Expert Systems With Applications 50 (2016) 100–106

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Methodology for automatic bioacoustic classification of anurans based on feature fusion

Juan J. Noda[a,*], Carlos M. Travieso[a,b], David Sánchez-Rodríguez[a,c]

[a] Institute for Technological Development and Innovation in Communications, Spain
[b] Signal and Communications Department, Spain
[c] Telematic Engineering Department, University of Las Palmas de Gran Canaria, Campus Universitario de Tafira S/N, 35017 Las Palmas de Gran Canaria, Spain

A R T I C L E   I N F O

Keywords:
Biological acoustic analysis
Bioacoustic taxonomy identification
Acoustic data fusion
SVM

A B S T R A C T

The automatic recognition of anurans by their calls provides indicators of ecosystem health and habitat quality. This paper presents a new methodology for the acoustic classification of anurans using a fusion of frequency domain features, Mel and Linear Frequency Cepstral Coefficients (MFCCs and LFCCs), with time domain features like entropy and syllable duration through intelligent systems. This methodology has been validated in three databases with a significant number of different species proving the strength of this approach. First, the audio recordings are automatically segmented into syllables which represent different anuran calls. For each syllable, both types of features are computed and evaluated separately as in previous works. In the experiments, a novel data fusion method has been used showing an increase of the classification accuracy which achieves an average of 98.80% ± 2.43 in 41 anuran species from AmphibiaWeb database, 96.90% ± 3.57 in 58 frogs from Cuba and 95.48% ± 4.97 in 100 anurans from southern Brazil and Uruguay; reaching a classification rate of 95.38% ± 5.05 for the aggregate dataset of 199 species.

1. Introduction

Anurans (frogs and toads) are remarkable biological indicators of environmental quality and stress (Beebee & Griffiths, 2005). Hypersensitive to chemical pollution, habitat degradation, pollution of rivers and surface water, climate change or even the sun's ultraviolet radiation (Alford & Richards, 1999; Egea-Serrano, Relyea, Tejedo, & Torralva, 2012), amphibians are one of the most endangered vertebrate groups by human activity, and abundance of wetlands is always one of the best indicators of good environmental conservation. Moreover, amphibian's secretions and toxins have a wide range of potential medical uses (Clarke, 1997) and the topic is currently widely researched.

Animals emit a rich variety of different signals and sounds to communicate for diverse purposes (Owings & Morton, 1998), being certain acoustic signals quite pure in tonal quality. In recent decades, advances in technologies that seek to automate the monitoring of wildlife using remote sensors and automated acoustic identification of species are transforming the way biologists study ecosystems (Gaston & O'Neill, 2004). For this purpose, various pattern recognition methods have been suggested to investigate the sound production of birds (Fagerlund, 2007), insects (Ganchev, Potamitis, & Fakotakis, 2007) and bats (Henríquez et al., 2014) among others. However, a robust machine learning technique to recognize frog and toad calls has still not been found. In anurans, the main goal of vocalization is advertisement (Duellman & Trueb, 1986) which presents unique acoustic properties per specie. Therefore, their calls can be used as an efficient parameter for taxonomic classification and survey.

There is no doubt of the effort made in the automatic acoustic recognition field to enable a reliable classification of species. However, there are few studies in literature that have been focused on amphibians and previous work limits the study to a small number of species, so an improvement in the state of the art is necessary to identify successfully a larger dataset. In this work, a new methodology for anuran sound recognition is proposed by applying a novel fusion technique of frequency domain features: Lineal Frequency Cesptral Coefficients (LFCC) and Mel Frequency Cesptral Coefficients (MFCC), with time domain features: Shannon entropy and call duration; in order to significantly increase the number of species able to be classified with a high success rate. This novel data fusion has been validated in three databases

* Corresponding author. Tel.: +34928624537; fax: +34928634021.
E-mail addresses: jnoda@ingetelca.com, jjnoda@gmail.com (J.J. Noda), carlos.travieso@ulpgc.es (C.M. Travieso), david.sanchez@ulpgc.es (D. Sánchez-Rodríguez).

http://dx.doi.org/10.1016/j.eswa.2015.12.020
0957-4174/

independently, where each one contains more species than that of any previous work. Moreover, to confirm the robustness of this methodology, these databases have been grouped together creating the most extended dataset of anuran calls automatically classify to date, keeping a high degree of success. Finally, three widely used pattern matching techniques: Hidden Markov Model (HMM) (Rabiner, 1989), Random Forest (RF) (Breiman, 2001) and Support Vector Machine (SVM) (Burges, 1998); have also been used to test this approach.

The remainder of this paper is organized as follows. Section 2 presents a review of previous research in this area. Section 3 describes the proposed methodology, the syllable segmentation and the feature extraction process in order to obtain rich information to feed the machine learning classifier. The SVM, RF and HMM classification methods used are described in Section 4 particularized for acoustic recognition. Then, in Section 5 the two databases employed are introduced. Section 6 contains the experimental methodology applied and the results obtained making a comparison of features and classification algorithms. Finally, in Section 7, the conclusions of this work are shown.

2. Related work

Intensive studies have been conducted in the field of bioacoustics classification by employing different features and methods, but only a few have regarded amphibians through intelligent systems. Taylor in (Grigg, Taylor, Mc Callum, & Watson, 1996) studied 22 frog species from the North Australia using features such as the peaks of the signal spectrogram and their frequencies to train a Decision Tree (DT) built with the C4.5 algorithm. However, this method was incapable of distinguishing all species and the process resulted time consuming. Lee, Chou, Han, and Huang (2006) studied 30 frog species and 19 cricket calls. They divided input signals into frames, calculated the averaged MFCCs and applied Linear Discriminant Analysis (LDA) for classification. Their work obtained a recognition rate of 96.8% over the frogs database but with a higher standard deviation. The average features on frames lose non-stationary information and it becomes difficult to recognize species with the same call frequencies. Brandes (2008) applied HMM on 9 bird, 10 frog and 8 cricket sound samples with an accuracy of 84.82%, 89.48% and 89.73%, respectively. This approach used peak frequencies and bandwidth from the spectrogram to parametrize vocalizations, though it has trouble dealing with broad band calls where frequency limits are not clear. Alternatively, Huang, Yang, Yang, and Chen (2009) calculated the spectral centroid, signal bandwidth and threshold-crossing rate as parameters of 5 anurans belonging to the Microhylidae family. Then, they employed k-Nearest Neighbor (KNN) and SVM for identification gaining just an 89.05 and 90.30% accuracy, in each classifier. In Acevedo, Corrada-Bravo, Corrada-Bravo, Villanueva-Rivera, and Aide (2009), a set of classification algorithms, SVM, DT and LDA were compared on 9 frog and 3 bird species. Their research used as features the call length duration, maximum and minimum frequencies in the spectrogram, maximum power and the frequency of maximum power in 8 segments of the call. In their work, SVM results outperform DT and LDA, achieving an accuracy of 94.95%.

On the other hand, Han, Muniandy, and Dayou (2011) presented a new method for animal sound identification combining Shannon, Rényi and Tsallis entropies using KNN for recognition. As a result, 7 frog species were successfully classified with 100% success, but two more couldn't be recognized properly due to their entropy features were similar. Another interesting approach can be found in Chen, Chen, Lin, Chen, and Lin (2012), where the authors applied a template based method to recognize 18 frog species with a identification rate of 94.3%, analyzing the length of the segmented syllables and applying a Multi Stage Average Spectrum (MSAS) method. However, it required a pre-classification stage because some anuran calls have similar syllable length. Yuan and Ramli (2013) introduced a recognition method based on MFCC and Linear Predictive Coding (LPC) with KNN to automate the identification of 8 frog specimens selected from the Internet database (AmphibiaWeb, 2015), obtaining a classification accuracy of 98.1%. In present research, this data collection of sound recordings has also been employed selecting 41 anuran species including those used in Yuan and Ramli (2013). Later, authors in Jaafar, Ramli, Rosdi, and Shahrudin (2014) made another interesting comparative study on two databases with 13 and 15 frog species respectively, employing MFCC coefficients to train three classifiers: SVM, Sparse Representation Classifier (SRC) and Local Mean KNN with Fuzzy Distance Weighting (LMkNN-FDW). The experimental results of LMkNN-FDW provided the best result with 98.4% on the first database but only 87.2% on the second, due to calls could not be characterized successfully, with some species below 50%. A more modern approach can be found in Bedoya, Isaza, Daza, and López (2014), where the authors used a fuzzy cluster classifier LAMDA (Aguilar-Martin & López de Mántaras, 1982) and MFCC on 13 anuran species from Colombia divided into two datasets by which they obtained accuracies between 99.38 and 100%. It was possible due to the small number of species in each dataset presented clearly distinguishable pitch frequencies. Recently, Xie et al. (2015) classified 16 Australian anurans by combining various acoustic parameters: dominant frequency, syllable duration, frequency modulation, oscillation rate and energy modulation. Then, Principal Component Analysis (PCA) and KNN were utilized for taxonomy cataloguing reaching only 90.5% success. Finally, in Colonna, Cristo, and Salvatierra (2015), an incremental segmentation technique was evaluated over 7 frog species, increasing the recognition by 37% with respect to sliding window approaches. However, they used KNN with $k = 1$ for classification which can lead to over-fitting.

It is not easy to find references on this topic. The literature is sparse and much of the previous work has limited studies to less than 20 species. In addition, most of the works cited are only based on temporal or frequency domain information but not both. In this paper, an intensive study has been conducted regarding the bioacoustics characteristics of anurans, over 199 species, enabling a broad range of identification. Furthermore, frequency and temporal acoustic attributes have been analyzed to seek the most discriminating features, fusing them to develop an effective classification system.

3. Proposed methodology

Anurans' call recordings are automatically segmented in syllables and grouped into sample sets by specie. Then, the feature parameters are extracted from each syllable and are fused into a single vector of characteristics per syllable. Afterwards, they are used to train a classification algorithm. In this paper, we have compared the results of three machine learning algorithms HMM, RF and SVM. Fig. 1 illustrates the proposed system technique.

3.1. Segmentation

The segmentation stage splits the file recordings into as many syllables as possible to yield useful information for the taxonomy identification. The syllable segmentation is obtained applying the algorithm proposed by Härmä in Härmä (2003). Härmä employed Short Time Fourier Transform (STFT) to obtain the spectrogram of the input signal and divided it into a set of N syllables by exploring the maximum amplitude peaks. In this work, the algorithm begins computing the STFT using a Hamming window of 512 samples and overlap of 25%. The window size and overlap have been selected considering the anurans' calls dominant frequency ranges and the