BIRD SONG IDENTIFICATION BY SPECTROGRAM CORRELATION

A Project Report

submitted by

YANGALA LAKSHMIPATHI

in partial fulfilment of the requirements
for the award of the degree of

MASTER OF TECHNOLOGY

under the guidance of
Dr. Anil Prabhakar

DEPARTMENT OF ELECTRICAL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY, MADRAS.

May 2015

THESIS CERTIFICATE

This is to certify that the thesis entitled **BIRD SONG IDENTIFICATION BY SPECTROGRAM CORRELATION**, submitted by **YANGALA LAKSHMIPATHI**, to the Indian Institute of Technology, Madras, for the award of the degree of **Master of Technology**, is a bona fide record of the research work carried out by him under my supervision. The contents of this thesis, in full or in parts, have not been submitted to any other Institute or University for the award of any degree or diploma.

Dr. Anil Prabhakar
Research Guide
Professor
Dept. of Electrical Engineering
IIT-Madras, 600 036

Place: Chennai

Date:28-04-2015

ACKNOWLEDGEMENTS

I take this opportunity to express my deepest gratitude to my project guide Dr. Anil Prabhakar for his valuable guidance and motivation throughout the project.I am very grateful to him for providing his valuable time to guide me during the project.

I would like to express my sincere gratitude and thanks to Mr. Robin Vijayan and his team, who have provided us with lots of audio recordings of a landscape required for the analysis.

It is a privilege to be a student in IIT Madras. I express special thanks to all my teachers for all the academic insight obtained from them. I also acknowledge the excellent facilities provided by the institute to the students.

I would like to thank all my lab members with whom I had various fruitful discussions. Special thanks to all my friends for making my stay in IIT Madras an enjoyable one.

I am indebted to my parents for their unconditional love, support and guidance. I dedicate this thesis to them.

ABSTRACT

Bird songs play a vital role in the communication between individuals and species. A Bird may listen to other birds and classify them as con-specific, neighbour or stranger, mate or non-mate, kin or non-kin. It may also sing to other birds for mate attraction, or territory defence.

Ecological and behavioural studies can benefit a lot from the study of bird calls. Ornithologists, biologists, computer scientists and electrical engineers would like to enhance their knowledge of bird songs and their behaviour. The study of bird songs involves many different disciplines that converge amongst recording birds provides a permanent record of the calls and allows comparison over time and between species and individuals. Bird song recognition approach is used to provide automatic investigation and remote monitoring of bird species population, which can provide the relevant agencies with sound information to habitat conservation as well as rare/endangered species survival plans and actions.

Generally Bird song recordings are huge and these recordings contain a lot of inactive period and also background noise. The structures of the songs of birds are usually so complex and variable that the human ear is unable to perceive them accurately. Manually going through all the recordings and identifying the particular bird song is a very tedious task and erroneous. This thesis is an attempt at proposing an algorithm to automatically identify the bird call from audio recordings.

KEYWORDS: Bird song; Inactive period; Spectrogram; Correlation coefficient.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ALGORITHMS

CHAPTER 1

Introduction

The monitoring of bird behaviour and diversity over a variety of spacial and temporal scales poses many challenges to human observers. A significant amount of knowledge on bird diversity and their interpreted behaviour is the result of field observations made by expert ornithologists. Bird species identification and the study of their interactions have been developed by means of the visual and acoustic abilities of these experts. Over the years, much of the information gathered from bird songs has been analyzed in order for the experts to effectively classify different species through their sounds. This has been done in a great manner through manually with minimal aid from computer algorithms. In the last decade, great advances have been developed both in the fields of computer science as well as in electronics, which can be applied in order to aid and automatize much of the work performed in the field. These kinds of limitations have caused our knowledge about some bird species with complex societies to be very scarce.

Collecting data from a land scape in outdoor environment and analysing them has been a tedious task that some field workers have little enthusiasm for, considering it difficult, dull and old fashioned work. However technology has enabled scientists to accomplish this task in a better way. Many sounds archives consisting of a huge number of bird sound recordings makes it difficult for the ornithologists to study the whole data. Birds are numerous and sensitive to environmental changes. Birds are a good indicator for assessing habitat changes because they are distributed over a wide range of areas and a significant amount of knowledge on bird diversity and behaviour was discovered through field observations by experienced birdwatchers and expert ornithologists.

Sound recordings of a land scape are huge, which makes it difficult to transfer this data to a remote location. This calls for the processing of the data in the field and transferring the necessary information only. A particular land scape sound recording contains more than one bird and it also consists of lots of inactive period(when no bird is making call). This data can be crunched by automatically identifying the bird songs from the recorded data of a landscape. In general bird song not only corresponds to short-time timbrel characteristics but also to temporal structure of the sound signal i.e., the time evolution of bird sounds will provide some discriminating information for bird's song recognition. Therefore it would be valuable to apply image analysis methods to identification of bird species based on spectrogram images.

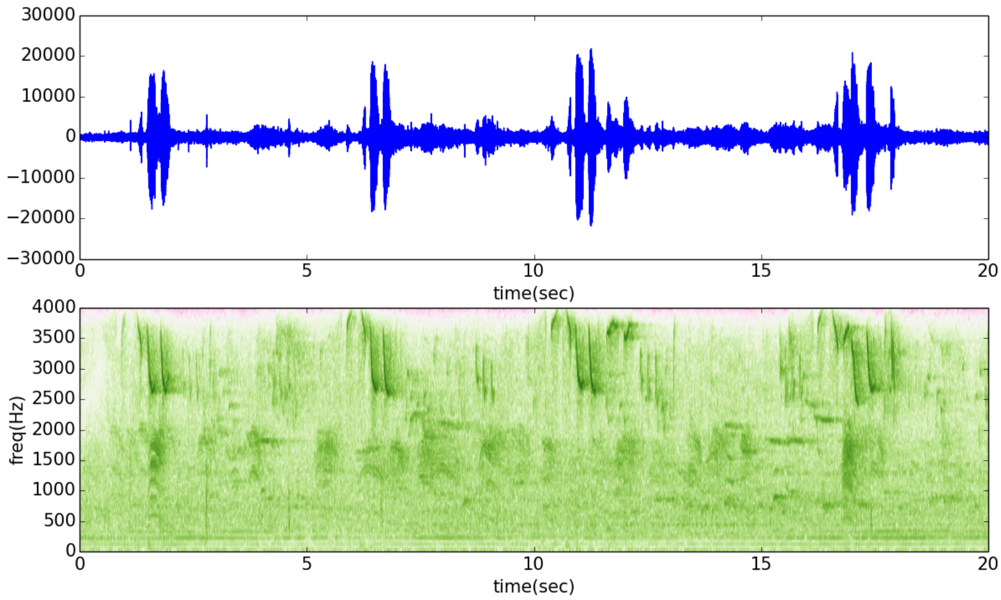

Figure 1.1: A typical 20sec sound recording of a land scape and it's spectrogram.

Traditionally visual inspection of the sound spectrogram or sonograms was one of the primary means for analysing bird song, typically relies on the subjective judgements of the experts. This process is extremely laborious, time consuming, and not entirely objective. Therefore it was impractical for large scale and long-term study of the population trends of different bird species. Automatic identification of bird song from a landscape audio recording emulates this process of manual identification.

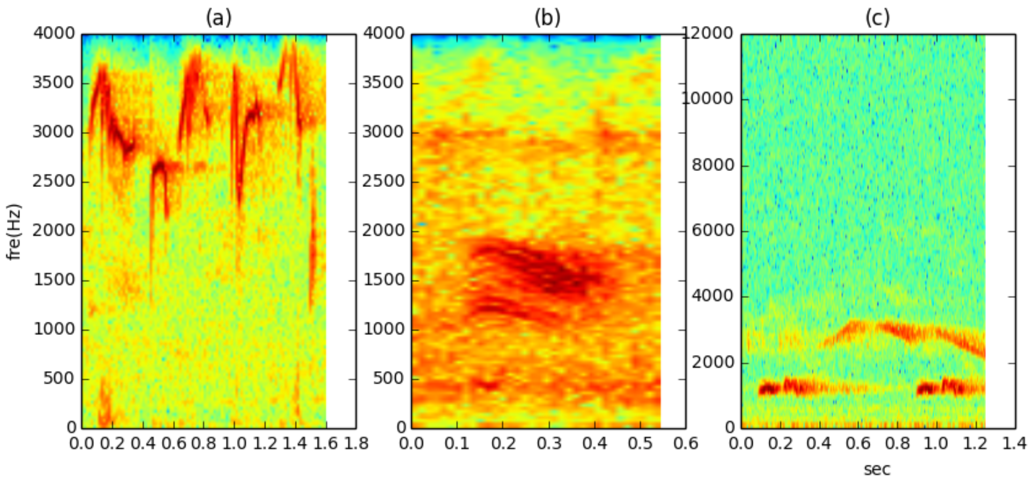

Figure 1.2: Different shapes of the spectrogram of different bird species. (a) Magpie robin (b) Crow (c) Unknown.

1.1 Organization of Thesis

Chapter 2 deals with the theory of spectrogram, correlation and changing the sampling rate of given audio signal.

Chapter 3 explains the Algorithm used in order to recognise the bird song. It also explains the process of extracting some of the variables required.

Chapter 4 discusses about the results obtained from the proposed algorithm and also presents the observations made on the results.

Chapter 5 concludes the thesis.

CHAPTER 2

Background

The hearing range of birds is from below 50 Hz (infrasound) to around 12 kHz, with maximum sensitivity between 1 kHz and 5 kHz. The only bird known to make use of infrasound (at about 20 Hz) is the western capercaillie. The average frequency of the songs of the songbirds is about 4,000 Hz. Many warbles, sparrows, waxwings, kinglets, and a number of other birds produce sounds that reach 8,000 Hz and beyond.

2.1 Niche hypothesis and Biophony

The 'niche hypothesis', an early version of the term, biophony(consists of Greek prefix and suffix meaning life sound), describes the acoustic bandwidth partitioning process that occurs in still-wild biomes by which non-human organisms adjust their vocalizations by frequency and time-shifting to compensate for vocal territory occupied by other vocal creatures. Thus each species evolves to establish and maintain its own acoustic bandwidth so that its voice is not masked. For instance, notable examples of clear partitioning and species discrimination can be found in the spectrograms derived from the biophonic recordings made in most un-compromised tropical and subtropical rain forests.

The study of biophony focuses on the collective impact of all sounds emanating from natural biological origins in a given habitat. The realm of study is focused on the intricate relationships;competitive and/or cooperative, generally between non-human biological sound sources taking into account seasonal variability, weather, and time of day or night, and climate change. It explores new definitions of animal territory as defined by biophony, and addresses changes in density, diversity, and richness of animal populations.

2.2 Spectrogram

A spectrum is always made over a finite time interval, which may be as long as the full length of the signal or may be only a short part of it. Therefore, an individual spectrum provides no information about temporal changes in frequency composition during the interval over which it is made. If this spectrum is made over a very short time interval (e.g.: few milliseconds), it shows an instantaneous frequency pattern, but again we can't get an idea of time evolution of our waveform. To see how frequency composition changes over time we need to examine a sound spectrogram, which is the representation of more spectra, computed on consecutive or overlapping segments of the signal. A spectrogram shows the evolution in time of sound frequency structure.

Spectrograms are graphical representations of audio files, with time on the horizontal axis and frequency(going from low frequency at the bottom of the image and high at the top) on the vertical axis. For each of the audio files, we divided the sound of the signal into frames and compute the spectrum for each of the frames. A spectrum represents the intensity of the signal as a function of frequency. The spectrogram is then created as a graph of the spectra of each frame in the sound. IN our case to create the spectra we divided the signals into frames with frames-size as 256 samples, with each frame starting a frame-step of 128 samples after the previous frame, we have 50 percent overlap between frames.

2.3 Correlation Coefficient

In statistics, the correlation coefficient also referred to as Pearson product-moment correlation coefficient is a measure of the linear correlation(dependence) between two variables X and Y, giving a value between +1 and -1 inclusive, where 1 is total positive correlation, 0 is no correlation, and -1 is total negative correlation. It is widely used in the sciences as a measure of the degree of linear dependence between two variables.

The formula for correlation coefficient can be expressed as

$$r = \frac{\sum[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y} \tag{2.1}$$

Where μ and σ represent the mean and variance respectively.

Spectrogram of the audio signal is a 2-dimensional matrix. The correlation coefficient between two spectrogram images is calculated using the following formula.

$$r = \frac{\sum_m \sum_n (A_{mn} - \mu_A)(B_{mn} - \mu_B)}{\sqrt{\sum_m \sum_n (A_{mn} - \mu_A)^2 \sum_m \sum_n (B_{mn} - \mu_B)^2}} \tag{2.2}$$

Here μ_A and μ_B represent the mean of matrices A and B respectively.

2.4 Sampling rate conversion

The sound recording archives of different landscapes might be recorded at different sampling rates. It is important to change the sampling rate of the input audio signal to that of reference signal. Decimation, or down-sampling, reduces the sampling rate, whereas expansion, or up-sampling, followed by interpolation increase the sampling rate.

2.4.1 Decimation

Decimation, consists of reducing the sampling rate by a factor M, here the output is defined as

$$y(m) = x(mM) \tag{2.3}$$

i.e., it consists of every Mth element of the input signal. It is clear that the decimated signal y does not in general contain all information about the original signal x. Therefore, decimation is usually applied in filter banks and preceded by filters which extract the relevant frequency bands.

2.4.2 Expansion

Expansion, consists of increasing the sampling rate by a factor L. Here, output is obtained by inserting $L - 1$ zeros between successive vales of the input $x(n)$, i.e.

$$y(m) = x(m/L), for\ m = (0, L, 2L, ...);\ 0, otherwise \tag{2.4}$$

The expansion operation is followed by interpolation leas to a representation of the signal x at a sampling rate increased by the factor L.

2.4.3 Sampling rate conversion by a non-integer factors

Sampling rate conversion by a non-integer factor F may be achieved by the use of expansion and decimation operations. If the conversion factor can be expressed as a rational number, i.e., the ratio of two integers, $F = L/M$, then the obvious ways to achieve the conversion is to apply expansion by the factor L followed by low-pass filtering and decimation with the factor M. A problem with this direct approach occurs when the integers L and M are large, in which case there will be high requirements on the anti-aliasing filters. In practice this problem is avoided in multirate signal processing be performing the sampling rate conversions in several stages with small integer factors at each stage.

CHAPTER 3

Proposed Bird Song Recognition Algorithm

Generally most birds are active during morning and evening hours, on average 5 hours per day. For the general values of $sampling\_rate = 24KHz, sample\_width = 16bits$ and stereo recording, the recorded data for 5 hours duration is around 1.6 GB, which makes it difficult to transfer to a remote location or even to analyse the manually collected data from the field. Each landscape will have limited variety of bird species. So, if we are able to identify particular bird song from the sound recording then we could extract all the Bird songs. This extracted output is expected to be lot less when compared to the total recorded data, as the birds won't be singing all the time during that 5 hours period.

3.1 Average bird song length

Birds produce a variety of sounds to communicate with flock members, mates, neighbours, and family. These sounds vary from short, simple call notes to surprisingly long, complex songs. The duration of these bird songs vary from few milli seconds to 4-5seconds. Superb Lyre bird has one of the longest and complex songs. Some birds also generate sounds by using substrates or special feathers or special wings.

We have taken 8 instances of a particular bird (reference bird) song to calculate the average bird song duration(length). Average song duration of Magpie Robin is 1.625 sec and crow has 0.32 sec. In the later sections, average bird song length is referred as Segment.

3.2 Reference Spectrogram

We have already seen that the spectrogram of an audio file is a 2-dimensional graph with time on the horizontal axis and frequency on the vertical axis. This can be represented using a 2-dimensional matrix, each element in the matrix representing the intensity of the signal at a particular frequency and time. Here also, we have considered 8 instances of a particular bird song to calculate the reference spectrogram. let $A_1, A_2, A_3, ...$ are the spectrograms of the individual songs of a bird, then reference spectrogram is

$$A_{ref} = \sum_n [A_i]; \quad n = 8 \; in \; this \; case. \tag{3.1}$$

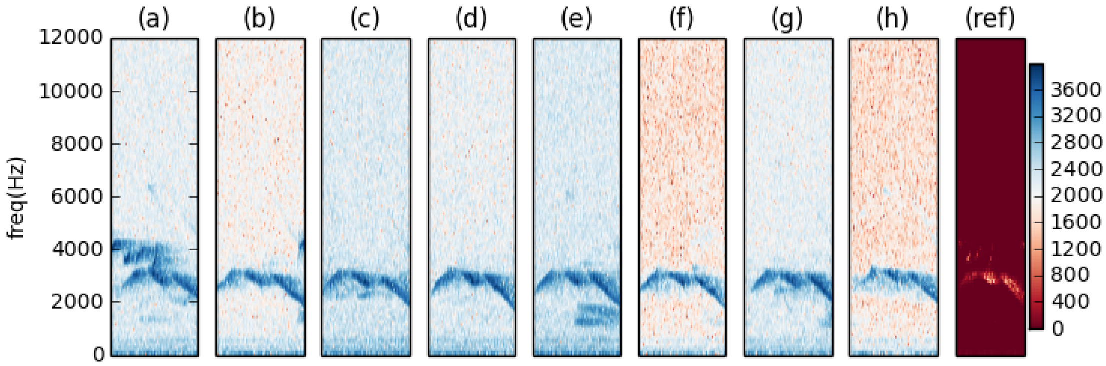

Figure 3.1: First 8 subplots are the individual spectrograms of an unknown bird song. (ref) is the reference spectrogram.

In figure 3.1, we can see that all the 8 individual spectrograms are in one color band and the reference is in other color band. This is because, while averaging mean of the intensity matrix is constant but variance changes a lot. Color map at the end represents the intensity values of each color. Clear observation is that first 8 subplots are in white - blue band and the reference spectrogram is in brown - white band. We can also notice that the noise gets cancelled out in the reference spectrogram.

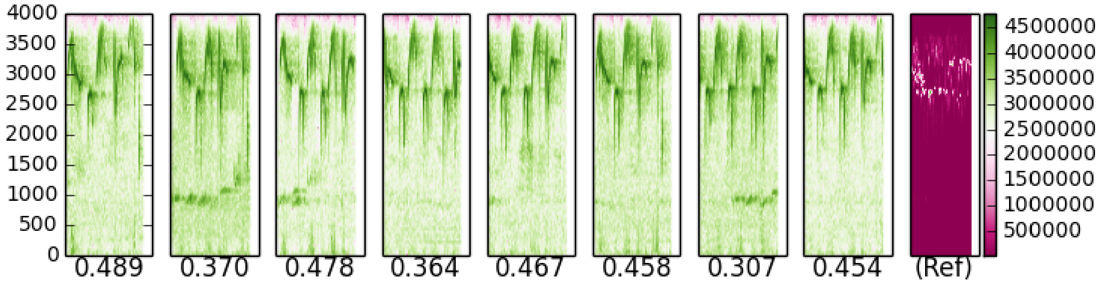

Figure 3.2: Reference spectrogram(Ref) of Magpie Robin song and it's individual spectrograms.

Most of the analysis presented in this paper is for Magpie Robin bird. Figure 3.2 is the reference spectrogram of the Magpie Robin bird song. The values present under the each subplot represent the correlation coefficient of that spectrogram with the reference spectrogram. Average duration of MR bird song is 1.625sec. Sampling rate in this case is only 8 kHz.

3.3 Method

Our main goal is to identify the particular bird song from a sound recording of a landscape. The idea behind the algorithm is that, if the correlation coefficient between the spectrogram of the segment from the recording and the reference spectrogram is higher, then the probability of it being the reference bird song is high.

Generally Bird song recordings are saved in Wave file format so there will not be any loss of information. These sound archives of the birds are collected from different places and different song meters. The differences in the sampling rate, gain and number of channels is certain. If the sampling rate of the input signal is different from the sampling rate of the reference, then either up sample or down sample the signal to match the sampling rate of the reference using Sampling rate conversion discussed in section 2.4.

Input is a continuous stream of an audio file. we read the first segment (average 'reference bird' song length) data and calculate the correlation coefficient with the reference spectrogram, then we move to the next segment of data and do the same. The problem with this is, the entire bird song won't be present in the single segment. So we can't move to next segment, but we can move by $\frac{segment}{f}$ fraction and consider the segment amount of data from that point. Previously we had one correlation coefficient for each segment, so it was easy to make decision either to retain the segment or to throw it away. In the later case, correlation coefficient's value increases while the segment approaching the 'correct song' and then decreases. The following figure 3.3 illustrates the same.

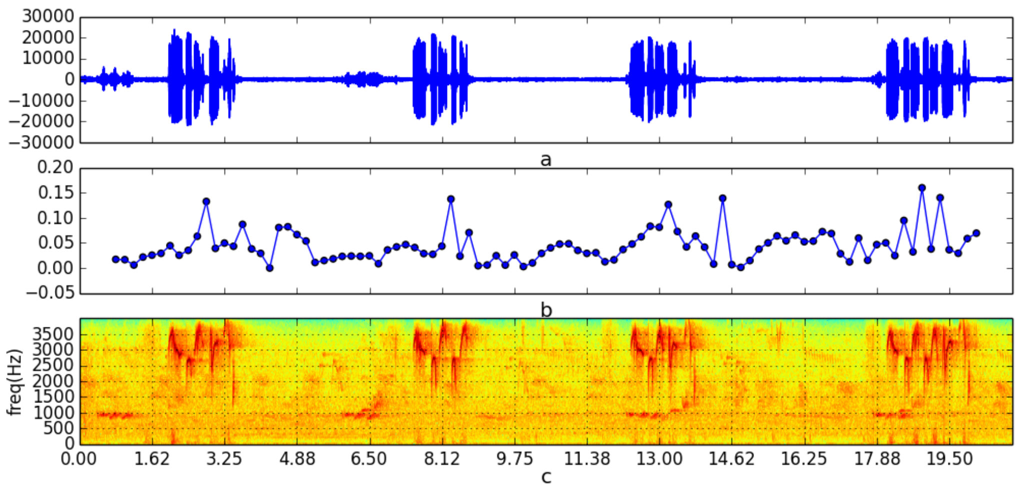

Figure 3.3: Magpie Robin bird call and It's analysis. (a) Input audio signal (b) correlation coefficients plot while moving the segment by segment/8 (c) spectrogram of the input signal.

The reference spectrogram for Magpie Robin bird is shown in figure 3.2. Different color maps are used to plot the spectrogram in figure 3.2 and figure 3.3. Now, we have to identify the segment that needs to be retained. If the correlation coefficient is greater than some threshold, mark that point as P_1, then look for next Q amount (preferably $1\ segment$) of data, mark the last point with correlation coefficient greater than threshold as P_2. Retain the data from point $P_1 - segment/f$ to $P_2 + segment$. Here, we are likely to retain extra data to the right and left of the song segment.

Choosing the correct threshold is also a primary concern. We do the analysis on the audio recordings of landscape, that might possibly contain the Magpie Robin bird. It is likely that we could pick up other bird calls or noises, could also miss Magpie Robin bird. Analyse this data and make a choice to select the threshold for the particular bird. This analysis part is presented in the next chapter, refer section 4.2. Note that the threshold, reference spectrogram and average bird song length depends on the bird song.

3.3.1 Algorithm

The algorithm for recognising a bird song from the continuous recording of a landscape is given below pseudo code 1. The algorithm takes reference spectrogram, reference sampling rate, threshold, segment, factor and input audio file as input and outputs the start time and end time of the recognised songs. Once the start and end times are known then, they can be either used to extract the song from the input audio file or these timings can also be used to analyse the birds behaviour in that particular landscape. It is important to make sure that the input data is re-sampled to sampling rate of the reference, else it will result in mismatch in the resolution of the spectrogram of the input audio signal and the reference spectrogram.

3.3.2 Pseudo code

Data: Reference spectrogram, reference sampling rate, threshold, segment, factor, input audio data
Result: Start and end points of the retained data
Read number of channels, sampling rate, number of frames, sample width from the audio file;
A_{ref} = read(Reference spectrogram) ;
i = 0;
P_1 = -1;
P_2 = -1;
Q = -1;
while *Audio file isn't finished* **do**
    if *i==0* **then**
        data = read frames(segment);
        i++;
    else
        data = data[$\frac{segment}{factor}$: end] + read frames($\frac{segment}{factor}$);
        i++;
    end
    data = resample(data);
    B_{xx} = specgram (data, sampling rate);
    corr = corrcoef(A_{ref},B_{xx});
    if *corr* $>$ *threshold* **then**
        if P_1 == *-1* **then**
            P_1 = i;
            P_2 = i;
            Q = 0;
        else
            P_2 = i;
            Q = Q + 1;
        end
    else
        if *Q==-1* **then**
            Do nothing!!;
        else
            Q = Q+1 ;
        end
    end
    if *Q==factor* **then**
        Print start = $P_1 - segment/f$, end = $P_2 + segment$;
        P_1= -1;
        P_2 = -1;
        Q = -1;
    else
        Do nothing!!;
    end
end

Algorithm 1: An algorithm for Bird song recognition

CHAPTER 4

Results and Discussions

The analysis part in the case of bird song recognition is an important and also a tedious task. It is not possible to listen to 4 or 5 hours of audio recording and see whether we have have picked up all intended bird songs or not. Traditionally, people use spectrogram or sonograph for the analysis of audio recording of bird songs.

4.1 Analysis

In the previous chapter 3, we have seen the method for recognising the bird song from a continuous sound recording of a landscape by spectrogram correlation. If we print the spectrogram of each segment at each stage using the pseudo code 1, then it takes a lot of time to analyse the total data. Because, at each stage we will be analysing only $\frac{segment}{factor}$ amount of new data.

We can make few changes to the above pseudo code such that every time we can print $constant * segment$ amount of data, it's spectrogram and the retained signal. Once we are done with analysing this part, the plot should be able to refresh and present the next data, which makes it easy for the analysis. The following pseudo code 2 does the same.

The below pseudo code 2 reads $constant * segment$ amount of data from the audio file each time. It runs the bird song recognition algorithm on this data for different values of threshold. we plot the input data, it's spectrogram and outputs of the all the threshold values in the same plot by using subplot command. After this we read the data from keyboard to make sure that the user is done with analysing this figure. If the user wishes to move to the next part of the data, we could erase this plot and redraw the next part of the input data on the same plot. This is done until the file is finished or the user can wish to end the process at any point.

In figure 4.1, subplot (a) represents the $8 * segment$ amount of input data. subplot (b) is the plot of correlation coefficients. Each point in this plot represents the correlation between reference spectrogram and the data from $point - \frac{segment}{2}$ to $point + \frac{segment}{2}$. The 6 subplots are the outputs obtained for different values of thresholds. Each segment in the output is equal to input segment if it is picked up by

```
Data: Reference spectrogram, reference sampling rate, threshold array, segment, factor, input
      audio data, Bird song recognition algorithm
Result: Plots the data one after the other
Read number of channels, sampling rate, number of frames, sample width from the audio file;
A_ref = read(Reference spectrogram ) ;
display factor = 8 ; # specifies number of segments needs to be displayed ;
i = 0;
while Audio file isn't finished do
    data = read frames(display factor * segment);
    for Analysis isn't done with all the thresholds do
        Bird song recognition algorithm(data,threshold);
        Store the correlation values of all the segments in an array;
    end
    # This point we have start and end points of recognized segments for different threshold
    values;
    if i == 0 then
        figure();
        i++;
    else
        # we already have the figure from the previous step;
    end
    subplot(data);
    subplot(correlation values);
    subplots(outputs);
    subplot(spectrogram of the data);
    check = Read data from keyboard(stdin);
    if check == 'enter' then
        clear the plot;
    else
        exit out of the while loop;
    end
end
```

Algorithm 2: Pseudo code for the analysis of the bird recognition algorithm

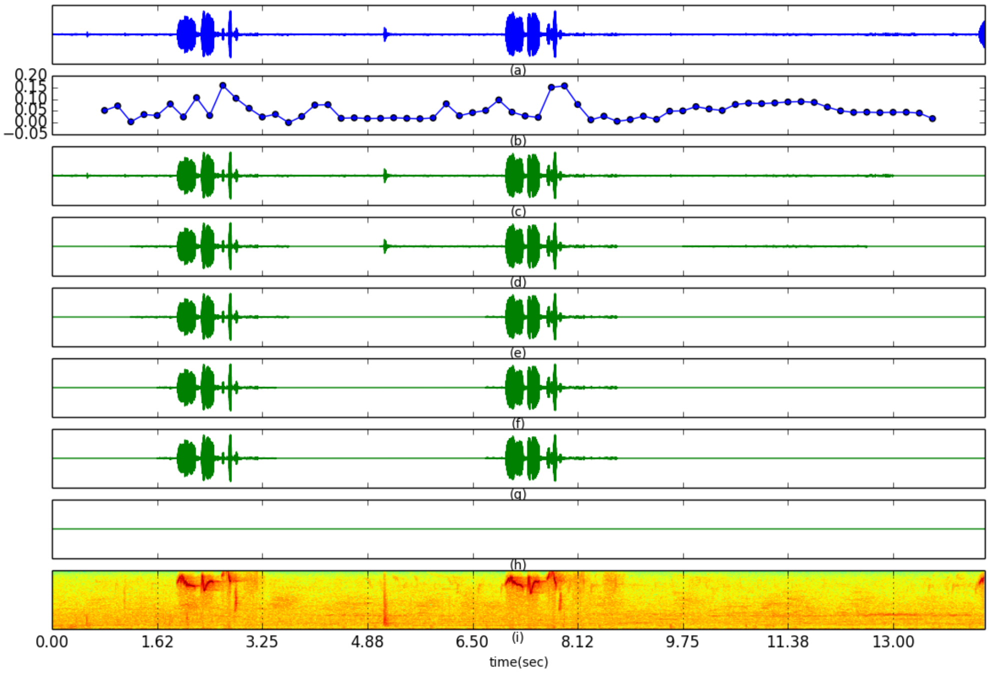

Figure 4.1: One instance of the output of the pseudo code 2. (a) input data, (b) correlation coefficients, sub-plots (c) to (h) are outputs of the algorithm with threshold values respectively 0.05, 0.08, 0.1, 1.2, 1.4, 1.6, (i) spectrogram of the input data

the algorithm else it is all zeros and the final subplot is the spectrogram of the input data.

4.2 Results and Observations

In the above figure 4.1, we have seen the analysis for the possible Magpie Robin data. It presents the output obtained using the bird recognition algorithm for different values of threshold. The output in the sub plots is plotted by placing zeros in place of ignored segments to make the understanding of the results easy. The recognised segments are divided into 4 categories.

1. True Positive(TP): Correct bird, i.e., Magpie Robin being identified
2. False Positive(FP): Other Bird songs or sounds being identified
3. True Negative(TN): Other Bird songs not being identified
4. False Negative(FN): Correct bird not being identified

The next few tables and figures emphasis on this analysis of the possible Magpie Robin data for few files. These sound recordings could be from different landscapes or from different recorders. Figures

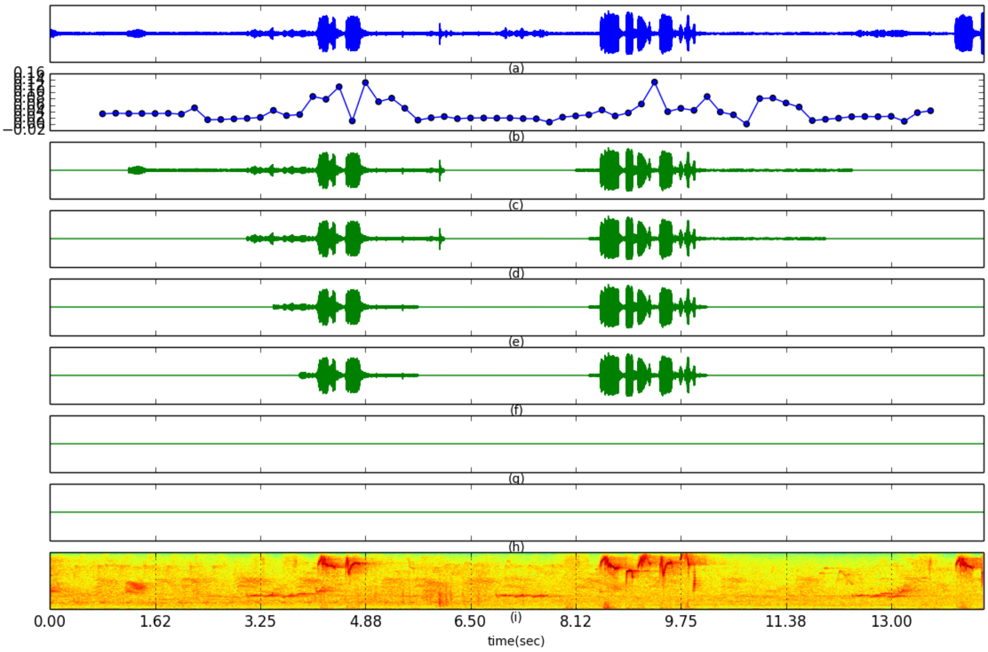

Figure 4.2: Effect of threshold on True positives, input is MR bird song in this case. (a) input data, (b) correlation coefficients, sub-plots (c) to (h) are outputs of the algorithm with threshold values respectively 0.05, 0.08, 0.1, 1.2, 1.4, 1.6, (i) spectrogram of the input data

4.2, 4.3, 4.4, 4.5 are the few instances of the analysis of the audio recordings of a possible Magpie Robin data. In figure 4.2, Magpie Robin songs(True positives) are recognised for the low values of threshold. Figure 4.3 and 4.4 have other bird songs(False positives) as inputs, in one case, it is not picking up the other bird songs for even smaller values of threshold, in the other case, it is picking up the other bird songs for low values of threshold. Figure 4.5 illustrates that there can be no bird songs but only background noise.

(a):

Threshold	TP	FP	TN	FN
0.05	23	30	0	0
0.08	23	15	0	0
0.1	20	2	0	2
0.12	19	0	0	4
0.14	12	0	0	9
0.16	7	0	0	13

(b):

Threshold	TP	FP	TN	FN
0.05	11	205	42	0
0.08	10	47	75	1
0.1	9	23	91	2
0.12	6	11	93	5
0.14	1	6	95	10
0.16	0	3	96	11

Table 4.1: Analysis of (a) mr2.220405.wav and mr1.220405.wav files (b) 20120518_150000.wav

From tables 4.1, 4.2 and 4.3, we observe that as the threshold increases, the number of true Magpie Robin songs(TP) being detected decreases. We can also see that as the threshold increases, number of

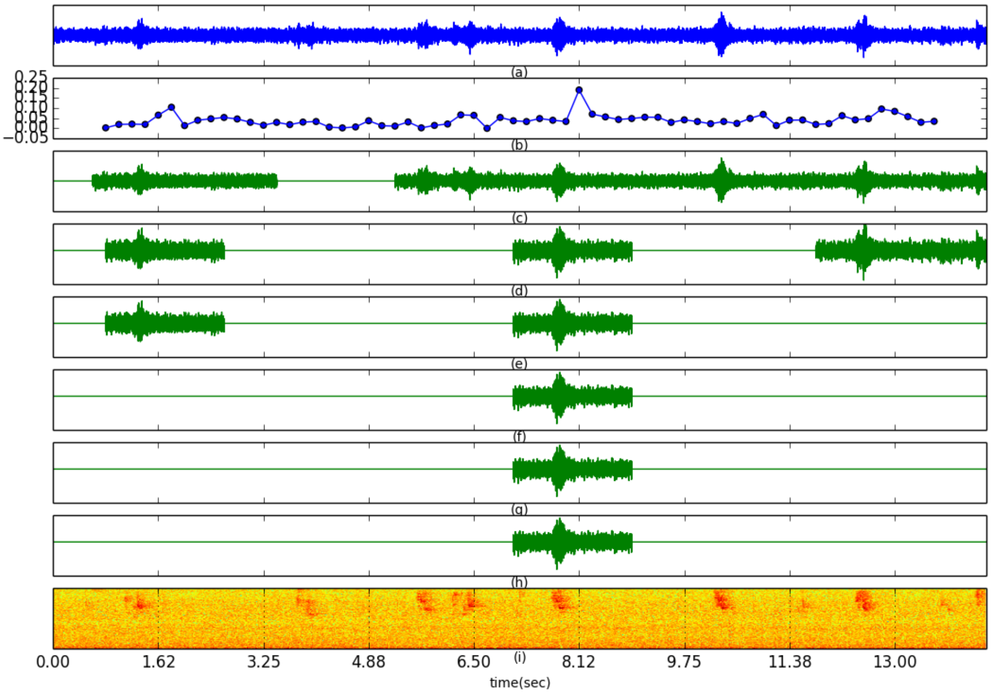

Figure 4.3: Algorithm picking up false positives, input is not MR bird song in this case. (a) input data, (b) correlation coefficients, sub-plots (c) to (h) are outputs of the algorithm with threshold values respectively 0.05, 0.08, 0.1, 1.2, 1.4, 1.6, (i) spectrogram of the input data

other songs and sounds(FP) being detected decreases and number of other songs(TN), not being identified increases. Note that sum of the false negative and true negative is not constant for all the thresholds, because silent recording segments recognised also come under false positive(FP) category. All these things intuitively make sense. Our next task is to decide the value of threshold for the correlation coefficient. While choosing the threshold, we should consider few things; the algorithm should be able to pick most of the reference bird songs and should be able to ignore all the other bird songs and silent recordings.

(a):

Threshold	TP	FP	TN	FN
0.05	0	418	20	0
0.08	0	120	20	0
0.1	0	0	20	0
0.12	0	0	20	0
0.14	0	0	20	0
0.16	0	0	20	0

(b):

Threshold	TP	FP	TN	FN
0.05	15	108	233	0
0.08	15	27	284	0
0.1	14	11	300	1
0.12	13	2	309	2
0.14	7	0	311	8
0.16	5	0	311	10

Table 4.2: Analysis of (a) 20120518_070000.wav (b)20120520_055000.wav

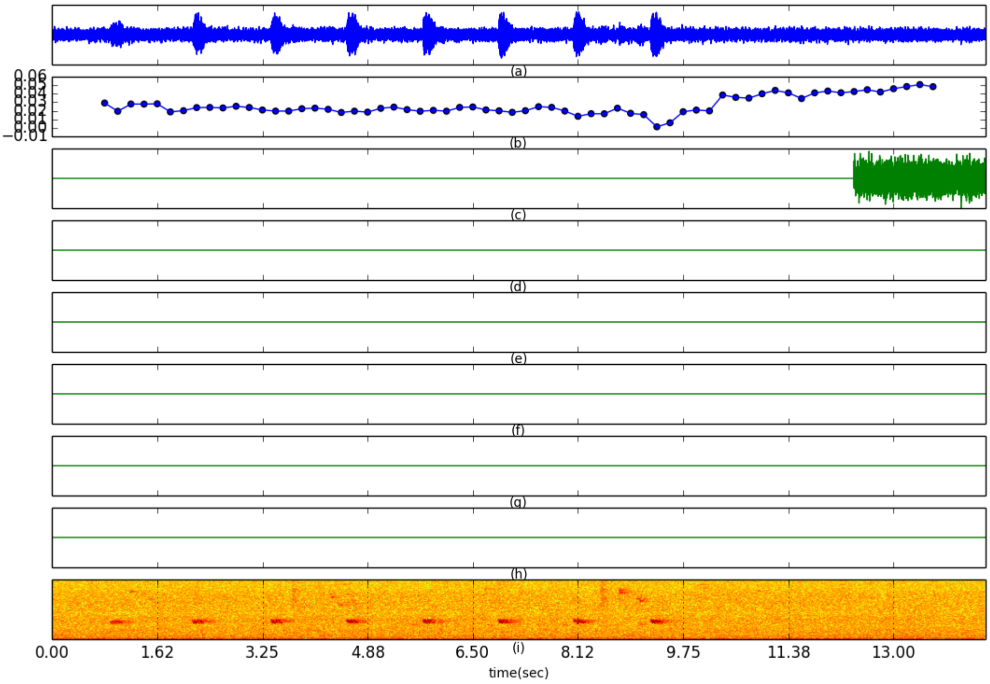

Figure 4.4: Algorithm ignoring other bird calls, input is not MR bird song in this case. (a) input data, (b) correlation coefficients, sub-plots (c) to (h) are outputs of the algorithm with threshold values respectively 0.05, 0.08, 0.1, 1.2, 1.4, 1.6, (i) spectrogram of the input data

Threshold, 0.05 for Magpie Robin bird picks all the True positives, but it also picks 1906 false positives which is 200Mb for the typical values of sampling rate = 16kHz. False negatives reduced by 90% for 0.1 threshold, but we lose roughly 10% of the true positives.

The performance of the algorithm can be objectively measured by two values, $\frac{Truepositives}{totalreferencebirdsongspresentinthedata}$ and $\frac{Falsepositives}{TotalBirdsongspresentinthedata}$. For Magpie Robin bird song and threshold = 0.1, these values are 90.6% and 15% respectively. Typically, a 4 hours audio recording is 920 MB($samplingrate = 1600$, $samplewidth = 16bits, stereo$) in size. Let us say the land scape has 10 different birds and 60 songs of each bird then,

(a):

Threshold	TP	FP	TN	FN
0.05	5	452	206	0
0.08	5	95	333	0
0.1	5	58	376	0
0.12	5	41	398	0
0.14	5	23	431	0
0.16	2	6	443	3

(b):

Threshold	TP	FP	TN	FN
0.05	10	693	236	0
0.08	10	138	382	0
0.1	10	74	428	0
0.12	9	37	459	1
0.14	9	18	473	1
0.16	5	12	479	5

Table 4.3: Analysis of (a) 20120519_150000.wav (b) 20120519_055000.wav

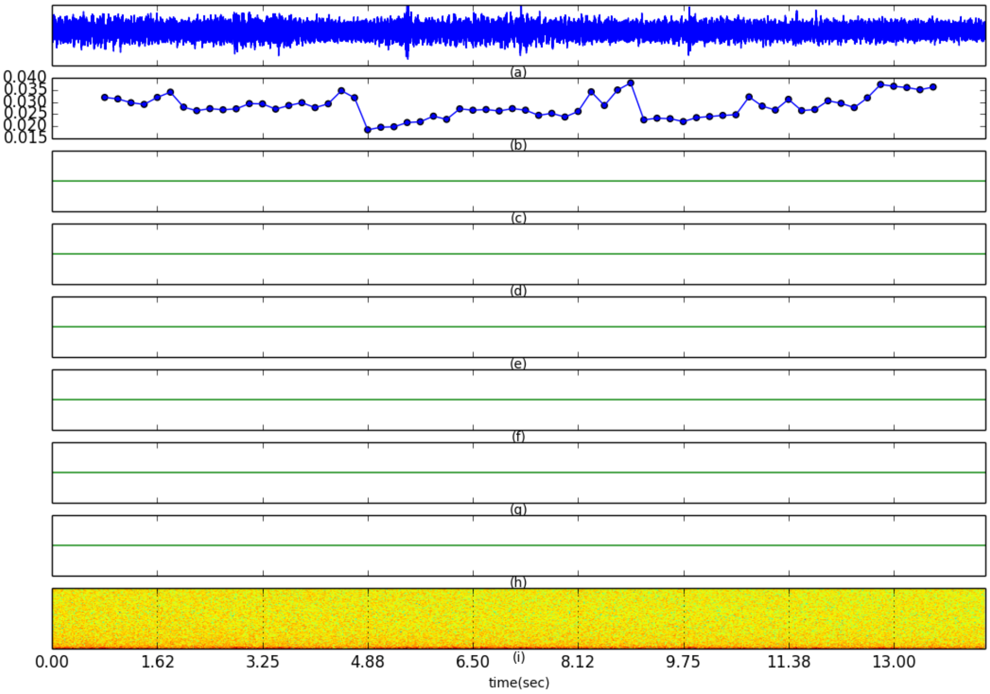

Figure 4.5: Input is just back ground noise in this case and the algorithm is able to ignore this data (a) input data, (b) correlation coefficients, sub-plots (c) to (h) are outputs of the algorithm with threshold values respectively 0.05, 0.08, 0.1, 1.2, 1.4, 1.6, (i) spectrogram of the input data

the output file after processing the data for all the 10 birds is going to be 173Mb $(16000 * 2(stereo) * 2(samplewidth) * 2sec * [0.9 * 60 * 10 + 0.15 * 10 * 540])$. We are able to reduce the recorded data by 80%, which is a lot when we are sending the data to a remote location.

Threshold	True positive	False positive	True negative	False negative
0.05	64	1906	737	0
0.08	63	442	1094	1
0.1	58	168	1215	5
0.12	52	91	1279	12
0.14	34	47	1330	28
0.16	19	21	1349	42

Table 4.4: Cumulative results of the total Magpie Robin files

CHAPTER 5

Conclusions

In this thesis, we have proposed a method to recognise the particular bird song from the audio recording of a landscape. Traditionally, people manually analyse the spectrogram of the audio recordings for the recognition of bird songs. We have used the similar technique, i.e., correlation of the reference spectrogram and the audio recording of a landscape. We have done the statistical analysis for choosing the value of threshold. It is important to make sure that the spectrograms of the reference and audio recording of a landscape are of same resolution in both time and frequency axis.

The future work may involve creating a database of reference spectrograms and threshold values for different birds. This algorithm can be implemented in a bird song meter to crunch the data in the field to make the transfer of audio recordings to a remote location feasible.

The entire implementation was done in PYTHON. The proposed method was tested on Magpie Robin bird song and the sound recording of the possible Magpie Robin data of a landscape.

REFERENCES

[1] David K. Mellinger, Christopher W. Clark. "Recognizing transient low-frequency wale sounds by spectrogram correlation," *J. Acoust. Soc. Am.*, Jun2000, Vol. 107, No. 6, 3518-3528.

[2] A. harma. "Automatic identification of bird species based on sinusoidal modelling of syllables," *International Conference on Acoustic speech signal processing,* 5:545-548, 2003.

[3] Katie Wolf. "Bird song recognition through spectrogram processing and labelling," *University of Minnesota - Twin Cities*.

[4] Deepika M, Nagalinga Rajan A. "Automatic identification of bird species from recorded bird song using ART approach," *International Journal of Innovative Research in Science, Engineering and technology*,2014, vol.3, special issue.3.

[5] Wave lab: Stanford University,
`http://www.statweb.stanford.edu/~wavelab`

[6] Krause, Bernie(1998). "Into a wild Sanctuary" *Berkeley, California: heyday Books*.