

Classification of Bird Species based on Bioacoustics

Arti V. Bang

Department of Electronics and Telecommunication
Vishwakarma Institute of Information Technology
University of Pune
Pune, India
arti.bang@viit.ac.in

Priti P. Rege

Department of Electronics and Telecommunication
College of Engineering Pune
Pune, India
ppr.extc@coep.ac.in

Abstract— This paper presents a study of automatic detection and recognition of bird species based on bioacoustics. The energy distribution in different frequency bands varies quite significantly among different birds' sounds. Energy in various frequency bands is extracted using wavelet packet transform and k-nearest neighbor algorithm is used for classification. In our experiment, classification accuracy of 91% is achieved for the classification of eight bird species. A total of 307 bird sound files are used to train/test the system.

Keywords— Bioacoustics, Syllables, Wavelet Packet Transform, k-nearest neighbor.

I. Introduction

In order to understand and evaluate the changes in environment, it is necessary to continuously obtain reliable information about the population of animals. Monitoring bird population is necessary to study decline in species population, since they are indicators of good ecosystem.

Existing methods of monitoring bird species through listening is by manual surveys, which are extremely laborious, and require observers trained in bird recognition. Recent developments in bioacoustics, computer science and pattern recognition & classification are providing new tools to meet this challenge. Automated bird population surveys could provide vast amounts of useful data of species while requiring less effort and expense than human surveys. Yet, there are many challenges to develop a system capable of monitoring birds indirectly by their songs.

Bird vocalization is usually considered to be composed of calls and songs [1]. The single distinct utterance by a bird is called as syllable, and serves as the basic building block of bird song. A song consists of a series of syllables arranged in a particular pattern. In this study our goal is to classify bird species from an interval of sound, containing one or more syllables, which correspond to song level organization.

Technical analysis of bird sound has a long history. A majority studies on bird recognition has been, based on visual inspection of spectrograms. For a large database, this is time consuming and dependent on observers.

II. Related Work

Pattern recognition techniques have been used in previous studies for sound classification. Different parametric representations for audio sounds have been proposed by various researchers. Few studies have been done on automatic recognition of bird species.

Anderson *et al.* [2] derived log magnitudes of fast Fourier transform and used dynamic time warping (DTW) for automated analysis of continuous recordings of bird song. Lee, Han and Chuang [3] derived two dimensional cepstral coefficients and modeled the same using vector quantization and Gaussian mixture models. They used nearest neighbor classifier to classify the bird species. In [4], comparison of various feature sets with different classifiers has been evaluated for classification of three bird species. Briggs *et al.* [5] devised algorithms to devise probabilistic model for audio features to classify six bird species. McIlraith and Card [6] had used Neural networks and statistical methods for classification of six bird species. Härmä [7] proposed a method for automatic identification of bird species based on sinusoidal modeling of syllables. Härmä and Somervuo [8] proposed a method to classify bird syllables into four classes according to their harmonic structure. Acevedo, Bravo *et al.* [9] classified calls of eight frogs and three birds using support vector machines and linear discriminant analysis. Calls were characterized by four variables.

Bird sound classification and recognition is essentially a pattern recognition problem. In this work, we have used wavelet packet transform to extract energy from different subbands, since birds' sound exhibit varying energy across band of frequencies.

III. Proposed Work

In this work, we used bird sound recordings from various internet websites [10], [11] & [12]. Proposed method is based on classification of bird species from an interval of sound, containing one or more syllables. Bird sounds range from being tonal whistles, to harmonic sounds, to inharmonic burst,

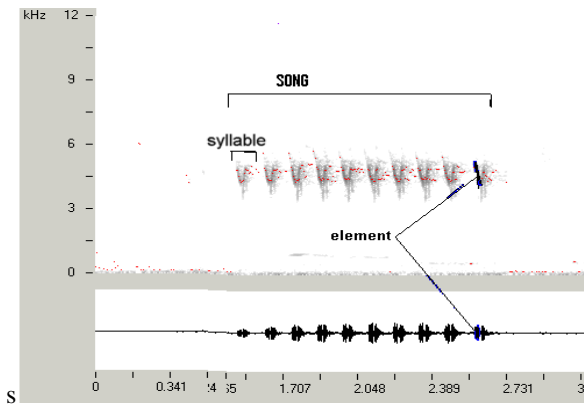


Figure.1: Spectrogram showing hierarchical levels of Bird Song.

or even being noise like. Typically the duration of a syllable ranges from few to few hundred milliseconds.

A mix of tonal, harmonic and inharmonic eight birds is selected. The bird sound files of eight species were divided so that about 65% of the files were used for training the classifier and 35% for testing the classifier. Duration of files varied from 1 sec. to 10 sec.

A. Preprocessing

The recordings obtained are in different formats with different sampling rates and were usually recorded in a noisy environment. The recordings were standardized to wave file format, sampling rate of 16 kHz with 16-bit resolution and a monotone type PCM format.

The bird sound frequency, in this database, ranges from 450 Hz (Common Wood Pigeon) to 8 KHz (Canada Goose). The recordings are first high pass filtered with a 300Hz cutoff in order to remove the low frequency background noise attributable to wind or mechanical sounds.

Sound files were recorded under unavoidable noisy environment like blowing wind sound, flowing water sound, insect sound, etc. This interferes with the bird sound frequency band and is removed using spectral subtraction algorithm [13] as shown in fig.2.

Table 1 contains list of species with their scientific names and abbreviations used in this paper. The silence between the syllables is removed using segmentation of syllables by computing the energy of overlapping frames in audio data. If the energy of the frame is below a threshold, the frame is discarded.

B. Wavelet Packet Decomposition

The energy distribution in different frequency bands varies quite significantly among different birds' sounds. Hence, the entire frequency range can be divided into subbands and

energy of the subbands can be used as a feature for classification. The wavelet analysis has gained a great deal of significance in the field of digital signal processing. The wavelet packet decomposition (WPD) method is a generalization of wavelet decomposition that offers a richer range of possibilities for signal analysis. In wavelet packet decomposition, [14] the approximations as well as the details can be split. WPD transforms the signal into wavelet coefficients which are usually large in number. Using all the wavelet coefficients as features will often lead to inaccurate results. Hence extraction of features is essential.

In the WPD the signal x is split into approximation (A) and detail (D) parts. N^{th} level WPD decomposes signal into 2^N number of frequency bands as against wavelet analysis that decomposes signal into $N+1$ frequency bands. The WPD, with a decomposition level of 3, is illustrated in fig.3, where the WPD tree is shown in an increasing frequency order from left to right. Out of the several wavelet families available, the Daubechies wavelet family dbN [15] is selected, because in this, both scaling and wavelet functions are compactly supported and they are orthogonal. The dB6 wavelet was selected, because the preliminary tests showed that it gave best results with the selected bird sounds. The preliminary tests showed that the best decomposition level was 4 which resulted into 16 subbands. Therefore, with a sampling frequency of 16KHz, each subband is of 500Hz. Fig. 4 shows the spectrograms and wavelet coefficients of eight species used in this study. In the spectrograms, the darker colors represent the higher energies of the sound. Correspondingly, the large absolute coefficient values are presented with darker color in adjacent Heisenberg wavelet coefficient boxes [16].

The energy of each subband is computed as

$$E(i) = \sum_{n=1}^p c^2(n) \quad i = 1, 2, \dots, 16$$

where the subband wavelet coefficients are denoted by c and p is the number of wavelet coefficients in a subband.

The total energy of all the bands (TE) is computed by adding energies of all the subbands. The ratio of energy of each subband to the total energy is computed as

$$r(i) = \frac{E(i)}{TE} \quad i = 1, 2, \dots, 16$$

A 32 length feature vector comprising of energies of 16 bands and their ratio with total energy is derived from all the training files.

C. Classification

The k-nearest neighbor (k-NN) is a simple, non parametric, instance based classifier [17]. The k-NN algorithm is preferred because it is well suited for multimodal classes and simple to

TABLE I Bird Species used in the study

Name of the Bird (Abbreviation)	Scientific Name
Canada Goose (CG)	Branta canadensis
Common Quail (CQ)	Coturnix coturnix
Common Wood Pigeon (CWP)	Columba palumbus
Indian Cuckoo (IC)	Cuculus micropterus
Little Tinamou (LT)	Crypturellus soui
Mallard (Mal)	Anas platyrhynchos
Rose ringed Parakeet (RRP)	Psittacula krameri
Undulated Tinamou (UT)	Crypturellus undulatus

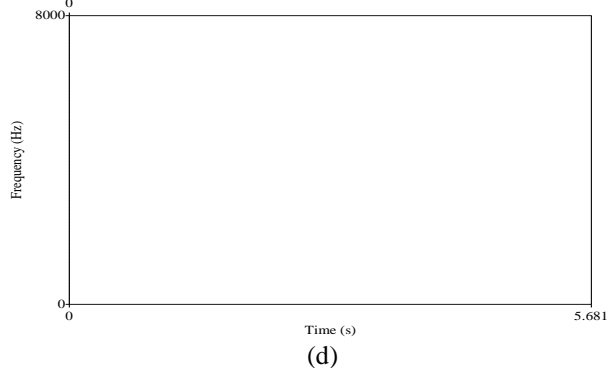
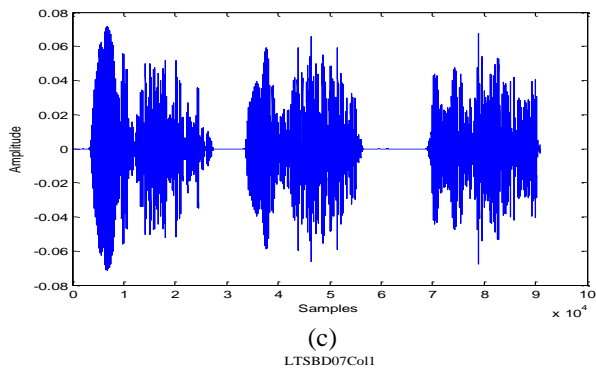
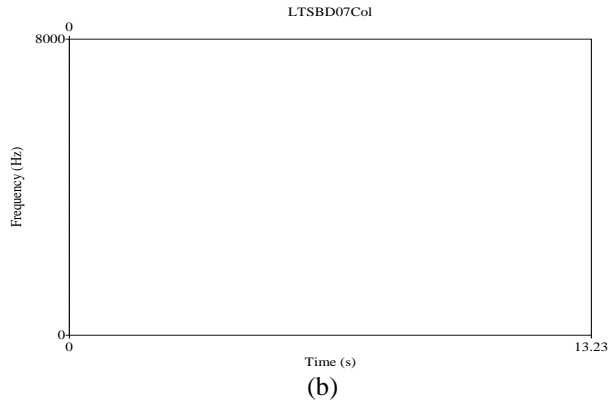
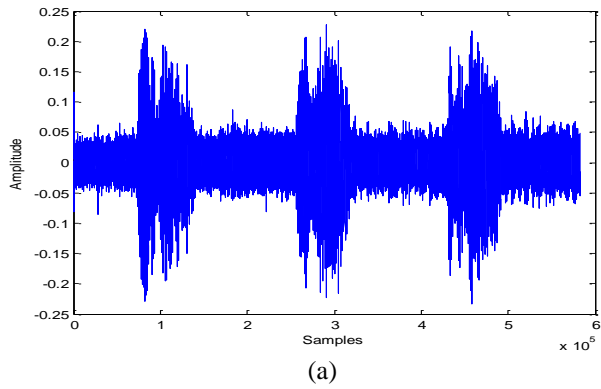


Figure 2: (a) & (b) are time domain plot and spectrograms of bird sound file before noise removal. (c) & (d) are time domain plot and spectrograms of bird sound file after applying spectral subtraction algorithm.

implement. All the training samples are stored in N-dimensional Euclidean space. There is no explicit training phase. The distance between unknown sample y and all the training samples of all classes x_i is computed using Euclidean distance

$$D(y, x_i) = \sqrt{(y - x_i)^T (y - x_i)}$$

Pick k-nearest neighbours to y and then choose the class with most nearest neighbours. The preliminary tests showed that the best results were obtained with $k = 3$.

IV. Results And Discussion

Table 1 contains list of species with their scientific names and abbreviations used in this paper. 65% of the available data was used as training data and 35% as testing data. A total of 307 files were used in this experiment. The data was preprocessed, noise removed and segmented as given in section III. Table II, shows the results in the form of the confusion matrix. The proposed method gave overall classification accuracy (OA) of 91%. The performance of the classification algorithm highly depends on the proper segmentation of syllables and removal of noise.

V. Conclusion And Future Work

In this paper, we have addressed the problem of bird species recognition. The work is based on wavelet packet decomposition. The algorithms were implemented in MATLAB environment and performance of 88% has been achieved.

There is no standard data set available for bird species recognition. Our aim is to collect more recordings and also increase the number of bird species. In order to improve the accuracy, more features will be derived from the wavelet coefficients.

Ss

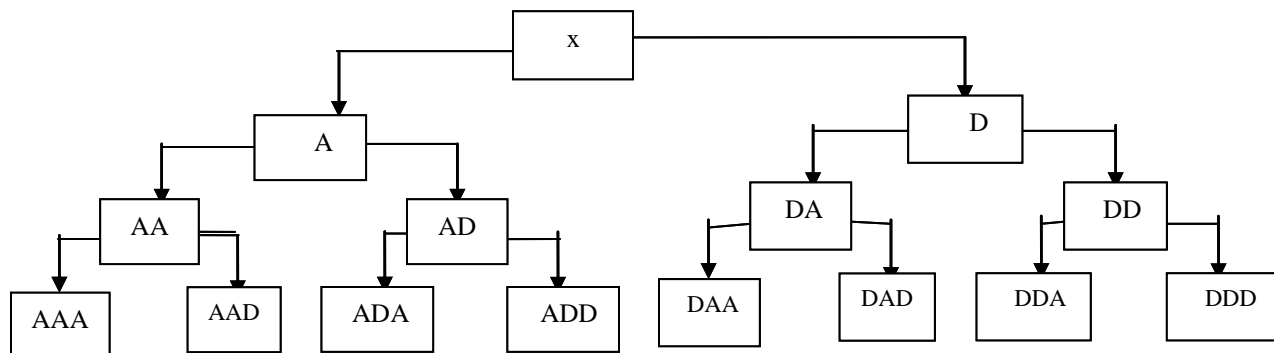


Figure 3: 3 – level Wavelet Packet Decomposition

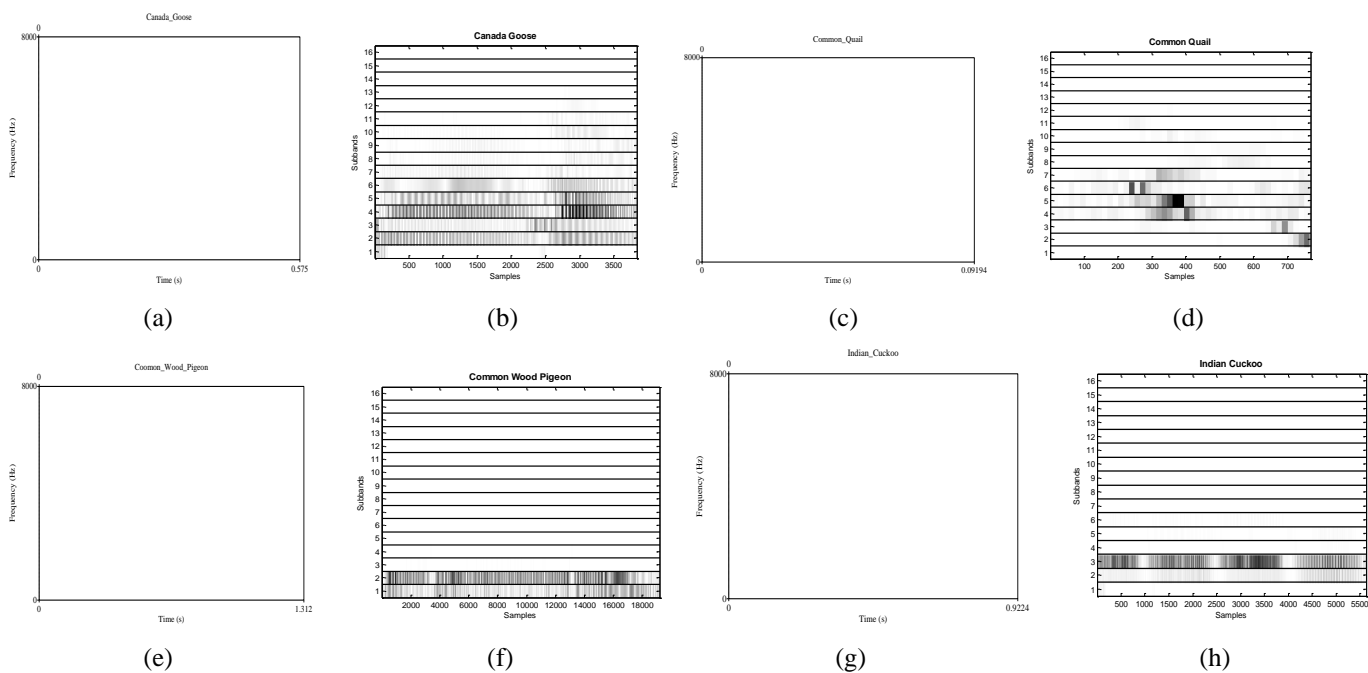


Figure 4 : (a), (c), (e) & (g), are typical spectrograms of few species used in this study. (b), (d), (f) & (h), are the Heisenberg boxes of corresponding wavelet coefficients.

TABLE II : Confusion Matrix

	Number of samples in the class								Acc(%)
	<i>CG</i>	<i>CQ</i>	<i>CWP</i>	<i>IC</i>	<i>LT</i>	<i>Mal</i>	<i>RRP</i>	<i>UT</i>	
<i>CG</i>	14	0	0	0	0	0	0	0	100
<i>CQ</i>	0	13	0	0	0	1	0	0	92.86
<i>CWP</i>	0	0	11	0	0	0	0	0	100
<i>IC</i>	0	0	0	12	0	0	0	4	75
<i>LT</i>	0	0	0	0	15	0	0	0	100
<i>Mal</i>	0	0	0	0	3	13	0	0	81.25
<i>RRP</i>	0	0	0	0	0	0	10	0	100
<i>UT</i>	0	0	0	3	0	0	0	13	81.25
%Rel	100	100	100	80	83.33	92.86	100	76.47	
Overall Accuracy: 91%									

References

- [1] Anil Kumar, “Acoustic Communication in Birds, Differences in Songs and Calls, their Production and Biological Significance”, *Resonance*, pp. 44-54, 2003.
- [2] S. E. Anderson, A. S. Dave, and D. Margoliash, “Template-based Automatic recognition of birdsong syllables from continuous recordings,” *J. Acoust. Soc. Amer.*, vol. 100, no. 2, pp. 1209–1219, Aug. 1996.
- [3] Lee, Han and Chuang, “Automatic Classification of Bird Species from their Sounds using Two-Dimensional Cepstral Coefficients”. *IEEE Transactions on Audio, Speech and Language processing*, vol. 16, no.8, pp. 1541-1550, Nov. 2008
- [4] Lopes, Koerich, Silla Junior and Kaestner, “Feature Set Comparision for Automatic Bird Species Identification”, *IEEE* pp. 965-970, 2011.
- [5] Briggs, Raich and Fern, “Audio Classification of Bird Species: A Statistical Manifold Approach”.
- [6] A. L. McIlraith and H. C. Card, “Bird song identification using artificial neural networks and statistical analysis,” in *Proc. Can. Conf. Elect. Comput. Eng.*, 1997, vol. 1, pp. 63–66.
- [7] A.Härmä, “Automatic identification of bird species based on Sinusoidal modeling of syllables,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2003, vol. 5, pp. 545–548.
- [8] A. Härmä and P. Somervuo, “Classification of the harmonic structure in bird vocalization,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2004, vol. 5, pp. 701–704.
- [9] Acevedo, Bravo et al., “Automated classification of bird and amphibian calls using machine learning: A comparison of methods”, *Ecological Informatics*, vol. 4, pp. 206–214, 2009.
- [10] www.zeno-canto.org
- [11] www.birding2asia.com
- [12] www.naturesongs.com
- [13] S.F. Boll, “Suppression of acoustic noise in speech using spectral subtraction”. *IEEE Transactions Acoustics Speech Signal Processing*, vol. 27, pp. 113:120, 1979.
- [14] Akansu, Haddad, “Multiresolution Signal Decomposition”, 2nd Ed. Academic Press.
- [15] I. Daubechies, “Ten Lectures on Wavelets”, SIAM, Philadelphia, Pa, USA, 1992
- [16] Stephen Mallat, “A Wavelet tour of Signal Processing”. Academic Press, Dec.2008
- [17] Duda, Hart, Stork, “Pattern Classification”, 2nd Ed., Wiley Edition.