Frog Sound Identification System for Frog Species Recognition

Clifford Loh Ting Yuan and Dzati Athiar Ramli

Intelligent Biometric Research Group (IBG), School of Electrical and Electronic Engineering, USM Engineering Campus, Universiti Sains Malaysia, 14300, Nibong Tebal, Pulau Pinang, Malaysia
claypod016@gmail.com, dzati@eng.usm.my

Abstract. Physiological research reported that certain frog species contain antimicrobial substances which is potentially and beneficial in overcoming certain health problem. As a result, there is an imperative need for an automated frog species identification to assist people in physiological research in detecting and localizing certain frog species. This project aims to develop a frog sound identification system which is expected to recognize frog species according to the recorded bio acoustic signals. The Mel Frequency Cepstrum Coefficient (MFCC) and Linear Predictive Coding (LPC) are used as the feature extraction techniques for the system while the classifier employed is k-Nearest Neighbor (K-NN). Database from AmphibiaWeb has been used to evaluate the system performances. Experimental results showed that system performances of 98.1% and 93.1% have been achieved for MFCC and LPC techniques, respectively.

Keywords: Frog Sound, Identification, Mel Frequency.

1 Introduction

Other than applications related to human identity recognition [1], biometric technology has been used on the identification of biological acoustic sounds which is imperative for biological research and environmental monitoring. This is particularly true for detecting and locating animals due to we often hear the animal sound rather than visually detect the animal [2]. In animals, the initiation of sound could be as a means of information transmission or as a by-product of their living activities such as moving, eating or flying. In general, animals make sounds to communicate with members of same species and thus their vocalizations have evolved to be species-specific. Therefore, identifying animal species from their vocalizations is meaningful to ecological research.

Interest towards automatic recognition of animal species based on their vocalization has increased and many researches based on these studies have been published. [3] investigated different types of animals includes birds, cats, cows and dogs according to the animal calls. In another research, 16 different classes of animal calls were successfully classified as reported in [4].

Apart of recognizing types of animals, recognizing of species of the same animals is also found especially for bird species identification as reported in [5], [6] and [7]. [5] proposed the segmentation sounds of bird syllable using spectrum over time

P.C. Vinh et al. (Eds.): ICCASA 2012, LNICST 109, pp. 41–50, 2013.

method and template matching was employed as classifier. Different representations of bird syllables have been studied in [6] while [7] executed the frequency information for syllables segmentation of bird species sounds.

However, studies on the frog species recognition is still in infancy. An automated frog call identification system for public online consultation has been developed by [8]. Three features i.e. spectral centroid, signal bandwidth and threshold crossing rate are extracted for the purpose of this study. This study revealed that, certain frog species can easily be recognized by proposed methods but some species such as Microhyla butleri Microhyla ornate needs further investigation. In another case, Dayou et al. [9] introduced three different types of entropy i.e Shannon entropy, Renyi entropy and Tsallis entropy as the features. This study is based on nine species of frog sound from Microhylidae family. The k-Nearest Neighbor, kNN was then used as classifier resulting identification accuracy more than 80%. Subsequently, Han et al. [10] used Fourier spectral centroid instead Tsallis entropy as feature extraction and the result has improved with an average accuracy more than 90%.

The interest of this project is to build a system which is able to identify the species of frog based on their sound. The acoustic signal of frog vocalizations can be represented as a sequence of syllables. Thus, syllable can be used as the acoustic component to identify the frog species. A syllable is basically a sound that a frog produces with a single blow of air from the lungs. Once the syllables have been properly segmented, a set of features will be calculated to represent each syllable. These features will later be used for training or identification purpose.

In this work, a frog call identification system is constructed based on audio signal sampled from recordings of frog call. This system can be divided into four modules, including digital frog call collection, feature extraction, matching module and decision module. This process can be illustrated as in Fig. 1.

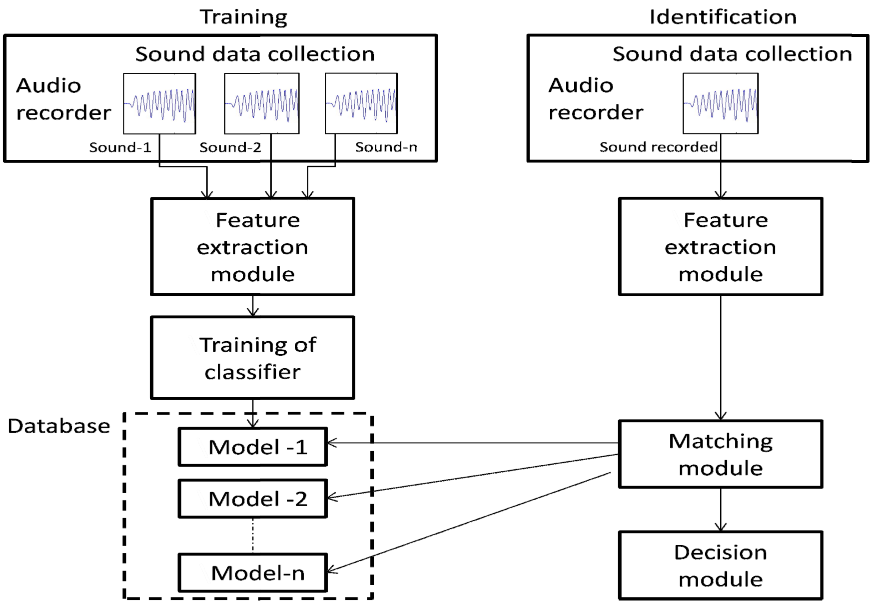

Fig. 1. Audio biometric identification system

2 Methodology

2.1 Data Acquisition

In this project, the digital frog call samples are obtained from AmphibiaWeb [11]. AmphibiaWeb is an online system enabling anyone with a Web browser to search and retrieve information relating to amphibian biology and conservation. AmphibiaWeb offers ready access to taxonomic information for every recognized species of amphibian in the world. Species accounts are being added regularly by specialist and volunteers and it also contains species descriptions, life history information, conservation status, audio signal, literature reference for many species [11].

The sampling rate and quantization bit of audio signals in AmphibiaWeb database are different due to the audio signals are offered by various contributors. Hence, the audio signals downloaded from AmphibiaWeb is converted and saved as 16-bit mono wav format. Table 1 lists the eight frog's audio signals that have been used in the project:

Table 1. List of frog call samples obtained from AmphibiaWeb

Types of Frog species
Adenomera_marmorata
Aglyptodactylus_madagascariensis
Ameerega_flavopicta
Anodonthyla_boulengerii
Aplastodiscus_leucopygius
Blommersia_wittei
Boophis_luteus
Boophis_miniatus

Each audio signal is then properly segmented as a syllable and a set of features can be calculated to represent each syllable as shown in Fig. 2.

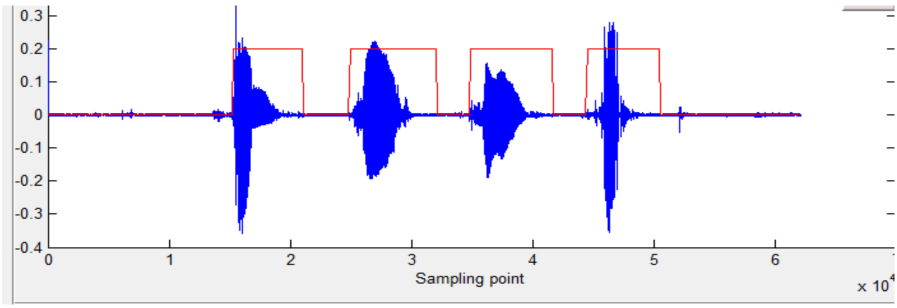

Fig. 2. Syllable segmentation

2.2 Feature Extraction

Mel-frequency cepstral coefficients (MFCC) which are commonly used as feature extraction for speech recognition and speaker recognition have been tested for frog sound recognition in this study. The computation of MFCC is based on short-term analysis [12]. The steps to implement MFCC are as follow:

1. Compute the Discrete Fourier Transform (DFT) of all frames of the signal. The DFT of all frames of the signal is

$$x_t(k) = X_t(e^{j\frac{2\pi k}{N}}) ,k= 0 ,1,\ldots, N-1 \tag{1}$$

Equation (1) is also known as signal spectrum.
2. The signal spectrum is then processed by filter bank processing. Filter bank is a set of 24 band-pass filters which emphasize on processing spectrum which is below 1kHz. Filter bank is generally used to simulate the human ear processing. The m-th filter bank output is $Y_t(m), 1 \leq m \leq M$ and M is number of band-pass filters.
3. Compute the energy of the logarithm of the square magnitude filter bank outputs, $Y_t(m)$. This can reduce the complexity of computing the logarithm of the magnitude of the coefficients.
4. Perform the inverse DFT on the logarithm of the magnitude of the filter bank output

$$y_t{}^{(m)}(k) = \sum_{m=1}^{m} \log\{|Y_t(m)|\} \cdot \cos\left(k\left(m - \frac{1}{2}\right)\frac{\pi}{M}\right), k = 0, \ldots, L \tag{2}$$

k is number of cepstral coefficients excluding 0’th coefficient. In this project, k = 12. Each feature set consists of 12 mel cepstrum coefficients, one log energy coefficient.

By using Linear Predictive Coding processing, the speech wave and spectrum characteristic can be precisely represented by a very few number of parameters. The LPC implementation is as follows:

1. Compute the Discrete Fourier Transform (DFT) of all frames of the signal. The DFT of all frames of the signal is shown in equation (3.11). This also defined as signal spectrum.
2. Compute the autocorrelation coefficient at equation (2.37) by using the following short-time autocorrelation function. That is
3. Use Durbin’s recursive solution for the autocorrelation equation.
4. Convert autocorrelation coefficient to complex cepstrum.

$$\dot{h}(n) = a_n + \sum_{k=1}^{n-1}\left(\frac{k}{n}\right)\dot{h}(k)a_{n-k} \qquad 1 \leq n \tag{3}$$

2.3 Identification Process

K-NN classifier requires a set of reference template to perform classification. The steps to assemble the reference template based on training data and identification process are listed as follows:

1. Training data is first sampled and continue with feature extraction. The extracted feature is then resized to 4096 feature points.
2. Assemble a vector whose distinct values define the grouping of the rows in training data based on their species.

3. Testing data is first sampled and continue with feature extraction. The extracted feature is then resized to 4096 feature points.
4. Perform classification based on the input testing data, training data, assembled vector, *k* value, distance metric, and the rule that used to classify sample. In this project, 1-nn with Euclidean distance and "nearest" rule are set.

3 Result and Discussion

To evaluate the system, eight species with 24 frog call syllable segmentation are divided into two sets of data i.e. training data and testing data. For each species, four syllables are randomly selected as training data while the rest as testing data. The accuracy of the classifiers is defined as follows:

Table 2. True positive and false positive of MFCC-KNN and LPC-KNN on frog calls obtained from AmphibiaWeb

MFCC-KNN		Predicted class								Recognition Accuracy (%)
		Ad	Ag	Am	An	Ap	Bl	Lu	Mi	
Actual class (4 training data)	Ad	20	0	0	0	0	0	0	0	100
	Ag	0	20	0	0	0	0	0	0	100
	Am	0	0	19	1	0	0	0	0	95
	An	0	0	0	20	0	0	0	0	100
	Ap	0	1	0	0	18	0	0	1	90
	Bl	0	0	0	0	0	20	0	0	100
	Lu	0	0	0	0	0	0	20	0	100
	Mi	0	0	0	0	0	0	0	20	100
Mean Recognition Accuracy										98.1
LPC-KNN		Predicted class								Recognition Accuracy (%)
		Ad	Ag	Am	An	Ap	Bl	Lu	Mi	
Actual class (4 training data)	Ad	19	0	0	0	1	0	0	0	95
	Ag	0	18	0	0	0	0	2	0	90
	Am	0	2	18	0	0	0	0	0	90
	An	0	0	0	18	0	0	0	2	90
	Ap	0	0	0	0	18	0	2	0	90
	Bl	0	0	0	0	0	18	1	1	90
	Lu	0	0	0	0	0	0	20	0	100
	Mi	0	0	0	0	0	0	0	20	100
Mean Recognition Accuracy										93.1

Ad: Adenomera_marmorata	Ap: Aplastodiscus_leucopygius
Ag:Aglyptodactylus_madagascariensi	Bl: Blommersia_wittei
Am: Ameerega_flavopicta	Lu: Boophis_luteus
An: Anodonthyla_boulengerii	Mi: Boophis_miniatus

Table 3. Performances of MFCC-KNN and LPC-KNN at four, three, and two training data on frog calls obtained from AmphibiaWeb

Species common name	K-NN					
	4 training data		3 training data		2 training data	
	MFCC	LPC	MFCC	LPC	MFCC	LPC
Adenomera_marmorata	20	19	20	19	20	19
Aglyptodactylus_madagascariensis	20	18	20	18	20	18
Ameerega_flavopicta	19	18	19	18	19	18
Anodonthyla_boulengerii	20	18	20	18	20	18
Aplastodiscus_leucopygius	18	18	18	18	18	18
Blommersia_wittei	20	18	20	18	19	18
Boophis_luteus	20	20	20	20	20	20
Boophis_miniatus	20	20	20	20	20	20
Matched syllable	157	149	157	149	156	149
total testing syllable	160	160	160	160	160	160
Accuracy (%)	98.125	93.125	98.125	93.125	97.5	93.125

Table 4. True positive and false positive of MFCC-KNN and LPC-KNN at different numbers of training data (a) 4 Training data, (b) 3 Training data, (c) 2 Training data

4 training data	MFCC-KNN		Predicted class								Recognition Accuracy (%)
			Ad	Ag	Am	An	Ap	Bl	Lu	Mi	
	Actual class	Ad	20	0	0	0	0	0	0	0	100
		Ag	0	20	0	0	0	0	0	0	100
		Am	0	0	19	1	0	0	0	0	95
		An	0	0	0	20	0	0	0	0	100
		Ap	0	1	0	0	18	0	0	1	90
		Bl	0	0	0	0	0	20	0	0	100
		Lu	0	0	0	0	0	0	20	0	100
		Mi	0	0	0	0	0	0	0	20	100
	Mean Recognition Accuracy										98.1

Table 4. (*continued*)

	LPC-KNN	Predicted class								Recognition Accuracy (%)
		Ad	Ag	Am	An	Ap	Bl	Lu	Mi	
Actual class	Ad	19	0	0	0	1	0	0	0	95
	Ag	0	18	0	0	0	0	2	0	90
	Am	0	2	11	0	0	0	0	0	90
	An	0	0	0	18	0	0	0	2	90
	Ap	0	0	0	0	18	0	2	0	90
	Bl	0	0	0	0	0	18	1	1	90
	Lu	0	0	0	0	0	0	20	0	100
	Mi	0	0	0	0	0	0	0	20	100
Mean Recognition Accuracy										93.1

(a)

3 training data	MFCC-KNN	Predicted class								Recognition Accuracy (%)
		Ad	Ag	Am	An	Ap	Bl	Lu	Mi	
Actual class	Ad	20	0	0	0	0	0	0	0	100
	Ag	0	20	0	0	0	0	0	0	100
	Am	0	0	19	1	0	0	0	0	95
	An	0	0	0	20	0	0	0	0	100
	Ap	0	1	0	0	18	0	0	1	90
	Bl	0	0	0	0	0	20	0	0	100
	Lu	0	0	0	0	0	0	20	0	100
	Mi	0	0	0	0	0	0	0	20	100
Mean Recognition Accuracy										98.1
	LPC-KNN	Predicted class								Recognition Accuracy (%)
		Ad	Ag	Am	An	Ap	Bl	Lu	Mi	
Actual class	Ad	19	0	0	0	1	0	0	0	95
	Ag	0	18	0	0	0	0	2	0	90
	Am	0	2	18	0	0	0	0	0	90
	An	0	0	0	18	0	0	0	2	90
	Ap	0	0	0	0	18	0	2	0	90
	Bl	0	0	0	0	0	18	1	1	90
	Lu	0	0	0	0	0	0	20	0	100
	Mi	0	0	0	0	0	0	0	20	100
Mean Recognition Accuracy										93.1

(b)

Table 4. (*continued*)

2 training data	MFCC-KNN	Predicted class								Recognition Accuracy (%)
		Ad	Ag	Am	An	Ap	Bl	Lu	Mi	
	Actual class: Ad	20	0	0	0	0	0	0	0	100
	Ag	0	20	0	0	0	0	0	0	100
	Am	0	0	19	1	0	0	0	0	95
	An	0	0	0	20	0	0	0	0	100
	Ap	0	1	0	1	18	0	0	0	90
	Bl	0	0	0	0	0	19	1	0	95
	Lu	0	0	0	0	0	0	20	0	100
	Mi	0	0	0	0	0	0	0	20	100
	Mean Recognition Accuracy									**97.5**
	LPC-KNN	Predicted class								Recognition Accuracy (%)
		Ad	Ag	Am	An	Ap	Bl	Lu	Mi	
	Actual class: Ad	19	0	0	0	1	0	0	0	95
	Ag	0	18	0	0	0	0	2	0	90
	Am	0	2	18	0	0	0	0	0	90
	An	0	0	0	18	0	0	0	2	90
	Ap	0	0	0	0	18	0	2	0	90
	Bl	0	0	0	0	0	18	2	0	90
	Lu	0	0	0	0	0	0	20	0	100
	Mi	0	0	0	0	0	0	0	20	100
	Mean Recognition Accuracy									**93.1**

Ad: Adenomera_marmorata
Ag:Aglyptodactylus_madagascariens
Am: Ameerega_flavopicta
An: Anodonthyla_boulengerii
Ap: Aplastodiscus_leucopygius
Bl: Blommersia_wittei
Lu: Boophis_luteus
Mi: Boophis_miniatus

(c)

$$\text{Accuracy (\%)} = \frac{N_c}{N_s} \times 100\% \tag{4}$$

Where N_c is the number of correctly recognized syllables, and N_s is the total number of test syllables. Besides, confusion matrix is also used to show the true positive and false positive of classifiers. True positive defined as the correct identification made by classifiers; false positive means the wrong identification made by classifiers [12].

In Table 2, the result shows that there are three false positives for MFCC-KNN based on frog calls obtained from AmphibiaWeb. This situation could be caused by the quality of recording. Higher quality of recording could increase the accuracy of K-NN classifier. In this case, K-NN classifier is able to identify all the species correctly since the numbers of false positive are much more fewer than number of true positive.

The accuracy of MFCC-KNN decreases with the decrement of the number of training data i.e. 98.1%, 98.1% to 97.5% as given in Table 3. On the other hands, the accuracy of LPC-KNN remains 93.1% when the numbers of training data change. However, as given in Table 4, the numbers of false positive and true positive maintain although the numbers of training data change. From the results, we can observe that the accuracy of LPC-KNN is not much affected by the numbers of training data.

4 Conclusion

In this study, a frog sound identification system which is expected to recognize frog species according to the recorded bio acoustic signals has been developed successfully. Experimental results revealed a very promising results and the developed identification system can be a viable approach in assisting people in physiological research in detecting and localizing certain frog species.

Acknowledgments. This project is supported by Short Term Grant USM, 304/PELECT/60311048.

References

1. Lip, C.C., Ramli, D.A.: Comparative Study on Feature, Score and Decision Level Fusion Schemes for Robust Multibiometric Systems. In: Sambath, S., Zhu, E. (eds.) Frontiers in Computer Education. AISC, vol. 133, pp. 941–948. Springer, Heidelberg (2012)
2. Lee, C.H., Chou, C.H., Han, C.C., Huang, Y.J.: Automatic recognition of Animal Vocalization using Averaged MFCC and Linear Discriminant Analysis. Pattern Recognition 27, 93–101 (2006)
3. Mitrovic, D., Zeppelzauer, M., Breiteneder, C.: Discrimination and Retrieval of Animal Sounds. In: 12th International on Multi-Media Modeling Conference Proceedings, p. 5 (2006)
4. Guodong, G., Li, S.Z.: Content-Based Audio Classification and Retrieval by Support Vector Machines. IEEE Transactions on Neural Network 14, 209–215 (2003)
5. Tyagi, H., Hegde, R.M., Murthy, H.A., Prabhakar, A.: Automatic Identification of Bird Calls using Spectral Ensemble Average Voiceprints. In: 13th European Signal Processing Conference Proceeding (2006)
6. Somervuo, P., Harma, A., Fagerlund, S.: Parametric Representations of Bird Sounds for Automatic Species Recognition. IEEE Transactions on Audio, Speech, and Language Processing 14, 2252–2263 (2006)
7. Harma, A.: Automatic Identification of Bird Species based on Sinusoidal Modeling of Syllables. In: International Conference on Acoustic Speech Signal Processing, vol. 5, pp. 545–548 (2003)

8. Huang, C.J., Yang, Y.J., Yang, D.X., Chen, Y.J.: Frog Classification using Machine Learning Techniques. Expert system with Application 36, 3737–3743 (2009)
9. Dayou, J., Han, N.C., Mun, H.C., Ahmad, A.H.: Classification and identification of frog sound based on entropy approach. In: International Conference on Life Science and Technology, vol. 3, pp. 184–187 (2011)
10. Han, N.C., Muniandy, S.V., Dayou, J.: Acoustic classification of Australian anurans based on hybrid spectral-entropy approach. Journal of Applied Acoustic 72, 639–645 (2011)
11. AMPHIBIAWEB: Information on amphibian biology and conservation. AmphibiaWeb, Berkeley, California, `http://amphibiaweb.org/`
12. Furui, S.: Comparison of Speaker Recognition Methods using Statistical Features and Dynamic Features. IEEE Transactions on Acoustics, Speech and Signal Processing 29, 342–350 (1981)