RESEARCH ARTICLE OPEN ACCESS

Gaussian Mixture Model Classification of Frogs

Damián A. Nicolalde Rodríguez*, Andrés E. Terneux **, Daniel P. Nicolalde Rodríguez***
** (Faculty Of Engineering, Pontificia Universidad Católica Del Ecuador,*
*** (Faculty Of Engineering, Pontificia Universidad Católica Del Ecuador,*
****(Faculty Of Engineering, Pontificia Universidad Católica Del Ecuador,*
Corresponding Auther : Damián A. Nicolalde Rodríguez

ABSTRACT
This Work Deals With Automatic Call-Independent Frog Species Identification. An Algorithm Is Designed To Process Field Recordings And Perform Automatic Identification Of 10 Species Of Anurans Inhabiting The Yasuní National Park In The Ecuadorian Amazon Region. First, End-Point Detection Using Short-Term-Energy (STE) With A Moving-Average Filter Is Applied To Isolate Frog Calls Over An SNR > 15 Db Threshold. Audio Segments With Background Noise And Silence Are Discarded. Isolated Segments Are Then Parametrized Using Cepstral Feature Vectors That Represent The Frog Acoustic Phenomenon. The Data Is Divided Into Two Groups From Which One Is Used To Train Gaussian Mixture Models And The Others Are Used For Testing Classification Accuracy For Each Species. GMM Models With Different Mixture Weights (Components) Are Generated In Order To Determine The Best Model Order. The Classification Task Is Based On The Maximum-Likelihood (ML) Rule Achieving The Maximum Average Success Rate Of 97.24% With GMM Models Of 64 Components.
Keywords - Frog Identification, Mel Frequency Cepstral Coefficients, Gaussian Mixture Models

---

Date Of Submission: 31-03-2018 Date Of Acceptance 16-04-2018

I. INTRODUCTION

In Nature Conservation, It Is Necessary To Quantify The Impact That Human Activities Have On Biodiversity And The Ecosystem As A Whole. One Common Way To Obtain Information Is To Measure Frog Population Sizes Since They Are Considered Accurate Indicators Of Environmental Stress Due To Their Aquatic And Terrestrial Habitat. Researchers Usually Record Anuran Audio Signals On The Field In A Labor Intensive Task, And Manual Analysis Of The Material Involves A Long And Tedious Process [1]. Therefore, The Main Challenge Is The Development Of Suitable Signal Processing Algorithms For Automatic Detection And Classification Of Frog Species Living In The Complex Acoustic Environment Of The Ecuadorian Rainforest.

Male Frogs Use Acoustic Signaling Mainly For Advertisement Purposes To Attract Potential Mates, Defend Their Territory And Show Distress. Anuran Vocalizations Are Commonly Composed Of A Call That Is Formed By One Or Many Sequenced Notes Also Known As Syllables. A Syllable Is An Acoustic Signal Produced By Air Blown Though Vocal Cords And Resonated By A Vocal Sac [2]. In This Work A Call Is Chosen As The Basic Element For Recognition. Most Of The Research Reported In The Literature Is Focused On Frog Species Recognition With A Call-Dependent Approach. Taylor Et Al. [3] Developed An Early Frog Recognition System For 22 Species Applying Spectrogram Analysis To Extract Frequency Peaks And Classify Frog's Species. However, Several Misidentifications On One Species And The Need To Lump 3 Species To Obtain Meaningful Results Due To Their Call Similarity Showed The Limitations Of This Approach. An Inspiring Work By Brandes [4] Introduced Feature Vectors Extracted From Spectrograms, And Modeled Bio-Acoustic Signals Of 10 Frogs Recorded In The Amazon Basin With Hidden Markov Models (HMM). Themethod Exhibited Low Performance When Faced With Broadband Frog Calls Since Less Intense Harmonics Are Ignored By The Algorithm. Huan Et Al. In [5] Developed A Frog Sound Identification System Extracting 3 Features Representing Frog Call Syllables Previously Segmented Reporting Up To 90.3% Recognition Rate Using Support Vector Machine (SVM) Classification. The Dataset Consisted Of 5 Species, 2 Of Which Were Clearly Misclassified Requiring Further Analysis. In [6] Lee Et Al. Proposed A Method Using Averaged Mel Frequency Cepstral Coefficients (MFCC) And Linear Discriminant Analysis (LDA) To Automatically Identify 30 Types Of Frogs. The Averaged MFCC Outperforms The Recognition Rate Reported Using Hmms But Losses The Dynamic Content Of The Frog Call. Chen Et

Al. In [7] Suggested A Method Based On Preclassification Of Syllable Lengths, And A Multi-Stage Averaged Spectrum (MSAS) With Template Matching. This Approach Reported The Best Recognition Rate On A Dataset Of 18 Frog Calls When Compared To Other Methods Based On Dynamic Time Warping (DTW), K-Nearest-Neighbor (Knn) And SVM. However, Misclassification Of Species With Similar Spectrum Was Reported.

Recently, Bedoya Et Al. In [8] Suggested An Unsupervised Methodology For Automatic Identification Based On A Fuzzy Classifier And Mfccs. The Method Was Tested Successfully With 13 Species Of Anurans Found In Colombia. The Call Dependent Nature Of The Approach Does Not Take Into Account The Individual Call Variations That Many Frog Species Exhibit [9].

A Call-Independent Frog Identification System Is Desired Since It Enables Species Recognition Despite The Call Type Produced [8]. Research In This Area Has Been Extensively Focused On Birds [10], [11] And Odontocetes [12]. Aboudan Et Al. In [13] Tested The Ability Of MFCC And Linear Predictive Cepstral Coefficients (LPCC) In The Frog Recognition Process Using GMM. However, Real Recordings Were Not Used At All. They Used Synthetic Sequences Of Frog Calls In Their Experiments, And The GMM Model Complexity Is Limited To One Mixture Weight.

In This Work, We Test The Ability Of MFCC With GMM To Recognize Calls Of 10 Frog Species Inhabiting The Yasuní National Park In Ecuador. The Recordings Were Made In The Rainforest Which Is Characterized By A Complex Acoustic Environment. Thus, It Is Possible To Find Different Sounds: Birds, Bats, Crickets, Mammals And Other Frog Species Sharing The Spectrum At The Same Time. Experimental Results Demonstrated The Effectiveness Of The Proposed Method To Achieve Call-Independent Recognition On Real Recordings Made In The Wilderness.

This Paper Is Organized As Follow. Section 2 Describes The Acoustics Database Used For This Work. Section 3 Details The Procedure To Isolate Frog Calls. Section 4 Explains The Generation Of GMM Models. Section 5 Describes The Identification Process Of Frog Species. The Proposed Method Is Evaluated On Real World Recordings In Section 6. Finally, A Discussion And Conclusions Are Summarized In Section

II. ACOUSTICS DATABASE

The Database Of Frog Calls Used In This Study Was Provided By Museo De Zoología Of Pontificia Universidad Católica Del Ecuador (PUCE) [14]. Recordings Were Made Using A Sennheiser K6-ME67TM Unidirectional Microphone Attached To Digital Recorders Olympus LS-10TM Or Marantz PMD660TM With Sampling Frequency Of 44100 Hz And 16-Bit Resolution. The Recording Schedule Was From 19h00 To 2h00 At Natural Ponds Located Within The Yasuní National Park In The Amazon Basin Of Ecuador. In The Study Zone More Than 128 Anuran Species Have Been Identified. For Our Experiments The 10 Frog Species Listed In Table 1 Were Chosen Based Upon Availability.

Table 1. Frog Species Database

Family	Gender	Code
Bufonidae	Rhinella margaritifera	f01
Craugastoridae	Pristimantis conspicillatus	f02
Hylidae	Dendropsophus bifurcus	f03
Hylidae	Dendropsophus Triangulum	f04
Hylidae	Hypsiboas alfaroi	f05
Hylidae	Hypsiboas calcaratus	f06
Hylidae	Hypsiboas cinerascens	f07
Hylidae	Osteocephalus fuscifacies	f08
Leptodactylidae	Engystomops petersi	f09
Leptodactylidae	Leptodactylus discodactylus	f10

Acoustic Environment At Yasuní National Park

At The Study Site, The Reliability Of Frog-Call Recognition Algorithms Is Affected By The Influence Of Noise That Degrades The Quality Of Field Recordings. First, Antrophogenic Sound Sources Like AC Generators, Traffic Noise From Trucks And Oil Drilling Activities Introduce An Important Component Of Noise Disturbing The Range Of Interest. Second, Biogenic Sound Sources Like Crickets Introduce A Noise Level That Is Present While Frog Sounds Are Active. Also, Natural Sound Sources Like Rain And Wind Are Present. Figure 1 Shows The Spectrogram Of A Rinhella Margaritifera Call Referring Antrophogenic And Biogenic Sound Presence.

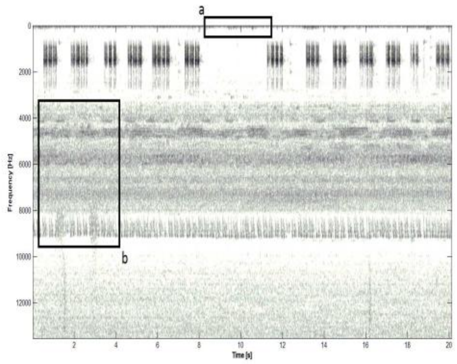

Fig 1. Spectrogram Showing A Rinhella Margaritifera Call In The Presence Of Noise Sources In The Study Zone (A) Antrophogenic (AC Generator) (B) Biogenic (Insects).

A Selection Of Bio-Acoustic Material Was Performed For Producing Ground-Truth For Training The GMM Models As Described In Section 4 As Well As For Testing The Created Models As Described In Section 5. In Addition, The Frogcall Activity Detector In The Segmentation Step Have To Be Carefully Tuned In Order To Reduce The Impact That Noise Sources Could Have On The Classification Stage.

Audio Selection And Annotation

The Bioacustic Material Used For The Experiments Was Selected By Specialists To Ensure That Only Best Quality Audio Was Utilized To Generate A Ground-Truth Corpus For Training And Testing The Algorithm. In General Terms, Audio Segments With Frog Calls SNR > 15 Db, No Multi-Species Overlap And Without Clipping Were Manually Selected. Field Recordings Containing Human Voice, Mechanical Artifacts Or Anthropogenic Sound Sources Were Discarded [15].

Segmentation Of Frog Calls Was Performed Automatically Applying The Signal Processing Algorithm Described In The Following Section. Automatic Call Segmentation Was Preferred For Training The Models Encouraged By The Experience Reported In [16] After Manual Segmentation Attempts Resulted In Specialist-Bias And Lack Of Consistency Among Different Annotators [15].

III. FROG CALL SEGMENTATION

Since A Frog Call Was Chosen As The Basic Element Of Species Identification, A Segmentation Technique That Detects Calls While Avoiding Portions Of Silence And Noise Is Required. The Technique Described Here Is Based On Short-Time-Energy (STE), And Endpoint Determination To Detect Audio Segments Containing Frog Calls. Figure 2 Shows The Frog Call Segmentation Diagram.

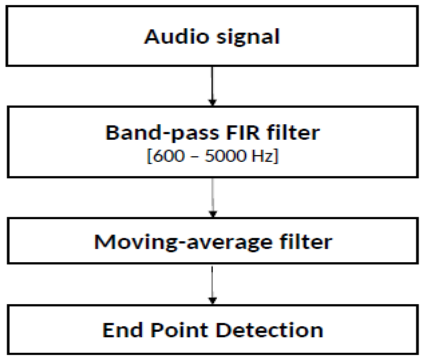

Fig. 2. Frog Call Segmentation Diagram.

An Algorithm For Automatic Segmentation Of Frog Calls Was Adapted Based On The Classic Endpoint Detection Algorithm Proposed For Human Voice Analysis In [17]. First, A Band-Pass FIR Filter Is Applied On The Original Audio Signal S With Cut-Off Frequencies 600 − 5000 Hz. The Filter's Bandwidth Is Selected In Order To Contain Most Of The Energy Present In The 10 Frog Calls Studied. The Filter Is Used For The Audio Segmentation Procedure, But Not For Training And Recognition Steps. The Filtered Signal Sf Is Then Divided Into 10 Ms Frames With No Overlapping In Order To Calculate A Modified Short-Time Energy (STE) Sequence According To:

$$E(n) = \sum_{m=(n-1)N+1}^{nN} s_f(m)^2, \qquad (1)$$

Where E(N) Is The Energy Of Frame N, Sf (M) Is The Filtered Discrete-Time Signal And N Is The Number Of Samples Of Each 10 Ms Frame.

A Moving-Average Filter Is Then Applied To E(N) To Get A Smooth Version Of The STE Sequence. A Whole Frog Call Is Then Delimited Rather Than Each Separated Note. The Moving Average Filter Is Applied According To The Following Formula:

$$E_s(n) = \frac{1}{2Na+1}(E(n+Na) + E(n+Na-1) + \cdots + E(n) + \cdots + E(n-Na)), \qquad (2)$$

Where Es(N) Is A Smoothed Version Of E(N), Na Is The Number Of Adjacent Points In Each Size Of E(N), And 2N + 1 Is The Total Numbers Of Data Points For The Moving-Average Calculation. In Our Experiments A Value Of Na = 10 Proved Sufficient To Detect The Frog Calls In Table 1.

End Point Detection

Endpoint Detection Of Frog Calls Is Performed According To The Decision Rule Suggested By Rabiner In [17] With Little Modification. The Algorithm Is Described In The Following Steps:

1) Compute The Mean Value Of 13 Values Of Es(N). The 13 Values Consider The 10 First And

The 3 Last Values Of The Sequence Es(N). This Mean Value Represents An Estimation Of The Background Noise Energy.

2) Verify If 3 Consecutive Values Of Es(N) Are Bigger Than An Established Threshold, To Determine The Starting Point Of A Call. Subsequently, Verify If 3 Consecutive Values Are Lower Than The Threshold To Determine The End Point Of The Call.

The Threshold Value Was Chosen Empirically To Ensure That Only Calls With SNR $\geq$ 15 Are Detected. For This Study A Threshold Value Of 40 Times The Estimation Of Background Noise Energy Is Used. Figure 3 Shows Calls Of Rinhella Margaritifera Detected Using The Proposed Segmentation Algorithm In A Field Recording.

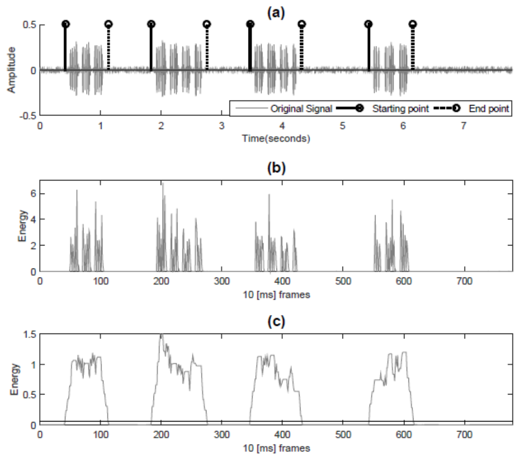

Fig. 3. End Point Detection In A Field Recording Of Rinhella Margaritifera. (A) Original Signal With End Point Detection. (B) Short Time Energy Of 10[Ms] Blocks, E(N). (C) Smooth Version Of E(N), Es(N) And Threshold Location.

IV. MODELING FROG CALLS AS AUSSIAN MIXTURE MODELS

The Procedure Followed To Model Frog Species Using Gaussian Mixture Models (GMM) Is Described In This Section. First, Frog Calls Are Characterized Using Mel Frequency Cepstral Coefficients (MFCC) [18]. Mfccs Are Expected To Model The Underlying Parameters Of The Mechanism Of Sound Production Of The Frogs. These Parameters Have Shown Inter-Species Variability In Tree Frogs.

Mel Frequency Cepstral Coefficients

In The Present Work, MFCC Coefficients Are Used To Represent The Audio Features That Describe The Acoustics Characteristics Of Frog Calls. In The Literature Different Kinds Of Audio Features Have Been Proposed For Audio Analysis With MFCC Achieving Best Results In Speaker Identification [18]. Moreover, Mel Cepstral Coefficients Have Shown A Robust Performance In Presence Of Non-Stationary Noise Which Is Commonly Found In Field Recordings In The Amazon Forest Environment. We Extracted 14 MFCC Coefficients And Formed A Feature Vector Of 13 Elements Discarding The First Coefficient. Mel Cepstral Features Were Extracted Using The Matlab Audio Analysis Library Available In [20] Which Is Implemented Based On The Auditory Toolbox By Slaney In [19].

Only Audio Segments Of The Original Audio Signal S That Resulted Of Applying The Procedure Described In Section 3 Were Considered For Feature Extraction Step. Each Audio Segment Was Divided Into 40 Ms Blocks With 50% Overlap And Hamming Windowed. MFCC Coefficients Are Extracted From Each Block Resulting In A Matrix Of MFCC Coefficients. The Consecutive Feature Vectors Represent The Spectral Characteristics Of A Frog Call, And The Sequence Of Vectors Contain Implicit The Time-Varying Features Of The Call.

Gaussian Mixture Model Description

The Probability Density Function Of The Frog-Calls Feature Vector Is Represented By A Gaussian Mixture Density Of M Components [21]:

$$p(\vec{x} \mid \lambda) = \sum_{i=1}^{M} p_i b_i(\vec{x}), \qquad (3)$$

Where $\vec{x}$ Is A D-Dimensional Feature Vector (In Our Case Containing 13 MFCC Coefficients), $b_i(\vec{x}),\ i = 1, \ldots, M$ Are The Component Densities, And $p_i,\ i = 1, \ldots, M$, Are The Mixture Weights. Each Component Density Is A Gaussian Function Of D Variables:

$$b_i(\vec{x}) = \frac{1}{(2\pi)^{D/2}|\Sigma_i|^{1/2}} \exp\left\{-\frac{1}{2}(\vec{x} - \vec{\mu}_i)'\Sigma_i^{-1}(\vec{x} - \vec{\mu}_i)\right\}, \qquad (4)$$

With Mean Vector $\vec{\mu}_i$ And Covariance Matrix $\sum_I$. The Mixture Weights Satisfy The Constraint $\sum_{i=1}^{M} p_i = 1$. The Density Model Is Denoted By The Mean Vector, Covariance Matrix And The Mixture Weights As:

$$\lambda = \{p_i, \vec{\mu}_i, \Sigma_i\},\ i = 1, \ldots, M. \qquad (5)$$

In The Automatic Identification Task, A GMM Model (λ) For Each Frog Species Of Yasuní National Park Was Generated. For The Experiments, We Chose A Unique Diagonal Covariance Matrix Per Generated GMM To Simplify The Models. It Is Important To Consider That Frog Calls Are Less Complex Than Human Utterances Which Are Composed Of Many Different Sounds. Frogs Produce Fewer Kinds Of Sounds And Posses A Limited Vocabulary. We Tested GMM Models With Different Number Of Components M In Order To Establish The Best Model Order In Terms Of

Performance While Keeping The Complexity Of The Model Relatively Low.

Training The Frog Models

Figure 4 Shows The Training Process Diagram. The Main Objective Is To Obtain A Database Of 10 GMM Models That Represents The Frog Species, f_k, $1 \geq k \geq 10$.

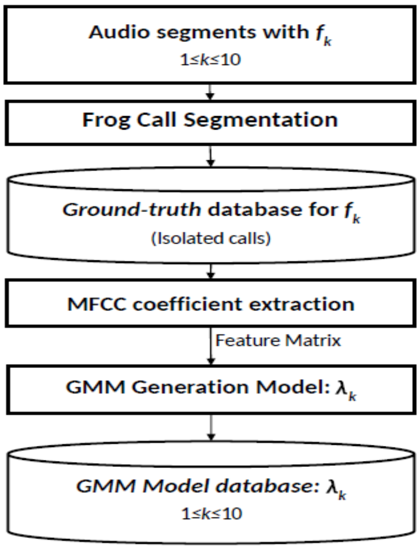

Fig. 4. Training Process Diagram.

First, It Is Necessary To Select Audio Segments That Contain Calls Of Each Species, F_k. The Selection Procedure Followed The Guidelines Proposed In [15]. Then, The Frog Call Segmentation Procedure Of Section 3, Was Applied To The Selected Audio Segments. As A Result, A Ground-Truth Database Of Isolated Frog Calls For The Studied Species F_k Was Generated. The Resultant Corpus Was Divided Into Two Groups. One Used To Train The GMM Models And The Other For Testing The Classification Accuracy Of The Proposed Algorithm. The MFCC Extraction Procedure Of Section 4.1 Was Applied To The Corpus To Obtain A Matrix Of Cepstral Features Of The Calls. This Matrix Of Cepstral Features Is Used To Estimate The Maximum-Likelihood Parameters Of The Gaussian Model Λ_k, Associated With The Species F_k, Using The Expectation-Maximization (EM) Method. We Followed The Guidelines Described In [21], But Modified Accordingly To Frog Species Recognition Based On Their Advertisement Calls.

V. MODELING THE FROGS SPECIES

Figure 5 Shows The Identification Process Diagram.

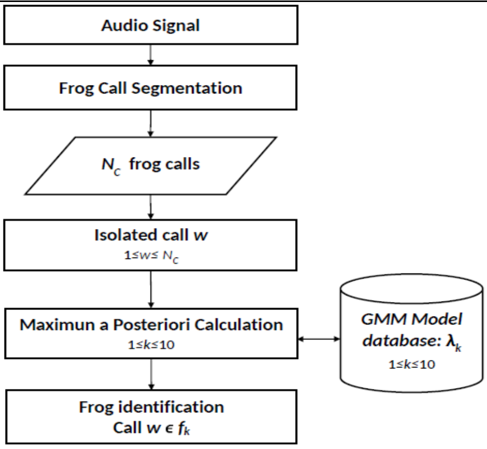

Fig. 5. Identification Process Diagram.

After Applying The Frog-Call Segmentation Algorithm To The Input Audio, NC Audio Segments Were Obtained For Classification. The Identification Procedure Of Figure 5 Was Applied To Each Isolated Call w, $1 \leq w \leq N_C$. Since For This Project We Required To Identify The 10 Species Of Yasuní National Park In Table 1, A Set Of Ten Frog Species $F = \{f_{01}, f_{02}, ..., f_{10}\}$ Was Established. Each Frog Species Is Represented By A Model λ_k, k = 1, 2, ..., 10. The Goal Is To Find The Frog Model Which Has The Maximum Posterior Probability For An Input Sequence $\vec{X} = \{\vec{x}_1, ..., \vec{x}_T\}$. For This Study, The Input Sequence Is A Matrix Of MFCC Coefficients Of Each Call W Of The Audio Signal. Minimum Error Bayes's Decision Procedure Was Applied To Tackle This Problem:

$$\hat{f} = \arg \max_{1 \leq k \leq 10} \Pr(\lambda_k | \vec{X}) = \arg \max_{1 \leq k \leq 10} \frac{p(\vec{X}|\lambda_k)}{p(\vec{X})} \Pr(\lambda_k), \tag{6}$$

$\hat{f}$ Is Considered The Identified Frog. Assuming Identical Prior Probabilities Of Frog Species $\Pr(\lambda_k)$ And The Value $p(\vec{X})$ Is The Same For All Models, The Decision Rule Becomes:

$$\hat{f} = \arg \max_{1 \leq k \leq 10} p(\vec{X}|\lambda_k). \tag{7}$$

Using Logarithms And Assumed Independence Between Observations, The Decision Can Be Expressed As:

$$\hat{f} = \arg \max_{1 \leq k \leq 10} \sum_{t=1}^{T} \log\ p(\vec{x}_t|\lambda_k). \tag{8}$$

Each Audio Segment On The Input W Is Composed Of T Blocks. So, The Expression $\sum_{t=1}^{T} \log\ p(\vec{x}_t|\lambda_k)$ In Eq. 8 Is Dependent On T Value. Normalizing Eq. 8 Based On T Value, We Have:

$$\hat{f} = \arg \max_{1 \leq k \leq 10} \frac{1}{T} \sum_{t=1}^{T} \log \ p(\vec{x}_t | \lambda_k). \qquad (9)$$

The Value Of Max $\left[\frac{1}{T}\sum_{t=1}^{T} \log \ p(\vec{x}_t|\lambda_k)\right], 1 \leq k \leq 10,$ Is Interpreted As The Maximum Likelihood Of The Model λ_k That Bestmatches The Input Signal. However, The Amazon Region Is An Environment With High Biodiversity, Where There Are Many Animal Sounds (Birds, Bats, Crickets, Mammals, Other Frog Species, Etc). It Is Important To Exclude Input Signals That Do Not Belong To The 10 Frog Species Of Table 1, Establishing A Threshold γ_k For The Maximum Likelihood Value. Thus, The Input Signal Will Be Accepted As A Frog Call Of Species In Table 1 As Long As:

$$\max_{1 \leq k \leq 10} \frac{1}{T} \sum_{t=1}^{T} \log \ p(\vec{x}_t | \lambda_k) \geq \gamma_k. \qquad (10)$$

VI. EXPERIMENTAL RESULTS

We Applied Real Field Recordings Attained At Yasuní National Park, In The East Of Ecuador Between 2001 To 2015 During Night Time, To Our Algorithms In Order To Evaluate The Performance Of The Proposed Method. The Experiments Are Performed Using Matlab R2014(A) In A Computer With Processor Intel Core^{tm} I5 CPU M520 @ 2.4 Ghz, 4G RAM And 64 Bits Operative System Windows 7 Professional. The Acoustic Database Described In Section 2 Was Used For The Experiments. The Call Frog Segmentation Procedure, As Explained In 3, Was Applied To The Original Recordings To Get An Audio Corpus Consisting Of 871 Calls Belonging To 10 Frog Species. These Calls Were Divided Approximately In 33% For Training And 66% For Testing As Shown In Table 2. The Algorithm Accuracy Was Tested Based On The Rate Of Correctly Recognized Calls Versus The Total Number Of Calls:

$$\text{Success rate}(\lambda_k) = \frac{\text{Calls successfully recognized (Frog } k)}{\text{Total calls (Frog } k)}. \qquad (11)$$

Table 2 Shows The Results Of Testing The Proposed Algorithm On The Evaluation Corpus. Also, There Is A Description Of Number Of Calls Used To Train The Model As Well As Number Of Calls Used To Test The Recognition Algorithm. Additionally, GMM Models With Different Components Values, M = 1, 2, 4, 8, 16, 32, 64, Were Generated In Order To Determine The Best Model Orders In Terms Of Recognition Performance.

Table 2 Also Describes The Obtained Success Rate For Each Frog Species Based On Testing The Identification Task In The Created 10 GMM Models With Different Component Values, M. Table 3 Shows The Average Success Rate For Frog Call Recognition Based On The Number Of GMM Components, M.

Results Of The Experiments Showed Good Average Success Rate With GMM Of 4 Or More Components. The Average Success Rate With GMM Of 4 Components Is 95.01%. The Maximum Average Success Rate In Our Experiments Is 97.24%. This Value Was Obtained With The GMM Model Of 64 Components. The Minimum Individual Success Rate Value Associated With F_{03} Was 89.18% And The Maximum Individual Success Value Is 100%. This Value Is Associated With 5 Frog Species. Based On The Results Found On Tables 2 And 3, The Performance Of The Classification Task Suggest That The Usage Of The Proposed Algorithm With GMM Models Of 4 Or More Components In The Complex Acoustic Environment Found In The Amazon Basin Of Ecuador Is Promising.

Accuracy Of The Automatic Frog Call Recognition Is Remarkable When Using 64-Component Gmms Obtaining A Maximum Rate Of 100% While Keeping A Minimum Of 89.1 % For F_{03}. The Drop In Performance For F_{03} Classification Is Due To Noisy Recordings. As Expected, The Accuracy Rate Improves When The Number Of GMM Components Increases. A Trade-Off Between Accuracy And Computational Complexity Was Observed And Should Be Taken Into Consideration For Practical Deployment Of The Algorithm.

Our Algorithm Is A Very Helpful Tool In Analyzing The Presence Of Frog-Calls In Recordings Of Frogs Made In The Wild. Since There Are Several Hundred Hours Of Frogs Bioacustic Material In PUCE Archive Without Identification, The Application Of Our Algorithm Is Expected To Save Time In Metadata Generation As Well As Improve The Inventory Procedure Without The Need Of Specialists Whom Are Scarce And Expensive.

Also, The Application Of Our Algorithm In Long Recordings Might Be Used For Wildlife Monitoring And Biodiversity Estimation Efforts Based In Acoustic Methods.

DISCUSSION BIOLOGY + RESULTS APPLIED TO BIOLOGY

Table 2 Success Rate For Frog Call Recognition.

Frog Code	# of Calls (Training)	# of Calls (Testing)	# of Calls (Total)	Model Order	Success Rate(%)
f_{01}	29	56	85	$M = 1$	100.00
				$M = 2$	100.00
				$M = 4$	100.00
				$M = 8$	100.00
				$M = 16$	100.00
				$M = 32$	100.00
				$M = 64$	100.00
f_{02}	40	88	128	$M = 1$	96.59
				$M = 2$	97.73
				$M = 4$	100.00
				$M = 8$	100.00
				$M = 16$	100.00
				$M = 32$	100.00
				$M = 64$	100.00
f_{03}	23	37	60	$M = 1$	89.18
				$M = 2$	89.18
				$M = 4$	89.18
				$M = 8$	89.18
				$M = 16$	89.18
				$M = 32$	89.18
				$M = 64$	89.18
f_{04}	8	16	24	$M = 1$	62.50
				$M = 2$	68.75
				$M = 4$	87.50
				$M = 8$	87.50
				$M = 16$	87.50
				$M = 32$	93.75
				$M = 64$	100.00
f_{05}	30	64	94	$M = 1$	100.00
				$M = 2$	100.00
				$M = 4$	100.00
				$M = 8$	100.00
				$M = 16$	100.00
				$M = 32$	100.00
				$M = 64$	100.00
f_{06}	9	15	24	$M = 1$	100.00
				$M = 2$	100.00
				$M = 4$	100.00
				$M = 8$	100.00
				$M = 16$	100.00
				$M = 32$	100.00
				$M = 64$	100.00
f_{07}	45	100	145	$M = 1$	91.00
				$M = 2$	91.00
				$M = 4$	94.00
				$M = 8$	98.00
				$M = 16$	99.00
				$M = 32$	99.00
				$M = 64$	99.00
f_{08}	16	29	45	$M = 1$	75.86
				$M = 2$	82.76
				$M = 4$	93.10
				$M = 8$	89.65
				$M = 16$	93.10
				$M = 32$	93.10
				$M = 64$	93.10
f_{09}	48	146	194	$M = 1$	50.00
				$M = 2$	71.91
				$M = 4$	86.30
				$M = 8$	91.78
				$M = 16$	93.15
				$M = 32$	94.52
				$M = 64$	95.89
f_{10}	33	40	73	$M = 1$	57.14
				$M = 2$	91.83
				$M = 4$	100.00
				$M = 8$	100.00
				$M = 16$	97.95
				$M = 32$	100.00
				$M = 64$	97.95

Table 3 Average Success Rate For Frog Call Recognition

GMM model order	Average Success Rate(%)
$M = 1$	82.23
$M = 2$	89.32
$M = 4$	95.01
$M = 8$	95.61
$M = 16$	95.99
$M = 32$	96.96
$M = 64$	97.24

VII. CONCLUSION

The Proposed Algorithms For Frog Recognition Based On Exploring The Acoustical Properties Of Frog Calls In Yasuní National Park - Ecuadorian Amazon Region Presents Good Results. Automated Evaluation Of Wildlife Recordings Introduces A Potent Technology That Is Complementary To Existing Survey Techniques Used Currently By Researchers In The Wild. Its Applications Range From Assessing Animal Populations In A Study Zone, Characterization And Inventory Of Unidentified Bioacustic Recordings Archived Creating Automatic Metatags, To Biodiversity Indexes Estimation Based In Acoustic Analysis. From The Performance Evaluation Of Our Algorithms, The Average Success Rate With GMM Models Of 4 O More Components Confirms The Positive Results. These Scores Are Obtanied For The Frog Species Found On Table 1. However, To Ensure The Highest Success Rate With Different Frog Species, It Would Be Recommended To Used GMM Models With 32 Or 64 Components. Moreover, It Is Notable That Many Frog Species Get 100% Of Individual Success Rate. It Is Important To Mention That The Obtained Results Are Based On Recordings With Frog Calls With Uniform Background Noise And Signalto-Noise Ratio Equal Or Greater Than 15 Db. The Selection Of Recordings Avoided Clipped Signals And Mechanical Noise (As Explained In Section 2).

MFCC Coefficients Have Been Used Successfully In Human Voice Characterization For Speaker Identification. The Main Reason For Using MFCC Coefficients In The Frog Recognition Task Is Due To Its Robust Performance When Faced With Non-Stationary Noise, And The Underlying Modeling Of The Sound Production Mechanism Of Frogs Which Might Enable Individual Recognition With More Analysis.

Gmms Performed Well In The Frog Species Recognition Task Using Their Advertisement Call. The Result Was Expected Since Frog's Vocabulary Is Smaller Than In Human Beings Where Gmms Have Been Applied Successfully.

The Implemented System Will Be Very Useful For Researchers Studying Environmental Changes Through Biodiversity Monitoring. Frog Presence Is An Evidence That The Ecosystem Has Not Been Altered. One Common Way To Obtain Information Is To Measure Frog Population Sizes And Presence. In Fact, Frogs Are Accurate Indicators Of Environmental Stress Due To Their Aquatic And Terrestrial Habitat. Many Important Applications For Biodiversity Monitoring And Wildlife Surveillance Are Envisioned Using The Proposed Algorithm Specially For Wireless Acoustic Sensors.

ACKNOWLEDGEMENTS

The Study Was Supported By Pontificia Universidad Católica Del Ecuador. The Authors Would Like To Thank Estación Científica Yasuní For Their Cooperation During The Execution Of Field Recordings. To Franco Cisneros, Jean Camino, Daniela Pareja And Daniel Rivadeneira For Their Support In The Acoustics Selection Material.

REFERENCES

[1] M. A. Acevedo And L. J. Villanueva-Rivera, "Using Automated Digital Recording Systems As Effective Tools For The Monitoring Of Birds And Amphibians," Wildlife Soc. Bulletin, Vol. 34, No. 1, Pp. 211–214, Mar. 2013.

[2] W. E. Duellman And L. Trueb, "Vocalization," In Biology Of Amphibians, 1st Ed., Baltimore, MD: Johns Hopkins Univ. Press, 1994, Ch. 4, Pp. 87-107.

[3] A. Taylor Et Al., "Monitoring Frog Communities: An Application Of Machine Learning," Proc. 8th Annu. Conf. Innovative Applicat. Artificial Intell., Portland–Oregon, 1996, P. 1564–1569.

[4] T. S. Brandes, "Feature Vector Selection And Use With Hidden Markov Models To Identify Frequency-Modulated Bioacoustic Signals Amidst Noise," IEEE Trans. Audio, Speech And Language Process., Vol. 16, No. 6, Pp. 1173–1180, Aug. 2008.

[5] C. J. Huang Et Al., "Frog Classification Using Machine Learning Techniques," Expert Syst. Applicat., Vol. 36, No. 2, Pp. 3737–3743, Mar. 2009.

[6] C. H. Lee Et Al., "Automatic Recognition Of Animal Vocalizations Using Averaged MFCC And Linear Discriminant Analysis," Pattern Recognition Lett., Vol. 27, No. 2, Pp. 93–101, Jan. 2006.

[7] W. P. Chen Et Al., "Automatic Recognition Of Frog Calls Using A Multistage Average Spectrum," Comput. And Math. Applicat., Vol. 64, No. 5, Pp. 1270–1281, Sep. 2012.

[8] C. Bedoya Et Al., "Automatic Recognition Of Frog Calls Using A Multistage Average Spectrum," Ecological Informatics, Vol. 24, Pp. 200–209, Nov. 2014.

[9] M. A. Bee Et Al., "Individual Variation In Advertisement Calls Of Territorial Male Green Frogs, Rana Clamitans: Implications For Individual Discrimination," Ethology, Vol. 107, No. 1, Pp. 65–84, Jan. 2001.

[10] J. Cheng Et Al., "A Call-Independent And Automatic Acoustic System For The Individual Recognition Of Animals: A Novel Model Using Four Passerines," Pattern Recognition, Vol. 43, No. 11, Pp. 3846–3852, Nov. 2010.

[11] E. J. S. Fox Et Al., "Call-Independent Individual Identification In Birds," Bioacoustics, Vol. 18, No. 1, Pp. 51–67, Jan. 2008.

[12] M. A. Roch Et Al., "Gaussian Mixture Model Classification Of Odontocetes In The Southern California Bight And The Gulf Of California," J. Acoust. Soc. America, Vol. 121, No. 3, Pp. 1737–1748, Jan. 2008.

[13] A. Aboudam, "Acoustic Monitoring System For Frog Population Estimation Using In-Situ Progressive Learning," M. S. Thesis, Dept. Elect. And Comp. Eng., Colorado State Univ., Fort Collins, 2013.

[14] S. R. Ron Et Al., (2016, Jan. 3). Amphibiawebecuador (2016.0) [Online]. Available: Http://Zoologia.Puce.Edu.Ec/Vertebrados/Anfibios

[15] A. Estrella Et Al., "Selection, Annotation And Labeling Of Ecuadorian Frog Calls," (In Spanish), Revista PUCE 102, Quito, Ecuador, 2016.

[16] H. Jaafar And D. A. Ramli, "Automatic Syllables Segmentation For Frog Identification System," In 2013 IEEE 9th Int. Colloq. Signal Processing And Its Applications, Kuala Lumpur, Malaysia, 2013, Pp. 224–228.

[17] L. R. Rabiner And M. R. Sambur, "An Algorithm For Determining The Endpoint Of Isolated Utterances," The Bell System Tech. J., Pp. 298–315, Feb. 1975.

[18] S. Davis And P. Mermelstein, "Comparison Of Parametric Representations For Monosyllabic Word Recognition In Continuously Spoken Sentences," IEEE Trans. Acoust., Speech, Signal Process., Vol. 28, No. 4, Pp. 357–366, Aug. 1980.

[19] M. Slaney, "Auditory Toolbox," Interval Research Corporation, Palo Alto, CA, Tech. Rep. #1998-010, 1998.

[20] T. Giannakopoulos And A. Pikrakis, "Introduction To Audio Analysis: AMATLAB Approach," San Diego, CA: Academic Press, 2014.

[21] D. A. Reynolds And R. C. Rose, "Robust Text-Independent Speaker Identification Using Gaussian Mixture Speaker Models", IEEE Trans. Speech Audio Process., Vol. 3, No. 1, Pp. 72–83, Jan. 1995.

Damián A. Nicolalde Rodríguez"Gaussian Mixture Model Classification of Frogs" InternationalJournal of Engineering Research and Applications (IJERA) , vol. 8, no.4, 2018, pp. 42-49