Expert Systems With Applications 50 (2016) 100–106

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Methodology for automatic bioacoustic classification of anurans based on feature fusion

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

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

ARTICLE INFO

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

ABSTRACT

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

1. Introduction

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

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

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

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

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

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

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

2. Related work

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

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

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

3. Proposed methodology

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

3.1. Segmentation

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

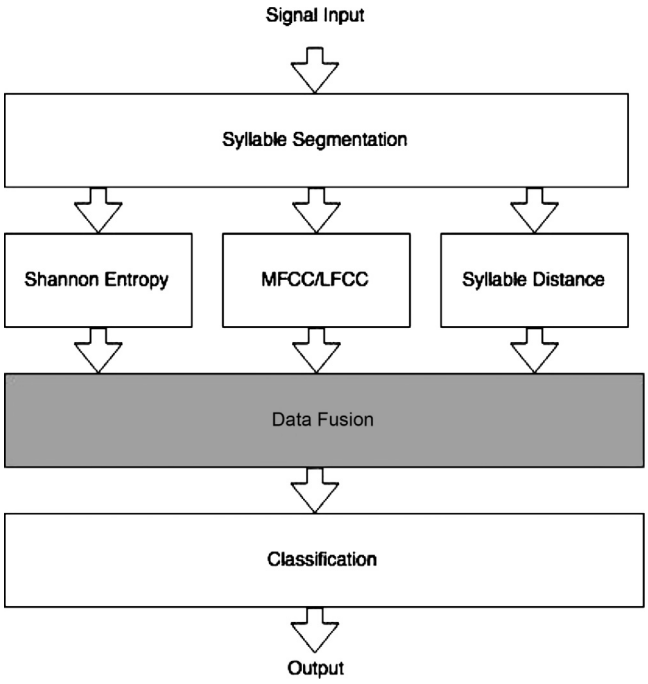

Fig. 1. System diagram.

time-frequency trade-off (Semmlow & Griffel, 2014), but for experimental purposes it has been kept constant due to having an advantage of speed when analyzing large dataset. They were established heuristically using a small subset of species and doing a blind test for the rest of the data. The resulting spectrogram is defined as a time-frequency matrix $S(f, t)$, where f is the frequency and t represents the time. Then, in each iteration the highest amplitude in the spectrogram $|S(f_n, t_n)|$ is located and its amplitude is calculated as (1). The matrix is explored from this point for $t > t_n$ and $t < t_n$ until $Y_n(t_n - t_0) < Y_n(0) - 20dB$, where 20 dB is the stopping criteria. When the edges of the syllable are reached the trajectory is stored and deleted from the matrix. The algorithm repeats this process until the end of the spectrogram.

$$Y_n(0) = 20log_{10}(|S(f_n, t_n)|) \tag{1}$$

3.2. Feature extraction

Once each audio recording file has been segmented, a group of characteristics are extracted per each resultant syllable in order to allow a suitable taxonomy classification. We proposed an acoustic parametrization technique of time-variable features, entropy and call duration, in conjunction with spectral features LFCC and MFCC cepstral representation of the sound.

MFCCs have been applied in automatic bioacoustic identification of frogs with varied success (Jaafar et al., 2014; Xie et al., 2015; Yuan & Ramli, 2013). Amphibians as well as humans heard lower frequencies better, Mel scale was designed precisely to emulate this human hearing response emphasizing lower frequencies. However, anurans are capable to articulating and hearing sounds above 20 kHz in the ultrasonic spectra (Feng et al., 2006). Therefore, in order to characterize this behavior, LFCCs have also been extracted to model the information of the high frequency regions (Zhou, Garcia-Romero, Duraiswami, Espy-Wilson, & Shamma, 2011). Both MFCC and LFCC are similarity computed based on short time analysis with the only difference of the Mel frequency scale transformation step as in (2), performed by a filter bank of 26 triangular band-pass filters. To calculate the coefficients, we have established a window size of 25 ms with an overlap of 50% on each frame. This value was determined experimentally by varying the window size from 10 ms to 1 s and evaluating the results. For MFCCs, the final vectors are the result of taking the lowest Discrete Cosine Transform (DCT) coefficients from the expression (3), where j denotes the index of the cepstral, B the number of triangular filters and N the number of coefficients to compute. Besides, X_i is the output response from the ith triangular filter. In this study, 18 coefficients have been retained for both features. Finally, the features of all frames within one syllable have been averaged individually for MFCCs and LFCCs.

$$m = 2595\log_{10}\left(\frac{f}{700} + 1\right) \tag{2}$$

$$C_j = \sum_{i=1}^{B}(\log|X_i|\cos\left[j\left(i - \frac{1}{2}\right)\right]\pi/B), 0 \le j \le N - 1 \tag{3}$$

With regard to time variable features, Shannon entropy (SE) defined as (4) is a measure that quantifies the randomness of the syllable providing information of the degree of voicing which has been used with success in voice detection (Renevey & Drygajlo, 2001). Entropy is uncorrelated with the frequency domain features as it represents the vocal source better, whereas MFCCs and LFCCs are more related to the vocal tract. For that reason, it has been selected to obtain a metric of the distinctive power variations inside the syllables. Additionally, syllable length (SL) has been adopted as a feature as in Acevedo et al. (2009); Chen et al. (2012); Xie et al. (2015) because in most frog species the call duration is different so it can be utilized for identification purposes.

$$SE(A) = -\sum_{i=1}^{n} p(a_i)\log_2 p(a_i), \quad \text{where } p(a_i) \text{ is the probability of } a_i \in A \tag{4}$$

After performing the feature extraction, a data fusion technique was applied to obtain a robust acoustic representation of each sound. In this work, we have fused MFCCs, LFCCs, SE and syllable length (SL) vectors forming a matrix where each row corresponds to a segmented syllable. The fusion is implemented concatenating the feature vectors horizontally as shown in (5). The number of features was experimentally calculated to be 38 coefficients (18 MFCCs + 18 LFCCs + 1 SE + 1 SL) by each vector. Thus, the number of coefficients is relatively small and they are almost independent of each other. Then, these vectors are used as input data for the classification stage. These sounds' attributes have been combined to hold information of higher and lower frequency regions and time variable characteristics instead of using only one of them as in previous state of the art works. This fusion approach may be used over other animal groups with similar acoustic characteristics such as insect or bird calls. However, for bird songs further research is required due to they present a more complex sound structure.

$$FUSION = \begin{pmatrix} MFCC_1 & LFCC_1 & SE_1 & SL_1 \\ \vdots & \dots & \dots & \vdots \\ MFCC_n & LFCC_n & SE_n & SL_n \end{pmatrix} \tag{5}$$

4. Classification system

In order to carry out the taxonomy identification, the performance of three machine learning algorithms has been compared. The following sections contain the implementation details of the algorithms used in this study.

4.1. Random Forest

RF (Breiman, 2001) is a machine learning algorithm able to model non-lineal input variables and robust to outliers in the

training samples. It bundles a bunch of randomly generated decision trees (DT) finding the best splinting for each subset of training features, so every DT is fit to a bootstrapped random training set. The prediction of the classes is made averaging the output votes from all the trees in the forest as is shown in (6). In this work, we have utilized $K = 200$ trees to characterize the toads' and frogs' calls, with predictor variables $m = \sqrt{N}$ where N is the length of the feature coefficients in a syllable.

$$Prediction = 1/K \sum_{n=1}^{K} y_i, \text{ where } y_n \text{ is the } n\text{th tree response} \quad (6)$$

4.2. Support Vector Machine

SVM (Burges, 1998) is a robust supervised learning technique which maps the data as elements of a higher dimensional space to create non-overlapping partitions. Ideally, SVM solves the classification of geometric parameters obtaining the optimal hyperplane from the training data which separates the data perfectly into two classes. However, there are situations in which the training data cannot be separated lineally. In this case, it is necessary to use a non-linear kernel function to project the data into a higher dimensional space to divide the classes. For the experiments, we have used an implementation based on libsvm (Chang & Lin, 2011) applying a C-Support Vector Classification (C-SVC) (Boser, Guyon, & Vapnik, 1992) which presents a decision function as defined in (7), where K in this work is a RBF kernel, $K(x, x') = \exp(-c\|x - x'\|^2)$, with $c = 1.5$. The multiclass classification is performed under the strategy "one-versus-one" generating one SVM for each pair of classes, therefore for N classes $N(N-1)/2$ classifiers are required to distinguish the samples.

$$f(x) = \text{sign}\left(\sum_{i=1}^{l} y_i \alpha_i K(x, x') + b\right), y_i \in \{1, -1\} \quad (7)$$

4.3. Hidden Markov Model

HMM (Rabiner, 1989) is an statistical model defined by two associated stochastic processes, an underlying process characterized by a string of N states which are not visible and an observation process that takes values in an alphabet of size M. The Markov chain is modeled by the probability of transition between hidden states, $A = \{a_{ij}\}$, $1 \le i, j \le N$. The observation process is modeled by the probability of obtaining an observed value given a hidden state $B = \{b_j(k)\}$, $1 \le j \le N; 1 \le k \le M$. Moreover, the initial state probability of the system is denoted as $\pi = \{\pi_i\}$ of size N. Thus, HMM is defined by $\lambda = [A, B, \pi]$ and the probability of an observation sequence V given λ can be calculated as shown in (8), where $S = [s_1, s_2, \ldots, s_N]$ represents the variables for the hidden states.

$$P(V/\lambda) = \sum_{S} \prod_{i=1:N} P(V_i/S_i, A) \cdot P(S_i/S_{i-1}, B) \quad (8)$$

In this paper, we have worked with a Bakis HMM also called left to right which is particularly appropriate for sequential sound data because the transitions between states are produced in a single direction (Rabiner, 1989). In particular, the time sequences for this study are toad and frog calls. In this case, the number of states, N, was swept between 2 and 100 states, in order to seek the best performance. Similarly, the number of observed symbols, M, was determined by experimentation. We have finally selected 23 states and 32 symbols in the experiments. The algorithm has been implemented using the toolbox defined in David, Ferrer, Travieso, Alonso, and y Comunicaciones (2004).

Table 1
Amphibiaweb database selected families.

No. species	Family
6	Bufonidae
2	Dendrobatidae
1	Hemiphractidae
9	Hylidae
2	Hyperoliidae
3	Leptodactylidae
7	Mantellidae
2	Microhylidae
6	Myobatrachidae
1	Ranidae
2	Scaphiopodidae

5. Databases

Three databases have been used in this work to validate the methodology: the AmphibiaWeb database (AmphibiaWeb, 2015), a sound guide of amphibians of Cuba (Alonso, Rodríguez, & Márquez, 2007) and a guide of the calls of frogs and toads from southern Brazil and Uruguay (Kwet & Márquez, 2010). AmphibiaWeb is an on-line database created by the University of California (Berkley) which contains information relative to amphibian biology and conservation. It offers access to reviews, audio and video content from a rich number of species to divulge taxonomy useful data to their conservation. The audio records were mainly recorded in their own habitats with significant background noise. In addition, the signals were saved with different sample rates and sample formats (bit depth). We have selected 41 anurans from this database of several taxonomy families, trying to choose anurans from previous state of the art works (Xie et al., 2015; Yuan & Ramli, 2013). Table 1 shows a list with the number of species by family which have been selected for the experiments, mainly from Australia and Madagascar.

The sound guide of the amphibians of Cuba includes the calls of 58 different species of which 55 are endemic of Cuba and belonging to the Bufonidae or Eleutherodactylus family. The guide is composed by 99 recordings which include several types of alert calls, advertisement and some chorus sounds.

Finally, the audio guide from Brazil and Uruguay contains 109 frogs and toads, we have selected 100 anurans from this audio guide discarding 9 species as they do not present enough number of samples to train and test the classification algorithms. The recordings are divided into two compact discs, the first contains 55 frogs belonging to the Hylidae family and the second 54 of 10 different families. In addition, some tracks contain multiple recordings of diverse call types. These databases give us a total of 199 species in the dataset.

6. Experimental methodology

The following experiments have been designed in order to explore the effectiveness of our feature data fusion approach for bioacoustic recognition. The proposed combination of characteristics is compared with the output of the system using the features individually to train the classification algorithm as in previous works. The experiments have been repeated 50 times for SVM and RF, and only 10 for HMM. The algorithms were run in an i7 Intel non-dedicated Windows computer with 8GB of RAM and 2GB of dedicated graphic memory. In addition, the total number of segmented syllables is 1564, 5201 and 10905 correspondingly to the datasets AmphibiaWeb, Cuba and Brazil-Uruguay. The feature vectors of these syllables have been split 50/50 for training and testing, shuffling the data in each iteration to obtain statistically significant results. The classification rates presented in the results are expressed as the mean accuracy per each class. Table 2 lists the

Table 2
Identification results due to experimentation.

Database	Features	Classification	Training time(sg)	Testing time(sg)	Accuracy mean % ± std
AmphibiaWeb(41 anurans)	MFCC	HMM	24.55	78.89	74.82% ± 24.07
		RF	0.68	0.04	96.10% ± 5.69
		SVM	0.11	0.08	97.82% ± 3.21
	LFCC	HMM	120	24.49	78.78% ± 20.68
		RF	0.69	0.04	95.83% ± 6.61
		SVM	0.11	0.09	96.81% ± 4.36
	MFCC+LFCC	HMM	187	39.26	86.85% ± 15.79
		RF	1.03	0.05	98.00% ± 3.92
		SVM	0.15	0.09	98.70% ± 2.58
	MFCC+LFCC+SE+SL	HMM	163	39.13	93.33% ± 10.51
		RF	1.39	0.07	98.27% ± 3.61
		SVM	0.16	0.09	**98.80% ± 2.43**
Cuba(58 frogs)	MFCC	HMM	349	101	55.43% ± 30.67
		RF	3.37	0.08	86.08% ± 16.76
		SVM	0.49	0.51	91.64% ± 8.85
	LFCC	HMM	359	100	62.77% ± 30.56
		RF	3.19	0.08	90.69% ± 10.59
		SVM	0.47	0.49	90.92% ± 10.02
	MFCC+LFCC	HMM	610	159	71.66% ± 26.81
		RF	4.94	0.08	92.54% ± 9.33
		SVM	0.81	0.57	96.40% ± 4.03
	MFCC+LFCC+SE+SL	HMM	474	173	77.16% ± 24.31
		RF	5.03	0.08	93.33% ± 8.10
		SVM	0.81	0.57	**96.90% ± 3.57**
Brazil and Uruguay(100 anurans)	MFCC	HMM	712	253	52.91% ± 25.47
		RF	10.13	0.17	84.74% ± 15.28
		SVM	1.73	4.33	90.53% ± 9.57
	LFCC	HMM	718	260	62.38% ± 21.23
		RF	10.48	0.18	88.03% ± 11.23
		SVM	1.64	4.51	91.69% ± 9.18
	MFCC+LFCC	HMM	1080	440	68.02% ± 24.14
		RF	15.54	0.17	91.18% ± 10.70
		SVM	4.86	5.97	95.30% ± 5.28
	MFCC+LFCC+SE+SL	HMM	1023	473	76.34% ± 20.34
		RF	16.74	0.17	92.73% ± 9.08
		SVM	4.41	5.54	**95.48% ± 4.97**
AmphibiaWeb+Cuba+Brazil- Uruguay(199 anurans)	MFCC+LFCC+SE+SL	SVM	13.39	31.65	**95.38% ± 5.05**

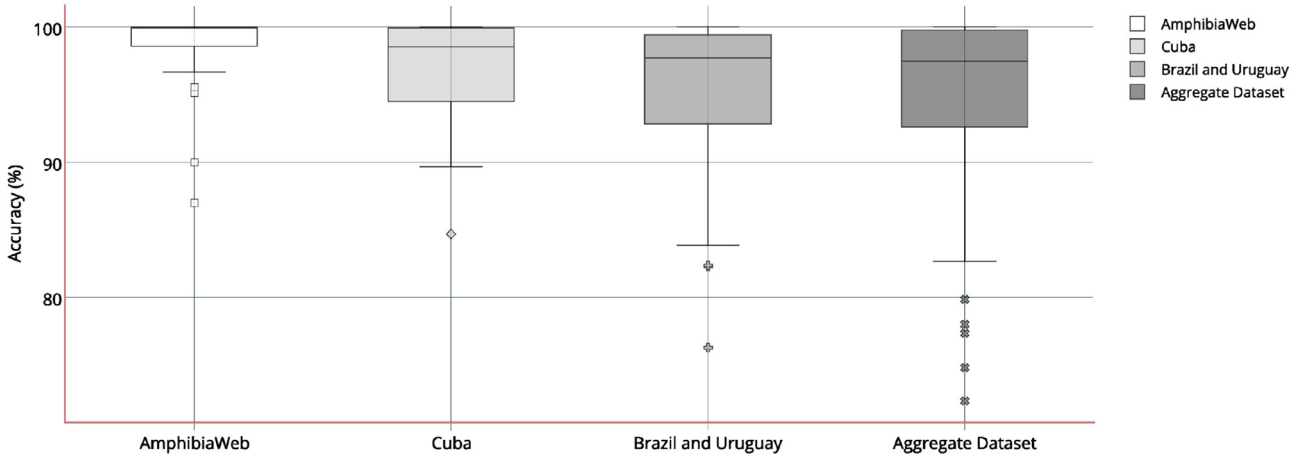

Fig. 2. Detailed accuracy by class for MFCC+LFCC+SE+SL fusion with SVM.

comparative results among the features and the three learning algorithms showing the times of training and testing by iteration, Fig. 2 shows a box plot with the performance evaluation of the proposed method, and finally in Table 3, a comparison is shown among the proposed system and some previous state-of-the-art methods.

6.1. Results and discussions

From Table 2, it can be observed that SVM clearly outperforms HMM in any experiment and slightly overcomes RF. The discrete HMM has serious classification problems in distinguishing anurans with few training syllables and several species presented less than 10 syllables after the segmentation stage. In fact, HMM is incapable of recognizing some species from the Brazil-Uruguay dataset and others present identification rates below 50%. These results are consistent with respect to Lee et al. (2006) where over 30 frogs applying HMM and MFCC only reached a classification rate of 78.9%, in our case 74.82% for 41 anurans. However, HMM gives us the chance to prove how the proposed data fusion technique significantly increases the classification rate showing an improvement by 20% of accuracy with respect to the MFCC approach in each

Table 3
Comparison of the methodology vs. the state-of-the-art.

Reference	Dataset	Segmentation	Parametrization	Classification	Accuracy
(Lee et al., 2006)	30 frogs and 19 cricket	Härmä	MFCC	LDA	96.8% and 98.1%
(Acevedo et al., 2009)	9 frogs and 3 birds from Puerto Rico	Manual	Call duration/Max. and min. frequency/Max. power/Frequency of max. power	SVM	94.95%
(Chen et al., 2012)	18 frogs	Energy and zero-crossing rate	Syllable length/MSAS	Template based	94.3%
(Yuan & Ramli, 2013)	AmphibiaWeb(8 frogs)	Manual	MFCC	KNN	98.1%
(Xie et al., 2015)	16 frogs from Australia	Härmä	MFCC	KNN	90.5%
This work	AmphibiaWeb(41 anurans)	Härmä	MFCC/LFCC/SE/SL fusion	SVM	98.80%
	58 frogs from Cuba				96.90%
	100 anurans Brazil-Uruguay				95.48%
	199 anurans from all datasets				95.38%

database. For instance, it was able to successfully identify a 93.33% of the syllables from the AmphibiaWeb dataset. The training and testing times for HMM are higher by far when compared to SVM and RF, due to HMM is implemented in Matlab while the two others are compiled in C++. RF output values reveal some of the disadvantages of this algorithm. It needs a larger set of training samples to capture additive structures, so it has a poor response to anurans with a low number of syllables. Nevertheless, RF poses the minimum computing times of testing, making it more suitable for implementations in portable field devices. Meanwhile, SVM classifier is able to separate classes more efficiently using the RBF kernel, reaching higher recognition results even with the Mel cesptral coefficients. Even so, it is perceptible that improving the set of training samples with the fusion approach, SVM gains efficiency and robustness for larger dataset.

With regard to features, it can be noticed that MFCC coefficients perform poorer than LFCC almost in all cases, due to anurans are capable of producing sounds at high frequencies. LFCC shows in general a superior performance especially when the number of anurans is increased, owing to its better characterization of higher frequencies but is less robust against the presence of background white noise. As a consequence, a first fusion is proposed by combining the MFCC and LFCC coefficients to parametrize the anuran calls in lower as well as higher frequencies. The tests confirm that the data fusion MFCC–LFCC enhances the system identification rate rising on all classifiers. Finally, a slight improvement in accuracy was achieved adding time variable information with SVM and RF, but there was clear progress at the HMM output because of the time-sequential nature of the data added.

When analyzing the results of the best approach (MFCC-LFCC-SE-SL fusion with SVM) in Fig. 2, it can be appreciated that the system was able to get a successful taxonomy classification in most of the species with an accuracy above 90%. The lower identification rate was 72.33% in the aggregate dataset for Uperoleia mimula belonging to the AmphibiaWeb database, it was misclassified by Hypsiboas pulchellus from the Brazil-Uruguay dataset because both presented a similar frequency distribution. For AmphibiaWeb database, 24 anurans reached an accuracy of 100%, among them 8 frogs used by Yuan and Ramli (2013), because their call frequency distributions are different. On this database, an accuracy of 98.8% was achieved surpassing other state of the art techniques in terms of number of species and classification rate (Acevedo et al., 2009; Bedoya et al., 2014; Brandes, 2008; Chen et al., 2012; Han et al., 2011; Huang et al., 2009; Jaafar et al., 2014; Lee et al., 2006; Xie et al., 2015; Yuan & Ramli, 2013). The database of frogs from Cuba contains various species with few syllables but the data fusion made a successful classification possible. The worst result was 84.9% for track 20, but 10 species reached a classification of 100%, with a total accuracy of 96.90% in 58 frog species. Moreover, on the Brazil–Uruguay database, the data fusion reached an identification rate of 95.48% over 100 anurans with only 16 species below 90%, mostly from tracks with multiple call types. The worst performance was 76.29% for Hypsiboas bischoffi, the recording presents a huge amount of background noise made up of frogs' chorus and singing insects which impedes a proper automatic segmentation. To conclude, the aggregate dataset with 199 different classes of anurans were successfully classified achieving a success of 95.38%. As far as we know, it represents the largest number of frogs and toads automatically identify by acoustic to date. In addition, 36 anurans reached an accuracy of 100% and only 5 of them (Uperoleia mimula, Anaxyrus punctatus, Aplastodiscus cochranae, Hypsiboas bischoffi and Leptodactylus chaquensis) performed poorly between 80 and 70%, as a result of overlapping frequencies among species and heavy background noise in some recordings.

In Table 3, the methodology presented in this work is compared with some references of the state-of-the-art. It shows that the proposed methodology is more robust than other techniques allowing more species to be identified. Moreover, in contrast with other works three public databases have been used so this approach can be contrasted in future works. Table 3 also displays the diverse methods which have been applied in literature to segment the audio files.

7. Conclusions and future work

The ability to locate and classify species in an area with a non-invasive method is extremely useful for biologists and conservationists. In this work, an automatic acoustic classification methodology of anurans through a data fusion of frequency domain and time-variable features has been introduced, which is capable of performing the recognition of these animal sounds. Their acoustic attributes have been analyzed to seek the best set of parameters to represent the anurans' calls. We have concluded that LFCC describes anuran vocalization more efficiently than MFCC due to they can produce sounds in high frequencies like other species. Therefore, we have fused both features to achieve a broad representation of the call frequency information and more robustness against noise. Furthermore, we have added time domain information to characterize the acoustic signal completely which has been particularly useful for improving the results of a time sequential algorithm as HMM. SVM has showed a better performance in the experiments since the acoustic features are non-linear distributed so the Gaussian kernel is able to separate the classes allowing an adequate classification. However, RF presents testing times five times inferior than SVM which can be employed for real time implementations. We have proved that this approach is efficient in recognizing a significant number of anurans discerning successfully 199 different species, the most extended dataset of anuran calls automatically

classify up to now which represents a significant improvement in respect to previous state-of-the-art works. The methodology has been tested in three public domain databases: AmphibiaWeb, a sound guide of amphibians from Cuba and a sound guide of calls of anurans from southern Brazil and Uruguay; achieving a classification success rate of 98.80% ± 2.43, 96.90% ± 3.47 and 95.48% ± 4.97, respectively. Finally, the aggregate dataset has been analyzed reaching 95.38% success. These results suggest that the number of species could be increased without losing much generalization performance, considering that it only decreases by 3.42% from 41 to 199 species. In addition, this solution does not require a huge number of samples per specie performing well with few training syllables. To conclude, the presented methodology is effective for taxonomic classification of anurans and might be applied to recognize individuals from the same species. However, more work should be done in automatic acoustic classification in order to diversify the kinds of species which can be identified by its family and suborder with a high success rate. The fusion technique has been proved effective over simple calls vocalizations, but more complex sounds as bird songs could required a different parametrization. The segmentation stage has a great influence in the system performance due to the presence of outliers in the segmented syllables. Furthermore, an improper windows size selection can lead to an increase in error rate as a consequence of poor resolution for peak detection. Therefore, it would be necessary to improve the segmentation stage in order to avoid undesirable samples and poor resolutions using an adaptive short time analysis technique. Yet, the number of known amphibian species is above 7000 so there still is a long way to go before a high quality solution can be reached.

References

Acevedo, M. A., Corrada-Bravo, C. J., Corrada-Bravo, H., Villanueva-Rivera, L. J., & Aide, T. M. (2009). Automated classification of bird and amphibian calls using machine learning: a comparison of methods. *Ecological Informatics, 4*(4), 206–214.

Aguilar-Martin, J., & López de Mántaras, R. (1982). The process of classification and learning the meaning of linguistic descriptors of concepts. *Approximate Reasoning in Decision Analysis, 1982*, 165–175.

Alford, R. A., & Richards, S. J. (1999). Global amphibian declines: a problem in applied ecology. *Annual review of Ecology and Systematics*, 133–165.

Alonso, R., Rodríguez, A., & Márquez, R. (2007). Sound guide of the amphibians from cuba (audio cd & booklet). *ALOSA sons de la natura, Barcelona, 2007*, 1–46.

AmphibiaWeb (2015). AmphibiaWeb.Berkeley, University of California. Accessed July 23, 2015. http://amphibiaweb.org.

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.

Beebee, T. J., & Griffiths, R. A. (2005). The amphibian decline crisis: a watershed for conservation biology? *Biological Conservation, 125*(3), 271–285.

Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on computational learning theory* (pp. 144–152). ACM.

Brandes, T. S. (2008). 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, 16*(6), 1173–1180.

Breiman, L. (2001). Random forests. *Machine learning, 45*(1), 5–32.

Burges, C. J. (1998). A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery, 2*(2), 121–167.

Chang, C.-C., & Lin, C.-J. (2011). Libsvm: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST), 2*(3), 27.

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

Clarke, B. T. (1997). The natural history of amphibian skin secretions, their normal functioning and potential medical applications. *Biological Reviews of the Cambridge Philosophical Society, 72*(03), 365–379.

Colonna, J. G., Cristo, M., Nakamura, E. F., & Salvatierra, M. (2015). An incremental technique for real-time bioacoustic signal segmentation. *Expert Systems with Applications, 42*, 7367–7374.

David, S., Ferrer, M. A., Travieso, C. M., Alonso, J. B., & y Comunicaciones, D. D. S. (2004). GPDSHMM: A hidden Markov model toolbox in the matlab environment. *CSIMTA, Complex Systems Intelligence and Modern Technological Applications*, 476–479.

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

Egea-Serrano, A., Relyea, R. A., Tejedo, M., & Torralva, M. (2012). Understanding of the impact of chemicals on amphibians: a meta-analytic review. *Ecology and evolution, 2*(7), 1382–1397.

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

Feng, A. S., Narins, P. M., Xu, C.-H., Lin, W.-Y., Yu, Z.-L., Qiu, Q., … Shen, J.-X. (2006). Ultrasonic communication in frogs. *Nature, 440*(7082), 333–336.

Ganchev, T., Potamitis, I., & Fakotakis, N. (2007). Acoustic monitoring of singing insects. In *Proceedings of IEEE international conference on Acoustics, speech and signal processing, 2007. ICASSP 2007: vol. 4* (pp. IV–721). IEEE.

Gaston, K. J., & O'Neill, M. A. (2004). Automated species identification: why not? *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 359*(1444), 655–667.

Grigg, G., Taylor, A., Mc Callum, H., & Watson, G. (1996). Monitoring frog communities: an application of machine learning. In *Proceedings of eighth innovative applications of artificial intelligence conference, portland oregon* (pp. 1564–1569).

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 *Proceedings of 2003 IEEE international conference on Acoustics, speech, and signal processing, 2003 (ICASSP'03): vol. 5* (pp. V–545). IEEE.

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

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

Jaafar, H., Ramli, D. A., Rosdi, B. A., & Shahrudin, S. (2014). Frog identification system based on local means k-nearest neighbors with fuzzy distance weighting. In *Proceedings of the 8th international conference on robotic, vision, signal processing & power applications* (pp. 153–159). Springer.

Kwet, A., & Márquez, R. (2010). Sound guide of the calls of frogs and toads from southern brazil and uruguay/guia de cantos das rãs e sapos do sul do brasil e uruguai/guía sonora de los sonidos de ranas y sapos del sur de brasil y uruguay. *Fonoteca, Madrid, Double CD and Booklet*, 1–55.

Lee, C.-H., Chou, C.-H., Han, C.-C., & Huang, R.-Z. (2006). Automatic recognition of animal vocalizations using averaged MFCC and linear discriminant analysis. *Pattern Recognition Letters, 27*(2), 93–101.

Owings, D. H., & Morton, E. S. (1998). *Animal vocal communication: A new approach*. Cambridge University Press.

Rabiner, L. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE, 77*(2), 257–286.

Renevey, P., & Drygajlo, A. (2001). Entropy based voice activity detection in very noisy conditions.. In *Proceedings of Eurospeech 2001 scandinavia, 7th european conference on speech communication and technology, 2nd interspeech event, aalborg, denmark, september 3–7, 2001* (pp. 1887–1890).

Semmlow, J. L., & Griffel, B. (2014). *Biosignal and medical image processing*. CRC press.

Xie, J., Towsey, M., Truskinger, A., Eichinski, P., Zhang, J., & Roe, P. (2015). Acoustic classification of australian anurans using syllable features. In *Proceedings of 2015 IEEE tenth international conference on Intelligent sensors, sensor networks and information processing (ISSNIP)*, (pp. 1–6). IEEE.

Yuan, C. L. T., & Ramli, D. A. (2013). Frog sound identification system for frog species recognition. In *Context-aware systems and applications* (pp. 41–50). Springer.

Zhou, X., Garcia-Romero, D., Duraiswami, R., Espy-Wilson, C., & Shamma, S. (2011). Linear versus mel frequency cepstral coefficients for speaker recognition. In *Proceedings of 2011 IEEE workshop on Automatic speech recognition and understanding (ASRU)* (pp. 559–564). IEEE.