Abstract—This work addresses the study of sounds produced by odontocetes using a brand-new approach based on signal modality. Characterizing the mechanisms used in sound generation, as well as classifying the repertory of sounds, is a complicated task given the complexity of the marine environment and the huge number of different marine mammal species. Several works in the last years have been raised to clarify how these sounds are generated. Understanding the behaviour of the organs responsible for the production and the mechanisms used in the generation is essential for the proposal of new classification features related to the underlying physics.

This complex panorama may gain leverage from the advances in signal processing to the design of new algorithms based on the nature of the signals. Reconstructing the phase space, obtaining Recurrence Plots (RPs) and quantifying the complexity of the resulting structures are processed used with the aim of discerning and characterizing the different sounds produced by the odontocetes. Specifically between sounds with strong (almost-)periodic components, where traditional algorithms of signal processing are unable to get information related to the corresponding underlying model.

I. INTRODUCTION

The way in which marine mammals, particularly odontocetes cetaceans, produce their sounds or vocalizations is a very complex field of study. The evolution has endowed them with extremely sophisticated organs adapted for producing underwater sounds. Knowledge about these organs and their functioning is a key factor in understanding the repertoire of sounds they are able to produce and what these sounds are used for. Characterizing the sounds to make a correct differentiation can shed some new light on the physiology of their sound production organs.

The sound production system of the odontocetes is composed of two air sacks linked by two nasal cavities [1], [2]. Within the cavities there are two pairs of phonic lips and their corresponding dorsal bursa membranes (Monkey-Lips-Dorsal-Bursa, MLDB). The air sacks placed at the extremes of the nasal cavities are in charge of moving the air from one sack to the other. There are sounds generated just by the resonance of the volume of air in the nasal cavities [3], [4]. Also, there are sounds generated by the vibrations of the MLDB, caused by air flowing through the nasal cavities [5], [6]. Therefore, the odontocetes take advantage of the same production system to generate two different kinds of sounds: resonant sounds and vibratory sounds.

Figure 1 compares two examples of the aforementioned types of sounds in order to understand the differences in the temporal and frequency domain. The quasiperiodic signal plotted in Figure 1a was generated by a resonance, and its spectrum shows the presence of some other harmonics apart from the resonant frequency (see Figure 1b). A signal generated by vibrations usually presents concatenated pulses in the temporal domain (Figure 1c), which corresponds to a large number of harmonics in the frequency domain (see Figure 1d).

Human voice modelling techniques begin with a first classification step based on the sound nature: voiced and unvoiced sounds. At the articulatory level, voiced sounds are those produced by the vibrations of the vocal cords (vocals), and the unvoiced ones those in which they do not vibrate (consonants). At the signal-processing level, Pitch Detection Algorithms (PDAs) [7] detect the presence or absence of a fundamental frequency (also named pitch), voiced and unvoiced sounds, respectively. These two types of sounds are easily identifiable in the spectrogram due to the different natures: quasiperiodic signals for voiced sounds (deterministic nature) and stochastic processes (random nature) in the case of unvoiced sounds (Table I).

In an analogous way to the human voice, the modeling of the vocalizations of the odontocetes requires a first classification based on the mechanism of production. More specifically, a classification between two deterministic sounds is proposed, each generated in a different way, either by a vibratory element or by a resonant tube, a more complex and particular scenario where PDAs do not work correctly. Traditional signal-processing PDAs try to obtain statistical characteristics or parameters from the spectrogram of the signal such as the level of the fundamental harmonic, the number of harmonics, etc. [8] without focusing on the underlying model of its generation. The complex speech apparatus of odontocetes requires the development of new algorithms able to distinguish between resonant and vibratory sounds, both of them rich in the number of harmonics but coming from different production mechanisms, and both deterministic (Table I). Analyzing the complexity of the dynamical system of the observed time series may reveal information about the production mechanism.

1The characteristic of determinism will be used throughout the text for those phenomena predetermined by the conditions in which they occur, and therefore are not free, but are necessarily preset.
from which the signal comes: a vibratory element or a resonant tube.

The remainder of this paper is structured as follows: Section II describes the phase space concept and the Recurrence Plots (RPs), the basis for the later Recurrence Quantification Analysis (RQA); Section III presents the results and the chosen features for the raised classification process. Finally, Section IV summarizes the conclusions of the work.

II. SIGNAL MODALITY APPROACH

The characterization of the signal modality is an emerging and interdisciplinary field that tries to address the problem of detecting the presence of underlying mechanisms of random/deterministic, linear/nonlinear generation in a given signal.

A. Phase Space Reconstruction (PSR)

The modality time series analysis discussed in this work is based on the reconstruction of the underlying dynamical system by means of the phase space concept: a vector space which collects all the possible system states useful for determining the future evolution of a signal [9]. For a time series $x(n)$, the phase space would be defined by

$$X_n = [x(n), x(n - \tau), \ldots, x(n - (E - 1) \cdot \tau)]^T,$$

where $N$ is the total number of points, $\tau$ is the discrete time lag, $E$ is the embedding dimension and $T$ refers to the transpose matrix [10]. The proper selection of $\tau$ and $E$ is crucial in the further analysis because it affects the correct representation of the data evolution in time. A common approach to determining the value of $\tau$ is the one proposed by Fraser and Swinney [11] that uses the first null of the time delayed mutual information. The selection of the minimum embedding dimension $E$ is based on the false nearest neighbor (FNN) algorithm proposed by Cao [12]. In this work, the value of embedding dimension $E$ is chosen as the fraction of false points smaller than $2\%$. Any other methods for the estimation of $E$ and $\tau$ might be used and the results are not strongly influenced if both parameters are properly computed [13].
B. Recurrence Plots (RPs)

In 1987, Eckmann et al. introduced a tool called Recurrence Plots (RPs) to visualize the recurrence of states which conform the phase space of a signal $x(n)$ [14]. Among the different variations of computing the RP, the most common way is using Equation (2).

$$R_{i,j} = \Theta(\varepsilon - \| \vec{X}_i - \vec{X}_j \|), \quad i,j = 1, \ldots, N_x$$  \hspace{1cm} (2)

where $N_x$ is the number of considered states $\vec{X}_n$, $\varepsilon$ is a threshold distance, $\| \cdot \|$ is the Euclidean distance, and $\Theta(\cdot)$ is the Heaviside step function. In this work, the value of the threshold $\varepsilon$ is chosen to be the 60% of the mean of the delay vectors (DVs) conforming the phase space (Eq. 1).

The RP representation converts the phase space diagram from a $\mathcal{R}^m$ space to a binary $\mathcal{R}^2$ space providing an easier way to analyse complex systems regardless of the embedding dimension $E$. At a glance, the RP visualization allows the identification of the multidimensional trajectories of the phase space, which can be classified as large and small scale patterns. Large-scale patterns can reveal information about homogeneity, periodicity, etc. Small scale patterns (texture) are conformed by single dots, vertical (and horizontal) lines as well as diagonal lines.

The presence of lonely points is due to rare states, i.e. states that are not persistent at any time or with large fluctuations. However, the presence of vertical (or horizontal) lines indicates a period of time where a state does not change or changes very slowly. Finally, the presence of diagonal lines corresponds to a segment of the phase space path which runs parallel to another segment. The length of the diagonal lines is determined by the duration of the evolution of the trajectory of these segments. The number and the duration of the recurrences are the basis of the Recurrence Quantification Analysis (RQA). Table II summarizes some of the measures for RPs, based on the morphology of existing lines: the recurrence rate of the states (RR), Eq. 3; the number of times in which the phase space path is in the same area, Trapping Time (TT), Eq. 4; the degree of determinism (DET), Eq. 7, etc. Some other characteristics of the underlying dynamical systems such as laminar phases [15], unstable periodic orbits [16] can also be quantified by means of RQA.

### III. Results

An extensive database of sounds produced by beluga whales *Delphinapterus Leucas* (odontocetes or toothed whales commonly called “sea canary” due their capacity to produce a huge repertory of sounds) was recorded at the facilities of the Oceanographic of Valencia in a controlled and repeatable experiment. In more than 5 hours of recording (sample rate of 96 KHz), 38 sounds were detected and later supervised by specialized biologists. The sounds were classified as vibratory (19 events) and resonant (19 events) considering this classification as ground truth for checking the results obtained by the RQA metrics.

<table>
<thead>
<tr>
<th>Symbol &amp; Description</th>
<th>Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>RR. Recurrence rate: density of recurrence points.</td>
<td>$\frac{1}{N^2} \sum_{i,j=1}^{N} R_{i,j}$</td>
</tr>
<tr>
<td>TT. Trapping Time: Averaged vertical line length$^a$.</td>
<td>$\frac{\sum_{v=v_{min}}^{N} v \cdot P(v)}{\sum_{v=v_{min}}^{N} P(v)}$</td>
</tr>
<tr>
<td>LAM. Laminarity: Percentage of recurrence points that form vertical lines$^a$.</td>
<td>$\frac{\sum_{v=v_{min}}^{N} v \cdot P(v)}{\sum_{v=1}^{N} v \cdot P(v)}$</td>
</tr>
<tr>
<td>$L_d$. Averaged diagonal line length$^b$.</td>
<td>$\frac{\sum_{l=l_{min}}^{N} l \cdot P(l)}{\sum_{l=1}^{N} P(l)}$</td>
</tr>
<tr>
<td>DET. Determinism: Percentage of recurrence points that form diagonal lines$^a$.</td>
<td>$\frac{\sum_{l=l_{min}}^{N} l \cdot P(l)}{\sum_{i,j=1}^{N} R_{i,j}}$</td>
</tr>
</tbody>
</table>

- $P(v)$ is the histogram of the lengths $v$ of the black diagonal lines, and $v_{min}$ is the minimal length of what should be considered to be a diagonal line (typically, $v_{min} = 2$).
- $P(l)$ is the histogram of the lengths $l$ of the black diagonal lines, and $l_{min}$ is the minimal length of what should be considered to be a diagonal line (typically, $l_{min} = 2$).

### A. Recurrence Quantification Analysis (RQA) results

An example of PSRs and RPs for the two sound types produced by odontocetes can be seen in Figure 2. The complexity of the signals of both phenomena can be observed in the reconstructed phase space (Figure 2a and Figure 2c). It is verified that both types of signal are deterministic dynamical systems, in view of the parallel trajectories followed by the points. The reconstruction parameters used are $E = 6$ and $\tau = 6$, in both cases. An optimal reconstruction step allows a further-detailed analysis. Figure 2b and Figure 2d illustrate the corresponding RPs, respectively. The value of the threshold $\varepsilon$ is chosen to be 60% of the mean delay vectors. It corresponds to a percentage of black points in the RP of approximately 2%.

It must be noted that Figure 2a and Figure 2c are 3-dimensional plots although the embedding dimension is $E = 6$ for graphical reasons. Despite the similarities seen in the phase space, the RPs (Figure 2b and Figure 2d) seem to unfold the complex structures which allow us to distinguish characteristic patterns for each phenomena. The easily-identified diagonal lines are closely related to the deterministic behaviour of both phenomena (refer to Table I). However, the vertical lines have a different distribution for each production mechanism. In the case of the resonant sounds, the vertical structures are closely related to the number points that correspond to a single state (the speed of the deterministic behaviour). On the other hand,
Fig. 2. a) Phase space reconstruction of a resonant sound (E=6, \( \tau = 6 \)). b) Recurrence plot of a resonant sound. c) Phase space reconstruction of a vibratory sound. b) Recurrence plot of a vibratory sound.

TABLE III
RESULTS OF THE RQA MEASURES FOR ODONTOCETES SOUNDS: RESONANT (RES.) AND VIBRATORY (VIB.) SOUNDS.

<table>
<thead>
<tr>
<th></th>
<th>( \text{RR} )</th>
<th>( \text{TT} )</th>
<th>( L_d )</th>
<th>( \text{DET} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Res.</td>
<td>0.13±0.04</td>
<td>5.77±1.67</td>
<td>0.99±0.01</td>
<td>19.54±15.20</td>
</tr>
<tr>
<td>Vib.</td>
<td>0.08±0.04</td>
<td>7.67±2.92</td>
<td>0.97±0.03</td>
<td>9.85±3.28</td>
</tr>
</tbody>
</table>

The vertical lines of the resonant sounds depend not only on the deterministic effect but also on particular structures, due to the harmonic richness of the studied phenomena and the different initial phases between them.

Although there are significant visual differences, the traditional RQA measures seen in Table II are not representative features for classification purposes. Table III summarizes the obtained results for the studied vibratory and resonant events.

B. Classification

Due to the predominant deterministic behaviour of both studied phenomena, the RQA measure computed over the black points recurrences cannot distinguish between the different production mechanisms. For classification purposes, it is necessary to study the statistics related to the non-neighbour points (white lines in the RPs). Figure 3 shows an example of the typical histograms obtained for each phenomena. The statistics of the diagonal lines are still quite similar due to deterministic behaviour, although the resonant sounds present some longer lines. However, the shape of the vertical lines distributions are completely different due to the aforementioned texture of the RPs. In order to quantify the complexity of the histograms, the Shannon entropy of the vertical and diagonal white lines is computed following Eq. 8 and 9, respectively.

\[
\text{ENTRWVL} = \sum_{v=v_{\min}}^{v_{\max}} -P(v) \cdot \ln(P(v))
\]

\[
\text{ENTRWDL} = \sum_{l=l_{\min}}^{l_{\max}} -P(l) \cdot \ln(P(l))
\]

The measures of entropy easily quantify the homogeneity/heterogeneity of the detected sound under study and classify it as a resonant phenomena or a vibratory one, respectively. Figure 4 compares the obtained values for the analysed database. The resonant sounds are classified as lower complexity events both in the diagonal and vertical dimensions. The vibratory sounds presents higher degree of complexity. Using these features as inputs for any simple classifier achieves good results.

IV. CONCLUSIONS

A new paradigm when modeling odontocetes vocalizations has been raised and studied: analysing the complexity of the underlying dynamical system of the observed time series reveals information about the production mechanism from which the signal comes. Under the hypothesis of two production
The white vertical lines of a vibratory sound. d) Length $l$ of the white diagonal lines of a resonant sound. c) Length $v$ of the white vertical lines of a vibratory sound. d) Length $v$ of the white vertical lines of a resonant sound.

**Fig. 3.** Histogram of the: a) Length $l$ of the white diagonal lines of a vibratory sound. b) Length $l$ of the white diagonal lines of a resonant sound. c) Length $v$ of the white vertical lines of a vibratory sound. d) Length $v$ of the white vertical lines of a resonant sound.

The reconstructed dynamical systems have shown similar properties in the phase space, where the paths mostly run parallel due to the predominant deterministic behaviour of the underlying production mechanisms. The computed RPs have unfolded the hidden trajectories, boosting the differences between a vibratory sound and a sound conforming by a sum of resonant frequencies. This complexity feature has been quantified by means of the RQA of the non-neighbour-point structures. The neighbour points have also been studied, however, the results did not warrant classification. The non-neighbour points of a resonant sound still have a homogeneous distribution, whereas, in the case of vibrant sounds, those points present a heterogeneous texture. The entropy of the distribution of the white points has been the feature chosen to analyze/classify the detected sounds as generated by a vibratory element or a resonant tube.

The reconstructed dynamical systems have shown similar properties in the phase space, where the paths mostly run parallel due to the predominant deterministic behaviour of the underlying production mechanisms. The computed RPs have unfolded the hidden trajectories, boosting the differences between a vibratory sound and a sound conforming by a sum of resonant frequencies. This complexity feature has been quantified by means of the RQA of the non-neighbour-point structures. The neighbour points have also been studied, however, the results did not warrant classification. The non-neighbour points of a resonant sound still have a homogeneous distribution, whereas, in the case of vibrant sounds, those points present a heterogeneous texture. The entropy of the distribution of the white points has been the feature chosen to get a first approach for classification.

A deep study of the physical production mechanisms and the features related to them are underestimated when designing sophisticated classifiers. This study based on the characterization of the signal modality and the techniques used open a new line of work when modeling any sound production. Further work is required to test the hypothesis of the work in a bigger database, including for different marine mammals.

**ACKNOWLEDGMENT**

This work has been partially supported by the Spanish Administration under Grant BIA2014-55311-C2-2-P, by the Fundación Biodiversidad and by European Commision DG Environment (QUIETMED).

**REFERENCES**


