Available online at www.ijarbest.com

International Journal of Advanced Research in Biology, Ecology, Science and Technology (IJARBEST)
Vol. 1, Issue 2, May 2015

Automatic Identification Of Birds Through Audio Spectral Analysis

Aparna P. C.[1]
Post-Graduate Scholar, ECE department, FISAT, Ernakulam, India[1]

***Abstract*— In this paper, we focus on identification of birds through their sounds. Bird species identification is gaining importance in field of ecological conservation and ornithology. . From three unknown bird sounds we identify a particular bird (Forest Owlet) through spectrum analysis. Since Forest Owlet is a critically endangered species in central India, its identification is important in their fragmented population census and thereby their conservation. Here, four different frequency domain analysis technique, viz., Mean Square Error (MSE) approach, Correlation analysis based on frequency shift and symmetry property, Wiener Filter theory and Mel Frequency Cepstral Coefficients (MFCC) approach are used. This paper present the comparison of these methods when implemented in MATLAB. Recorded bird calls from xentocanto website have been used in the above analysis.**

Index Terms*— Mean Square Error (MSE) approach; Correlation Analysis; Mel Frequency Cepstral Coefficients (MFCC) approach

I. INTRODUCTION

A variety of creature's sounds, including human speech, dog barks, bird songs,frog calls, cricket calls, etc can be heard. People often fail to see animals but they hear them. It is noted that vocalization of each animal is different from another. Therefore, this unique property of vocalizations can be used to automatically identify animal species .Birds are numerous and easier to monitor than other species. There are many practical reason behind the monitoring and locating a particular bird. First, ornithologists are always curious to locate a bird and study about it. Second, birds are endangered by different human activities like deforestation, poaching and overgrazing. Identification of a bird by their songs help in population census and thereby acts as an aid for conservation of bird species. Third, birds create security issue near airports. Hence identification of birds and their population census/control is important in present scenario.

The basic principle is that each bird song is different when analyzed in frequency domain. In this project, six audio samples are used, i.e., one unknown and five reference samples and our task is to identify to which reference samples the unknown sample belong is. Four methods are used for audio analysis viz., Mean Square Error (MSE) approach, Correlation analysis based on frequency shift and symmetry property, Wiener Filter theory and Mel Frequency Cepstral Coefficients (MFCC).

Here bird under study is Forest Owlet. It is a rare species which was thought to extinct but fortunately rediscovered after 113 year in Maharashtra(1997). Since their rediscovery the population census of this species is a challenge to the forest authorities and ecologists. Setting up of this real time system in forest will be a helping hand to those who are interested in conservation and rehabilitation of these Owlets.

II. BIRD SONG

A. Hierarchical levels of bird song

Spectrogram is representation of power distribution with respect to time and frequency axis. The resolution of the color indicates the gradient of the power distribution. The deeper color means the higher power distribution. Fig 3.3 shows spectrogram representation of a Common Chaffinch song. Phrases, syllables and elements or notes are hierarchical levels of bird song. Building block of bird song is called element, which is the smallest separable element in spectrogram. Syllables are produced by one or more elements or notes. Structure of syllables varies a lot and the number of elements in syllables also varies. When a series of syllables occur together in a particular pattern it is called a phrase. Syllables in a phrase are typically similar to each other. A song is constructed of a series of phrases. Each song is different from another in terms of their spectrogram composition.

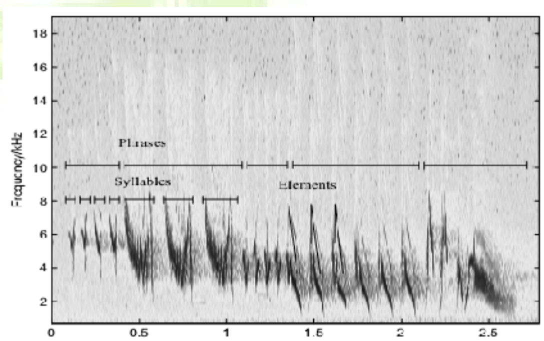


Fig. 1. Hierarchical levels of common chaffinch song

International Journal of Advanced Research in Biology, Ecology, Science and Technology (IJARBEST)
Vol. 1, Issue 2, May 2015

B. Recordings and their sources.

Bird recording are obtained from Xeno-Canto which is a citizen science project. In this volunteers record, upload and annotate recordings of birdsong and bird calls. All the recorded signals that are used i.e., both reference and unknown signals are from Melghat Tiger Reserve and owl species that are listed below often misrecognized. The audio analysis of the below enlisted five species and their comparison yields better recognition than manual observations.

Signal used	Name of Owl
Signal 1	Forest Owlet
Signal 2	Mottled Wood Owl
Signal 3	Jungle Owlet
Signal 4	Collared Scops Owl
Signal 5	Oriental Scops Owl

TABLE I LIST OF REFERENCE SIGNALS

The Forest owlet (Athene blewitti) is an endemic owl that belongs to the shrinking forests of central India. This bird is on the verge of extinction and is listed as critically endangered. Population is very low and fragmented in Central India and their identification is very difficult. The major threat to forest owlet population is the loss of habitat due to conversion of habitat to agricultural land, frequent forest fires, encroachment and poaching and also due to bling beliefs.

Response to Call Back methods are used for the species which are not easily heard in their habitat. A prerecorded song is played and we wait for bird's response. If a response is found, then the bird is identified. But it's a traditional and old technique which has many disadvantages. It creates ecological imbalance and miscommunication between different species. This paper aims at providing a relevant and reliable technique for identification of birds through audio spectral analysis.

III. METHODOLOGY

MSE based approach, Correlation analysis based on frequency shift and symmetry property, Wiener Filter theory and MFCC based approach are the methods used for audio analysis.

A. Common Steps for 3 Analysis

1) Use "wavread" command in MATLAB to read the audio file.
2) Use "spectrogram" function to process audio signals and get returned matrix signals.
3) Transpose the matrix signals for rows and columns, take "sum" operation of the matrix and get a returned row vector for each column summation result. This row vector is the frequency spectrum signal.
4) Normalize the frequency spectrums by the linear normalization by using the equation.

$$y=(x-MinValue)/(MaxValue-MinValue) \quad (1)$$

Here for the simplicity of drawing flow chart 2 reference signals are only considered but 5 reference signals are used in actual audio analysis .This is true for the rest of flowcharts too. After these steps, we obtain spectra and these spectra are given as input to formerly mentioned three methods.

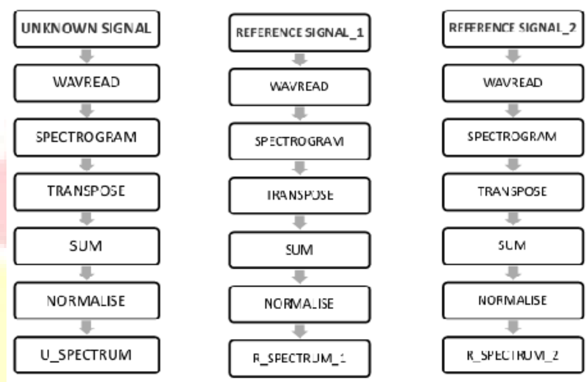


Fig. 2. Flow Chart showing initial steps of 3 analysis

B. MSE Approach

Basic idea behind this analysis is different bird songs have different spectra. One of reference spectra is taken and maximum value is found. The maximum so found is shifted to the maximum of unknown spectrum. Then we calculate error between the spectra (subtracting the spectra) followed by the calculation of mean square value of error. The procedure is repeated for the rest of reference spectra. One of the reference spectrum yield minimum mean square error, then it is concluded that unknown signal matches to that reference signal.

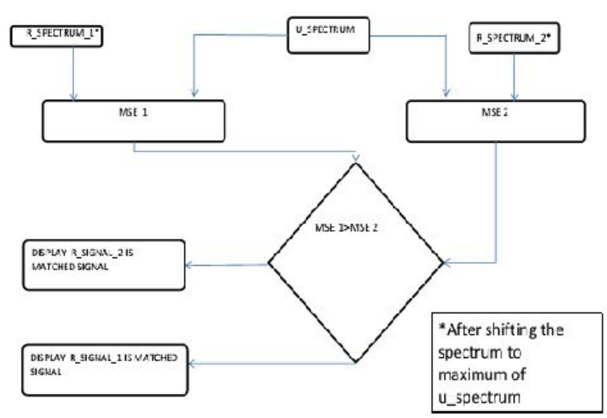


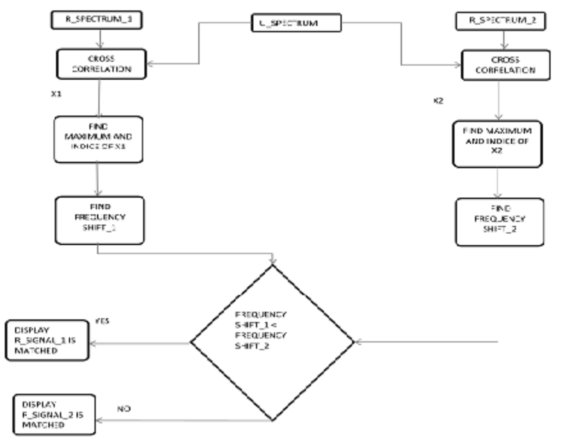
Fig. 3. Flow Chart showing MSE approach

C. Correlation Analysis

To realize recognition of bird through audio spectrum analysis , there is a need to compare the unknown spectrum with five reference spectra. A check is made to look which reference spectrum matches with unknown spectrum. In this method cross correlation between unknown signal w.r.t five reference signals are found and then either of the following methods are chosen for analysis:

1) Correlation Analysis based on Frequency Shift: Frequency shift is difference between middle point and maximum value point. Correlation between two similar spectra are found then frequency shift found is minimum and the vice versa is true in the case of two different spectra.

Fig. 4. Flow Chart showing Correlation Analysis based on Frequency

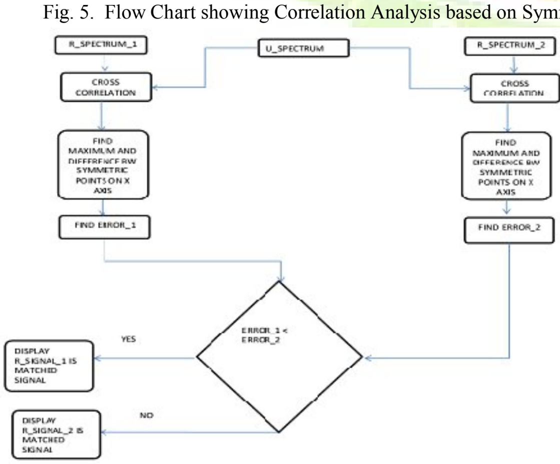


Shift

In this technique first cross correlation between reference and unknown spectrum is found, secondly frequency shift is found. The reference signal the unknown signal at is matched with u is the one that gives least frequency shift value.

2) Correlation Analysis based on Symmetry Property: Correlation between two similar spectra results in a new spectrum which is symmetric w.r.t middle point.

Fig. 5. Flow Chart showing Correlation Analysis based on Symmetry



Property

In this technique first cross correlation between reference and unknown spectrum is found, secondly we find the middle point and then we calculate error due to asymmetry. The matched signal is the one that gives minimum error.

D. Wiener Filter Theory

Wiener filter is used to estimate the desired signal d (n) from the observation process x (n) to get the estimated signal d (n)'. Assumption is that d (n) and x (n) are correlated and jointly wide-sense stationary.

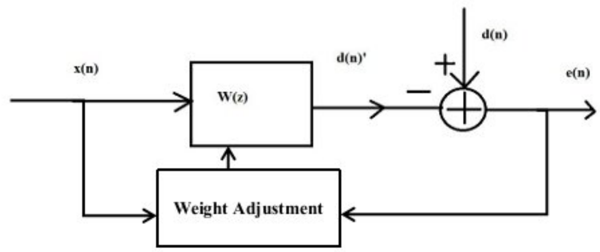


Fig. 6. Wiener Filter

1) Output d (n)' is the convolution of x (n) and w (n) and error of estimation is:

$$e(n)=d(n)-d(n)'=d(n)-\sum_{l=0}^{p-1} w(l)x(n-l) \qquad (2)$$

2) The purpose of Wiener filter is to choose the suitable filter order and find the filter coefficients with which the system can get the best estimation i.e. minimize the mean square error:

$$\xi=E\{|e(n)|^2\}=E\{|d(n)-d(n)'|^2\} \qquad (3)$$

3) Wiener Hopf Equation

$$r_{dx} = R_x W \qquad (4)$$

4) where R_x is input auto correlation matrix ,r_{dx} is cross correlation between input and desired signal and W is weight vector.

5) Minimum MSE

$$\xi_{min}=r_d(0)-\sum_{l=0}^{p-1} w(l) r_{dx}^{*}(l) \qquad (5)$$

Steps involved in this approach includes :

1) Find the auto-correlations of reference signals and find the cross-correlations for the unknown signal with the five reference signals

2) According to the Wiener-Hopf equation calculate the filter coefficients for both five reference signals.

3) Find the mean values of the minimum mean square-errors using (5) for each of the five filter coefficients.

4) As shown in the flowchart, comparing the minimum mean square-errors, the reference signal which gives the least error value will chosen as the matched one.

International Journal of Advanced Research in Biology, Ecology, Science and Technology (IJARBEST)
Vol. 1, Issue 2, May 2015

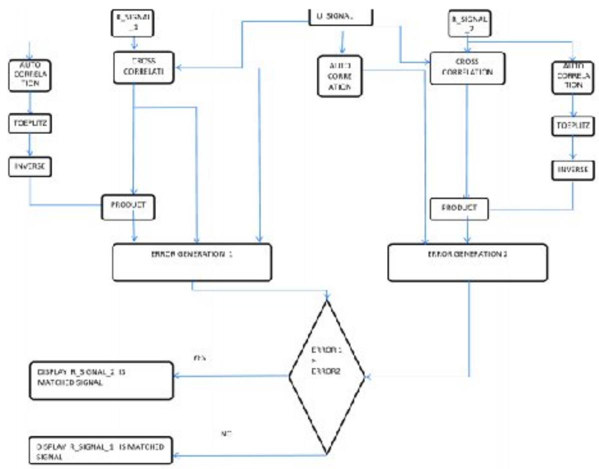


Fig.7. Flow Chart based on Wiener Filter Theory

E. MFCC Approach

This is a technique for recognition of a bird based on feature extraction. Here a new scale called Mel scale which characterizes the human ear perception of frequency is introduced. This scale is a linear frequency spacing below 1000 Hz and a logarithmic spacing above 1000 Hz.

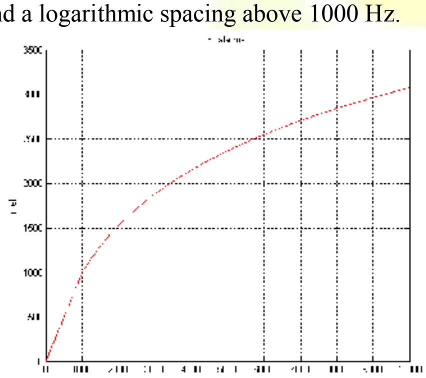


Fig. 8. Graph showing relation between mel and hertz

- for converting from frequency to Mel scale is:
  $M(f)= 1125\ln(1+f/100)$ (6)
- for converting from Mel scale to frequency is:
  $M^{-1}(m)=700(\exp(m/1125)-1)$ (7)

.

1) Frame the signal into short frames.

2) Multiply the frames by hamming window
   Hamming window is given by equation (8)
   $h(n)=0.54-0.46\cos(2*\pi*n/N) \quad 0\leq n\leq N$ (8)

3) Calculate the FFT of windowed signal

4) Calculate f(m)
   $f(m)= (N/Fs) *Mel (Mel(fl)+m*p$ (9)

$p= (\{Mel(fh)-Mel(fl)\}/M+1)$ (10)

where N is the length of FFT, fl and fh are lowest and highest frequency, Fs is the sampling frequency, M is number of filters.

5) Apply the mel filterbank to spectrum obtained from step 3. The impulse response of this filter bank is

$$H_m(k)=\begin{cases} 0 & k<f(m-1) \\ k-f(m-1)/f(m)-f(m-1) & f(m-1)\leq k\leq f(m) \\ f(m+1)-k\ /f(m+1)-f(m) & f(m)\leq k\leq f(m+1) \\ 0 & k>f(m+1) \end{cases} \quad (11)$$

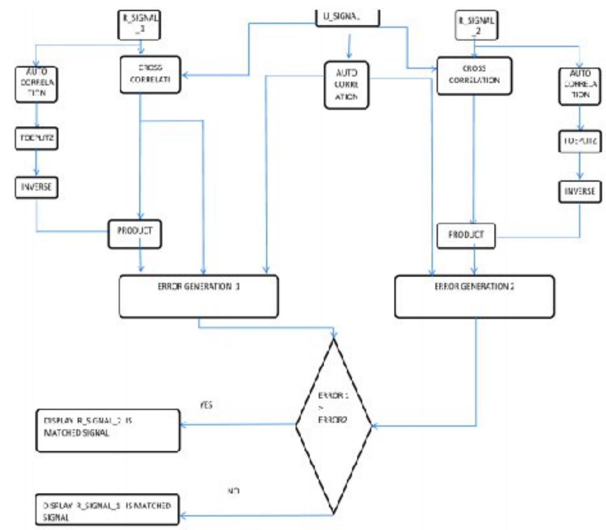


Fig. 9. Flow chart for MFCC comparison

MFCCs of reference signals and unknown signal is calculated by systematically following the above listed steps. Then Euclidean distance of MFCCs of unknown signal from each reference signal is calculated and the minimum value of Euclidean distance is found. It is concluded that unknown signal matches to the reference signal that yield minimum value of Euclidean distance.

IV. RESULTS AND DISCUSSIONS

Five inputs for the audio analysis were collected and fed into a computer in wave file format. These audio file were taken into Matlab through a wavread command. Then MSE based approach, Correlation analysis based on frequency shift and symmetry property, Wiener Filter theory and MFCC based approach were coded in Matlab and following results were obtained. The final graph were plotted and tables were drawn to indicate results in a more compact manner.

Results shows that signal-1 (Forest Owlet) matches with unknown signal. This implies that the unknown audio belong to Forest Owlet and thus recognition of Forest Owlet is done by audio spectrum analysis.

Available online at www.ijarbest.com

International Journal of Advanced Research in Biology, Ecology, Science and Technology (IJARBEST)
Vol. 1, Issue 2, May 2015

A. MSE Approach

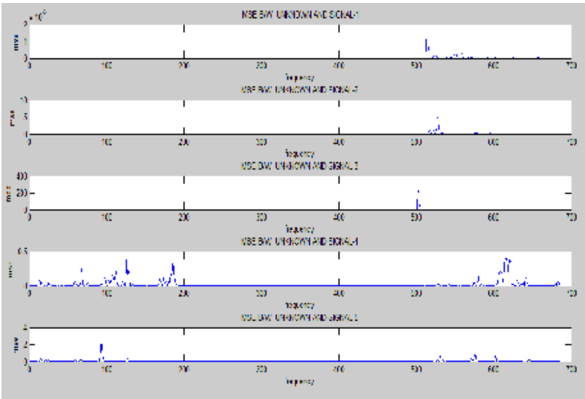


Fig. 10. MATLAB output for MSE based method

TABLE II LIST OF AVERAGE MSE FOR DIFFERENT REFERENCE SIGNALS

Average MSE	Value
signal -1	0.0000
signal -2	0.0380
signal -3	0.8037
signal -4	0.0192
signal -5	0.0341

From the table it is clear that signal -1 is giving the lowest average of mean square error and hence it is concluded that the unknown signal is matching to forest owlet.

B. Correlation Analysis

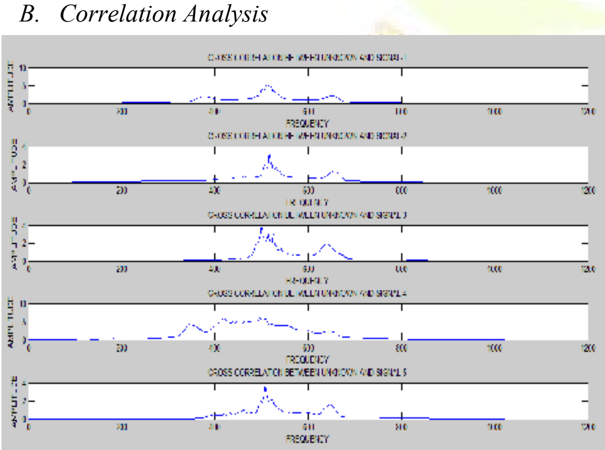


Fig. 11. MATLAB output for Correlation Analysis

Two methods based on frequency shift and symmetry property find the cross correlation between reference spectra and unknown spectrum and the technique for matching is quite different. From the table it is clear that signal -1 is giving the lowest error due to asymmetry and minimum frequency shift hence the unknown song is identified as Forest Owlet's song.

Cross Correlation	Frequency Shift	Error due to asymmetry
signal -1	1	0.0132
signal -2	5	0.0566
signal -3	12	0.3696
signal -4	3	1.2826
signal -5	4	0.1254

TABLE III. LIST OF FREQUENCY SHIFT ERROR DUE TO ASYMMETRY FOR DIFFERENT REFERENCE SIGNALS

C. Wiener filter theory based Approach

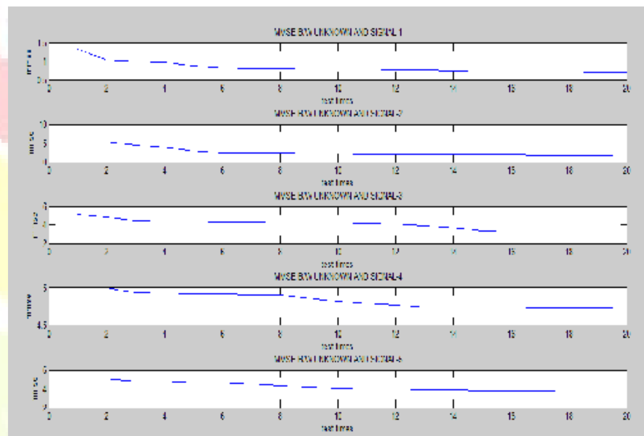


Fig. 12. MATLAB output for Wiener Filter Theory Based Method

TABLE IV. LIST OF MINIMUM MEAN SQUARE ERROR(MMSE) FOR DIFFERENT REFERENCE SIGNALS

From the table it is clear that signal -1 is giving the lowest

MMSE	Value
MMSE b/w unknown and signal -1	0.8323
MMSE b/w unknown and signal -2	2.6451
MMSE b/w unknown and signal -3	4.0128
MMSE b/w unknown and signal -4	4.8340
MMSE b/w unknown and signal -5	4.2158

MMSE and hence the unknown signal is identified as forest Owlet's song.

International Journal of Advanced Research in Biology, Ecology, Science and Technology (IJARBEST)
Vol. 1, Issue 2, May 2015

D. MFCC Approach

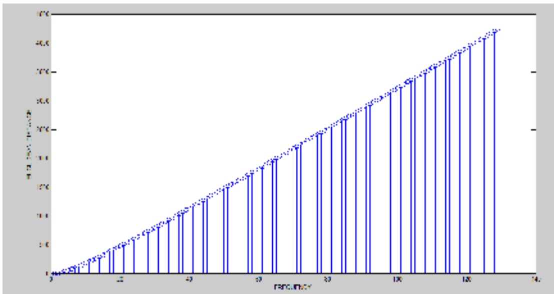
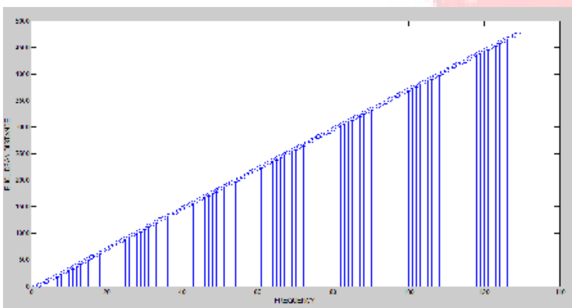


Fig. 13. MATLAB output showing Euclidean distance b/w MFCCs of reference signal-1 and unknown signal

Fig. 14. MATLAB output showing Euclidean distance b/w MFCCs of



reference signal-1 and unknown signal

Figure 13 shows Euclidean Distance between MFCC if signal -1 and unknown signal and the distance value is found to be 5.6827e+006 and Figure 14 shows Euclidean Distance between MFCC if signal -2 and unknown signal and the distance value is found to be 7.6730 e+006. From these results it is found that signal 1 matches to unknown signal more than signal-2. Hence it can be concluded that unknown signal belongs to Forest Owlet

V. CONCLUSION

Identification of a bird based on its song/call was a difficult task. This paper throws light into the study of audio spectrum of a particular bird through different frequency analysis techniques. Here Forest Owlet species identification is given emphasis and hence its conservation is highlighted. The reason why this bird is chosen is that it is a rare and unique species and its population census is very important for its protection. Four unique techniques were discussed and results derived from these methods were quite satisfactory. Here five reference signals were used and these methods helps the observers to identify to which reference signal the unknown signal belong to. It was only possible for a keen bird watcher or well-experienced researchers to do so. But today the technology have evolved in such a way that experience and skill is not necessary for bird identification, all we require is a chirp of a bird and tool for frequency analysis .The major disadvantage of this technique is that the execution time varies from seconds to minutes depending on the size of audio file. So in future this time can be reduced through creation of database and already training the signal as in the case of neural networks.

Acknowledgment

I would like to express my sincere gratitude towards Dr. Karthikeyan Chittayil, Visiting Professor , Electronics & Communication Department, Federal Institute of Science And Technology for his kind support and motivation which filled in me confidence to complete this work.

REFERENCES

[1] Pramanik A. , Raha R. ,"Automatic Speech Recognition using Correlation Analysis", in IEEE Conference on Information and Communication Technologies (WICT),pp.670-674, Oct. 30 2012-Nov. 2 2012

[2] Deepak , Vikas Mittal ,"Speech Recognition using FIR Wiener Filter", in International Journal of Application or Innovation in Engineering & Management (IJAIEM), pp. 204-208,Volume 2, Issue 5, May 2013.

[3] Seppo Fagerlund ,"Automatic Recognition of Bird Species by Their Sounds", Master's Thesis Report, Nov 8, 2004.

[4] A. Harma," Automatic identication of bird species based on sinusoidal modeling of syllables," in Proceedings of IEEE Int. Conf. Acoust., Speech, and Signal Processing, Hong Kong, April 2003, vol. 5,pp. 545-548

[5] Marcelo T. Lopes, Lucas L. Gioppo, Thiago T. Higushi, Celso A. A. Kaestner,Carlos N. Silla Jr.," Automatic Bird Species Identification for Large Number of Species "in IEEE International Symposium on Multimedia,2011,pp. 117-122

[6] Xeno-Canto Web Site, <http://xeno-canto.org/>

Aparna P. C. received the B.Tech degree in Electronics and Communication Engineering from Kannur University, Kerala, India, in 2013. Currently, she is post graduate student with the Department of Communication Engineering, Federal Institute of Science and Technology (FISAT), Kerala, India. Her current research area includes speech processing.