We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900
Open access books available

186,000
International authors and editors

200M
Downloads

154
Countries delivered to

Our authors are among the

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS

BOOK CITATION INDEX

INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

A Methodology Based on Bioacoustic Information for Automatic Identification of Reptiles and Anurans

Juan J. Noda, David Sánchez-Rodríguez and
Carlos M. Travieso-González

Additional information is available at the end of the chapter

Abstract

Nowadays, human activity is considered one of the main risk factors for the life of reptiles and amphibians. The presence of these living beings represents a good biological indicator of an excellent environmental quality. Because of their behavior and size, most of these species are complicated to recognize in their living environment with image devices. Nevertheless, the use of bioacoustic information to identify animal species is an efficient way to sample populations and control the conservation of these living beings in large and remote areas where environmental conditions and visibility are limited. In this chapter, a novel methodology for the identification of different reptile and anuran species based on the fusion of Mel and Linear Frequency Cepstral Coefficients, MFCC and LFCC, is presented. The proposed methodology has been validated using public databases, and experimental results yielded an accuracy above 95% showing the efficiency of the proposal.

Keywords: acoustic data fusion, bioacoustic processing, biological acoustic analysis, anurans identification, reptiles identification, pattern recognition, cepstral coefficients

1. Introduction

The technological advances open the door to develop and implement tools in different and wide fields of science. In particular, the use of specific devices to acquire sound, the use of big computational load, the implementation on programming languages of feature extraction algorithms, and machine learning systems give the option to develop a novel approach to identify different kinds of animal species from their sounds. This type of tool will do easier Biologist's task on their studies about the environment and the behavior of those animal species.
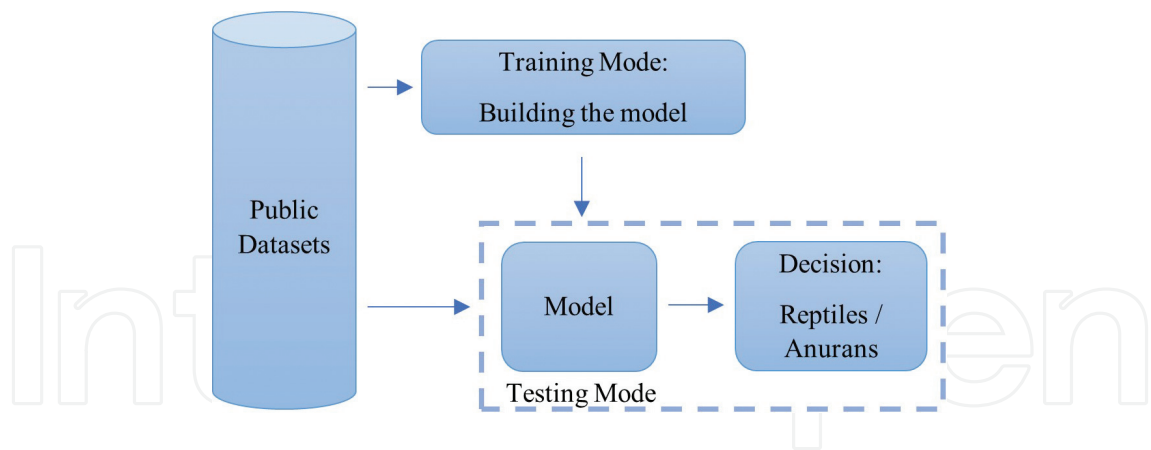
Figure 1. Block diagram of the recognition system based on reptiles and anurans.

Many features can be used for these tools, but it can be closed according to the kind of species to analyze. The main features are from videos, images, or sounds. For this work, the idea is to propose a methodology to identify reptile and anuran species; therefore, all previous features can be applied. Nevertheless, the main activity of these species is during the night. Hence, the sound is the most useful feature to know the daily activity and to carry out the species identification.

In this chapter, the bioacoustic information will be the feature used to this development; and a robust and novel proposal based on the fusion of MFCC and LFCC for the identification of different reptile and anuran species is presented. The proposed approach has been validated according to **Figure 1** using public datasets, and experimental results show the efficiency of the proposal. Based on a supervised classification system, this approach is composed by two modes, training and testing modes. This methodology follows a hold-out cross validation method.

In addition, a feature extraction technique with the highest classification capacity and minimal computation complexity is implemented. To face this challenge, a set of experiments allowing the comparison between the performances of the different feature extraction techniques to apply are shown. The goal is to specify which features are the most effective, obtaining the bioacoustic characteristics and the identification of reptile and anuran species.

The rest of this chapter is organized as follows: Section 2 shows related works. Section 3 describes the methods for automatic identification of reptile and anuran species. In Section 4, the experimental methodology and the results are described. Finally, in Section 5, the conclusions derived from this work are summarized.

2. Related works

There are numerous previous studies on the spectral-temporal characteristics of the acoustic emissions produced by animals, which attempt to analyze the frequency and time parameters of these emissions to identify patterns in their communications and social iterations [1, 2].

In recent years, various efforts have been made to automate the processing of acoustic information using intelligent systems. However, most of the studies conducted in this field have focused their research on a single animal group and, in most cases, these studies have been carried out on just a few species. For instance, one of the first attempts to automatically recognize animal species can be found in [3], where neural networks were used to classify the vocalizations of two false killer whales.

The sounds of insects have also been studied. As an example of this, in [4], their emissions were characterized by using LFCC, their fundamental harmonic and the distance of each call. Authors achieved an 86.3% identification result at species level using Gaussian mixture model (GMM) as a classifier. On the other hand, in [5], the identification of 14 species of birds using two different sets of parameters was proposed. The first data set represented sounds using Mel coefficients and the second consisted of a set of signal parameters such as frequency range, spectral flow, and Wiener's entropy. For the classification of the vocalizations, authors proposed a decision tree (DT) where each node of the tree was formed by a support vector machine. In their experiments, MFCC achieved the best results but separated the species into two sets of data.

Many other methods have been applied to other groups of animals such as primates [6, 7], bats [8], fishes [9, 10], elephants [11], dolphins [12], but birds [13–16] have been especially studied for their wide variety of vocalizations. Recently, however, the acoustic characteristics of the anurans have managed to attract the attention of the scientific community, due to their relatively simple vocalizations and abundant sound production, which make them ideal test subjects for automatic recognition. Therefore, several studies have been carried out with varying degrees of success extracting different types of acoustic signal parameters to characterize the amphibian vocalizations.

One example of this can be found in [17], where four anuran species were classified using neural networks (NN), applying a discrete wavelet transform (DTW) to get the main features of each frog call and Fisher's optimization criterion of reducing the data dimensionality. This method was able to identify the species with a success rate of 71%, but it required a high computational cost. Instead, in [18], five frog species were analyzed, computing the threshold-crossing rate, signal bandwidth, and spectral centroid. With these features, they achieved an accuracy of 89.05% by using K-Nearest Neighbors (KNN) and 90.30% by applying support vector machine (SVM). A different approach was proposed by Han et al. in [19] that combined three types of entropy (Shannon, Rényi, and Tsallis) to recognize nine Microhylidae frog species. This method managed to correctly identify only seven of the nine frog species due to the similar entropy values among these species.

Low-level acoustic attributes have been also used to discriminate frog vocalizations at genus level with a significant rate of success [20]. Coefficient of variation of root-mean-square energy, dominant frequency, and spectral flux were computed for short-time frames to distinguish between the advertising calls of four genera, *Bufo*, *Hyla*, *Leptodactylus*, and *Rana*.

On the other hand, MFCC have been widely used in the recognition of anurans and reptiles in combination with a variety of pattern recognition techniques, due to their noise robustness and

computational efficiency. For instance, an interesting approach was developed by [21], that achieved the classification of 30 frogs and 19 cricket species with success rates above 96% with a large standard deviation. For this they split the acoustic signal into frames and calculated the average of the MFCC to train a linear discriminant analysis (LDA) algorithm. Another example can be found in [22], where the MFCC were tested in three algorithms: Local Mean KNN with Fuzzy Distance Weighting (LMkNN-FDW), sparse representation classifier (SRC), and SVM. LMkNN-FDW outperformed SRC and SVM, obtaining the highest-performance results on 20 frog species.

At present, deep learning techniques are being employed in frog acoustics classification [23–25], applying convolutional neural networks (CNN). However, most of these works also use MFCC as parameters, relying on the discriminatory capacity of the classifier without looking for a better representation of the acoustic signal information. **Table 1** summarizes some different techniques and algorithms that have been used in the recognition of anurans.

Lastly, the class *Reptilia*, however, has received little attention due to its limited sound production. In fact, to the best of our knowledge, the acoustic signals for reptile's automatic identification have been poorly considered in literature, being this work one of the first research to address this approach. Although only a few species such as crocodiles present an important repertory of calls. In this research, the sound emitted by reptiles has been intensively studied to verify the capacity of their calls for inter-species identification.

References	Parametrization	Classifier
Yen and Fu [17]	Discrete wavelet transform (DTW)	NN
Lee et al. [21]	MFCC	LDA
Brandes [43]	Peak frequencies and bandwidth	HMM
Acevedo et al. [41]	Call length, maximum and minimum frequencies, maximum power, and the frequency of maximum power	SVM, DT and LDA
Huang et al. [18]	Spectral centroid, signal bandwidth, and threshold-crossing rate	KNN and SVM
Han et al. [19]	Shannon, Rényi, and Tsallis entropies	KNN
Yuan et al. [27]	MFCC and linear predictive coding (LPC)	KNN
Bedoya et al. [42]	MFCC	Learning algorithm for multivariate data analysis (LAMDA)
Chen et al. [44]	Length of the segmented syllables	Multi stage average spectrum (MSAS)
Xie et al. [28]	Dominant frequency, syllable duration, frequency modulation, oscillation rate, and energy modulation	PCA and KNN
Hassan et al. [25]	MFCC	Convolutional neural networks (CNN)

Table 1. Summary of amphibian acoustic identification methods proposed by literature.

3. Methods

The proposed methodology in this chapter is illustrated in **Figure 2** and is composed by the following methods. First, both reptile and anuran audio recordings are processed by a segmentation algorithm to separate the acoustic signal in syllables. Next, the cepstral feature parameters, MFCC and LFCC, are extracted and fused in a vector standing for the main characteristics for each syllable. Then, these vectors are used as inputs in the classification phase for training and testing a classifier implemented by a machine learning algorithm. Next, a detailed description of each method is defined.

3.1. Signal segmentation

In order to obtain useful features for the automatic identification, the audio recordings are split into as many syllables as possible. This process is based on the work of Härmä [26] for acoustic signal segmentation. The algorithm makes use of the signal spectrogram to detect each sound and separate it into syllables. The spectrogram was determined by short-time Fourier transform (STFT) with the following Hamming window sizes which have been heuristically computed: 256 samples and 33% overlap for reptiles, and 512 samples and 25% overlap for anurans corpus. As a result, the matrix $H(f, t)$ represents the computed signal spectrum where f is the frequency and t the time. The segmentation procedure performs the following steps repeatedly until the end of the spectrogram is reached:

1   Find t_n and f_n such that $\left|H(f_n, t_n)\right| \geq |H(f, t)|$ computing the amplitude in t_n as $\Upsilon_n(0) = 20 \log_{10}(|H(f_n, t_n)|)$.

2   From t_n, seek the highest peak between $t > t_n$ and $t < t_n$ until $\Upsilon_n(t - t_n) < \Upsilon_n(0) - \beta\ dB$, where β is the stopping criteria. For reptile and anuran sounds, β has been set to 25 and 20 dB, respectively. The time interval $[t_n - t_s, t_n + t_e]$ represents the limits of the syllable.

3   This trajectory which represents a syllable is stored and then, is deleted from the matrix. The index n is updated to $n + 1$.

3.2. Features extraction and fusion

After carrying out the segmentation process, frequency domain characteristics are computed to gather useful information for the automatic classification. MFCC and LFCC have been applied in
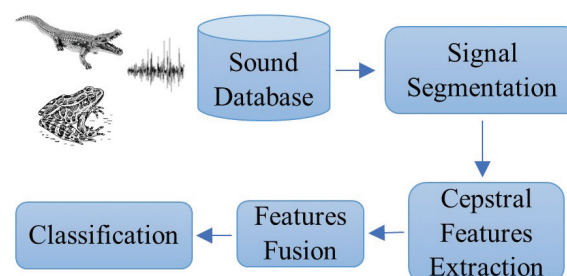

Figure 2. The proposed methodology for automatic acoustic identification of reptiles and anurans.

animal bioacoustic classification [4, 22, 27, 28], because they have a low computational cost and their implementations are easy. On the other hand, low frequency sounds are emitted by most species of reptiles and anurans, that is, in a 0–20 kHz interval, such as human auditory range. Thus, to reinforce the low frequency range, MFCC have been considered. Nevertheless, both corpuses can produce sounds above 20 kHz, and hence, to get a characterization in high frequency ranges, LFCC have also been used [29]. Thus, both cepstral coefficients are computed to parametrize the audio signal, because they contain information of lower and higher frequency.

These features are computed via STFT using 25 milliseconds Hamming window overlapping at 50%. In order to get this value, a set of experiments were carried out where the window size was modified from 10 ms to 1 s. After that, the discrete Fourier transform (DFT) is computed in each signal frame, and a bank of 40 and 26 triangular band pass filters for reptiles and anurans, respectively, are wrapped to the resultant spectrum. The MFCC are obtained by applying the discrete cosine transform (DCT) to log-magnitude filter outputs, $\log |\Upsilon_i|$, and taking the lowest values. MFCC features are calculated as follows, Eq. (1):

$$MFCC_j = \sum_{i=1}^{B} (\log |\Upsilon_i| \cos [j(i - 0.5)]\pi/B), \qquad 1 \leq j \leq N \tag{1}$$

where j indicates the MFCC index, B is the number of triangular filters, and N is the MFCC to calculate.

On the other hand, LFCC are calculated using Eq. (2), where K is the number of DFT magnitude coefficients X_i.

$$LFCC_j = \sum_{i=1}^{K} \left(\log |X_i| \cos \left[\frac{ji\pi}{K}\right] \right), \qquad 1 \leq j \leq N \tag{2}$$

For both features, the coefficients number has been obtained by carrying out a set of experiments to achieve the highest accuracy in the classification phase. Thus, 18 coefficients have been taken for both MFCC and LFCC.

Finally, the cepstral coefficients are fused, concatenating the features as in Eq. (3), where each syllable is represented by a row. Hence, each row contains 36 coefficients, and the full matrix represents the coefficients extracted for all syllables of a species. Thus, a broad spectral representation of a call is used as input to the classification phase.

$$Features = \begin{pmatrix} MFCC_1 & LFCC_1 \\ \dots & \dots \\ MFCC_n & LFCC_n \end{pmatrix} \tag{3}$$

3.3. Classification

To validate the robustness of the proposed methodology based on cepstral coefficients fusion, three machine learning algorithms have been evaluated in the classification stage: K-nearest neighbor, random forest, and support vector machine.

3.3.1. K-nearest neighbor

KNN was proposed by Cover [30], and infers the new data classification based on the closest training samples. The machine learning algorithm considers the K-nearest point distances to the observation to predict which class is similar. Then, to calculate the class prediction, simple majority of neighbors is used. In this chapter, the number of nearest neighbors has been fixed to $k = \sqrt{N}$, where N denotes the length of the cepstral coefficients.

3.3.2. Random forest

RF is a machine learning algorithm presented by Breiman [31]. It is able to model non-lineal input variables, and in addition, it is robust to outliers in the training dataset. RF is an ensemble of decision trees. The generalization error converges to a limit when the number of trees in the forest becomes large. An average of the output votes from all the trees in the forest is computed for the prediction of the classes, Eq. (4). In this study, a value of K = 200 trees was utilized because it returns better results, with predictor variables $m = \sqrt{N}$, where N is the length of the cepstral coefficients.

$$Prediction = \frac{1}{K}\sum_{i=1}^{K} y_{i}, \quad where\ y_n\ is\ the\ nth\ tree\ response \qquad (4)$$

3.3.3. Support vector machine

SVM [32] is a robust supervised learning technique and has been used to resolve the acoustic signal classification. The aim is to create non-overlapping partitions mapping the data as elements of a higher-dimensional space. SVM computes the classification of geometric parameters getting the optimal hyperplane from the training data which separates the data perfectly into two classes. Nevertheless, sometimes the training data cannot be separated lineally. In those cases, and in order to divide the classes, a non-linear kernel function is used to project the data into a higher dimensional space. In this chapter, an implementation based on LIBSVM library [33] was used implementing a C-Support Vector Classification (C-SVC) [34], which uses a decision function as showed in Eq. (5), where K is a radial basis function (RBF) kernel, $k(x, x') = e^{\left(-c\|x-x'\|^2\right)} k(x, x') = e^{\left(-\gamma\|x-x'\|^2\right)}$. In order to carry out the multiclass classification, the strategy "one-versus-one" is performed generating one SVM for each pair of classes. Thus, for N different classes, $N(N-1)/2$ classifiers are necessary to identify the samples.

$$f(x) = \text{sign}\left(\sum_{i=1}^{l} y_i \alpha_i K(x, x') + b\right), \ \ y_i \in \{1, -1\} \qquad (5)$$

Lastly, a grid-search was implemented to adjust the SVM parameters ($\gamma = 2^{-12}, 2^{-11}, \dots, 2^2$; C = $2^{-2}, 2^{-1}, \dots, 2^{10}$) using cross-validation to find the optimum kernel gamma parameter, γ, and the value of the penalty parameter of the error term (C). The values obtained for the kernel gamma were 0.45 and 1.45 for reptile and anuran corpus, respectively. For the penalty error term, the values were 30 and 20.

4. Experimental methodology and results

In this section, the datasets and the experimental results obtained from experiments carried out to evaluate the effectiveness of the proposed methodology are described and discussed. Experiments were focused onto comparing accuracy using the following different features: MFCC, LFCC, and MFCC/LFCC fusion. The syllables generated by the segmentation phase have been randomly rearranged and split in half—one for training the model and the rest for testing (k-fold cross-validation with k = 2). For each class, the accuracy has been evaluated as in Eq. (6), and then the results have been averaged. Using the feature with the best accuracy results, experiments varying the training size were also carried out, from 5 to 50% of the full dataset. The aim is to validate the performance and the robustness of proposed methodology.

In order to validate the experimental results, and to ensure statistical independence, all experiments have been repeated 100 times. The acoustic classification system was implemented in Matlab, and two classifiers were used for each dataset: KNN and SVM classifiers for reptile identification, and RF and SVM classifiers for anuran identification. The experiments were run in a non-dedicated Windows machine based on an Intel Core i7 4510 with a clock speed of 2 GHz, and 16 GB of RAM.

$$Accuracy = \frac{Syllables\ Correctly\ Identified}{Total\ Number\ of\ Syllabes} \times 100 \qquad (6)$$

4.1. Datasets

Two different datasets have been built to validate the proposed methodology in this chapter. Each dataset contains audio content of anurans and reptiles, respectively.

4.1.1. Anurans dataset

The following three databases of anurans have been used to build the anurans dataset: the AmphibiaWeb database [35], a compilation of audio recordings of the amphibians of Cuba [36] and a sound guide of frogs and toads from southern Brazil and Uruguay [37]. AmphibiaWeb was created by the University of California (Berkley), where on-line information related to amphibian conservation and biology is stored. The recordings contain significant background noise and were mainly gathered in their own habitats. In addition, the signals were recorded with different sample formats and rates. From this database, a total of 41 anurans of several taxonomy families were selected, where most of them are anurans from previous literature studies [27, 28]. On the other hand, the collection of amphibians of Cuba contains 99 recordings of several types of advertisement and alert calls of 58 species, most of them endemic. Finally, the sound guide from Brazil and Uruguay is composed by 109 frogs and toads. From them, nine species have been rejected because they do not have enough samples to fit and test the model. Hence, a total of 199 species compose the whole anurans dataset. **Table 2** shows the number of segmented syllables grouped by taxonomic family.

Dataset	Family	Number of Species	Number of syllables
AmphibiaWeb	Bufonidae	6	270
	Dendrobatidae	2	36
	Hemiphractidae	1	34
	Hylidae	9	309
	Hyperoliidae	2	84
	Leptodactylidae	3	110
	Mantellidae	7	241
	Microhylidae	2	52
	Myobatrachidae	6	239
	Ranidae	1	19
	Scaphiopodidae	2	170
Cuba	Bufonidae	10	1141
	Eleutherodactylidae	42	2951
	Hylidae	4	737
	Ranidae	2	372
Brazil and Uruguay	Alsodidae	1	210
	Bufonidae	10	1500
	Brachycephalidae	2	124
	Centrolenidae	1	33
	Cycloramphidae	2	46
	Hemiphractidae	1	32
	Hylidae	49	4633
	Hylodidae	5	353
	Leptodactylidae	23	2971
	Microhylidae	1	54
	Odontophrynidae	5	914

Table 2. Anurans dataset.

4.1.2. Reptiles dataset

Sound repositories of reptiles is quite limited, because they have not been acoustically and exhaustively analyzed. Thus, reptile recordings form three Internet sound collections have been extracted to build the dataset. The Animal Sound Archive at the Museum für Naturkunde in Berlin [38] was the principal source of reptile audio recordings. It stores 120,000 tracks of diverse species which are freely available from their database. The second collection used was California Herps [39] which contains some Squamata sounds. Finally, a

small number of tortoise vocalizations from the California Tortoise Club [40] collection was added to the dataset. Therefore, the whole dataset used in this work is formed by 1895 samples matching to 27 different reptile species and six family groups. **Table 3** shows the number of segmented syllables grouped by taxonomic family.

4.2. Analysis of accuracy

4.2.1. Anurans results

Table 4 indicates the accuracy results for each set of features and the computation time for training and testing by iteration. As it can be observed, Mel coefficients perform better results than LFCC when the number of anurans is small. Nevertheless, when it is increased, LFCC shows a superior performance because higher frequencies are better characterized. Hence, a MFCC and LFCC fusion is proposed to characterize the anuran sounds in lower as well as higher frequencies. The experiments reinforce that this approach improves the classification rate on all databases and the aggregate dataset. As it can be appreciated, RF is clearly outperformed by SVM in all experiments. Furthermore, a successful classification with an accuracy above 95% using the aggregate dataset was achieved. Regarding the training time, RF takes more computation time than SVM. Nevertheless, RF is clearly faster when testing is carried out. It is more noticeable when the species number increases.

A detailed analysis indicates that an accuracy of 98.70% was achieved for AmphibiaWeb database, outperforming other research in terms of number of species identified and accuracy [21, 22, 27, 28, 41–44]. Furthermore, 100% accuracy was reached for 24 anurans. On the other hand, the Cuba database stores some species with a reduced number of syllables, but even in this situation, the features fusion achieved a successful classification improving the accuracy about 5%. An accuracy of 84.90% was the worst result obtained, and a 100% classification rate was reached by 10 species. The mean total accuracy was 96.40% in 58 frog species. Regarding the Brazil–Uruguay dataset, the MFCC and LFCC fusion yielded an identification rate of 95.30% over 100 anurans, where only 16 species achieved an accuracy below 90%. Finally, a success rate of 95.29% was successfully achieved using the aggregate dataset. To the best of our knowledge, it is the largest number of toads and frogs identified using acoustic signals. The proposed methodology in this chapter was compared with some research in literature, **Table 5**. As can be seen, this approach is more robust than other research reaching a higher success rate.

Family	Number of species	Number of syllables
Alligatoridae	3	28
Gekkonidae	2	215
Helodermatidae	1	383
Viperidae	12	950
Elapidae	1	10
Testudinidae	8	309

Table 3. Reptiles dataset.

Database	Features	Classifier	Training Time(s)	Testing Time (s)	Accuracy ± std
AmphibiaWeb (41anurans)	MFCC	RF	0.68	0.04	96.10% ± 5.69
		SVM	0.11	0.08	97.82% ± 3.21
	LFCC	RF	0.69	0.04	95.83% ± 6.61
		SVM	0.11	0.09	96.81% ± 4.36
	MFCC+LFCC	RF	1.03	0.05	98.00% ± 3.92
		SVM	0.15	0.09	98.70% ± 2.58
Cuba (58 frogs)	MFCC	RF	3.37	0.08	86.08% ± 16.76
		SVM	0.49	0.51	91.64% ± 8.85
	LFCC	RF	3.19	0.08	90.69% ± 10.59
		SVM	0.47	0.49	90.92% ± 10.02
	MFCC+LFCC	RF	4.94	0.08	92.54% ± 9.33
		SVM	0.81	0.57	96.40% ± 4.03
Brazil and Uruguay (100 anurans)	MFCC	RF	10.13	0.17	84.74% ± 15.28
		SVM	1.73	4.33	90.53% ± 9.57
	LFCC	RF	10.48	0.18	88.03% ± 11.23
		SVM	1.64	4.51	91.69% ± 9,18
	MFCC+LFCC	RF	15.54	0.17	91.18% ± 10.70
		SVM	4.86	5.97	95.30% ± 5.28
AmphibiaWeb+Cuba+Brazil-Uruguay (199 anurans)	MFCC+LFCC	RF	69.78	0.42	90.29% ± 12.85
		SVM	56.4	38.95	95.29% ± 5.63

Table 4. Accuracy results for anurans dataset.

Furthermore, in this work, three public datasets were used, and therefore, this approach can be validated and contrasted.

4.2.2. Reptiles results

Table 6 shows the accuracy results for each set of features and the computation time for training and testing by iteration. As it can be observed, both MFCC and LFCC features obtain similar results. As is known, most of the reptile sounds are from 0.1 to 4 kHz. Therefore, Mel coefficients reinforce the lowest frequencies because those spectrum regions are enhanced. Nevertheless, some reptiles, such as lizards, emit high-frequency components even into the ultrasound range (>20 kHz). MFCC features contain poor information at these frequencies, because the area under the Mel-filter bank grows at higher frequencies. Hence, LFCC are more appropriate to parametrize those reptile sounds. Thus, in some experiments, LFCC surpasses MFCC when the best classifier is used, SVM. The experiments confirm that the MFCC/LFCC data fusion enhances the identification rate. As it be appreciated, SVM slightly outperforms

Reference	Dataset	Features	Classifier	Accuracy (%)
Lee et al. [21]	30 frogs and 19 crickets	MFCC	LDA	96.8 and 98.1
Acevedo et al. [41]	9 frogs and 3 birds from Puerto Rico	Call duration/max. and min. Frequency/max. power/frequency of max. power	SVM	94.95
Chen et al. [44]	18 frogs	Syllable length/MSAS	Template based	94.3
Yuan et al. [27]	8 frogs (AmphibiaWeb)	MFCC	KNN	98.1
Xie et al. [28]	16 frogs from Australia	MFCC	KNN	90.5
In this work	41 anurans (AmphibiaWeb)	MFCC/LFCC	SVM	98.7
	58 frogs from Cuba			96.4
	100 anurans from Brazil-Uruguay			95.3
	199 species from all datasets			95.29

Table 5. State of the art comparison.

Features	Classifier	Training time (s)	Testing time (s)	Accuracy
MFCC	KNN	0.08	0.03	96.00% ± 7.20
	SVM	0.13	0.06	95.84% ± 7.74
LFCC	KNN	0.07	0.03	92.98% ± 9.95
	SVM	0.15	0.05	96.15% ± 5.35
MFCC+LFCC	KNN	0.12	0.05	97.78% ± 3.33
	SVM	0.23	0.06	98.52% ± 3.22

Table 6. Accuracy results for reptiles dataset.

KNN in all experiments. Furthermore, this approach yielded a successful classification with an accuracy above 98%. Regarding the training time, both KNN and SVM take similar computational cost time because the number of reptile species is small.

A detailed analysis reveals that a 100% accuracy was reached in 9 species, regardless of the cepstral coefficients employed. It is due to the spectral distribution of calls in those reptile species is clearly different from others. Nevertheless, the best classification results were achieved by using MFCC/LFCC feature fusion, outperforming both MFCC and LFCC, and independently of the used classifier. Thus, it is confirmed that this methodology achieves a better parametrization of the reptile sounds by keeping in account important information of low- and high-frequency zones. It allows to increase the system accuracy. Finally, it should be noted that the MFCC/LFCC fusion identified 13 species with an accuracy of 100%.

4.3. Analysis of training dataset size

In order to validate the robustness of this methodology, the efficiency of the system was tested by varying the training dataset size from 5 to 50%. All experiments were carried out by using the MFCC/LFCC features fusion and SVM as classifier.

4.3.1. Anurans results

Table 7 shows the experimental results varying the training size of the whole dataset of anurans, 199 species. As can be seen, larger training datasets are useful to improve the performance of this approach. In addition, an accuracy above 90% is yielded using only a 20% training size. It should be noted that the recordings of some species have very few syllables, even with only three samples. Therefore, when training size is considerably reduced, the classifier is modeled with only one sample. Precision, recall, and F-Measure measurements have been also computed by varying the training dataset size. As shown, these measurements follow a similar behavior related with accuracy, increasing when the training size does, and keeping values above 0.9 using only a 20% training size and close to 0.95 using a 40% of training size. Thus, small training datasets allow to make less the time needed and the computational cost to calculate the classifier model. This evidence that the fusion of both MFCC and LFCC features is efficient for modeling the discriminant information in the anurans sounds. Furthermore, the data fusion allows to obtain classification results above 80% in all cases, demonstrating the robustness of the feature fusion method.

4.3.2. Reptiles results

Table 8 shows the experimental results varying the training size of the whole dataset of reptiles, 27 species. As can be seen, the system accuracy increases when the training size does, and the proposed methodology can obtain good results with a low number of training samples. Thus, when the number of samples is close to 5%, this approach decreases in effectiveness, but even in these circumstances, the system yields an accuracy above 85%, keeping in mind that only one syllable characterizes most of the reptile species. For other training cases, the accuracy is above 90%. Furthermore, precision, recall, and F-Measure measurements also follow a similar behavior related with accuracy, that is, when the training size increases, the measurements also do, keeping values close to 0.9 using only a 5% training size and close to

Training size (%)	Accuracy (%) ± std	Precision	Recall	F-Measure
5	80.01% ± 18.05	0.86	0.80	0.83
10	86.94% ± 12.94	0.91	0.87	0.89
20	91.38% ± 8.71	0.94	0.91	0.92
30	93.53% ± 7.14	0.95	0.93	0.94
40	94.48% ± 6.72	0.96	0.94	0.95
50	95.29% ± 0.16	0.96	0.95	0.96

Table 7. Classifier performance with different training size for anurans dataset.

Training size (%)	Accuracy (%) ± std	Precision	Recall	F-Measure
5	85.50% ± 20.06	0.91	0.85	0.88
10	91.03% ± 14.06	0.94	0.91	0.92
20	94.81% ± 8.01	0.96	0.94	0.95
30	96.86% ± 5.39	0.97	0.96	0.97
40	97.88% ± 3.76	0.98	0.97	0.98
50	98.52% ± 3.26	0.98	0.98	0.98

Table 8. Classifier performance with different training size for reptiles dataset.

0.97 using a 30% of training size. Furthermore, a lower training dataset size offers savings in computational cost and time needed to compute the classifier model. This evidences that the fusion of both cepstral coefficients can be used in an effective way for discerning important information in the reptile sounds. Hence, the data fusion achieves the classification results above 85% in all cases, validating the robustness of the MFCC/LFCC features fusion.

5. Conclusions

Automatic species identification based on bioacoustic information has become an attractive research topic due to growing interest among biologists for sampling populations and controlling the conservation of these living beings in large and remote areas where environmental conditions and visibility are limited. In this chapter, a methodology based on the fusion of cepstral coefficients, MFCC and LFCC, was proposed and validated using public datasets of reptile and anuran species. This data fusion allows to characterize the acoustic signal with both low- and high-frequency components, being more robust against noise and increasing the classification rate. The results of the proposed methodology are encouraging with a mean accuracy of 95.29 and 98.52% for anurans and reptiles, respectively.

Regarding the anurans identification, the proposed methodology was collated with some research of literature, being more robust and identifying more species than the other techniques. Furthermore, public databases have been used, and therefore, this approach can be validated and contrasted. On the other hand, as far as authors know, the anurans dataset contains the largest number of toads and frogs automatically identified by acoustic characteristics. For reptile identification, the authors are not keeping in mind about other researches that have considered the use of reptile acoustic signals for species classification. Even so, the experimental results have demonstrated that the MFCC/LFCC feature fusion achieves a broad characterization of the acoustic signal, yielding a high identification rate.

Finally, the proposed methodology described in this chapter has been analyzed using scenarios with reduced training dataset, validating the robustness of the system. It declines in effectiveness when the training dataset size decreases, but even so, with only a 5% of the samples for training, this approach yields an accuracy above 80%, keeping in mind that many species are only characterized by only one syllable.

Author details

Juan J. Noda, David Sánchez-Rodríguez* and Carlos M. Travieso-González

*Address all correspondence to: david.sanchez@ulpgc.es

Institute for Technological Development and Innovation in Communications, University of Las Palmas de Gran Canaria, Las Palmas de Gran Canaria, Spain

References

[1] DiMattina C, Wang X. Virtual vocalization stimuli for investigating neural representations of species-specific vocalizations. Journal of Neurophysiology. 2006;**95**(2):1244-1262. DOI: 10.1152/jn.00818.2005

[2] Ziegler L, Arim M, Narins PM. Linking amphibian call structure to the environment: The interplay between phenotypic flexibility and individual attributes. Behavioral Ecology. 2011;**22**(3):520-526. DOI: 10.1093/beheco/arr011

[3] Murray S, Mercado E, Roitblat H. The neural network classification of false killer whale (*Pseudorca crassidens*) vocalizations. The Journal of the Acoustical Society of America. 1998;**104**(6):3626-3633. DOI: 10.1121/1.423945

[4] Ganchev T, Potamitis I, Fakotakis N. Acoustic monitoring of singing insects. In: IEEE International Conference on Acoustics, Speech and Signal Processing; 16–20 April 2007; p. IV-721-IV-724

[5] Fagerlund S. Bird species recognition using support vector machines. EURASIP Journal of Advances in Signal Processing. 2007:1-8. DOI: 10.1155/2007/38637

[6] Turesson HK, Ribeiro S, Pereira DR, Papa JP, de Albuquerque VHC. Machine learning algorithms for automatic classification of marmoset vocalizations. PLoS One. 2016;**11**(9): e0163041. DOI: 10.1371/journal.pone.0163041

[7] Mielke A, Zuberbühler K. A method for automated individual, species and call type recognition in free-ranging animals. Animal Behaviour. 2013;**86**(2):475-482. DOI: 10.1016/j.anbehav.2013.04.017

[8] Agranat I. Bat Species identification from zero crossing and full spectrum echolocation calls using hidden Markov models, fisher scores, unsupervised clustering and balanced winnow pairwise classifiers. In: Proceedings of Meetings on Acoustics ICA2013; ASA; 2013. p. 010016

[9] Noda JJ, Travieso CM, Sánchez-Rodríguez D. Automatic taxonomic classification of fish based on their acoustic signals. Applied Sciences. 2016;**6**(12):443. DOI: 10.3390/app6120443

[10] Sattar F, Cullis-Suzuki S, Jin F. Identification of fish vocalizations from ocean acoustic data. Applied Acoustics. 2016;**110**:248-255. DOI: 10.1016/j.apacoust.2016.03.025

[11] Clemins P, Johnson M. Application of speech recognition to African elephant (Loxodonta Africana) vocalizations. In: Acoustics, speech, and signal processing, 2003. Proceedings (ICASSP'03); IEEE; 2003. p. I-484-I-487

[12] Gillespie D, Caillat M, Gordon J. Automatic detection and classification of odontocete whistles. The Journal of the Acoustical Society of America. 2013;**134**(3):2427-2437. DOI: 10.1121/1.4816555

[13] Somervuo P, Harma A, Fagerlund S. Parametric representations of bird sounds for automatic species recognition. IEEE Transactions on Audio, Speech, and Language Processing. 2006;**14**(6):2252-2263. DOI: 10.1109/TASL.2006.872624

[14] Adi K, Johnson M, Osiejuk T. Acoustic censusing using automatic vocalization classification and identity recognition. The Journal of the Acoustical Society of America. 2010;**127**(2):874-883. DOI: 10.1121/1.3273887

[15] Potamitis I, Ntalampiras S, Jahn O, Riede K. Automatic bird sound detection in long real-field recordings: Applications and tools. Applied Acoustics. 2014;**80**:1-9. DOI: 10.1016/j.apacoust.2014.01.001

[16] Zhao Z, Zhang S, Xu Z, Bellisario K, Dai N, Omrani H, Pijanowski B. Automated bird acoustic event detection and robust species classification. Ecological Informatics. 2017;**39**: 99-108. DOI: 10.1016/j.ecoinf.2017.04.003

[17] Gary Y, Fu Q. Automatic frog call monitoring system: A machine learning approach. Applications and Science of Computational Intelligence V; International Society for Optics and Photonics. 2002:188-200

[18] Huang C, Yang Y, Yang D, Chen Y. Frog classification using machine learning techniques. Expert Systems with Applications. 2009;**36**(2):3737-3743. DOI: 10.1016/j.eswa.2008.02.059

[19] Han N, Muniandy S, Dayou J. Acoustic classification of Australian anurans based on hybrid spectral-entropy approach. Applied Acoustics. 2011;**72**(9):639-645. DOI: 10.1016/j.apacoust.2011.02.002

[20] Gingras B, Fitch W. A three-parameter model for classifying anurans into four genera based on advertisement calls. The Journal of the Acoustical Society of America. 2013;**133**(1):547-559. DOI: 10.1121/1.4768878

[21] Lee C, Chou C, Han C, Huang R. Automatic recognition of animal vocalizations using averaged MFCC and linear discriminant analysis. Pattern Recognition Letters. 2006;**27**(2): 93-101. DOI: 10.1016/j.patrec.2005.07.004

[22] Jaafar H, Ramli DA, Rosdi BA, Shahrudin S. Frog identification system based on local means k-nearest neighbors with fuzzy distance weighting. In: International Conference on Robotic, Vision, Signal Processing & Power Applications; Lecture Notes in Electrical Engineering, vol 291. Springer, Singapore; 2014. p. 153-159

[23] Colonna J, Peet T, Abreu C, Jorge A, Ferreira E, Gam J. Automatic classification of anuran sounds using convolutional neural networks. In: Proceedings of the Ninth International

C* Conference on Computer Science & Software Engineering; 20–22 July 2016; Portugal; ACM; 2016. p. 73-78

[24] Strout J, Rogan B, Mahdi Seyednezhad S.M, Samrt K, Ush M, Ribeiro E. Anuran call classification with deep learning. In: Acoustics, Speech and Signal Processing (ICASSP), 5–9 March 2017; IEEE; 2017. p. 2662-2665

[25] Hassan N, Athiar D, Jaafar H. Deep neural network approach to frog species recognition. In: Signal Processing & its Applications (CSPA), 10–12 March 2017; IEEE; 2017. p. 173-178

[26] Härmä A. Automatic identification of bird species based on sinusoidal modeling of syllables. In: IEEE International Conference on Acoustics, Speech, and Signal Processing; Hong Kong, China. 6–10 April 2003; p. V545-V548

[27] Yuan C, Ting L, Athiar D. Frog sound identification system for frog species recognition. In: International Conference on Context-Aware Systems and Applications; Berlin. Springer; 2012. p. 41-50

[28] Xie J, Towsey M, Truskinger A, Eichinski P, Zhang J, Roe P. Acoustic classification of Australian anurans using syllable features. In: IEEE Tenth International Conference on Intelligent Sensors, Sensor Networks and Information Processing; 7–9 April 2015; Singapore. p. 1-6

[29] Feng AS, Narins PM, Xu C-H, Lin W-Y, Yu Z-L, Qiu Q. Ultrasonic communication in frogs. Nature. 2006;**440**:333-336. DOI: 10.1038/nature04416

[30] Cover TM, Hart PE. Nearest neighbor pattern classification. IEEE Transactions on Information Theory. 1967;**13**:21-27. DOI: 10.1109/TIT.1967.1053964

[31] Breiman L. Random forests. Machine learning. 2001;**45**(1):5-32. DOI: 10.1023/A:101093

[32] Burges CJ. A tutorial on support vector machines for pattern recognition. Data Mining and Knowledge Discovery. 1998;**2**:121-167. DOI: 10.1023/A:100971

[33] Chang CC, Lin CJ. LIBSVM: A library for support vector machines. ACM Transactions on Intelligent Systems and Technology. 2011;**2**(3):27. DOI: 10.1145/1961189.1961199

[34] Boser BE, Guyon IM, Vapnik VN. A training algorithm for optimal margin classifiers. In: Fifth Annual Workshop on Computational Learning Theory; 27–29 July 1992; Pittsburg. p. 144-152

[35] AmphibiaWeb [Internet]. University of California, Berkeley. Available from: https://amphibiaweb.org/ [Accessed: November 14, 2017]

[36] Alonso R, Rodríguez A, Márquez R. Sound guide of the amphibians from Cuba (audio cd & booklet). ALOSA Sons de la Natura, Barcelona. 2007:1-46

[37] Kwet A, Márquez R. Sound guide of the calls of frogs and toads from southern Brazil and Uruguay. Fonoteca, Madrid, Double CD and Booklet. 2010:1-55

[38] Berlin Natural Museum. The Animal Sound Archive [Internet]. Available from: http://www.tierstimmenarchiv.de [Accessed: October 15, 2017]

[39]   California Reptiles and Amphibians [Internet]. Available from: http://www.californiaherps. com [Accessed: October 11, 2017]

[40]   California Turtle and Tortoise Club [Internet]. Available from: http://www.tortoise.org [Accessed: October 16, 2017]

[41]   Acevedo M, Corrada-Bravo C, Corrada Bravo H, Villanueva-Rivera L, Aide M. Automated classification of bird and amphibian calls using machine learning: A comparison of methods. Ecological Informatics. 2009;**4**(4):206-214. DOI: 10.1016/j.ecoinf.2009.06.005

[42]   Bedoya C, Isaza C, Daza J, López J. Automatic recognition of anuran species based on syllable identification. Ecological Informatics. 2014;**24**:200-209. DOI: 10.1016/jecoinf.2014.08.009

[43]   Brandes TS. Feature vector selection and use with hidden Markov models to identify frequency-modulated bioacoustic signals amidst noise. IEEE Transactions on Audio, Speech, and Language Processing. 2008;**16**(6):1173-1180. DOI: 10.1109/TASL.2008.925872

[44]   Chen W-P, Chen S-S, Lin C-C, Chen Y-Z, Lin W-C. Automatic recognition of frog calls using a multi-stage average spectrum. Computers & Mathematics with Applications. 2012;**64**(5):1270-1281. DOI: 10.1016/j.camwa.2012.03.071