Acoustic Classification of Bird Species

Rashmika Patole and Priti Rege

Abstract The main focus of this paper is identification of bird species or even individual birds on the basis of their sounds. This work compares an audio signal of an unknown bird to a database of known birds. The system has two modes of operation: training mode, and recognition mode. In the training mode, ate a feature model of the available bird sounds in the database is created. The recognition mode will use the information obtained from the training mode to isolate and identify the bird. Mel frequency Cepstral Coefficients and Gammatone frequency Cepstral Coefficients have been employed as feature sets for classification. The classification accuracies are evaluated using Support Vector Machine and Artificial Neural Networks.

Keywords Bird identification · Acoustic classification

1 Introduction

Birds have always been of great interest to people since ages because of their social as well as ecological importance. Bird monitoring has always been of great importance because of many practical reasons. In the context of ecological concern, birds play an important role, since they are one of the classes of living beings that have direct contact with humans. Reasons such as changes in habitat, nest egg loss, mortality during migration human and animal predators, etc. have caused decline in the population of bird species over the last few years [1]. It can be possible to correct the population decline and reduce future risk of extinction by understanding the connection between the bird vocalizations and their behavior patterns. The identification of birds can also aid in the monitoring of migration and population of birds in the ecosystem.

There are numerous engineering applications where identification of birds in real-time is necessary. One such application is used in aircraft monitoring systems where they need to avoid collision between birds and the aircraft. There may be birds in

R. Patole (✉) · P. Rege
College of Engineering, Pune, India
e-mail: rkp.extc@coep.ac.in

the neighbourhood of wind turbine generators which may need to be tracked. Also, identification of birds is necessary to understand their seasonal migratory patterns and behaviour, especially at night and when the weather conditions are unfavourable. To study the impact of human development on plants and animals, ornithologists have to identify and count the number of birds in particular site. To identify birds in a particular area, it is easy to rely on their sounds because they are often easier to locate a particular bird by hearing its sound instead of seeing it physically. Hence, it is advantageous to rely on the bird sounds to identify bird species in a particular area. Thus, ornithologists must study the bird sounds and identify the birds in an area by sound alone. To monitor the bird sounds in real time can be a difficult task. Therefore, it can be useful to record unknown sounds so that they can be identified later. Thus, there is a need for automated methods for bird species identification to monitor and also to evaluate the diversity and quantity of birds [2].

Classifying bird species has been a research topic since many years. Different feature sets and classification algorithms for the task of bird classification have been discussed in the literature. In [3], spectrograms are used to represent the bird sound recordings and Dynamic Time Warping (DTW) has been used to measure the difference between the spectrograms [4] uses different feature sets such as Linear Predictive Coding (LPC) coefficients, LPC-derived cepstral coefficients, LPC reflection coefficients, Mel-Frequency cepstral coefficients (MFCCs), log mel-filter bank channel, and linear mel-filter bank channel. DTW and Hidden Markov Model (HMM) are used to form the acoustic models and classify the bird sounds. Neural networks and multivariate statistics have been employed in [5] to identify the bird species [6] gives an overview of previous works in the area of bird classification from vocalizations. A recent study includes recognition of bird species based on a hybrid model including HMM and Deep Neural Networks [7].

In [8], the author uses a decision tree along with support vector machine (SVM) for classification. Some prior work is concerned with classification of bird species from individual syllables [9], while other work is also concerned with identifying species from songs composed of sequences of syllables [8, 10]. The algorithms that have been applied to classifying syllables include nearest-neighbour and distance-based classifiers [8, 11, 12], multi-layer perceptrons [13], and support vector machines [9].

This paper is organized into the following sections: Sect. 2 discusses the sound mechanism in birds. The computation of Mel-Frequency Cepstral Coefficients, implementation of Vector Quantization and K-means algorithm is given in Sect. 3. Section 4 provides the results of the experiment followed by conclusion in Sect. 5.

2 Elementary Concepts and Organization of Bird Sounds

The mechanism through which sound is produced in birds is very similar to the human sound production mechanism. In humans, the vocal chord are responsible for the production of sound. A similar organ is present in birds, which is called Syrinx.
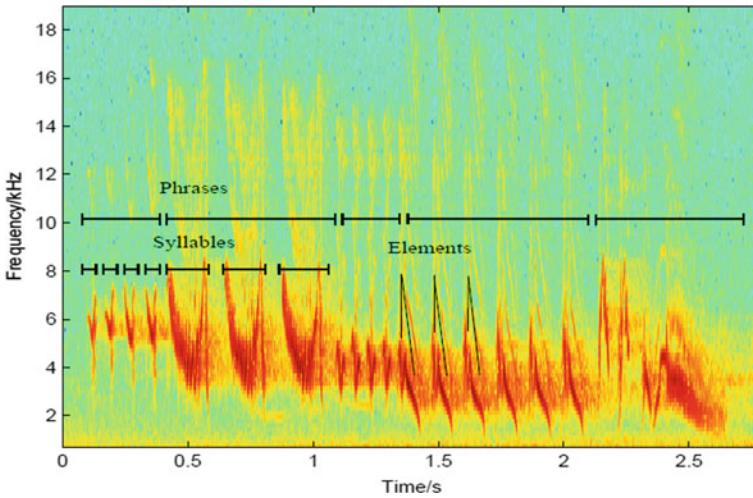
Fig. 1 Hierarchical levels of bird song

Bird sounds can be divided into either songs or calls, which can be further divided into phrases, syllables, and elements or notes as shown in Fig. 1. Similar to human speech, bird sounds can also be divided into voiced sounds and unvoiced sounds. Voiced sounds in birds are similar to the human vowel sounds in structure as well as the way they are produced. Sounds that do not contain any harmonics, e.g. pure tonal or whistled sounds can also be produced by birds. Such sounds are closely related to the unvoiced sounds in human speech. Bird sounds can be also noisy, broadband, or chaotic in structure [14]. Figure 2 shows examples of bird songs and calls from different bird species.

3 System Overview

The problem of Bird Species Classification is similar to other audio or speech classification problems like classification of general audio/speech content, auditory scene recognition, music genre classification, speech recognition, etc. that have been extensively studied during the last few decades. This project involves two modules namely (1) Training module (*feature extraction)* and (2) Testing module (*feature matching)* and classification.

The feature extraction step aims to extract acoustic features from the audio signal waveform. This module converts the audio signal waveform of the bird to some type of parametric representation for further analysis and processing. Feature extraction is about reducing the dimensionality of the input-vector but the discriminating power of the signal is maintained. These features carry the characteristics of the bird sound which are unique to a specific bird. Similar to the human speech signal, the audio
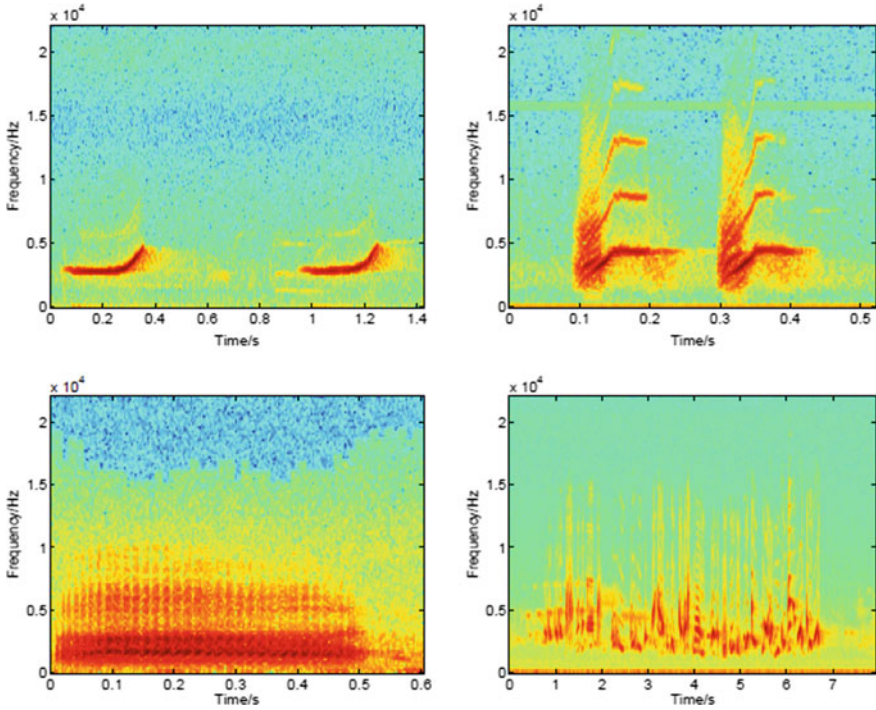
Fig. 2 Bird sounds from Willow Warbler (*Phylloscopus trochilus*) (upper row), Common Chaffinch (*Fringilla coelebs*), Hooded Crow (*Corvus corone cornix*) (second row)

signal of birds is a slowly varying signal. This can be seen in Fig. 3. It is not stationary. Therefore, the signal processing techniques which are commonly used cannot be applied to our signal because of its non-stationary nature. If the audio signal is analyzed over a short period of around 5–50 ms, the characteristics of the signal remain fairly stationary. Therefore, short-time analysis is needed to analyze the audio signal.

Mel-frequency Cepstral Coefficients (MFCCs) and Gammatone filter Cepstral coefficients (GFCCs) are used as features. The pre-processing done and the filter banks for extracting MFCCs and GFCCs have been described below.

3.1 Pre-processing and Filter Banks

The audio recordings of bird sounds available are first framed into short intervals of 25 ms size. The frames have an overlap of 50% and are windowed using a Hamming window. Short Time Fourier Transform (STFT) converts the frames into frequency domain. STFT uses 1024 FFT points. Two filter banks are used in this work, the
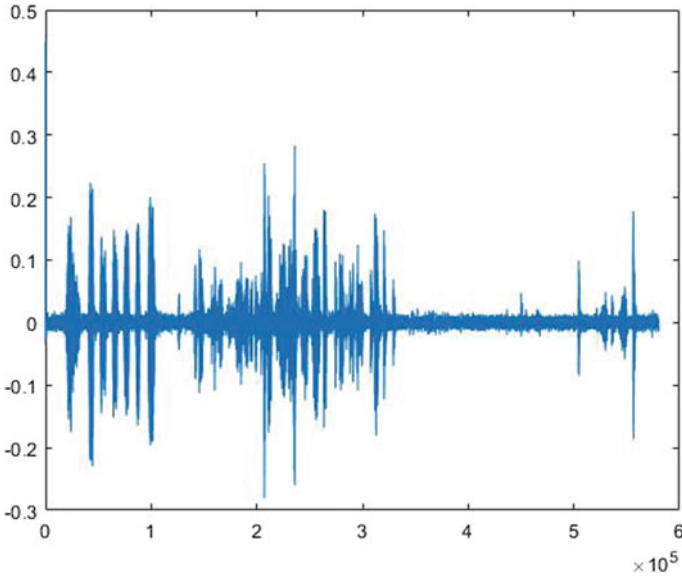
Fig. 3 Audio signal of Greater Racket-tailed Drongo

Mel-bank for obtaining MFCCs and the Gammatone filter bank for obtaining the GFCCs. 32 filters have been used in the Mel-bank. The linearly spaced frequencies are converted to the Mel frequencies, using the formula in Eq. (1).

$$\text{mel}(f) = 2595\log_{10}\left(1 + \frac{f}{700}\right) \tag{1}$$

The first filter is narrower while the filters become broader with increasing frequency and they are triangular in shape.

The Gammatone (GT) filter bank [14] is a biologically inspired bank with ERB (equivalent rectangular bandwidth) especially for effective representation of spectral properties at lower frequencies. The authors have used GT filter bank for another application as given in [15]. The magnitude response of the GT bank is similar to the ReOx function which closely models the human auditory system. Gammatone filter bank has its impulse response similar to Gamma distribution function. 64 filters are used with an ERB scale ranging from frequencies $\frac{fs}{2}$ to 100 Hz. ERB scale used in this paper is calculated using the Glasberg and Moore parameters of EarQ $= 9.26449$, minimum B.W. $= 24.7$ and order $= 1$. GT filter of fourth-order is implemented using four cascaded filters of order one.

Table 1 Comparison of Classification accuracies for MFCCs and GFCCs

Features/Classifiers	SVM (%)	ANN (%)
MFCCs	91.7	95.6
GFCCs	89.4	90.2
MFCCs + GFCCs	93.5	97.6

3.2 Classification

Support Vector Machine (SVM) and Artificial Neural Network (ANN) are employed for the classification of bird sounds. SVM is a supervised learning model that classifies the data points by finding the best hyperplane to separate the data points of one class from the other. In this paper, SVM is used for multi-class classification. An artificial neural network consists of input layer, hidden layers, and output layer. The hidden layer nodes firing are dependent on the activation function. Sigmoid hidden neurons and softmax output neurons are used to serve the purpose. The algorithm used to train the network is scaled conjugate gradient back-propagation.

4 Dataset and Experimental Results

Our data set consists of bird sounds taken from the Xeno-Canto [16] dataset which consists of bird sounds from all over the world. The data set consists of sounds from 70 different bird classes and 10 recordings from each class. The duration of each audio recording is approximately 3–4 s.

The feature sets consist of MFCCs, GFCCs and MFCC + GFCC features. Table 1 shows the classification accuracies using the above feature sets and classifiers. MFCC features give a good accuracy for SVM as well as ANN. GFCC features when employed alone give accuracy values less than MFCC features. It can be observed that highest classification accuracy of 97.6% is given by ANN when MFCC and GFCC features are used in combination.

5 Conclusion

This paper discusses a methodology for bird species classification based on its sound. In this paper, two sets of classifiers and feature sets have been employed for the classification of bird species. ANN outperforms SVM and gives the highest accuracy with both the feature sets. The accuracy can be further improved by exploiting more feature sets and classifiers. Also, the future work will look into scaling up the database by including more number of bird sounds. Furthermore, the recordings available were free from noise. Real bird recordings will definitely include environment noise. The performance of our system will be assessed in the presence of different noises.

References

1. Dale S (2001) Causes of population decline in Ortolan Bunting in Norway. In: Proceedings in 3rd international Ortolan symposium, pp 33–41 (2001)
2. Brandes TS (2008) Automated sound recording and analysis techniques for bird surveys and conservation. Bird Conserv Int 18:163–173
3. Anderson SE, Dave AS, Margoliash D (1996) Template-based automatic recognition of birdsong syllables from continuous recordings. J Acoust Soc Am 100(2):1209–1219
4. Kogan J, Margoliash D (1998a) Automated recognition of bird song elements from continuous recordings using dynamic time warping and hidden Markov models: a comparative study. J Acoust Soc Amer 103(4):2187–2196
5. McIlraith AL, Card HC (1997a) Birdsong recognition using backpropagation and multivariate statistics. IEEE Trans Signal Process 45(11):2740–2748
6. Rassak S, Nachamai M (2016) Survey study on the methods of bird vocalization classification. In: 2016 IEEE international conference on current trends in advanced computing (ICCTAC). IEEE, pp 1–8
7. Jancovic P, Köküer M (2019) Bird species recognition using unsupervised modeling of individual vocalization elements. IEEE/ACM Trans Audio Speech Lang Process 27(5):932–947
8. Somervuo P, Harma A, Fagerlund S (2006) Parametric representations of bird sounds for automatic species recognition. IEEE Trans Audio Speech Lang Process 14(6):2252–2263
9. Fagerlund S (2007) Bird species recognition using support vector machines. EURASIP J Adv Signal Proces
10. Kogan JA, Margoliash D (1998b) Automated recognition of bird song elements from continuous recordings using dynamic time warping and hidden markov models: A comparative study. J Acoust Soc Am 103(4):2185–2196
11. Lee C-H, Lee Y-K, Huang R-Z (2006) Automatic recognition of bird songs using cepstral coefficients. J Inf Technol Appl 1(1):17–23
12. Chen Z, Maher RC (2006) Semi-automatic classification of bird vocalizations using spectral peak tracks. J Acoust Soc Am 120(5 Pt 1):2974–2984
13. Fletcher NH (2000) A class of chaotic bird calls. J Acoust Soc Am 108(2):821–826
14. Valero X, Alias F (2012) Gammatone cepstral coefficients: Biologically inspired features for non-speech audio classification. IEEE Trans Multimedia 14(6):1684–1689
15. Patil R, Patole R, Rege P (2019) Audio environment identification. In: 2019 10th International conference on computing, communication and networking technologies (ICCCNT). IEEE
16. https://www.xeno-canto.org/
17. Härmä A (2003) Automatic identification of bird species based on sinusoidal modeling of syllables. In: IEEE international conference on acoustics, speech, and signal processing, 2003. Proceedings (ICASSP'03), vol 5, pp V–545–548
18. Härmä A, Somervuo P Classification of the harmonic structure in bird vocalization. In: IEEE international conference on acoustics, speech, and signal processing, 2004. Proceedings. (ICASSP'04), vol 5, pp V–701–704
19. Calvo de Lara JR (2005) A method of automatic speaker recognition using cepstral features and vectorial quantization. In: 10th Iberoamerican conference on pattern recognition, CIARP 2005 proceedings
20. Furui S (1994) An overview of speaker recognition technology. In: ESCA workshop on automatic speaker recognition identification and verification, pp 1–9
21. Song FK, Rosenberg AE, Juang BH (1987) A vector quantisation approach to speaker recognition. AT&T Tech J 66–2:14–26