Expert Systems With Applications 72 (2017) 83–92

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Automatic anuran identification using noise removal and audio activity detection

Jesús B. Alonso[a,*], Josué Cabrera[a], Rohit Shyamnani[a], Carlos M. Travieso[a], Federico Bolaños[b], Adrián García[b,1], Alexander Villegas[b], Mark Wainwright[c]

[a] Instituto Universitario para el Desarrollo Tecnológico y la Innovación en Comunicaciones (IDeTIC), Universidad de Las Palmas de Gran Canaria, 35017, Las Palmas de Gran Canaria, Spain
[b] Escuela de Biología, Universidad de Costa Rica, San Pedro de Montes de Oca, San José, 2060, Costa Rica
[c] La Casa del Hueco, Monteverde, Puntarenas, Costa Rica

A R T I C L E I N F O

Article history:
Received 23 December 2015
Revised 11 October 2016
Accepted 10 December 2016
Available online 13 December 2016

Keywords:
Bioacoustical identification
Biodiversity monitoring
Species richness
Ecological indices
Environmental audio
Frog identification

A B S T R A C T

The use of bioacoustics to identify animal species has huge potential for use in biology and conservation research. Fields that could be greatly enhanced by the use of bioacoustical techniques include the study of animal behavior, soundscape ecology, species diversity assessments, and long-term monitoring - for example to further our understanding of the conservation status of numerous species and their vulnerability to different threats. In this study, we focus primarily, but not exclusively, on the identification of anuran vocalizations. We chose anurans both because they tend to be quite vocal and because they are considered indicators of environmental health. We present a system for semi-automated segmentation of anuran calls, based on sound enhancement method that uses Minimum-Mean Square Error (MMSE) Short-Time Spectral Amplitude (STSA) estimator and noise suppression algorithm using Spectral Subtraction (SS), and an automated classification system for 17 anuran species based on Mel-Frequency Cepstrum Coefficients (MFCC) and the Gaussian Mixture Model (GMM). To our knowledge this is the first study that applies this combination of methods for animal identification. This technique achieves accuracies of between 96.1% and 100% per species. Experimental results show that the semi-automated segmentation technique performs better than automated segmentation systems, improving the average success rate to 98.61%. The effectiveness of the proposed anuran identification system in natural environment is thus verified. This work presents a first approach to future tools which can signify a significant advance in the procedures to analysis in a semiautomatic or even in an automatic way to analysis indicators of environmental health based on expert and intelligent systems.

1. Introduction

Evaluating the destruction and degradation of the environment through, for example, the impacts of climate change, agriculture, and other human activities has become a central and urgent task in conservation biology. To this end, biologists are trying to generate a better understanding of factors such as species richness distribution, changes in ecosystem composition, presence or absence of indicator species, shifts in animal migration patterns, and population dynamic of rare or endangered species (Bedoya, Isaza, Daza, & López, 2014; Chen & Li, 2013; Chen, Chen, Lin, Chen, & Lin, 2012; Han, Muniandy, & Dayou, 2011; Lee, Hsu, Shih, & Chou, 2013; Potamitis, Ntalampiras, Jahn, & Riede, 2014; Ventura et al., 2015; Wagner, Züghart, Mingo, & Lötters, 2014).

Bioacoustics is the science of the animal communication and associated behavior through acoustic signals. Many animals communicate acoustically. The sounds produced by many birds, frogs, bats, and insects, for example, contain species-specific features that facilitate communication and the recognition of conspecifics (Brumm & Slabbekoorn, 2005; Cheng, Sun, & Ji, 2010; Lee, Lee, & Huang, 2006). Acoustic recordings of the environment play an increasingly important role in the biodiversity monitoring and soundscape ecology of terrestrial and aquatic ecosystems in

* Corresponding author. Universidad de Las Palmas de Gran Canaria, Campus de Tafira, Ed. de Electrónica y Telecomunicaciones, Pabellón B, Despacho D-102, 35017, Las Palmas de Gran Canaria, Las Palmas, Spain.

E-mail addresses: jesus.alonso@ulpgc.es, jalonso@dsc.ulpgc.es (J.B. Alonso), jcabrera@idetic.eu (J. Cabrera), rohit.vashdev@gmail.com (R. Shyamnani), carlos.travieso@ulpgc.es (C.M. Travieso), federico.bolanos@ucr.ac.cr (F. Bolaños), garciar.adrian@gmail.com (A. García), alexmont@costarricense.cr (A. Villegas), markwainwright@racsa.co.cr (M. Wainwright).

[1] Departamento de Ecología, Universidade Federal do Rio Grande do Norte, Natal, RN, 59078–900, Brazil

http://dx.doi.org/10.1016/j.eswa.2016.12.019
0957-4174/

relation to human activities (Fagerlund, 2007; Jaafar, Ramli, & Shahrudin, 2013; Lee et al., 2013; Towsey et al., 2014b; Towsey, Wimmer, Williamson, & Roe, 2014a). Research of animal sounds has importance in a variety of fields, such as population monitoring (Bardeli et al., 2010; Bedoya et al., 2014; Hödl, 1977; Juang & Chen, 2007; Patti & Williamson, 2013; Potamitis et al., 2014), migration monitoring (Härmä, 2003; Juang & Chen, 2007), environmental monitoring (Chen & Li, 2013; Jaafar & Ramli, 2013; Lee et al., 2006), the study of animal behavior (Bardeli et al., 2010; Hödl, 1977; Patti & Williamson, 2013; Zsebők, Czabán, Farkas, Siemers, & von Merten, 2015), prevention of harmful human/animal interactions (Bardeli et al., 2010; Hödl, 1977; Patti & Williamson, 2013). The field of bioacoustics has also been used in ornithological (Bardeli et al., 2010; Chen & Li, 2013; Hödl, 1977; Jaafar & Ramli, 2013; Juang & Chen, 2007; Patti & Williamson, 2013; Wielgat, Zieliński, Potempa, Lisowska-Lis, & Król, 2007), and agricultural studies (Wielgat et al., 2007). Moreover, it can be used for educational and pedagogical purposes (Huang, Yang, Yang, & Chen, 2009; Wielgat et al., 2007), helping to avoid bird strikes with airplanes (Juang & Chen, 2007) and for recreation by non-professional birdwatchers and other naturalists (Juang & Chen, 2007).

Traditional monitoring protocols consisting of repeated site visits for several minutes at certain times of the day (i.e. each morning, noon and dusk) over several days (Ganchev, Jahn, Marques, de Figueiredo, & Schuchmann, 2015; Hödl, 1977; Wimmer, Towsey, Planitz, Williamson, & Roe, 2013a), by knowledgeable specialists can achieve accurate results (Bedoya et al., 2014; Dong et al., 2015; Han et al., 2011; Towsey et al., 2014b). However, the effectiveness of such methods is limited by the economic cost of keeping experts in the field (Chen et al., 2012; Colonna, Cristo, Salvatierra, & Nakamura, 2015; Dong et al., 2015; Wimmer et al., 2013a), the fact that they are time-consuming (Chen et al., 2012; Dayou et al., 2011), extremely laborious and not entirely objective (Lee et al., 2013). The increased availability and affordability acoustic sensors over recent years greatly facilitates the recording of large volumes of continuous acoustic data in a passive and noninvasive manner, over extended periods of hours or even months (Bardeli et al., 2010; Bedoya et al., 2014; Dong et al., 2015; Towsey et al., 2014a, b; Wimmer et al., 2013a; Zsebők et al., 2015). This technology also facilitates data collection in remote areas where access is difficult (Bardeli et al., 2010; Ventura et al., 2015). It is claimed that it required 2 minutes of listening for an expert to identify species in 1 minute of audio it is estimated that on average it required 2 minutes of listening for an expert to identify species in 1 minute of audio (Wimmer, Towsey, Roe, & Williamson, 2013b), it is impractical for researchers to analyze manually the large volumes of acoustic recordings. For this reason, it is imperative to develop automated or semi-automated systems that simplify and speed up the task of scanning recordings for vocalizations of interest (Bardeli et al., 2010; Bedoya et al., 2014; Chen et al., 2012; Dayou et al., 2011; Dong et al., 2015; Patti & Williamson, 2013; Towsey et al., 2014a, b; Truskinger, Towsey, & Roe, 2015). Aside from reducing costs and human hours to manageable levels, these systems can also process a large amount of data with minimal habitat disturbance (Bedoya et al., 2014; Chen et al., 2012).

Automated systems (Chen et al., 2012; Lee et al., 2013; Somervuo, Härmä, & Fagerlund, 2006; Towsey et al., 2014a, b) hold out the promise of being fast and to detect a higher number of events in the recordings that the traditional census methods (Wimmer et al., 2013b). Nonetheless they do not have the accuracy currently required for ecological studies (Potamitis et al., 2014; Truskinger et al., 2015), usually require preliminary data on the structure of the vocalizations being studied (Dong et al., 2015), in many cases have insufficient training data and are unable to deal with the variability in the calls of most animal species (Wimmer et al., 2013a) and may incorrectly identify the calls of other species. A semi-automated approach is a hybrid that combines the advantages of humans and computers, giving ecologists more flexibility when analyzing acoustic data (Potamitis et al., 2014; Truskinger et al., 2015; Wimmer et al., 2013a). Computers can detect 50% more species than the traditional in-person search methods, but expert listeners can identify species calls that computers do not recognize (Wimmer et al., 2013a, b) such as different call types from the same species, different dialects, or calls masked by the presence of environmental noise. This is because human analysis capabilities are still superior to that of automated computational analysis tools (Wimmer et al., 2013a).

The main problem for automated and semi-automated systems are recognizing the segment of the recording where a focal animal's call starts and ends (Bardeli et al., 2010; Colonna et al., 2015; Jaafar & Ramli, 2013; Ventura et al., 2015), since undefined and unconstrained noise can mask vocalizations of interest (Jaafar et al., 2013; Patti & Williamson, 2013; Ryan, 1988; Towsey et al., 2014a). Unwanted sounds, generated by geophony (non-biological natural sound sources as wind, rain, leaf rustle, etc.) antrophony (human-induced noise sources as traffic, airplanes, machines, etc.) and biophony (sounds from other animals), can be considered noise (Brumm & Slabbekoorn, 2005; Cheng et al., 2010; Dong et al., 2015; Towsey et al., 2014b). In the state-of-art, various methods have been used to eliminate noise and segment the vocalizations of interest. Some studies have solved the noise problem using techniques tailored for specific researches (Bardeli et al., 2010), recording with directional microphones (Wielgat et al., 2007) or using band-pass filters with optimal frequency ranges for species-specific prefiltration (Patti & Williamson, 2013; Potamitis et al., 2014; Wielgat, Swietojanski, Potempa, & Król, 2012). Other authors use noise attenuation techniques of general application in bioacoustics analyses such as adaptive filtering algorithms based on energy function (Lee et al., 2013), adaptive level equalization (Dong et al., 2015; Towsey et al., 2014a), iterative time-domain algorithms (Bedoya et al., 2014; Cheng et al., 2010; Fagerlund, 2007; Härmä, 2003; Huang et al., 2009; Huang et al., 2013; Lee et al., 2006; Somervuo et al., 2006; Vaca-Castano & Rodriguez, 2010), time-domain energy functions (Juang & Chen, 2007), R-S method (Chou, Liu, & Cai, 2008), adaptive energy detection (AED) (Zhang & Li, 2015), Hilbert follower (Potamitis et al., 2014), Short Time Energy (STE) and Short Time Average Zero Crossing Rate (STAZCR) approaches (Chen et al., 2012; Colonna et al., 2015; Jaafar & Ramli, 2013; Jaafar et al., 2013; Tyagi, Hegde, Murthy, & Prabhakar, 2006), morphological filtering applied on the spectrogram seen as an image (Ventura et al., 2015) and high-pass filters that reduce the influence of low-frequency interferences from the environment such as wind and traffic noise (Brumm & Slabbekoorn, 2005; Dong et al., 2015; Towsey et al., 2014b; Ventura et al., 2015).

Previous literature in bioacoustical species identification has focused on insects (Ganchev & Potamitis, 2007; Gerhardt & Huber, 2002), bats (Alonso et al., 2015b; Armitage & Ober, 2010; Henríquez et al., 2014), birds (Bardeli et al., 2010; Chen & Li, 2013; Chou & Liu, 2009; Chou et al., 2008; Dong et al., 2015; Fagerlund, 2007; Ganchev et al., 2015; Härmä, 2003; Juang & Chen, 2007; Lee et al., 2013; Mitrovic, Zeppelzauer, & Breiteneder, 2006; Patti & Williamson, 2013; Somervuo et al., 2006; Towsey et al., 2014a; Truskinger et al., 2015; Tsai, Xu, & Lin, 2013; Tyagi et al., 2006; Vaca-Castano & Rodriguez, 2010; Ventura et al., 2015; Wielgat et al., 2007; Wielgat et al., 2012; Zhang & Li, 2015), anurans (Bedoya et al., 2014; Chen et al., 2012; Colonna et al., 2015; Dayou et al., 2011; Han et al., 2011; Hödl, 1977; Huang et al., 2009; Huang et al., 2013; Huang et al., 2014; Jaafar & Ramli, 2013; Jaafar et al., 2013) and other (Mitrovic et al., 2006; Zsebők et al., 2015). Of them, three major groups: birds, insects and anurans are also considered as