Automatic Classification of Frogs Calls based on Fusion of Features and SVM

Juan J. Noda Arencibia, Carlos M. Travieso & David Sánchez-Rodríguez
Institute for Technological Development and Innovation in Communications,University of Las Palmas de Gran Canaria.
Las Palmas, Spain
jnoda@ingetelca.com,
carlos.travieso@ulpgc.es,david.sanchez@ulpgc.es

Malay Kishore Dutta, Garima Vyas
Department of Electronics and Communication Engineering
Amity University,Noida, India
malaykishoredutta@gmail.com, gvyas@amity.edu

***Abstract*—This paper presents a new approach for the acoustic classification of frogs' calls using a novel fusion of features: Mel Frequency Cepstral Coefficients (MFCCs), Shannon entropy and syllable duration. First, the audio recordings of different frogs' species are segmented in syllables. For each syllable, each feature is extracted and the cepstral features (MFCC) are computed and evaluated separately as in previous works. Finally, the data fusion is used to train a multiclass Support Vector Machine (SVM) classifier. In our experiment, the results show that our novel feature fusion increase the classification accuracy; achieving an average of 94.21% ± 8,04 in 18 frog's species.**

Keywords—sound classification, frogs call recognition, MFCC, data fusion , SVM

I. INTRODUCTION

Amphibians populations have been declined globally and several species are at risk of extinction [1]. Anurans (frogs and toads) are high sensitivity to environmental changes, pollution, water quality and other environmental stresses due, among other factors, to their thin skin and water permeability [2][3]. Therefore, they are used as nature bioindicators to determine the environmental health and habitat quality.

Monitoring biological indicators and species identification are vital tasks for ecological studies and conservation applications. The uses of bioacoustics monitoring technologies are very efficient to confirm the presence of species [4], particularly when visibility is limited. In the anurans the main purpose of vocalization is advertisement; the calls follow an acoustical pattern which is species-specific and are affected by changes in climatic conditions [5]. Furthermore, in most of the frog species the length of their calls is different, so it can be used to classification purposes.

Many research papers have been conducted in the field of bioacoustics animal classification. In [6], 22 frog's species samples obtained in the north of Australia were parameterized using as features the peaks of the signal spectrogram and theirs frequencies to train a C4.5 machine learning algorithm. However, the process results time consuming and not all the species couldn't be recognized. Brandes [7] studied 9 bird's, 10 frog's and 8 cricket's calls achieving an accuracy of 96%, 88% and 89% respectively. This approach used as features the peak frequencies and bandwidth from a spectrogram, then applied Hidden Markov Model (HMM) for classifying. In [8], the authors present an interesting comparison of methods using a set of features (call duration, max. and min. frequencies, maximum power and the frequency of maximum power in 8 segments) with Decision Trees (DT), Support Vector Machine (SVM) and Linear Discriminant Analysis (LDA); to automate the identification of 9 frog's and 3 bird's species. SVM method on this work outperforms the other two algorithms achieving 94.95% of accuracy with a low rate of false positives. Chen et al. [9] proposed a template based method which is extracted analyzing the length of the segmented syllables and applying a Multi Stage Average Spectrum (MSAS) method. It was successfully tested on 18 frog's species with detections ratios of 91.9% and 94.3%. Yuan et al. [10] applied MFCC and Linear Predictive Coding (LPC) with k-Nearest Neighbor (KNN), obtaining an identification accuracy of 98.1% in a dataset of 8 anurans selected from the Internet database AmphibiaWeb [11]. In contrast with other works, this manages a public database of sound recordings which we have used to evaluate our solution. More recently, in [12] syllable duration, frequency modulation, dominant frequency and oscillation rate was applied with a KNN classifier to recognize 16 species of Australian frogs' achieving a classification accuracy of 90.5%.

There are still few studies in the automatic acoustic classification field about anurans, and most of previous works have been limited to less than 10 species. In this paper, we proposed a novel fusion of acoustic features using Mel Frequency Cesptral Coefficients (MFCC) features, Shannon entropy and syllable distance; combining time and frequency domain information with a SVM classifier in order to successfully identify 18 anurans selected from AmphibiaWeb database.

The paper is organized as follows: Section 2 explains the description proposed technique. Next, in Section 3, we describe the database used for experiments. In Section 4, we

978-1-4673-7948-9/15/$31.00 ©2015 IEEE

evaluate our approach. Finally, Section 5 contains the conclusions and future work.

II. Proposed Technique

The proposed system is shown in Fig. 1. In the first stage, the audio signals from the database are automated segmented in syllables of different lengths. Next, from each syllable we extract the MFCC vectors of coefficients and computed its entropy and length, which are used into the classification stage. Finally, the multi-class SVM procedure produces the final object classification results.

A. Syllable Segmentation

The syllable segmentation procedure is performed following the algorithm of Härmä in [13], decomposing the input signal into a set of frequency and amplitude modulated sinusoidal pulses where each pulse corresponds to one syllable. The algorithm computed the spectrogram of the signal using a Short Time Fourier Transform (STFT), in our case using a Hamming window = 512 and overlap = 25% (Fig 2.). The spectrogram is represented as a time-frequency matrix $M(f,t)$. First, the algorithm finds the highest peak amplitude in the spectrogram $|M(f_n,t_n)|$, placing the n^{th} syllable in t_n and computing the amplitude as follow:

$$A_n(0) = 20\log_{10}(|M(f_n,t_n)|) \qquad (1)$$

Starting from this point, it traces the maximum adjacent peaks of $M(f,t)$ for $t>t_n$ and $t<t_n$ until $A_n(t_n\text{-}t_0)<A_n(0) - \beta dB$, where β is the stop n^{th} syllable segmentation criteria and their value is 20dB. The frequencies and amplitudes trajectories are stored, and the area from the spectrogram matrix is deleted:

$$M(f_n,[t_n - t_l,\cdots,t_n + t_r]) = 0 \qquad (2)$$

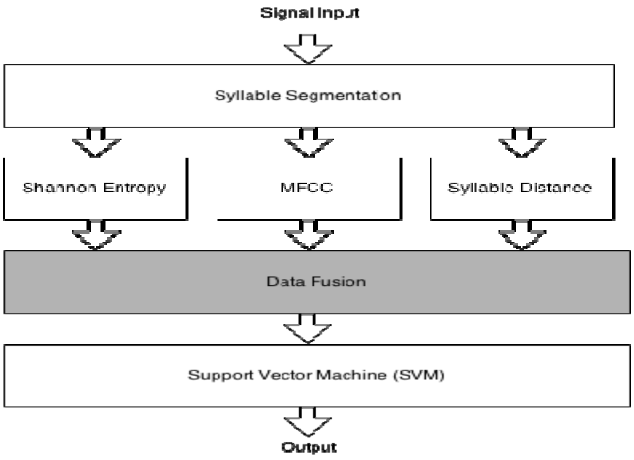

Fig 1. System description.

These steps are repeated until the end of the spectrogram.

B. MFCCs Extraction

After the segmentation process, the MFCC feature vectors are calculated per each syllable. Cepstral features are widely used in speech recognition field [14] and have also been applied in bioacoustics recognition with varied success [7][10][12]. Mel scale characterizes better the information to lower frequencies emulating the frequency response of the human auditory system; the amphibians in a similar way hear lower frequencies much better. Equation (3) shows the transformation from frequency to Mel scale.

$$m = 2595\ \log_{10}\left(\frac{f}{100}+1\right) \qquad (3)$$

MFCCs are calculated based on short-time analysis. First, a pre-emphasis filter is applied to compensate the high frequency. Then, each syllable is divided into frames; in our system we have established a window of size 25 milliseconds with 50% overlap. After segmenting, the frequency response of each frame is computed by Fast Fourier Transform (FFT), and their power spectrum is mapped to Mel scale; to do this we have used 26 triangular band-pass filters. Finally, features are obtained by taking log of the outputs of a Mel-frequency filter bank and conducting the Discrete Cosine Transform (DCT) as in (4). The final vectors are obtained by holding the lowest DCT coefficients, in our approach the best results were obtained with 18 coefficients.

$$C_n = \sum_{k=0}^{K-1}\left(\log|S_k|\cdot\cos\left[n\left(k-\frac{1}{2}\right)\frac{\pi}{K}\right]\right), \qquad (4)$$
$$0 \le n \le N-1,$$

where N is the number of MFCCs and K is the number of triangular filters.

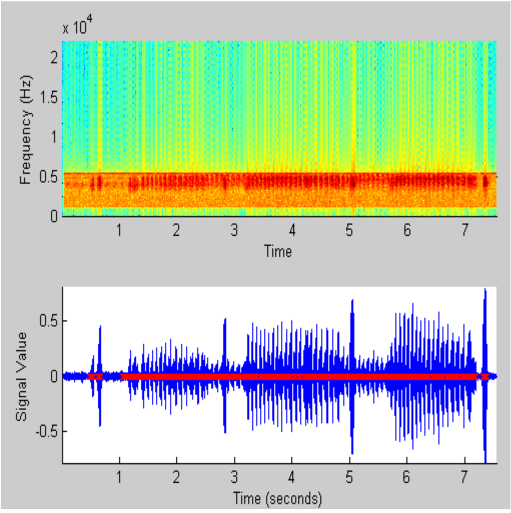

Fig 2. Boophis miniatus audio segmentation

C. Shannon Entropy

Shannon Entropy (SE) is also extracted for each syllable. The entropy is a measure of the average information content into a signal and it has been applied previously with success in speech recognition [15]. Let A be a discrete finite signal such as $A=\{a_1,a_2,a_3,...a_n\}$, Shannon Entropy is defined as follow (5):

$$SE(A) = -\sum_{i=1}^{n} p(a_i) \log_2 p(a_i) \qquad (5)$$

where $p(a_i)$ is the probability of a_i □ A.

SE is adopted in order to measure significant vocal changes of the frogs' call defining similar syllable models. In this paper, we have selected to fuse the entropy, MFCC and syllable length to add temporal and frequency domain information to the classification stage.

D. Support Vector Machine (SVM)

In order to carry out the classification a Support Vector Machines (SVM) [16] has been utilized. The SVM classifier is a strong classifier, so much it accepts a great number of parameters. The SVM solves the classification of geometric parameters or measures, to calculate exactly the separate hyperplane from the training data [16]. Besides, this solution introduces methods to be able to work with no separable and separable lineally cases of the data. The decision boundary is decided with the calculation of a separate hyperplane that discriminates between the positive and the negative samples; therefore a SVM treats the classification as two different classes. This decision boundary or separate hyperplane can be designed with different kernels, lineal, polynomial, gaussian, etc. For the following experiments, we have used a RBF kernel. We have also implemented a multiclass classifier for identification under the strategy "*one-versus-all*" according to [17], due to SVM only recognizes two classes.

III. Database

The AmphibiaWeb database [11] was created by the University of California, Berkeley, as an online project to share the knowledge relating to amphibians, so that this information could be used to help their conservation. AmphibiaWeb actually contains 7.416 species detailed descriptions with 606 digital sound recordings and 117 videos. The audio files come from various contributors and mainly recorded in natural environments which a significant noise background. In addition, the signals were saved with different samples rate and sample format (bit depth).

In this work, we have selected 18 anurans from this database; the selected species with their family taxonomy and origin are shown in Table 1. After downloading the audio files, each of them is converted to a 16 bit mono WAV format.

TABLE I: Selected Species

No.	Scientific Name	Family	Country
1	Adenomera marmorata	Leptodactylidae	Brazil
2	Aglyptodactylus laticeps	Mantellidae	Madagascar
3	Aglyptodactylus madagascariensis	Mantellidae	Madagascar
4	Ameerega flavopicta	Dendrobatidae	Brazil
5	Anaxyrus microscaphus	Bufonidae	USA
6	Anaxyrus punctatus	Bufonidae	USA,Mexico
7	Anaxyrus retiformis	Bufonidae	USA,Mexico
8	Anodonthyla boulengeri	Microhylidae	Madagascar
9	Aplastodiscus leucopygius	Hylidae	Brazil
10	Austrochaperina robusta	Microhylidae	Australia
11	Blommersia wittei	Mantellidae	Madagascar
12	Boophis miniatus	Mantellidae	Madagascar
13	Dendropsophus microps	Hylidae	Brazil
14	Gastrotheca ovifera	Hemiphractidae	Venezuela
15	Rana sierrae	Ranidae	USA
16	Rhinella schneideri	Bufonidae	Argentina
17	Spea intermontana	Scaphiopodidae	USA,Canada
18	Spea multiplicata	Scaphiopodidae	USA,Mexico

IV. Experimental results

To evaluate the effectiveness of the system, the proposed data fusion of features is compared with the output of the system using only the MFCCs coefficients to train the SVM classifier as in previous works. The experiments have been repeated 50 times shuffling the dataset to obtain statistically significant results. Table II lists the performance evaluation of proposed method, the accuracy rate represents the percent of test syllables which were correctly identified by the system, is defined as in (6). In Table III, a comparison is shown among the proposed system and the previous state-of-the-art methods.

$$Accuracy(\%) = \frac{\text{Syllables Correctly Identified (Nc)}}{\text{Total Number of Syllables (Ns)}} \qquad (6)$$

For MFCC features, the experiments reveal a poor classification performance in some species, in fact two species: Anodonthyla boulengeri and Aplastodiscus leucopygius hardly reached 60% of accuracy per experiment due to their spectrogram distributions are similar to other species. Adenomera marmorata, Ameerega flavopicta and Anaxyrus microscaphus presented a classification rate of 100%, since theirs call frequencies distribution are different respect the rest of the anurans. For the data fusion, the results suggest that the proposed technique provide rich complementary information about the frog call vocalization increasing the recognition accuracy in almost all the species. Two more species reach 100% of accuracy compared with MFCC approach, Aglyptodactylus madagascariensis and Dendropsophus microps. Furthermore, the anurans which presented a low classification rate, Anodonthyla boulengeri and Aplastodiscus leucopygius, increase significantly their accuracy because both have a distinctive average syllable length of 0.06 and 0.09 seconds respectively. We can also observe that Spea intermontana accuracy doesn’t improve.

It is due to the audio presents a strong cricket background noise and the diversity of the training samples is not large enough to classify their call variance. In last two columns, we have analyzed our approach using only 20% and 30% of the samples to train the SVM to prove the robustness of the system. The system appears to perform well with low number of training data achieving an acceptable level of accuracy.

V. CONCLUSIONS

In this work, we have presented an automatic identification system of anurans based on their vocalization call, introducing a new approach of parameterization through a data fusion of Mel Frequency Cepstral Coefficients (MFCC), Shannon Entropy (SE) and syllable length, combining time and frequency domain characteristics. The data fusion has been proved to be efficient to improve the performance of MFCC, achieving a highest classification rate per specie. We have tested our solution on the public domain database AMPHIBIAWEB reaching a classification success rate of 94.21% with a standard deviation of 8.04%. However, more work remains to be done in order to increase the number of anurans that can be successfully detected and identified. Furthermore, an improvement in the segmentation stage to deal with the background noise is required. We hope to apply the techniques in other species, such as whales and insects, to prove the strength of our approach.

TABLE II. EXPERIMENTAL RESULTS

No.	Scientific Name	Total Syllables Ns	MFCC Mean % ± std	Fusion MFCC&Shanon&Syllable Distance Mean % ± std		
			50% Training	**50% Training**	**30% Training**	**20% Training**
1	Adenomera marmorata	17	100±0.0	100±0.0	100±0.0	100±0.0
2	Aglyptodactylus laticeps	19	84.88±0.35	86.00±0.34	80.18±0.32	74.92±0.43
3	Aglyptodactylus madagascariensis	15	99.71±0.05	100±0.0	99.55±0.06	99.60±0.06
4	Ameerega flavopicta	22	100±0.0	100±0.0	100±.0.00	100±0.00
5	Anaxyrus microscaphus	77	100±0.0	100±0.0	99.95±0.02	99.66±0.05
6	Anaxyrus punctatus	19	72.00± 0.49	80.44±0.39	69.00±0.46	64.00±0.48
7	Anaxyrus retiformis	32	89.75±0.30	97.37±0.26	93.05±0.25	85.54±0.35
8	Anodonthyla boulengeri	21	62.80± 0.48	94.00±0.23	67.66±0.46	59.14±0.49
9	Aplastodiscus leucopygius	18	60.00± 0.49	78.66±0.41	85.00±0.35	80.33±0.39
10	Austrochaperina robusta	31	99.33± 0.08	99.33±0.08	99.11±0.09	98.28±0.12
11	Blommersia wittei	23	99.81± 0.04	99.63±0.06	99.23±0.08	99.12±0.09
12	Boophis miniatus	25	89.00± 0.31	95.00±0.21	92.00±0.27	88.94±0.31
13	Dendropsophus microps	53	93.00± 0,25	100±0.0	95.35±0.21	94.05±0.23
14	Gastrotheca ovifera	34	90.23± 0.29	97.76±0.14	92.40±0.26	88.34±0.32
15	Rana sierrae	19	96.66±0.17	97.11±0.16	96.18±0.19	94.30±0.23
16	Rhinella schneideri	94	99.82± 0.04	99.95±0.02	99.96±0.01	99.78±0.04
17	Spea intermontana	19	74.00± 0.43	74.00±0.43	69.81±0.45	68.15±0.46
18	Spea multiplicata	151	96.26±0.18	96.69±0.17	95.93±0.19	95.79±0.20
	Average % ± std		**89.29±13.29**	**94.21±8.04**	**90.80±11.46**	**88.33±13.50**

TABLE III. COMPARISON OF THE PROPOSAL VS. STATE-OF-THE-ART

Reference	Dataset	Segmentation	Features	Classification	Accuracy
[6]	22 frogs from Australia	Manual	Spectrogram peaks and frequencies	C4.5 DT	Not all species coundn’t be reconized
[7]	9 birds, 10 frogs and 8 crickets	Manual	Peak frequency and bandwidth	HMM	96%,88% and 89%
[8]	9 frogs and 3 birds from Puerto Rico	Manual	Call duration/Max. and min. frequency/Maximum Power/Frequency of maximun power	SVM	94.95%
[9]	18 frogs	Energy and zero-crossing rate	Syllable length/ MSAS	Template based	91.9% and 94.3%
[10]	AmphibiaWeb (8 frogs)	Manual	MFCC	KNN	98.1%
[12]	16 frogs from Australia	Härmä	Syllable length /Frequency modulation/Dominant frequency/Oscillation rate	KNN	90.5%
This Work	AmphibiaWeb (18 frogs)	Härmä	MFCC/Entropy/Syllable length	SVM	94.21%±8.04

REFERENCES

[1] Alford, R.A., Richards, S.J., Global amphibian declines: a problem in applied ecology. Annual Review of Ecology and Systematics 30, 133–165, 1999

[2] A. Serrano,R.A. Relyea, M. Tejedo and M. Torralva, Understanding of the impact of chemicals on amphibians: a meta-analytic review. Ecology and evolution.vol 2.,pp 1382-1397,July 2012.

[3] Sueur J., Pavoine, S. Hamerlynck O. and Duvail, S., Rapid acoustic survey for biodiversity appraisal. PLoS One, 3(12), e4065,2008.

[4] Duellman, William E., Biology of amphibians. JHU Press, 1986.

[5] Grigg, G., Taylor, A., Mc Callum, H., and Watson, G, "Monitoring frog communities: an application of machine learning." *Proceedings of Eighth Innovative Applications of Artificial Intelligence Conference, Portland Oregon*. 1996.

[6] Brandes, T. S.., "Feature vector selection and use with hidden Markov models to identify frequency-modulated bioacoustic signals amidst noise." *Audio, Speech, and Language Processing, IEEE Transactions on* 16.6 (2008): 1173-1180.

[7] Acevedo, M. A., Corrada-Bravo, C. J., Corrada-Bravo, H., Villanueva-Rivera, L. J., and Aide, T. M., Automated classification of bird and amphibian calls using machine learning: A comparison of methods, Ecological Informatics, 2009, vol. 4, no 4, p. 206-214.

[8] Chen, W. P., Chen, S. S., Lin, C. C., Chen, Y. Z., and Lin, W. C., “Automatic recognition of frog calls using a multi-stage average spectrum”, Computers & Mathematics with Applications, vol. 64, no 5, p. 1270-1281, Jan. 2012.

[9] Yuan, C. L. T., and Ramli, D. A, Frog Sound Identification System for Frog Species Recognition, *Context-Aware Systems and Applications*. Springer Berlin Heidelberg, 2013. p. 41-50, 2013

[10] AmphibiaWeb: Information on amphibian biology and conservation. Berkeley, University of California, http://amphibiaweb.org/. (Accessed: Jun. 2015).

[11] Xie, J., Towsey, M., Truskinger, A., Eichinski, P., Zhang, J., and Roe, "Acoustic classification of Australian anurans using syllable features." *Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP), 2015 IEEE Tenth International Conference on*. IEEE, April 2015.

[12] A. Harma,“Automatic identification of bird species based on sinusoidal modeling of syllables,”presented at the 2003 5th International Conference on Acoust. Speech Signal Process, pp. 545–548.

[13] Tiwari, V. (2010). MFCC and its applications in speaker recognition. *International Journal on Emerging Technologies*, *1*(1), 19-22.

[14] Hosseinzadeh, D., and Krishnan, S, Combining vocal source and MFCC features for enhanced speaker recognition performance using GMMs. In *Multimedia Signal Processing, 2007. MMSP 2007. IEEE 9th Workshop on* (pp. 365-368). IEEE, October 2007

[15] Burges, C. J. C., “A Tutorial on Support Vector Machines for Pattern Recognition”, Data Mining and Knowledge Discovery, 2(2): 121-167, 1998.

[16] C. Travieso, J. Alonso, M. Ferrer, “Sign Language to Text by SVM”, in Proceeding of 7th International Symposium on Signal Processing and its Applications, Paris, France, Vol. II, pp. 435-438, July 2003.