Master Thesis
Electrical Engineering
October 2016

Master of Science in Electrical Engineering with Emphasis on Signal Processing

Bird Chirps Annotation Using Time-Frequency Domain Analysis

Submitted by,

Suveen Kumar Vundavalli

Sri Rama Srinivasa Varma Danthuluri

Department of Applied Signal Processing
Blekinge Institute of Technology
SE–371 79 Karlskrona, Sweden

This thesis is submitted to the Department of Applied Signal Processing at Blekinge Institute of Technology in partial fulfillment of the requirements for the degree of Master of Science in Applied Signal Processing.

Contact Information:
Author(s):
Suveen Kumar Vundavalli
E-mail: suvu15@student.bth.se
Sri Rama Srinivasa Varma Danthuluri
E-mail: srda15@student.bth.se

University adviser:
Dr. Benny Lövström
Department of Applied Signal Processing

University examiner:
Dr. Sven Johansson
Department of Applied Signal Processing

Dept. Applied Signal Processing	Internet	:	www.bth.se
Blekinge Institute of Technology	Phone	:	+46 455 38 50 00
SE–371 79 Karlskrona, Sweden	Fax	:	+46 455 38 50 57

Abstract

There are around 10,426 bird species around the world. Recognizing the bird species for an untrained person is almost impossible either by watching or listening them. In order to identify the bird species from their sounds, there is a need for an application that can detect the bird species from its sound. Time-frequency domain analysis techniques are used to implement the application. We implemented two time-frequency domain feature extraction methods.

In feature extraction, a signature matrix which consist of extracted features is created for bird sound signals. A data-base of signature matrix is created with bird chirps extracted features. We implemented two feature classification methods. They are auto-correlation feature classification method and reference difference feature classification method. An unknown bird chirp is compared with the data-base to detect the species name. The main aim of the research is to implement the time-frequency domain feature extraction method, create a signature matrix database, implement two feature classification methods and compare them.

At last, bird species were identified in the research and the auto-correlation classification method detects the bird species better than the reference difference classification method.

Keywords: Bird Species Detection, Correlation, Identification, Time-Frequency Analysis, Signature Matrix

Acknowledgement

Firstly, we would like to thank our supervisor, Dr. **Benny Lövström** for his unconventional help to us. He made us learn from our mistakes in the process of thesis implementation. He provided us with many project related books and materials. He is also our teacher for the first lecture we attended in Sweden. We will be forever in debt of your help.

Thanks to our parents (**Satyanarayana Vundavalli & Srirama Vundavalli**), (**Danthuluri Dasaradha Uday Bhaskara Varma & Danthuluri Vani**) for encouraging us in reaching our goals.

Above all, we thank the universe for providing us such an opportunities, creating situations that makes us harder and blessings in achieving our life goals.

Contents

Abstract **i**

1 Introduction **1**
1.1 Motivation . . . 1
1.2 Aim and Objectives . . . 1
1.3 Research Questions . . . 2
1.4 Overview of Research . . . 2
1.5 Report Outline . . . 4

2 Related Work **5**

3 Methodology **7**
3.1 Creating Signature Matrix Database . . . 7
3.1.1 Selection of Bird's Sound . . . 7
3.1.2 Phrase Detection . . . 8
3.1.3 Decimation . . . 10
3.1.4 Feature Extraction Method . . . 12
3.1.4.1 Filter Bank . . . 12
3.1.4.2 Segmentation . . . 13
3.1.4.3 Feature Extraction . . . 14
3.1.5 Signature Matrix Creation . . . 16
3.2 Detection of Bird Species . . . 17
3.2.1 Auto-Correlation Feature Classification Method . . . 17
3.2.2 Reference Difference Feature Classification Method . . . 18
3.2.3 Detecting Bird Name . . . 20
3.2.3.1 Threshold Calculation . . . 20
3.2.3.2 Species Detection using Threshold Calculation . . . 21

4 Results & Analysis **23**
4.1 Phrase Separation . . . 23
4.2 Decimation . . . 25
4.3 Energy and Dominance Frequency Features . . . 27
4.4 Auto-correlation Classification Method . . . 33
4.5 Reference Difference Classification Method . . . 36

4.6 Hit Rate 39

5 Conclusion **41**

6 Future Work **42**

References **43**

A Appendices **45**

List of Figures

1.1 Signature matrix database creation 3
1.2 Detection of bird species 4

3.1 Components of bird sound recording spectrogram 9
3.2 Time domain bird sound signal 9
3.3 N^{th} order FIR filter 11
3.4 Block diagram of feature extraction method 12
3.5 Block diagram of stage - 1 : Filter Bank 12
3.6 Frequency filter bank in frequency domain 13
3.7 Block diagram of feature extraction method stage - 2 14
3.8 Block diagram of feature extraction method stage - 3 15
3.9 Sliding window technique 19
3.10 Block diagram for threshold calculation 21
3.11 Block diagram for species detection 22

4.1 First three phrases of bird sound signal 23
4.2 Plot of first detected phrase of Chaffinch 24
4.3 Plot of second detected phrase of Chaffinch 24
4.4 Plot of third detected phrase of Chaffinch 25
4.5 Frequency plot of a phrase of Chaffinch audio signal 26
4.6 Decimated frequency plot of phrase of a Chaffinch audio signal . . 26
4.7 Dominant frequency features of Chaffinch in Spectrogram 27
4.8 Energy features of Chaffinch in Spectrogram 28
4.9 Dominant frequency features of Whistle in Spectrogram 28
4.10 Energy features of Whistle in Spectrogram 29
4.11 Dominant frequency features of Owl in Spectrogram 30
4.12 Energy features of Owl in Spectrogram 30
4.13 Dominant frequency features of Song in Spectrogram 31
4.14 Energy features of Song in Spectrogram 32
4.15 Dominant frequency features of Warbler in Spectrogram 32
4.16 Energy features of Warbler in Spectrogram 33
4.17 Representation of hit and miss values between database data and test database for sum of mean and maximum correlation values . 34

4.18 Representation of hit and miss values between database data and test database for maximum correlation values 35
4.19 Representation of hit and miss values between database data and test database for mean correlation values 36
4.20 Representation of hit and miss values between database data and test database for sum of mean and minimum difference values . . 37
4.21 Representation of hit and miss values between database data and test database for mean difference values 38
4.22 Representation of hit and miss values between database data and test database for mean difference values 39

A.1 Sliding window technique for frequency band 6 45
A.2 Sliding window technique for frequency band 11 46
A.3 Frequency plot of human whistle before decimation 47
A.4 Frequency plot of human whistle after decimation 47
A.5 Frequency plot of owl before decimation 48
A.6 Frequency plot of owl after decimation 48
A.7 Frequency plot of chaffinch before decimation 49
A.8 Frequency plot of chaffinch after decimation 49
A.9 Frequency plot of human song before decimation 50
A.10 Frequency plot of human song after decimation 50
A.11 Frequency plot of Warbler before decimation 51
A.12 Frequency plot of Warbler after decimation 51

List of Tables

4.1 Minimum similarity between test and database required to identify the signal 40
4.2 Hit rate of correlation classification method 40
4.3 Hit rate of reference difference classification method 40

Chapter 1

Introduction

1.1 Motivation

Bird sounds can be used under several circumstances. It can be used as a recognition signal to indicate the individual, the kinship and the species of the birds [1] [2] [3] [4]. There are many reasons to identify the bird species [5] [6] [7] [8] [9]. There are many bird species that are on the edge of extinction. There is also a need to register number of species of birds or to make an account of population size of specific species of birds living in a particular area. By taking this survey, it could be easy to estimate the variation of population in the bird species from year to year and to identify the impact of surroundings on the bird species.

There is a need for a method that is easy to detect the bird species. But it is difficult to identify the bird species directly by seeing at it, because there are many species of birds living around us and also they mostly live on trees high above the ground.

So, if one can identify the bird species which are on the line of extinction just from the bird chirps, it can be easily noted and necessary action can be taken. But for more accuracy, a software identification of bird species from their chirps is important. It is easy to identify the bird species if one can be able to detect the bird species on the move.

1.2 Aim and Objectives

The main aim of this research is to propose a method to detect an unknown bird's name from its sound. The objectives of this research are:

1. Selecting three bird species, a whistle and a song.

2. Extracting phrases from the selected bird's chirps, whistle and song.

3. Extracting features from the phrases using two feature extraction methods.

4. Forming a database of signature matrices obtained from feature extraction methods.
5. Implementing two feature classification methods for unknown bird's chirp using database.
6. Selecting threshold values for 3 bird species chirps and other two sound sources.
7. Detecting the name of unknown bird chirp using thresholds.

1.3 Research Questions

1. Which one of the two feature extraction methods performs better in detection?
2. How processing speed of the feature extraction methods can be increased?
3. Which one of the two feature classification methods perform better in classification of signature matrices?
4. How accurate will be the detection?
5. How detection accuracy changes with single syllable bird and poly syllable bird.

1.4 Overview of Research

Bird species detection using audio signal involves feature extraction method, feature classification and species detection [2]. In our research, the bird species detection from audio signals involves two stages:

1. Creating a database of signature matrices of different bird species.
2. Detecting the name for the unknown bird sound using signature matrices database.

The processing of first stage involves selecting bird sounds extracted in laboratory, decimating the audio data without removing the feature data, feature extraction from audio data using feature extraction methods and creating a signature matrix database [3] [1] [10] [11] [2] as shown in Fig. 1.1.

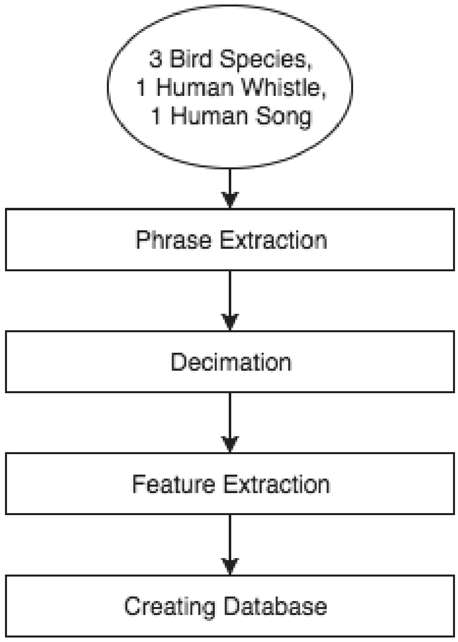

Figure 1.1: Signature matrix database creation

Whereas the processing in second stage involves loading unknown bird sound, creating a signature matrix just like in the first stage, classifying the signature matrix of unknown bird sound using signature matrix database and finally detecting name of bird species using values obtained in classification methods [3] [6] as shown in Fig. 1.2.

Mostly feature extraction techniques uses time domain analysis, frequency domain analysis and time-frequency analysis to extract features from a bird audio data [12]. Feature extraction using a time domain analysis is very useful for stationary data [12]. As bird's sound data is highly non-stationary, time domain feature extraction techniques are not useful for bird audio data extraction. Feature extraction using frequency analysis is helpful for non-stationary signals as well as stationary signal. This analysis extraction technique is useful to an extent. Due to the lack of temporal evaluation, this type of feature extraction is limited, whereas time-frequency analysis feature extraction techniques provide information about both temporal and spectral features [12] [10]. This reason leads to select the time-frequency analysis for feature extraction.

The final steps in this topic are classification and detection. Using classification algorithms, the newly obtained bird's sound data classified or processed with the signature matrix to obtain the commonness or similar features. This similar features data leads to detection of bird species.

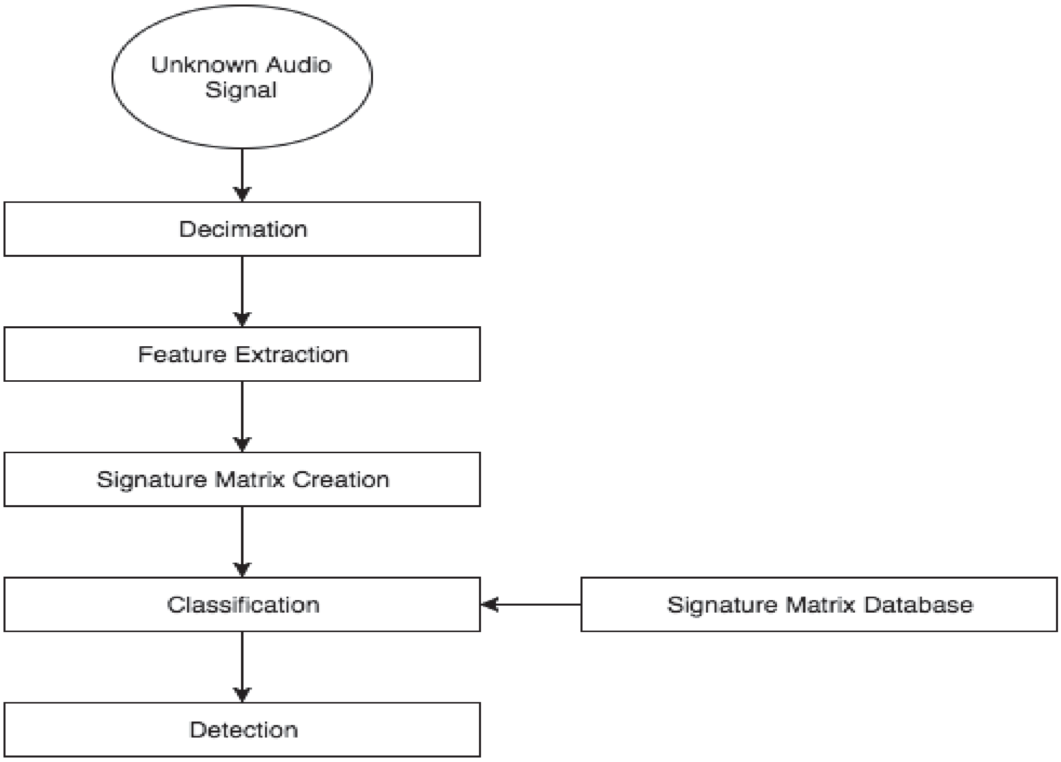

Figure 1.2: Detection of bird species

1.5 Report Outline

In this report, related works section describes how other researchers approached the bird species detection, their contribution and tools used from their work.

Methodology section describes about signature matrix database creation and detection of bird species. Signature matrix database creation describes about phrase extraction method, decimation, feature extraction method, signature matrix database. Detection of bird species part describes about classification methods, threshold values calculation and bird species detection.

Results section describes about results of phrase detection, decimation, feature extraction methods, auto-correlation classification method, reference difference classification method and hit ratio.

Conclusion and discussion section describes about the conclusions obtained after analysis of results.

Finally future works section describes about the research gap in our research.

Chapter 2
Related Work

The concept of syllable detection and feature extraction plays a major role in bird species identification. In [1], the authors analyzed male great reed warbler's song based on its extracted syllables. The analysis on the syllables extracted were done using time frequency analysis such as Wigner domain representation, Doppler domain, ambiguity domain, Instantaneous auto-correlation function and also with multitaper techniques. They concluded that the multitaper technique works way better than the other domain analysis.

From [1] we came to know that the feature extraction using only syllables can yield better results and also reduce the computation and increase the efficiency of the detection. In [1], they used single bird sound syllables for analysis and detecting the efficiency of different methods. But in our research analysis and detection is done with more than one bird.

In [13], the authors proposed an "algorithm to detect a multi component linear chirp signal, where the number of components is assumed to be unknown. The author demonstrated the performance of the proposed method on both real and simulated data and compared the results with the Cramer-Rao lower bound (CRLB) as well as competing algorithm. He used bat sound signal as real data. Finally concluded that it outperforms all.

In [13], the authors segmented each chirp signal to get the features out of it. We took this as basis and implemented our feature extraction method.

In [10], the author compared the quantified parameters of seven time-frequency representation algorithms for marine bio sonar research. The seven algorithms are "spectrogram (Sp), Wigner-Ville distribution (WV), Choi-Williams distribution (CW), Reassigned multitaper spectrogram (RMSp), locally stationary process multitaper spectrogram (LMSp), Re-assigned smoothed pseudo Wigner-Ville distribution (RSmWV) and spectrogram multiplication (SpM). The author conclude that RSmWV was most capable and robust technique in localizing closed spaced signal components. And also SpM method is good technique where there are no extreme demands when it comes to signal component localization precision."

In [10], the author implemented spectrogram (Sp) analysis, in our research the feature extraction methods were implemented based on Spectrogram (Sp) analysis.

In [11], the author implemented a technique for detection of bird species. "In this, parametric representation of sinusoidal modeling, Mel-cepstrum and vectors of various descriptive features were compared. The author concluded that for some species, the accuracy of detection is 70 %. But overall, the average detection is 40%-50% based on the method used".

In [11], the author suggested that automatic syllable segmentation can be implemented with 93% of accuracy. In the same way, we implemented an automatic phrase segmentation.

In [3], the aim of the author is to find out the detection accuracy with entire signal and with high amplitude parts of the signal. He implemented an algorithm to automatically detect the high amplitude parts of the signal from the entire recording. At last he concluded that detection of bird species is better for selected high amplitude parts of signal than the entire signal. His conclusion made us to implement the detection with phrase extracted signal.

Chapter 3

Methodology

The methodology of this research involves 2 stages. They are

1. Creating signature matrix database.
2. Detecting unknown bird species name.

3.1 Creating Signature Matrix Database

3.1.1 Selection of Bird's Sound

The sources needed for this research are bird sounds. The sources are mainly taken from online databases [14]. The other way of collecting is by recording the bird sounds directly from the forests, parks etc. But this requires lot of expertise and high definition equipment. By considering the limitations, collecting sources from online database is better than recording the bird's sound sources in real time.

The sound sources collected from the online sources should be laboratory generated or should be noiseless [1]. Otherwise feature extracted values will be polluted. A noise removal algorithm can be implemented to the sound sources. But the noise removal algorithms take away the less amplitude feature extracted values and that results in less efficiency and less accuracy of the detection.

In this research, three bird's species sounds were collected from online website [14]. A Chaffinch, an Owl and a Willow warbler bird's sound were collected. The collected bird sounds contains 50 phrases of Owl bird sounds, 240 phrases of Chaffinch bird sounds and 197 phrases of Willow warbler. The reason for selecting more than one audio data of same bird is that some birds can sing in different sounds. Bird sounds contain calls and songs. Songs are melodious sounds of the bird and calls are transient sounds [3]. For example, a Lyrebird can sing in 20 different sounds. It's harder to find the name of the bird species by having a single type of sound in our database. In order to increase the accuracy of the detection, more than one type of song of bird species should be collected.

In order to detect the real time accuracy of the research, a human whistle and a human song were recorded in real time. Again 84 recordings of 5 seconds whistles and 83 recordings of 5 seconds song were recorded. The collected recordings of the sound from the sources are in different types like MP3, MKV etc. It's difficult to process in MATLAB when the source files are in different formats. On a common basis, the different audio formats are changed to .WAV format [5].

Sampling frequency of collected data must be 44,100 samples per second. Otherwise, there will be problems with decimation factor. The collected audio data can be of different size. The feature extracted techniques using in the research captures the spectral and temporal components of audio data. So, the size of the audio data does not matters.

3.1.2 Phrase Detection

Usually bird sound are composed of three different components as shown in Fig. 3.1. These components are phrases, syllables and notes [15]. Phrases contains both syllables and elements. Phrase is a pattern of occurring syllables and elements. Syllable is a composition of several notes. Syllables of different bird's sounds vary from single syllable to many syllables [13] [10] [11]. Single syllable sounds are called mono syllable sounds. Many syllable sounds are called polysyllable sounds. Owl bird sound in our database is a mono syllable sound. Chaffinch, Willow Warbler, Human Whistle and Human Song sounds are poly syllable sounds. Note is relative time and pitch of a bird sound that contains highness and lowness.

In our sources, recordings available are more than 10 minutes of data. This data contains several silent regions as shown in the below figure. There is no use of these silent regions in processing. These silent regions create problems in detection phase [3] [11]. The unknown bird sound that needed to be detected may contains these silent regions. The silent regions in unknown sound sources and sound sources available in our database may correlate and gives wrong detection [5] [12]. In addition, these silent regions takes extra processing speed and computation. These silent regions are needed to be removed.

In our research, an algorithm is implemented to detect the phrases. A single recording contains several phrases as shown in Fig. 3.2. The algorithm is implemented in such a way that it separates all the phrases in recording. These separated phrases are stored individually in the database under file name of particular bird sound. This algorithm is only implemented for the recordings of bird sounds that are used for database creation. Implementing this algorithm for unknown bird species does not give any change in the results.

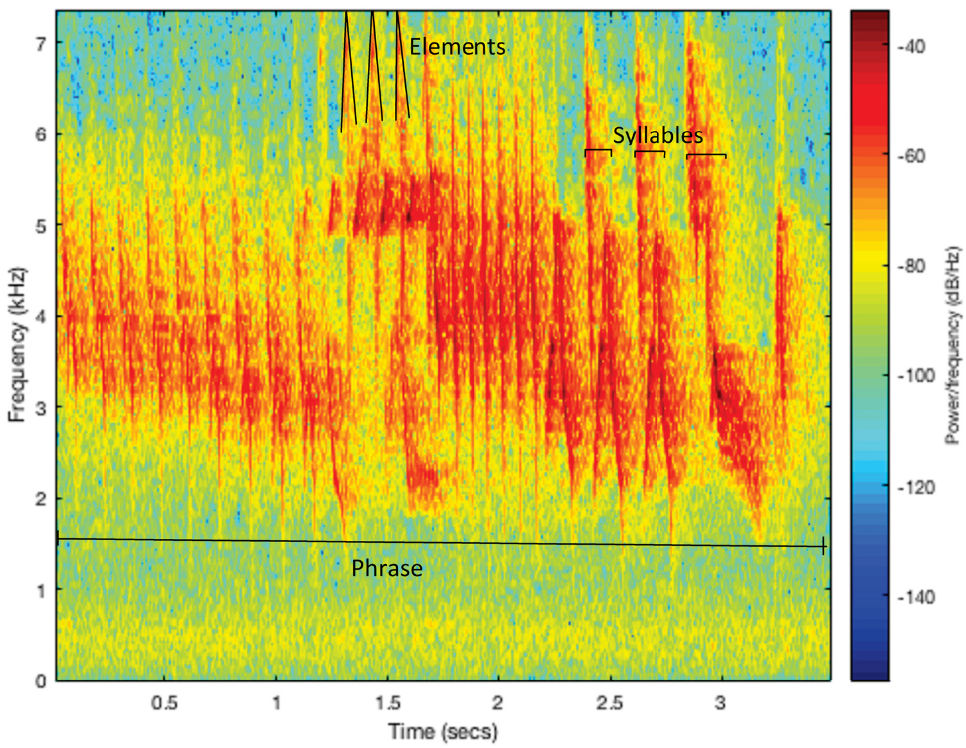

Figure 3.1: Components of bird sound recording spectrogram

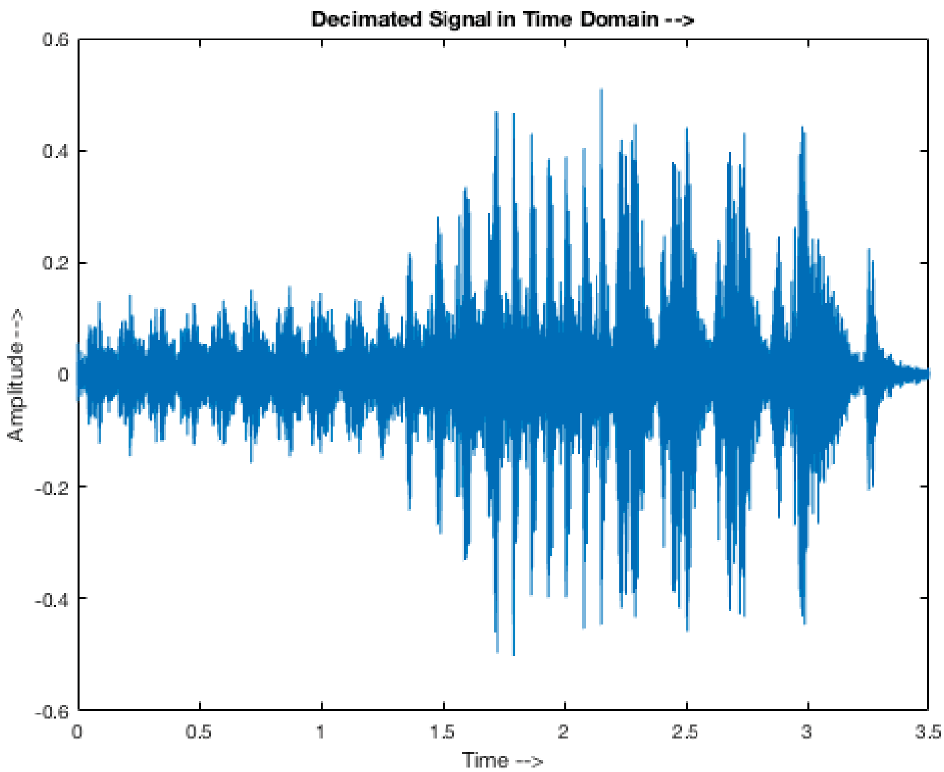

Figure 3.2: Time domain bird sound signal

3.1.3 Decimation

The process by which the sampling rate of signal can be decreases with a factor is called down sampling. By implementing a low-pass filtering before implementing the down sampling, the resultant process is called decimation. The decimation factor is the ratio of input signal to that of output signal in samples [1].

Usually for birds, the frequency range is 1,000 HZ to 8,000 HZ. Mostly, all the electronic audio recorders, record the signal with a 44,100 of sampling frequency [1] [11] [9]. According to Nyquist criteria (3.2), sampling frequency (fs) must be greater than the two times of maximum frequency ($2f_m$) i.e., $f_s \geq 2 * f_m$. Otherwise, sampled signal will be effected by aliasing.

New sampling frequency $= \frac{f_s}{Decimation factor} = \frac{44{,}100}{2} = 22,050$

According to Nyquist criteria,

$$F_s \geq 2 * maximum frequency \tag{3.1}$$

$$22,050 \geq 2 * 8000Hz \Rightarrow 22,500 \geq 16,000$$

By decimating signal with decimation factor 2, signal avoids from aliasing effect. The natural frequency of Chaffinch, Texas Barn Owl, Willow Warbler, Human Whistle and Human Song is below 6000 Hz.

i.e., new sampling frequency $= \frac{f_s}{Decimation factor}$

Assuming decimation factor $= 3 \Rightarrow \frac{44{,}100}{3} \Rightarrow 14,700$

According to Nyquist criteria, $F_s \geq 2 * maximum frequency$

$\Rightarrow 14,700 \geq 2 * 6000Hz \Rightarrow 22,500 \geq 12,000$

For our sources, decimation factor of 3 is suitable to avoid aliasing effect.

After decimation some higher frequencies components will be left. In order to remove those high frequency components, the decimated signal is passed through low pass filter as in Fig. 3.3. In MATLAB by default, the decimate command uses Chebyshev IIR filter. IIR filter computes in regression method. So, it gives the output samples depending on the removed samples. Due to this, computational complexity of IIR is higher when compared to other filters like FIR filter.

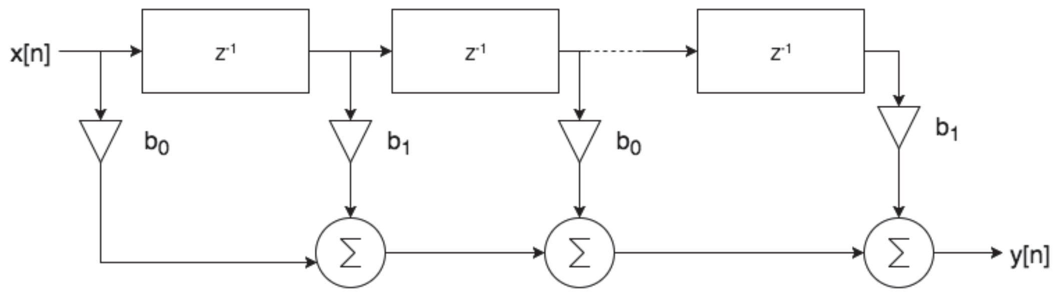

Figure 3.3: N^{th} order FIR filter

The FIR filter's cut-off frequency is in a way that it suppress those high frequency components that causes aliasing effect. Unlike IIR filter, FIR filter does not follow any regression methods in executing. But the Fir filter's delay is half of the original. Here computational complexity is compromised with accuracy. As the rate of computational complexity decreases with the increase in the size of the signal to be down sampled.

The decimation using FIR filter equation is shown in equation (3.2),

$$y(n) = \sum_{k=0}^{M-1} x[nM-k].h[k] \tag{3.2}$$

Where, h(k) is impulse response of the FIR filter
K is the length of the filter,
M is the decimation factor,
(.) is the dot product
x(nM-k) is input signal down sampled by factor M.

The cut-off frequency of IIR filter is $\frac{0.8}{decimationfactor}$ and a pass band ripple 0.05 dB. In order to create transfer function, there will be pass band distortions due to round-off produced in the convolution. Whereas in FIR filter, the cut-off frequency is $\frac{1}{decimationfactor}$. When using IIR filter in decimator, the filter to sound signal applied in forward as well as in reverse directions. But with FIR filter, the filter applied in only one direction. This process saves space and is very useful in long sequence signals.

By default FIR filter order is 30. In Fir, for order greater than 13 the decimation factor divides into several factors. The advantages of decimating in several times, the processing speed increases. If the IIR filter order is greater than 13, decimate process results are unreliable, which is not suitable in our research. Due to these advantages we opted for FIR filter in our decimation process.

3.1.4 Feature Extraction Method

The feature extraction is carried out in three stages as in Fig. 3.4.

1. Applying filter bank
2. Segmentation
3. Feature extraction

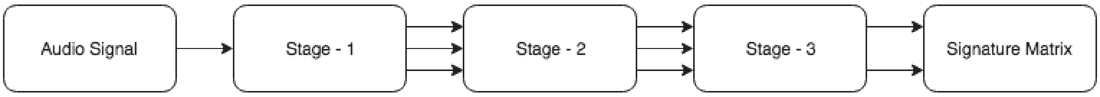

Figure 3.4: Block diagram of feature extraction method

3.1.4.1 Filter Bank

Normally filter banks are an array of Band-pass filters. These band-pass filters divides the signal into several frequency components and each band-pass filter carries a single frequency sub-band of the whole signal.

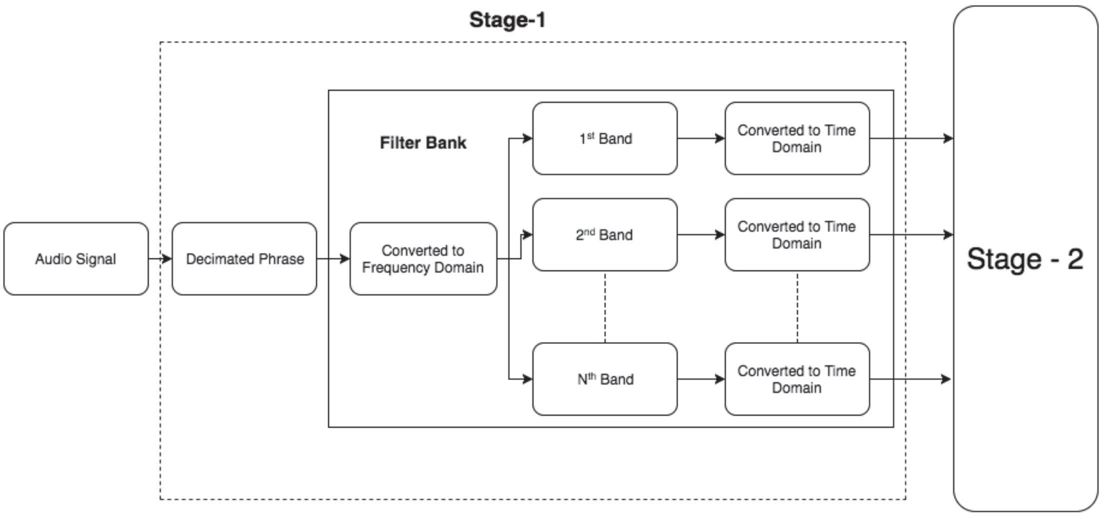

Figure 3.5: Block diagram of stage - 1 : Filter Bank

There are some problems in using these array of band-pass filter banks. One of some is aliasing. By using this type of filter bank, the values extracted from the extraction methods will be polluted. In order to avoid extracted values pollution in this experimentation we opted to a different filter bank method.

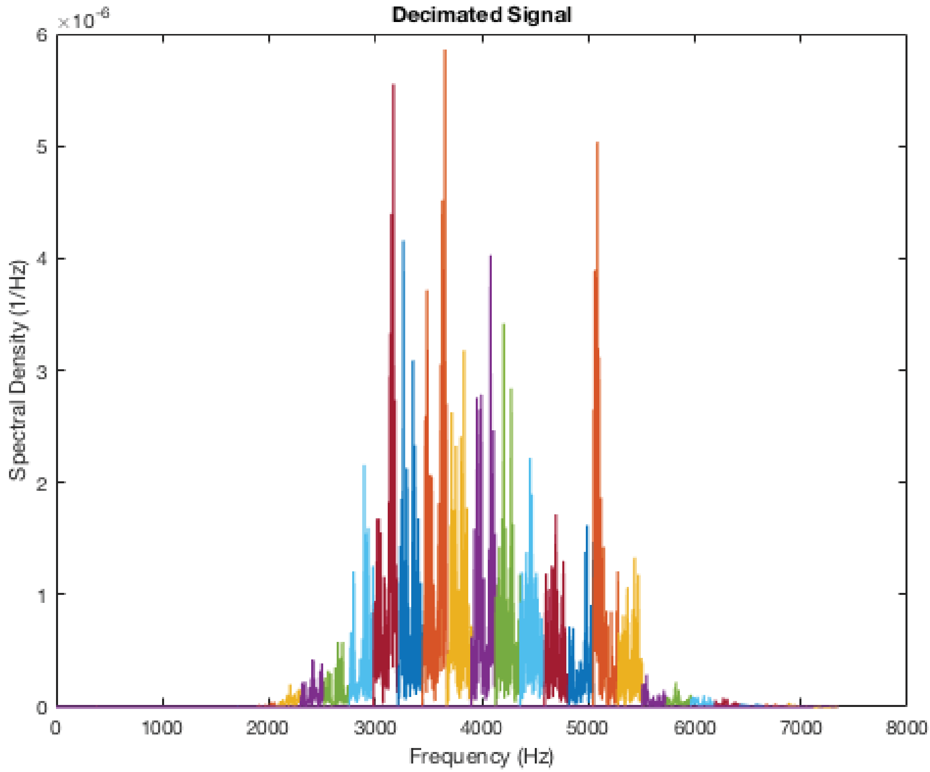

Figure 3.6: Frequency filter bank in frequency domain

In this opted filter bank, the frequency domain of the signal will be separated into a range of frequencies as shown in Fig. 3.5 by passing through a loop processing. In this method, there is no extraction values pollution when compared to the array of band-pass filter banks.

In this method, the time domain audio signal is converted into frequency domain. Then the frequency spectrum is divided into 32 sub-bands of equal band width as shown in the Fig. 3.6 which is divided into 32 sub-bands. Again each frequency sub-bands is converted into time domain. Finally these 32 time domain signals are passed to the segmentation stage.

3.1.4.2 Segmentation

Segmentation is nothing but cutting the signal into several parts as shown in Fig. 3.7. The 32 frequency domain signals taken from the filter bank is segmented in this section. The 32 frequency bands are converted individually into time domain signals. Each signal in 32 time domain signals is segmented into 40 ms using MATLAB processing techniques. Each time domain signal can be segmented to below 40 ms [12] [10]. But, by segmenting the signal into less than 40 ms will result in unnecessary evaluation of too many segments.

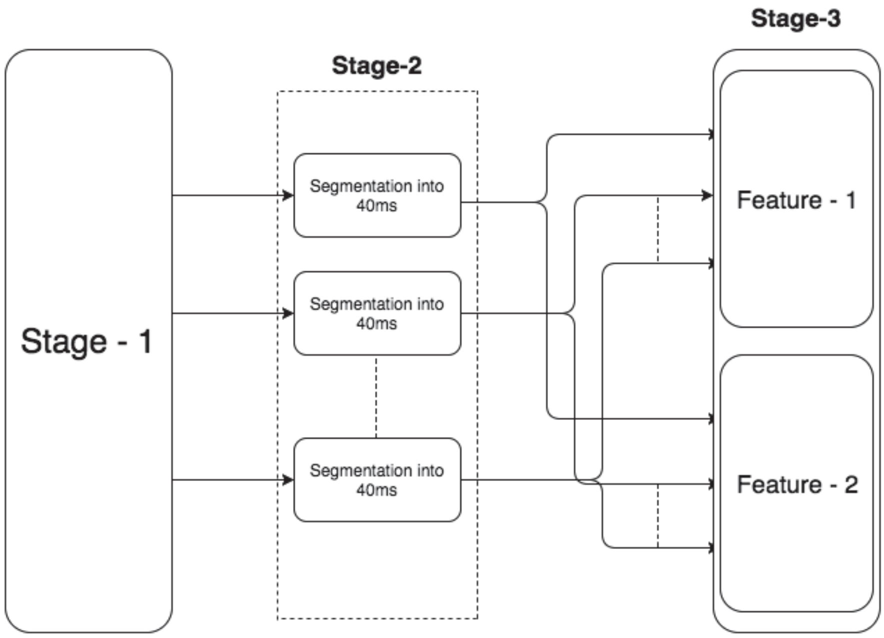

Figure 3.7: Block diagram of feature extraction method stage - 2

Segmenting signal to below 40 ms will also make the segmented signal stationary. This bird species detection method is kind of similar to pattern detection in [12]. A bird species is detected based on the time and frequency patterns in the audio signal. By segmenting the signal above 40 ms will result in the loss of resolution of the non-stationary parts. This further leads to wrong detection.

3.1.4.3 Feature Extraction

The features are the special characteristics of bird signals. Every audio signal has different feature values [1] [13] [10]. These features are unique for every bird sound signal. In our research, energy values and dominance frequency values are taken as features as shown in Fig. 3.8.

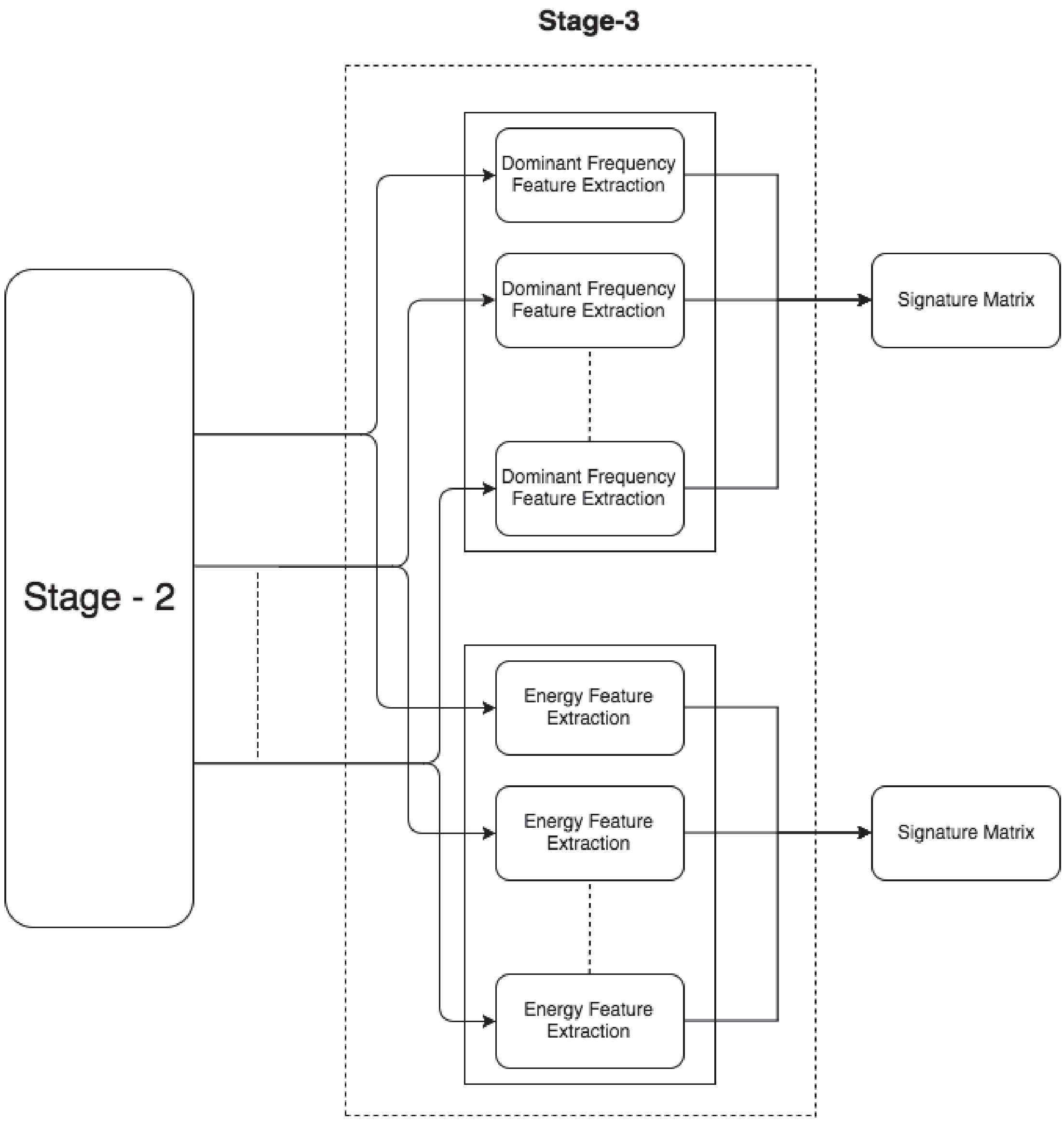

Figure 3.8: Block diagram of feature extraction method stage - 3

In order to extract Energy and dominance frequency features each 40 ms signal calculated in the previous stage is converted into frequency domain. The maximum frequency value is calculated for each 40 ms signal. These maximum frequencies in 32 time domain signals are stored as signature matrix and called as dominance frequencies features.

In the same way as Dominance frequency extraction, energy values are calculated for every 40 ms signal in 32 time domain signals. These energy values are stored as signature matrix and called as energy features [3].

3.1.5 Signature Matrix Creation

The next step in this topic involves creating a signature matrix. Signature matrix is nothing but a matrix form of extracted feature values [12]. This extracted features of a particular bird species were stored in a database as signature matrix.

In a similar away, the signature matrix for 3 bird species, 1 human whistle and a lyrical song is created and stored in the database. The extracted features were stored individually with a file name as name of bird species [12]. In our research, we are extracting energy and dominance frequency features. The number of phrases for Chaffinch is 240, Owl has 50 phrases, Willow Warbler has 197 phrases, human whistle has 84 phrases and human song has 84 phrases. So, there will be 2 signature matrices for every phrase of every sound source. Then the number of signature matrices a Chaffinch has 480, an Owl has 100, Willow Warbler has 394, human whistles has 168 and human song has 168. The total number of signature matrices in the database are 1310.

For example, in our database there is a signature matrix of a bird named "Chaffinch" and the file of that signature matrix of all the phrases is stored in the database as "lyrebird.mat". In a similar way, extracted features of 4 other sound sources were stored in the database as "respective bird name.mat".

Location mapping is plays another important role in detection. The location mapping is keeping track of signature matrices of a particular sound source in the database. After storing the signature matrices in the database, location mapping is important to detect the name of the identified bird [7].

The main reason for creating this database is to provide the signature matrices of 5 sound sources to the classification methods [12]. Later in the classification methods, the signature matrix of unknown bird signal is compared with the 655 signature matrices of bird species in our database and name of the unknown bird species can be detected.

This implementation can only detect the unknown bird species that are only present in our database. The Matlab Code was implemented in such a way that it shows "not present in our database" message, if the unknown bird signal signature matrix is not compatible with the signature matrices in our database. Audio sources takes lot of physical space in electronic devices. From extraction methods, extracting features from audio sources and stored as signature matrix. Signature matrices replace sound sources. This leads to large memory audio sources compacting into small memory data features [8].

3.2 Detection of Bird Species

3.2.1 Auto-Correlation Feature Classification Method

Correlation is used to calculate the commonness between two variables. The correlation values will be in between +1 and -1. When the correlation co-efficient is nearer to ±1, the two variables will have high quantity of common elements. When the correlation co-efficient is nearer to 0, the commonness between them is very low.

If the time series $x(t)$ is correlated with another time series $h(t)$. The resultant correlation value will be $r(\tau)$ as in Eqn. (3.3) .

$$i.e., r(\tau) = x(t) * h(t) \tag{3.3}$$

Where t is the time series
τ is time lag
$(*)$ is convolution factor

Correlation of a matrix or function with itself at different lags in time is called Auto-correlation.

If the time series $x(t)$ is correlated with itself, then the resultant will be r_x as in Eqn. (3.4).

$$i.e., r_x = x(t) * x(t - \tau) \tag{3.4}$$

where, r_x is the auto-correlation value
$x(t - \tau)$ is delayed time series signal by τ
τ is time lag
(*) is convolution factor

By applying auto-correlation to our data, we can violate the assumption of instance independence. In our thesis, we need this specification. Because, the bird sounds recorded are starting at different instances. This auto-correlation method helps us in relational learning and inference.

In our auto-correlation perfect multicollinearity is almost impossible. Perfect multicollinearity is getting the auto-correlation value to exact ±1. Perfect multicollinearity is impossible. Though the sounds of a same bird same but the electronic equipment used to reduce these sounds are different and also due to the differences in specifications of components in different recorders, this condition occurs. And also our source files are taken from the online database. The auto-correlated co-efficient must be nearer to ± 1 in order to say that the given audio signal is similar i.e. classified values are equal to a particular bird values in our database. When the auto-correlated co-efficient is nearer to 0 (zero), the taken signal is not similar to that of audio signal in our database.

3.2.2 Reference Difference Feature Classification Method

A sliding window is a technique with different uses. Usually in sliding window technique, the window slides across the signature matrices. In our case, we are using this for classification of features of extracted values taken from the feature extracted methods.

The signature matrix of one of the feature extraction methods is considered and it is correlated with the same feature extracted signature matrix of database files. Though, the number of frequency bands are same in test and database signature matrices, but the time segments are not equal in real-time scenario. Hence there is a need for sliding window technique.

The database signature matrix is taken as reference signature matrix and the test signal signature matrix is considered as source signature matrix. The signature matrix with smallest time segments is later swapped to small matrix and the signature matrix with largest time segments is swapped to large matrix to use for sliding and for automated code generalization. Assume the size of the smaller signature matrix is 100×32 and the size of the larger signature matrix is 140×32, then the classification is applied between smallest signature matrix and the first 100 time segments among 140 time segments of larger signature matrix. Then the window is moved by one time segment and the classification algorithms are applied on it. Then it is moved by one time segment at a time until the window reaches the last 100^{th} time segment as shown in Fig. 3.9.

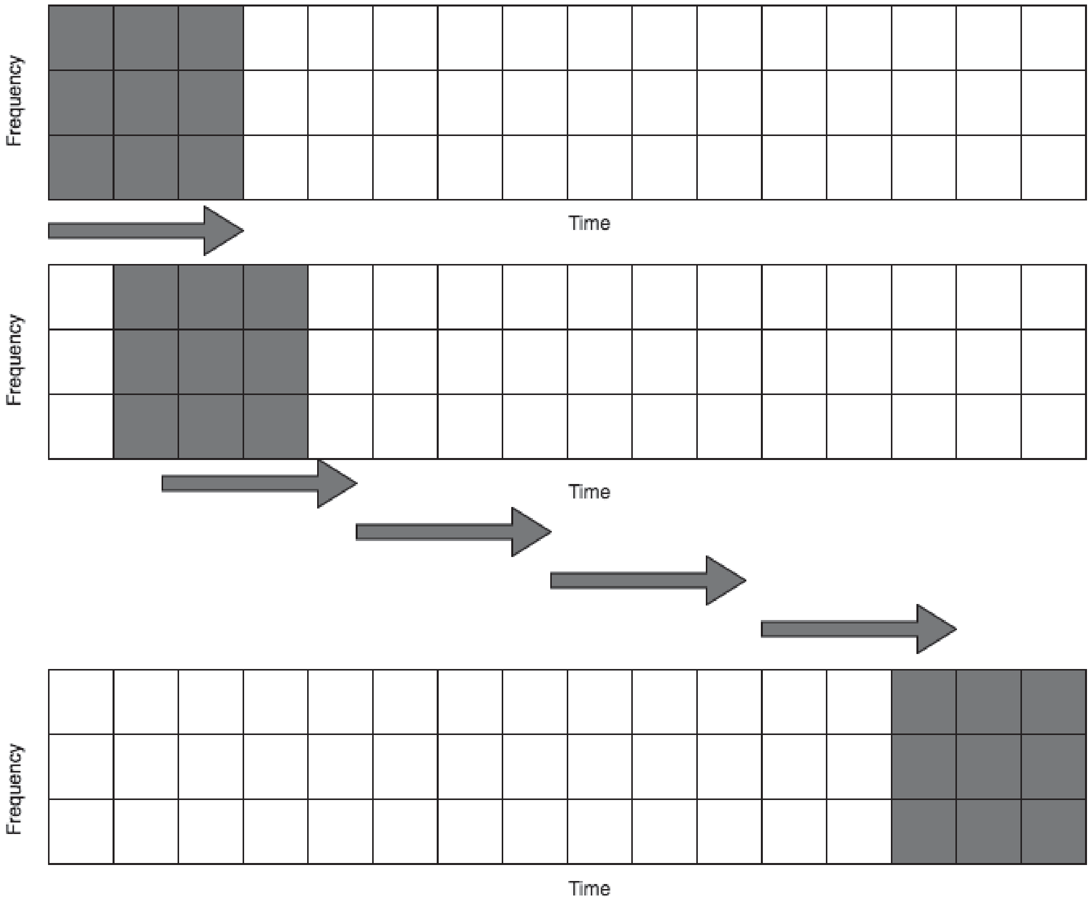

Figure 3.9: Sliding window technique

In this project, we are using this to take out the difference between the audio signal's extracted values and the extracted values of different bird sound signal's present in our database. When the mean values of difference is nearer to zero, there is a high chance of bird sound being similar to that of our bird sound in our database. When the mean values of difference is far from zero, then there is a very low chance of bird sound being similar to that of our bird sound in our database.

Getting zero mean difference value is almost impossible. Because, the source files we are using are taken from online. Where the audio files of same bird species is recorded in different geographical locations and with different electronic equipment's. Due to this reason there is a difference in frequency values of same bird species audio signal. So, zero mean difference is almost impossible.

3.2.3 Detecting Bird Name

Detecting the bird species is done at this stage. This is a novel method to calculate the minimum thresholds of a sound source to detected. Unknown bird chirp's signature matrix is classified with signature matrices in the database using classification methods. i.e., signature matrices of Owl, Whistle, Song, Chaffinch and Warbler were processed with unknown bird chirp using classification methods. This detection phase takes two stages. Firstly, testing for the threshold values and secondly, detecting the bird species according to threshold values.

3.2.3.1 Threshold Calculation

In this stage, as shown in Fig. 3.10 the preferable threshold values for real time implementation will be found. In order to find these threshold values, the following steps to be followed:

1. Take test sources of Chaffinch, Owl, Willow Warbler, human whistle and human song.
2. Apply feature extraction method to new test sources and creating signature matrices for all of them.
3. Apply classification methods for all the test sources using signature matrices in the database.
4. Create a matrix1 with maximum correlation values and a matrix2 with mean correlation values obtained in the Auto-correlation classification methods.
5. Create a matrix3 with sum of matrix1 and matrix2.
6. Change matrix1, matrix2 and matrix3 by keeping 10 highest values and making remaining values to zero.
7. Create a matrix4 with mean difference values and a matrix5 with minimum difference values obtained in the reference difference classification methods.
8. Create a matrix6 with sum of matrix4 and matrix5.
9. Change matrix4, matrix5 and matrix6 by keeping 10 highest values and making remaining values to zero.
10. Calculate sum of all the elements in matrix1, matrix2, matrix3, matrix4, matrix5 and matrix6 individually.

11. Then for every new test source find out the three minimum values from matrix1, matrix2, matrix3, matrix4. Matrix5 and matrix6.

12. From the above results in auto-correlation classification method, set the thresholds for matrix1, matrix2, matrix3, matrix4, matrix5 and matrix6 in such a way that the unknown sound source should be detected.

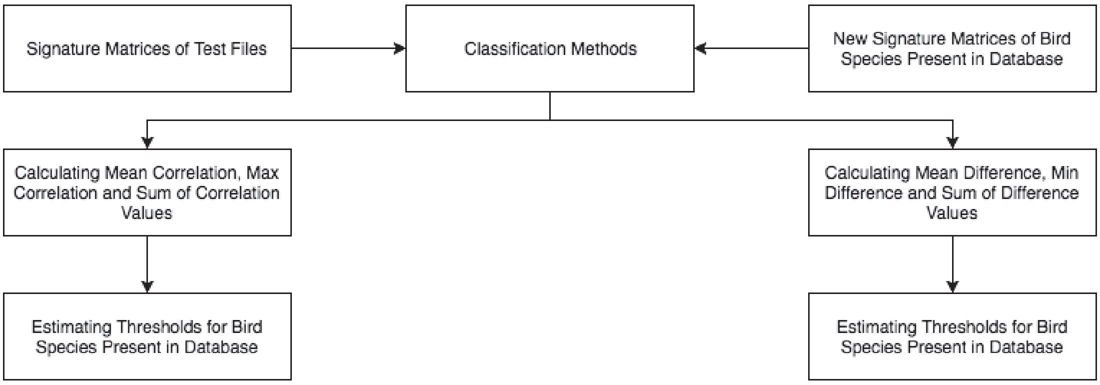

Figure 3.10: Block diagram for threshold calculation

3.2.3.2 Species Detection using Threshold Calculation

In the second stage, the detection of unknown bird species in real time will be done. The first 9 steps from the previous stage is implemented to unknown chirp. The next steps as follows:

1.
    - In auto-correlated classification method,

      If the matrix3 value of unknown sound source is less than the threshold of a particular sound source in the database, the message "sorry unable to identify the sound source" shows up. Otherwise, it moves to the next step.

    - In Reference difference classification method,

      If the matrix6 value of unknown sound source is greater than the threshold of a particular sound source in the database, the message "sorry unable identify the name of sound source" shows up. Otherwise, it moves to the next step.

2.
    - In auto-correlated classification method,

If the matrix-1, matrix-2 and matrix-3 values of unknown bird chirp is greater than the threshold values of a particular sound, the sound name shows up. Otherwise, the message "sorry! Seems like known signal, but unable to identify correctly" shows up.

- In Reference difference classification method,

  If the matrix4, matrix5 and matrix6 values of unknown bird chirp is less than the threshold values of a particular sound, the sound name shows up. Otherwise, the message "sorry! Seems like known signal, but unable to identify correctly" shows up.

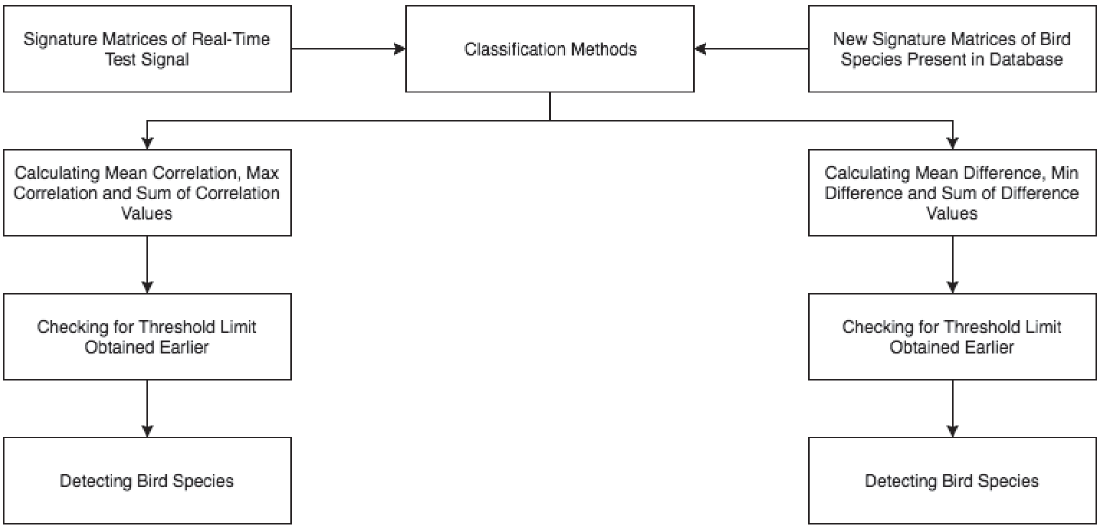

Figure 3.11: Block diagram for species detection

As every bird chirp in the database is processed with itself and with all the other birds' chirps, there will be lot of computations, the first stage takes lot of time to process. The second stage takes very less time to process. As there is only one unknown bird chirp is processed with all the other bird chirps in database.

Chapter 4

Results & Analysis

4.1 Phrase Separation

The recordings of the Chaffinch, Texas Owl, Willow Warbler, Human Whistle and Human lyrical song audio data were analyzed for phrase separation as shown in Fig. 4.1.

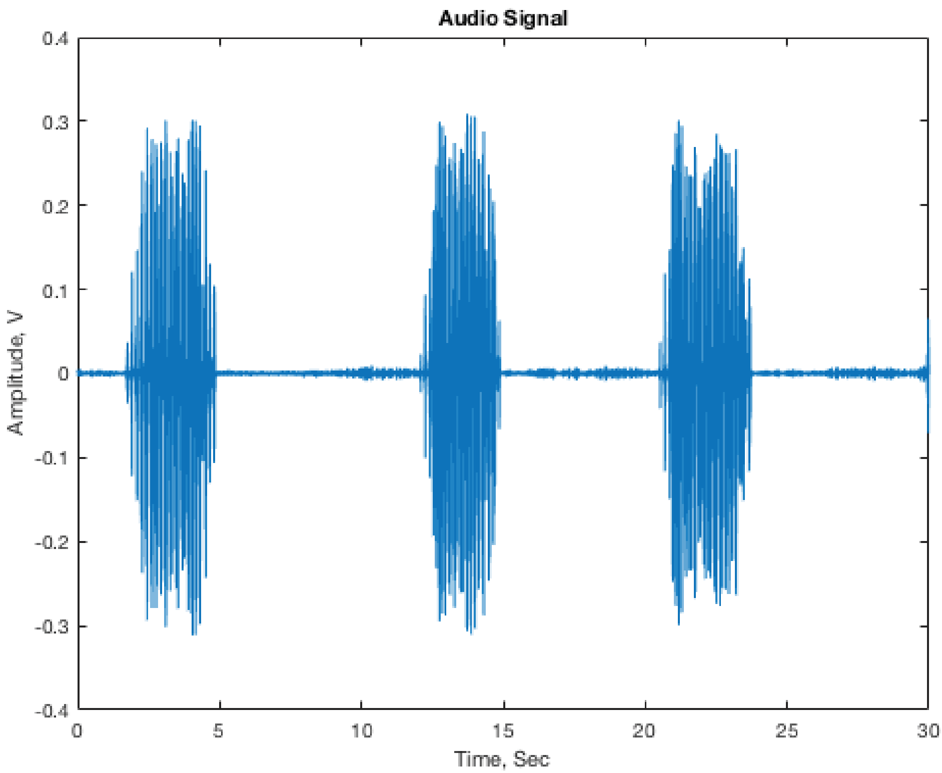

Figure 4.1: First three phrases of bird sound signal

The recordings of the Chaffinch, Texas Owl, Willow Warbler, Human Whistle and Human lyrical song audio data were processed to extract as discussed in section 3.1.2 the syllables from the audio file and the output is as shown in Fig. 4.2, Fig. 4.3 and Fig. 4.4

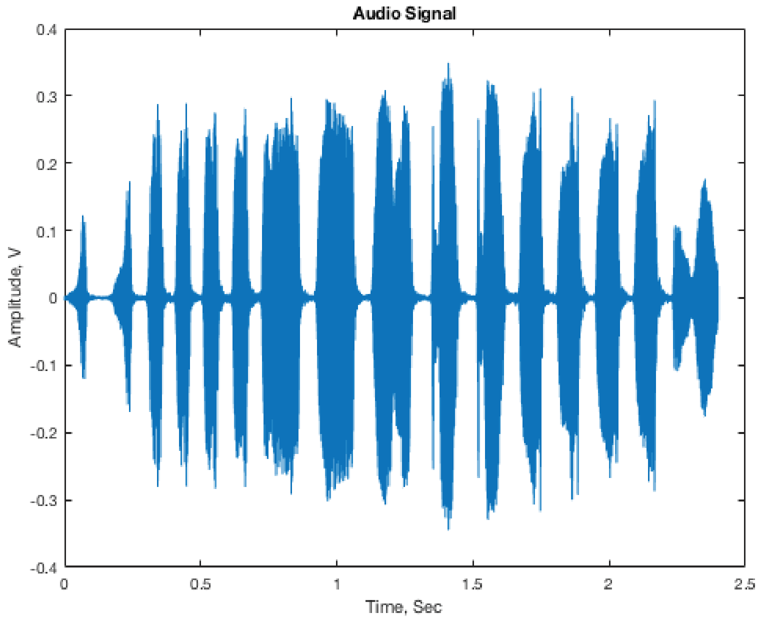

Figure 4.2: Plot of first detected phrase of Chaffinch

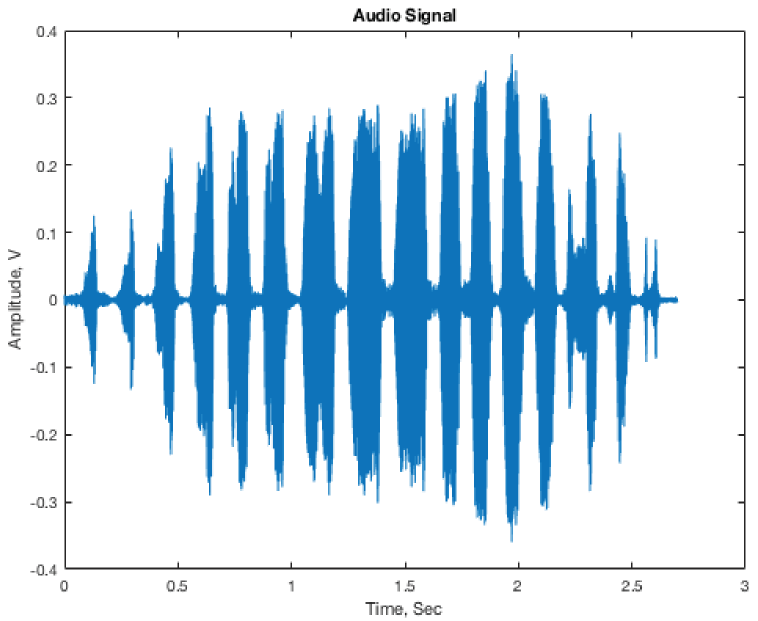

Figure 4.3: Plot of second detected phrase of Chaffinch

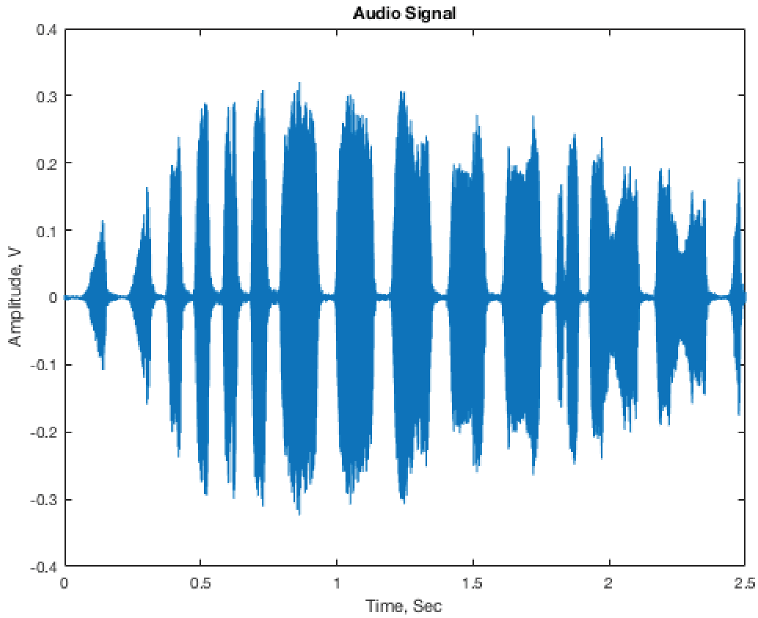

Figure 4.4: Plot of third detected phrase of Chaffinch

The number of samples used to process the detection implementation decreased because of the removal of silent regions in the original signal using phrase detection method.

4.2 Decimation

The each phrase from each individual sound source obtained from the 3.1.3 are converted into frequency domain. One phrase of Chaffinch is taken as example. The frequency domain of plot of that phrase is shown in the Fig. 4.5

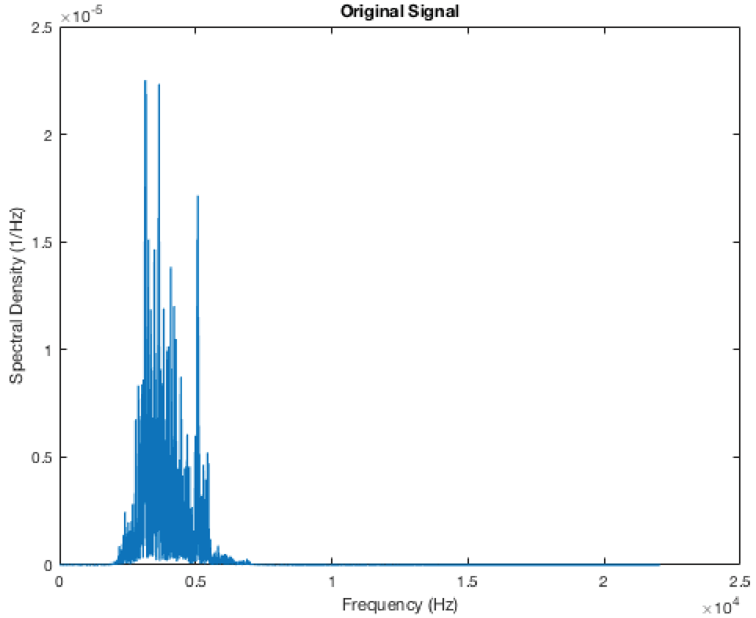

Figure 4.5: Frequency plot of a phrase of Chaffinch audio signal

The decimation is applied to the Each and every phrase of every sound source with a decimation factor 3 as described in the section 3.1.3. The decimated Chaffinch phrase signal is plotted and shown in Fig. 4.6

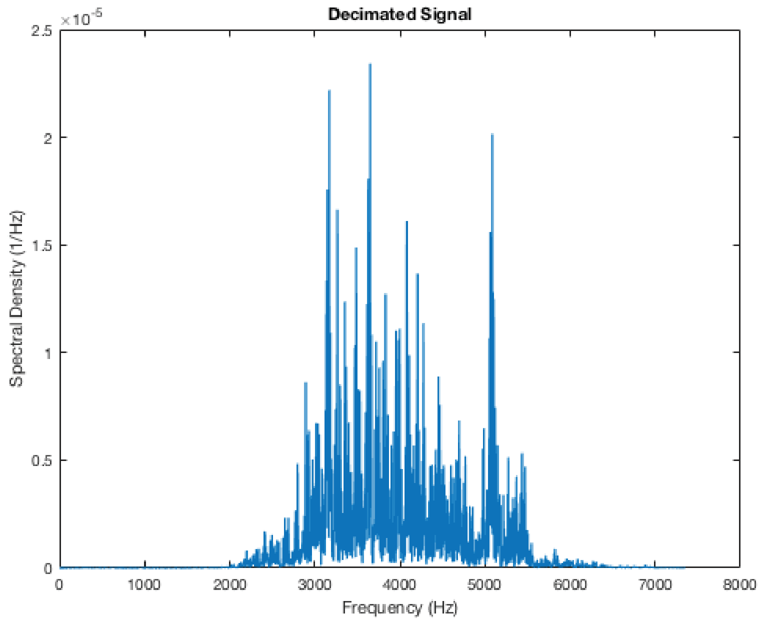

Figure 4.6: Decimated frequency plot of phrase of a Chaffinch audio signal

The number of samples in the Fig. 4.5 decimated by a factor of 3 as shown in Fig. 4.6.

4.3 Energy and Dominance Frequency Features

The feature extraction method is applied to each and every decimated signal obtained in the section 3.1.3 as described in section 3.1.4. Energy and dominance frequency features of every phrase of Chaffinch, Whistle, Owl, Song and Warbler is stored as a signature matrix in the database as shown in Fig. 4.7, Fig. 4.8, Fig. 4.9, Fig. 4.10, Fig. 4.11, Fig. 4.12, Fig. 4.13, Fig. 4.14, Fig. 4.15 and Fig. 4.16 respectively.

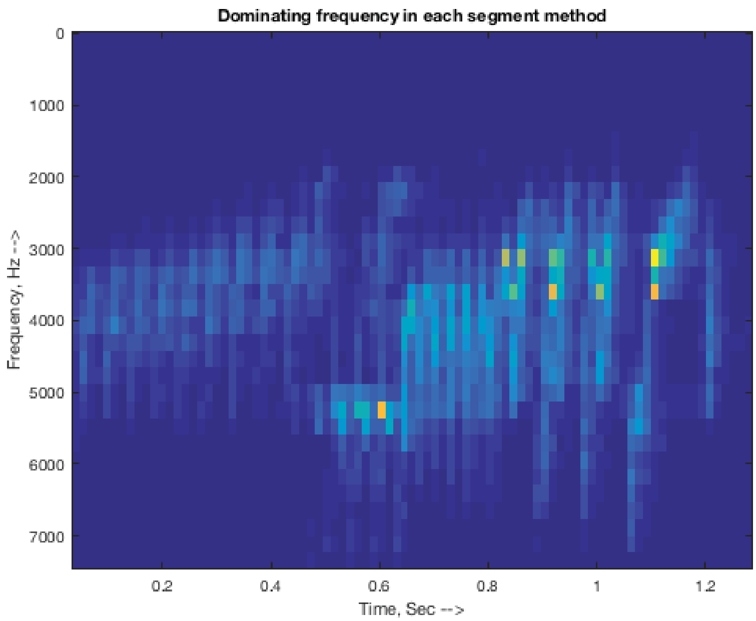

Figure 4.7: Dominant frequency features of Chaffinch in Spectrogram

Dominance frequency features of one phrase of a Chaffinch audio data of one phrase is plotted in spectrogram and shown in Fig. 4.7

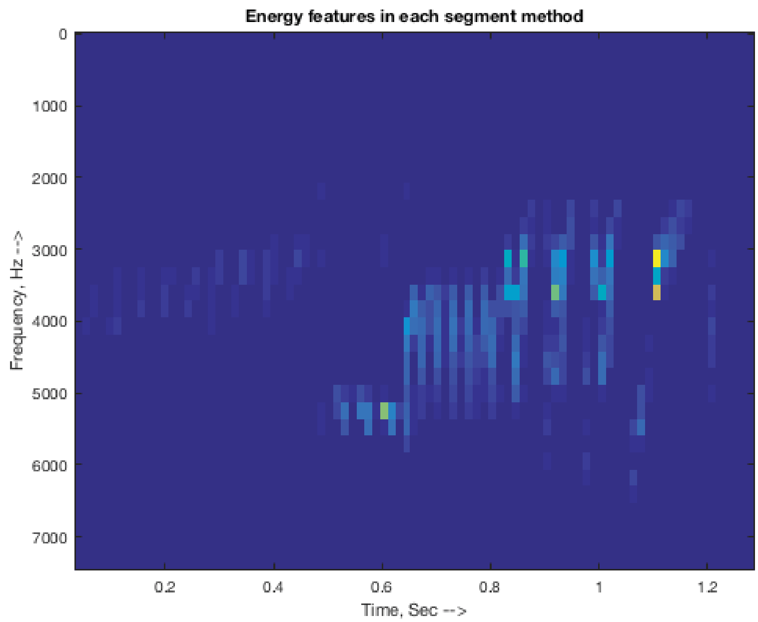

Figure 4.8: Energy features of Chaffinch in Spectrogram

Energy features of Chaffinch audio data of one phrase is plotted in spectrogram and shown in Fig. 4.8.

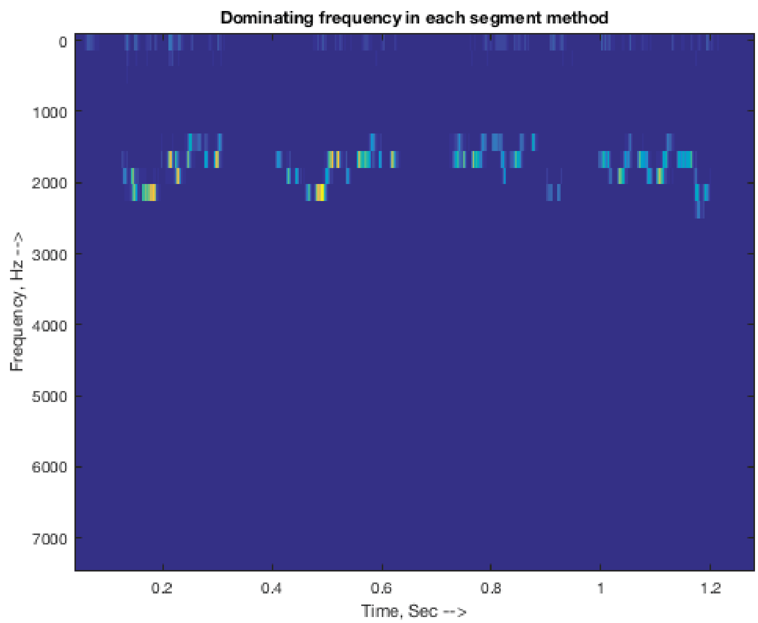

Figure 4.9: Dominant frequency features of Whistle in Spectrogram

Dominance frequency features of one phrase of a Whistle audio data of one phrase is plotted in spectrogram and shown in Fig. 4.9

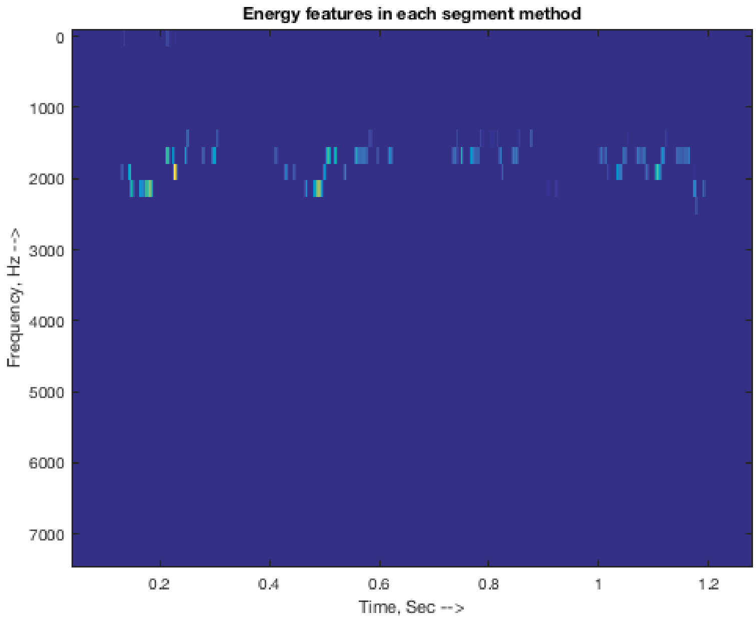

Figure 4.10: Energy features of Whistle in Spectrogram

Energy features of Whistle audio data of one phrase is plotted in spectrogram and shown in Fig. 4.10.

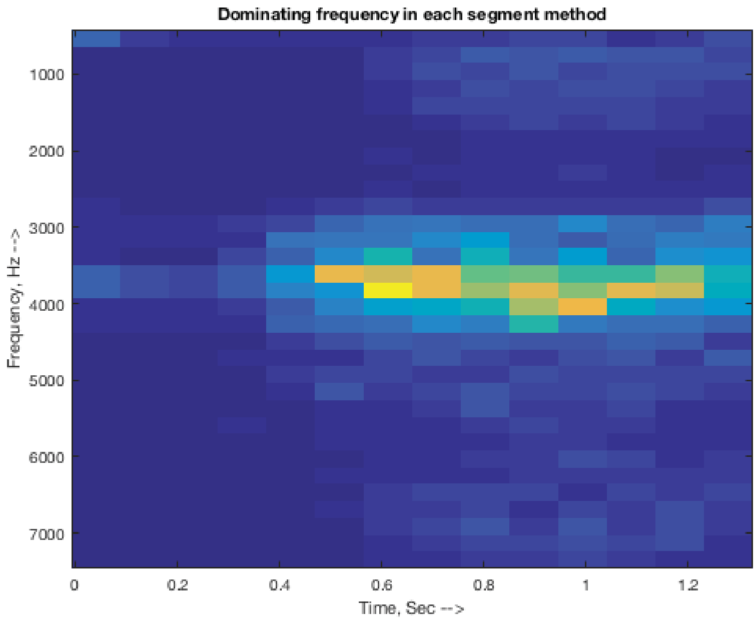

Figure 4.11: Dominant frequency features of Owl in Spectrogram

Dominance frequency features of one phrase of a Owl audio data of one phrase is plotted in spectrogram and shown in Fig. 4.11

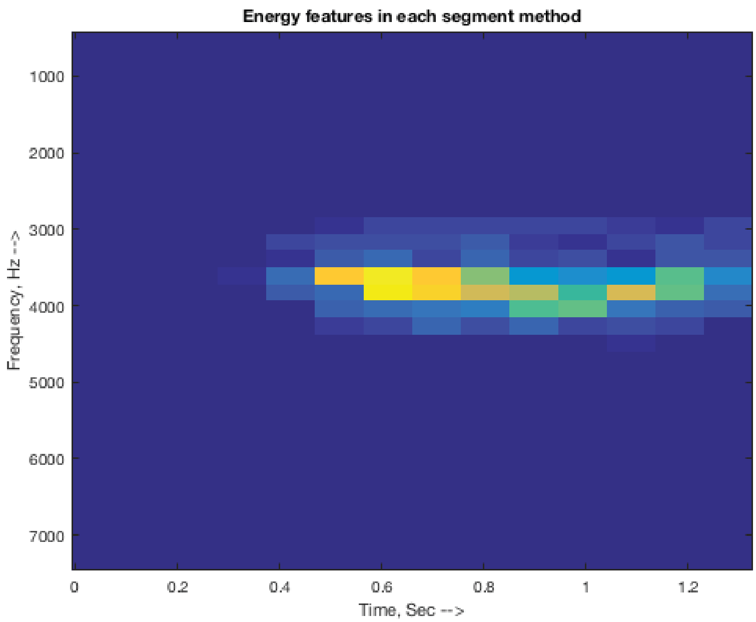

Figure 4.12: Energy features of Owl in Spectrogram

Energy features of Owl audio data of one phrase is plotted in spectrogram and shown in Fig. 4.12.

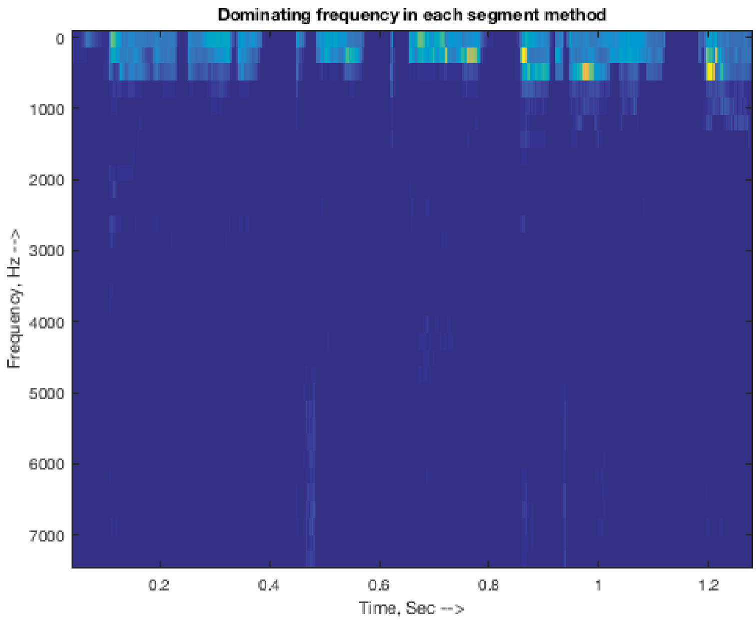

Figure 4.13: Dominant frequency features of Song in Spectrogram

Dominance frequency features of one phrase of a Song audio data of one phrase is plotted in spectrogram and shown in Fig. 4.13

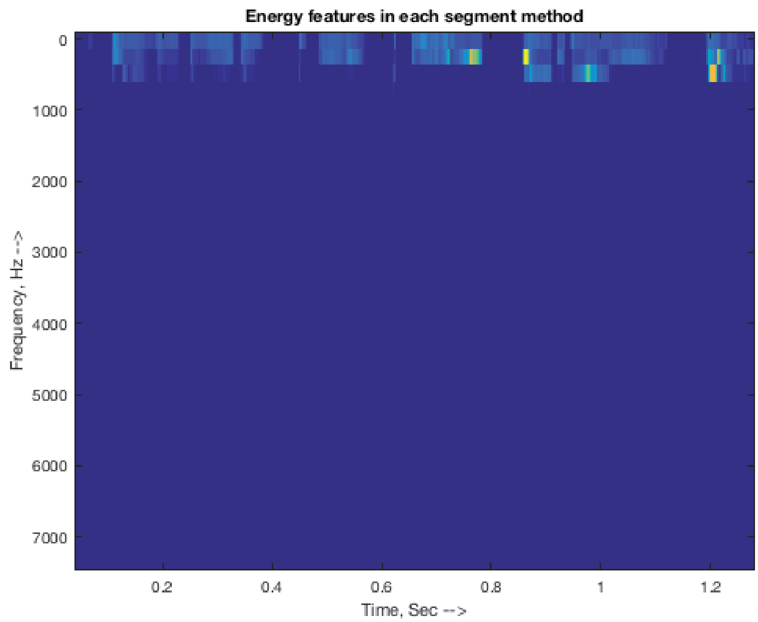

Figure 4.14: Energy features of Song in Spectrogram

Energy features of Song audio data of one phrase is plotted in spectrogram and shown in Fig. 4.14.

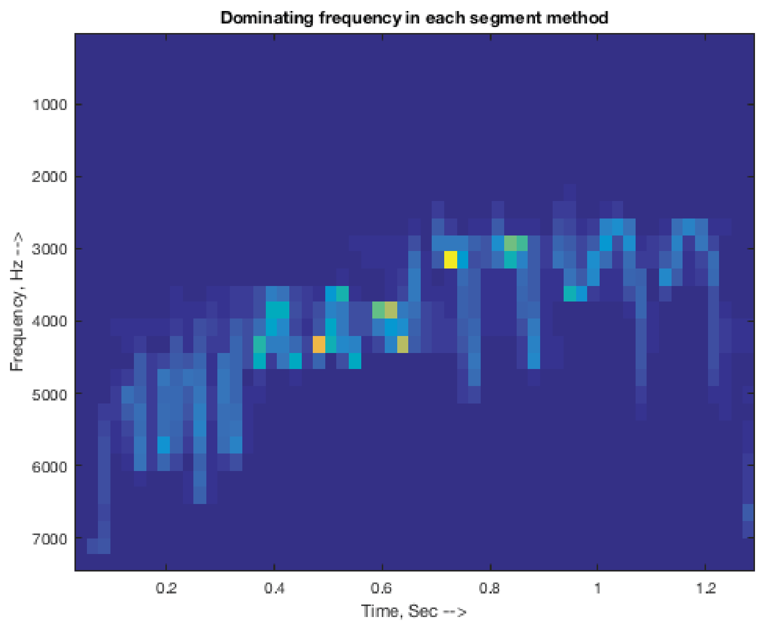

Figure 4.15: Dominant frequency features of Warbler in Spectrogram

Dominance frequency features of one phrase of a Warbler audio data of one phrase is plotted in spectrogram and shown in Fig. 4.15

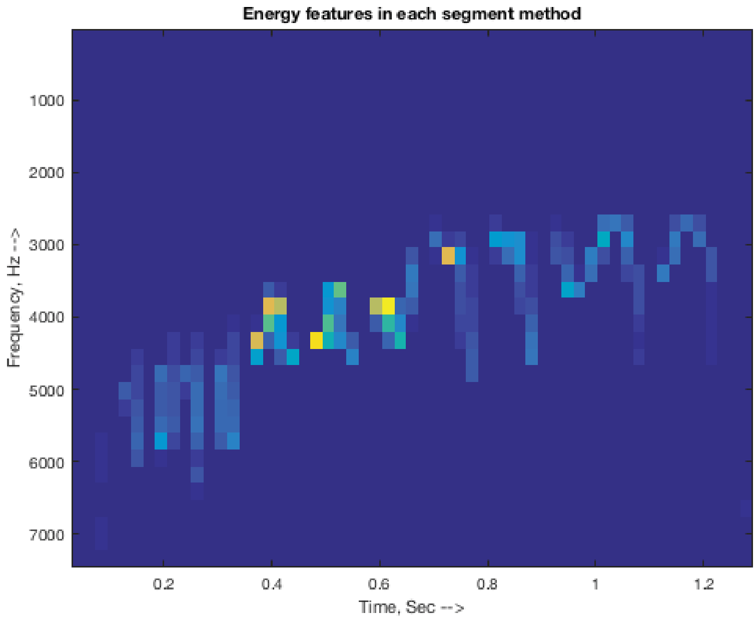

Figure 4.16: Energy features of Warbler in Spectrogram

Energy features of Warbler audio data of one phrase is plotted in spectrogram and shown in Fig. 4.16.

4.4 Auto-correlation Classification Method

Chaffinch, Texas Owl, Willow Warbler, Human Whistle and Human lyrical song audio data sources Signature matrices are created. As described in section 3.1.5 a database of signature matrix is created.

Test sources signature matrices are classified with signature matrices of sound sources in database using auto-correlation classification method as described in the section 3.2.1 . The threshold values for maximum correlation, minimum correlation and sum of maximum and minimum values are calculated as described in section 3.2.3.1.

The test sources are classified with database sources and detection process is implemented as described in the section 3.2.3.2 and every test sources hit and miss chances were labeled in Fig. 4.17 with sum of mean and correlation values.

In the same way, hit and miss chances of mean correlation values are labeled in Fig. 4.19 and hit and miss chances of maximum correlation values are labeled in Fig. 4.18

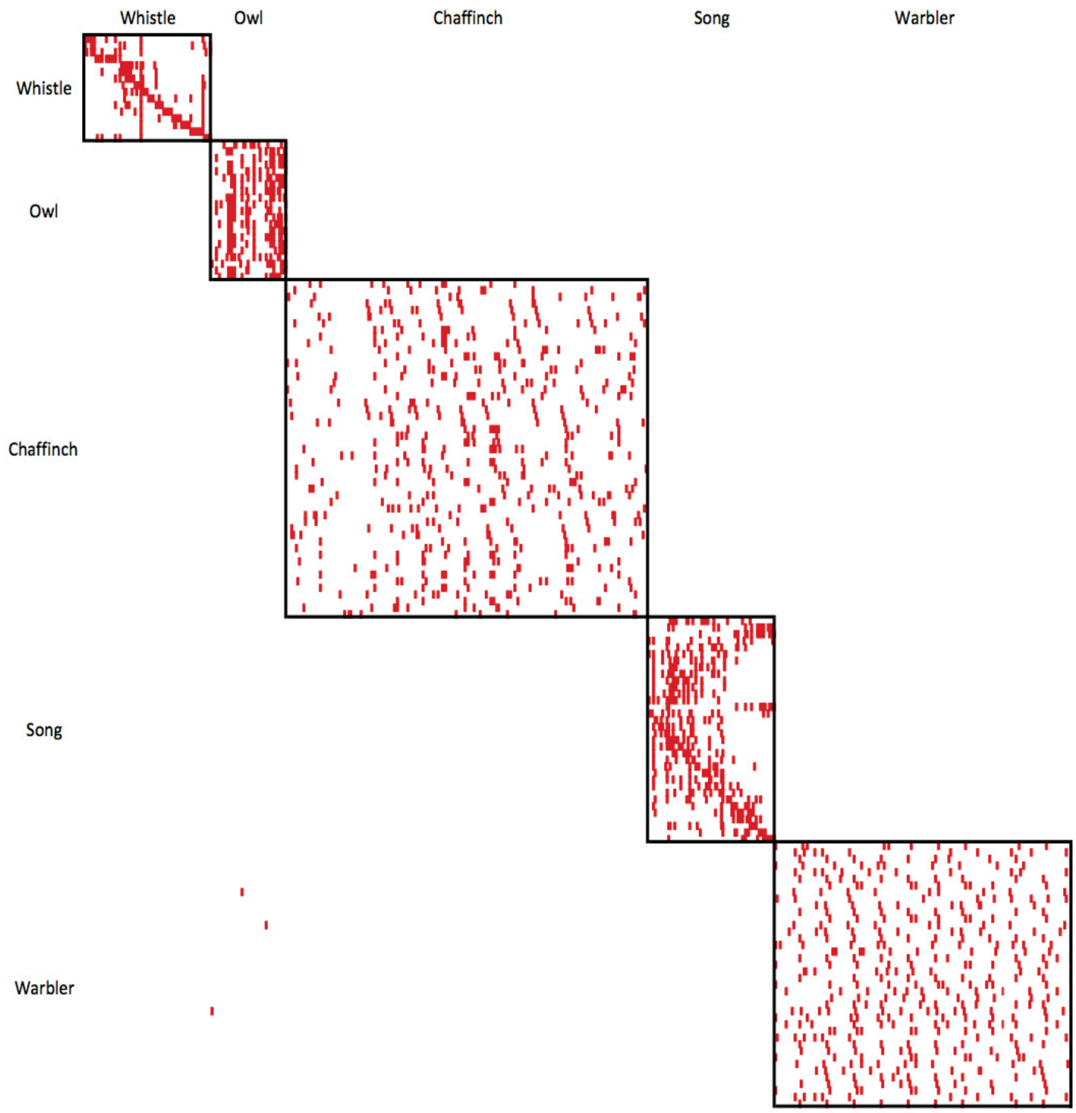

Figure 4.17: Representation of hit and miss values between database data and test database for sum of mean and maximum correlation values

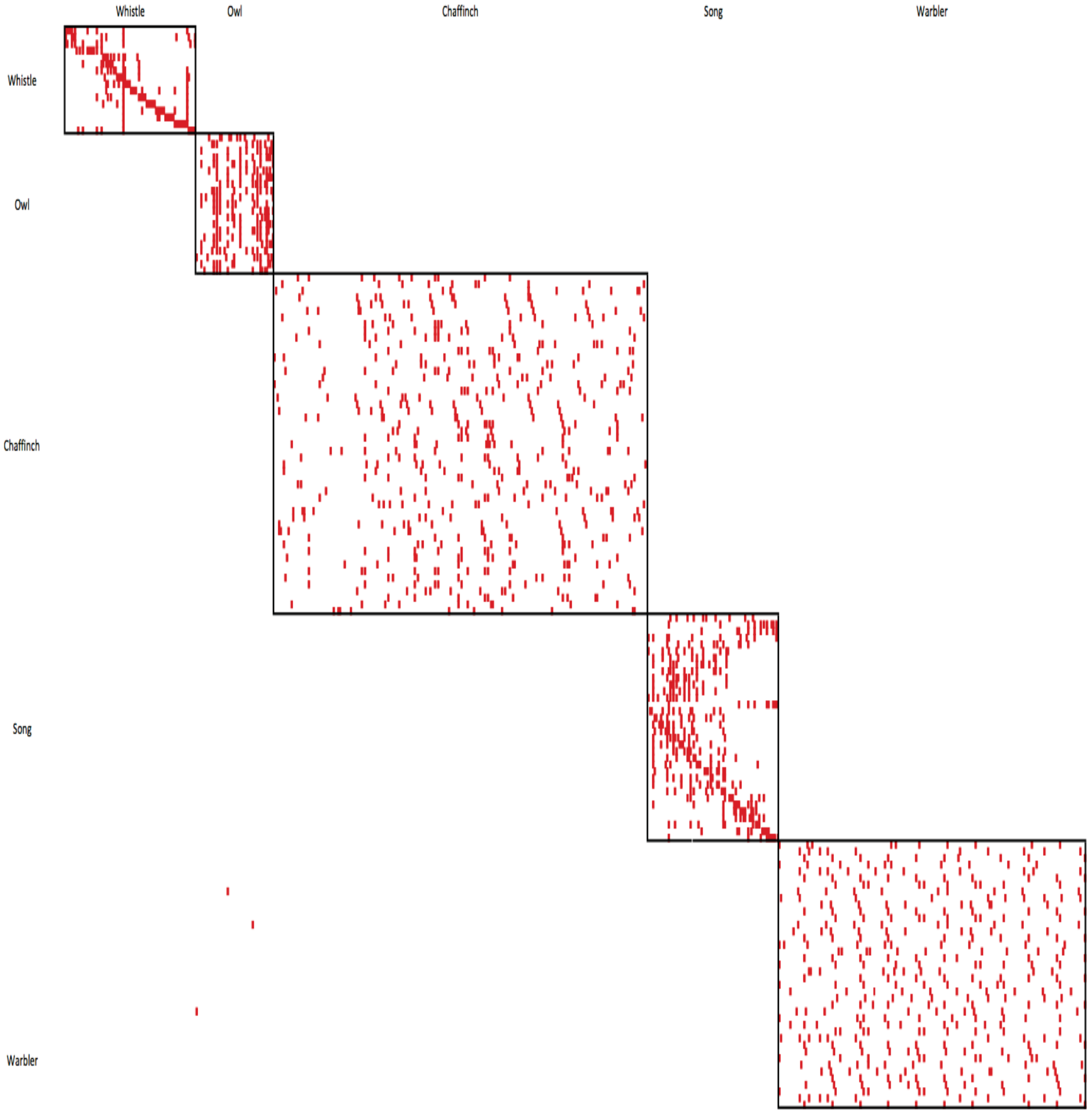

Figure 4.18: Representation of hit and miss values between database data and test database for maximum correlation values

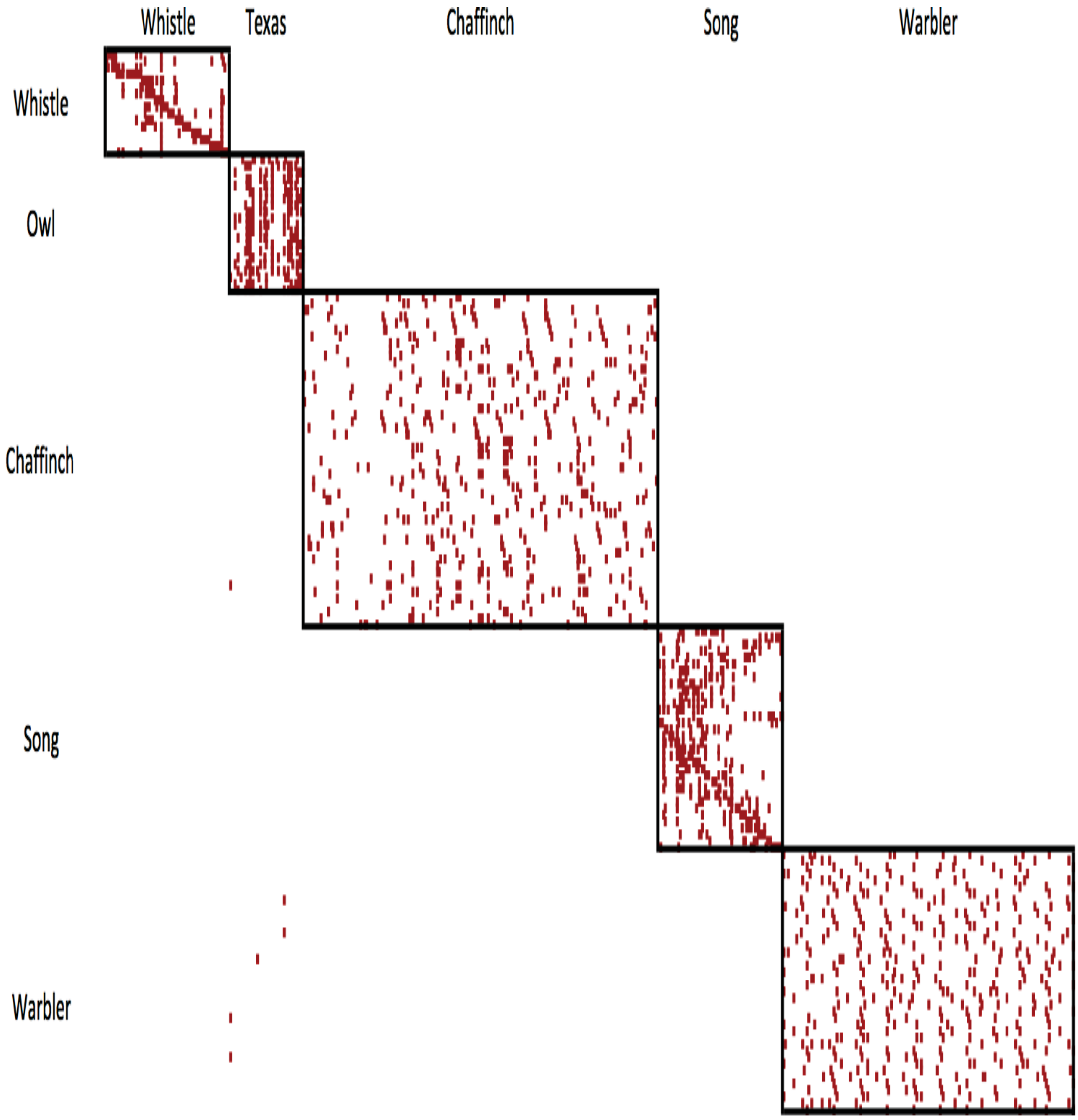

Figure 4.19: Representation of hit and miss values between database data and test database for mean correlation values

4.5 Reference Difference Classification Method

The Test sources signature matrices are classified with database signature matrices using reference classification method as described in section 3.2.2. The hit and miss chances for sum of mean and minimum difference values are labeled and shown in Fig. 4.20 and for mean difference values are labeled and shown in Fig. 4.21 and for minimum difference values are labeled and shown in Fig. 4.22

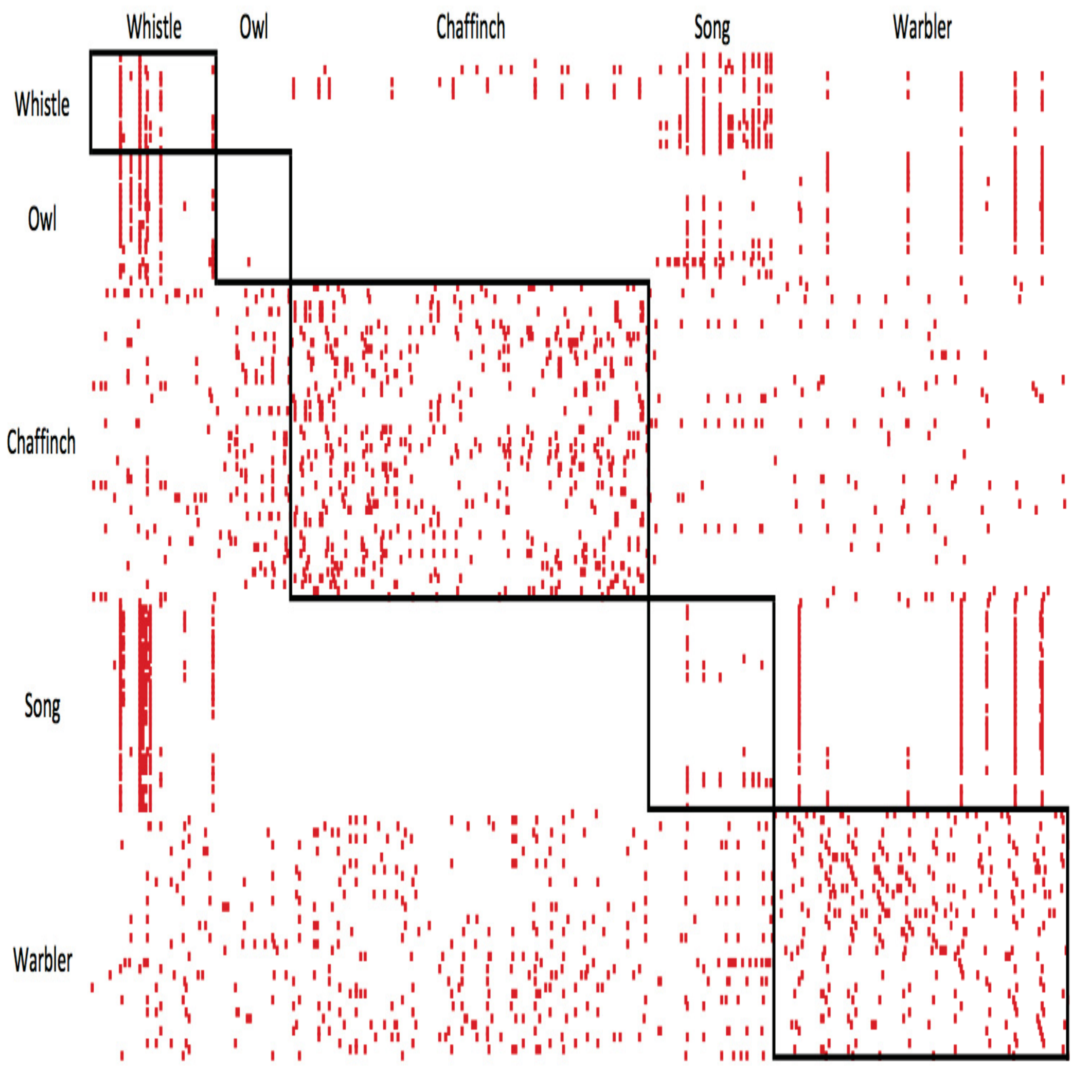

Figure 4.20: Representation of hit and miss values between database data and test database for sum of mean and minimum difference values

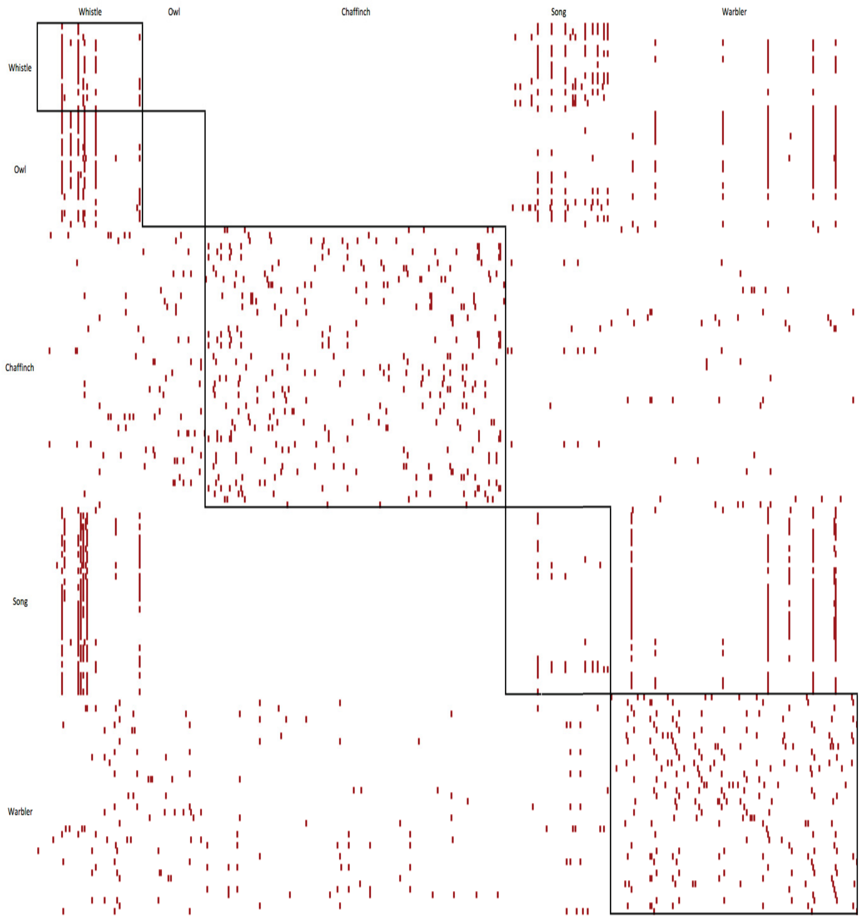

Figure 4.21: Representation of hit and miss values between database data and test database for mean difference values

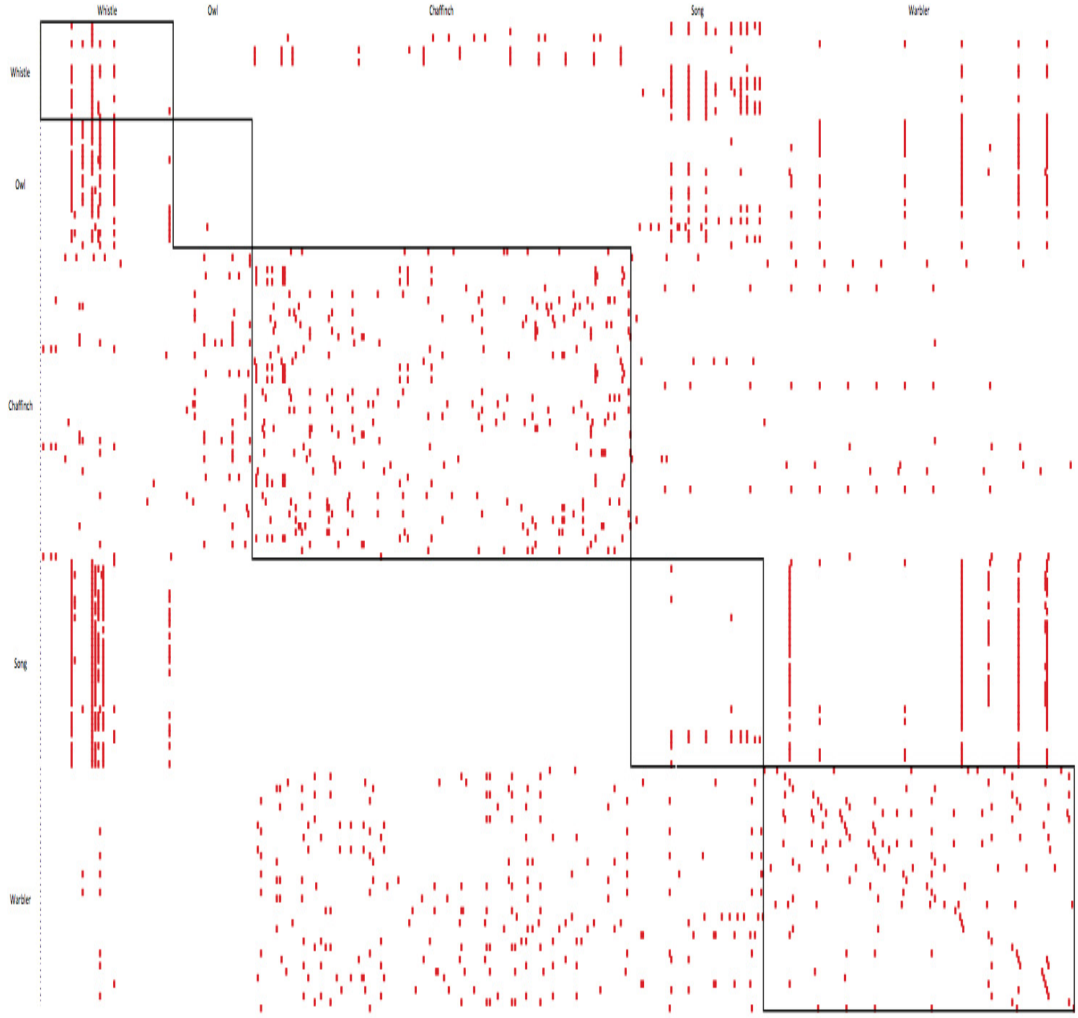

Figure 4.22: Representation of hit and miss values between database data and test database for mean difference values

4.6 Hit Rate

The correlation algorithm's efficiency can be improved by changing the threshold values. The considered threshold values for all the bird species is as defined in table 4.1. Though the observed threshold values that are obtained with test files are larger than the actual set threshold because implementation of this project in real-time gives lesser correlation values.

	Sum of Correlations	Max. Correlation	Mean Correlation
Whistle	70%	70%	65%
Owl	75%	75%	75%
Chaffinch	70%	70%	65%
Song	70%	70%	65%
Warbler	70%	70%	65%

Table 4.1: Minimum similarity between test and database required to identify the signal

The Hit Rate is calculated for the top ten correlation values obtained with all test files. The top ten hits for the sum of correlation values, the maximum correlation values and the mean correlation values were considered and for all the test signals. The number of rightly identified correlation values among them are considered and hit rate is calculated as shown in the table 4.2.

	Sum of Correlations	Max. Correlation	Mean Correlation
Whistle	100%	100%	100%
Owl	100%	100%	100%
Chaffinch	99.90%	100%	99.80%
Song	100%	100%	100%
Warbler	99.25%	98.5%	98.5%

Table 4.2: Hit rate of correlation classification method

The Hit Rate is calculated for the top ten reference difference values obtained with all test files which are recorded in the same environmental conditions as in database signals. The top ten hits for the sum of reference difference values, the minimum reference difference values and the mean reference difference values were considered and for all the test signals. The number of rightly identified correlation values among them are considered and hit rate is calculated as shown in the table 4.3.

	Sum of Differences	Min. Difference	Mean Difference
Whistle	47.19%	21.88%	32.50%
Owl	0.48%	0.48%	0%
Chaffinch	69.22%	67.84%	69.02%
Song	48.38%	8.53%	48%
Warbler	48%	36.50%	58.75%

Table 4.3: Hit rate of reference difference classification method

Chapter 5

Conclusion

The main focus of this research is to detect the species of a sound source. The more concentrated feature values results in better classification. From section 4.3, dominant frequency features are slightly efficient than energy features for 5 sound sources. To detect sources higher than five, the detection accuracy decreases drastically with dominant features than with energy features.

From section 4.1 and section 4.2, the number of samples used for detection process decreased. The processing speed for the detection increased with phrase extraction followed by decimation. Through this process, this implementation can be used in devices with small computational capacities.

From section 4.4 and section 4.5, the hit ratio of auto-correlation classification method is more than 95% and the hit ratio of reference difference classification is less than 50%. The auto-correlation classification method detects sound sources better than reference difference classification method.

From section 4.4 and section 4.5, the hit ratio of Texas owl which is a single syllable sound source has less detection accuracy than Chaffinch, Willow Warbler, Human whistle and lyrical song which are poly syllable sound sources because the correlation of small single syllable phrase with poly syllable phrase is falsely higher hence results in false detection.

Chapter 6

Future Work

1. Implementing this research with more than five bird species.

   In this research, we are able to differentiate between three bird species. It can be further extended to detect more number of bird species.

2. Developing an android application.

   By developing an android application, this research can be made available to may users.

3. Animal species detection

   Not only for birds, with further analysis and thorough examination about the sounds of animals we can extract features of animal's species sounds and can be classified.

4. Identify the birds emotions

   Birds can predict the nature calamities way before than our electronic technologies. If one can find the particular emotions of a bird that tells about the natural calamities, there will be more time to face the challenges.

5. By further extraction and classification, one can find the differences between the same bird species.

6. By comparing the physical behavior and the extracted values, there is a chance to find the language of birds.

References

[1] M. Hansson-Sandsten, M. Tarka, J. Caissy-Martineau, B. Hansson, and D. Hasselquist, "Svd-based classification of bird singing in different time-frequency domains using multitapers," in *Signal Processing Conference, 2011 19th European*, pp. 966–970, IEEE, 2011.

[2] F. Briggs, B. Lakshminarayanan, L. Neal, X. Z. Fern, R. Raich, S. J. Hadley, A. S. Hadley, and M. G. Betts, "Acoustic classification of multiple simultaneous bird species: A multi-instance multi-label approach," *The Journal of the Acoustical Society of America*, vol. 131, no. 6, pp. 4640–4650, 2012.

[3] T. L. Evangelista, T. M. Priolli, C. N. Silla Jr, B. A. Angélico, and C. A. Kaestner, "Automatic segmentation of audio signals for bird species identification," in *Multimedia (ISM), 2014 IEEE International Symposium on*, pp. 223–228, IEEE, 2014.

[4] R. Bardeli, D. Wolff, F. Kurth, M. Koch, K.-H. Tauchert, and K.-H. Frommolt, "Detecting bird sounds in a complex acoustic environment and application to bioacoustic monitoring," *Pattern Recognition Letters*, vol. 31, no. 12, pp. 1524–1534, 2010.

[5] E. Vilches, I. A. Escobar, E. E. Vallejo, and C. E. Taylor, "Data mining applied to acoustic bird species recognition," in *18th International Conference on Pattern Recognition (ICPR'06)*, vol. 3, pp. 400–403, IEEE, 2006.

[6] M. Hansson-Sandsten, "Classification of bird song syllables using singular vectors of the multitaper spectrogram," in *Signal Processing Conference (EUSIPCO), 2015 23rd European*, pp. 554–558, IEEE, 2015.

[7] I. Potamitis, S. Ntalampiras, O. Jahn, and K. Riede, "Automatic bird sound detection in long real-field recordings: Applications and tools," *Applied Acoustics*, vol. 80, pp. 1–9, 2014.

[8] X. Dong, J. Xie, M. Towsey, J. Zhang, and P. Roe, "Generalised features for bird vocalisation retrieval in acoustic recordings," in *Multimedia Signal Processing (MMSP), 2015 IEEE 17th International Workshop on*, pp. 1–6, IEEE, 2015.

[9] K. Qian, Z. Zhang, F. Ringeval, *et al.*, "Bird sounds classification by large scale acoustic features and extreme learning machine," in *2015 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pp. 1317–1321, IEEE, 2015.

[10] J. Starkhammar and M. Hansson-Sandsten, "Evaluation of seven time-frequency representation algorithms applied to broadband echolocation signals," *Advances in Acoustics and Vibration*, vol. 2015, 2015.

[11] P. Somervuo, A. Harma, and S. Fagerlund, "Parametric representations of bird sounds for automatic species recognition," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 6, pp. 2252–2263, 2006.

[12] B. Ghoraani and S. Krishnan, "Time–frequency matrix feature extraction and classification of environmental audio signals," *IEEE transactions on audio, speech, and language processing*, vol. 19, no. 7, pp. 2197–2209, 2011.

[13] J. Swärd, J. Brynolfsson, A. Jakobsson, and M. Hansson-Sandsten, "Sparse semi-parametric chirp estimation," in *2014 48th Asilomar Conference on Signals, Systems and Computers*, pp. 1236–1240, IEEE, 2014.

[14] "Xeno-canto, sharing bird sounds from around the world." `http://www.xeno-canto.org/`. Accessed September 1, 2016.

[15] C.-H. Chou, P.-H. Liu, and B. Cai, "On the studies of syllable segmentation and improving mfccs for automatic birdsong recognition," in *Asia-Pacific Services Computing Conference, 2008. APSCC'08. IEEE*, pp. 745–750, IEEE, 2008.

Appendix A

Appendices

The Fig. A.1 and Fig. A.2 represents the dominant frequencies for all time segments for Chaffing in frequency band 6 and frequency band 11 respectively. the signature matrix is calculated by first calculating the dominant frequency in each time segment.

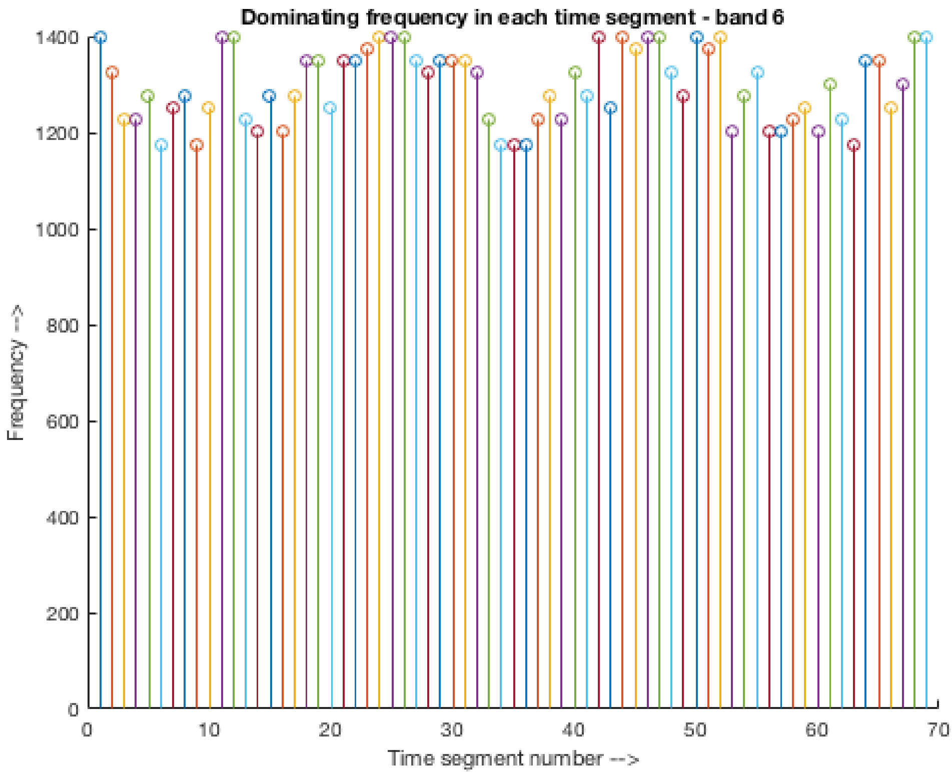

Figure A.1: Sliding window technique for frequency band 6

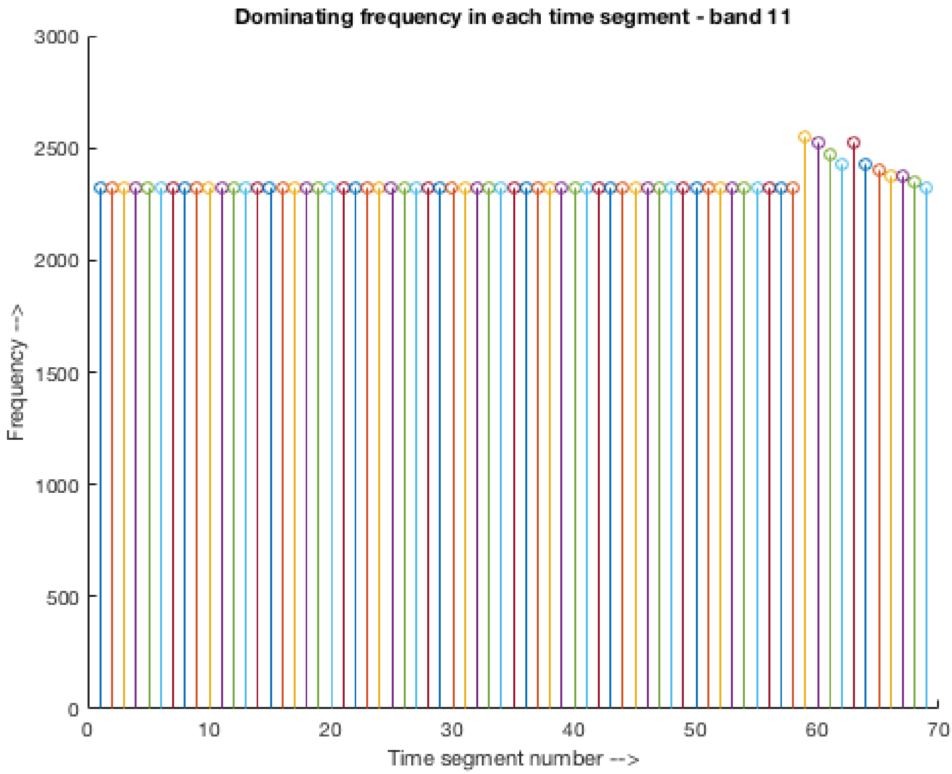

Figure A.2: Sliding window technique for frequency band 11

The Fig. A.3, A.4, A.5, A.6, A.7, A.8, A.9, A.10, A.11 and A.12 represents the frequency plots of human whistle, owl, chaffinch, human song and willow warbler, before and after decimation respectively.

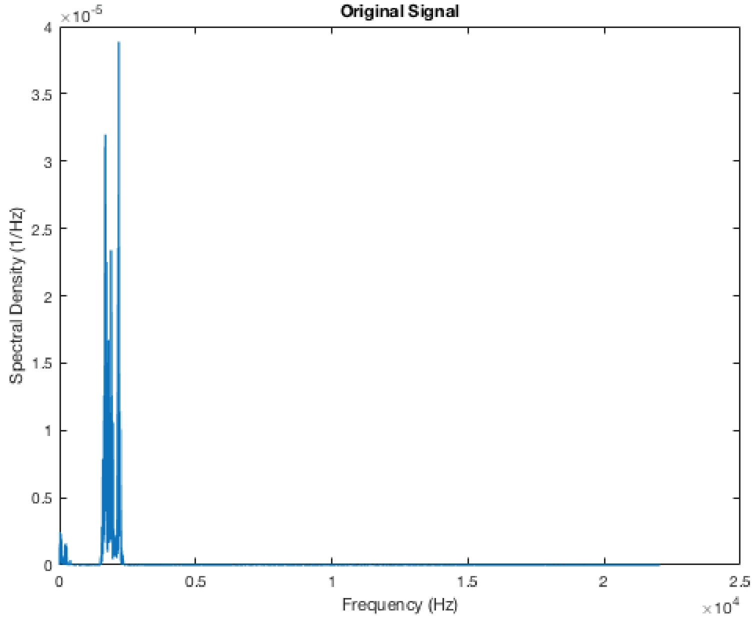

Figure A.3: Frequency plot of human whistle before decimation

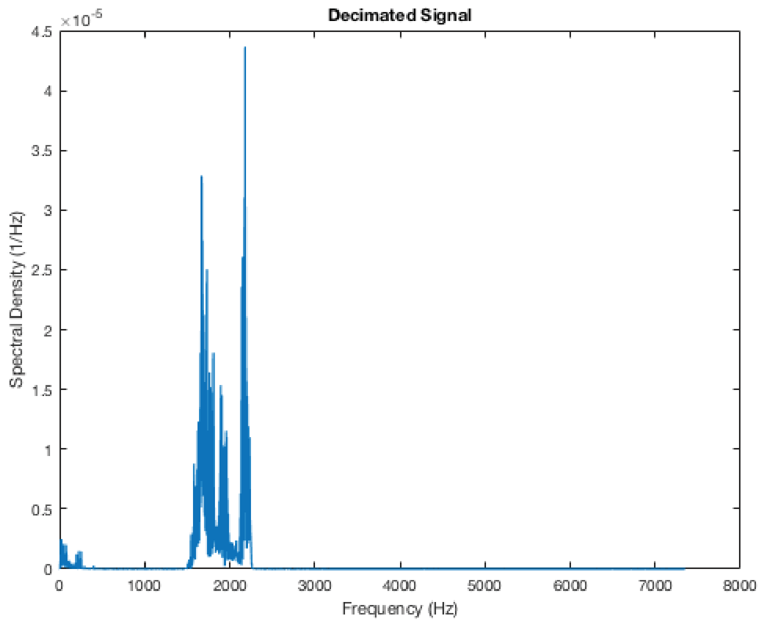

Figure A.4: Frequency plot of human whistle after decimation

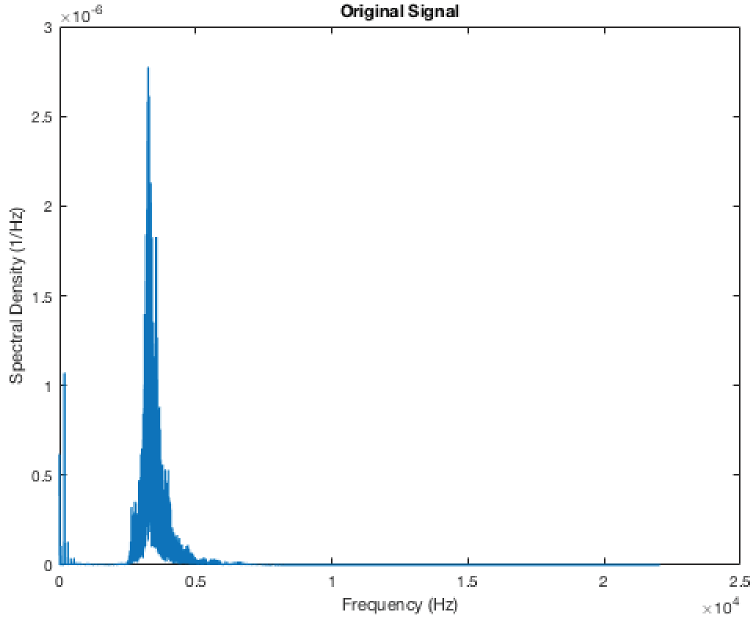

Figure A.5: Frequency plot of owl before decimation

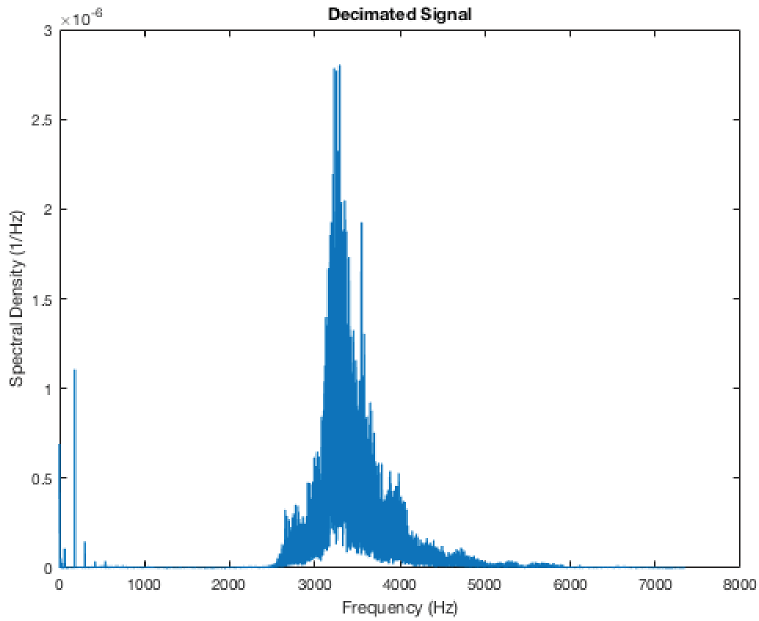

Figure A.6: Frequency plot of owl after decimation

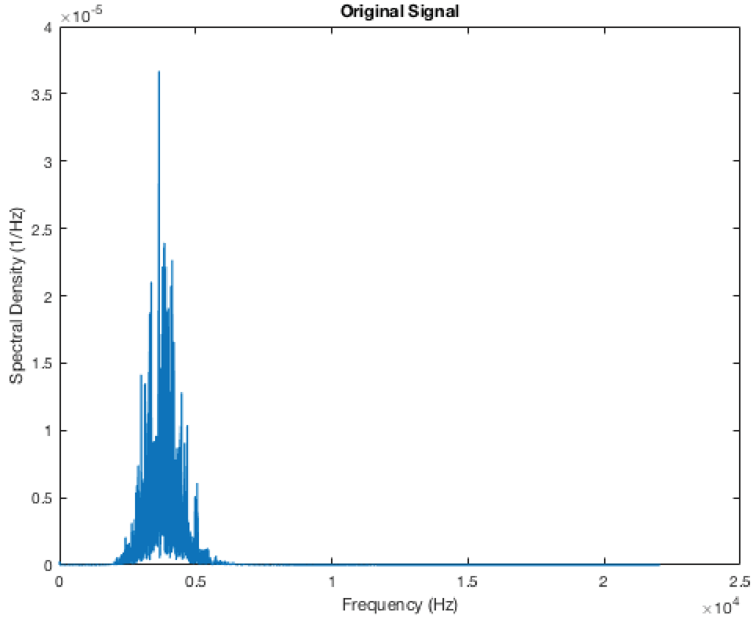

Figure A.7: Frequency plot of chaffinch before decimation

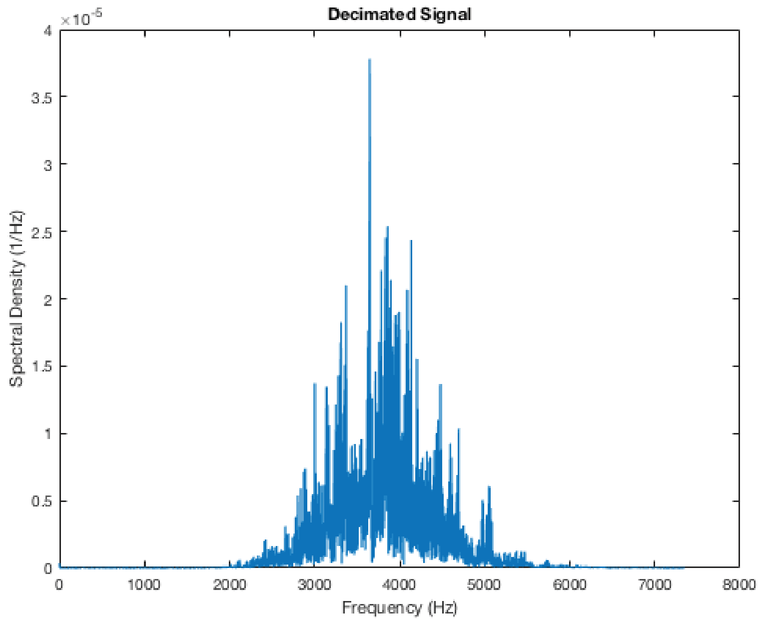

Figure A.8: Frequency plot of chaffinch after decimation

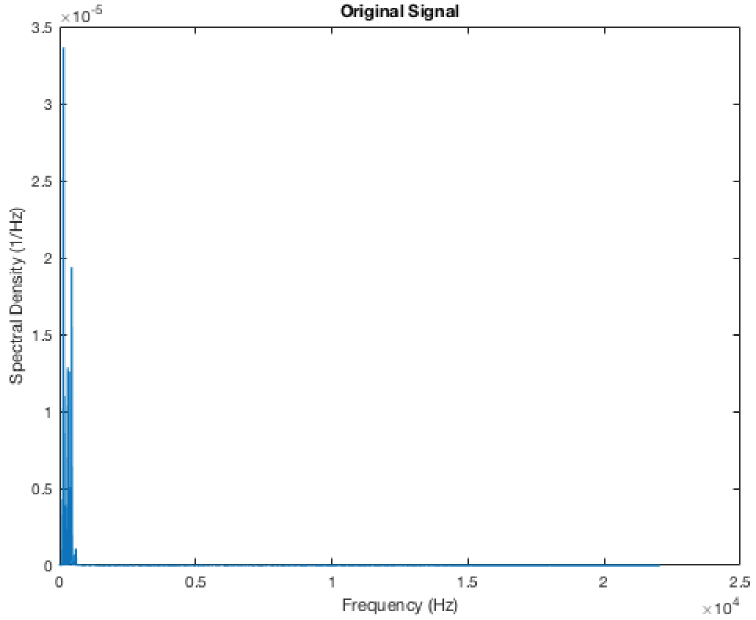

Figure A.9: Frequency plot of human song before decimation

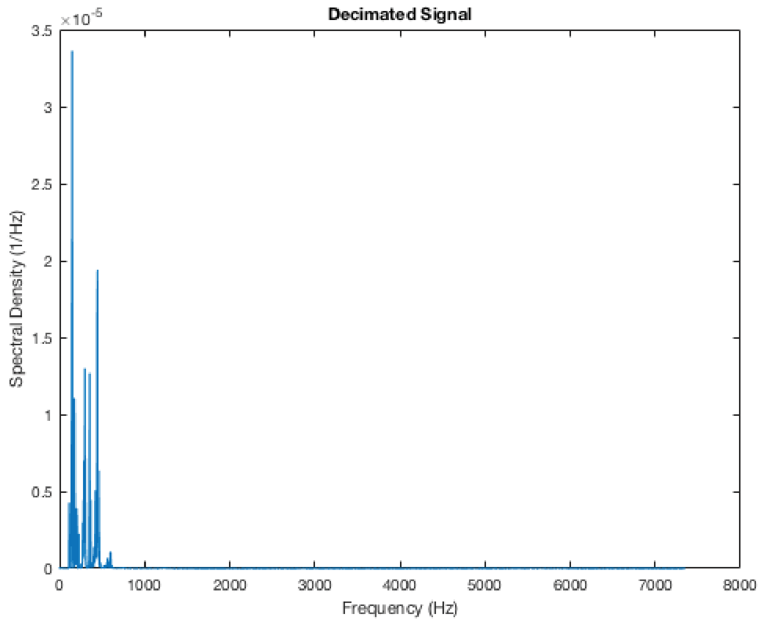

Figure A.10: Frequency plot of human song after decimation

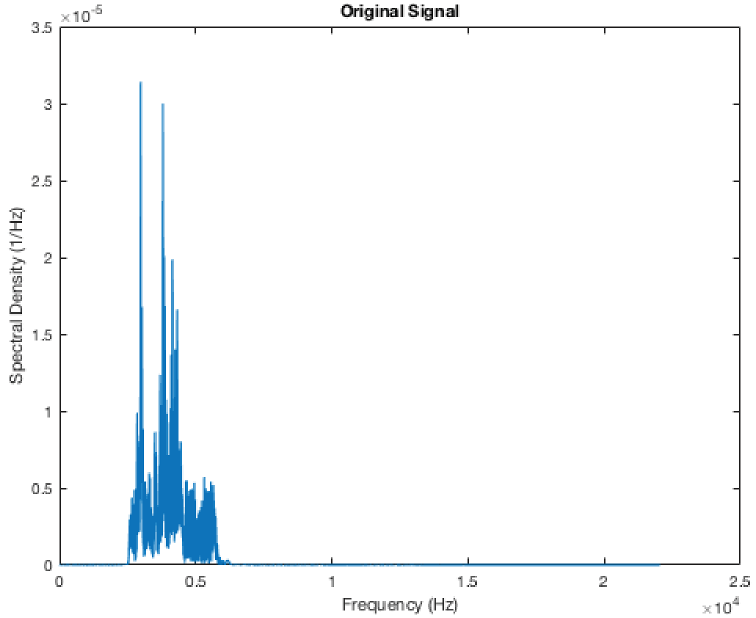

Figure A.11: Frequency plot of Warbler before decimation

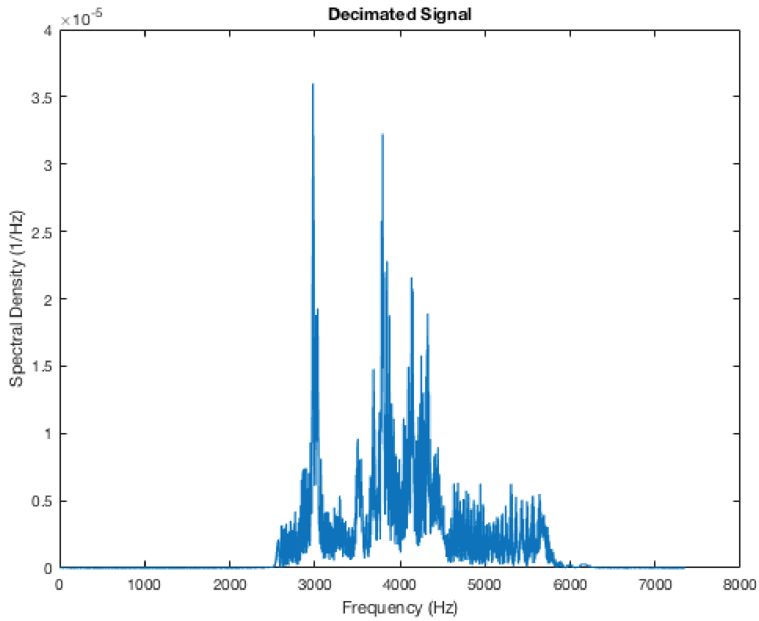

Figure A.12: Frequency plot of Warbler after decimation