NAIVE BAYES CLASSIFIER FOR AUTOMATIC ANALYSIS OF BELUGA WHALE SONGS

Guillermo Lara, Ram´on Miralles, Mariam Torres
Instituto de Telecomunicaciones y Aplicaciones Multimedia, Universitat Polit`ecnica de Val`encia
Edificio 8G, 46022, Valencia, Spain
phone: + (34) 963879737, email: rmiralle@dcom.upv.es

Abstract - Very little is known about the way cetaceans and particularly beluga whales communicate. However as scientists and biologists investigate it is observed the extraordinary adaptation of the cetacean communication system to the underwater medium and the surprising communication skills. The ultrasonic sound emitted by beluga whales can be identified in what some scientists call vocalizations and related to animal behavior. The signal processing group GTS (iTEAM) of the Universitat Polit`ecnica de Val`encia alongside with the Oceanogr`afic have developed an automatic system for continuously monitoring beluga whale sounds. This system is intended to establish new behaviour patterns and help biologists to obtain a better understanding of beluga whales. The present work is devoted to the comparison of the different Naive Bayes classifiers into the automatic monitoring system.

Keywords: Statistical signal processing, Signal detection and classification, Bioacustics.

1. INTRODUCTION

Previous studies have shown that the analysis of the beluga vocalizations patterns is a good tool to evaluate their communication and welfare state [1]. The correlation obtained between vocalizations and behaviors show up that their vocalizations are strongly influenced by external stimulus. The first impression when beluga vocalizations are heard for the first time is that there are an endless variety of sounds. Nevertheless, time-frequency analysis shows that there are a limited number of patterns and they are repeated. The researchers from the Signal Processing Group (iTEAM) trained an automatic algorithm, after having worked with large amount of data picked up from long time periods of direct visualization. The classification system is based on statistical analysis from the acoustic signal received by the hydrophone [2]. The developed system was appropriated for designing experiments that help to understand a bit more the beluga behavior.

This paper is focused on analyzing the best vocalization classifier for the system. In order to do that, a comparative among Naive Bayes distributions (Gaussian, Kernel, Multinomial and Multivariate multinomial) is presented.

2. BELUGA WHALE SOUNDS AND THEIR RELATION TO ANIMAL BEHAVIOR

Studies of the vocalizations emitted by the Oceanogr`afic beluga whales (Kairo and Yulka) have allowed scientists to obtain a large collection of sounds. These studies started in 2003 when both whales arrived at the installations of the Ciudad de las Artes y las Ciencias from the Mar del Plata, in Argentina. A comparison between vocalizations rate and animal welfare was done. This comparative study demonstrated that during stress periods (such as those produced by air transport to new facilities or beluga pregnancy) acoustic activity decreased significantly [1]. Additionally, a set of classification categories for beluga sounds was created. All this work was done manually listening one by one a large number of records and analyzing with the aid of the spectrogram how the energy was distributed in time and frequency. This process is tedious and time consuming and can not be maintained 24 hours a day. Instead of continuous inspection, researchers analyze only a few minutes a day of the recordings. Recently, the Instituto de Telecomunicacion y Aplicaciones Multimedia (iTEAM) has begun to collaborate with the Oceanogr`afic researchers to employ automatic classifiers which allow a continuous examination of the emitted sounds.

We pretend to use a simple classification scheme that will help to establish relationships among sounds and behaviors. The quantity of vocalizations produced by belugas, or cetacean in general, reaches an extensive number of data with complex and rapidly repeated clicks [3]. In order to create a simple set of categories all the vocalizations will be classified in three groups: tonal, pulsed and Jawclap sounds. In this study the echolocation clicks are not taken into account and they have been manually removed from the recordings. Echolocation is broadly studied in many species and seems to be used by the animals as a biological sonar instead of having communicative purposes.

Tonal sounds are characterized by narrow bandwidth squeals and whistles, giving out a very concrete component clearly detected in frequency. Differences such as the number of frequencies during the same period of time can be found. When belugas produce more than one frequency simultaneity is called multitonal vocalization in contrast when an isolated frequency component is detected, called as a simple tonal one. In most cases tonal sounds seem to have communicative nature.

Pulsed sounds look like short broadband clicks (see figure 1). These kind of vocalization can be related to communicative or aggressive behaviour. A different kind of pulsed sound is the Jawclap. The mechanisms that beluga whales use to generate this sound is completely different to the mechanism employed to generate other pulsed sounds. Due to the aggressive meaning associated to this sounds a specific category has been created. Zookeepers and biologists from Oceanogr`afic were fundamental in the development of this classification scheme.

In addition to single vocalization it is frequently found combinations of two or more individual vocalizations. These combinations are usually associated with special situations and specific stimulus which make difficult to distinguish one category or another. This kind of signals, that will be re ferred in this work as mixed signals, are composed of various vocalizations concatenated or overlapped in time. The classification algorithm must consider this extraordinary capability of belugas. The mixed signal can be composed of a tonal-pulsed mix or a pulsed-jawclap mix. An example of this flexibility to produce sounds is illustrated in the figure 1 where the combination of a pulsed and a jawclap are really close in time.

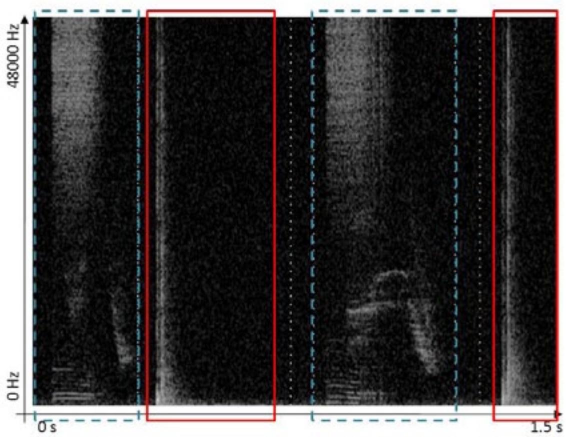

Fig. 1: Spectrogram of beluga whale vocalizations showing two alternated pulsed vocalizations (blue square) and Jawclaps (red square)

3. BRIEF DESCRIPTION OF THE AUTOMATIC CLASSIFIER OF BELUGA WHALE SOUNDS

The recorded vocalizations were provided by the Oceanografic biologists. Sounds were acquired in the Oceanografic facilities using a single hydrophone, an acquisition card and a computer. For this preliminary study an Excel table with the relative time when each vocalization started. Their duration and the manual classification was also supplied.

The classification has been done using MATLAB and the MATLAB Statistic toolbox. This work has been focused on comparing for this specific application a set of Naive Bayes classifiers and choosing the most appropriated. A Naive Bayes classifier is a probabilistic classifier based on Bayesian statistics with strong independence assumptions. It classifies data in two steps; firstly, using the training samples estimates the parameters of a probability distribution and secondly, it predicts the probability of that sample belonging to each class. The class-conditional independence assumption simplifies the training step. It allows a better estimate of the Naive Bayes parameters required for accurate classification, and uses less training data than other classifiers.

All the steps and decisions involved in sampling data affect the pattern, so the choice of the distinguishing features is a critical step to design [4]. In order to improve the accuracy of classification, it is essential to choose the features that can capture the temporal and spectral characteristics of signals. After several empirically combinations, table 1 shows the features chosen: features 1-9 are frequency parameters related to resonances, bandwidths and power spectral amplitudes. Features 10-13 give statistical information related to higher order moments of the vocalization [5]. Feature 18 is inspired in the human voice parameters obtained in LPC (vocoder) models [6]. This parameter gives information of the residual prediction error. Finally features 14-17 give higher order statistical information of the residual prediction error (feature 18).

NUMBERTS	FEATURES
1	Fundamental frequency f_0
2	Q-Factor of f_0 = Δf_0= f_0
3	Power Spectral Density at frequency f_0 S_x(f_0)
4	Fundamental frequency f_1
5	Q-Factor of f_1 = Δf_1= f_1
6	Power Spectral Density of frequency S_x(f_1)
7	Fundamental frequency f_2
8	Q-Factor of f_2 = Δf_0= f_2
9	Power Spectral Density of frequency S_x(f_2)
10	"Skewness" of the vocalization (as described in [5])
11	"Kurtosis" of the vocalization (as described in [5])
12	Autocovariance test of the vocalization (as described in [5])
13	Temporal reversibility of the vocalization (as described in [5])
14	"Skewness" of the sonority signal
15	"Kurtosis" of the sonority signal
16	Autocovariance test of the sonority signal
17	Temporal reversibility of the sonority signal
18	Sonority signal (as described in [6])

Table 1: Brief description of the features vector employed in the automatic classifier.

4. COMPARATION AMONG CLASSIFIERS DISTRIBUTIONS

In order to optimize the performance of the classifier, all the characteristics were tested by a training set. The objective was reducing the eighteen features shown in table 1. The algorithm of sequential selection was created for selecting and ordering the most representative characteristics for each distribution [7]. The results are shown in table 2. The steps were followed by:

- An objective criterion to minimize all the possible characteristics of the subsets.

- A sequential searching algorithm which adds or eliminates the characteristics subsets when the criterion is evaluated. This sequential searching allows testing feature to feature, using the called Sequential Forward Selection (SFS) [7].

The classification set was composed of 313 vocalizations where the most representative 50 were chosen for the training set. During the training phase a category label or cost for matching the pattern was provided. It was seek to reduce the sum of the costs for the pattern (tonal, pulsed and jawclap sounds). Figure 2 evaluates the error when classifying Gaussian, Kernel, Multinomial and Multivariate multinomial density functions as the number of features is increased. In addition to these classifiers (based on Naive Bayes), two discriminant analysis classifiers were also compared because of their covariance matrices similarities with Naive Bayes (diaglinear an diagquadratic). A brief outline of the compared distributions are:

- The 'normal' distribution is appropriate for features that have normal distributions in each class. For each feature you model with a normal distribution, the Naive Bayes classifier estimates a separate normal distribution for each class by computing the mean and standard deviation of the training data in that class.

- The 'kernel' distribution is appropriate for features that have a continuous distribution. It does not require a strong assumption such as a normal distribution and you can use it in cases where the distribution of a feature may be skewed or have multiple peaks or modes. It requires more computing time and more memory than the normal distribution. For each feature you model with a kernel distribution, the Naive Bayes classifier computes a separate kernel density estimate for each class based on the training data for that class.

- The multinomial distribution is appropriate when all features represent counts of a set of words or tokens. The classifier counts the set of relative token probabilities separately for each class. The classifier defines the multinomial distribution for each row by the vector of probabilities for the corresponding class.

- The multivariate multinomial distribution is appropriate for categorical features. For each feature you model with a multivariate multinomial distribution, the Naive Bayes classifier computes a separate set of probabilities for the set of feature levels for each class.

For each one, the optimal ordination to minimize the training error was solved (see table 2). The best classifier during the training set was the one based on Naive Bayes with Multivariate multinomial configuration. It achieved less than 1% error with just only for features (1, 4, 3 and 6). The "Kernel" distribution was the second best distribution which uses 11 features to get less than 5% error. The others classifiers have similar behavior with up to 20% error, being the Multinomial distribution the worst classifier during the training set.

It is important to emphasize that the errors obtained in the training test in table 2 will be lower than the errors that will be later obtained in the test set (next section).

5. RESULTS

In order to check the classifier efficiency, results were compared on a different vocalization set processed and manually classified by the Oceanogr`afic biolo-

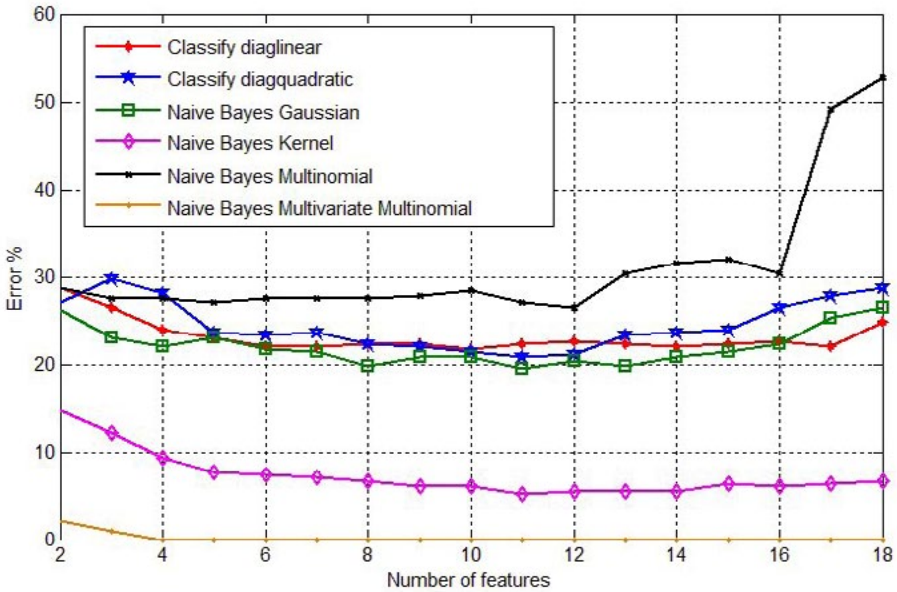

Fig. 2: Comparative among the different classifiers distributions and the number of features used for their evaluation.

gists. This set was tested for all the classifiers and the results are shown in table 3. The table presents the classification rate for all Bayes classifiers for the optimal number of features and for the whole set of eighteen features. It also shows the percentage number of vocalizations that are not classified (missing rate). When the order and the number of features is optimum, the classification rate will be the best. Table 3 shows also that if mixed signals are not considered classification percentages increase for all classifiers excepting the Naive Bayes Multivariate Multinomial. Naive Bayes Multivariate Multinomial classifier reaches a 98.7% of classification rate, but it has a 30% missing rate (vocalizations that was not able to assign to any category). It shows up that if it does not have a clear decision about each vocalization, it will not classify the signal. In spite of when the classifier does it, the algorithm classifies successfully. In addition, this classifier has the distinction to obtain the same classification rate when it uses all the eighteen features or when it uses the optimum order number of features. This is because the features added do not cause any confusion, unlike the other classifiers do. The Naive Bayes kernel classifier using the optimum order and number of features without mixed signals obtains the best results, specifically a 89.2% classification rate and just a 0.2% missed signals. On the other hand, the classifier Naive Bayes Multinomial has the worst classification rate (66.6%), but the Gaussian obtains a good one (81.8%). Either of them do not have missed signals. The percentages for the diaglinear and diagquatratic classifiers (not shown in table 3) are quite similar to those given for the Naive Bayes Gaussian distribution.

Clasificator	Sorted Significative Features	Opt. #	Error
Diaglinear	6,15,2,10,5,17,12,7,9,14,11,1,4,16,18,3,8,13	10	21.43
Diagquadratic	9,15,3,10,2,18,11,7,1,17,14,16,4,13,8,5,16,12	11	20.77
N.B. Gaussian	1,15,6,11,8,17,9,14,18,10,2,7,3,4,13,5,16,12	11	19.49
N.B. Kernel	12,16,6,2,10,13,7,15,1,4,18,8,17,3,11,14,9,5	11	5.11
N.B. Multinomial	9,15,6,3,10,12,16,18,13,1,14,17,4,7,11,2,8,5	12	26.52
N.B. Multivariate Multinomial	1,4,3,6,7,8,5,9,10,11,12,13,14,2,15,16,17,18	3	0.96

Table 2: Optimal number of features and most significative features sorted from more to less significance (N.B.=Naive Bayes).

6. CONCLUSIONS AND FUTURE WORK

The study presented here showed that the best classifier for the system was the Naive Bayes following a kernel distribution with the optimum number of features. The percentages achieved for this classifier for the whole set of vocalizations (including consecutive and partially overlapped vocalizations) was 84% of detection probability with just 1.7% of vocalizations that the algorithm was not able to classify. This study has shown that it is possible to create a real time classifier to analyze beluga sounds the 24 hours a day.

The percentages achieved are quiet good, but they could be improved if a more precise classifier that could take into account the mixed vocalizations was employed. It has to be known that the number of mixed vocalizations is not negligible. In fact, in the vocalizations files we have analyzed approximately the 20-25% of vocalizations were marked by the biologists as mixed vocalizations. In order to get this goal the authors propose as future lines:

- Pyramidal analysis in time and fusion of the different classifiers to decide whether it is a single vocalization or a mixed one.
- Combining a low resolution classifier based on spectrogram correlation with the proposed Naive Bayes classifier.

Increasing the number of categories will be done as a more precise beluga behavior knowledge is achieved. Open sea tests of the proposed algorithm for different species (dolphins) and the possibility of integrating the proposed algorithm in underwater buoys is also planned.

7. ACKNOWLEDGEMENT

The authors thanks very much to Jose Antonio Esteban from Oceanogr`afic who prepared all experimental analysis for deployment. This work was supported by the C´atedra Telef´onica - UPV, the national R + D program under Grant TEC2008-02975 (Spain), FEDER programme and Generalitat Valenciana PROMETEO 2010/040.

REFERENCES

[1] M. Castellote and F. Fossa,"Measuring acoustic activity as a method to evaluate welfare in captive beluga whales (delphinapterus leucas)",Aquatic Mammals, vol. 32, pp. 325-333 (2006).

[2] S. Kay, Fundamentals of statistical Signal Processing: Estimation Theory, 1 edition.Prentice Hall (1998) [3] Roger S. Payane and Scott McVay, "Songs of Humpback Whales" Science, vol. 173, pp. 585-597, Aug. 1971.

[4] Christopher M. Bishop, Pattern Recognition and Machine Learning. Springer; 1st ed. 2006.

[5] R. Miralles and L. Vergara and A. Salazar and J. Igual, "Blind Detection of Nonlinearities in Multiple-echo Ultrasonic Signals" IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, v.55,3, pp. 1-11(2008).

[6] E. Verteletskaya and K. Sakhnov and B. Simak, "Pitch Detection Algorithms and Voiced/Unvoiced Classification for Noisy Speech", Systems, Signals and Image Processing, 2009. IWSSIP 2009. 16th International Conference on, pp. 1–5 (2009)

[7] Richard O. Duda, Peter E. Hart, David G. Stork, Pattern Classification: Second Edition. John Wiley and Sons, 2001.

Naive Bayes Mutivariate multinomial Distribution				
	Optimal Features		All Features	
	All Signals	No mixed sign	All Signals	No mixed sign
Noise	100%	100%	100%	100%
Tonal	97.3.%	97.3.%	97.3.%	97.3.%
Pulsed	100%	100%	100%	100%
Jawclap	100%	100%	100%	100%
No classification	40.1%	30.0%	40.1%	30.0%
Total	98.7%	98.7%	98.7%	98.7%

Naive Bayes Kernel Distribution				
	Optimal Features		All Features	
	All Signals	No mixed sign	All Signals	No mixed sign
Noise	96.2%	96.2%	97.1%	97.1%
Tonal	85.4.%	89.1.%	78.4.%	80.7.%
Pulsed	66.0%	74.7%	67.0%	76.3%
Jawclap	63.9%	83.3%	71.4%	97.1%
No classification	1.7%	0.2%	2.1%	0.2%
Total	84.2%	89.2%	82.1%	86.7%

Naive Bayes Gaussian Distribution				
	Optimal Features		All Features	
	All Signals	No mixed sign	All Signals	No mixed sign
Noise	91.3%	91.8%	93.1%	93.6%
Tonal	71.2%	76.4%	76.0%	84.2%
Pulsed	60.5%	69.2%	59.3%	67.5%
Jawclap	50.3%	60.0%	37.0%	32.1%
No classification	0.0%	0.0%	0.0%	0.0%
Total	77.0%	81.8%	74.3%	79.4%

Naive Bayes Multinomial Distribution				
	Optimal Features		All Features	
	All Signals	No mixed sign	All Signals	No mixed sign
Noise	69.1%	75.2%	69.6%	73.6%
Tonal	58.8%	60.4%	56.6%	53.6%
Pulsed	32.1%	40.5%	31.5%	37.1%
Jawclap	31.3%	50.0%	66.7%	83.3%
No classification	0.0%	0.0%	0.0%	0.0%
Total	60.5%	66.6%	58.6%	62.6%

Table 3: Comparative distribution percentages among different distributions. (Up-Down) N.B. multivariate multinomial, N.B. kernel, N.B. Gaussian and N.B. multinomial