Humpback whale song theme recognition tool

Julie Patris, Franck Malige, D. Djokic, Renata Sousa-Lima, Hervé Glotin

Humpback whale songs and themes recognition tool

Julie Patris(1,4,5) **Franck Malige**(1,4,5) **Divna Djokic**(2,3)
Renata Sousa-Lima(2,3) **Hervé Glotin**(1,4,5)

(1) Université de Toulon, Aix Marseille Université, CNRS, LIS, DYNI team, Marseille, France

(2) Laboratory of Bioacoustics, Department of Physiology and Behavior, Universidade Federal do Rio Grande do Norte (UFRN), Natal, Brazil

(3) Graduate Program of Psychobiology, Biosciences Center, UFRN, Natal, Brazil

(4) BRILAM, STIC AmSud

(5) EADM GDRS CNRS MADICS, France

divna.divna@yahoo.com, julie.patris@univ-amu.fr

Introduction

Here we present a tool to visualize the structures in recordings of humpback whales (*Megaptera novaeangliae)*. This tool runs on inputs of the sound transcribed into a list of letters, named after each type of unit in the song (see Payne and McVay, 1971 [4]), but it is independent of the method of this transcription.

This is a very visual way of representing the sequence of units, allowing immediate identification of the structures of the highly organized vocalization emitted by male humpback whales. It is also used for automatic unsupervised selection of different structure levels in the recording and could be used for monitoring changes in the songs.

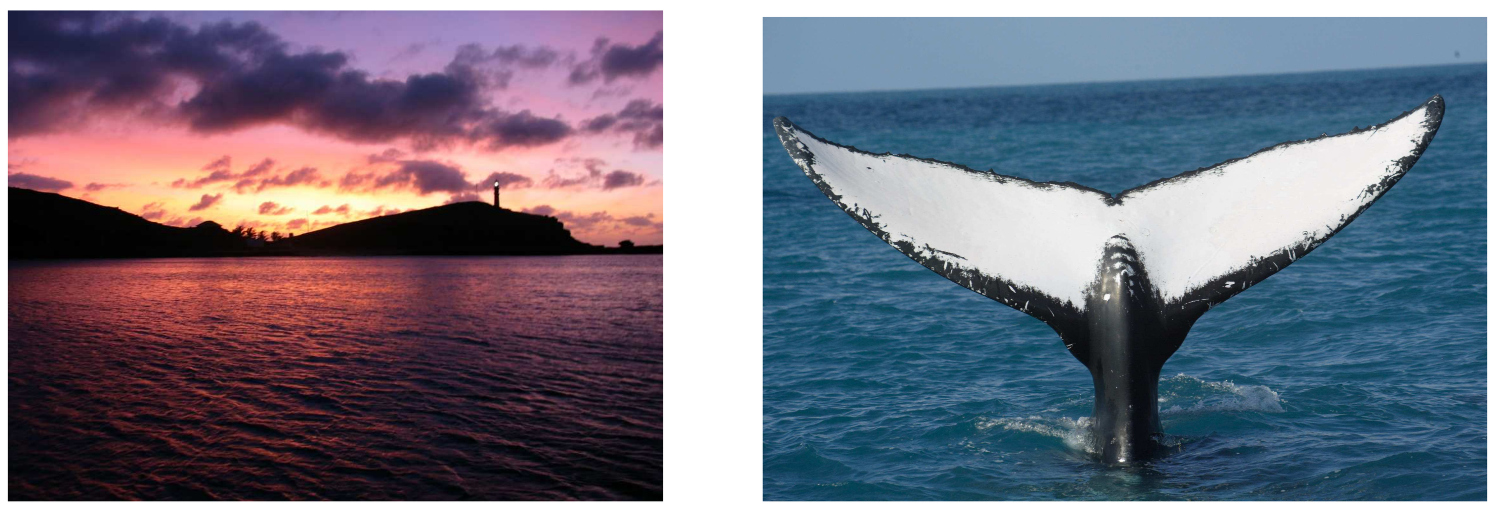

Figure 1: Left: Abrolhos archipelago sunset, Brazil, where most of the data were recorded. Right : a humpback whale. Pictures : R. Sousa-Lima .

Data

Two sets of data were used. One was recorded around Abrolhos archipelago, Brazil (see figure 1), in 2000 and 2001 and the other around Cano Island, on the Pacific coast of Costa Rica in 2016 (thanks to a collaboration with Laura May-Collado, Vermont university). Data from Brazil are owned by the Laboratory of Bioacoustics, UFRN (Brazil) and Cornell University (USA). Costa Rica data were sampled at 44100 Hz, and data from Brazil at 48000 Hz.

In total, 520 minutes of recordings were transcribed into lists of units (340 min. from Brazil and 180 min. from Costa Rica). All the data were reviewed visually and aurally using spectrograms created in software Raven Pro 1.5, using an FFT window size of 1024 pt., Hann window, and 50% overlap.

Song structure was characterized at the level of units, which are defined by Payne and McVay, 1971 [4] as "the shortest sound that is continuous when evaluated at real speed". The units were transcribed and labeled manually into letters (each different unit is a different letter). The string of consecutive letters represents the transcription of each recording.

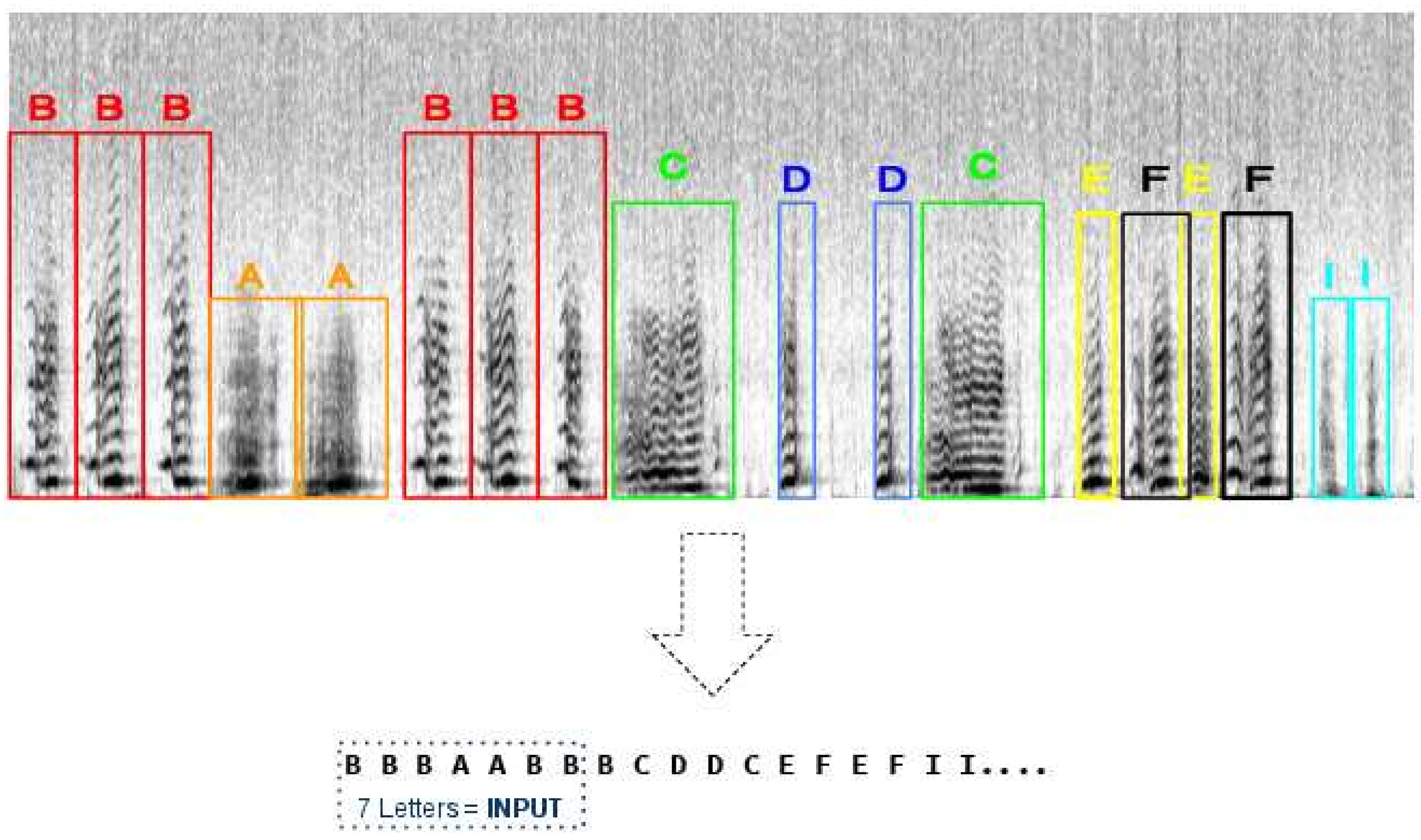

Figure 2: Example of manual transcription of whales songs used in this work.

The Levenshtein matrix representation

Our method to visualize the information contained in a recording is the computation of a distance matrix. We use the Levenshtein distance between two strings (Levenshtein, 1965 [3]) as previously proposed by Garland et al., 2013 [2], but we apply it on letters rather than any other level of structure. The Levenshtein distance between two strings computes the minimum number of insertions, deletions and substitutions necessary to transform one string into the other.

We perform an auto-correlation matrix using Levenshtein distance between every n-letters parts of the recording. n was set to $n = 7$ since in our data the information is basically the same when n varies from 4 to 20 (see figure 3).

The left image of figure 3 shows obvious structures in the recording. On the diagonal of the (symmetric) matrix, squares represents auto-similar structures. Rectangles out of the diagonal show whether theses structures correlates with each other. The high contrast of this representation is due to sparsity of input representation (letters) added to the efficiency of the distance operator.

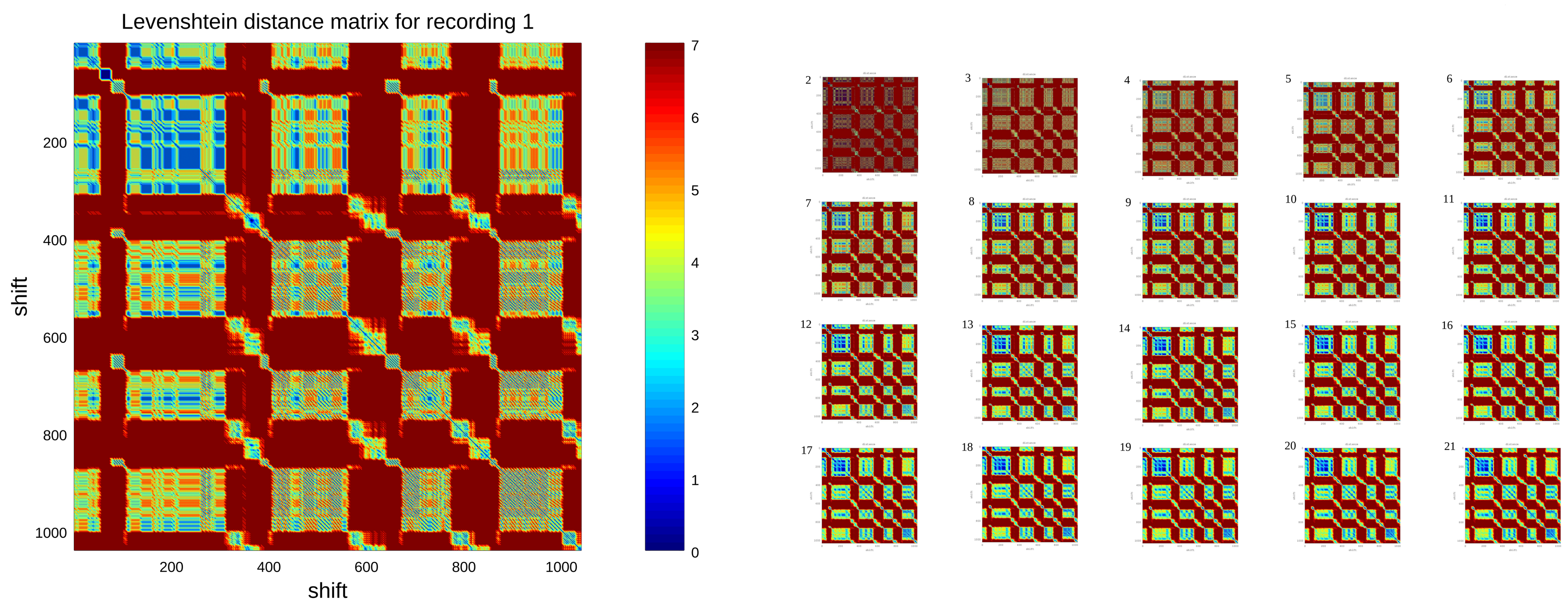

Figure 3: Left : the Levenshtein distance matrix on the Abrolhos 2000 recording for 7-letters scale of correlation. Right : the matrix when the size of the correlation changes from 2 to 21 letters.

Two levels of evident structures

On figure 4, we annotated the two different scales of structures in several recordings. First, a global pattern can be seen that is reproduced periodically. We thus define a 'song' as **the largest structure that can be found in a recording**. The song is repeated with a large amount of similarity, apparent in the non-zero correlation with other songs. This structure is obvious in our representation, it is also in accordance with the general literature on humpback whale bioacoustics (Cholewiak et al., 2013 [1]).

The second scale of structure that is visible in the distance matrix representation is the theme-level. Here we define it as the second structure, in scale, that can be found in the recording. Inside a song, the themes are not usually repeated (theme 2 does not correlate with theme 1).

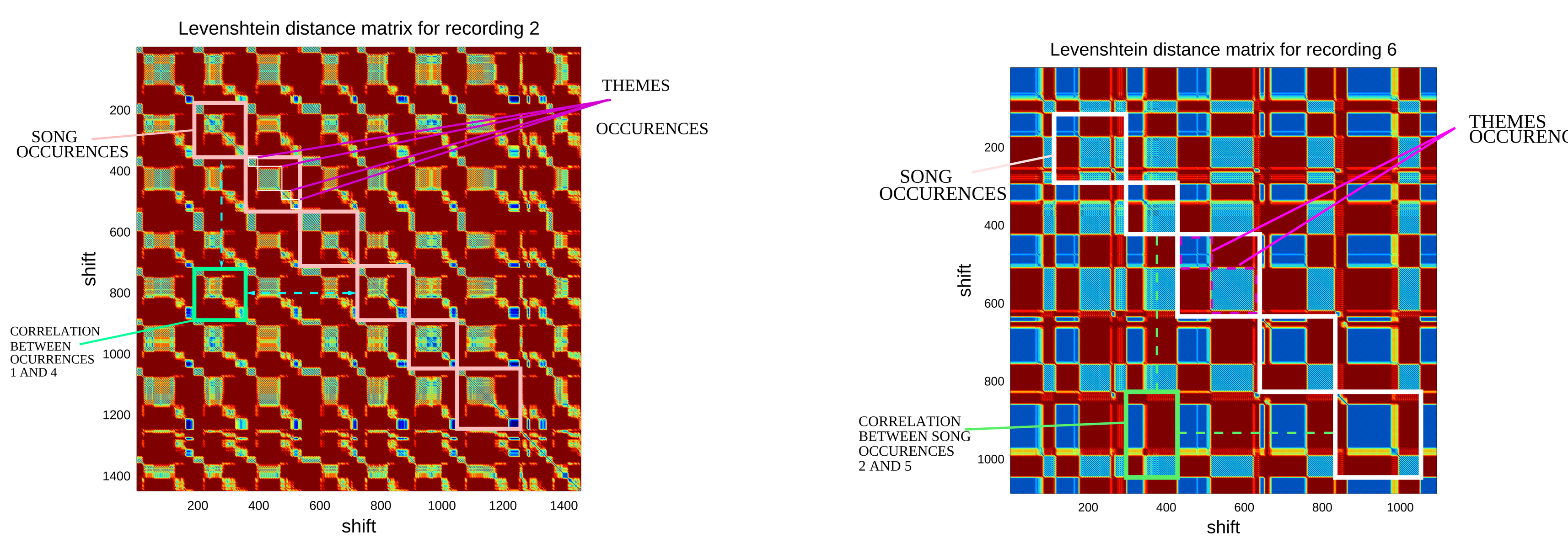

Figure 4: Visual interpretation of the Levenshtein distance matrix representation. On two examples (`left` : Abrolhos 2000 recording 2 and `right`: recording 6, Costa Rica, 2016), songs are highlighted as structures reproduced along the recording, and themes as the components of songs. Far from the diagonal, the rectangles show the level of similarity from one song to the other. .

Song and Themes automatic extraction

Based on the above we wrote a quick routine that can separate songs and themes in a fully automatic way (figure 5). This is a step towards automatic treatment of large scale recordings. Automatic tools to transcribe sounds into a string of letters such as the one developed in Razik et al. 2015 [5] could be used beforehand, and then our method could be applied to extract, classify and compare the themes and songs.

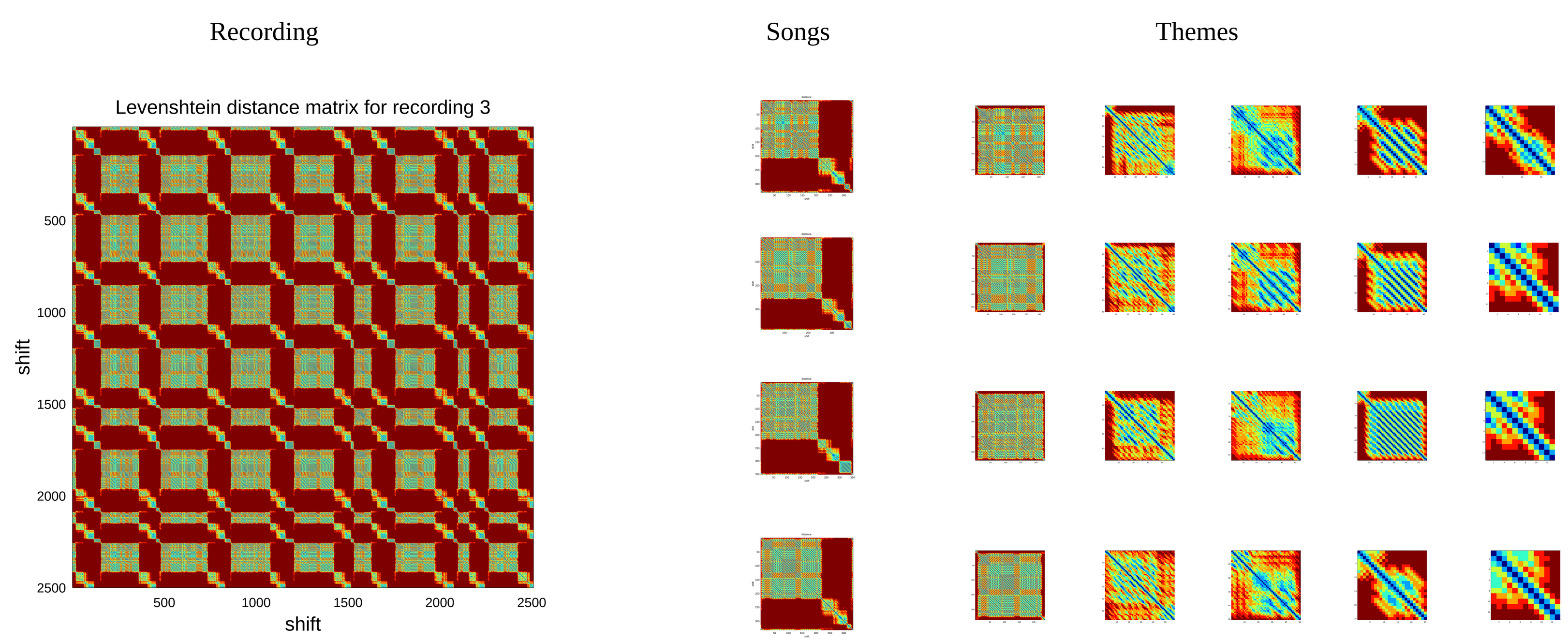

Figure 5: A first example (recording 3, Abrolhos 2001) of automatic extraction of songs and themes. Only the first 4 songs and their themes are represented. .

References

[1] D.M. Cholewiak, R.S. Sousa-Lima, and S. Cerchio. Humpback whale song hierarchical structure: Historical context and discussion of current classification issues. *Marine Mammal Science*, 29(3):312–332, 2013.

[2] E.C. Garland, M.J. Noad, A.W. Goldizen, M.S. Lilley, M.L. Rekdahl, C. Garrigue, R. Constantine, N. Daeschler Hauser, M. Poole, and Jooke Robbins. Quantifying humpback whale song sequences to understand the dynamics of song exchange at the ocean basin scale. *J. Acoust. Soc. Am*, 133(1):560–569, 2013.

[3] V. I. Levenshtein. Binary codes capable of correcting deletions, insertions, and reversals. *Doklady Akademii Nauk SSSR*, 163(4):845–848, 1965.

[4] R. S. Payne and S. McVay. Songs of humpback whales. *Science*, 173:585–597, 1971.

[5] Joseph Razik, Hervé Glotin, Maia Hoeberechts, Yann Doh, and Sébastien Paris. Sparse coding for efficient bioacoustic data mining: Preliminary application to analysis of whale songs. In *IEEE International Conference on Data Mining (IEEE ICDM) workshop Env. Acoustic Data Mining Forum*, page 7, Atlantic City, NJ, USA, nov 2015.