International journal of application or innovation in engineering and management

ijaiem.com/Dec 2022/ Volume 11/Issue 2/Article No-1/29-37
ISSN: 2319-4847

Naive Bayes Algorithm for Audio-Based Bird Species Recognition

Meher .k Meherkg7650@gmail.com,Jaipur ,India.

Article Info

Received: 11-09-2022 Revised: 04-10-2022 Accepted: 13-11-2022 Published:09/12/2022

Abstract— Small of the smallest living creature is the part of the whole ecosystem. Monitoring and maintaining this environmental health has become a task of prime importance, since human activity and climate change has placed the greater pressure on biodiversity. This helps in monitoring dangerous situations to planes caused by birds near airports and also in saving endangered species. Powerful audio signal processing techniques makes it possible to introduce automated methods for the detection of such vocalizations. The complete identification procedure essentially involves collection of recorded vocalizations of different species, audio pre-processing to remove the noise and silence, segmentation to select the most representative characterizing elements of the signal called syllable, extraction of different features commonly found in sound classification and speech recognition to obtain relevant characteristics, and decision procedures for identification, where Machine Learning (ML) algorithms areused to train employed classifiers using labelled database of previously known species of interest. Naïve Bayes algorithm is used for classification and hence recognition of bird species. The required class probabilities are found to be as good as 100% and as worst as 97% in some cases.

Keywords – Naïve Bayes algorithm; Syllable Segmentation; Mel- Frequency Cepstral Coefficients; Forward Selection algorithm

INTRODUCTION

Birds can positively impact the raising of crops and live- stock. They scavenge carcasses and recycle nutrients back into the earth. Other services provided by birds to agricultural area include pollination, seed dispersal, fertility, pest control, rodent control, etc. Many nature centers and nonprofit environmental organizations create revenue through taking visitors on bird watching tours. Birds also provide services such as carbon sequestration, waste decomposition, and air purification. Birds are the part of the whole ecosystem and provide early indication of the state of the environment. Monitoring andmaintaining this environmental health has become a task of prime importance, since human activity and climate change has placed the greater pressure on global biodiversity.

Powerful audio signal processing techniques makes it possible to introduce automated methods for the detection of bird vocalizations which will save time of ecologists for similar task. Thesetechniques does not needs re- searchers/ornithologists to actually see or in contact with the bird, and hence has Invasiveness Degree- A [1]. However these Automatic Bird Identification (ABI) tasks presents various difficulties from initial stage itself, since the acoustic sensors used for recording bird sound are subject to extraneous noise such as wind and rain. Also there are varying levels of backgroundnoises, variation in species vocalizations and some overlapping vocalizations. Overlapping vocalizationsproblem can be considerable because birdsong is almost always connected with others (birds answer to each other) [2]. Distance of recording sensor from source of sound to be recorded is also important.

This paper uses database that contains bird songs recorded in specific geographic region in and around India and applies Naïve Bayes Algorithm with Averaged Mel Frequency Cepstral Coefficients (AMFCCs) 1 to 12 and their velocity (δ) and acceleration (δδ) coefficients (making a 36 dimensional feature vector)

International journal of application or innovation in engineering and management

ijaiem.com/Dec 2022/ Volume 11/Issue 2/Article No-1/29-37
ISSN: 2319-4847

to recognize the species in a given recording of a bird song.The rest of this paper is organized as follows. The section II presents related research studies and briefly describes the bird species recognition problem; Section III outlines the database used in the recognition experiments and describes initial signal pre-processing, syllable segmentation, the feature extraction procedures and indicate the classification algorithm; Section IV presents the results obtained in experiments; finally, Section V presents conclusions and indicates future research directions.

BIRD SPECIES RECOGNITION USING BIRD SONG

A. Related works

Research community has been attracted to Automatic Bird Species Recognition task since past few years. Numerous research studies and applications on automated bioacoustics monitoring have been published.

Marcelo T. Lopes et al. [3] carried out experiments in the two databases, one with full audio recordings and other with pulses. They employ the MARSYAS framework for feature extraction, which was already used in several audio applications. Best results using pulses were obtained (with the MLP and SMO classifiers) indicating a fact that pulses better characterize bird vocalization, and their use outperforms the use of the complete sound records. The obtained maximum F-measure values were only 85.7% (with SMO-Pearson) and 89.7% (with SMO-Pearson) for specific and random classes with maximum of 8 species found in Southern Atlantic Coast of South America.

Jason Wimmer et al. [4] used approach of building comparatively few recognizers capable of recognizing generic features such as oscillations, whistles, whips and stacked harmonics without building a universal classifier. MFCC features followed by HMM were found suitable only for high quality single-syllable calls. Algorithms such as Discrete Co- sine Transform to find repeating or oscillating elements of calls within a user specified bandwidth, the Event Pattern Recognition (EPR) technique which modelled a call as a 2D distribution of acoustic events in the spectrogram, the Syntactic Pattern Recognition (SPR) technique modelled a call as a symbol sequence were used. However this is a Human-in- the-loop analysis which provided a hybrid approach in which selection of an appropriate algorithm must be done by an experienced user.

Dorota Kamiska and Artur Gmerek presented fully auto- mated algorithm in [2]. SOM and kNN classifiers, have been chosen and compared in their paper for sounds of 10 bird species downloaded from different internet sources (Disjoint sets with 70% - for training, 30% - for testing). The classification accuracy for different features showed that spectral features are the best for Automatic Bird Recognition (ABR) task. Their best results had mean Classification accuracy of 69.94% with kNN classifier and 52.92% with SOM classifier.

Study of Felix Weninger and Bjorn Schuller [5] has given major contribution in their work to evaluate SVMs, HMMs with different topologies, and recurrent neural networks (RNNs) with Long Short-Term Memory (LSTM) on the HU- ASA database. Both cyclic and left-right types of HMMs outperform static classification by SVM. Additionally, comparison of traditional Cepstral features which are commonly used in sound classification to an enhanced feature set derived from speech emotion recognition was carried out. But no clear picture was established in the comparison of standard Cepstral features with an enhanced feature set containing pitch information. Also there was no clear picture concerning the performance of the LSTM-RNN.

Chang-Hsing Lee et al. [6] used frequency information to extract the syllables exactly. Averaged MFCCs in a syllable were used as features to identify animals from their sounds. Experiments concluded that AMFCC greatly outperforms HMM and ALPC, and alternate training and testing of syllables perform better than progressive training and testing. The average classification accuracy was up to 96.8% and 98.1% for 30 kinds of frog calls and 19 cricket calls, respectively. Only frog and cricket calls in which the syntactical arrangements of the sounds in a call does not change significantly are used in their experiments.

Iosif Mporas et al. [7] evaluated the appropriateness of six classifiers in the bird species recognition task with the help of real-field audio recordings of seven bird species, which are common for the Hymettus Mountain in Attica, Greece. Two temporal and sixteen spectral audio descriptors computed using the open SMILE acoustic parameterization tool were used. For high SNR the boosting algorithm outperformed all the rest classification algorithms, while for low SNR the bagging meta- classifier offered slightly better performance than the boosting algorithm with maximum classification accuracy of 92.89%.

Thiago L. F. Evangelista et al. [8] showed that automatic segmentation of audio into short segments with high amplitude- called *pulses* outperforms the use of the complete audio records and manually segmented records in the species identification task.

R. Bardeli et al.[9] through their study reported about an approach using pattern recognition techniques based on simple event detection and repetition rate estimation of bird song for continuous bird monitoring, specifically for only two endangered target species namely Eurasian bittern and Savis warbler. Poor to satisfactory results are achieved with this approach for only two specific species.

Panu Somervuo, Aki Harma, and Seppo Fagerlund [10] segmented a recording into individual syllables using an iterative time-domain algorithm and then parameterized each segmented region using three models such as Sinusoidal Model, Mel-Cepstrum Model, and Descriptive Parameters. Dynamic time warping (DTW) algorithm was used for comparing variable length sequences.

Gaussian mixtures were used for modelling probability density functions in pattern recognition. The average recognition accuracy for single syllable was only around 40% to 50% depending on the method. The recognition results improved significantly in song-based recognition.

Marcelo T. Lopes et al. [11] employed and compared three different feature sets which are (a) the one produced by the MARSYAS framework, (b) the IOIHC feature set, that taps into rhythmic properties of sound signals, and (c) the feature set obtained from the audio processing tool Sound Ruler.Six different classification algorithms are used to evaluate diverse possibilities. Results showed that MARSYAS and Sound Ruler feature sets present good performance for almost every classifier.

From this study it is evident that several studies have attacked the bird species identification problem with various audio conditioning methods and different machine learning [12] algorithms. Up to now the obtained conclusions are somewhat unclear and difficult to compare, since there is no standard and the employed databases are specific containing few species only.

B. Bird species recognition problem

The automatic bird species recognition using audio signal processing can be defined as the problem of identifying the species of a specific bird from its recorded songs. Birds songs (related to mating) are more melodious compared to bird calls (short transient alert signals) and also considered as ideal for species identification by experts and are considered here.

In a digital recording device the sound is stored as a sequence of numeric values in a convenient scale, obtained by a sampling procedure.

The complete signal can be pre-processed and segmented in order to select its most representative part. Several features can be extracted from these segments to obtain the feature vector. Identification relies on previous known audio records which are manually labelled by experts. If this scenario occurs then it is easy to put the focused problem as a standard ML classification task.

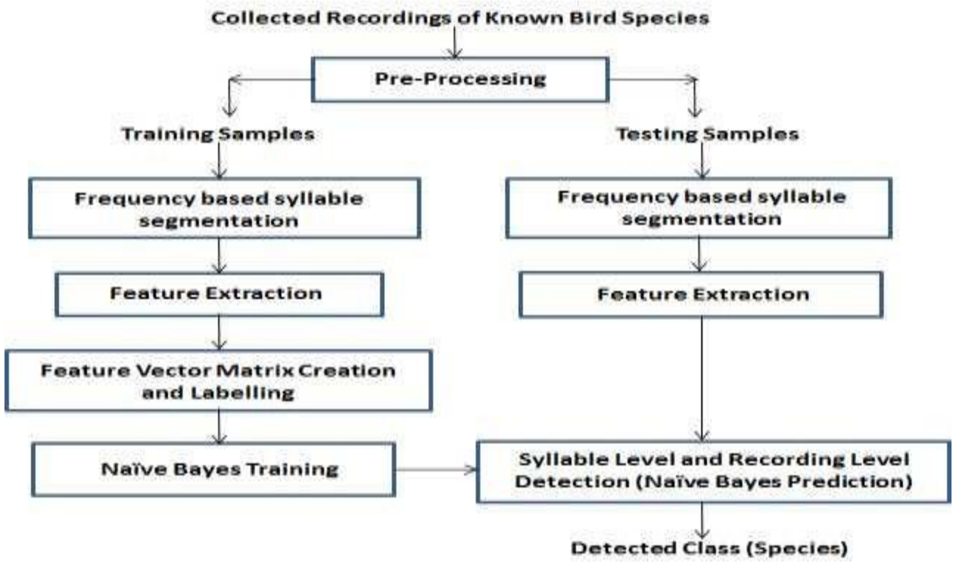

Figure 1 The block diagram of proposed bird species recognition algorithm

I. DATABASE AND PROPOSED RECOGNITION METHOD FOR BIRD SONG

A. Database

The database composed of songs from bird species which are commonly found in geographic region which encompasses India. This database containing over 200 bird song recordings from 4 bird species is collected from xeno-canto website [13]. The audio records in this database were obtained directly in real environments, without any pre-processing or audio cleaning, and thus contain sounds from other birds and animals as well as environmental noise such as waterfall, wind, rain.

B. Pre-processing

All the recordings which were .mp3 are first converted to simplest .wav file format and then converted to mono audio followed by re-sampling at 44100 Hz. Precision of 32 bits per sample is used to guarantee sufficient resolution of details for the subsequent processing. The recordings are then processed through Audacity (v2.1.2, free, open source, cross- platform software for recording and editing sounds) [14] for background noise reduction by setting required noise reduction limit. Database thus created is then extended by separating single lengthy recordings into multiple recordings. The final database is labelled. One of the preprocessed silence removed signal for *Asian Koel* song is shown in Fig. 2.

C. Syllable segmentation

Each input bird song is first segmented into a set of syllables using the frequency based syllable segmentation method proposed by Harma [15]. Every syllable is treated as the fundamental recognition unit of bird song. Syllable segmentation is necessary step since features extracted from syllables are less variant to territorial variations in bird sounds. The syllable segmented *Asian Koel* song is shown in Fig. 3. From this figure it is evident that frequency based syllable segmentation extracts almost complete syllables from recording.

D. Feature (AMFCC, DELTA, DELTA-DELTA) Extraction

Mel Frequency Cepstral Coefficients (MFCCs) have been the most widely used features for speech recognition [16], bird song recognition [6] [17], audio retrieval and other audio classification tasks [18] since they can represent the audio signal spectrum in a close-packed form. Here MFCCs for each syllable are obtained separately. In general, an input syllable is first divided into a set of frames.

Then MFCCs are computed for each frame with first MFCC coefficient replaced by logarithm of frame energy. The first 12 MFCC coefficients are used here for each individual frame. MFCCs thus obtained for different syllables may have varying number of feature vectors, since numbers of frames per syllable are not fixed. To handle this issue, the Averaged MFCCs (AMFCCs) of all the frames in a syllable are computed and used as features to represent the syllable. Thus features size is fixed (12 features per syllable) regardless of the syllable length. The regression coefficients allow integrating past and future context information. Hence velocity coefficients (Deltas) and acceleration coefficients (Delta-Deltas) can be obtained by using regression equation for each 12 dimensional feature vector, thus resulting in 36-dimensional feature set per syllable.

E. Classification algorithm

Naïve Bayes is a learning method for which a simple scheme-specific attribute selection approach has shown good results. Although this method deals well with random attributes, it has the potential to be misled when there are dependencies among attributes, and particularly when redundant ones are added. However, good results have been reported using the forward selection algorithm- which is better able to detect when a redundant attribute is about to be added than the backward elimination approach- in conjunction with a very simple, almost ―naïve,‖ metric that determines the quality of an attribute subset to be simply the performance of the learned algorithm on the training set.

1) Naïve Bayes Algorithm

Naive Bayes algorithm makes two ―naive‖ assumptions that, all attributes are a priori equally important and all attributes are statistically independent given the target class. This is a very strong assumption that is most unlikely in real data, i.e. that the attributes do not interact. Nevertheless, the approach performs surprisingly well on data where this assumption does not hold. The target and predictor variables are assumed to be categorical. A case is assigned to the class which has largest posterior probability.

By using Naïve Bayes' algorithm, the posterior probability of H given E is:

$$P(H|E) = \frac{P(E_1|H) * P(E_2|H) * \ldots * P(E_n|H) * P(H)}{P(E)} \quad \ldots..(1)$$

where,

- H – hypothesis
- E – evidence related to the hypothesis H, i.e., the data to be used for validating (accepting/rejecting) the hypothesis H
- P(H) – probability of the hypothesis (prior probability)
- P(E) – probability of the evidence i.e., the state of the world described by the gathered data
- P(E|H) – (conditional) probability of evidence E given that the hypothesis H holds
- P(H|E) – (conditional) probability of the hypothesis H given the evidence E

Class (Species) Detection

The recognition of species begins with species detection for each individual syllable. Class probabilities for each syllable are first obtained by using Naïve Bayes Prediction. The probability of occurrence of each class is counted. The class which has maximum class probability among the syllables is detected as the output class of that test recording. Thus making it a recording level recognition rather than syllable level.

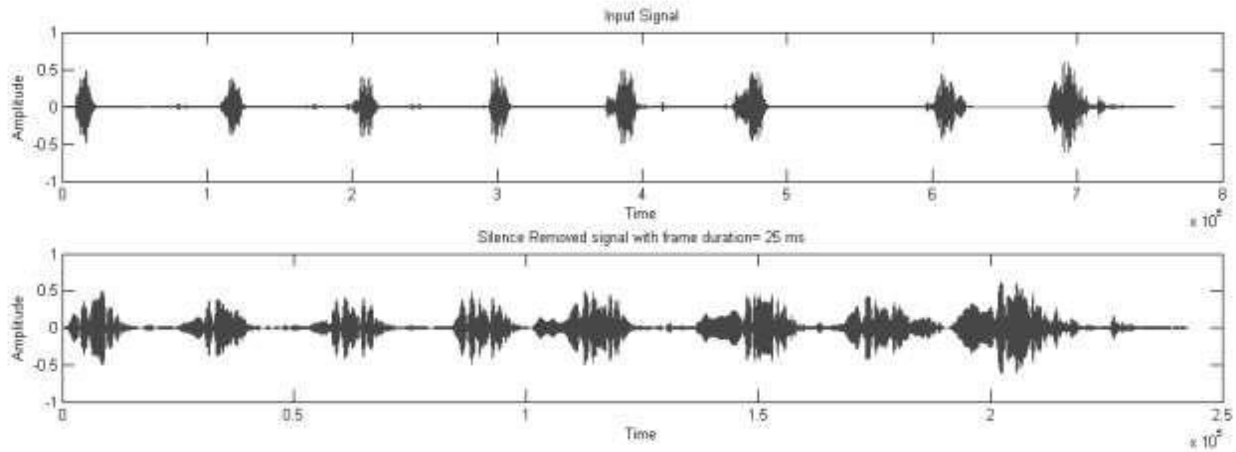

Figure2 Input and corresponding silence removed result of *Asian Koel* song

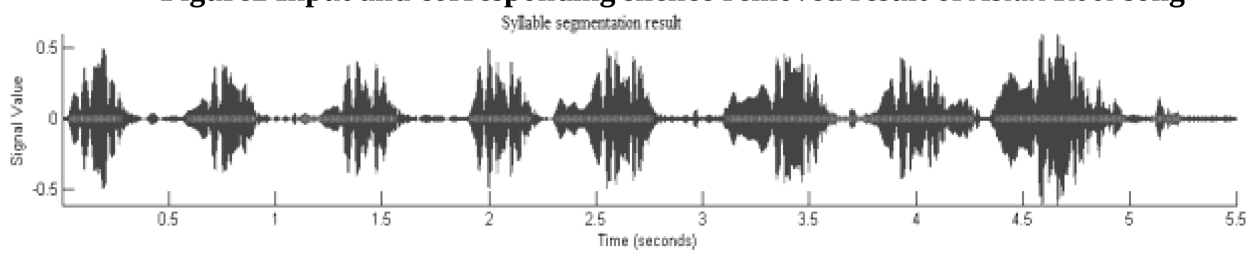

Figure 3Result of syllable segmentation when applied on silence removed *Asian Koel* song

II. EXPERIMENTAL RESULTS

The machine learning experiment was carried out on database containing over 200 bird song recordings. 70% of the available recordings from each species type are used for training and remaining 30% are used for testing purpose. The result of testing, are as shown in the confusion matrix in Table I. The advantage of using Confusion matrix is that, it is unbiased method of error rate estimation which not only tells us how many recordings got misclassified but also exactly where misclassification occurred.

where N_C is the number of recordings which were recognized correctly and N_{total} is the total number of test recordings for that class. From Table I it can be seen that Classification Accuracy of 100 % is obtained for all the classes (Species) of *Brahminy Kite* (BK), *Grey Francolin* (GF), *Hill Patridge* (HP), and *Spotted Dove* (SD). This accuracy measure will be a good accuracy estimate only if the number of test recordings N_{total} are large enough and representative of the population.

The class specific performance is also important in many fields. Hence we make use of the following types of test results; A True Positive (TP) (which detects the correct class), A True Negative (TN) (which does not detect the class when it is absent), A False Positive (FP) (which detects the class even if it is absent), and A False Negative (FN) (which detects the wrong class). Using these types of test results we define two performance metric as appear in Table II namely

TABLE 1 A CONFUSION MATRIX FOR 4 CLASSES WITH CLASSIFICATION ACCURACIES IN %

	Detected Class				
True Class	***Class***	***BK***	***GF***	***HP***	***SD***
	BK	100	0	0	0
	GF	0	100	0	0
	HP	0	0	100	0
	SD	0	0	0	100

TABLE 2 CLASS PERFORMANCE METRICS IN %

Class	TP	FP	TN	FN	Recall	Precision
BK	16	0	54	0	100	100
GF	22	0	48	0	100	100
HP	16	0	54	0	100	100
SD	16	0	54	0	100	100

Recall (Sensitivity) relates to the test's ability to identify a class correctly and Precision (Predictive value positive) is the proportion of positives that correspond to the presence of the class. There may be a problem in identification using this algorithm, when two or more birds sing simultaneously with comparable amount of intensity captured by recording device. This problem can be considerable because bird songs are most of the times connected with others (birds respond each other).

CONCLUSION AND FUTURE WORK

In this paper Bird species recognition using AMFCC, DELTA (δ), and DELTA-DELTA (δδ) features (obtained from frequency based syllable segments) and Naïve Bayes Algorithm for classification has been carried out for 4 bird species which are common to Indian Territory. The Recall (%) obtained is extremely good for all the considered species. The recognition based on class probabilities and recording level detection is more suitable than syllable level detection. At this point itis worth to note that, birds tend to call in clusters and reporting on a song class with individual syllables may not be as accurate as with complete song, with three or more syllables. Furthermore 30% of recordings were designated to test set and not 30% of syllables. Hence amount of bird syllables is different for different species.

As a future work we intend to test and include new species having features, which are not connected to spectral construction of syllables. Additional features, extracted from phrases and songs, which show connections between syllables could also be used. The algorithm needs to be tested for various other species for verifying its consistency in results.

REFERENCES

[1] http://www.ccac.ca/en /standards/policies/policy-categories of invasiveness#moderate>, Canadian Council on Animal Care Website, retrieved November 24, 2016.

[2] Dorota Kaminska, Artur Gmerek, Automatic Bird Species Identification: A Comparison of KNN and SOM Classifiers, New Developments in Audio & Video/Signal Processing Algorithms (NTAV/SPA), Applications, Architectures, and Arrangements September 27–29, 2012.

[3] "Automatic bird species identification for large number of species" by Marcelo T. Lopes, Lucas L. Gioppo, Thiago T. Higushi, Celso A. A. Kaesther, Carlos N. Silla Jr., and Alessandro L. Koerich, IEEE International Symposium on Multimedia, 2011.

[4] Jason Wimmer, Birgit Planitz, Ian Williamson, Paul Roe, Michael Towsey, Collectively and Automatedly Analyzing Environmental Acoustic Data, Future Generation Computer Systems 29, 560-568, 2013

[5] Bjorn Schuller and Felix Weninger, "Audio recognition in the wild: static and dynamic classification on a real-world database of animal vocalizations,"2011, IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).

In May 2006, Chang-Hsing Lee, Chih-Hsun Chou, Chin Chuan Han, and Ren Zhuang Huang published "Automatic recognition of animal vocalizations using averaged MFCC & linear discriminant analysis" in Pattern identification Letters 27(2006), 93-101 No. 1, pp. 17–23.

[7] Mporas Iosif, Ganchev Todor, Kocsis Otilia. Karl-L. Schuchmann, Klaus Riede, Olaf Jahn, Nikos Fakotakis, IEEE 24th International Conference on Tools with Artificial Intelligence, 2012, Automatic auditory categorization of bird species using real-field recordings.

[8] Bruno, A. Angelico, Celso A. A. Kaestner, Carlos N. Silla Jr., Thiago L. F. Evangelista, Thales M. Priolli, IEEE International Symposium on Multimedia, 2014. Automatic Segmentation of Audio Signals for Bird Species Identification.

[9] Detecting bird sounds in a complex auditory environment and application to bioacoustic monitoring, by R. Bardeli, D. Wolff, F. Kurth, M. Koch, K.-H. Tauchert, and K.-H. From-molt, Pattern Recognition Letters 31 1524-1534, 2010.

In their paper "Parametric representations of bird sounds for automatic species recognition," Panu Somervuo, Aki Harma, and Seppo Fagerlund cited IEEE Transactions on Audio, Speech, and Language Processing, Vol. 14, No. 6, November 2006.

[11] Feature set comparison for automated bird species identification, IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2011; Marcelo Teider Lopes, Alessandro Lameiras Koerich, Carlos Nascimento Silla Junior, and Celso Antonio Alves Kaestner.

[12] T. M. Mitchell, McGraw-Hill, 1997, Machine Learning.

[13] The most recent visit to the Xeno-Canto website, http://xeno-canto.org/, was made on January 19, 2017.

[14] On June 28, 2016, the Audacity website, http://www.audacityteam.org/download/windows/, was visited.

[15] Harma, Automatic bird species identification using sinusoidal syllable modeling, Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003 (ICASSP '03), vol. 5, pp. 545548.

[16]Speech recognition using MFCC and DTW, Bhadragiri Jagan Mohan and Ramesh Babu N., International Conference on Advances in Electrical Engineering (ICAEE), Vellore, pp. 1-4, 2014.

[17] C. H. Lee, C. C. Han, and C. C. Chuang used two-dimensional cepstral coefficients to automatically classify bird species based on their sounds in IEEE Transactions on Audio, Speech, and Language Processing, vol. 16, no. 8, pp. 1541-1550, November 2008.

[18] Young, The Technical Writer's Handbook, 2007. C. N. Silla Jr., C.A.A. Kaestner, and A. L. Koerich, Automatic music genre categorization using ensemble of classifiers, Proceedings of the IEEE Int. Conf. on Syst., Man, Cybern., Montreal, Canada, pp. 16871692. University Science, Mill Valley, CA, 1989.