Automatic prosodic clustering of humpback whales song

G. Picot[1], O. Adam[2], M. Bergounioux[1], H. Glotin[3], and F-X. Mayer[4]
[1]UMR 6628-MAPMO, University Orléans, France
[2]LiSSi-iSnS, University Paris Est, France
[3]UMR6168-LSIS, University of Toulon, France
[4]Megaptera Association, Ste Marie Island, Madagascar

***Abstract-* We automatically segmented a humpback whale song using the Roger Payne's principle of sound unit and we patterned and automatically classified intonations of sound units thanks to an unsupervised algorithm to describe recurrent patterns. We determined 6 different patterns of intonation. Some different sound units present the same pattern of intonation and one sound unit can be submitted to different patterns of intonation. Intonation could be related to information's transmission between humpback whales while they are singing.**

I. INTRODUCTION

Humpback whale songs are long cyclical sequences produced by males during the reproduction season which follows their migration from high-latitude to low-latitude waters. Singers of one geographical population share parts of the same song. This leads to the idea of dialect [1,2,3]. Different hypotheses of these songs were emitted [4,5,6], even as used as sonar [7,8,9].

In 1982, in a handmade work without using any software, Roger Payne exhibited a hierarchical structure of humpback whales songs. Defining a *sound unit* as any continuous sound between two silences, he showed that a song is an ordered sound units sequence. A so called subphrase is a particular sequence of 4 to 6 sequences. Subphrases compose phrases that compose themes. A song is composed of 3 to 9 themes and its length is about 20 to 40 minutes. Variations of songs perpetually occur; they are related to modifications of sound unit (duration, frequency) or to modifications of the vocalizations order. The explanation of this evolution remains unknown but it seems that one modification is shared by all the singers of of a population. Some long-term similarities between songs of neighbors populations exist that refutes any hypothesis of random modification [2,3,10].

We studied acoustic characteristics of songs by developing signal processing automatic methods. We present recurrent distinctive features of humpback whales songs and put forward new hypothesis about syntax or semantic content. We look for an automatic classification of specific intonations of different sound units of a same song. This goal is not new [11] but the originality of our work is to focus on prosody of humpback whale songs that is the role of intonation in the transmission of information via acoustic communication between different speakers.

II. METHODS

Using MATLAB© software, we have implemented an automatic segmentation algorithm based on Payne's principle to build up a database of sound units of a song. To analyze these vocalizations, we focus on the frequency range [0-10kHz] and set the minimal duration of a vocalization to 0.1s. Songs were sampled at 44.1kHz and encoded with a vectorial form. We calculate the spectrogram matrix using samples of 4096 coefficients and a shifting window of 2048 samples [12]. We kept the frequency curve corresponding to the maximal energy level in each column of the spectrogram matrix. We called it the characteristic frequency curve denoted $F0$. It is important to note that $F0$ is a priori different from the fundamental frequency f (Fig. 1).

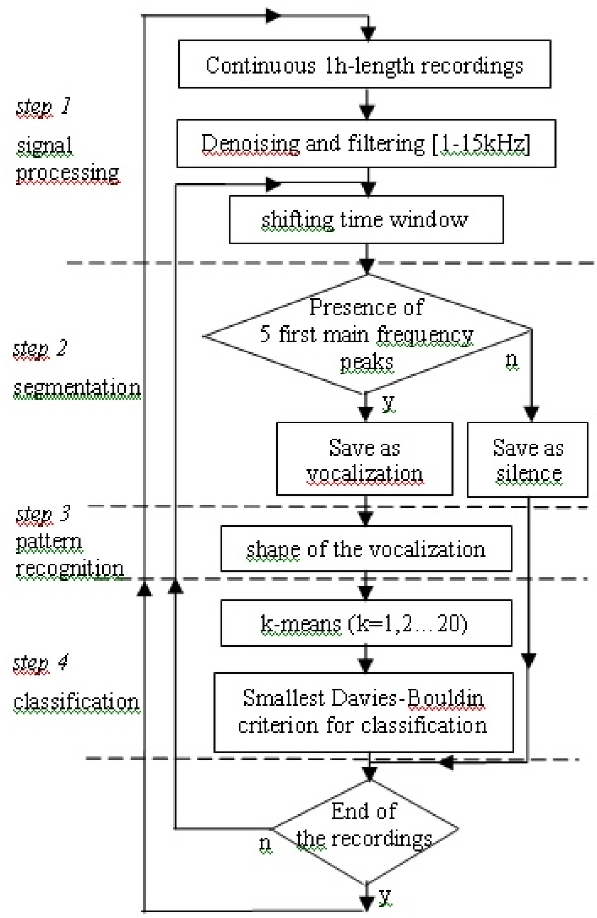


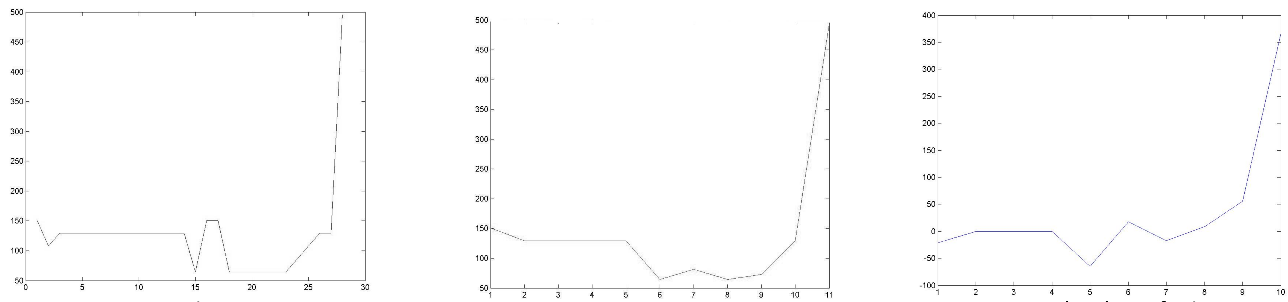
Fig. 1 Algorithm for the HW vocalizations analysis

978-1-4244-2816-8/08/$25.00 ©2008 IEEE

We extracted indexes of $F0$ corresponding to values superior to the 20Hz threshold of minimal frequency and distinguished sound units by listing couples of difference superior to 0.1sec. Vocalizations with too oscillating characteristic frequency or too weak energy or characteristic frequency were deleted to improve the algorithm accuracy.

We obtained a pattern of the intonation of a vocalization, assuming that humpback whales songs are harmonic and considering evolution of characteristic frequency. We listed frequency corresponding to each tenth of the total duration of a unit sound to approximate characteristic frequency with a piece-wise linear function. Then we calculated the slope of the segment joining collected frequencies to get an estimate of the "derivative curve" of the characteristic frequency (Fig. 2). The time evolution of the characteristic curve is an acceptable pattern of the intonation of vocalization. Thus we constituted the matricial database of intonations.

An automatic classification of intonations was performed using the unsupervised algorithm k-means which splits a database in k distinct clusters. We used the predefined function of Matlab© called *Kmeans_clusters* [13,14,15,16]. The number of clusters was varying between 1 and 20. The accuracy of an estimate was computed with the corresponding Davies-Bouldin index that gives information on the ratio between intra-clusters and inter-clusters similarity [17].



$F0$ Estimate of $F0$ Derivative of $F0$

Fig. 2: Variation of the characteristic frequency $F0$.

Let note the k^{th} classes of the set E: $E_1, E_2, \ldots, E_k$ with their respective kernel: $x_1, x_2, \ldots, x_k$. For $1 \leq i \leq k$, the diameter of E_i is defined by eq. 1:

$$\Delta(E_i) = \left[\frac{1}{card(E_i)} \sum_{x \in E_i} \|x - x_i\|_2^2 \right]^{\frac{1}{2}} \tag{1}$$

The distance between 2 classes is given by eq. 2:

$$\delta(E_i, E_j) = \|x_i - x_j\|_2 \tag{2}$$

We use the Davies-Bouldin criterion for the classification. This criterion is:

$$DB = \frac{1}{k} \sum_{i=1}^{k} \max_{1 \leq j \neq i \leq k} \left\{ \frac{\Delta(E_i) + \Delta(E_j)}{\delta(E_i, E_j)} \right\} \tag{3}$$

A low criterion (eq. 3) means a strong similarity intra-classes. Thus we deduced an optimal classification.

III. RESULTS

The automatic segmentation and classification of intonations algorithms were tested on a 24 minutes long song. 505 sound units were detected. The time reference points for silences and vocalizations were compared and turned out to be perfectly complementary. The listening of discerned sound units turned out to be satisfactory.

TABLE I
DAVIES-BOULDIN INDEXES

k	1	2	3	4	5	6	7	8	9	10
DB	Nan	1.48	1.59	1.77	1.49	1.44	1.42	1.43	1.27	1.51

k	11	12	13	14	15	16	17	18	19	20
DB	1.39	1.39	1.20	1.37	1.38	1.52	1.60	1.40	1.58	1.30

The intonations of the 505 sound units were modeled, pooled under matricial form and classified. The table I reports the Davies-Bouldin indexes of the different classifications.

We determined 13 clusters. The figure 3 reports the temporal organization of intonations.

72% of sound units are submitted to one intonation pattern. 5 other patterns gather more than 2.5% of the sound units. The last 7 patterns are too poorly represented to be significant and often seem to be very similar to main patterns. The surrounding marine noise may have induced a perturbation for the performed classification. We report below the 6 types of the listed intonations. They correspond to the 4, 7, 13, 12, 8 and 2 clusters classes.

The main pattern is the number 4 which gathers different sound units of frequencies between 0 and 5 kHz (Fig. 4). The characteristic frequency first slightly decreases; then it increases slowly showing a slight intonation. Modulations of rhythm and variations of the amplitude are picked out but global shapes of the 4 listed characteristic frequencies are similar. Below the characteristic frequency of each unit sound is juxtaposed to the spectrogram with a continuous line thanks to the software Praat (http://www.fon.hum.uva.nl/praat/). Its values depends on the [0-500Hz] range at the right-hand of the picture.

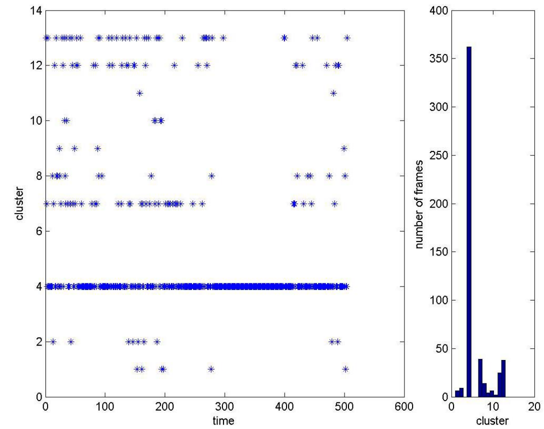


Fig. 3: time repartition of the intonations

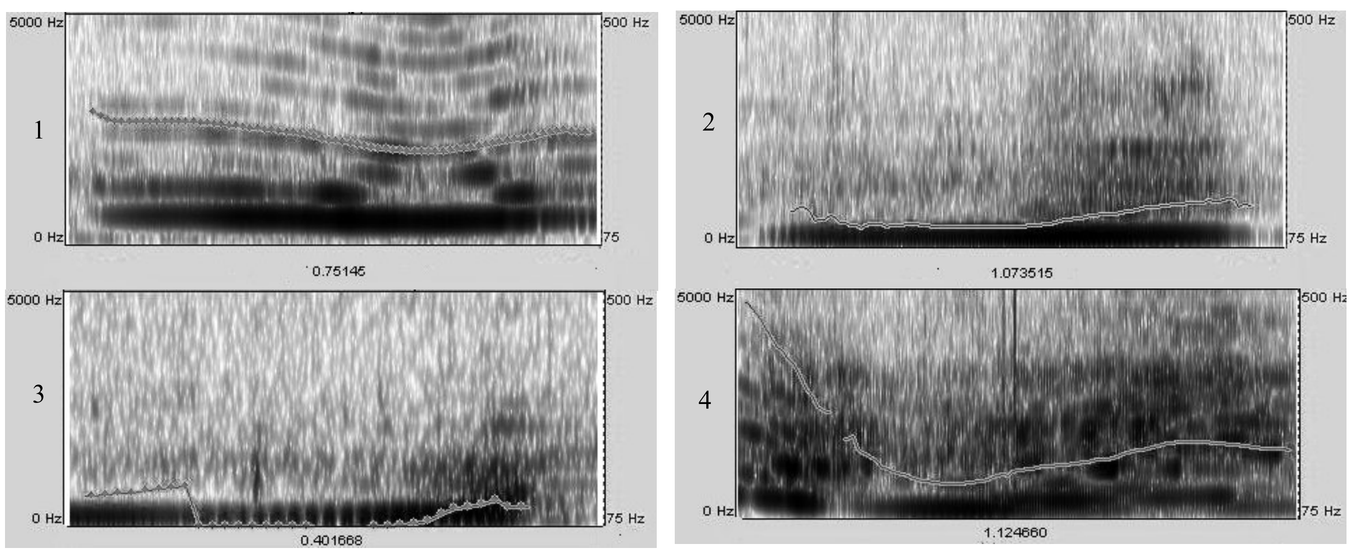


Fig. 4 Intonation 4. Vocalizations 1, 2, 3 and 4

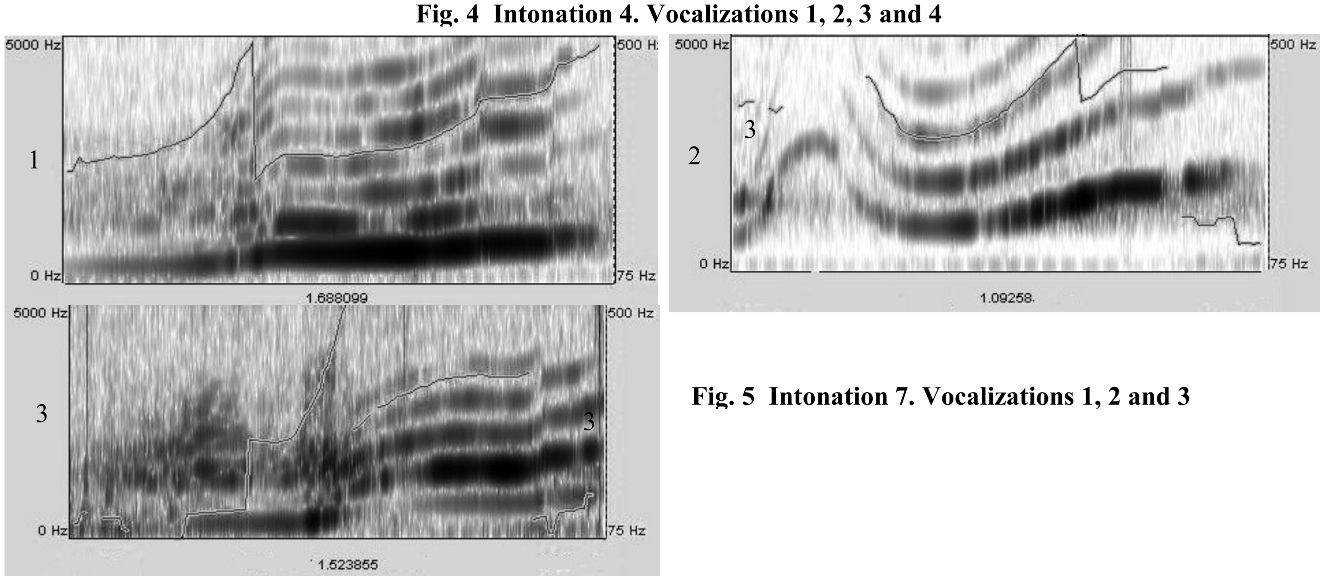


Fig. 5 Intonation 7. Vocalizations 1, 2 and 3

Pattern 7 gathers 7.5% of sound units. It is characterized by an increasing of the slope of the characteristic frequency starting at the half of the sound unit; it is followed by a sharp readjustment of the height of the note. The corresponding spectrum of unit sounds seems to be large in view of the 3 examples below (Fig. 5).

Pattern 13 describes a brutal falldown of the characteristic frequency at the half of the sound unit after a phase of frequency modulation of variable amplitude. It remains constant, falls again then remains constant before increasing. 2 models of sound units have this particularity which gathers 5.5% of the sound units of the song (Fig. 6).

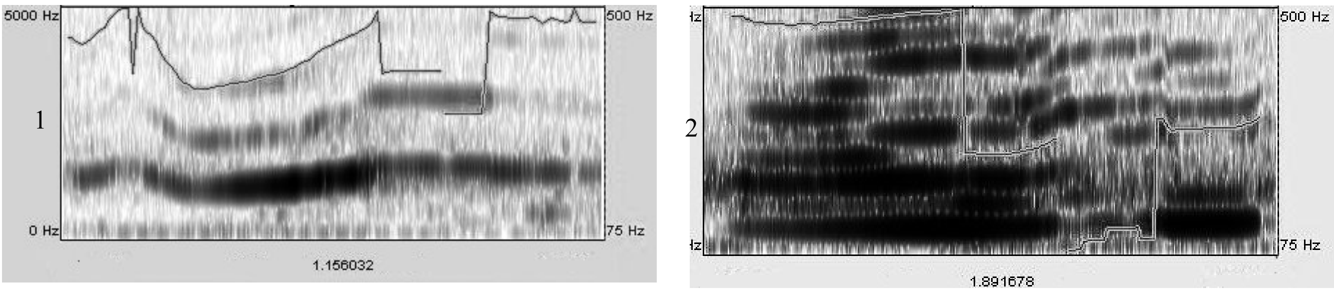


Fig. 6 Intonation 13. Vocalizations 1 and 2

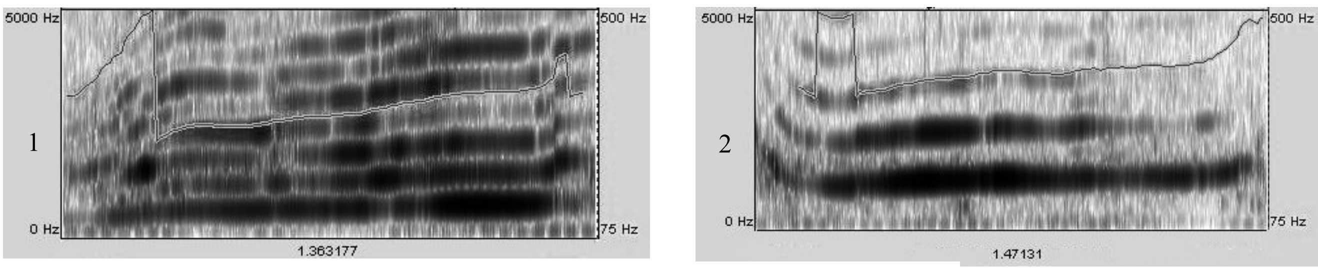


Fig. 7 Intonation 12. Vocalizations 1 and 2

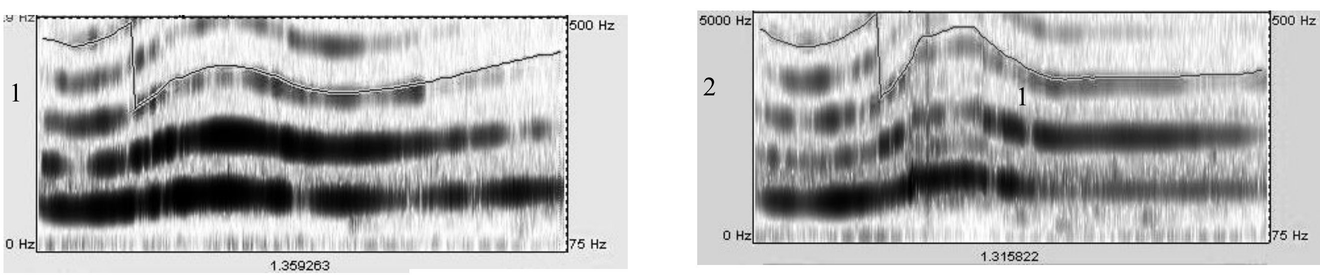


Fig. 8 Intonation 8. Vocalizations 1 and 2

The characteristic frequencies of vocalizations of the class 12 show a quick jump followed by a slow increasing ended by a sharp speeding up. 2 distinct sound units are thus identified (Fig. 7); it gathers 5% of total sound units. Note the likeness with the class 2. Moreover it seems that the first example of vocalizations of the class 4 is similar to the first example of vocalizations of the class 2. This makes us believe that the singer has changed the intonation of this sound unit during the song.

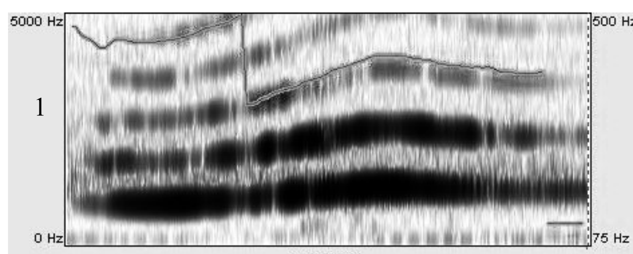


Fig. 9 Intonation 11. Vocalizations 1

Pattern 8 gathers 2.5% of the sound units (Fig. 7). There is one unique example of vocalization characterized by a sudden reduction of the characteristic frequency before a long phase of frequency modulation as in the class 4. The difference between the two patterns shows the importance of a slight modification of intonation of a sound unit.

Let us have a look on one of the unit sound that belongs to the class 11 (Fig. 8). It is very similar to vocalizations of the former class but with slightly different intonation: the duration before the fall of the characteristic frequency is longer. This observation gives a good illustration that one vocalization can be associated to distinct intonations.

TABLE II
DURATION OF VOCALIZATIONS OF EACH CLASS

Intonation	2	4	7	8	12	13
Duration (sec)	1.85	0.84	1.83	2.1	1.8	1.1

Pattern 2 shows several similarities with t pattern 7. There is a frequency peak at the half of the unit sound. The distinction between the two patterns is hard to establish but it could be explained by the less important increasing of the characteristic frequency in the pattern 2 than in the number 7. It gathers 2.5% of the sound units.

The table II gives the mean durations of vocalizations of each class of intonation.

The mean duration of sound units associated to intonation 4 is the only one inferior to 1sec. The shortest vocalizations are associated to the same pattern of intonation. The modulations of intonation are mainly applied on long vocalizations. The table below gives the probability that one pattern of intonation is followed by another. The way goes from intonation in the row to intonation in the column (Table III).

It is delicate to identify a coherent organisation of intonations. It appears that patterns of intonations 2, 7, 8, 12, and 13 phrase irregularly the song and that pattern 4 is the basic intonation. A more accurate pattern is required to try to explicit a structure of the song related to intonations.

TABLE III
TRANSITION MATRIX BETWEEN THE 6 MOST VISITED CLUSTERS (IN %). CLUSTER 4 IS DOMINANT.

From \ To	2	4	7	8	12	13
2	8	69	2	8	0	0
4	2	79	6	2	4	3
7	0	68	5	3	13	3
8	0	27	27	0	27	18
12	8	44	12	4	0	16
13	0	66	4	4	7	18

IV. . CONCLUSION

Automatic segmentation of the song was successfully tested. The characterization by the so-called characteristic frequency is useful to distinguish sound unit and ambient marine noise. This may be a good base to implement more efficient automatic segmentation algorithm. The study of the characteristic frequency of unit sound gave pertinent patterns of sound units. 6 distinct and recurrent patterns have been distinguished. We can note that this number also corresponds to the different vocalizations that Roger Payne distinguished in a whole song. Humpback whales seem to use as many vocalizations as intonations. Different vocalizations (duration, frequency, fundamental frequency, harmonics) showed similarities and we obtained examples of vocalizations occurring several times but with different intonations. Intonation could have a structural importance and transmit information. The signification of a sound unit could be altered with intonation. The mean duration of sound units of each class seems to be relevant that modifications of intonations affect the longest vocalizations. We can suppose that differences of intonation are more perceptible when they are related to long vocalizations. Correlations between intonation and duration of the vocalization have to be statistically studied. Moreover accurate patterns of intonation have also to be obtained. Other clustering methods will be tested as far as k-means is known to be quite dependant to the initial conditions. It will be also interesting to study the role of silences in the song and the energy of the vocalizations.

ACKNOWLEDGMENT

This work was supported in part by Association Megaptera (France), Association Dirac (France) and by University of Orléans (France).

REFERENCES

[1] D.A Helweg, D.H Cato, P.F Jenkins, C. Guarrigue and R.R McCauley, "Geographic variation in south pacific humpback whales songs," *Behavior*, 135, pp. 1-27, 1997.
[2] R.S. Payne, "Behavior and vocalizations of humpback whales (Megaptera sp)" pp. 56-78. *In:* K.S.Norris and R.R. Reeves (eds) Report on a workshop on problems related to humpback whales (Megaptera novaeangliae) in Hawaii, NTIS, PB-280794, Report No. MMC-77/03. Springfield, VA. 90pp. , 1978
[3] K. Payne, P. Tyack and R. Payne, "Progressive changes in the songs of humpback whales (Megaptera novaeangliae) : a detailed analysis of two seasons in hawaii", in *Communication and behavior of whales*, pp. 9-58, Westview Press, Boulder, 1983.

[4] L. Medrano, M. Salinas, I. Salas, P. Ladron de Guevara, A. Agayo, J. Jacobsen, and C.S Baker, "Sex identificationd of humpback whales, Megaptera novaeangliae, on the wintering grounds of the pacific ocean," *Canadian journal of zoology*, 72, pp. 1771-1774, 1994.
[5] A.S frankel, C.W Clark, L.M Herman and C.M Gabriele, "Spatial distribution, habitat utilization and social interactions oh humpback whales, Megaptera novaeangliae, off Hawaii, determined using acoustic and visual techniques", *Canadian journal of zoology*, 73, pp. 1134-1146, 1995.
[6] C.S Baker and L.M Herman, Aggressive behavior between humpback whale (Megaptera novaeangliae) wintering in Hawaiian waters, *Canadian journal of zoology*, 62, pp. 1922-1937, 1984.
[7] E. Mercado III and L.Frazer, "A sonar model for humpback whale song," *IEEE J. Oceanic Eng.,* vol 25, pp.160-183, Jan. 2000.
[8] W.W.L. Au, A.Frankel, D.A Helweg, and D.H Cato, "Against the humpback whale sonar hypothesis", *IEEE J. Oceanic Eng.*,26, pp. 295-300, 2001.
[9] E. Mercado III and L. N. Frazer, "Humpback Whale Song or Humpback Whale Sonar? A Reply to Au et al.," Jounal of the acoustical society of America, 100, p. 2644, 2001.
[10] H.E Winn, T.J Thompson, W.C Cummings, J. Hain, J. Hudnall, H.Hays and W.W Steiner, "Song of the Humpback Whale-Population comparison," *Behavioral Ecology and Sociobiology*, 8, pp. 41-46, 1981.
[11] E. Mercado III and A.Kuh, "Classification of Humpback Whale Vocalizations Using a Self-Organising Neural Network," ICASPP, 1998.
[12] C.Gasquet et P.Witomski, "Analyses de Fourier et Applications," Masson, 1995.
[13] P. Guillaume, "Musique et acoustique, de l'instrument à l'ordinateur", Hermes Science, 2005.
[14] P. Escudier, and J.L Schwartz, "La parole, des modèles cognitifs aux machines communicantes", Hermes Science, 2005.
[15] R. O. Duda, P.E Hart, and D. G. Stork, "Pattern Classification," John Wiley & Sons inc., 2001.
[16] J. P. Haton, C. Cerisara, D. Fohr, Y. Laprie, and K.Smaili, "Reconnaissance automatique de la parole,du signal à son interprétation, Dunod, 2006.
[17] D.L. Davies, and D.W. Bouldin, "A cluster separation measure," in *IEEE Transactions on Pattern Analysis ans Machine Intelligence*, Vol.1, No2,1979.