Expert Systems with Applications 41 (2014) 5451–5465

Contents lists available at ScienceDirect

Expert Systems with Applications

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

An automatic acoustic bat identification system based on the audible spectrum

Aarón Henríquez [a], Jesús B. Alonso [a,*], Carlos M. Travieso [a], Bernal Rodríguez-Herrera [b], Federico Bolaños [b], Priscilla Alpízar [b], Karmele López-de-Ipina [c], Patricia Henríquez [a]

[a] *Instituto para el Desarrollo Tecnológico y la Innovación en Comunicaciones (IDeTIC), Universidad de Las Palmas de Gran Canaria, Despacho D-102, Pabellón B, Ed. de Eletrónica y Comunicaciones, Campus de Tafira, 35017 Las Palmas, Spain*
[b] *Escuela de Biología, Universidad de Costa Rica, Costa Rica*
[c] *Systems Engineering and Automation Department, Universidad del País Vasco/Euskal Herriko Unibertsitatea, Spain*

ARTICLE INFO

Keywords:
Bat identification
Feature extraction
Spectrogram
Audio adaptation

ABSTRACT

Nowadays the task of monitoring bat species is a very difficult task because of several factors. The main ones are the difficulty of creating databases automatically and the particularities of the vocalizations of bats. For this reason, it is common to extract bat calls manually from a recording and treat them individually. We propose a new form of identification and labeling process based on adapting bat calls to the audible spectrum and significantly reducing the noise of its spectrogram. This process can be performed automatically from a recording made in a natural area. Our database consists of 189 h of recordings obtained in various natural areas in Costa Rica. 50 bats calls of 7 different classes are extracted from this database. We have obtained an average error of 2.7% and 3 of the 7 classes have an error below 1%.

1. Introduction

Bats are essential to the conservation of terrestrial ecosystems because they collaborate in many functions such as seed dispersal and pollination, as well as control the populations of their prey, being an indispensable link in the food chain (Galindo 1998). In recent decades, our planet is experiencing an extraordinary climate change, causing a significant decline in habitats worldwide. The bats are present in the vast majority of our planet's ecosystems and are extremely sensitive to climate changes and habitat destruction (Jones, Jacobs, Kunz, Willig, & Racey, 2009). They are, therefore, a natural indicator of the health of ecosystems. For this reason, the researchers have tried to monitor the populations of these animals for decades. At the beginning, due to the limitation of technology, this task was carried out by specimens capturing and manual identification by researchers. Thus, for many species, although the identification is accurate, the task was performed slowly, and ineffectively. An automated system for monitoring populations enables specialists to work much more efficiently, in less time and with less cost (Agranat, 2009).

An automatic identification of bat species would allow biologists to save a lot of time, since currently the acoustic bat identification is done manually (Waters & Barlow, 2013). Biologists could focus on studying the causes of migration and population changes instead of detecting them. Therefore, not only researchers would save time but also they would reduce the costs of the actions they decide to undertake following the information analysis. It opens the door to monitor environmental conditions in real time and in a remotely form; a request of biologists for years.

Most researches work with bat echolocation calls, but there are others that works with images (Azmy et al., 2012). The first works in acoustic identification of bats were at the beginning of the eighties (Skowronski & Harris, 2006). Many authors have concluded that it is possible to reach a species-level identification from echolocation calls (O'Neill, 2004). However, they have identified a number of constraints to be considered so this task can be performed effectively. Firstly, calls should be used only in the search phase (Griffin, 1958), that is echolocation calls in free flight. Secondly, you have to consider all possible variations in the locutions of same species. Depending on the sex (Neuweiler et al., 1987), age (Moss, Redish, Gounden, & Kunz, 1997), habitat or the level of stress of the specimen (Xu et al., 2008), the locutions can present significant variations, both in duration and in frequency and intensity. Thirdly, while taking recordings, bats vary their position from the microphone, that causes variations in the time that successive calls reach recording device and the Doppler Effect might have an influence (Hiryu, Katsura, Lin, Riquimaroux, & Watanabe, 2005). Finally, we

* Corresponding author. Tel.: +34 928452863; fax: +34 928 451243.
E-mail address: jalonso@dsc.ulpgc.es (J.B. Alonso).

http://dx.doi.org/10.1016/j.eswa.2014.02.021
0957-4174/

have to avoid noisy calls where several specimens issued at a time, be reflected or have excessive noise. All these restrictions are a handicap that has avoided obtaining a fully automatic system of recognition of species.

Traditionally, creating an audio repository is performed in a supervised way in order to eliminate or reduce the influence of the error sources discussed (Britzke, 2003). The correct way is to capture the specimens and to record their calls while release. Thus, this way ensures that the record is made in free flight although the specimen can be stressed. Also, age, sex and trajectory of the specimens are known which ensures that all samples are taken under almost identical conditions. The recordings are labeled and classified manually avoiding errors. However, there remains the problem of needing much intervention by the researchers.

As an alternative to the supervised way, the audio repositories can be created from recordings acquired in an unsupervised way in which the bats are recorded in freedom without being influenced by humans. The goal is the obtaining of the audio samples to classify them in the same way they have been done with the capture of individuals. Today, thanks to technological development and treatment techniques it is possible to generate audio recordings repositories automatically.

Nowadays intelligent systems that identify species of bats automatically from recordings of their vocalizations (O'Neill, 2004) are beginning to be used. These systems are usually composed of three steps: parameterization, modeling and classification (Lopez-de-Ipina et al., 2013). The parameterization is to obtain a set of features (measures) of the audio signal to characterize and identify different species of bats. In the modeling process, a mathematical representation is created of each class (family, genus or specie) from a set of audio samples which a priori are known to which species belongs. Finally, given a particular audio sample that corresponds to a class, a priori unknown, it is compared between different models and automatically classified in one of the classes modeled.

In the parameterization, measures from the signal are extracted (the features). In this process it is common to get some physical values such as the maximum and minimum frequency, duration or other time and frequency parameters (Brigham, Kalko, Jones, Parsons, & Limpen, 2002). This process should be made so that the extracted feature helps to identify the class and therefore differentiate it from the rest. If we look at the spectrograms of the locutions of different species, we realize that there are modulated tones that are distinguishable by the physical characteristics mentioned, so it is logical to use in identification. Previous to that, it is common to adapt the signal to facilitate this step. It is normal to reduce noise and make a representation of the audio signal in the frequency domain, typically using the Fourier transform (FFT). As a result, improvements in the process were based on adapting the shape signal to obtain very precise parameters that help to differentiate the classes.

The modeling process consists of creating a mathematical representation of each class. So far, a large number of parametric and nonparametric methods have been used (Britzke, Duchamp, Murray, Swihart, & Robbins, 2011). Among the parametric models, Discriminant Function Analysis (DFA) and Multinomial Logistic Regression (MLR) are the most utilized. Other parametric models that have been successfully employed are the Gaussian Mixture Models (GMM) and Hidden Markov Models (HMM) (Skowronski & Harris, 2006). The most widely used non-parametric model is the Artificial Neural Networks (ANN) (Mirzaei et al., 2011). Different results can be obtained with these models depending on the previous parameterization (Jennings, Parsons, & Pocock, 2008; Redgwell, Szewczak, Joseph, Jones, & Parsons, 2009). Generally, more flexible models perform better (Britzke et al., 2011) when the number of samples in the training process is adequate. Gaussian Mixture Models fulfill these assumptions because it is very flexible and converge fast to the solution.

The classification process consists of deciding which class belongs to the input sample. Traditionally it has been done by comparing the sample of the unknown class with the models of each class, and obtaining a value that represents the similarity of the sample to each class. The most common thing is to understand that the class whose comparison has achieved the highest value of similarity is chosen by the system; however depending on the class to identify may not be adequate.

Table 1 shows the results obtained in other studies; some have obtained a high accuracy due to a detailed parameterization and modeling processes. The method used in the segmentation step is mentioned because the ultimate goal is to automate the entire process so that the system can automatically classify a taken sample. Automatic segmentation consists of extracting samples automatically from a recording, but later a subset with the best quality can be selected. Manually segmentation means that also the samples are "cutted" manually from a recording, that can be monitored or not.

In this study our first aim is to design a new audio adaptation process so that bat locutions remain in the audible band without distortion. The time-expansion and frequency-division conversions do not accomplish this second condition. Time-expansion extends the signal in time at the same time as it compress the spectrum whereas frequency-division loses the information of the harmonics. With the heterodyne conversion, information is lost if the bandwidth is very large, and also not all the bandwidth of the audible band is always available when the position of the characteristic frequency of the locution is not in the middle of the band. However, in many cases, the locution is not distorted so we decided to improve this conversion method to obtain more clean locutions in the audible band.

This new conversion method consists, firstly, of a noise reduction phase, secondly, the modulation phase and finally, resampling. Before modulating, the histogram of the spectrogram is computed, in order to know exactly the center of the band. After the modulation the locution will be centered in the audible band. Nevertheless, if we apply directly the modulation using the characteristic frequency as the modulation frequency, a part of the locution could be lost, especially when the distance between the center frequency and the characteristic frequency is big.

In classical parameterization, the physical parameters are extracted from the spectrogram of the locution. There are also some works that parameterize with information about the shape. We have considered mixing the two methods for two reasons. Firstly, most classes with similar shape differ in frequency and viceversa. Secondly, it is very simple and fast to get all the information from the audible spectrogram obtained using our conversion method. Furthermore, we apply a feature not used so far from the shape curves, such as the kurtosis.

This paper is divided into 6 sections. The second section describes the methods and materials to create the database and the strategy followed by the segmenter and the automatic classification system is showed. The third section presents the results obtained. After that, the discussion section is presented, in which our results are compared with those in the state of the art. The final section presents the conclusions.

2. Materials and methods

2.1. Materials

Our database consists of 189 h of continuous recording made in the original natural areas of Costa Rica, a country with greatest