

CLASSIFICATION OF BLUE WHALE USING TIME-FREQUENCY SIGNATURES

Nicolas JOSSO¹, Cornel IOANA¹, Jack MCLAUGHLIN²

¹ Laboratoire de Grenoble Image Parole Signal Automatique (GIPSA-Lab), Grenoble National Polytechnic Institute, Grenoble, France

² Applied Physics Laboratory (APL) University of Washington, Seattle, WA
E-mail : ¹[nicolas.josso, cornel.ioana]@gispa-lab.inpg.fr, ²jackm@apl.washington.edu

Keywords: Signal classification, Acoustic signal processing, Time-frequency analysis.

Abstract: *Characterization of marine mammal vocalizations is of great help for understanding underwater issues or for species monitoring. The vocalizations of the North-East Pacific (NEPAC) blue whales are known to be made of at least three different call types. 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. Using time-frequency tools and analyses to create features, we show that a simple Gaussian Mixture Model classifier can be used to accurately track and identify the call types in long vocalizations.*

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 duration and frequency content of these vocalisations have been found to vary across population groups [1] [2].

Because of their loudness and their very low frequency, blue whale’s calls can be heard at hundreds of miles from the emitter. Thanks to evolution, underwater mammal’s calls such as blue whale’s vocalizations are especially designed for underwater propagation and echolocation. That is why it could be interesting for submarines to be able to emit blue whale’s calls for sonar localisation and communication purposes. This approach also has the second advantage of being undetectable by other people listening to the sound underwater which is primary for military purposes. Characterizing 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. We adapted APL-UW’s existing speaker recognition technology to the problem of classification of blue whale vocalizations by population group. 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 PROCESSING

2.1 Data collection

An initial set of data were provided by seaGliders, autonomous underwater vehicles

(AUVs) built by APL-UW. These vehicles can travel underwater for many months along a preprogrammed path recording ocean parameters (*e.g.* temperature, salinity) as they go. The seaGliders can also make acoustic recordings at a sampling rate of 3 kHz. Blue whales calls of the North-eastern Pacific group were represented in a collection performed in Monterey Bay in 2006, and in data recorded in Kauai. We also worked with data given to us by 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. A final set of data were obtained from the NOAA T-Phase collection which are organized into 1-Day files. The file names reflect the Hydrophone General Location, Year and Julian Date of the data.

In total, at least 15 000 hours of recordings which were likely to contain blue whales calls from four different populations (North-eastern

Pacific, South-eastern, Antarctic and Atlantic) were gathered.

2.2 Call detection

With this great amount of data, an automatic way to proceed to the call detection was necessary. Kathleen M. Stafford and Marc Stewart recommended us to work with a software package called Ishmael [11] which is used by most marine mammal biologists. Calls were detected automatically for each of the four different populations using the spectrogram correlation of Ishmael with the following settings: a 256-point Hanning window shifted by 64 samples.

-North-eastern population:

The vocalizations of the North-eastern population can consist of the A-C-B call sequence [Fig. 1] but the A-B type was most commonly recorded.

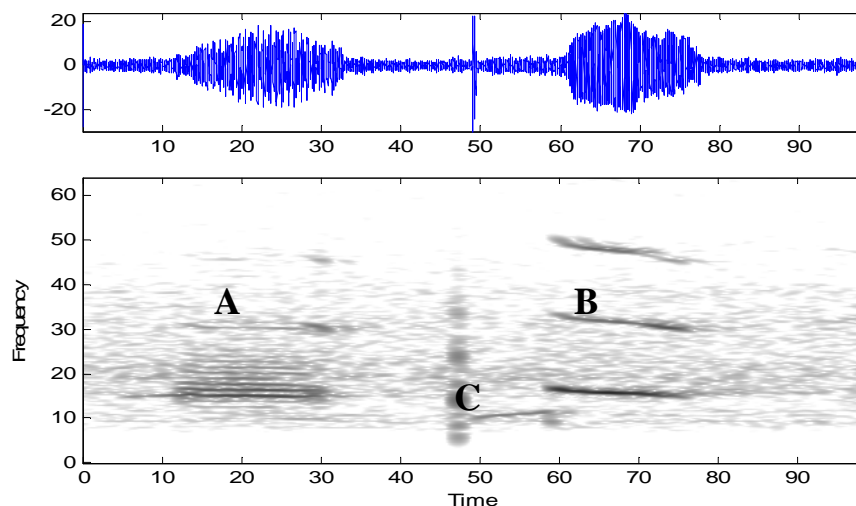


Fig. 1: A-C-B type North-eastern Pacific call detected in data recorded by the Keck Foundation. The first figure represents the time series representation and the abscissa is graduated in seconds. The second one is the spectrogram of the call. The peak that occurs just before the C call at around 50 seconds is not part of the call. It must come from the hard drive recording data.

The A call is the first part of the call and its pulsive nature may be seen in the time series representation. Then the C call is the FM up sweep between 10 and 12 Hz and the sequence is ended by the B call which is the FM down sweep. The first and the second harmonics of the B call are shown in the spectrogram. Each call is detected by a correlation between the spectrogram and linear FM down sweep. To begin, a linear FM down sweep between 16 and 15 Hz of 8 seconds was

used for automatic detection of the B call using a spectrogram correlation. However, each call must be detected only once, and it happened that the A call was detected so that the A-C-B call was detected twice. This is why the linear FM down sweep used for the detection sweeps down from 49 to 48 Hz in 8 seconds. So 3310 North-eastern Pacific calls were detected in this collection of data with a low signal to noise ratio (SNR).

-South-eastern population:

The calls of the North-eastern population are complex and made of 4 main parts A-B-C-D and can be seen in the spectrogram [Fig. 3]. The A call is a long moan with slight frequency modulation and has often upper sidebands at around 3 Hz intervals. The B part of the call is a FM pulsive down sweep with sidebands at around 7 Hz intervals. The third part of the call is called the C call and is short with no significant modulation, is very pulsive and has sidebands at around 6 Hz intervals. The last part, the D call, is a very pulsive FM moan with sidebands at around 6 Hz too. The

D call consists of two parts, the first one is a unmodulated segment followed by a modulated down segment.

Each call is detected by a correlation between the spectrogram and linear FM down sweep of 8 seconds between 26.5 and 25.5 Hz in order to detect the D part of each call. It was not as easy to detect a lot of call of this population as for the North-eastern population. Around 8400 hours have been processed using this automatic way of detecting calls and 1855 calls of the South-eastern population have been detected in NOAA data.

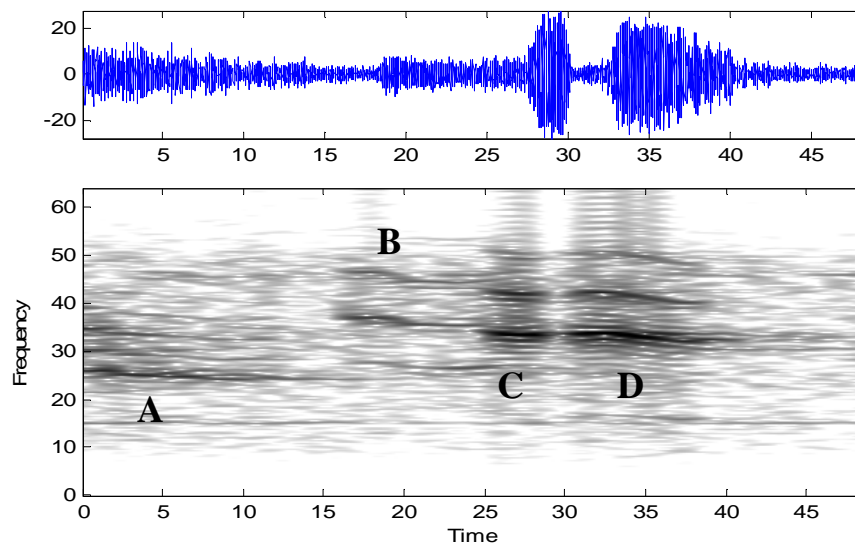


Fig. 3: A typical South-eastern call detected on NOAA data which is made of four parts. The first figure represents the time series representation and the abscissa is graduated in seconds. The second one is the spectrogram of the call.

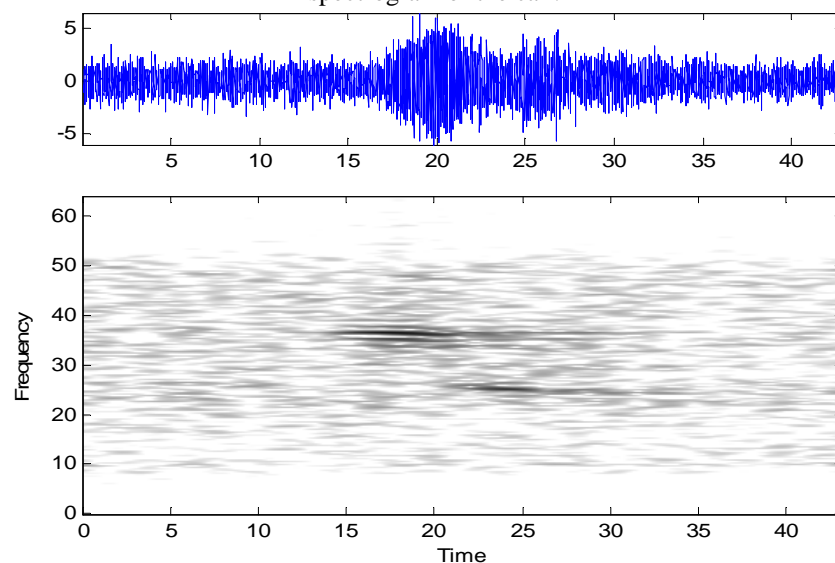


Fig. 4: A typical Antarctic call detected on NOAA data. Once again, the first figure represents the time series representation and the abscissa is graduated in seconds. The second one is the spectrogram of the call. The SNR of the vocalization is poor and the call is faint in the time series domain. Though, it can be easily seen in the time frequency representation

-Antarctic population:

The calls of the Antarctic population are made of 28Hz moans which are followed by constant frequencies at 20 Hz and 28 Hz [figure 4]. An Antarctic call is made of a 28Hz moan which is followed by constant frequencies at 20 Hz and 28 Hz.

This time, calls are detected by a correlation between the spectrogram and linear FM down sweep of 8 seconds between 28.5 and 28 Hz in order to detect the 28Hz moans. Around 4000 hours have been processed using

this automatic way of detecting calls. Finally 1826 calls of the Antarctic population have been detected in NOAA data.

-Atlantic population:

The calls of the Antarctic population are made of two sequences. The first sequence is the A call which is an unmodulated pure frequency at around 18.5 Hz, the second sequence is called the B call and is made of a modulated frequency between 18.5 and 15 Hz [figure 5].

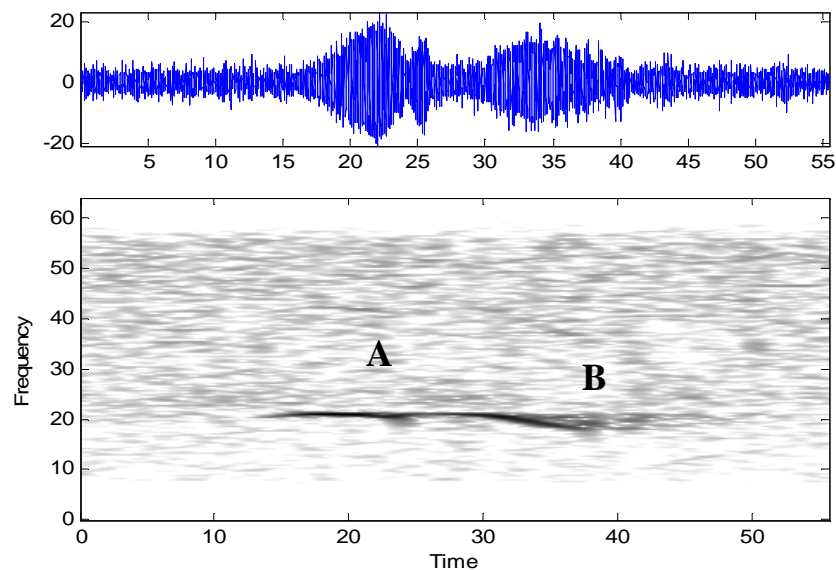


Fig. 5: A typical Atlantic call detected on NOAA data. The first figure represents the time series representation and the abscissa is graduated in seconds. The second one is the spectrogram of the vocalization. The vocalizations of the Atlantic population are the simplest we were given to see among blue whales' calls.

A pure frequency at 18.5 Hz and 9 seconds long was used for spectrograms correlations in order to detect the A calls. Around 3500 hours have been processed using this automatic way of detecting calls and 3524 calls have been detected for the Atlantic population in NOAA data.

2.3 Processing and sorting of calls

When using the automatic detection with the spectrogram correlation on Ishmael, it is possible to write a text file where the starting and the ending time of the detected call and the name of the file are written. Then, we create a file per call detected from our detection text

file. For each population, calls have been sorted into 3 different classes which are:

- Training
- Development testing
- Testing

This leads to a certain repartition of the blue whale calls. For the population where there were less calls, the training class is more important than the development testing and testing ones. In fact, the training class is always bigger than other classes in order to train correctly the classifier. Though, we paid attention not to over train the classifier because by having training set too large.

	NE Pacific	SE Pacific	Antarctic	Atlantic
Training	1300	933	899	1324
Dvpt testing	1005	459	476	1100
Testing	1005	463	453	1100

Table 1: Number of calls for each population and each class

3. DETECTION AND CLASSIFICATION

3.1 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. Over the past decade, systems for performing speaker authentication have been gradually refined. Though early attempts at speaker ID employed a diversity of classifiers and feature sets, most state-of-the-art systems today utilize some form of mel frequency cepstral coefficient (MFCC) vectors (computed every 10 milliseconds or so) as features and a Gaussian mixture model-based classifier (GMM). The GMM is simply a sum of weighted Gaussians which can be described by

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

where \vec{x} is an N -dimensional random vector (the MFCC 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

$$b_i(\vec{x}) = \frac{1}{(2\pi)^{N/2} |\sum_i|^{1/2}} \exp\left\{-\frac{1}{2}(\vec{x} - \vec{\mu}_i)^T \sum_i^{-1} (\vec{x} - \vec{\mu}_i)\right\}$$

where $\vec{\mu}_i$ and \sum_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 speaker of a test utterance is normally determined by calculating the probability of the test utterance given models for many putative speakers, and selecting the speaker whose model scores the highest probability.

The popularity of GMMs for speaker authentication is largely due to the small number of parameters that must be estimated from training data to determine the components; the ability of GMMs to accurately describe training data in a high dimensional feature space through the inclusion of additional components; and the existence of the Expectation Maximization (EM) algorithm, an iterative process enabling rapid calculation of GMM parameters from training data.

3.2 Classification

The training class of each population will be used only to train the classifier, the development testing one will be used to test the features and the testing will only be used at the very end of the classification to make sure that the classification is working correctly. The most significant modification was to identify features to discriminate among the groups in this extremely low-frequency domain.

According to studies [4] [5] blue whale calls are not supposed to vary a lot in a same population. So the key was to find features that are able to make a difference between calls of different populations while calls from a same population must appear similar. However, calls are in the same range of frequencies so simplistic energy detection in frequency bands is not supposed to work for the classification. Calls are so similar in a same population that it

appears possible to use a correlation between each call and a template of the population as a feature. Moreover, calls of the different populations also appear to be far enough one from the other to classify calls using the correlation. As calls are easier to see and to interpret in the time frequency, we decided to use the correlation between the spectrogram of each call and the spectrogram of a template for each population as features.

For each population, a template was created in the time frequency domain. This template has

to match as much as possible with all the spectrograms of the calls of all the population. That is why the template of a population is not always made of all the call but only of the most important part of the call. For example, in North-east Pacific population, the A and the C calls always disappeared quickly when the signal to noise ratio (SNR) increases [figure 5]. That is why the template of the North-east Pacific population does not comport either the A and the C call.

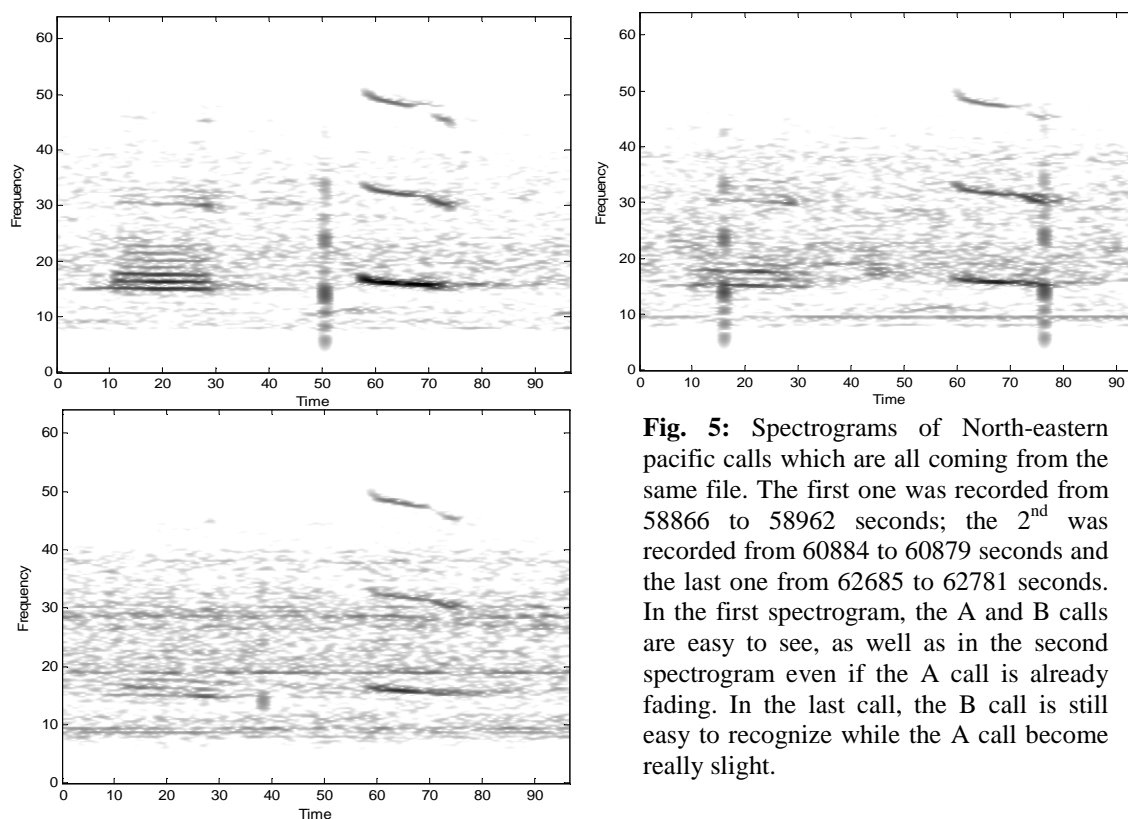


Fig. 5: Spectrograms of North-eastern Pacific calls which are all coming from the same file. The first one was recorded from 58866 to 58962 seconds; the 2nd was recorded from 60884 to 60879 seconds and the last one from 62685 to 62781 seconds. In the first spectrogram, the A and B calls are easy to see, as well as in the second spectrogram even if the A call is already fading. In the last call, the B call is still easy to recognize while the A call become really slight.

For each population a template was created following the maximum of energy after filtering in the time frequency domain. Then, templates are created from pure frequencies and chirps without any noise. In the [figure 6], A°) is the spectrogram which led to the template of the North-eastern Pacific template. B°) corresponds to the spectrogram reconstructed from A°) spectrogram. It is not the real template because the template of this population will not have an A call as previously explained. C°) is the template of the Antarctic population and D°) correspond to the template of South-eastern population.

-The features:

The features which were used are all based on the correlation between the spectrogram of a template of each population and the spectrogram of each call. Spectrogram settings were a 256-point Hanning window shifted by 8 samples prior to Fourier transforming. However, the bi-dimensional correlation between two spectrograms takes a lot of time and that is why only 99 mono-dimensional correlations have been calculated. The sampling rates of the detected calls are not the same for all data so in order to have meaningful features; the correlation was calculated between 0 Hz and the smallest

sampling rate divided by two. For each call, a correlation has been computed between the absolute value of its spectrogram and the absolute value of each template's spectrogram for 99 frequencies between 0 and 49.5 Hz. Only the maximum values of the correlation for each frequency were kept, so this lead to a matrix of 99 by 4 values for each detected call. The whole matrix was kept in order to be able to create a lot of different features if necessary. As a start, only the maximum of the columns of each matrix has been kept as a feature. So

the classifier was trained giving him four features (which corresponds to the maximum of the maxima of the 99 correlations with each of the four templates) for all the training sets and using a text file for the parameters. The important parameter is the number of Gaussians which will be used to model the different populations. This leads to the creation of a model file for each population, and the next step was to score the development testing data with the model files.

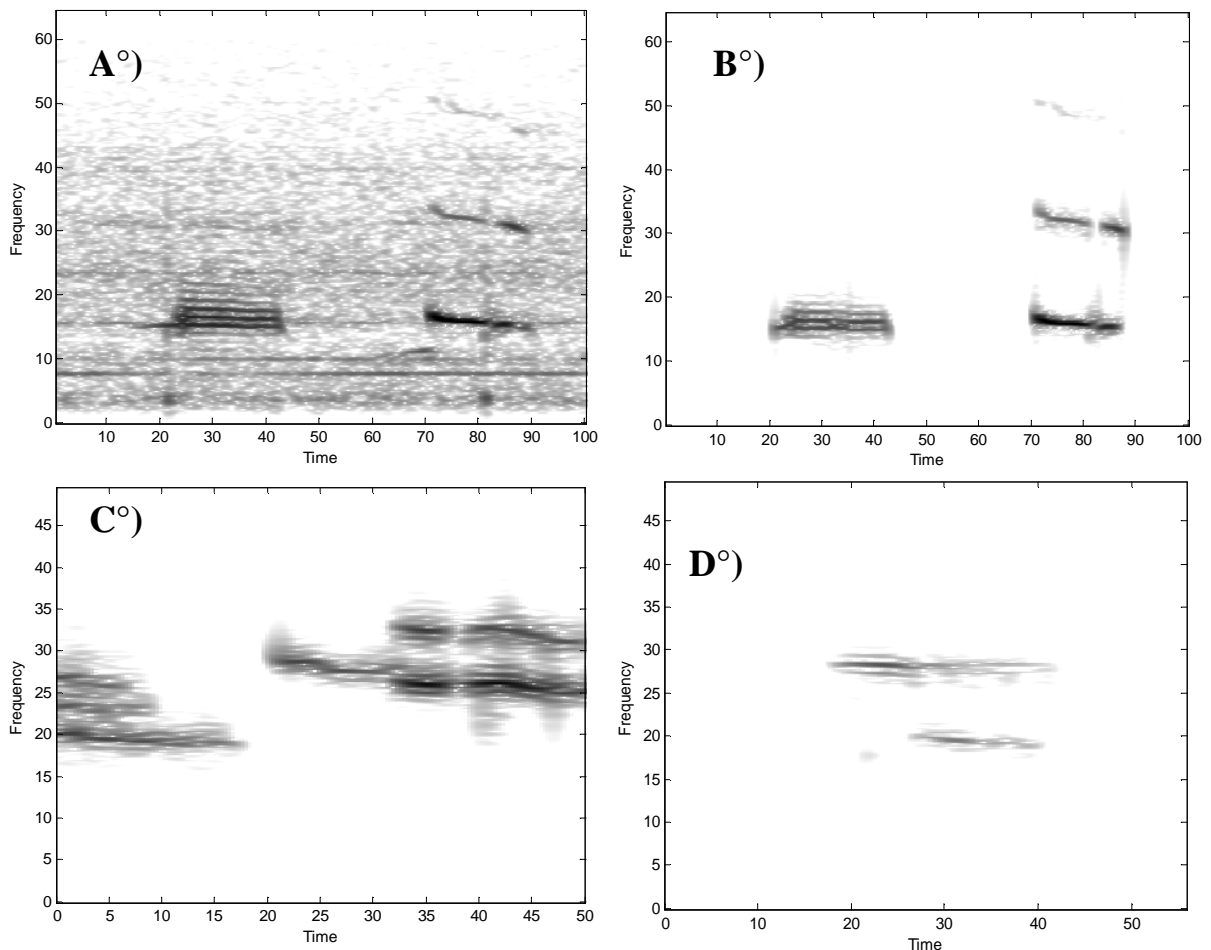


Fig. 6: Spectrograms of the templates the different populations' call.

4. RESULTS

Results are shown using confusion matrixes so lines correspond to all the calls of a given population and the columns correspond

to the population where calls have been classified. The figures represent the number of calls of one population (the lines) which were classified in one given population (the columns).

	NE Pacific	SE Pacific	Antarctic	Atlantic
Training	1300	933	899	1324
Dvpt testing	1005	459	476	1100
Testing	1005	463	453	1100

Table 2: Number of calls for each population and each class

The features of the development testing data were scored with the model of the four populations, scoring the features of each detected call independently. For each call, four scores were given by the classifier; calls were classified in the population correspondent to

the model with which they had the highest score. Results depend on the number of Gaussians used to model the populations so an empiric method was necessary to find the best number of Gaussians.

Number of Gaussians to estimate the model: 27 Dimensionality: 4					
	NE Pacific	SE Pacific	Antarctic	Atlantic	% of call sorted in the right population
NE Pacific	992	6	2	5	99%
SE Pacific	1	433	24	1	95%
Antarctic	2	10	462	2	97%
Atlantic	15	0	1	1084	99%
Number of Gaussians to estimate the model: 25 Dimensionality: 4					
	NE Pacific	SE Pacific	Antarctic	Atlantic	% of call sorted in the right population
NE Pacific	995	3	2	5	99%
SE Pacific	0	434	25	0	95%
Antarctic	2	14	458	2	96%
Atlantic	15	2	0	1083	99%
Number of Gaussians to estimate the model: 26 Dimensionality: 4					
	NE Pacific	SE Pacific	Antarctic	Atlantic	% of call sorted in the right population
NE Pacific	993	6	2	4	99%
SE Pacific	1	433	24	1	94%
Antarctic	2	12	460	2	97%
Atlantic	16	0	0	1084	99%

Table 3: Confusion matrixes obtained for different parameters with the development testing data

When the results were good enough with the development data, the testing group was classified to make sure the classification is working with another set of data. The results

obtained with the test group are almost the same as the ones obtained with the development testing group which confirms that the classification operation is correct.

	NE Pacific	SE Pacific	Antarctic	Atlantic	% of call sorted in the right population
NE Pacific	975	3	24	3	97%
SE Pacific	0	463	0	0	100%
Antarctic	1	15	434	2	96%
Atlantic	1	0	0	1099	99.9%

Table 4: Confusion matrix obtained with the testing data, a number of Gaussians to estimate the model of 24 and a dimensionality of 4.

The results are really good and more than 98% of the vocalizations studied are classified in the right population. This shows that the features extracted are representative of each class and allows differentiating calls of four different populations. 15 000 hours of recordings have been studied so we suppose

that different blue whales have been emitting vocalizations for each population so the classifier is not only able to recognise one whale per population. With this great amount of data we can consider these results are general and the classification should work the same way with any other set of data.

It should be considered that only quite high SNR vocalizations were kept after the detection process because of a high threshold. This way to proceed allowed us to have a very low false alarm rate. That is why it could be interesting to test the classification process with vocalizations having a poor SNR which should lead to more confusion between the different populations. However, it is important to know that the calls studied were not all with a really high SNR as it can be seen in the [figure 5].

5. CONCLUSION

In this work, four different populations of the same specie (blue whale) have been classified using only there vocalizations. This process is wholly automatic from the detection and the creation of a file for each call to its classification. All the populations of blue whales have not been classified because of the lake of data for the Indian Ocean and the Madagascar pygmy populations; however, the same technique should be able to classify those populations the same way. This will prove that all the populations of blue whales can be automatically classified only using their vocalizations.

A study by the same authors will show that the different part of the vocalizations of one pod can be accurately tracked and identified for future purposes. The creation of synthetic blue whale vocalizations could be used by submarines for communication or echolocation. This work shows it is possible with the creation of the template of the call of each population. Finally, it is important to know which population may emit vocalizations in one area and it can be achieve by this wholly automatic process. It is also possible to know if a vocalization that was emitted in one place could have been emitted by one of the population leaving in the area or must have been synthesised.

6. ACKNOWLEDGEMENT

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.

7. REFERENCES

- [1] Marine Mammal Sounds - Blue whale, http://www.pmel.noaa.gov/vents/acoustics/whales/sounds/sounds_whales_blue.html
- [2] Blue Whale — Bioacoustics Research Program www.birds.cornell.edu/brp/listen-to-project-sounds/blue-whale
- [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] Mellinger, D.K., and C.W. Clark. 2003. Blue whale (*Balaenoptera musculus*) sounds from the North Atlantic. *J. Acoust. Soc. Am.* 114(2):1108–1119.
- [5] Julie A. Rivers (1997) "Blue whale, *Balaenoptera musculus*, vocalizations from the waters off central California"
- [6] Kathleen M. Stafford, Sharon L. Nieu Kirk, Christopher G. Fox (1999) "Low-frequency whale sounds recorded on hydrophones moored in the eastern tropical Pacific"
- [7] Kathleen M. Stafford, Sharon L. Nieu Kirk, 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. Nieu Kirk, 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, D.K., and C.W. Clark. 2006. MobySound: A reference archive for studying automatic recognition of marine mammal sounds. *Appl. Acoust.* 67(11-12):1226-1242.