2015 IEEE 15th International Conference on Data Mining Workshops

Sparse coding for efficient bioacoustic data mining: Preliminary application to analysis of whale songs

Joseph Razik*, Hervé Glotin*†, Maia Hoeberechts‡, Yann Doh* and Sébastien Paris*
*Aix-Marseille Université, 13397 Marseille, ENSAM, France
Université de Toulon, 83957 La Garde, France
UMR CNRS LSIS 7296, Equipe DYNI
†Institut universitaire de France, 75005 Paris, France
‡Ocean Networks Canada, University of Victoria, BC, Canada

***Abstract*—Bioacoustic monitoring, such as surveys of animal populations and migration, needs efficient data mining methods to extract information from large datasets covering multi-year and multi-location recordings. This paper introduces a method for sparse coding of bioacoustic recordings in order to efficiently compress and automatically extract patterns in data. We demonstrate the proposed method on the analysis of humpback whale songs. Previous work suggests that the structure of these songs can be characterized by successive vocalizations called sound units. Most of these analyses are currently done with expert intervention, but the volume of recordings drive the need for automated methods for sound unit classification.**

This paper proposes that sparse coding of the song at different time scales supports the distinction of stable song components versus those which evolve year to year. The approach is summarized as: first, an unsupervised method is used to encode the entire bioacoustic dataset into a dictionary; second, sparse coding is used to limit the number of elements in the dictionary; third, salient features are identified using the Lasso algorithm; and finally, an interpretation of the evolving and stable components of the songs is derived, supporting an analysis of year to year variation. It is shown that shorter codes are more stable, occurring with similar frequency across two consecutive years, while the occurrence of longer units varies across years as expected based on the prior manual analysis. 250 ms segments appear to be an appropriate length for encoding stable features of whale songs, possibly corresponding to subunits. We conclude by exploring further possibilities of the application of this method for biopopulation analysis.

I. Introduction

Different kinds of vocalizations (moans and screams) are emitted by Humpback whales [1], reported as songs by Roger and Katy Payne [2], [3]. These songs are predominant in the breeding zone but have also been recorded during migration and occasionally in the feeding area [4], [5].

These sounds are emitted by male individuals [6]. Different hypotheses about these songs is that they possibly play a role in female attraction [7], [8], [9] and/or for strong interactions between males as territorial defense or challenges [10], [11]. Noad et al. [12] highlighted song copying between males from the Australian East coast and those from the Australian West coast.

Songs are cyclic and composed of a structured and continuous sequence of sounds that can be repeated several times without interruption. These short continuous sounds between 2 silences are called sound units [13]. The complex structure of these songs are based on successive specific sound units forming a sequence, several sequences forming a phrase, and several phrases forming a theme-song [13].

Current challenging objectives in the analysis of humpback whale songs include:

1) detection of the different kinds of vocalizations;
2) automatic classification of the sound units;
3) extraction of phrases of the songs;
4) localization of individuals and characterization of interaction of the singers.

Objectives 1. and 2. are particularly challenging because:

- there is a large diversity of the sound units (moans, growls, sets of pulses, cries and trumpet sounds). Sound units' features vary both in the time domain and in the frequency domain. The main frequency is from 100 Hz up to 20 kHz and the source levels could be more than 170 dB re 1 u Pa at 1 m (see examples in Fig. 1);
- singers emit sounds simultaneously;

978-1-4673-8493-3/15 $31.00 © 2015 IEEE
DOI 10.1109/ICDMW.2015.232

- varied underwater ambient noise is present.

For objectives 1. and 2., researchers proposed new approaches based on temporal and spectral features [14]. Some variations in these features were reported and could contain part of the information needed for detection and classification [15].

Methods used to analyze human speech have been applied to Humpback whale calls since they present several similarities including the presence of voiced and unvoiced type vocalizations, as defined in Mercado and Kuh[16]. Humpback whale calls have been analyzed using linear prediction coding (LPC) [16], energy content in specific time windows [17], spectrographic analysis [14], Mel-Frequency Cepstral Coefficients (MFCCs) [18], [19], [20], [21], affinity propagation [21], K-means [22], and classified with self-organizing maps (SOM) [16], [14], Hidden Markov Models (HMM) [23], the sliding window match length (SWML) entropy estimation [14] and neural networks. The great variety of methods used by researchers to analyze Humpback whale vocalizations reflects the great diversity of the features of these sounds [24].

The main drawback of most of these methods is that the number of sound units seems unknown, maybe unlimited, because songs are changing during the season, from one year to another and in the different breeding areas [25], [26], [27].

To better analyze these songs and to improve performance of classification, we recently introduced the concept of subunit [21], [28], [23]. We suggest that one or more subunits are present in one sound unit. The interest of this approach is to show that a number of subunits could be used for characterizing sound units, meaning that a sound unit could be built from a combination of these subunits.

This paper proposes for the first time a fully automatic sparse coding of humpback whale songs, to determine their stable components versus their evolving ones, at different time scales. We also propose a definition of code complexity which separates the song components from the background sea noise. We then explore the method's applicability to analyze the relative contribution sound units *vs.* subunits to song decomposition and evolution.

II. MATERIALS AND METHODS

A. Humpback whale recordings

For this paper we focus on recordings of humpback whales from Sainte Marie Channel (Madagascar). The recordings vary in length from several years to months or days, as they were collected by different collaborators (Megaptera, LAM, Cesigma). The datasets used were collected in 2008 and 2009. The frequency sampling is between 16 kHz and 44.1 kHz with 16 bit sample encoding. Figure 1 shows a spectrogram of a typical sample of a Madagascar song from our corpus. The duration for the different recording sites is presented in Tab. I.

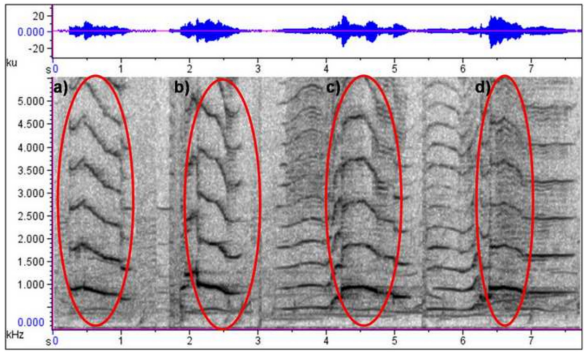

Fig. 1: Spectrogram of Madagascar song segments (from [23]).

Corpus name		Sampling info	Duration
Madagascar	2008	16 kHz, 16 bits	03:02:19.54
Madagascar	2009	44.1 kHz, 16 bits	04:38:24.27

TABLE I: Duration of available data according to dataset.

The hydrophone used for the recordings is a ColmarItalia GP280 (omni-directional, $[5Hz, 90kHz]$, sensitivity -170 dB *re 1 V/u Pa* (see datasheet on www.colmaritalia.it). The hydrophone was deployed from a motor boat (motor off), positioned $\approx$100 m in front of the singers at depth 20 m (the water column depth was between 40 m and 50 m).

In order to normalize the data according to other recording parameters, we have down sampled all the sound files to 16 kHz sampling frequency, 16 bits.

B. Cepstral representation

The first step in the analysis is to characterize the recorded songs by Mel-Frequency Cepstral Coefficients (MFCC) [29], [30], [22]. The use of the cepstral scale is motivated by the fact that mammals perceive frequency on a logarithmic scale along the cochlea [31], [32]. In our approach, we apply a method developed for human speech analysis to humpback whale vocalizations [28]. Rather than directly duplicating the method, we demonstrate its application for analyzing the harmonic parts of each sound unit or subunit of whale songs. We then build

codebooks at different time scales and explore their properties in characterizing song evolution.

We compute the 12 first static Mel-Frequency Cepstral Coefficients (MFCC), $M_1, M_2, \ldots, M_{12}$. To these 12 coefficients $M_1, M_2, \ldots, M_{12}$ we add a M_0 coefficient that captures the energy of the signal, thus yielding 13-dimensional MFCC vectors. These coefficients are computed with a 512 point Fast Fourier Transform (32 ms), with a window length of 250 ms and a frameshift of 10 ms.

On these resulting vectors, Cepstral Mean Subtraction (CMS) and variance normalization were applied. The extraction of these parameters is done with the SPro toolkit [33].

As songs' patterns are longer than a 10 ms scale, we form super-vectors by concatenating MFCC vectors. These super-vectors form the dictionary words in the sparse analysis (see Section II-C). We consider words of length 250 ms, 500 ms, 1 s, 2 s, and 4 s, which are formed by concatenating 25, 50, 100, 200 and 400 MFCC vectors respectively. In order to be sure to capture sound units, the vectors are concatenated with a 50% overlap; for example, the MFCC vectors we manipulate for 500 ms scale are 650-dimensional vectors (13×50 component), one every 250 ms.

C. Dictionary and sparse coding

In order to support efficient bioacoustic data mining, the large MFCC vectors are encoded by a learned dictionary and a sparse code. The sparse code identifies how dictionary words are recombined to produce a reconstructed representation of the original vectors. This dictionary was learned on the union of the MFCC representations of the original humpback whale song datasets. In this section we explain the details of the sparse coding method.

In order to obtain a global robust representation of the signal $\mathbf{X} = [\mathbf{x_1}, \ldots, \mathbf{x}_N] \in \mathbb{R}^{n \times N}$, each MFCC vector $\mathbf{x_i}$ (n = 650 in the case of 500 ms) are first linearly encoded as the vector $\mathbf{c_i} \in \mathbb{R}^k$ such that $\mathbf{x_i} \approx \mathbf{D c_i}$ where $\mathbf{D} \triangleq [\mathbf{d}_1, \ldots, \mathbf{d}_k] \in \mathbb{R}^{n \times k}$ is a preliminary trained dictionary with the constraint $\|\mathbf{d}_j\|_2 = 1$. In a first attempt to solve this linear problem, $\mathbf{c_i}$ can be the solution of the Ordinary Least Square (OLS) problem:

$$l_{OLS}(\mathbf{c_i}|\mathbf{x_i};\mathbf{D}) \triangleq \min_{\mathbf{c_i} \in \mathbb{R}^k} \left\{ \frac{1}{2} \|\mathbf{x_i} - \mathbf{D c_i}\|_2^2 \right\} \quad (1)$$

OLS formulation can be extended to include a regularization term avoiding data overfitting. Thus, we obtain the ridge regression (RID) formulation:

$$l_{RID}(\mathbf{c_i}|\mathbf{x_i};\mathbf{D}) \triangleq \min_{\mathbf{c_i} \in \mathbb{R}^k} \left\{ \frac{1}{2} \|\mathbf{x_i} - \mathbf{D c_i}\|_2^2 + \beta \|\mathbf{c_i}\|_2^2 \right\} \quad (2)$$

This problem can be analytically solved and then $\mathbf{c_i} = (\mathbf{D}^T\mathbf{D} + \beta \mathbf{I}_k)^{-1}\mathbf{D}^T\mathbf{x_i}$. In order to decrease reconstruction error and to have a sparse solution, this problem can then be reformulated as a constrained Quadratic Problem (QP):

$$l_{QP}(\mathbf{c_i}|\mathbf{x_i};\mathbf{D}) \triangleq \min_{\mathbf{c_i} \in \mathbb{R}^k} \frac{1}{2} \|\mathbf{x_i} - \mathbf{D c_i}\|_2^2 \; s.t. \; \|\boldsymbol{c_i}\|_1 = 1 \quad (3)$$

To solve this problem, we can use a QP solver involving high combinational computation to find the solution. Under the RIP assumption [34], a greedy approach can be used efficiently to solve Eq. 3. Finally, the sparse code (SC) is defined by:

$$l_{SC}(\mathbf{c_i}|\mathbf{x_i};\mathbf{D}) \triangleq \min_{\mathbf{c_i} \in \mathbb{R}^k} \frac{1}{2} \|\mathbf{x_i} - \mathbf{D c_i}\|_2^2 + \lambda \|\mathbf{c_i}\|_1, \quad (4)$$

where λ is a regularization parameter which controls the level of sparsity of the sparse code c_i. This problem is also known as basis pursuit [35] or Lasso [34] problem. To solve this problem, we can use the popular Least Angle Regression (LARS) algorithm.

The dual part of the training of the dictionary $\mathbf{D}$ and the computation of the projection sparse codes $\mathbf{C} = [\mathbf{c_i}, \ldots, \mathbf{c_N}]$ supports the reconstruction of the MFCC vectors. For an MFCC vector $\mathbf{x} \in \mathbb{R}^n$ and the associated sparse code vector $\mathbf{c} \in \mathbb{R}^k$, the reconstructed MFCC vector $\hat{\mathbf{x}}$ is the linear combination of the dictionary codebook vector $\mathbf{d}_i$ according to c_i values of the sparse code $\mathbf{c}$. More formally, $\hat{\mathbf{x}}$ is given by the following equation:

$$\hat{\mathbf{x}} = \mathbf{D}.\mathbf{c} = \sum_{i=1}^{k} \mathbf{d}_i.c_i \quad (5)$$

D. Dictionary size

An important aspect of the encoding is the choice of an appropriate size for the dictionary. The goal of sparse coding is to create an encoding of the larger dataset which maintains structure in the data to facilitate analysis, but at the same time reduces the size of the encoding to permit efficient computation.

One drawback of sparse coding is that the size of the dictionary has to be fixed manually and this size should not over complete the expected number of classes after clustering. In this experiment, we learned three dictionaries with $K = 16$, $K = 32$

and $K = 1024$ words respectively. The $K = 32$ dictionary is used for the analyses in presented in III, as it was empirically determined to support the best discriminative representation of the full vectors.

E. Relationship between MFCC signal and sparse code

In this section we present an approach to verifying that a given sparse code is representative of its original set of MFCC vectors. Each MFCC vector corresponds one sparse code vector. We expect that patterns appearing with MFCC auto-correlation to still appear with the sparse code vector auto-correlation. In Fig. 2, we present an example for the dictionary with $K = 1024$ words. The figure shows the 2009 recording auto-correlation with MFCC vectors on the top and sparse codes vectors on the bottom over a subset of 400 randomly chosen samples. We can note that if there is information in the MFCC space, this information also appears in the sparse code space. This is seen in the figure as the corresponding structure in the auto-correlations.

F. Codebook complexity estimation

The estimation of complexity of time-frequency plane can include moment-based measures such as time-bandwidth product and the Shannon and Renyi entropies [36]. In order to analyze the dictionary we generate, we extend the time-frequency complexity definition to cepstral pattern complexity, based on the principle that a concentration of energy in the time-frequency plane will also generate an energy concentration in the cepstral plane. We investigate a quantitative measure of complexity inspired from existing work [36]. This measure is closely related to the assumption that signals of high complexity (and therefore high information content) must be constructed from a large numbers of elementary components. We thus define the complexity measure of the sparse vectors $\mathbf{d}_i$ of the dictionary $\mathbf{D}$ as the Shannon entropy:

$$H(\mathbf{d}_i) = -\sum_{t,j} p(\mathbf{d}_i(t, M_j)).\log(p(\mathbf{d}_i(t, M_j))), \tag{6}$$

where $p(\mathbf{d}_i(t, M_j))$ is the estimate of the energy distribution at time t for the cepstral coefficient M_j. The codebook for dictionary size $K = 32$, time scale 250 ms, sorted by their complexity measure is shown in Fig. 3.

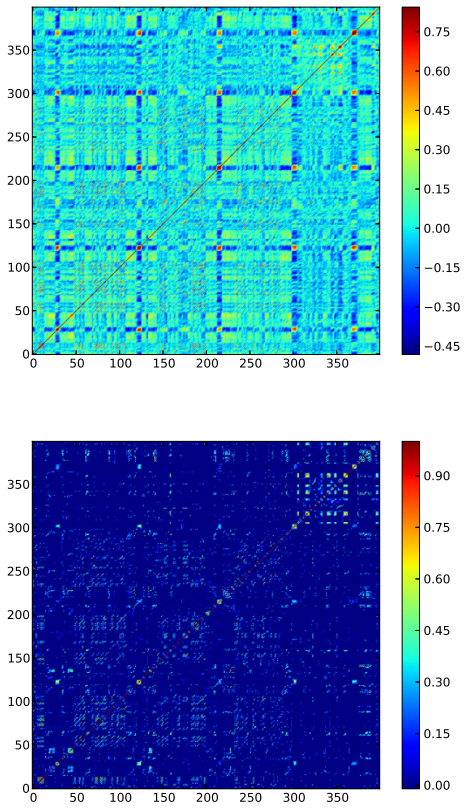

Fig. 2: Representation of the auto-correlation matrix between MFCC vectors (a), and sparse code vectors (b). Results obtained on the 4 s encoding of the corpus with a dictionary of 1024 vectors.

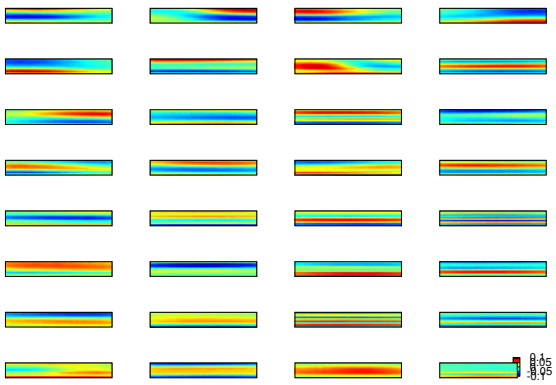

Fig. 3: The codebook composed of 32 codes, sorted by degree of complexity (from top left to bottom right), computed at the time scale of 250 ms, learned from the union of 2008 and 2009 song sets.

G. Divergence measure of the codes

Information Theoretic methods support the analysis of structure and the organization within a communication system [37]. As our goal is to analyze differences between whale songs (communication system) over different years, we propose to use an information theoretic measure to estimate the song divergence.

In order to get a diachronic analysis, *i.e.* to determine which code is more or less used from one year to one other, we compute the Kullback-Liebler distance [38] over song components as represented in the sparse coding. We interpret a difference in the average of the Kullback-Leibler distance for a song encoding subset between 2008 and 2009 recordings as an evolution of the song, assuming that higher this distance is, the more the songs evolved from one year to the other.

Therefore, the song distance is defined as follows. Let be $A_{\mathbf{d}_i}$ (resp. $B_{\mathbf{d}_i}$) be the discrete probability distribution over R bins $r = \{1, \ldots, R\}$, of the 2008 $\mathbf{C}$ sparse codes for the sparse vector $\mathbf{d}_i$ (resp. 2009). Then the distance for the sparse vector $\mathbf{d}_i$ is:

$$dist_{KL}(A_{\mathbf{d}_i}, B_{\mathbf{d}_i}) = \sum_{r=1}^{R} (A\mathbf{d}_i^r - B\mathbf{d}_i^r).\log_2(A\mathbf{d}_i^r / B\mathbf{d}_i^r) \tag{7}$$

Finally, the final song distance is the average of the $dist_{KL}$ of target code subsets.

III. RESULTS

The results in this section are computed over a sparse code and corresponding $K = 32$ word dictionary. As explained in II-B, the MFCC input vectors are composed of concatenations of 13 coefficient vectors to form super-vectors with length of 250 ms, 500 ms, 1 s, 2 s to 4 s.

A. Analysis of song code complexity

Fig. 4 shows a graph of the sorted complexity (as defined in Section II-F) of the quefrequency words of the learned dictionary. We hypothesize an interpretation of the complex code words, which exhibit energy variation in time and frequency, as encodings of whale song subunits and of the less complex words, which exhibit more uniformity, as components of sea noise.

The next section will compare the evolution between 2008 and 2009 of the songs with reference to the sparse coding.

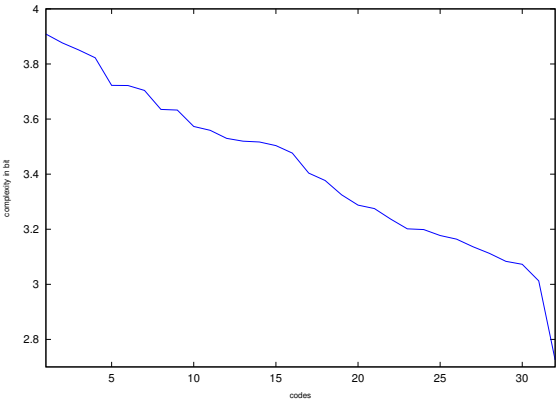

Fig. 4: Complexity values of the 32 codes (sorted) of the codebook illustrated previously (time scale 250 ms learned on 2008 union 2009). The difference between highest and lowest complex codes is significant.

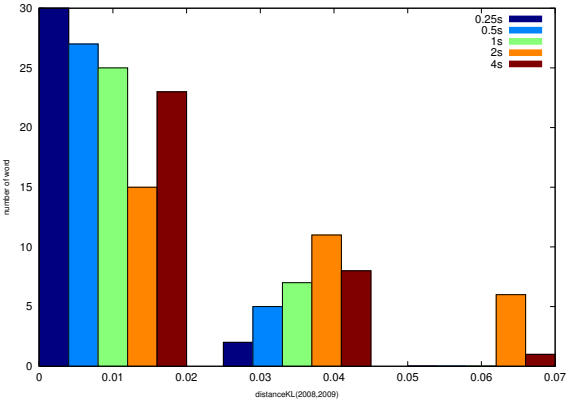

Fig. 5: Histogram of the KL distance (2008, 2009) computed over the 32 sparse vectors of the dictionary.

B. Analysis of song evolution

Using the formula given in Section II-G, we compute the Kullback-Liebler (KL) distance between the 2008 and 2009 code words, resulting in 32 distances for each set of super-vectors of increasing length (250 ms, 500 ms, 1 s, 2 s, 4 s). The histogram of the distribution of these distances grouped into bins of average distance is shown in Fig. 5. It can be seen from the distribution that short-duration (250 ms) representations are more stable across years than longer ones.

In order to determine whether the evolution can be attributed to changes in the more or less complex code words, we compute the KL distance between

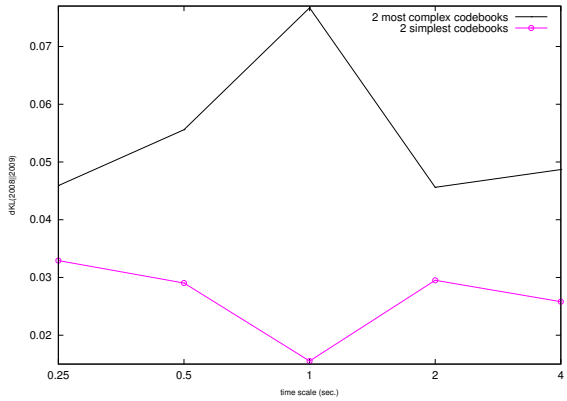

Fig. 6: Distance between 2008 and 2009 songs, in average over the two most complex codes, versus the two simplest ones. In abscissa the time scale of the code, from 250 ms to 4 s.

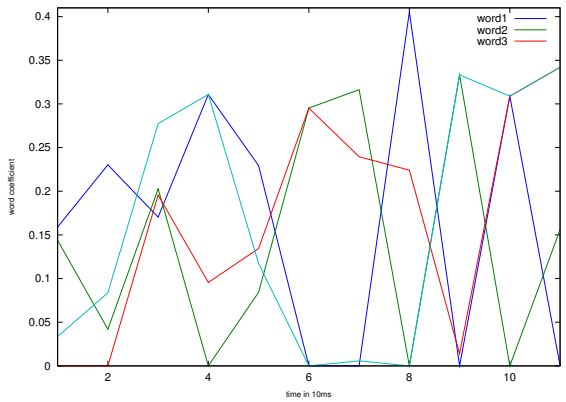

Fig. 7: Contributions of four code words to 2008 song segment encoding for dictionary size $K = 16$.

2008 and 2009 for the 2 most complex codes, versus the 2 least complex code words, for all lengths. The divergence analysis (Fig. 6) illustrates that the simplest code words show far less variation across years than the most complex code words. Furthermore, the complex code words are similar for short durations (the KL distance is low for 250 ms), but differ at longer durations with the largest variability at the time scale of 1 s.

We can interpret this result to means that the code of 1 s scale are vary year by year, and may be composed of stable codes on a shorter time scale which exhibit less variation. This result is compatible with the subunit concept [22], which postulates that evolving whale songs are composed from shorter stable song elements. Our results suggest that the subunit could be coded at the 250 ms time scale, while the units would be coded at the 1 second scale. The longer time scale (2 and 4 seconds) are less diverging, possibly due to the fact that this time scale is relative to global song structure that may vary less than the unit level.

Note that the distance computation directly on the raw MFCC yields, as expected, to insignificant difference in distance, whatever the size or the year of the units. This effect is known as the "dimensionality curse effect" which makes the KL-distance metric inefficient in high-dimensional spaces. Indeed, the dimensionality of the MFCC vectors is 650 (13x50), and according to [39], any simple distance computation of any pair of vectors in such a high dimensional space results will result in a similar distance.

IV. DISCUSSION

We presented an unsupervised dictionary learning algorithm for generating a proto-lexicon of the songs of Humpback whales at different time scales. These dictionaries are used as the basis for a sparse encoding of the original datasets. These representations are more generic and efficient than obtained in our previous method [22].

We show in this paper the utility of a sparse representation of complex bioacoustic patterns for efficient data mining. We presented the hypothesis that the long and short time duration sequences might represent varying sound units and stable subunits of whale songs, respectively. In order to support this hypothesis, in future work we intend to further analyze the structure of the representation of the sparse coding with respect to the original acoustic dataset. Two ideas for this analysis are presented here, computed on a dictionary with $K = 16$.

Fig. 7 shows the relative contributions of code words to the encoding of a song segment from the 2008 dataset. This type of analysis is suggested as a means of characterizing the song structures. We see clearly the variation of the code word activity. For example at 60 ms, words 2 and 3 contribute simultaneously to generate a complex pattern, whereas at 80 ms, word 1 is more active than any other word. Note that code words do not appear sequentially in the song, but rather contribute relatively to the overall song structure. They do not represent independent temporal components of the original acoustic recording. This implies that a human listener would be unlikely to hear individual code words, but rather mixtures of code words.

To determine whether the code words actually

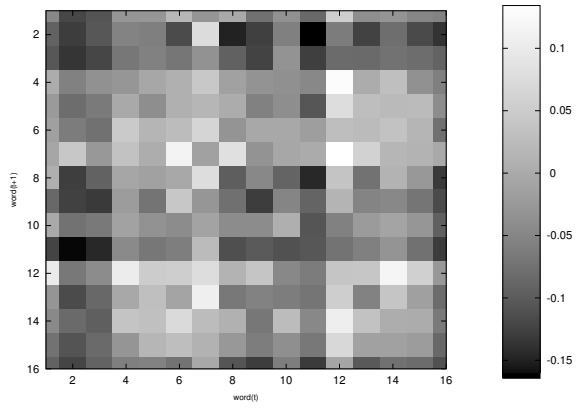

Fig. 8: Log ratio of the probabilities of sparse vector pairs from the 2008 dataset versus the 2009 dataset. This is illustrated for $K = 16$, scale 250 ms.

represent meaningful sound subunits, we could consider a pairwise analysis of the probability of the occurrence of bigrams in the encoded representation. As an example, again presented on the dictionary with $K = 16$, we computed the probability of occurrence each consecutive word (sparse vector) pair for the 2008 and 2009 datasets. The probability of the ordered occurrence for the pair (w_1, w_2) (bigram), is written as $P(w_1, w_2)$. In a random system, we would not expect to see patterns across years in the probabilities.

We then computed the log ratio of these probabilities from the 2008 and 2009 encoded datasets at the 250 ms time scale, as shown in Fig. 8.

The results show that there are differences in bigram occurrence across years. For example we see that pair $(6, 7)$ is more frequent in 2008, while pair $(6, 2)$ is less frequent in 2008. This could be interpreted as song evolution (encoded as order of words) from one year to another, while subunits (encoded as words) remain stable.

The sparse coding technique shows promise as a means to provide objective insight into evolution of song structure, although further development is needed to support strong conclusions about the interpretation of the encoding.

V. Conclusion

This paper demonstrates the promise of the sparse coding method to learn features in an unsupervised manner for humpback whale songs and to support their analysis. The sparse dictionary that is presented in this paper is automatically learned from the recordings, and captures variations in sound units and subunits with a limited number of elements. This approach is suitable for analyzing signals with which contain both stable and variable features. The acoustic datasets processed in this work exhibit those characteristics, where the variability is found among singers and between years, but stable features are present for example, in the ambient noise.

In summary, we presented:

1) an unsupervised method for encoding a large, variable bioacoustic dataset into a dictionary;
2) a method which establishes criteria for sparse coding to limit the number of the elements of this dictionary;
3) using the Lasso algorithm, the sparse coding distinguishes the salient features of the signal from the noise components;
4) an interpretation of the approach in characterizing evolving sound units and stable subunits;
5) an analysis of year to year variation.

The results establish sparse coding as a promising method for analyzing humpback whale songs.

This paper lends new support for the concept of units versus component subunits in humpback whale songs. We show that the shortest units (subunits) are the most stable, occurring with similar frequency across two consecutive years, while the longest units exhibit more variation from one year to one other. 250 ms segments appear to be an appropriate length for encoding stable features of whale songs, possibly corresponding to subunits.

In future work, a systematic information theoretic analysis will be used to characterize the evolution of sound units. The approach will be applied to multiple geographic locations across multiple years to further explore population differences and song evolution.

Another potential application would be to model the vocal identity of individual whales, which could provide a basis for singer authentication or dialect identification. We also intend to explore the method's applicability to analyzing sounds from other species beyond humpback whales.

Acknowledgment

The authors would like to thank Cetamada NGO (www.cetamada.org), LAM and Cesigma for recordings and Fondation Total for the recordings during our project BAOBAB. Partly supported by Institut Universitaire de France, and by MASTODONS SABIOD projet from Mission Interdisciplinarité (MI) du CNRS.

REFERENCES

[1] W. Schevill, "Underwater sounds of cetaceans," *Marine Bio-Acoustics, ed by W.N. Tavolga (Pergamon, Oxford)*, pp. 307–316, 1964.

[2] Anonymous, "Singing whales," *Nature*, vol. 224, p. 217, 1969.

[3] H. E. Winn, P. J. Perkins, and T. C. Poulter, "Sounds of the humpback whale," in *7th Annual Conference on Biological Sonar and Diving Mammals*, 1970, pp. 39–52.

[4] P. J. Clapham and D. K. Mattila, "Humpback whale songs as indicators of migration routes," *Marine Mammal Science*, vol. 6, no. 2, pp. 155–160, 1990.

[5] C. W. Clark and P. J. Clapham, "Acoustic monitoring on a humpback whale (megaptera novaeangliae) feeding ground shows continual singing into late spring," *Proceedings - Royal Society of London. Biological sciences*, vol. 271, no. 1543, pp. 1051–1057, 2004.

[6] D. A. Glockner, "Determining the sex of humpback whales (megaptera novaeangliae) in their natural environment," in *R. Payne, ed. Communication and behavior of whales*, 1983, pp. 447–464.

[7] H. Winn and L. Winn, "The song of the humpback whale megaptera novaeangliae in the west indies," *Mar. Biol.*, vol. 47, pp. 97–114, 1978.

[8] L. M. Herman and W. N. Tavolga, "The communication systems of cetaceans," *Cetacean behavior: Mechanisms and function*, pp. 149–209, 1980.

[9] L. Medrano, M. Salinas, I. Salas, P. L. D. Guevara, A. Agayo, J. Jacobsen, and C. Baker, "Sex identification of humpback whales, megaptera novaeangliae, on the wintering grounds of the pacific ocean," *Canadian Journal of Zoology*, vol. 72, pp. 1771–1774, 1994.

[10] J. Darling, "Migrations, abundance and behavior of hawaiian humpback whales (megaptera novaeangliae)," Ph.D. dissertation, University of California Santa Cruz, 1983.

[11] D. Cholewiak, "Evaluating the role of song in the humpback whale (megaptera novaeangliae) breeding system with respect to intra-sexual interactions," Ph.D. dissertation, Cornell University, 2008.

[12] M. Noad, D. Cato, M. Bryden, M. Jenner, and K. Jenner, "Cultural revolution in whale songs," *Nature, London*, vol. 408, p. 537, 2000.

[13] R. Payne and S. McVay, "Songs of humpback whales," *Science*, vol. 173, no. 3997, pp. 585–597, 1971.

[14] P. Suzuki, J. Buck, and P. Tyack, "Information entropy of humpback whale songs," *J. Acoust. Soc. Am.*, vol. 119, no. 3, pp. 1849–1866, 2006.

[15] W. Au, M. Lammers, A. Stimpert, and M. Schotten, "The temporal characteristics of humpback whale songs," *J. Acoust. Soc. Am.*, vol. 118, no. 3, p. 1940, 2005.

[16] E. Mercado III and A. Kuh, "Classification of humpback whale vocalizations using a self-organizing neural network," in *IEEE World Congress on Computational Intelligence*, vol. 2, 1998, pp. 1584–1589.

[17] P. Rickwood and A. Taylor, "Methods for automatically analyzing humpback song units," *J. Acoust. Soc. Am.*, vol. 123, no. 3, pp. 1763–1772, 2008.

[18] D. Helweg, "Geographic and temporal variation in songs of humpback whales," *J. Acoust. Soc. Am.*, vol. 100, no. 4, p. 2609, 1996.

[19] S. Mazhar, T. Ura, and R. Bahl, "An analysis of humpback whale songs for individual classification," *J. Acoust. Soc. Am.*, vol. 123, no. 5, p. 3774, 2008.

[20] G. Picot, O. Adam, M. Bergounioux, H. Glotin, and F. Mayer, "Automatic prosodic clustering of humpback whales song," in *PASSIVE 08*, I. explorer, Ed., 2008, p. 6p.

[21] H. Glotin, L. Gauthier, F. Pace, F. Benard, and O. Adam, "New automatic classification for humpback whale songs," in *PASSIVE 08*, P. university and ONR, Eds., 2008, p. 93.

[22] F. Pace, F. Benard, H. Glotin, O. Adam, and P. White, "Subunit definition and analysis for humpback whale classification," *Journal of Applied Acoustics*, vol. 71, november 2010.

[23] F. Pace, P. R. White, and O. Adam, "Classification of humpback whale (megaptera novaeangliae) calls using hidden markov models," in *5th International Workshop on Detection, Classification, Localization, and Density Estimation of Marine Mammals using Passive Acoustics*, 2011, p. 29.

[24] J. G. Harris and M. D. Skowronski, "Automatic speech processing methods for bioacoustics signal analysis: a case study of cross-disciplinary acoustic research," in *ICASSP*, vol. 5, 2006, pp. 793–796.

[25] R. Payne and L. N. Guinee, "Humpback whale (megaptera novaeangliae) songs as an indicator of "stocks"," *R. Payne, ed. Communication and behavior of whales*, pp. 333–358, 1983.

[26] D. A. Helweg, L. A. Herman, S. Yamamoto, and P. H. Forestell, "Comparison of songs of humpback whales (megaptera novaeangliae) recorded in japan, hawaii, and mexico during the winter of 1989," Cetacean Research Institute, Tech. Rep. 1, 1990.

[27] S. Cerchio, J. K. Jacobsen, and T. F. Norris, "Temporal and geographical variation in songs of humpback whales, megaptera novaeangliae: Synchronous change in hawaiian and mexican breeding assemblages," *Animal Behaviour*, vol. 62, pp. 313–329, 2001.

[28] F. Pace, P. White, and O. Adam, "Characterisation of sound subunits for humpback whale song analysis," in *4th International Workshop on Detection and Localization of Marine Mammals using Passive Acoustics*, 2009, p. 56.

[29] S. Davis and P. Nermelstein, "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," *IEEE Trans ASSP*, vol. 28, pp. 357–366, 1980.

[30] L. Rabiner and B. H. Huang, *Fundamentals of Speech Recognition*. Englewood Cliffs, NJ: Prentice Hall, 1993.

[31] D. D. Greenwood, "Auditory masking and the critical band," *Journal of the Acoustical Society of America*, vol. 33, pp. 484–502, 1961.

[32] ——, "Critical bandwidth and the frequency coordinates of the basilar membrane," *Journal of the Acoustical Society of America*, vol. 33, no. 1344–1356, 1961.

[33] G. Gravier, "Spro: a free speech signal processing toolkit," 2010, vers. 5.0. https://gforge.inria.fr/projects/spro (date last viewed 06/07/12).

[34] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society, Series B*, vol. 58, pp. 267–288, 1994.

[35] S. S. Chen, D. L. Donoho, Michael, and A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, pp. 33–61, 1998.

[36] P. Flandrin, R. G. Baraniuk, and O. Michel, "Time-frequency complexity and information," in *IEEE International Conference on Acoustics, Speech, and Processing*, vol. 3, 1994, pp. 329–332.

[37] C. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379–423,623–656, 1948.

[38] S. Kullback and R. Leibler, "On information and sufficiency," *Annals of Mathematical Statistics*, vol. 22, no. 1, pp. 79–86, 1951.

[39] K. S. Beyer, J. Goldstein, R. Ramakrishnan, and U. Shaft, "When is "nearest neighbor" meaningful?" in *Proc. of the 7th International Conference on Database Theory (ICDT)*. Springer-Verlag London, 1999, pp. 217–235.