Proceedings of the Technical Sessions, 28 (2012) 100-105
Institute of Physics – Sri Lanka

Classification of Birds using FFT and Artificial Neural Networks

Anuradha Abewardana and Upul Sonnadara
Department of Physics, University of Colombo, Sri Lanka

ABSTRACT

The use of feed-forward artificial neural network to categorize a selected set of Sri Lankan bird species based on their vocalization is presented. The inputs to the neural network were frequencies of bird vocalizations where each vocalization was characterized by a frequency range. Out of the selected birds, only two birds showed peak frequency values below 1,000 Hz. The Sri Lanka Scaly Thrush has the maximum average peak frequency of 7,761 Hz and the Green Billed Coucal has the lowest of 334 Hz. The preliminary results show that the artificial neural network which was trained to classify individual birds based on their frequency features had an accuracy of greater than 90% for several bird types.

1. INTRODUCTION

Almost all the birds use sound (bird calls or bird songs) to communicate with their own species as well as different species. Bird songs which are longer and more complex compared to bird calls are usually emitted by male birds. Songbirds are a popular model for the study of vertebrate acoustic communication [1-2].

The purpose of this paper is to present preliminary results of learned call types of 10 species of Sri Lankan resident birds, namely, Brown-Capped Babbler (Pellorneum fuscocapillus), Black Naped Monarch (Hypothymis azurea), Chestnut Backed Owlet (Glaucidium castanonotum), Green Billed Coucal (Centropus chlororhynchos), Emerald Dove (Chalcophaps indica), Sri Lanka Orange Billed Babbler (Turdoides rufescens), Sri Lanka Scaly Thrush (Zoothera imbricata), Sri Lanka Myna (Gracula ptilogenys), Yellow Browed Bulbul (Iole indica) and Sri Lanka Yellow Fronted Barbet (Megalaima flavifrons).

There are a number of related studies available in the literature on using neural networks to study vocalizations. Murray et. al. [3] successfully carried out a study by using unsupervised, self-organizing neural networks to categorize the repertoire of false killer whale vocalizations. They have used two-dimensional characterization of false killer whale vocalizations (each vocalization was characterized by a sequence of short-time measurements of duty cycle and peak frequency) as inputs for the neural networks. In another study, an artificial neural network was used to classify black-capped chickadee (Poecile atricapillus) call note types [2]. The authors have used 9 acoustic features as the input to the network to identify call types with more than 98% accuracy.

Selin et. al. [4] discusses the use of wavelets in recognizing inharmonic and transient bird sounds efficiently. They have used shift invariant feature vectors calculated from the wavelet coefficients as the inputs for two neural networks; the unsupervised self-organizing map (SOM) and the supervised multilayer perception (MLP). Their results

Proceedings of the Technical Sessions, 28 (2012) 100-105
Institute of Physics – Sri Lanka

show that SOM and MLP networks are capable of classifying test sounds with an accuracy of 78% and 96% respectively.

In a recent study, the use of artificial neural networks in identifying different call types of Black Lemurs have been discussed [5]. Linear productive coding was used to extract the features from the sound tracks. Their results show that neural networks were able to recognize all vocal categories with high accuracy (92.5–95.6%) and outperform statistical techniques (76.1–88.4%). They also suggest that neural networks can be used as an effective and robust tool to understand the primate vocal communication.

2. METHODOLOGY AND IMPLEMENTATION

2.1 Data sample:

The sound tracks of different birds were provided by the Department of Zoology, University of Colombo. All recordings had been collected from the Sinharaja Natural World Heritage Site (6°21–26′ N; 80°21–24′ E) in Sri Lanka, which encompass 11,187 ha of lowland evergreen rainforest characterized as a Mesua–Shorea community [6]. The recordings have been collected using Marantz (PMD222) Digital Recorder and Unique directional mike standard parabola. All recorded sound files were in .wav format.

2.2 Noise reduction:

Since the sound tracks were recorded in a rainforest, high level of background noise is present in the data set. In these types of forests, the vegetation is densely packed causing sound reverberations. There are many different bird species as well as a number of other creatures such as cricketers produce noise. The climate is also a crucial factor; rain can cause significant interference and wind causes leaves to fall and interfere through most of the acoustic frequencies. All of these factors limit the quality of the sound recordings, making the bird species recognition a more complicated process and requiring the introduction of different filtering techniques in order to obtain suitable results. The background noise and the unwanted vocalizations were removed using the software package “Camtasia Studio[1]”.

2.3 Segmentation:

After removing the noise, sound clips were segmented into smaller pieces where each segment contains a single call of the bird. This task was achieved by manually listening to the audio clip. Then the filtered audio clips were normalized using a sound editor. Normalization is carried out to adjust the volume so that the loudest peak is equal to (or a percentage) the maximum signal that can be used in the digital audio. Usually the sound file should be normalized to 100% at the last stage in production to make it clear without distortions. This makes all sound clips equal in amplitude eliminating the sound attenuation due to distance.

[1] http://download.cnet.com/Camtasia-Studio/3000-13633_4-10665109.html

Fifty (50) calls from each bird were selected for the study. Each sound segment was loaded into the Matlab environment. All sound files which were in the form of .wav format were converted into .m files for further processing.

2.4 Feature extraction:

In order to extract features, power spectrums were calculated through Fast Fourier Transformation (FFT) for different call types of different birds. The frequency range was converted into log scale and categorized into 10 sub groups of each being 0.5. Since the frequency range was considered in log scale instead of linear scale the categorization became easier. After dividing into sub sets, sum of amplitude values corresponds to each frequency was obtained for each subset. Prior to being presented to the network the feature values were normalized to avoid the potential problems when training the network.

2.5 Network training:

A feed-forward neural network was constructed using the Matlab neural network toolbox. Initially, the network was constructed with a five hidden layers. The neural network output had 10 nodes (each representing a different bird) and 10 inputs for each bird (which represented the normalized amplitude values in different frequency ranges). Out of 50 samples available, 35 were used as the training set. As the study was carried out to classify 10 birds [10×35] matrix was inputted to the network.

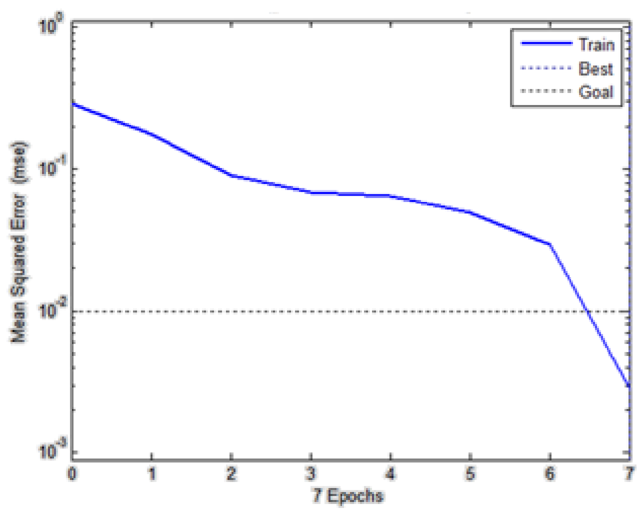

Fig. 1: Training performance of the neural network

Tan sigmoid transfer function was applied to the first layer (input layer) while log sigmoid transfer function and pure linear transfer function was applied to the hidden layer and the output layer respectively. Supervised learning was used with Levenberg-Marquardt (trainlm) training algorithm. The network architecture was tested with different transfer functions, different training algorithms (trainrp, traincgf, trainscg) and different number of output layers. Number of epochs was set to 1000 and the goal (Mean Squared Error) was set to 0.01 in each training process. Fig. 1 shows the training performance of the neural network.

Proceedings of the Technical Sessions, 28 (2012) 100-105
Institute of Physics – Sri Lanka

Network was trained to identify the vocalization types of different birds. To determine the vocalization types, output of the network was rounded off and compared with the target values. Testing set consisted with the remaining 15 samples.

In the first trial, the training set was introduced in a particular order (first 35 belonged to training set and next 15 testing set). In the second trial, the network was trained by selecting the training set randomly by keeping the same composition. Final result did not show a significant difference in both these cases.

3. RESULTS AND DISCUSSION

For each sample, six acoustic features, namely, SF-start frequency, EF-end frequency, PF-peak frequency, AD-ascending duration, DD-descending duration and TD-total duration were calculated. These measurements were employed by Dawson et al. [2] to classify black-capped chickadee call note types successfully. Average values of these parameters for selected 50 sound samples for each bird is shown in Table 1.

Table 1: Acoustic Features of the 10 birds (average value)

	Brown-Capped Babbler	**Black Naped Monarch**	**Chestnut Backed Owlet**	**Green Billed Coucal**	**Emerald Dove**	**Sri Lanka Orange Billed Babbler**	**Sri Lanka Scaly Thrush**	**Sri Lanka Myna**	**Yellow Browed Bulbul**	**Sri Lanka Yellow Fronted Barbet**
SF(Hz)	2,395	1,396	646	288	89	1,097	4,390	1,693	1,203	898
EF(Hz)	3,392	5,886	1,607	898	1,009	1,980	14,250	6,535	4,015	1,898
PF(Hz)	2,675	3,394	1,065	334	453	1,943	7,761	3,614	2,021	1,406
AD(s)	0.440	0.605	0.088	0.957	0.158	0.588	0.750	0.092	0.624	0.068
DD(s)	0.501	0.603	0.093	1.230	0.221	0.412	0.187	0.281	0.615	0.213
TD(s)	0.920	1.204	0.202	2.197	0.380	1.020	0.218	0.372	1.243	0.285

Fig. 2 shows the average starting frequency, end frequency and peak frequency for the 10 birds. This clearly shows the variation in frequency ranges of birds. Sri Lanka Scaly Thrush has the maximum peak frequency of 7,761 Hz while the Green billed Coucal has the lowest peak frequency of 334 Hz and longest duration of 2.197 s. Even though two separate birds do not have exactly the same frequencies, a clear separation or direct relationship cannot be seen among the most prominent frequency parameters.

A significant difference can be observed between the call types of 6 birds; Brown Capped Babbler, Black Naped Monarch, Chestnut Backed Owlet, Green Billed Coucal, Sri Lanka Scaly Thrush and Sri Lanka Yellow Fronted Barbet. Therefore, in this work, FFT was used to obtain futures to identify only the 6 birds from their vocalizations.

Proceedings of the Technical Sessions, 28 (2012) 100-105
Institute of Physics – Sri Lanka

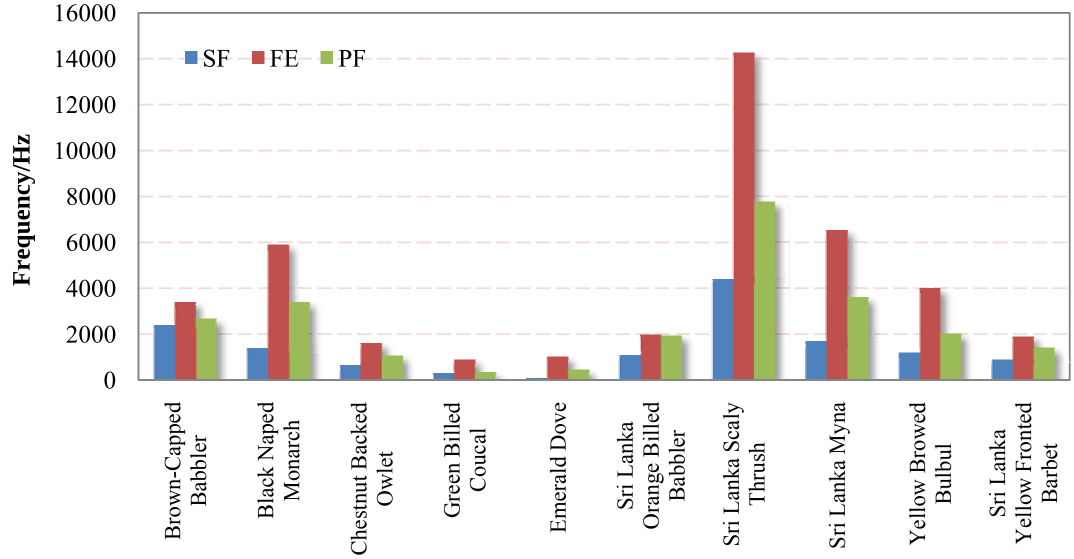

Fig. 2: Average frequency of the birds

Similarities between SF, FE and PF could be seen for Emerald Dove and Green Billed Coucal; Sri Lanka Orange Billed Babbler and Sri Lanka Yellow Fronted Barbet; Sri Lanka Myna and Black Naped Monarch; and, Yellow Browed Bulbul and Brown-Capped Babbler.

Fig. 3a and 3b shows Fourier spectrum of Sri Lanka Orange-Billed Babbler and Brown Capped Babbler respectively.

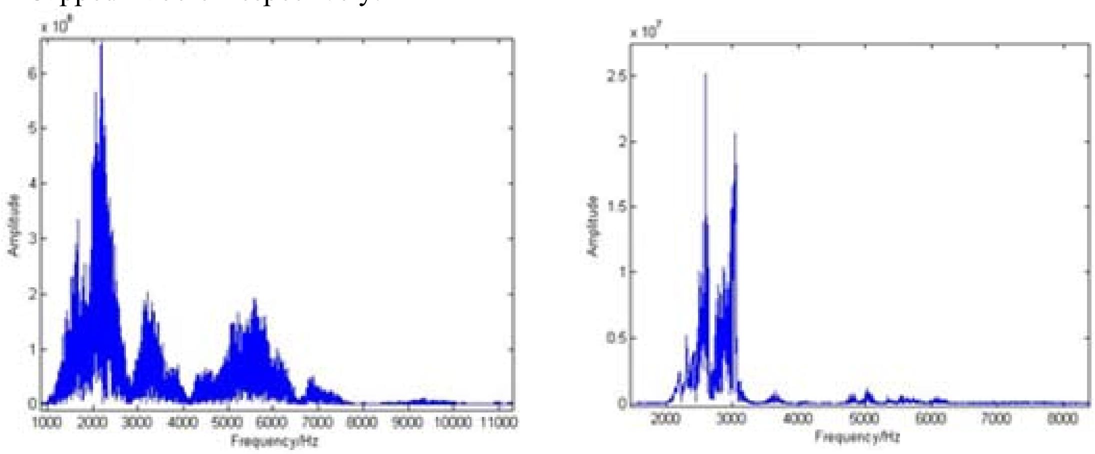

Fig. 3: (a) Fourier spectrum of Sri Lanka Orange-Billed Babbler (b) Fourier spectrum of Brown Capped Babbler

When training the neural network, with 7 iterations the neural network reached the expected goal for the selected six different bird species. The network produced a linear correlation coefficient of 0.994 between the input and output values. After the testing set was introduced, simulated output of the network was rounded off and the output of the network was compared with the expected (actual) output. The results are summaries in Table 2.

Proceedings of the Technical Sessions, 28 (2012) 100-105
Institute of Physics – Sri Lanka

Table 2: Detection accuracy

Bird	Accuracy
Brown-Capped Babbler	93%
Black Naped Monarch	93%
Chestnut Backed Owlet	93 %
Green Billed Coucal	100%
Sri Lanka Scaly Thrush	100 %
Sri Lanka Yellow Fronted Barbet	100%

4. CONCLUSION

The preliminary results discussed in this paper show that the developed neural network is capable of classifying different bird types with limitations. The network which uses FFT to extract frequency features was able to classify the Brown-Capped Babbler, the Black Naped Monarch, the Chestnut Backed Owlet, the Sri Lanka Scaly Thrush, the Green Billed Coucal and the Sri Lanka Yellow Fronted Barbet with an accuracy of more than 90%. However, improvements are required to separate the Emerald Dove and the Green Billed Coucal; the Sri Lanka Orange Billed Babbler and the Sri Lanka Yellow Fronted Barbet; the Sri Lanka Myna and the Black Naped Monarch; and, the Yellow Browed Bulbul and the Brown-Capped Babbler. Further development to this work is currently in progress.

ACKNOWLEDGEMENT

The authors wish to thank Mr. C.P. Ratnayake and Prof. S.W. Kotagama for providing the sound tracks of birds used in this study.

REFERENCES

1. Slater, P.J.B., *Fifty years of bird song research: a case study in animal behaviour*, Animal Behaviour, 65 (2003) 633–639.
2. Dawson, M.R.W., Charrier, I. and Sturdy, C.B., *Using an artificial neural network to classify black-capped chickadee (Poecile atricapillus) call note types*, J. Acoust. Soc. Am., 119 (2006) 3161–3172.
3. Murray, S.O., Mercado, E., and Roitblat, H.L., *The neural network classification of false killer whale (Pseudorca crassidens) vocalizations*, J. Acoust. Soc. Am., 104 (1998) 3626–3633.
4. Selin. A., Turunen. J. and Tanttu J.T., *Wavelets in Recognition of Bird Sounds*, J. Advances in Signal Processing, (2006) 2-9.
5. Pozzi, L., Gamba, M., and Giacoma, C., *The Use of Artificial Neural Networks to Classify Primate Vocalizations: A Pilot Study on Black Lemurs*. Am. J. Primatol., 72 (2010) 337–348.
6. Ratnayake. C. P., Goodale. E. and Kotagama. S. W., *Two sympatric species of passerine birds imitate the same raptor calls in alarm contexts*, Naturwissenschaften, 97 (2009) 103-108