iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 6 **Issue:** IV **Month of publication:** April 2018

DOI: http://doi.org/10.22214/ijraset.2018.4126

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue IV, April 2018- Available at www.ijraset.com

Singer Identification in Indian Hindi Songs: A Review

Ajay Jayswa[1], Hetal Guadani[2], Pranay Patel [3]
[1]*P.G.Student, Department of Computer Engineering, B.V.M. Engineering College, V. V. Nagar, India*
[2]*Assistant Department of Computer Engineer, G.H. Patel Engineer College, V. V. Nagar, India*
[3]*Assistant Professor, Department of Computer Engineering, B.V.M. Engineering College, V. V. Nagar, India*

Abstract: Singer Identification Systems are used to identify a singer from the song, which in turn is used as a sub-part of Music Information Retrieval (MIR) Systems. There are lot of Indian Hindi Songs on the Internet, and hence it is important to devise the system which can efficiently identify the singer from Hindi songs. In this paper, we have summarized frequently used audio features and classification methods to identify a singer from audio song.
Index Terms: Audio Features, MFCC, LPCC, ZCR, Singer Identification.

I. INTRODUCTION

Day by day, huge amount of Indian Hindi audio songs in digital format is stored on the Internet, hence the need for organizing the metadata in databases, querying and retrieving information about particular song, singer, genre etc. has increased significantly. Music information retrieval (MIR) is define as a science of retrieving information from music. MIR is innovative field of research with many real world applications like the knowledge of background in musicology, psychology, academic music study, signal processing, machine learning or combination of them. Singer Identification (SID) is one of the sub field of MIR that has also grown rapidly in recent times. SID means identifying which singer has sung a particular song or part of song (if sung by multiple singers). Human brain can easily identify particular singer from songs with little training. Human have intelligent brain which organizes the songs as input in such a way that they may be recognised easily. But for a machine, singer identification (SID) is difficult because of reasons like presence of different instrument music, audio of other singer, if more than two singers are singing at the same time. Also speech identification is quite unlike singing voice. The voice is crafted from flesh and bone rather than metal or wood, and thus the physical characteristics of each voice are as unique as one's fingerprint or eye retina. Further, singers use their voices in the ways which enhance the uniqueness of the singer voice. Both of these factors, biology and technique contributes to SID.

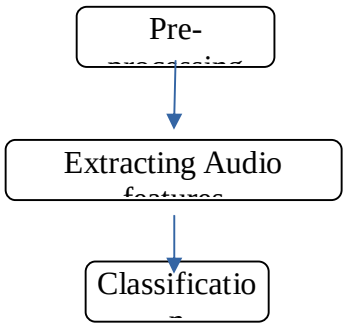


Figure 1: General Flow of SID systems

Overall, SID system has three steps: 1) Pre-processing 2) Extracting Audio features and 3) Classification. Classification contains two stages i.e. the training stage, in which model is trained by given data, and the testing stage, in which unknown data is given as input to trained c1assifier model and the model then c1assify the data into different c1asses. Portion of the testing data correctly c1assified, is considered as measure of performance.

II. LITERATURE SURVEY

Sarfaraz Masood et al. [1] presented the paper, in which they used audio sample of 2 seconds from five well known singers' solo songs. Owing to difficulty in separation of singer's voice and instrument music, they manually selected samples having low

©IJRASET (UGC Approved Journal): All Rights are Reserved

intensity background music. They used 8 audio features namely mel-frequency cepstral coefficients (MFCC) and Spectral features like root mean square energy, brightness, roughness, spectral roll off, skewness, flatness, and spectral centroid for classification. Their system got overall 92.5% accuracy using multi-level feed-forward artificial neural network.

Tushar Ratanpara and Narendra Patel [2] used different approach. From Indian Video Songs, they extracted perceptual features of an audio signal cepstral coefficients to identify a singer. They used 53 audio features i.e. 12 dimensional timbre audio feature vectors, 12 pitch classes, 13 MFCC coefficients, 13 LPC coefficients, and 3 loudness feature vector of an audio signals. Five different classifiers were used namely AdaBoost.M2, k-nearest neighbour (KNN), GMM, BPNN, and NBC model on various divisions of datasets for classification. Their research shows that AdaBoost.M2 provides more accurate results than other classification model used. Particularly when learning cycles are increased from 50 to 5000, AdaBoost.M2 gives more accurate results.

Unnikrishnan V M and Rajeev Rajan [3] presented a research in which they mimicked the quality of tespeech using MFCC. For Classification, they used GMM. The experiment evaluated the competence of 5 artists in mimicking 5 target speakers and ranked them according to the scores of a classifier and identified the best mimicking performance.

Annamaria Mesaros et al. [4] presented methods and applications for singer identification based on lyrics content of the singing. They dealt with analysis of audio signals. MFCCs are extracted in frames of 20-30 ms, by applying a windowing function at fixed intervals. At the point where singing is detected, a 25 seconds fragment was selected and represented using frame based cepstral coefficients derived from LPC of order 12, and used to train GMMs for singers. The authors simply assumed that the 25 seconds following detection will contain voice, but nevertheless the performance of the method is 82% in classifying 45 songs belonging to eight singers.

Namrata Dave et al. [6] compared various methods like MFCC, LPC, PLP and their prone and cons for Speech Recognition and used HMM and Neural Network techniques for pattern recognition.

III.FEATURES OF AUDIO SIGNAL

A. MFCC (Mel-Frequency Cepstral Coefficients)

The most prevalent and dominant method used to extract spectral features is calculating MFCC. In 1980 Davis and Mermelstein developed MFCC features which is widely used in automatic speech and speaker recognition. Fig. 2 shows MFCC derivation block diagram. MFCC is one of the most popular feature extraction techniques used in speech recognition based on frequency domain using the Mel scale which is based on the human ear scale. [4]

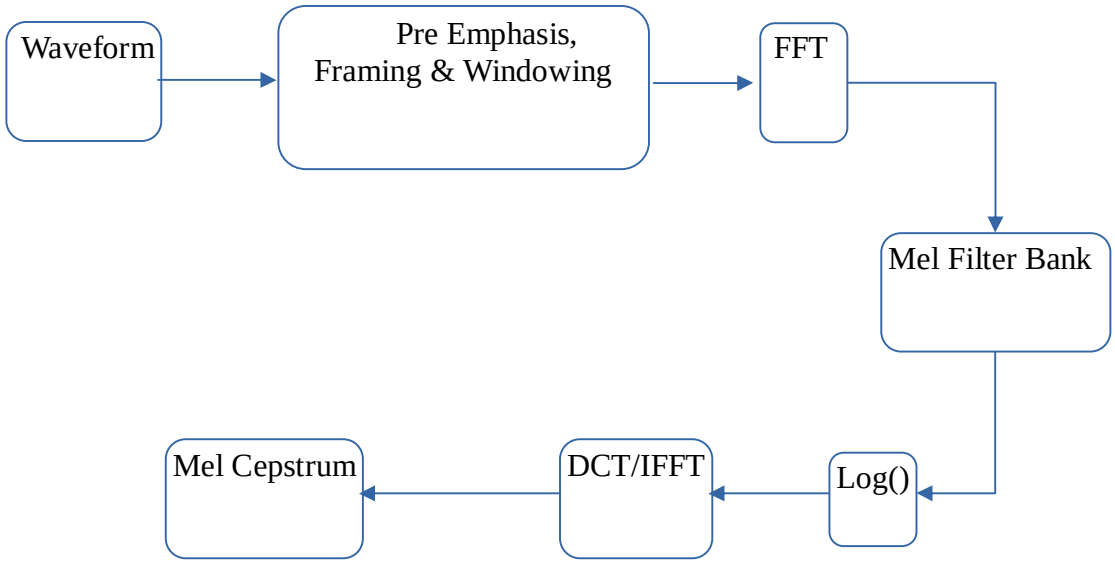


Figure-2 MFCC derivation

©IJRASET (UGC Approved Journal): All Rights are Reserved

The following figure-3 shows the MFCC coefficients graph for two singers.

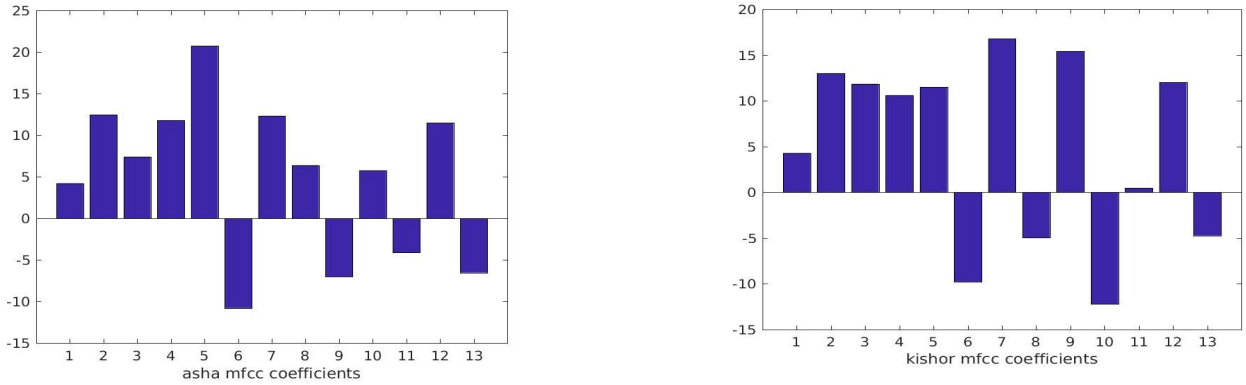


Figure-3 MFCC graph for two singers

B. LPCC

It is desirable to compress signal for efficient transmission and storage. For efficient utilization of channels on wireless media transmission Digital signal in compress form, LPC is most widely used for medium or low bit rate coder. Power spectrum of the Digital signal is calculated By LPC using formant analysis. LPC is one of the most powerful speech analysis techniques and it has gained popularity as a formant estimation technique. LPCC calculation is as shown in following figure.

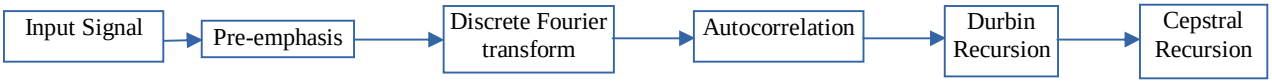


Figure-4 LPCC derivation

The speech signals can be seen as linear combination of the previous p samples. Therefore, the speech production model can be often called linear prediction model, or the autoregressive model. From here, p indicates the order of the LPC analysis.

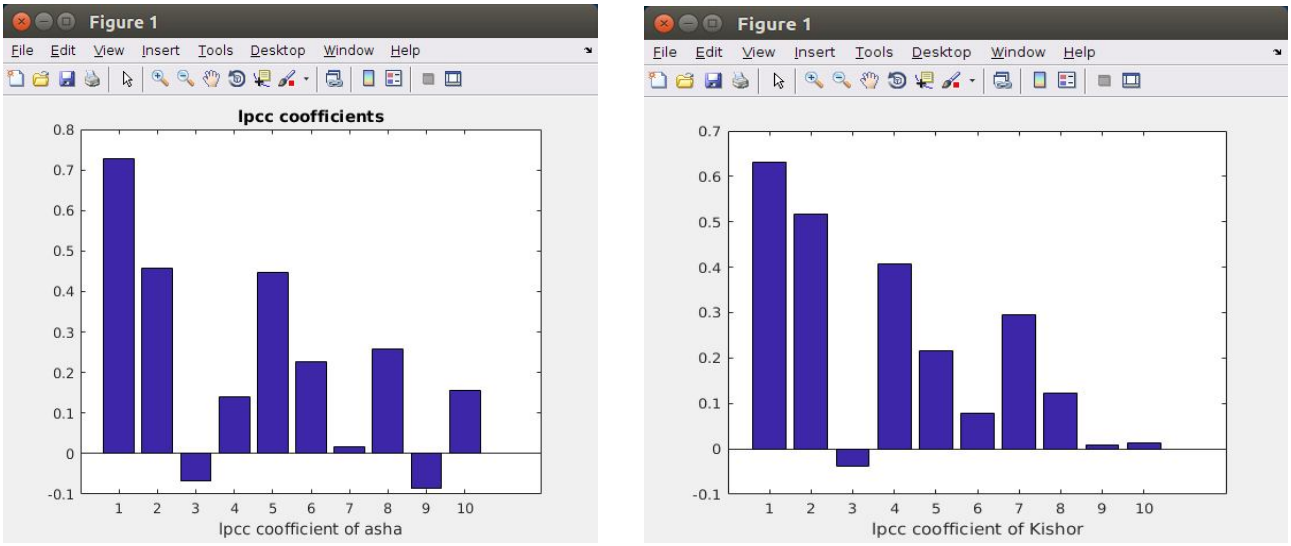


Figure-5 LPCC graph for two singers

C. Zero Crossing Rate (ZCR)

The Number of zero crossing of the audio signal within a given frame is defined by zero crossing rate. Characteristics of ZCR is following: (1) Unvoiced sound and environment noise usually greater than ZCR of voice.

©IJRASET (UGC Approved Journal): All Rights are Reserved

(2) It is very difficult to separate unvoiced sounds from environmental noise by using only Zero Crossing Rate, because both have similar Zero Crossing Rate. (3) ZCR is often used in conjunction with energy (or volume) for end-point detection. In particular, ZCR is used for detecting the start and end positions of unvoiced sounds.

D. Pitch [4]

Pitch is related to fundamental frequency of a sound. Pitch is said to range from low or deep to high sounds. Higher the frequency higher the pitch of sound. Different Pitch range of male voices is observed to be from 209Hz to 360Hz, while that of the female voices lies between 243Hz to 774Hz. The pitch range helps to identify the gender of the singer.

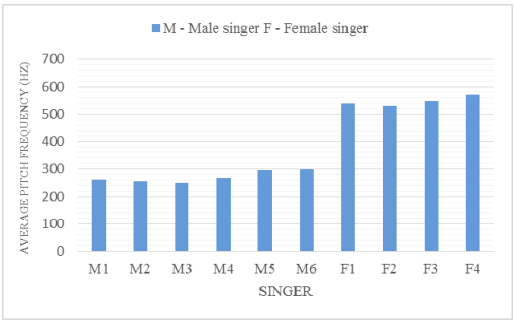


Figure-6 Plot of average pitch values of singers obtained by Cepstrum technique from different Cappella song[4]

E. Timbre

It is the tonal quality of sound. Each sound has its own tone which is known as timbre. Voices of different females differ with different tone quality. It is a good feature to classify an instrument but it does not give a good result to classify human voice.

F. Loudness

It is the energy or the vibration of the sound. Sound intensity is also defined as loudness.

G. Classification

After extraction of features from an audio signal, next step is identification of a singer using different classifiers. It is two step train the classifier. Classification can be based on supervised or unsupervised learning methods.

H. Random Forest

Random Forest Classification algorithm which is mostly used in recent years for speaker Identification System. In this algorithm, first all the samples are divided into subsamples. All these sub samples are sub decision tree. Basically random forest uses concept of decision tree to test and train samples. For different sub samples the labels are decided and from the random split test set for each sample all the decision trees are visited and based on majority vote final label gets decided. Each tree is grown as follows:

If the number of cases in the training set is N, sample N cases at random but with replacement, from the original data. This sample will be the training set for growing the tree.

If there are M input variables, a number m<<M is specified such that at each node, m variables are selected at random out of the M and the best split on these m is used to split the node. The value of m is held constant during the forest growing.

Each tree is grown to the largest extent possible. There is no pruning.

I. K-nearest neighbor (KNN) [2]

K-nearest neighbor is lazy algorithm. We do not process or learn a model, instead we just store the examples. For each point, the corresponding x value and the y value is stored. When given a new instance, we find what the closest instance in terms of the x value is and suppose this is the closest, find the y value of the instance. This is the basic nearest neighbor algorithm. It is used to predict singer using values of K. The training samples are audio feature vectors which are distributed in a multidimensional feature space. Each training sample contains a class label. Feature vectors and class labels of training samples are stored in the training phase. Euclidean distance is computed for each test sample. Test samples are classified by assigning the class labels using k nearest training samples. Using this model usually provides 50% to 60% accuracy in Indian Video Songs System shown in figure-7.

©IJRASET (UGC Approved Journal): All Rights are Reserved

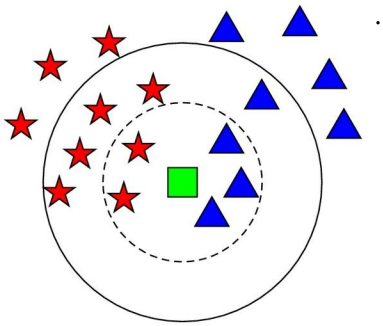


Figure-6 KNN classification algorithm [7]

J. K-Mean Cluster

K-mean cluster which is type of unsupervised machine learning algorithm which is used in [5]. The goal is to organize data into cluster such that data points having high similarity falls into same cluster and to find grouping of the objects.

K. Gaussian Mixture Model (GMM)[9]

This model is widely used for singer identification systems. Gaussian mixture model is all about fitting a particular distribution using a mixture of Gaussian's linear combination of several Gaussian distributions. And the algorithm is trained by data using a particular algorithm which is called the expectation maximization algorithm, which is an efficient way of doing approximate maximum likelihood estimation. This model provides 40% to 50% accuracy.

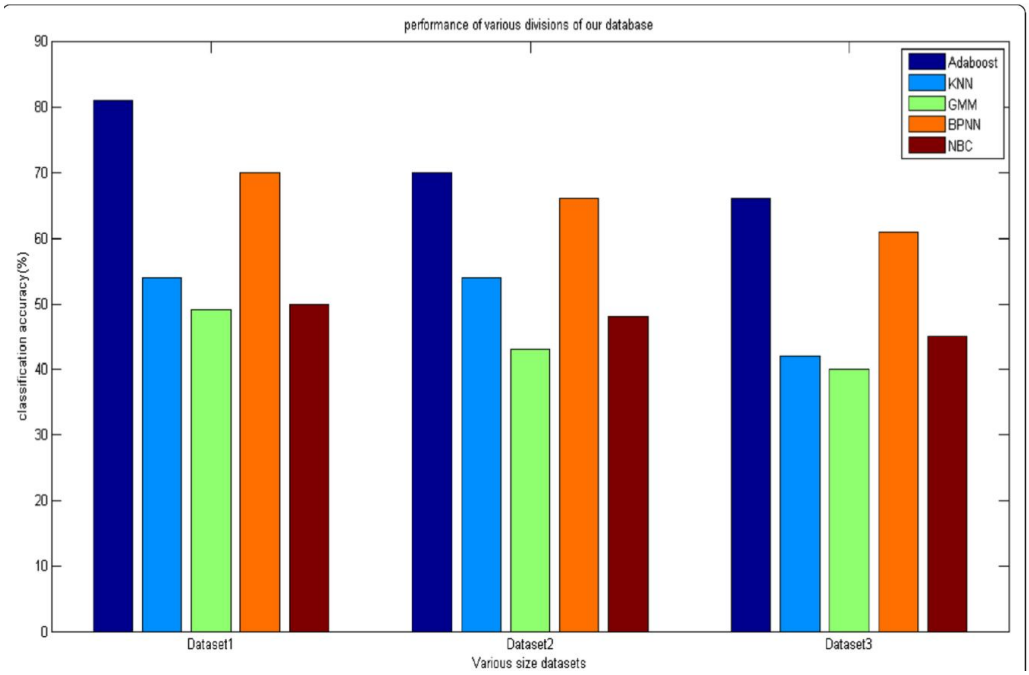


Figure-7- Performance of various classification algorithm of same dataset [2]

IV. CONCLUSION

From the literature survey, different features and classification methods, we can conclude that the fields of speech recognition and speaker verification applications used the most frequently features i.e. LPCC and MFCC. But only LPCC is not so acceptable because of its linear computation nature. KNN,K-mean Cluster ,GMM and Random Forest are considered as the most dominant pattern recognition techniques used in the field of speech recognition and Singer Identification System. We also discussed some basic audio features like Pitch, ZCR, Timber, and Amplitude.

©IJRASET (UGC Approved Journal): All Rights are Reserved

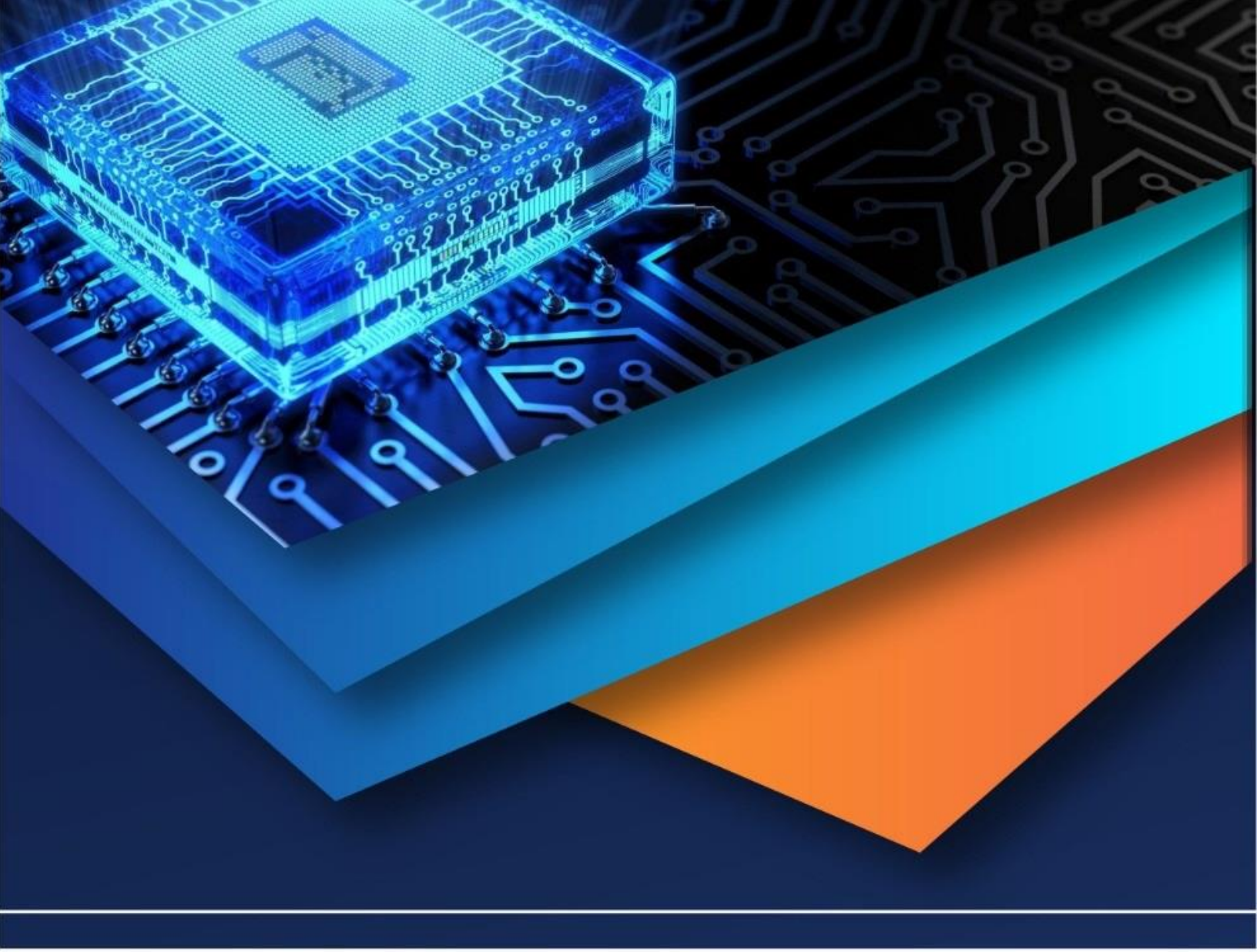
V. ACKNOWLEDGMENT

I am very grateful and would like to thank my guide Prof. Hetal Gaudani, Computer Engineering Department, G.H. Patel College of Engineering & Technology and Prof. Pranay Patel, Computer Engineering Department, B.V.M Engineering College for their advice and continuous support.

REFERENCES

[1] Sarfaraz Masood , Jeevan Singh Naya1 and Ravi Kumar Jain , Singer Identification in Indian Hindi Songs using MFCC and Spectral Features, 1s IEEE International Conference on Power Electronics. Intelligent Control and Energy Systems (ICPEICES-2016)
[2] Tushar Ratanpara and Narendra Patel , Singer identification using perceptual features and cepstral coefficients of an audio signal from Indian video songs,EURASIP Journal on Audio,speech,and Music Processing ,2015 Springe
[3] Unnikrishnan V M and Rajeev Rajan ,"Mimicking Voice Recognition Using MFCC-GMM Framework ",International Conference on Trends in Electronics and Informatics,ICEI 2017
[4] Annamaria Mesaros ∗ , Simina Moldovan , "Methods for singing voice identification using energy coefficients as features" ,
[5] Saurabh H. Deshmukh , Dr. S.G.Bhirud ,Analysis and application of audio features extraction and classification method to be used for North Indian Classical Music's singer identification problem,
[6] Deepali Yoginath Loni , Dr. Shaila Subbaraman ,Singing Voice Identification Using Harmonic Spectral Envelope,2015 International Conference on Information Processing (ICIP) Vishwakarma Institute of Technology
[7] Jian Wu, 1 Zhiming Cui, 1 Victor S. Sheng, 2 Yujie Shi, 1 and Pengpeng Zhao 1"Mixed Pattern Matching-Based Traffic Abnormal Behavior Recognition" Hindawi Publishing Corporation The Scientific World Journal Volume 2014, Article ID 834013, 12 pages.
[8] Aathreya S. Bhat ,Amith V. S.,Namrata S. Prasad,Murali Mohan D.,An Efficient Classification Algorithm For Music Mood Detection In Western and Hindi Music Using Audio Feature Extraction,2014 Fifth Fifth International International Conference Conference on on Signals Signal and and Image Image Processing Processing
[9] An introduction to machine learning with scikit-learn (April 2018))Rettrived from scikit-learn.org/stable/_downloads/scikit-learn-docs.pdf
[10] Classification Algoritham in Machine Learning (April 2018) Retrived From nptel.ac.in/courses/117106100/Module
[11] Audio Features And Its Matalab Code Help(March 2018) ,Retrieved from jcbrolabs.org/speech-processin
[12] Matalabcode Tollbox (march 2018) Retrived from practicalcryptography.com/miscellaneous/
[13] Wei-Ho Tsai, Member, IEEE, and Hsin-Chieh Lee, Singer Identification Based on Spoken Data in Voice Characterization,IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 20, NO. 8, OCTOBER 2012

©IJRASET (UGC Approved Journal): All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)