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

ARTICLE INFO

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

ABSTRACT

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 important indicators of environmental health (Bardeli et al., 2010; Colonna et al., 2015; Ganchev et al., 2015; Jaafar & Ramli, 2013; Jaafar et al., 2013; Lee et al., 2013; Towsey et al., 2014b; Vaca-Castano & Rodriguez, 2010; Wagner et al., 2014).

These studies have used feature extraction methods developed for human speech recognition applied to species and individual recognition in animals. The features include Mel-frequency Cepstrum Coefficients (MFCC) (Bedoya et al., 2014; Chou & Liu, 2009; Chou et al., 2008; Colonna et al., 2015; Fagerlund, 2007; Jaafar & Ramli, 2013; Jaafar et al., 2013; Lee et al., 2006; Mitrovic et al., 2006; Patti & Williamson, 2013; Somervuo et al., 2006; Tsai et al., 2013; Vaca-Castano & Rodriguez, 2010; Wielgat et al., 2007; Wielgat et al., 2012; Zhang & Li, 2015), entropy (Dayou et al., 2011; Dong et al., 2015; Han et al., 2011; Towsey et al., 2014a), Acoustic Complexity Index (ACI) (Towsey et al., 2014b), threshold-crossing rate (Fagerlund, 2007; Huang et al., 2009; Huang et al., 2013; Huang et al., 2014; Somervuo et al., 2006), short time energy (Fagerlund, 2007; Huang et al., 2013; Mitrovic et al., 2006; Somervuo et al., 2006), spectral centroid (Fagerlund, 2007; Han et al., 2011; Huang et al., 2009, 2013, 2014; Somervuo et al., 2006), signal bandwidth (Fagerlund, 2007; Huang et al., 2009, 2013, 2014; Somervuo et al., 2006), spectral roll-off frequency (Fagerlund, 2007; Huang et al., 2013; Huang et al., 2014; Somervuo et al., 2006), spectral flux (Fagerlund, 2007; Somervuo et al., 2006), spectral flatness (Fagerlund, 2007; Huang et al., 2013, 2014; Somervuo et al., 2006), Human-Factor Cepstral Coefficients (HFCC) (Wielgat et al., 2007), Linear Predictive Coding (LPC) coefficients (Juang & Chen, 2007; Lee et al., 2006; Mitrovic et al., 2006), Bark Frequency Cepstral Coefficients (BFCC) (Mitrovic et al., 2006), and Multi-Stage Average Spectrum (MSAS) (Chen et al., 2012).

These features may be used directly in models capable of classifying sequences of data. To date, researches have used mainly classifiers such as k-Nearest Neighbors ($k-NN$) (Dayou et al., 2011; Han et al., 2011; Huang et al., 2009; Jaafar & Ramli, 2013; Jaafar et al., 2013; Mitrovic et al., 2006; Vaca-Castano & Rodriguez, 2010), Decision Three (DT) (Fagerlund, 2007; Huang et al., 2013, 2014; Towsey et al., 2014a), Gaussian Mixture Model (GMM) (Cheng et al., 2010; Ganchev et al., 2015; Lee et al., 2013; Somervuo et al., 2006; Tsai et al., 2013), Hidden Markov Model (HMM) (Kogan & Margoliash, 1998; Lee et al., 2006, 2013; Patti & Williamson, 2013; Somervuo et al., 2006; Wielgat et al., 2012), Neural Network (NN) (Chou & Liu, 2009; Chou et al., 2008; Huang et al., 2013), Dynamic Time Warping (DTW) (Kogan & Margoliash, 1998; Somervuo et al., 2006; Wielgat et al., 2007), Support Vector Machine (SVM) (Fagerlund, 2007; Huang et al., 2009; Mitrovic et al., 2006; Zhang & Li, 2015; Zsebők et al., 2015), Linear Vector Quantization (LVQ) (Mitrovic et al., 2006), Singleton-type Recurrent Neural Fuzzy Networks (SRNFN) (Juang & Chen, 2007), Spectral Ensemble Average Voice Print (SEAV) (Tyagi et al., 2006), and Learning Algorithm for Multivariate Data Analysis (LAMDA) (Bedoya et al., 2014).

Motivated by previous related research in animal identification, we tested an animal vocalization identification system, primarily but not exclusively to anurans, composed of two subsystems: a semi-automated segmenter based on the combination of the noise removal method and the activity detector proposed by a paper from Ephraim and Malah (1984) and Boll (1979) respectively, which detects animal vocalizations in the environment recordings; and an automatic identification system using MFCC and GMM, which identifies the vocalizing species.

The sound enhancement method using Minimum-Mean Square Error (MMSE) Short-Time Spectral Amplitude (STSA) estimator, developed by Ephaim and Malah, and noise suppression algorithm using Spectral Subtraction (SS), developed by Steven F. Boll, have been successfully applied in Automatic Speech Recognition (ASR) researches (Alonso, Cabrera, Medina, & Travieso, 2015a; Mak & Yu, 2014). More than 2700 and 4000 papers have cited these methods, respectively, yet they have barely been exploited for use in animal identification (Patti & Williamson, 2013). Here we investigate for the first time the combination of both algorithms with MFCC and GMM methods that are already well known in ASR and animal identification.

This work focuses on the identification of anurans, the order of amphibians commonly known as frogs and toads. Of the approximately 7432 species of amphibians in the world, about 6531 are anurans, distributed in 54 families (Frost, 2015). Anurans are excellent environmental health indicators, due in part to their biphasic life cycle that in most species includes both aquatic and terrestrial stages.

Anurans vocalize by inhaling air through their nostrils into their lungs, and then pumping it back across vocal cords in the larynx. We mammals produce sound in a similar fashion. While mammals rely primarily on the larynx for amplification, however, the males of most frog species are able to amplify sound further in one or two inflatable vocal sacs located near the throat. Frogs with vocal sacs keep their mouth and nostrils closed while emitting most types of calls. Thus, instead of escaping, air passes from the larynx through a pair of openings in floor of the mouth known as vocal slits, and into the vocal sac or sacs. The vocal sacs help transmit sound to the environment. Some frogs repeat this calling process multiple times before reopening their nostrils, by shunting air back and forth between the sacs and the lungs (Duellman & Trueb, 1986; Hödl, 1977; Ryan, 1988). Anurans vocalize primarily during breeding activity, which for most species is restricted to certain seasons and for many just a few days or weeks (Duellman & Trueb, 1986; Hödl, 1977; Jaafar & Ramli, 2013; Ryan, 1988). Anurans can produce several different types of call. The type encountered most often by far, however, is the "advertisement call", a signal given almost exclusively by males that may be used in both male-male and male-female interactions (Duellman & Trueb, 1986). Among those the most commonly produced are the distress calls emitted both by females and males when they are in danger and the advertisement calls, produced only by males to provide information on its location and availability for mating (Duellman & Trueb, 1986; Hödl, 1977; Ryan, 1988). This last type of call has the advantage of being simple, repetitive, with a small vocalization repertoire and is distributed in a small low frequency range (Duellman & Trueb, 1986; Hödl, 1977; Ryan, 1988). Calls can be seen as an organized sequence of brief sounds known as syllables which basically consist of a sound that a anuran produces with a single blow of air from the lungs (Dayou et al., 2011; Han et al., 2011; Huang et al., 2009, 2014; Jaafar & Ramli, 2013; Jaafar et al., 2013; Vaca-Castano & Rodriguez, 2010). The anuran identification system proposed in this paper is focused on syllable identification.

This paper is presented as follows: Section 2 explains the methodology related with this work and used materials, Section 3 shows the experimental results and discusses the comparison with other similar researches. Finally, in Section 4 conclusions and future work are expressed.

2. Material and methods

2.1. Materials

There are 6531 species of anuran in the world, which 325 species are located in Central America. Panama is the country with the greatest diversity of anuran in the region with 173 species, followed by Costa Rica with 146 (Frost, 2015).

The database includes 75 Costa Rican species in 11 families, shown in the Table 1, and it is comprised of 95 unsupervised recordings recorded at 44.1 kHz in natural environments from Costa Rica, with a total duration of 1 h and 44 minutes. All

Table 1
Summary of the family and scientific names (Frost, 2015) of the 76 anuran species recorded in the database.

Family	Scientific name	Family	Scientific name
Aromobatidae	*Allobates talamancae*	Hylidae	*Agalychnis annae*
Bufonidae	*Incilius coccifer*		*Agalychnis callidryas*
	Incilius coniferus		*Agalychnis saltator*
	Incilius luetkenii		*Anotheca spinosa*
	Incilius melanochlorus		*Agalychnis spurrelli*
	Incilius periglenes		*Dendropsophus ebraccatus*
	Rhaebo haematiticus		*Dendropsophus microcephalus*
	Rhinella marina		*Dendropsophus phlebodes*
Centrolenidae	*Cochranella euknemos*		*Duellmanohyla uranochroa*
	Cochranella granulosa		*Ecnomiohyla miliaria*
	Espadarana prosoblepon		*Hyloscirtus colymba*
	Hyalinobatrachium chirripoi		*Hyloscirtus palmeri*
	Hyalinobatrachium colymbiphyllum		*Hypsiboas rosenbergi*
	Hyalinobatrachium fleischmanni		*Hypsiboas rufitelus*
	Hyalinobatrachium talamancae		*Isthmohyla picadoi*
	Hyalinobatrachium valerioi		*Isthmohyla pseudopuma*
	Sachatamia albomaculata		*Isthmohyla lancasteri*
	Sachatamia ilex		*Isthmohyla zeteki*
	Teratohyla pulverata		*Scinax boulengeri*
	Teratohyla spinosa		*Scinax elaeochrous*
Craugastoridae	*Craugastor crassidigitus*		*Scinax staufferi*
	Craugastor fitzingeri		*Smilisca baudinii*
	Craugastor podiciferus		*Smilisca phaeota*
	Craugastor stejnegerianus		*Smilisca puma*
	Craugastor talamancae		*Smilisca sila*
	Craugastor underwoodi		*Smilisca sordida*
	Pristimantis altae		*Tlalocohyla loquax*
	Pristimantis cruentus		*Trachycephalus typhonius*
	Pristimantis ridens	Leptodactylidae	*Engystomops pustulosus*
	Dendrobates auratus		*Leptodactylus insularum*
Dendrobatidae	*Oophaga granulifera*		*Leptodactylus melanonotus*
	Oophaga pumilio		*Leptodactylus poecilochilus*
	Phyllobates lugubris		*Leptodactylus savagei*
	Phyllobates vittatus	Microhylidae	*Hypopachus variolosus*
	Diasporus diastema	Ranidae	*Lithobates taylori*
Eleutherodactylidae	*Diasporus hylaeformis*		*Lithobates vibicarius*
	Diasporus vocator		*Lithobates warszewitschii*
Hemiphractidae	*Gastrotheca cornuta*		

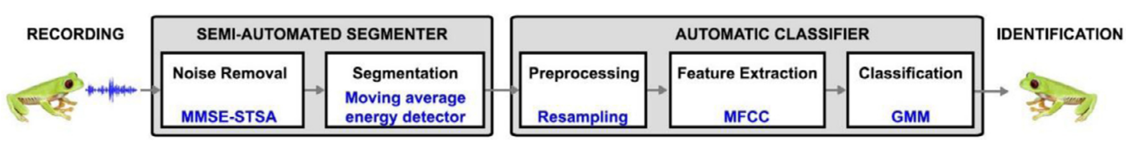

Fig. 1. Anuran syllables identification system.

recordings were made in natural environments in Costa Rica by co-authors A. Villegas, M. Wainwright, and A. García.

2.2. Methodology

The proposed system analyzes recordings in which anurans are likely to call. It consists of five stages, grouped in two subsystems: a semi-automated segmenter and automatic classifier (Fig. 1).

The first subsystem is divided into two stages. The first stage is a noise removal that estimates and reduces the background noise such as rain or wind noise. This facilitates the second stage, in which calls are detected and segmented into syllables. The second subsystem is dived into three stages. First, pre-processing stage is required in order to obtain a pattern space able to distinguish among species. Second stage performs the extraction of acoustic features of the syllables with the purpose of maximizing differences among anuran species. Third, a classification stage in which the previously extracted features for each syllables are analyzed in order to determine whether the selected syllables belongs or not to the pattern of pre-established species cluster.

2.2.1. Segmenter

Automated call segmentation approach are based on ASR methods that function adequately only when background noise is limited. In an environmental recordings is characterize by the present of a wide variety of noises, animal calls and by the recording condition cannot be controlled: distance from the microphone to sources, the movement of the target that we want to record, etc. (Wimmer et al., 2013a).

For these reasons, automated analysis systems often fail to segment field recordings correctly. The best strategy for maximizing the performance of animal identification systems while minimizing manual effort is the use of a semi-automated segmentation system supervised by human experts who can correct obvious errors and prepare the training or testing data set used in the automatic identification system (Kogan & Margoliash, 1998).

We have developed a semi-automated segmentation tool (Fig. 2) that combines the manual and automatic approach. The tool enables users create and manage a database, remove noise from recordings, systematically segment (activity detector) likely anuran call syllables (sample) within recordings, and facilitates recording playback, spectrogram display, and data storage.

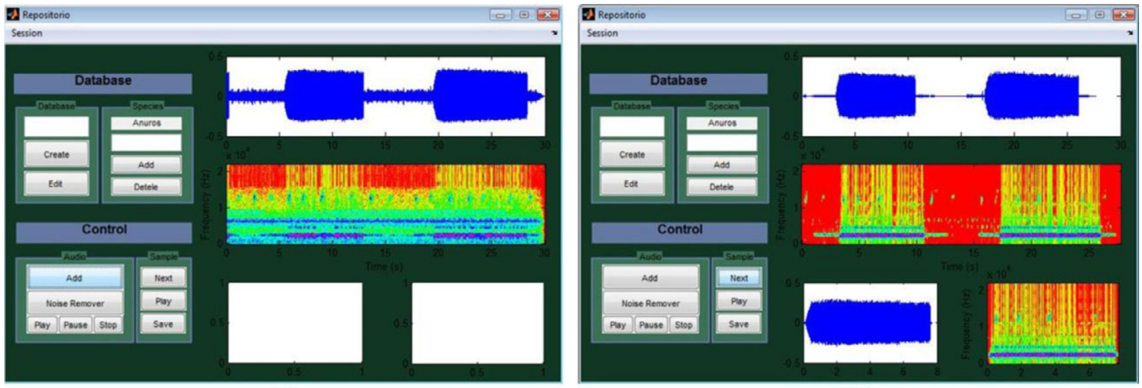

Fig. 2. Semi-automated segmentation tool. Example of application of noise removal (a) and segmentation (b) in a recording with noise.

This semi-automated segmentation tool, divided into noise removal stage and activity detector stage, provides users with the ability to analyze large volumes of acoustic data interactively and systematically. Semi-automated approach has the advantage that a human expert can associate many different vocalizations with a single species though the species have a broad range of vocalizations and these vocalizations may have significant regional variation (Wimmer et al., 2013a). Also, environmental factors such as wind, rain, vegetation and topography can attenuate, muffle and distort vocalizations considerably, a human expert can decide if these samples are or are not significant for the research (Wimmer et al., 2013a). For these reasons, semi-automated approach has the advantage that the ability of a human expert in visual analysis of sound spectrographs supported by auditory playback cannot currently be outperformed automated computational analysis tools (Cheng et al., 2010; Wimmer et al., 2013a) and we believe is the best approach currently.

2.2.1.1. Noise removal. Background noise is attenuated using the Minimum Mean-Square Error (MMSE) Short-Time Spectral Amplitude (STSA) estimator proposed by Ephraim and Malah, (1984). Here, the vocalization and noise spectral components are modeled as Gaussian random variables only once from an initial noise segment and the signal to noise ratio estimator is calculated a priori. Each recording is spectrally decomposed by means of a short-time discrete Fourier transform (DSTFT) using a Hamming window with overlap. The short-time spectral amplitude of the signal is then estimated and combined with the complex exponential of the noisy phase. The estimated DSTFT samples in each analysis frame are used for synthesizing the enhanced vocalization signal by using the weighted overlap and add method (Crochiere, 1980).

2.2.1.2. Activity detector. Once the noise removal stage is complete, vocalizations can be detected and segmented. As an activity detector for segmenting the recording into syllables, here we use a method of noise suppression through spectral subtraction proposed by Boll, (1979). This was done by using a moving average energy detector, which detects the presence of a syllable when the energy is above a certain threshold.

2.2.2. Automatic identification system

Once an advertisement call has been segmented into syllables, these are used in an automatic identification system. Syllables are preprocessed, resampled and windowed, before the feature extraction. Then acoustic features are extracted such that each syllable is transformed into a feature set. Feature sets are divided into two sets to train and test the classifier. Classifier is trained to distinguish between feature sets and is testing to classify new recordings as belonging to one of the target classes (species) or to an unknown class.

2.2.2.1. Signal preprocessing. The recorded sound signal is first resampled to sampling rate of 8 kHz. Then, from each syllable signal the DC component is removed and the z-score normalization is made. After that, the syllable signal is windowed by a Hamming window, which reduces discontinuity on both ends of a window, with a window size of 256 samples and overlapping size of 10% for each pair of successive windows.

2.2.2.2. Feature extraction. Non-statistical features such as Mel-Frequency Cepstral Coefficients (MFCC) have been used widely in human speech and speaker recognition systems. Such features have been used to identify animal species accurately in other research (Bedoya et al., 2014; Cheng et al., 2010; Colonna et al., 2015).

In our study, the MFCC features are extracted from each syllable. After signal preprocessing, a discrete Fourier transform (DFT) is executed over each frame. The Fourier coefficients obtained are squared and the result is filtered by a set of Mel-scaled triangular filters. Afterwards, a discrete cosine transform (DCT) is applied to convert the filter banks energies to the cepstral domain.

2.2.2.3. Classifier. Feature classification methods developed for human speech recognition have been applied to species and individual recognition in animals. Research into speech recognition has shown that probabilistic models provide a better model of acoustic speech events and framework for dealing with noise and channel degradation than non-probabilistic models (Cheng et al., 2010). We selected Gaussian Mixture Models as classification technique, with parameters of the GMM obtained by maximizing the likelihood function and obtained iteratively using the expectation–maximization algorithm.

3. Experimental results & discussion

Syllable segmentation is one of the important steps in animal identification system. Currently, most researches in animal identification system are focused mainly on automated segmentation due to its quick to process larger volumes of data. Nevertheless, in this paper and in other researches, a semi-automatic approach is defended as the best method to perform the complex segmentation in natural environments with multiple noise sources and with a predictable variability in the species vocalization. Semi-automated segmentation efficiency was evaluated by selecting randomly 33% recording used in this study, having expert biologists perform both and semi-automated segmentation on those recordings, and comparing the results. Manual segmentation was performed by visual inspection of the waveform graphics with digital audio editing

Table 2
Segmentation results.

Research	Species	Segmentation	Syllables detected	
			Manually	Auto / Semi
Jaafar and Ramli (2013)	Anuran	Automatic	100%	93%
Somervuo et al. (2006)	Birds	Automatic	82.33%	97%
Proposed segmentation	Anuran	Semi-automatic	85%	98%

Table 3
Performance of anuran identification system.

Species	Accuracy	Sensitivity	Specificity	PPV	NPV	FPR	FDR
1	96.1%	94.25%	97.95%	97.87%	94.46%	2.05%	2.13%
2	100%	100%	100%	100%	100%	0%	0%
3	99.79%	99.7%	99.88%	99.88%	99.70%	0.12%	0.12%
4	97.85%	96.81%	98.89%	98.87%	96.88%	1.11%	1.13%
5	100%	100%	100%	100%	100%	0%	0%
6	100%	100%	100%	100%	100%	0%	0%
7	99.49%	99.27%	99.71%	99.71%	99.27%	0.29%	0.29%
8	97.03%	95.59%	98.47%	98.42%	95.71%	1.53%	1.58%
9	96.31%	94.57%	98.05%	97.98%	94.75%	1.95%	2.02%
10	100%	100%	100%	100%	100%	0.00%	0%
11	97.03%	95.59%	98.47%	98.42%	95.71%	1.53%	1.58%
12	96.46%	94.74%	98.18%	98.12%	94.91%	1.82%	1.88%
13	96.82%	95.27%	98.37%	98.32%	95.41%	1.63%	1.68%
14	100%	100%	100%	100%	100%	0%	0%
15	99.54%	99.34%	99.74%	99.74%	99.34%	0.26%	0.26%
16	100%	100%	100%	100%	100%	0%	0%
17	100%	100%	100%	100%	100%	0%	0%
Mean	98.61%	97.95%	99.28%	99.25%	98.01%	0.72%	0.75%

software, while semi-automated segmentation was done using the system described in this paper.

The results, shown in the Table 2, were also compared, with the few researches in animal identification with automatic segmentation that presented experimentation between manual segmentation and automatic segmentation proposal in each one.

In this experiment, our approach shows similar mild better accuracy rate than the compared researches with automatic segmentation. This is probably due to the semi-automated segmentation tool developed, which uses noise removal and activity detector stage combined with visual analysis of sound spectrographs and auditory playback, improve the visual and auditory ability of a human expert to identify syllables species in noisy environments that is complex for the fully automated systems current.

In order to investigate anuran identification system performance using MFCC-GMM with respect to the number of Gaussian densities per model, between 1 to 24 MFCCs was computed from a data set of 17 species at random and 1, 2, 4, 8, 16, 32, 64 and 128 component Gaussian densities was modeled. We used three sizes of the training sets to perform the experiment: 33%, 50% and 66% of samples, and the rest of samples were used to testing. In order to obtain accuracy results each process was repeated 100 times.

The results obtained in the study of the optimal number of Gaussian mixture components, shown in the Fig. 3, are quite clear. For high number of Gaussian components (8, 16, 32, 64 and 128), classification needs a larger size of training set (66%) to achieve the best possible results, which however do not exceed a success rate of 70%. On the contrary, the success rate increases when number of Gaussian mixture components is lower (1, 2 and 4) - up to 90.25% for 1 Gaussian mixture component with a reasonable training set size (33% and 50%). For its part, the number of mel-frequency cepstrum coefficients is optimum in the range of 12 to 16 MFCC.

Once the optimal configuration of the system was established, we used recordings of 17 anuran species to train a model with 1 Gaussian mixture component using 16 MFCC. The performance of the classification system was evaluated through the use of a testing set. Despite having optimized system configuration, sometimes the model with the highest score may not be the most suitable. Therefore, a unique threshold for each species was established, so that if the maximum probability value of a species was not higher than the threshold, the system would reject the sample. The threshold selected was in the point of "equal error rate" (EER). The EER is the cross point between the false acceptance curve, accepting a sample as a species but was not, and the false rejection curve, reject a sample that belongs to species.

The Table 3 shows the system performance analysis. Seven parameters are evaluated: accuracy (measures the reliability of the system for a high activation), sensitivity (ratio of right answers in genuine comparisons respect the total of comparisons. It measures the ability of the system to detect the high activation tested), specificity (ratio of right answers in non-genuine comparisons respect the total of comparisons. As higher the value is, more difficult is to replace the high activation by another), Positive Predictive Value (PPV) (a measure of the probability of a true positive result is a true positive), Negative Predictive Value (NPV) (it is a measure of the probability that a negative result is really a true negative), False Positive Rate (FPR) (false acceptance ratio to the total of non-genuine comparisons. It represents the validity of the system) and False Discovery Rate (FDR) (false acceptance ratio to the total of genuine comparisons. FDR procedures are designed to control the expected proportion of incorrectly accepted false positives).

The results show an overall accuracy over 99% for a relative wide amount of species indicating the reliability of the system described. In terms of single species' syllables, the system show excellent potential of recognition among individuals of the same species, with 47% of species having a success rate of 100% and another 6% of species with a success rate over 99%. Therefore the efficiency of the system in the field of anuran identification is confirmed.

The high success rate indicates that the semi-automatic segmentation tool provides high noise immunity to the system, thus preventing effectively use samples degraded by noise in the

Fig. 3. Study of optimal number of Gaussian mixture components and MFCCs.

training and testing data set. In any case, it is import to note that in this study the acoustic conditions of the train and test datasets are the same, i.e. the train and test datasets were drawn from the same recording or in few cases from different recordings.

Comparison of these results to those of others should take into consideration that many factors may vary between studies, such as noises conditions or recording quality, the number of species used, the types of experiments performed, and the format in which results are presented. However, in the Table 4 we have summarized to the degree possible a comparison of the results of different studies on anuran acoustical identification.

Regarding the comparison of similar researches in the state-of-art, accuracies between 85% and 99% were achieved; dominate the syllables feature extraction by mel-frequency cepstrum coefficients (MFCC) and k-Nearest Neighbors method for the anuran syllables classification. Furthermore, it has shown a widespread use of automatic segmentation approach based on the energy function.

The system proposed in this work using similar feature extraction over the previous results obtained by the comparative studies, despite using a larger number of species that complicates the classification task. Only one study with automatic segmentation (Bedoya et al., 2014) and one study with manual segmentation (Han et al., 2011) outperform 98% of success rate, with success

Table 4
Comparison of accuracy (Acc.) results in anuran identification research.
Segmentation analysis: automatic (Auto), manual, not available (N/A), semi-automatic (Semi).

Research	Species	Segmentation		Features	Classifier	Acc.
		Analysis	Method			
Bedoya et al. (2014)	6	Auto	Iterative time-domain algorithms	MFCC	LAMDA	99.61%
Dayou et al. (2011)	9	N/A	N/A	Shannon entropy Rényi entropy Tsallis entropy	k-NN	90%
Han et al. (2011)	9	Manual	Raven Lite software	Spectral centroid Shannon entropy Rényi entropy	k-NN	98.15%
Huang et al. (2009)	5	Auto	iterative time-domain algorithm	Spectral centroid Signal bandwidth Threshold-crossing rate	SVM	90.3%
Huang et al. (2014)	9	Auto	Adaptive end-point detection	Spectral centroid Signal bandwidth Spectral roll-off Threshold-crossing rate Spectral flatness Average energy	DT+NN	93.4%
Jaafar et al. (2013)	15	Auto	STE-STAZCR	MFCC	k-NN	85.78%
Vaca-Castano and Rodriguez (2010)	20	Auto	Iterative time-domain algorithms	MFCC	k-NN	86.78%
Chen et al. (2012)	18	Auto	STE-STAZCR	MSAS	Template matching method	94.3%
Proposed	17	Semi	MMSE-STSA+SS	MFCC	GMM	98.61%

rate of 99.61% and 98.15% respectively. Therefore, the approach detailed in this work with a success rate of 98.61% is presented as a good anuran identification system, capable of identifying up to three times more species than the study (Bedoya et al., 2014) with the best results in the state-of-art and with the same reliability that offers manual segmentation in studies (Han et al., 2011) with similar results but providing a segmentation task faster and more scalable for expert biologists to use semi-automatic segmentation. In any case, it is import to note that in this study the acoustic conditions of the train and test datasets are the same, i.e. the train and test datasets were drawn from the same recording or in few cases from different recordings. Also the size of the recording used in the study is limited to considerate conclusive and it is needed to evaluate the proposal method with other databases used in the state-of-art.

4. Conclusions

In this work an identification system for anuran syllables with semi-automated segmentation is proposed to effectively recognize the anuran syllables in natural environments. The approach has been developed with the purpose to provide a tool for analyzing of larger volumes of recorded acoustic data that help biologists and ecologists in the fields of biodiversity monitoring and soundscape ecology.

In this paper, a semi-automated segmentation tool based on the use of minimum-mean square error short-time spectral amplitude estimator (Ephraim & Malah, 1984) combined with noise suppression algorithm using spectral subtraction method (Boll, 1979) is implemented. A semi-automatic segmentation approach assisted by noise removal and activity detector is used compared to the trend of automatic approaches used in the majority of researches in the state-of-art. Semi-automatic segmentation provide hybrid approach to the ecologists with flexibility when analyzing acoustic data interactively and systematically, minimizing effort of segmenting task, allowing associate many different vocalizations with a single species, allowing discard degraded samples due to the noise that are not used in the classification systems (training and testing) by improving their efficiency, avoiding the main limitations of the automatic systems which tend to reliably generate correct and erroneous segmentation and need training data in the form of a wide range of prior knowledge species vocalizations.

Motivated by previous related researches in animal identification this paper exploring segmentation based on sound enhancement method using minimum-mean square error short-time spectral amplitude estimator and noise suppression algorithm using spectral subtraction, which have been successfully applied in ASR researches and however there is little evidence in the state-of-art of its use in the important step of syllable segmentation for animal identification system. The experiments were developed with a database of 75 anuran species from several families recorded in natural environments from Costa Rica. The experimental results obtained using the proposed system showed an accuracy of 98% and were compared with the manual segmentation and other researches in animal identification with automatic segmentation, showing a slightly better accuracy rate. This slight increase in the syllables detection is due to the best ability of human experts to identify syllables supported by visual analysis of sound spectrographs and auditory playback.

An automatic classification system of anuran syllables based on semi-automatic segmentation has also been studied. In our study, the MFCC features are extracted from each syllable and GMM classification technique are applied. These methods are well known in ASR and animal identification but have not been studied in combination with the proposed segmentation system to date. The experiments were developed with 17 anuran species to determine the optimal number of Gaussian densities components and number of MFCC, getting high success rate for low number of Gaussian mixture components and a number of MFCC between 12 and 16 coefficients. The system performance using a model with 1 Gaussian mixture component and 16 MFCC was evaluated. The results showed accuracies between 96.1% and 100% per species and excellent potential of anuran identification with average success rate of 98.61%. In comparison with other similar researches, the proposed approach has been shown to have a better accuracy rate than most the rest of researches in spite of using more amount of anuran species in their classification and the same acoustic features. In conclusion, the semi-automated segmentation has improved the classification system.

In view of the performance system and the experimental results, the proposed recognition system based on semi-automated segmentation is adequate for the anuran identification and their potential is not limited only to the anuran acoustic research if not that can be extended to the identification of others complex animal vocalizations such as birds, insects or mammals. In future works, this approach will be implemented with other animal species and complemented with the search of acoustic features and classification methods that may improve the recognition in a range of species more.

The system proposed in this paper opens the door to global web systems based on expert and intelligent systems that would allow to monitor indicators of changes in the environmental health of our planet on real time and to reaction with the adequate environmental policies.

References

Alonso, J. B., Cabrera, J., Medina, M., & Travieso, C. M. (2015a). New approach in quantification of emotional intensity from the speech signal: Emotional temperature. *Expert Systems with Applications, 42*(24), 9554–9564.

Alonso, J. B., Henríquez, A., Henríquez, P., Rodríguez-Herrera, B., Bolaños, F., Alpízar, P., et al. (2015b). Advance in the bat acoustic identification systems based on the audible spectrum using nonlinear dynamics characterization. *Expert Systems with Applications, 42*(24), 9528–9538.

Armitage, D. W., & Ober, H. K. (2010). A comparison of supervised learning techniques in the classification of bat echolocation calls. *Ecological Informatics, 5*(6), 465–473.

Bardeli, R., Wolff, D., Kurth, F., Koch, M., Tauchert, K., & Frommolt, K. (2010). Detecting bird sounds in a complex acoustic environment and application to bioacoustic monitoring. *Pattern Recognition Letters, 31*(12), 1524–1534.

Bedoya, C., Isaza, C., Daza, J. M., & López, J. D. (2014). Automatic recognition of anuran species based on syllable identification. *Ecological Informatics, 24*, 200–209.

Boll, S. F. (1979). Suppression of acoustic noise in speech using spectral subtraction. *Acoustics, Speech and Signal Processing, IEEE Transactions On, 27*(2), 113–120.

Brumm, H., & Slabbekoorn, H. (2005). Acoustic communication in noise. *Advances in the Study of Behavior, 35*, 151–209.

Chen, S., & Li, Y. (2013). Automatic recognition of bird songs using time-frequency texture. In *Computational intelligence and communication networks (CICN), 2013 5th International conference On* (pp. 262–266).

Chen, W., Chen, S., Lin, C., Chen, Y., & Lin, W. (2012). Automatic recognition of frog calls using a multi-stage average spectrum. *Computers & Mathematics with Applications, 64*(5), 1270–1281.

Cheng, J., Sun, Y., & Ji, L. (2010). A call-independent and automatic acoustic system for the individual recognition of animals: A novel model using four passerines. *Pattern Recognition, 43*(11), 3846–3852.

Chou, C., & Liu, P. (2009). Bird species recognition by wavelet transformation of a section of birdsong. In *Ubiquitous, autonomic and trusted computing, 2009. UIC-ATC'09. Symposia and workshops On* (pp. 189–193).

Chou, C., Liu, P., & Cai, B. (2008). On the studies of syllable segmentation and improving MFCCs for automatic birdsong recognition. In *Asia-Pacific services computing conference, 2008. APSCC'08* (pp. 745–750). IEEE.

Colonna, J. G., Cristo, M., Salvatierra, M., & Nakamura, E. F. (2015). An incremental technique for real-time bioacoustic signal segmentation. *Expert Systems with Applications*.

Crochiere, R. E. (1980). A weighted overlap-add method of short-time fourier analysis/synthesis. *Acoustics, Speech and Signal Processing, IEEE Transactions On, 28*(1), 99–102.

Dayou, J., Han, N. C., Mun, H. C., Ahmad, A. H., Muniandy, S. V., & Dalimin, M. N. (2011). Classification and identification of frog sound based on entropy approach. In *International conference on life science and technology: 3* (pp. 184–187).

Dong, X., Towsey, M., Truskinger, A., Cottman-Fields, M., Zhang, J., & Roe, P. (2015). Similarity-based birdcall retrieval from environmental audio. *Ecological Informatics, 29*, 66–76.

Duellman, W. E., & Trueb, L. (1986). *Biology of amphibians*. JHU Press.

Ephraim, Y., & Malah, D. (1984). Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator. *Acoustics, Speech and Signal Processing, IEEE Transactions On, 32*(6), 1109–1121.

Fagerlund, S. (2007). Bird species recognition using support vector machines. *EURASIP Journal on Applied Signal Processing, 2007*(1), 64.

Frost, D. (2015). *Amphibian species of the world: An online reference* version 6.0 (17 january 2014).

Ganchev, T. D., Jahn, O., Marques, M. I., de Figueiredo, J. M., & Schuchmann, K. (2015). Automated acoustic detection of vanellus chilensis lampronotus. *Expert Systems with Applications, 42*(15), 6098–6111.

Ganchev, T., & Potamitis, I. (2007). Automatic acoustic identification of singing insects. *Bioacoustics, 16*(3), 281–328.

Gerhardt, H. C., & Huber, F. (2002). *Acoustic communication in insects and anurans: common problems and diverse solutions*. University of Chicago Press.

Han, N. C., Muniandy, S. V., & Dayou, J. (2011). Acoustic classification of australian anurans based on hybrid spectral-entropy approach. *Applied Acoustics, 72*(9), 639–645.

Härmä, A. (2003). Automatic identification of bird species based on sinusoidal modeling of syllables. In *Acoustics, speech, and signal processing, 2003. Proceedings. (ICASSP'03). 2003 IEEE International conference On, 5V-545-8: vol. 5*.

Henríquez, A., Alonso, J. B., Travieso, C. M., Rodríguez-Herrera, B., Bolaños, F., Alpízar, P., et al. (2014). An automatic acoustic bat identification system based on the audible spectrum. *Expert Systems with Applications, 41*(11), 5451–5465.

Hödl, W. (1977). Call differences and calling site segregation in anuran species from central amazonian floating meadows. *Oecologia, 28*(4), 351–363.

Huang, C., Chen, Y., Chen, H., Jian, J., Tseng, S., Yang, Y., et al. (2014). Intelligent feature extraction and classification of anuran vocalizations. *Applied Soft Computing, 19*, 1–7.

Huang, C., Lin, C., Hsu, P., Lee, Y., Chen, H., Lien, Y., et al. (2013). Application of fast learning neural-networks to identification of mixed anuran vocalizations. In *Conference anthology* (pp. 1–5). IEEE.

Huang, C., Yang, Y., Yang, D., & Chen, Y. (2009). Frog classification using machine learning techniques. *Expert Systems with Applications, 36*(2), 3737–3743.

Jaafar, H., & Ramli, D. A. (2013). Automatic syllables segmentation for frog identification system. In *Signal processing and its applications (CSPA), 2013 IEEE 9th International colloquium On* (pp. 224–228).

Jaafar, H., Ramli, D. A., & Shahrudin, S. (2013). MFCC based frog identification system in noisy environment. In *Signal and image processing applications (ICSIPA), 2013 IEEE International conference On* (pp. 123–127).

Juang, C., & Chen, T. (2007). Birdsong recognition using prediction-based recurrent neural fuzzy networks. *Neurocomputing, 71*(1), 121–130.

Kogan, J. A., & Margoliash, D. (1998). Automated recognition of bird song elements from continuous recordings using dynamic time warping and hidden markov models: A comparative study. *The Journal of the Acoustical Society of America, 103*(4), 2185–2196.

Lee, C., Hsu, S., Shih, J., & Chou, C. (2013). Continuous birdsong recognition using gaussian mixture modeling of image shape features. *Multimedia, IEEE Transactions On, 15*(2), 454–464.

Lee, C., Lee, Y., & Huang, R. (2006). Automatic recognition of bird songs using cepstral coefficients. *Journal of Information Technology and Applications, 1*(1), 17–23.

Mak, M., & Yu, H. (2014). A study of voice activity detection techniques for NIST speaker recognition evaluations. *Computer Speech & Language, 28*(1), 295–313.

Mitrovic, D., Zeppelzauer, M., & Breiteneder, C. (2006). Discrimination and retrieval of animal sounds. *Paper presented at the Multi-media modelling conference proceedings, 2006 12th International* (p. 5).

Patti, A., & Williamson, G. (2013). Methods for classification of nocturnal migratory bird vocalizations using pseudo wigner-ville transform. In *Acoustics, speech and signal processing (ICASSP), 2013 IEEE International conference On* (pp. 758–762).

Potamitis, I., Ntalampiras, S., Jahn, O., & Riede, K. (2014). Automatic bird sound detection in long real-field recordings: Applications and tools. *Applied Acoustics, 80*, 1–9.

Ryan, M. J. (1988). Constraints and patterns in the evolution of anuran acoustic communication. *The Evolution of the Amphibian Auditory System*, 637–677.

Somervuo, P., Härmä, A., & Fagerlund, S. (2006). Parametric representations of bird sounds for automatic species recognition. *Audio, Speech, and Language Processing, IEEE Transactions On, 14*(6), 2252–2263.

Towsey, M., Wimmer, J., Williamson, I., & Roe, P. (2014a). The use of acoustic indices to determine avian species richness in audio-recordings of the environment. *Ecological Informatics, 21*, 110–119.

Towsey, M., Zhang, L., Cottman-Fields, M., Wimmer, J., Zhang, J., & Roe, P. (2014b). Visualization of long-duration acoustic recordings of the environment. *Procedia Computer Science, 29*, 703–712.

Truskinger, A., Towsey, M., & Roe, P. (2015). Decision support for the efficient annotation of bioacoustic events. *Ecological Informatics, 25*, 14–21.

Tsai, W., Xu, Y., & Lin, W. (2013). Bird species identification based on timbre and pitch features. In *Multimedia and expo (ICME), 2013 IEEE International conference On* (pp. 1–6).

Tyagi, H., Hegde, R. M., Murthy, H., & Prabhakar, A. (2006). Automatic identification of bird calls using spectral ensemble average voice prints. In *Signal processing conference, 2006 14th European* (pp. 1–5).

Vaca-Castano, G., & Rodriguez, D. (2010). Using syllabic mel cepstrum features and k-nearest neighbors to identify anurans and birds species. In *Signal processing systems (SIPS), 2010 IEEE workshop On* (pp. 466–471).

Ventura, T. M., de Oliveira, A. G., Ganchev, T. D., de Figueiredo, J. M., Jahn, O., Marques, M. I., et al. (2015). Audio parameterization with robust frame selection for improved bird identification. *Expert Systems with Applications, 42*(22), 8463–8471.

Wagner, N., Züghart, W., Mingo, V., & Lötters, S. (2014). Are deformation rates of anuran developmental stages suitable indicators for environmental pollution? Possibilities and limitations. *Ecological Indicators, 45*, 394–401.

Wielgat, R., Swietojanski, P., Potempa, T., & Król, D. (2012). On using prefiltration in HMM-based bird species recognition. In *Signals and electronic systems (ICSES), 2012 International conference On* (pp. 1–5).

Wielgat, R., Zieliński, T. P., Potempa, T., Lisowska-Lis, A., & Król, D. (2007). HFCC based recognition of bird species. *Processing Algorithms, Architectures, Arrangements and Applications, 2007*, 129–134.

Wimmer, J., Towsey, M., Planitz, B., Williamson, I., & Roe, P. (2013a). Analysing environmental acoustic data through collaboration and automation. *Future Generation Computer Systems, 29*(2), 560–568.
Wimmer, J., Towsey, M., Roe, P., & Williamson, I. (2013b). Sampling environmental acoustic recordings to determine bird species richness. *Ecological Applications, 23*(6), 1419–1428.
Zhang, X., & Li, Y. (2015). Adaptive energy detection for bird sound detection in complex environments. *Neurocomputing, 155*, 108–116.
Zsebők, S., Czabán, D., Farkas, J., Siemers, B. M., & von Merten, S. (2015). Acoustic species identification of shrews: Twittering calls for monitoring. *Ecological Informatics, 27*, 1–10.