Nandwana, M., Robin, V. V., Sinha, A. and Prabhakar, A.
2014
Identifying birds with species detection algorithms.
Final Technical Report
submitted to
CEPF-ATREE Western Ghats Program

CHAPTER 1

Introduction

1.1 Introduction

Our aim is to build an automated system which can identify the bird call present in a continuous recording and can classify which species it belongs to. It has always been a challenging task to design such system which can identify the bird calls without manual intervention.

Generally manual inspection of calls requires multiple numbers of experts, which involves a lot of money and is not reliable too. Human expert inspects spectrogram of calls for identification of the bird for that call. While recording some background noise or clubbing of bird calls makes it difficult for manual identification of bird calls, this gives rise to develop a reliable automatic system for the identification of calls.

Figure 1: Basic block diagram of an automated bird call identifier.

A basic block diagram of an automated bird call identifier is shown in Figure 1. This system takes bird call as an input and after processing gives us the name of the bird species for that call as output.

Earlier it was not an active field so there is no common framework has been proposed to do this research. One more thing is research into the issue of bird call identification has not gained much traction in the scientific community.

Recently interest towards automatic recognition of bird species based on their vocalizations has increased. Some methods for bird call identification have been proposed. Techniques based on speech recognition, speaker verification or speaker identification was used for the identification of bird calls. Some of them are dynamic time wrapping (DTW), Gaussian mixture model (GMM), hidden Markov models (HMM), FFT spectrum, spectrograms, and Wigner-Ville distribution (WVD).

All the techniques which were mentioned above were originally developed for human speech recognition and at later stage people used them for identification of bird calls. But we will see that human speech production mechanism is much more difficult as compare to the bird call production mechanism. One more problem in using all these techniques is that they are very complicated and they need a high end processor for their implementation, which makes is much more expensive too.

Thus by keeping all the above points in our mind we have developed a completely new algorithm only for bird call recognition. Our algorithm is fast, simple and easy to implement on a low end processor. Thus makes it much more suitable for bird call identification.

In this report a new technique of bird call analysis is presented, which is based on the spectral ensemble average voice prints (SEAV) developed for bird call analysis. SEAV computes the ensemble averages based on FFT spectrum of each bird, so it is called spectral ensemble average voice print of that particular bird. We proposed and implemented some important modification in SEAV based on our analysis because this time the experiments were done with the field recordings.

In our proposed method, with the help of recorded data we first created the templates of different bird species. All these different templates were used for the identification and classification of the calls. The results were encouraging.

1.2 Application of Bird Call Identification

Applications of bird calls identification can be found in the following areas:

- Meaningful research on taxonomy
- Monitoring of bird migration in ornithology: - An ornithologist may be interested in finding out whether a particular species has appeared in a particular region. He may also interest in finding out the population of a particular species in a region or number of different species present in that region.
- Significance in biological studies
- Bird watching: - Many people have bird watching as their hobby.

1.3 Tools and System Specification

All the experiments were run on 32 bit OS Microsoft windows XP. The system was having Intel Core 2 Duo CPU E7500 at 2.93 GHz and 2GB RAM. Mathworks MATLAB® R2009b was used for all the simulations. Audacity 1.3 was used for separating the calls from the recordings.

1.4 Organization of Report

We have divided our work in six chapters. First chapter deals with the introduction, applications and tools used for experiments.

In second chapter the vocalizations of human and birds are compared and complexity of human speech compare to bird calls is shown. A brief introduction of all the previous approaches is given we have also discussed about the need of a completely new method especially for bird call recognition.

Third chapter deals with SEAV and its modification for processing of field recorded data. Modifications of SEAV are discussed in details.

Implementation of modified SEAV on field recorded data is discussed in fourth chapter. Implementation of SEAV is done in two steps source template creation and then tests template creation and classification.

In fifth chapter the results of MATLAB tests are presented. The accuracy of modified SEAV is also given for overall system.

Last chapter is mainly about future scope of proposed work. A number of ideas with are given in this chapter. Appendix includes MATLAB codes.

CHAPTER 2

Call Production Mechanism and Past Approaches

This chapter deals with vocalizations of human and bird. Before we discuss our bird call recognition technique it is important to know about human speech production mechanism and bird call production mechanism. So in this chapter first will discuss the human speech production mechanism followed by bird call production mechanism. After that with the help of spectrogram comparison of both bird call and human speech we will try to show the complexity of human speech with respect to bird call. At the end past approaches used for identification are also discussed.

2.1 Human speech production mechanism

Human speech production mechanism is very complex. In Figure 2 a schematic diagram of human speech production mechanism is shown.

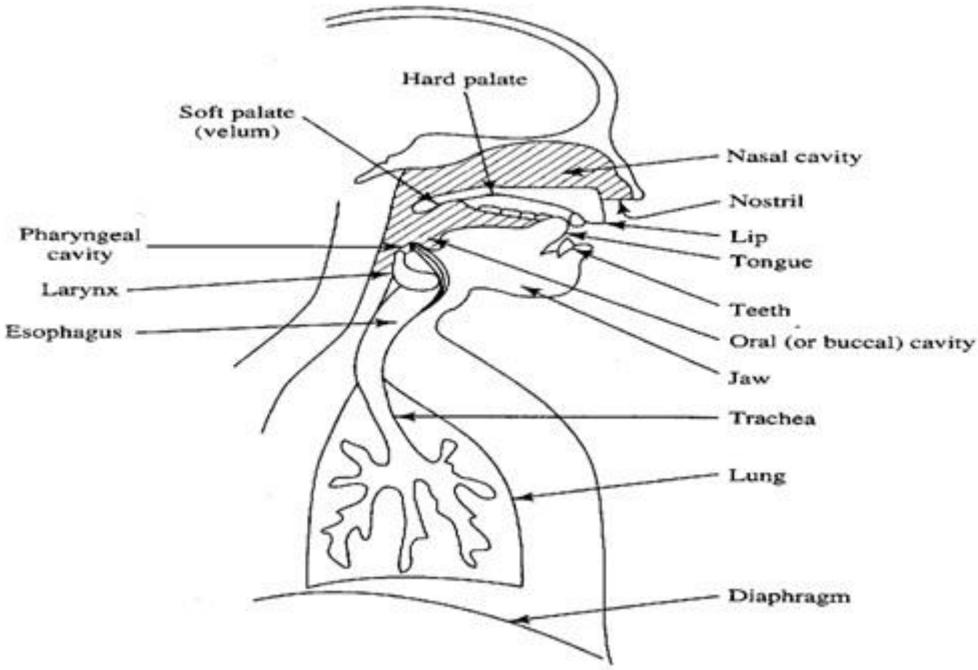

Figure 2: A schematic diagram of human speech production mechanism.

The vocal tract begins at the opening of the vocal cords, or glottis, and ends at the lips. The vocal tract consists of the pharynx (the connection from the esophagus to the mouth) and the mouth, or the oral cavity. In the average male, the total length of the vocal tract is about 17 cm. The cross-sectional area of the vocal tract, determined by the position of the tongue, lips, jaw, and velum, varies from zero (complete closure) to about 20 cm^2. The nasal tract begins at the velum and ends at the nostrils. When the velum, (a trapdoor-like mechanism at the back of the mouth cavity) is lowered, the nasal tract is acoustically coupled to the vocal tract to produce the nasal sounds of the speech.

Air enters the lungs via the normal breathing mechanism. As the air is expelled from the lungs through the trachea (or windpipe), the tensed vocal chords within the larynx are caused to vibrate (in the mode of relaxation oscillator) by the air flow. The air flow is chopped into quasi-periodic pulses which are then modulated in frequency in passing through pharynx (the throat cavity), the mouth cavity, and possibly the nasal cavity. Depending on the positions of the various articulators (i.e. jaw, tongue, velum, lips, mouth), different sounds are produced.

2.2 Bird vocalizations

Bird vocalizations are divided into two subcategories:

(1) Bird songs
(2) Bird calls

2.2.1 Bird songs

Bird songs are very long in length and they generally use it to attract their mates. Bird songs can be further subdivided into phrases, syllables and elements. Often, syllables will consist of a single element, but sometimes they are more complex. Phrases are usually made up of repetitions of a syllable, and the song itself is a series of phrases.

2.2.2 Bird calls

Bird calls are short and they are used for communication such as alerting other birds of a predator. Various bird species have unique calls. These bird calls are distinct based on inflection, length, and context, meaning the same bird may have more than one call. Our work is focused on bird calls only.

2.3 Bird call production mechanism

Majority of birds produce sounds as a result of sound waves originating from channels of air flow within the syrinx which is an organ located in the intersection of the main bronchi of the lungs and the trachea. The vibrations of the membranes within a birds syrinx which produce bird calls is very similar to the action of the human vocal chords which produces vowel sounds.

In Figure 2 a schematic bird call production mechanism is shown:

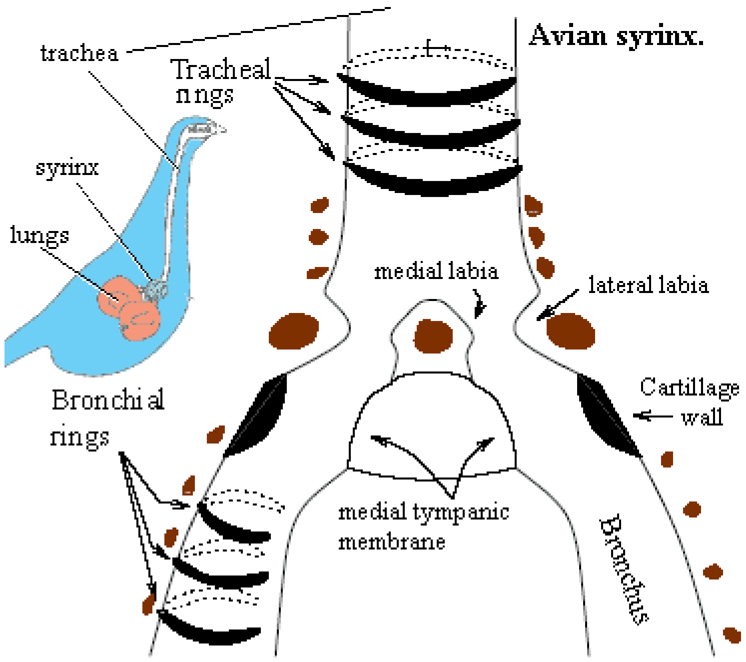

Figure 3: A schematic diagram of bird calls production mechanism.

Also, birds can produce sounds with a structure in the lungs, called a lyrinx, which allows them to produce two different tones at the same time. Various species of birds have unique bird calls. These bird calls are distinct based on inflection, length, and context, meaning the same bird may have more than one call.

2.4 Spectrogram comparison of human and bird vocalizations

The spectral properties of bird vocalizations are quite different from that of human speech. Bird vocalizations are generally very tonal, consisting of either a single frequency at a given point in time, or of a frequency and a few harmonics. This means most of the energy is located in a narrow region of frequency, whereas in human speech, the high energy regions are often distributed over a larger spectral range, due to fricatives and the formants in vowels.

In the following figures the spectrograms of human speech signal and bird call signal are shown.

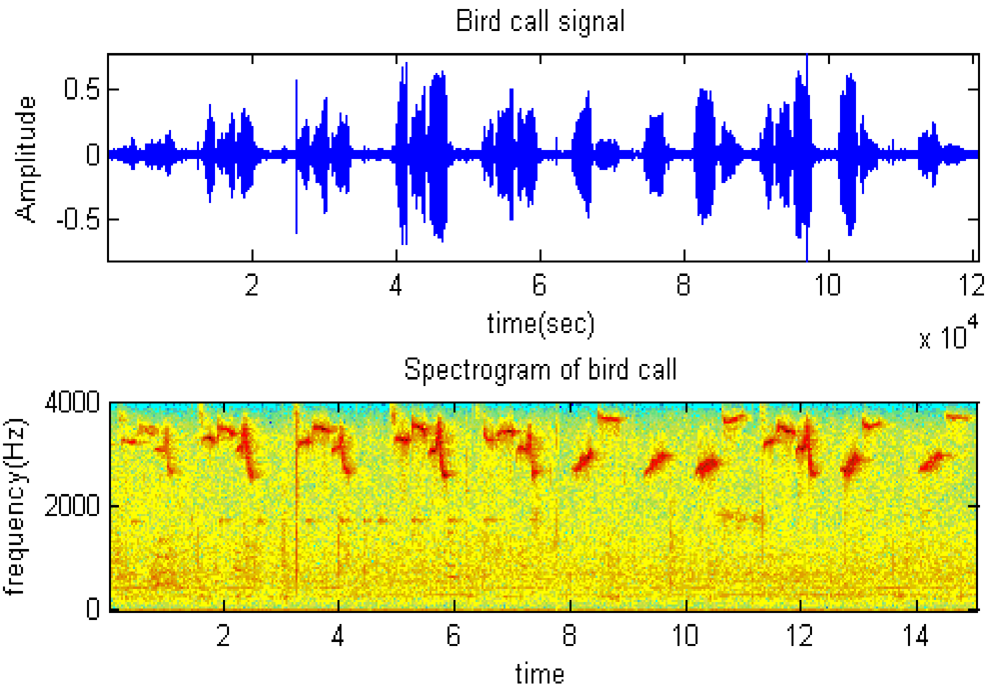

Figure 4: Bird call signal and its spectrogram.

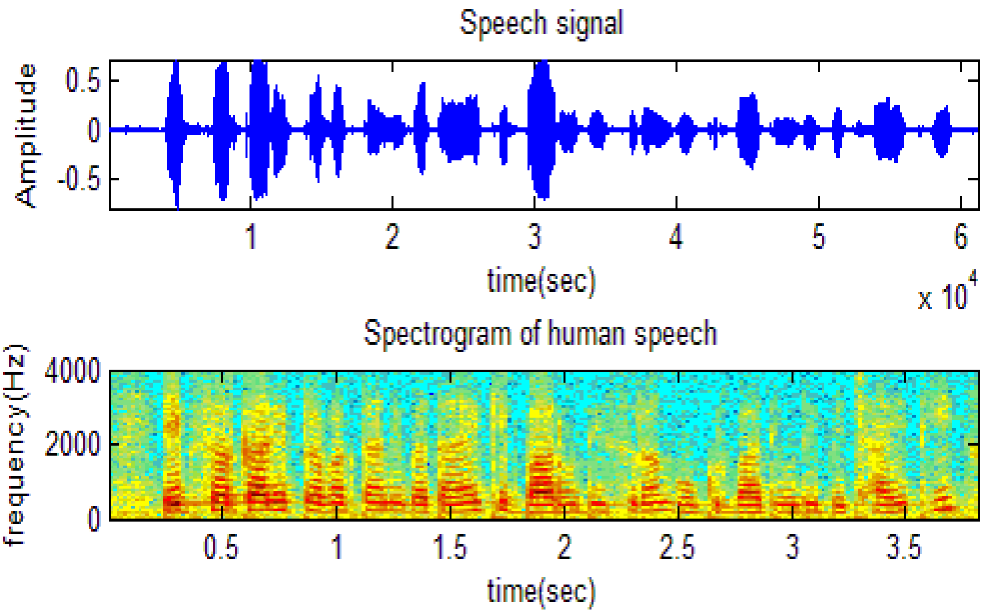

Figure 5: Human speech signal and its spectrogram.

In Figure 4 a bird call signal of White Bellied Short-Wing (WBS) with its spectrogram is shown. Where as a human speech signal with its corresponding waveform is also shown in Figure 5.

From the spectrograms we can see that birds have only one frequency at a time where as human has more than one frequencies occurring together. The frequency distributions corresponding to human vocalizations are much more complex than the bird vocalizations. From the above spectrograms we can easily find the complexity of human vocalization compared to bird calls. This shows that the bird calls are easier compare to human speech and can be processed with easy algorithms.

2.5 Previous approaches

The methods used for signal recognition in bioacoustics range from trained human listening to recording and/or visually inspecting spectrograms, to complex machine-based detections, measurement and classification algorithms. Each of these methods has advantages and limitations.

Trained observer can cue on suitable pattern differences and reliably identify and discriminate relevant sounds in acoustic recordings. However, given the quantity of data frequency collected during acoustic studies, relying on human experts is a rate-limiting step and is often impractical.

There are some of the techniques which were used in speech recognition were also used for bird call identification. Various signal processing techniques used in human speech recognition like FFT spectrum, spectrograms, Dynamic Time Wrapping (DTW), Wigner-Ville Distribution (VWD), Hidden Markov Models (HMM), and many other techniques have been used for identification of bird calls. After applying successfully these technique in speech recognition people directly used these techniques for identification of bird calls.

2.6 Need of a new algorithm

As discussed earlier from the production mechanism as well as from the spectrogram analysis it has been shown that human speech is much more complex compare to bird calls. So instead of using complex algorithms which were used for speech signals we can develop a new algorithm especially for bird call analysis.

The proposed algorithm should have following merits in it:

- Use lesser computation
- Easy to implement on a low end processor
- Cheap
- Accurate

By keeping all the points in mind, SEAV has been developed and modified. SEAV is discussed in detail in next chapter.

CHAPTER 3

Spectral Ensemble Average Voiceprints

3.1 Spectral Ensemble Average Voice Print (SEAV)

As discussed in the previous chapter generally Dynamic Time Wrapping (DTW) and Hidden Markov Models (HMM) are the techniques which are used in automatic identification of bird calls. These techniques are computationally expensive. It has already been mentioned that bird calls are easy for analysis compared to human speech. So they should be processed by some easy algorithms which require lesser computation.

In our approach towards the identification of calls we have used a new technique named Spectral Ensemble Average Voice Print (SEAV). The idea behind this technique is that each bird has different spectral content in its calls, so by use of this property we can identify their calls.

For each frame an *N* point FFT is computed. The value of *N* is computed as equal to or greater than the product of sampling rate and the window length. An ensemble average is then computed on the FFT spectrum, by taking the average of the corresponding FFT co-efficient of each frame of the bird call. Hence an ensemble average of vector length *N/2* is computed for an *N* point FFT spectrum. The ensemble average vector thus derived is called the Spectral Ensemble Average Voice Prints of that particular bird.

3.1.1 Algorithm

The algorithm for computing the SEAV is as follows:

(1) En frame the bird call signal with a frame size of 20ms and frame rate of 10ms.

(2) Compute the *N* point FFT of each frame of the windowed bird call signal x(m) as

$$X(k) = \sum_{m=0}^{N-1} x(m) e^{-j2\pi m \frac{k}{N}}$$

where k = 0,1,……, (N-1)

(3) Compute the ensemble average of the FFT spectrum across all the frames. If there are *J* frames in the bird call signal then the spectral ensemble average (SEAV) is computed as

$$X_{seav}(0) = X_1(0) + X_2(0) + X_3(0) + \ldots\ldots\ldots\ldots + X_J(0)$$
$$X_{seav}(1) = X_1(1) + X_2(1) + X_3(1) + \ldots\ldots\ldots\ldots + X_J(1)$$
$$X_{seav}(2) = X_1(2) + X_2(2) + X_3(2) + \ldots\ldots\ldots\ldots + X_J(2)$$
$$\ldots\ldots\ldots\ldots = \ldots\ldots + \ldots\ldots + \ldots\ldots + \ldots\ldots\ldots\ldots + \ldots\ldots$$
$$X_{seav}\left(\frac{N}{2}-1\right) = X_1\left(\frac{N}{2}-1\right) + X_2\left(\frac{N}{2}-1\right) + X_3\left(\frac{N}{2}-1\right) + \ldots\ldots\ldots\ldots + X_J\left(\frac{N}{2}-1\right)$$

(4) The vector $\{X_{seav}(0), X_{seav}(1), X_{seav}(2), \ldots\ldots\ldots, X_{seav}\left(\frac{N}{2}-1\right)\}$ of length *N/2* is the SEAV corresponding to each bird.

But this algorithm was originally used for the cage recording of the birds. But in our work we are using the field recording of bird calls and we are also interested in implementing that on a low end processor. So in order to do all this we made following modifications to the proposed SEAV algorithm.

3.2 Modifications over the previous SEAV approach:

Following modifications were made in the above algorithm for its implementation on a low end processor as well as for processing the field recordings of bird call.

(1) Frame size correction
(2) Dropping out zero crossing rate criterion
(3) Fixing of energy threshold
(4) Use of front end filter for cleaning the data
(5) Windowing concept

Now, we will discuss them in detail.

3.2.1 Frame Size Correction

Bird call production is much more spontaneous compare to human speech production. So it is always better to use smaller frame length for processing of calls. But as small the frame length is the processing of data will be slower. So we have to optimize the frame size in order to get faster speed of processing as well as to implement our algorithm on a low end processor.

Earlier the frame length was proposed as 20ms with an overlapping of 10ms. But this was having 160 points in each frame for a sampling frequency of 8 kHz. It was difficult to implement on a low end processor because a lot of time of processor was getting invested into dealing with this number. It is difficult to represent 160 in form of binary number.

So we have used a frame length of 32ms with and overlapping of 16ms in which each frame contains 256 numbers of points (2^8). It was very easy for a processor to deal with this number because now it can be implemented on a low end processor easily. This also results in processing speed.

3.2.2 Dropping out Zero Crossing rate criterion

Zero crossing is a point where sign of a function changes either from positive to negative or vice versa.

Zero crossing rate of a signal is defined as the rate of sign changes along a signal i.e. the rate at which the signal changes from positive to negative or negative to positive. It is defined as

$$zcr = \sum_{t=1}^{T-1} \prod \{ x(t-1)x(t) < 0 \}$$

where x denotes the signal with a period of T and

$\prod\{A\}$ is an indicator function whose value is equal to 1 if argument A is true otherwise 0.

From the analysis made, it has been observed that the number of positive frames was greatly affected by the energy threshold criterion, and negligibly by the zero-crossing rate. So this criterion can be dropped from the frame selection criteria. This would contribute to a small reduction in the computational complexity. The results of the analysis are shown in the table. For this analysis thresholds were decided by taking the average of energy and zero crossing rates of frames.

S. No.	File Name	Enthr	Zcrthr	Total no. of frames in file	No. of positive frames	% of positive frame
1.	mr1	4.2037	-	3754	137	3.65
2.	mr2	4.6887	-	15229	1305	8.57
3.	mr5	1.3149	-	5423	227	4.18
4.	mr6	0.7502	-	17960	1411	7.86
5.	mr7	0.6720	-	16419	1336	8.14

Table 1: when only energy threshold was taken into account.

S. No.	File Name	Enthr	Zcrthr	Total no. of frames in file	No. of positive frames	% of positive frame
1.	mr1	-	94.1440	3754	461	12.28
2.	mr2	-	104.24	15229	2340	15.36
3.	mr5	-	93.144	5423	1851	34.13
4.	mr6	-	85.136	17960	3477	19.36
5.	mr7	-	81.944	16419	5006	30.49

Table 2: when only zero crossing thresholds was taken into account.

S. No.	File Name	Enthr	Zcrthr	Total no. of frames in file	No. of positive frames	% of positive frame
1.	mr1	4.2037	94.1440	3754	130	3.46
2.	mr2	4.6887	104.24	15229	1184	7.77
3.	mr5	1.3149	93.144	5423	221	4.08
4.	mr6	0.7502	85.136	17960	1336	7.44
5.	mr7	0.6720	81.944	16419	1264	7.69

Table 3: when both energy and zero crossing thresholds were taken into account.

3.2.3 Fixing of Energy Threshold

Earlier, in SEAV, the energy threshold was being decided by computing the energy of some initial frames and taking the average of them. It needed more computation and thus the overall time taken for the execution was more. From our analysis, it was found that the energy of noise in the signal was having a very less order compared to the energy of bird calls. The energy of noisy frames was also in the same order. So we separated out the noisy part from the signal and computed the average energy of the frames. It was by this, we were able to fix the energy threshold value to 0.1.

3.2.4 Use of frontend filter for cleaning the data

From the spectrogram analysis of recorded bird calls, we observed that both of the species were having their calls in the frequency band above 2 kHz. In the lower band, we found various kinds of noises which were making our identification harder. So there was a need to use a front-end high-pass filter to filter out the noise present in the lower frequency range. We used the filter with the following parameters, and the response is shown in the figure.

- Order of the filter = 34
- Stopband Frequency = 1500
- Passband Frequency = 2000
- Stopband Attenuation (dB) = 60
- Passband Ripple (dB) = 1
- Sampling Frequency = 8000

These entire filter parameters were decided so that the method can be easily implemented on a low-end processor.

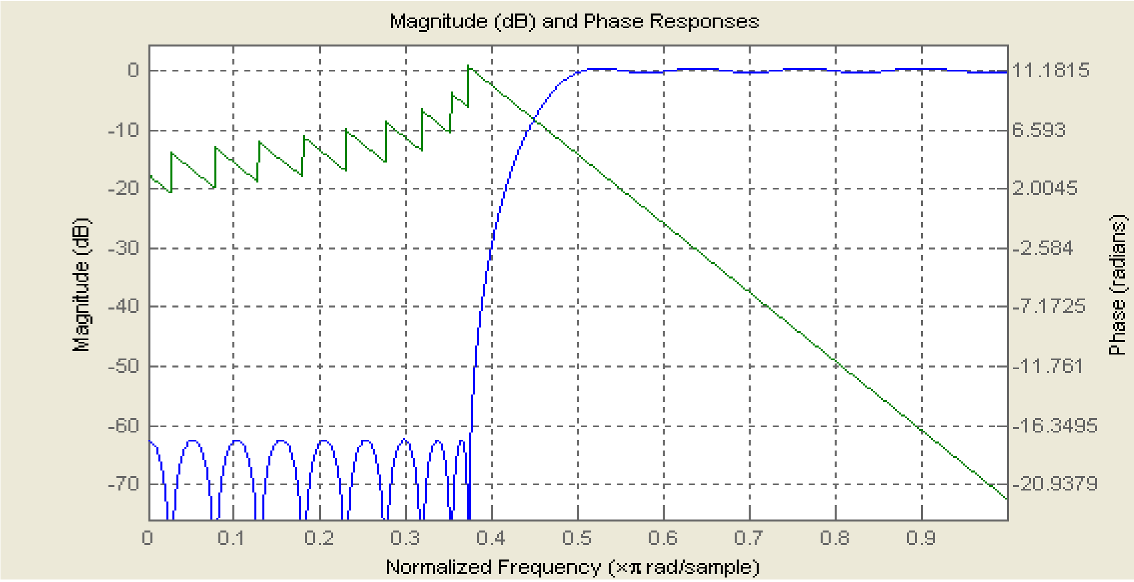

Figure 6: Response of front end filter.

3.2.5 Windowing concept

We used energy threshold for the identification of positive frames in the test signal. But it was resulting in very sharp end points which may affect the characteristics of the original call. Finally it may create a big problem in making our system unreliable for the identification of bird calls. In following figure a signal is shown in which the positive calls are identified using the energy threshold.

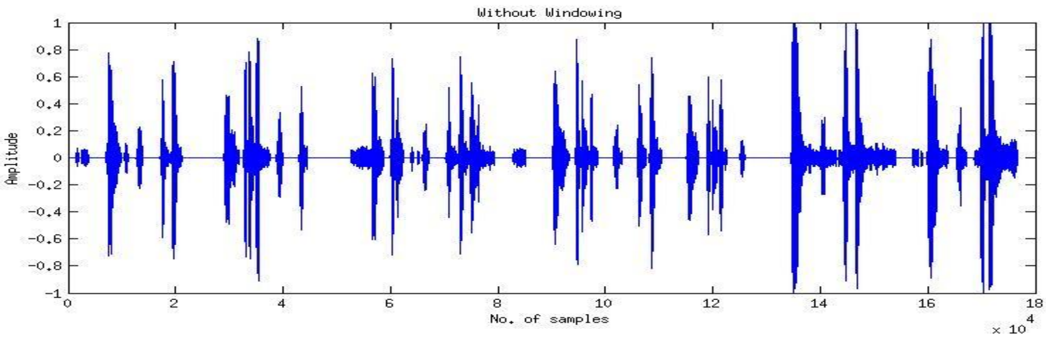

Figure 7: Bird call signal without windowing.

Each bird call is a collection of localized amplitude variations. It can be seen from the above figure that one bird call is being split into some discrete parts, owing to the less energized, but important frames lying in the middle of two peak-amplitude portions of the same bird call. So an attempt has been made to combine the discrete portions and get the true bounds of the bird call. In this context, it has been implemented that if a frame has been marked positive, few of its neighboring frames, before and after it are considered to be part of the bird call. This has been called 'windowing'.

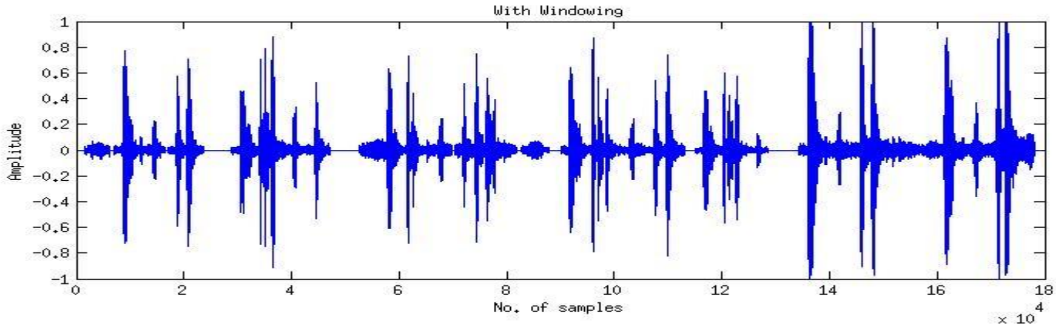

Figure 8: Bird call signal after windowing.

The above figure shows the corresponding output. By keeping in mind all the above modifications, we applied modified SEAV on the recorded data for the bird call identification, which will be discussed in the next chapter.

CHAPTER 4

Modified SEAV Implementation

4.1 Database preparation

Earlier, cage recordings of birds were used for the implementation of SEAV. Those samples differed greatly from the real-time recordings because they were free from background noises. The results obtained using the cage recorded data were also good enough during testing, which might not work precisely in real-time.

Researchers from National Institute of Advanced Studies (NIAS), Bengaluru, recorded calls for two species, viz., Magpie Robin and White Bellied Short-Wing in the forests of Western Ghats.

We used this database that was recorded in field for our analysis. Different kinds of background noises were observed in the recordings such as other bird chirps, insect sounds and instrument noises. Preprocessing is needed to suppress that background noise and enhance the target bird call before template creation. For that purpose, we have used front-end filter, as discussed earlier. The details of bird call database are shown below.

S. No.	Species	Total # recordings	Rejected (bad) recordings	# recordings used for template creation	# recordings used for testing
1.	Magpie Robin	16	3	6	13
2.	White Bellied Short-Wing	20	5	7	15

Table 4: Details of bird call database.

As a part of preprocessing we made following changes in the database.

(1) All the bird calls are recorded and sampled at the sampling rate of 8 kHz.
(2) The bird call is first windowed into a number of frames with a frame rate of 32ms and an overlap of 16ms. Frame size of 32ms with 8 kHz sampling frequency includes 256 points (2^8 points) in each frame which makes it faster for computation.

4.2 Implementation of SEAV

We have implemented modified SEAV for bird call identification in the following two steps:

(1) Source Template Generation
(2) Test template generation and classification.

Source template generation involves the extraction of structured sounds of interest from random background noise and generation of source template using SEAV followed by its storage in the system as a bird's identity.

In test template generation and classification step the test template is generated after removing background noise from the test signal using SEAV. By comparing this test template against the source templates we can classify the signal into its relevant group.

Now, we will discuss the implementation algorithm in detail.

4.2.1 Source Template Generation

As a first part of our recognition process, we first analyzed the recorded data and with the help of this we created some source templates for each species. Source template is basically the signature of a birds call. Source template is stored in system which later will be used for classification of the test template.

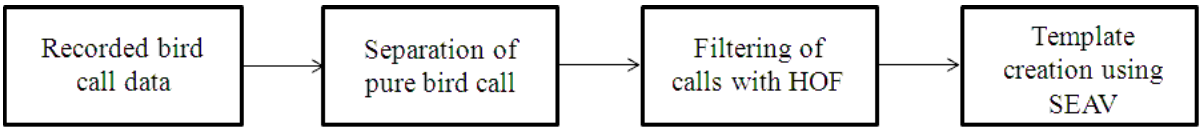

Figure 9: Block diagram of source template creation.

The source template creation was done in four steps, which we can see in the above flow chart.

STEP I:

In the first step towards the source template creation we listened all the audio files of each bird species. By observing the pattern of the calls based on spectrogram analysis we separated out the calls from main file and arranged them in the same sequence.

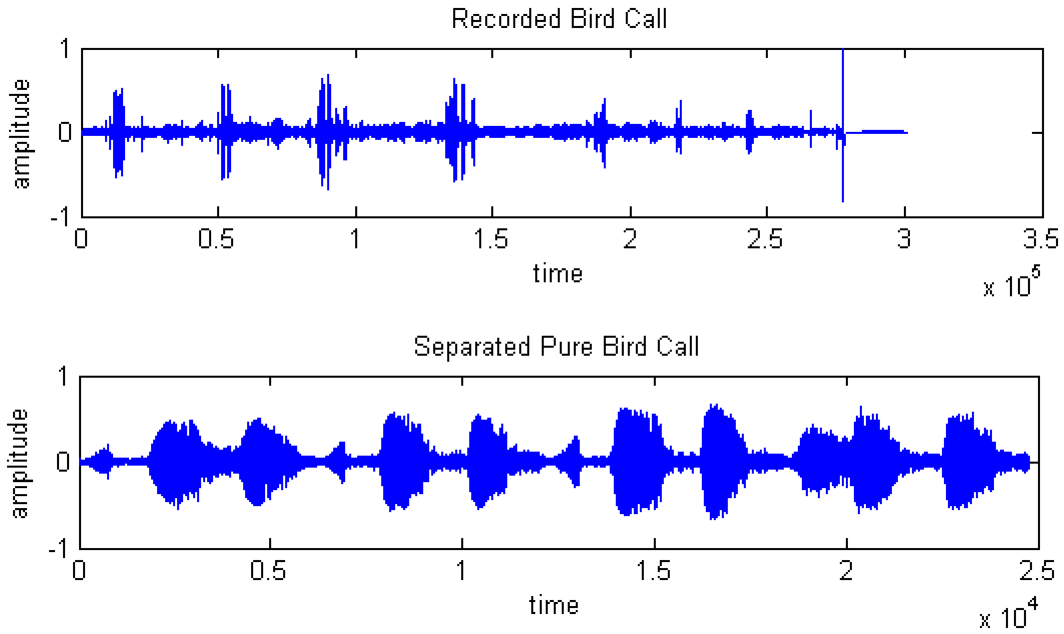

Figure 10: Recorded bird call signal (up) and its separated pure calls (down).

STEP II:

Since the recordings are made in forest, the quality of calls is not good. They have a low value of Signal to Noise Ratio (SNR) due to background noise. The background noise may be due to many reasons such as wind, chirping of other birds in the background, human sources, other animals etc.

From spectrogram analysis we observed that all the bird calls were in the frequency band of 2 kHz to 4 kHz and the lower band was containing the background noise. To remove this noise and to improve quality of calls, we created a higher order high pass filter to reject the lower band.

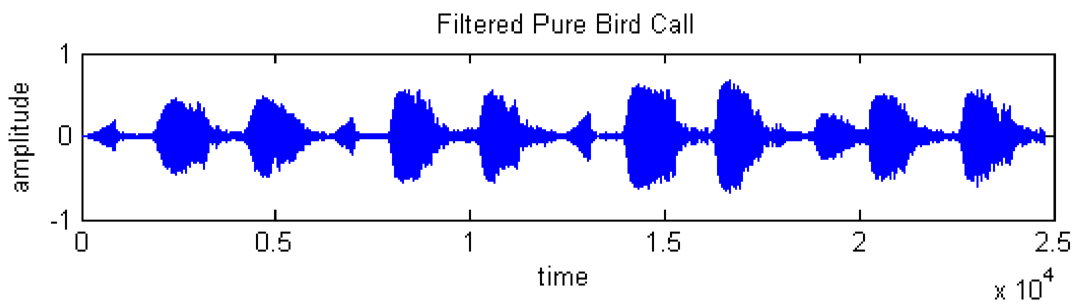

Figure 11: Filtered pure bird call signal.

STEP III:

After getting pure filtered calls from the filter we applied SEAV on that. As a result of SEAV we got a vector of length *N/2*. We stored this vector in a file as a bird's identity.

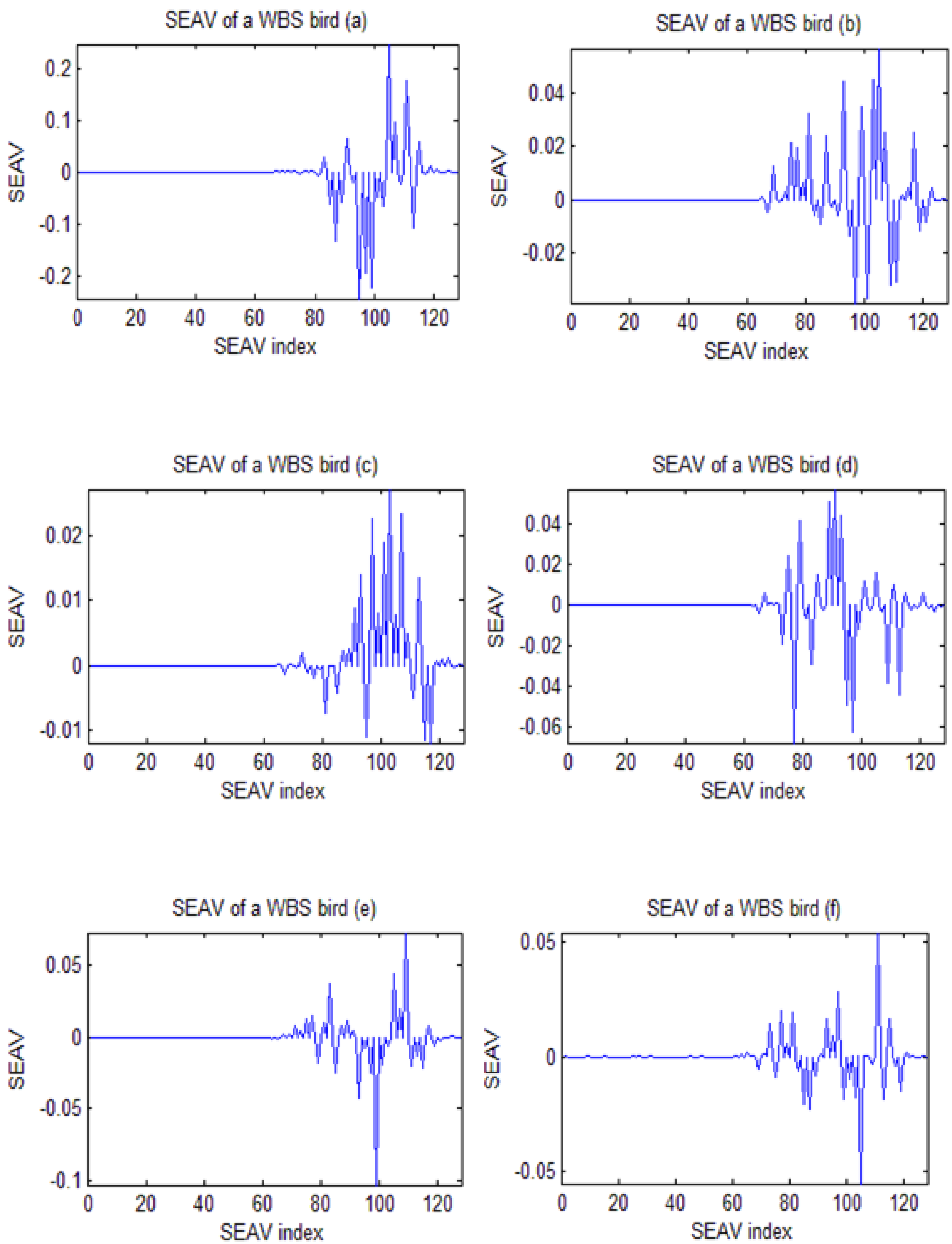

Figure 12: SEAV of WBS birds.

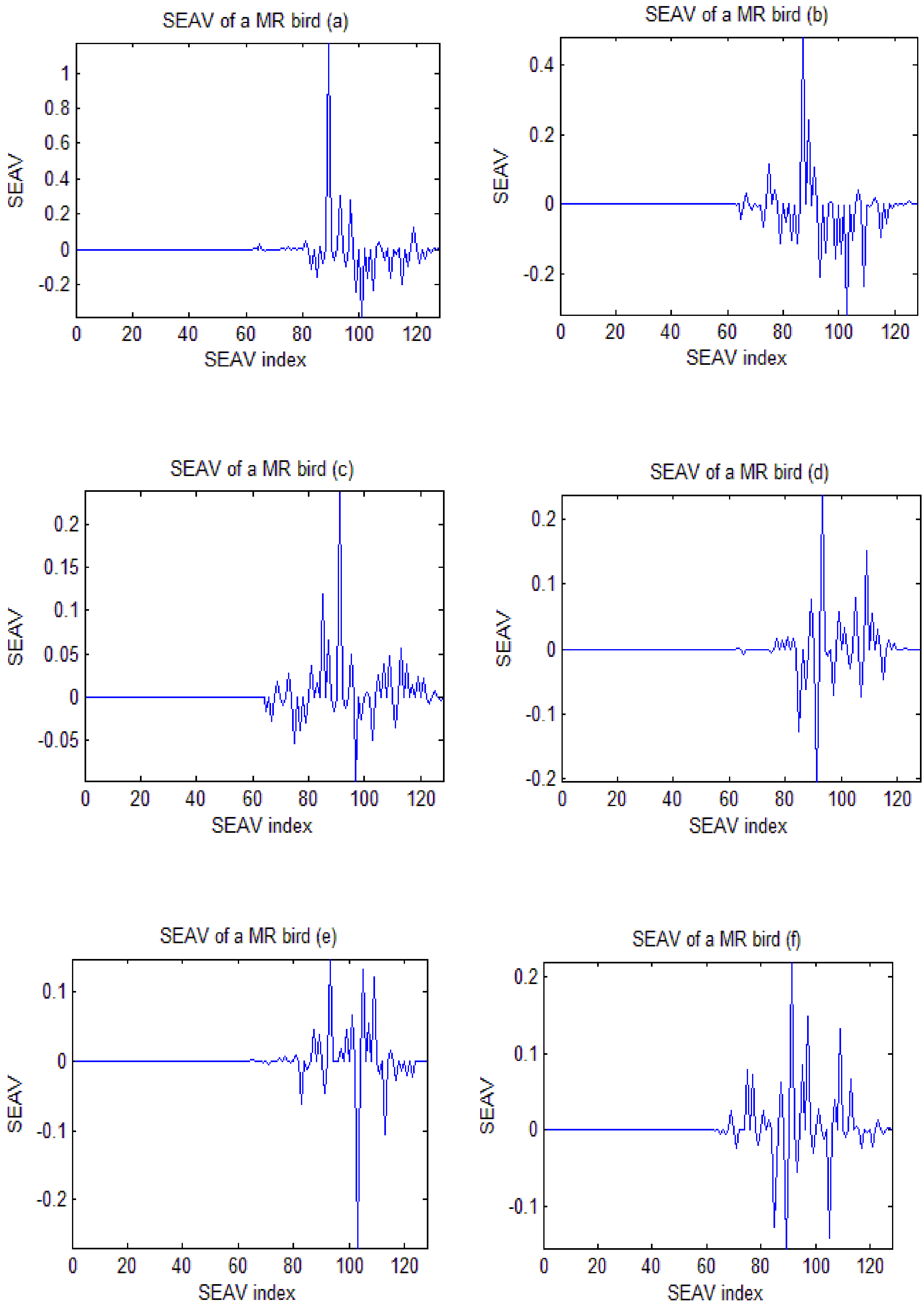

Figure 13: SEAV of MR birds.

From above analysis and the SEAV plots, we can see that calls of different birds include different spectral content, and SEAV is an efficient tool for bird call identification.

4.2.2 Test template creation and classification

The second part of the implementation involves test template creation. After the source templates were created and stored in the system, they are used for the identification of birds in real time. The block diagram of the process is shown below.

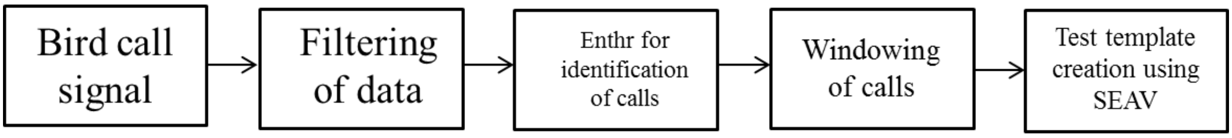

Figure 14: Block diagram of test template creation.

Test templates were created using the following steps:

STEP I:

The recorded bird call was having a lot of noise. We used a front-end filter as a preprocessing step for removing the noise. We get a filtered call as the output, which will be used for further processing.

STEP II:

By our analysis on recordings, we were able to set an energy threshold for identification of positive frames and removal of noise.

STEP III:

After getting a signal that is free from noise and which contains much of the bird call content, we use windowing before actually computing the ensemble averages. Windowing is introduced to ensure that all the spectral contents of a bird call are taken as a complete bird call. Due to energy thresholding, it might be possible that one bird call be split into several discrete regions, owing to the presence of less energized frames in the middle of a single call. So in windowing, we tried to group all such regions as follows: if a frame has been marked positive, few of its neighboring frames, before and after it are considered to be part of the bird call.

STEP IV:

After getting windowed calls after windowing from step III we applied SEAV on those calls. As a result of SEAV we got a vector of length *N/2*. This vector is called test template.

STEP V:

We do classification here. It is based on Euclidean distance between a source template and a test template. The Euclidean distance between two vectors $X = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix}$ and $Y = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix}$ is defined as

$$d = \sqrt{(X_1 - Y_1)^2 + (X_2 - Y_2)^2 + (X_3 - Y_3)^2}$$

This distance is calculated between the test template and all the source templates. The source template corresponding to the least distance is declared as the correct match to the test template and the classification is done accordingly.

CHAPTER 5

Results and Conclusion

5.1 Results

This chapter deals with the results which we got after the implementation of modified SEAV in MATLAB. All the results were for the field recorded data. The results were stored in the form of an excel sheet in which top row corresponds to the source template. These source templates were stored in our system as a bird's identity. We have created 6 templates for Magpie Robin and 7 templates for white bellied short wings.

The first column of the sheet represents the no. of calls identified in a test signal. In the respective cells the distance between corresponding call and source template is given.

The correct match is highlighted by green whereas red is used for a negative result. In the following two diagrams the two sample sheets are shown first one is of magpie robin as a test signal and second one is WBS as a test signal.

	Mr1	Mr2	Mr3	Mr4	Mr5	Mr6	WBS1	WBS2	WBS3	WBS4	WBS5	WBS6	WBS7
Call 1	2.861248	2.834142	2.740388	2.724331	2.771388	2.966973	2.741017	2.783126	2.765761	2.738143	2.741973	2.749943	2.753101
Call 2	2.164807	2.025953	1.958361	1.966829	1.922442	2.084739	1.987627	1.957797	1.968236	1.957914	1.981717	1.984569	1.966117
Call 3	4.005592	3.53129	3.632866	3.619267	3.570128	3.705674	3.61361	3.60389	3.59339	3.62128	3.602793	3.563036	3.583433
Call 4	2.966548	2.877569	2.881569	2.913323	2.927665	2.870229	2.960969	2.859866	2.90335	2.867373	2.93743	2.913617	2.91881
Call 5	3.033895	2.781598	2.780947	2.774692	2.847164	2.663748	2.694251	2.72148	2.694255	2.718652	2.664548	2.70472	2.7103
Call 6	3.108291	3.031002	3.005021	2.998495	2.948342	2.939213	2.98681	3.00568	2.981353	3.006733	3.011147	2.985901	2.968479
Call 7	2.205123	2.168206	2.160302	2.17204	2.220382	2.317626	2.247717	2.198935	2.163362	2.155179	2.148349	2.169024	2.185754
Call 8	3.801068	3.509408	3.465678	3.434666	3.584006	3.538789	3.456811	3.467255	3.456856	3.439198	3.44908	3.457337	3.464725
Call 9	2.646107	2.029353	2.140209	2.121574	2.205606	2.232096	2.222324	2.142052	2.127831	2.1388	2.125634	2.121462	2.122313
Call 10	2.367334	1.884324	1.89256	1.910611	1.909681	2.03509	1.931546	1.895012	1.872828	1.866163	1.891173	1.879293	1.887476
Call 11	3.840291	3.474941	3.374697	3.360118	3.491187	3.449008	3.438849	3.361532	3.398288	3.39057	3.442674	3.434905	3.409153
Call 12	2.1938	1.989581	1.973911	1.881635	1.991003	1.857871	2.209541	1.924886	1.90206	1.93507	1.937545	1.913994	1.948478
Call 13	1.990903	1.713346	1.650111	1.708217	1.645	1.659039	1.667844	1.644014	1.645275	1.668549	1.649247	1.68433	1.666677
Call 14	2.627794	2.328296	2.215847	2.161894	2.410688	2.283057	2.272408	2.197343	2.231627	2.246562	2.263964	2.249294	2.199736
Call 15	2.444577	2.259263	2.232071	2.181626	2.164986	2.155485	2.260593	2.180316	2.175659	2.173439	2.184079	2.209724	2.225578
Call 16	1.789658	1.617712	1.536977	1.501197	1.553349	1.419753	1.47923	1.463042	1.459974	1.46748	1.483139	1.48678	1.475633
Call 17	2.422897	2.085765	2.196006	2.271384	2.154093	2.409547	2.250439	2.243448	2.208403	2.171073	2.182094	2.182045	2.205638

Table 5: Distances between source templates (first row) and test templates of the identified calls (first column) of Magpie Robin

	MR1	MR2	MR3	MR4	MR5	MR6	WBS1	WBS2	WBS3	WBS4	WBS5	WBS6	WBS7
Call 1	2.094105	1.320157	0.854603	1.007587	0.896419	0.882479	0.960615	0.741228	0.736418	0.736097	0.800764	0.779048	0.709602
Call 2	2.133603	1.388485	1.048695	1.099066	1.237163	1.039208	1.199276	0.947037	0.985676	0.995668	0.999604	0.953692	0.983261
Call 3	1.977884	1.236745	0.992436	1.054948	0.977661	1.0527	1.162451	0.804409	0.827043	0.80718	0.826174	0.83108	0.849163
Call 4	1.975456	1.316187	0.742505	0.900695	0.68237	0.712837	0.952133	0.744238	0.694375	0.712891	0.683222	0.712617	0.744226
Call 5	2.129425	1.531022	1.082157	1.164327	1.073088	1.040214	0.94841	0.962257	0.910236	0.955935	0.887765	0.936567	0.872521
Call 6	1.84092	1.119121	0.613649	0.719332	0.673664	0.73256	0.763596	0.524072	0.462979	0.465147	0.464399	0.452338	0.479342
Call 7	2.146781	1.170939	0.576565	0.910301	0.900941	0.720139	0.921368	0.654446	0.571144	0.630728	0.574928	0.586271	0.580513
Call 8	2.280508	1.569674	1.226574	1.161696	1.066928	1.228746	1.264419	1.065939	1.048018	1.077315	1.156127	1.086323	1.063546
Call 9	2.247587	1.355658	0.826476	1.254288	1.08157	0.920758	1.099135	0.983727	0.954818	0.938938	0.985646	0.959132	0.941556
Call 10	2.076031	1.462758	0.891052	1.000076	0.97903	1.027073	1.113513	0.922536	0.899732	0.902181	0.862653	0.882398	0.970569
Call 11	1.855903	1.127384	0.524307	0.598748	0.569126	0.612381	0.65339	0.337908	0.242292	0.323898	0.322504	0.293643	0.258004
Call 12	2.138912	1.469556	0.957139	1.015173	1.025051	1.00245	1.158771	1.001249	0.952249	0.967268	0.927544	0.980417	0.917621
Call 13	2.625616	1.761918	1.578676	1.769201	1.760232	1.706977	1.675797	1.669567	1.643414	1.651848	1.585896	1.671081	1.665994
Call 14	1.845994	1.20333	0.559969	0.586909	0.610073	0.686513	0.676737	0.344234	0.333493	0.334978	0.374774	0.3532	0.321265
Call 15	1.774448	1.178301	0.709493	0.712633	0.738276	0.760207	0.762453	0.532918	0.450205	0.49684	0.501521	0.485027	0.473253
Call 16	2.043436	1.341981	0.775096	0.825414	0.712471	0.802562	0.855354	0.670358	0.658349	0.630628	0.648772	0.627217	0.689867
Call 17	1.8047	1.119393	0.471704	0.553901	0.503389	0.551989	0.662	0.276752	0.17577	0.233018	0.22847	0.238926	0.27013
Call 18	1.893007	1.267107	0.754168	0.725758	0.658545	0.797329	0.85008	0.547008	0.567078	0.578767	0.597793	0.52967	0.566746
Call 19	1.783962	1.283224	0.716834	0.68179	0.753659	0.818113	0.865749	0.574384	0.495857	0.562894	0.559986	0.476654	0.552445
Call 20	1.877351	1.113352	0.463331	0.601325	0.556173	0.57866	0.678107	0.29025	0.187856	0.269082	0.242717	0.253314	0.261209
Call 21	1.866386	1.120983	0.542685	0.612747	0.54842	0.566154	0.615281	0.384587	0.311549	0.351191	0.355881	0.356035	0.36384

Table 6: Distances between source templates (first row) and test templates of the identified calls (first column) of Wide Bellied Short-Wing.

5.2 Accuracy of SEAV

Accuracy of SEAV is defined as:

$$\% \; Accuracy = \frac{No.of\; positive\;\; calls\;\; identified}{Total\;\; no.of\;\; calls\;\; identified} \times 100$$

By the formula above, we have calculated the accuracy given by our algorithm in the detection of both the test species of birds, and the corresponding results are shown in the bar graph below.

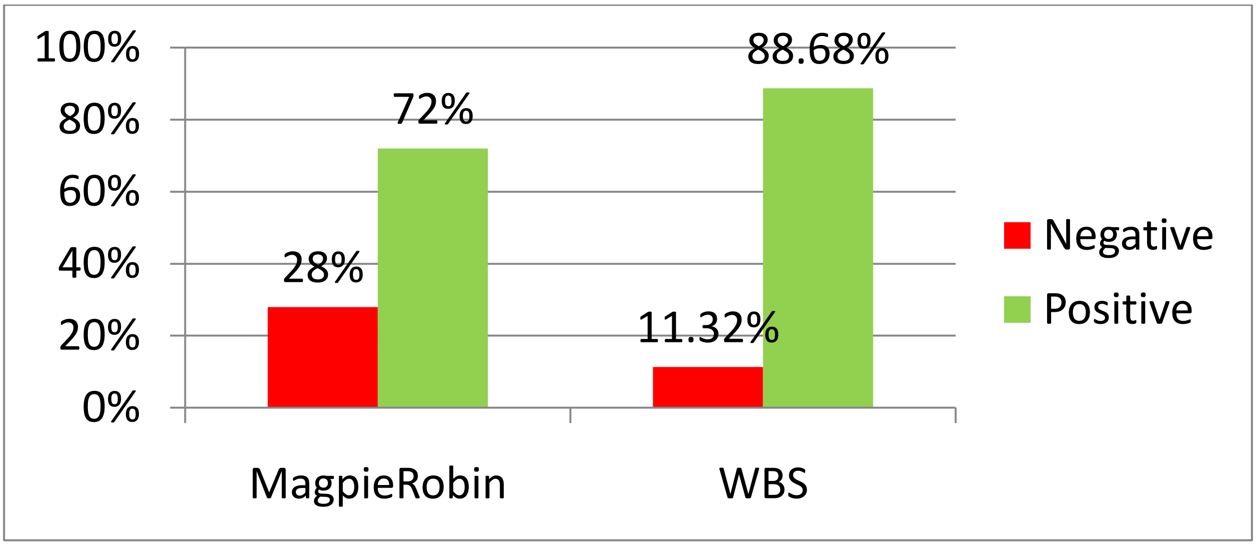

Figure 15: SEAV results for various species.

We can see that modified SEAV worked very significantly on real time field recorded data. It showed 72% positive result for Magpie Robin and 88.68% for WBS. Thus the overall efficiency of the system is around 80%, which is an encouraging result, and shows the potential of modified SEAV over the conventional speech recognition methods.

CHAPTER 6

Future Scope

6.1 Future scope

This section deals with the future scope of the presented work. A number of possible extensions to the work are mentioned here.

1. Larger database
2. Hardware implementation
3. Development of an iPhone/Android app.
4. Development of bird call identifier software
5. Implementation to insects and other animals

6.1.1 Larger Database

In this work, identification has been performed on two species of birds. It can be further extended to identify more number of birds. When extended, it should be kept in mind that the recognition of each individual involves more time, fixing of energy threshold will need more attention, and involves more manual effort in selecting some best calls of each bird for template creation. It would be of more use when extended to a larger database.

6.2.2 Hardware Implementation

The algorithm has been written in such a manner that it can be easily implemented on hardware. A portable light-weight device which can be used for this purpose essentially needs the following components in it:

i. ***Microphone:*** To capture the bird call data in the field.

ii. ***Microcontroller:*** To digitalize the signal and analyze it. This depends on the capability of the processor being used. If the processor is powerful enough, the complete analysis can be carried out at the microprocessor level itself, and the report of the observed birds can be prepared. If it is not, then only the noise removal and preparation of useful bird data for further analysis can be one at this level.

iii. ***Wireless Transmitter:*** To transmit the prepared report/data to a centrally monitored system.

The block diagram of such a device is shown in the figure below.

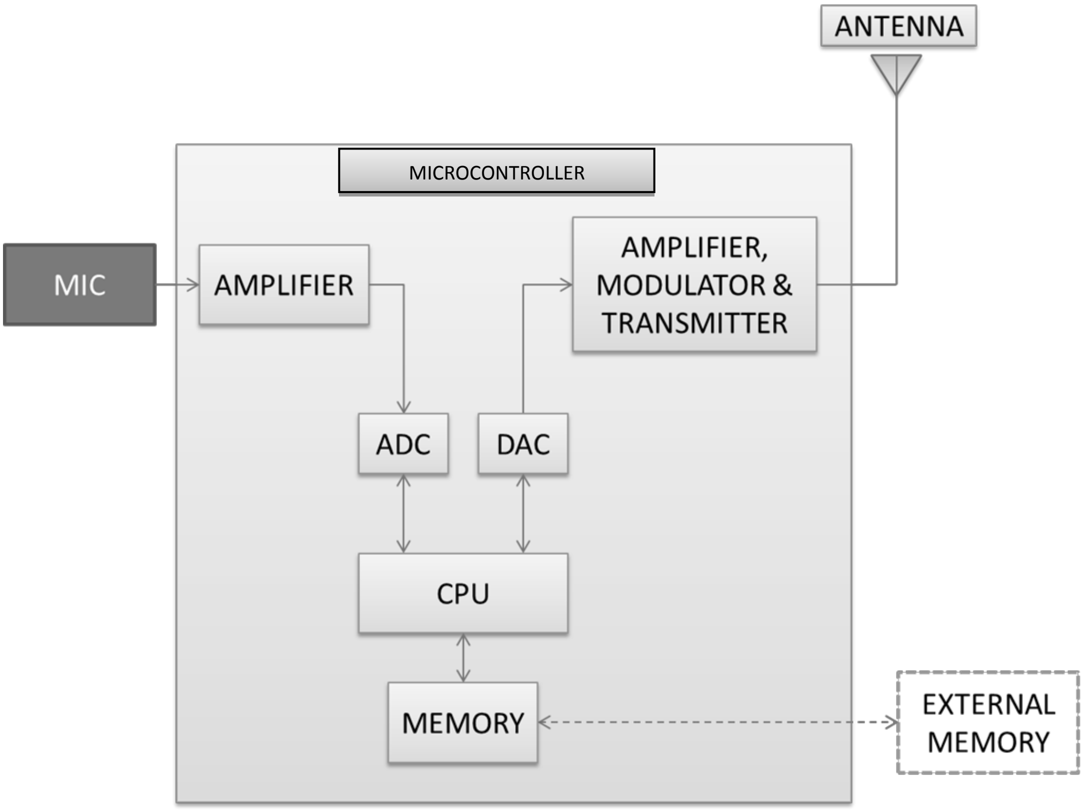

Figure 16: Block diagram showing hardware implementation.

6.1.3 Development of iPhone/Android app.

Usually mobile phones have the above mentioned components of hardware integrated together, so it is easier to develop an iPhone/Android app that implements the whole algorithm. The data can be sent through GSM/GPRS to a centrally monitored system.

6.1.4 Development of Bird Call Identifier software

This is a very useful application because future developments to the code can be easily updated when using over a computer. Software can be developed using this algorithm and which will be extremely helpful to ornithologist for their research.

6.1.5 Implementation to insects and other animals

In farm fields, a major problem that would reduce the output of crops is pests. If we could recognize pests by the sound produced by them using a recognition device, we could use the same device for emitting a repellant sound consisting of a particular frequency corresponding to the type of pest, to keep it away.

CHAPTER 7

BIODIVERSITY IMPLICATIONS

7. 1. Private Forests:
Private tea estate owners were visited to discuss the value of the proposed project. We received considerable support and enthusiasm from the owners visited and they permitted and facilitated us to conduct field visits and conduct recordings using purchased automated recorders. These recordings have bird song data from the private forests. Since our algorithms could detect only seven bird species, we could not provide a list of the species at these forests. However, this will be possible once the ongoing software development (with new funding) is complete. We had attempted to manually identify birds from these recordings, but we had to give up this effort since it took over four months to get through just parts of recordings from one estate.

7.2. Presence of endemic, threatened bird species outside protected areas:
Towards understanding this, data has been collected, but with only partial success of our species identification algorithms, a final list of birds has not been produced. However, since the data exists and algorithm development is ongoing (with funding from other agencies described in the main body of the report), we will be able to provide this information at a later date.

7. 3. Potential for use for threatened, endemic mammals and tree frogs:
Towards this outcome data has been collected, but with only partial success of our species identification algorithms, a final list of species has not been produced. However, since the data exists and algorithm development is ongoing (with funding from other agencies described in the main body of the report), we will be able to provide this information at a later date. However, both tree frog and mammal researchers were consulted during trial deployment of the recorders. Our current collaboration as a continuation of the project funded by CEPF-ATREE includes researchers who are planning to use this system to monitor various other taxa including monitoring of human-elephant conflict.

7. 4. Open access for biodiversity database:
A preliminary website has been developed and is at www.birdsongs.in , a server for the project has been purchased (with other funding) and is at the NCBS server room. After further development of the species recognition algorithm, the website will be hosted through NCBS server at www.bioacoustics.in. We are also discussing modalities of making our data also available through www.indiabiodiversity.org. At present the data has not been shared due to the large file sizes. For example data from just one season's recording is about one TB while more meaningful splits of the dataset, without background noise but with animal sounds, with smaller file sizes generated through an algorithm, will considerably reduce the logistical constraints. Such an algorithm is being developed (with other funding) and the data will be made available after this is completed.