whale in the Gulf of Alaska (L)"

[9]Mark A. McDonald, John Clalambokidis, Arthur M. Teranishi and John A. Hildebrand (2001) "The acoustic calls of blue whales off California with gender data"

[10]Sharon L. Nieukirk, Kathleen M. Stafford, David K. Mellinger and Robert P. Dziak, Christopher G. Fox (2004) "Low frequency whale and seismic airgun sounds recorded in the mid-Atlantic Ocean".

[11]Mellinger Dave,

http://cetus.pmel.noaa.gov/cgi-bin/MobySoft.pl