Applied Acoustics 71 (2010) 771–776

Contents lists available at ScienceDirect

Applied Acoustics

journal homepage: www.elsevier.com/locate/apacoust

Real time automatic detection of *Orcinus orca* vocalizations in a controlled environment

Jonas Philipp Luke [a,*], José G. Marichal-Hernández [a], Fernando Rosa [a], Javier Almunia [b]

[a] Dto. Física Fundamental y Experimental, Electrónica y Sistemas. Universidad de La Laguna, Astrofísico F^{co} Sánchez s/n, La Laguna, Spain
[b] Loro Parque Fundación, Puerto de La Cruz, Spain

ARTICLE INFO

Article history:
Received 27 November 2008
Received in revised form 9 April 2010
Accepted 12 April 2010
Available online 13 May 2010

Keywords:
Bioacoustics
Signal detection
Real time
Zero crossing rate
Orcinus orca

ABSTRACT

The use of passive acoustic observation is a useful tool in bioacoustic studies of cetaceans, such as killer whales (*Orcinus orca*). Such studies require the recording, detection, and classification of vocalizations of animals during long periods of time. The manual processing of the recordings is an extremely time consuming task because of the large amount of data to process. Automatic detection and classification techniques are useful to improve the processing, increase the amount of information and, as a consequence, provide information for the conservation of these species.

The *Orca Ocean* facilities of Loro Parque in the Canary Islands, Spain, were used as an experimental platform for developing devices to perform bioacoustic studies. Detection methods with low computational complexity were tested in order to capture vocalizations of four *O. orca* specimens in real time. The algorithms were also tested in other scenarios in order to determine their global performance. The sensitivity to noise was also tested. The most accurate method in this study was implemented and integrated with a continuous recording system generating an event database in real time, downsizing the storage demand to 7%. This allowed the storage of all the sound events produced in about one month on a standard computer harddisk and the generation of basic statistics on vocal activity of the animals.

1. Introduction

Any research on bioacoustics involves the recording, detection and classification of sound events produced by animals. These tasks have been performed manually for many years. Since thousands of vocalizations have to be analyzed it would be useful, and interesting, to automate this process. This gives the chance to process greater amounts of data and detect patterns unnoticed by human observers.

Many researchers have described that killer whales communicate using a limited number of stereotyped vocalizations, referred to as dialects. The structure of these dialects is different between populations and also differs between family groups within the same population [1,2]. The study of such complex vocal behavior will benefit from the automatic compilation of bioacoustic information. The information obtained is interesting for behavioral scientists and for conservation management [3,4].

Like for other cetaceans, killer whale vocalizations are classified in three categories: Clicks, whistles and calls. Clicks are very short pulses typically given in series. The frequency content can be relatively narrow to broad, with emphases peaking as high as 85 kHz. The duration of clicks ranges from 0.1 to 25 ms [5]. Whistles are characterized by a single narrow-band tone with little or no harmonic structure. Their frequencies range between 1.5 and 18 kHz. The duration of whistles ranges from 50 ms to 10–12 s [5]. Calls are the most spectrally rich category. They are harmonically structured and consist of a low frequency component whose fundamental frequency ranges between 80 and 2400 Hz. Many call types also contain a high frequency component with the fundamental frequency ranging between 2 and 12 kHz [6]. Both components contain several harmonics, so that frequency content may reach very high frequencies. Multiple temporal components can be found in calls and their duration ranges between 0.5 and 1.5 s [5].

The development of automatic techniques for the processing of these vocalizations requires the solving of three main issues: Recording, detection and classification. For this purpose an experimental platform, consisting of the necessary hardware and software, was created at the facilities of Loro Parque in the Canary Islands, Spain.

This paper focuses on automatic signal detection and signal storage techniques which are the step previous to the classification system. Taking into account that in a continuous recording system the detector should work in real time, some signal detection techniques with low computational complexity were compared, and their sensitivity to noise was studied. First, in Section 2, related

* Corresponding author.
E-mail addresses: jpluke@ull.es (J.P. Luke), jmariher@ull.es (J.G. Marichal-Hernández), frosa@ull.es (F. Rosa), adjunto@loroparque-fundacion.org (J. Almunia).

0003-682X/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.apacoust.2010.04.003

works are described. In Section 3, materials and methods are presented. Section 4 explains the results of the comparative study, and in Section 5 the integration with an event database is described. Finally, in Section 6, some concluding remarks are given.

2. Related works

The detection of vocalizations can be performed manually, either listening to sounds or observing spectrograms. However, the amount of data to be processed favors the use of automatic detection systems. As a consequence, several methods to detect marine mammal vocalizations have been developed and tested in recent decades. The algorithms can operate either in time domain or on time–frequency distributions of the signal (e.g. spectrogram). The proposed automatic detection techniques go from simple energy thresholds to more elaborate methods like cross-correlation [7], image processing techniques on spectrograms, wavelet decompositions [8], bayesian inference [9] or neural networks [10].

The processing time is not a critical point in offline computation of datasets. Some previous work has been done on killer whales with offline processing [11]. However, a continuous vocal behavior tracking system needs real time processing, because processing time becomes critical. As a consequence, techniques with low computational complexity have to be used or the techniques have to be implemented on faster specific hardware.

3. Materials and methods

3.1. The Orca Ocean facilities

The recording of killer whale vocalizations and the testing of detection techniques was done in the *Orca Ocean* pools at Loro Parque. Currently, these facilities are used as an experimental platform to develop bioacoustic signal processing devices. The facilities consist of three pools with four *O. orca* individuals, two males and two females.

3.2. Recording system

The recording system is supported by eight built-in hydrophones strategically located in the pools. The hydropones are connected to a personal computer supported on the Debian GNU/Linux 4.0 operating system and provided with a PowerDAQ acquisition card from United Electronics. This system allows the play back of one of the channels through a loud speaker in order to hear in situ what is happening in the pools. It allows continuous recording of all eight channels at 200 kSamples/s providing a maximum detectable frequency of 100 kHz. Taking into account the spectral content of orca vocalizations discussed in Section 1, this sampling frequency seems to be enough to avoid the aliasing effect. Another reason to choose this sampling rate was that source location techniques will also be tested in the pools and the spatial resolution of such techniques depends on the sampling frequency.

3.3. Detection methods

Signal detection consists of the discrimination between signal plus noise and background noise intervals. More formally, the detection problem can be formulated as a simple exclusive binary test between null hypothesis H_0 (absence of signal) and alternative hypothesis H_1 (presence of signal) as shown in Eq. (1).

$$\begin{aligned} H_0: &\quad x(n) = \text{noise}(n) \\ H_1: &\quad x(n) = \text{signal}(n) + \text{noise}(n) \end{aligned} \tag{1}$$

where x is an observation process, in this case the underwater recording. It can be composed of only background noise under hypothesis H_0 or signal (*orca* vocalizations, anthropogenic noise, water splashes, ...) plus background noise under hypothesis H_1. Using the observation process some kind of evidence variable has to be computed. The evidence variable is then compared with a threshold to decide between the null and the alternative hypothesis.

The goal is to detect the vocalizations in real time on a continuous data stream in order to create an event database on the fly. For this reason, detection algorithms with low computational complexity were chosen.

Here some functions to compute evidence variables are presented and their performance is evaluated:

- Energy in a sliding window: This consists of computing the average of the square of the samples in a sliding window over the data stream using Eq. (2), where $x(n)$ is the observed value at instant n and N is the window length.

$$E(n) = \frac{1}{N}\sum_{i=-\frac{N}{2}}^{\frac{N}{2}} x(n+i)^2 \tag{2}$$

For faster calculation a circular buffer with N samples is used and the sum is calculated recursively.

- Energy in a sliding window filtering out low frequencies (Filtered Energy for short): The method is the same as the previous, but filtering out the low frequencies of the signal. It can be computed using Eq. (3), where $|X(n,k)|^2$ is the value of the spectrogram at frequency index k, and the data is averaged over a time window of length N centered at time n, as for Eq. (2).

$$E_f(n) = \frac{1}{k_2 - k_1}\sum_{k=k_1}^{k_2} |(n,k)|^2 \tag{3}$$

- Instantaneous frequency: This uses Eq. (4) to compute first spectral moment.

$$IF(n) = \frac{2\pi}{N}\frac{\sum_{k=k_1}^{k_2} k|X(n,k)|}{\sum_{k=k_1}^{k_2} |X(n,k)|} \tag{4}$$

$|X(n,k)|$ is the square root of value of the spectrogram at frequency index k, and the data is averaged over a time window of length N centered at time n, as for Eq. (2). If the signal durations are longer than the window length N, then Eq. (4) can be used to approximate the instantaneous frequency of the signal [12].

- Zero crossing rate in a sliding window filtering out low frequencies: This technique consists of counting the number of zero crossings of the signal in a sliding window of size N. Note that a filtered version of the signal with suppressed low frequencies is used in order to avoid influence of water noise. The computation of this evidence variable is performed following Eq. (5):

$$zcr(n) = \frac{1}{N}\sum_{i=-\frac{N}{2}}^{\frac{N}{2}} \mathbf{I}\{x_f(n+i)x_f(n+i-1)\} \tag{5}$$

where $x_f(n)$ is a filtered version of the observation at instant n and $\mathbf{I}(\cdot)$ is a function that is 1 if the argument is a negative value and 0 otherwise.

3.4. Experiments

In order to compare performance of the evidence variables presented above, ROC analysis was used [13]. This technique is based on depicting the true positive rate (TPR) in relation to the false po-