JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue

JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)

An International Scholarly Open Access, Peer-reviewed, Refereed Journal

GENDER BASED CLASSIFICATION OF HINDI SONGS USING AUDIO FEATURES

[1]Mittal Darji, [2]Vinita Shah

[1]Assistant Professor, [2]Assistant Professor
Department of Information Technology,
G H Patel College of Engineering & Technology, Anand, India

Abstract : Hindi movies contain many songs which can be classified in various ways based on the properties of audio and video. This paper presents the classification of Hindi movie songs based on the gender of singer. Songs are classified into 3 classes: Male solo, Female solo and Duet. The main 2 steps of the algorithm are feature extraction and classification. Male and female generally differ in their frequency and pitch. Hence, pitch and other frequency based audio features like Zero Crossing Rates (ZCR), spectral centroid and energy entropy are used for this purpose. Based on these features classification is performed using a neural network.

IndexTerms **- Song Classification, Audio Features, Pitch, Zero Crossing Rate, Spectral Centroid, Energy Entropy.**

I. INTRODUCTION

A large number of Hindi movies are released every year leading to a huge collection of movies. This intern will create a large collection of songs as well. Popularity of Hindi cinema keeps on increasing. There are other industries related to the film industry. There are people like singers and dancers who are interested only in songs from movies. CD or cassette making is another business that deals with songs. Songs are thus having a large number of repeated viewers. The revenue generated by the music and songs of a movie is around 4-5% of the total as per [11]. The importance of film music and songs can be realized as well from the fact that 48% of India's music sales are from film music says [12]. These applications and popularity of songs have inspired researchers to develop classification techniques which can ease the job of many.

Such a huge collection of songs can be classified according to genre, artist, instruments used, emotions etc. In this paper, gender of the singer is used for classification purpose. Classification as per desire from a big collection is a tedious and time consuming process when done manually. Several classification techniques are developed over the time by researchers to make this job easier. These techniques are based on video features and/or audio features of songs. This paper works on audio features as audio based classification is well suited for our purpose [10]. Classification in general uses main 2 steps: 1) feature extraction 2) classification.

Audio has various features which shows specific characteristics for male and female. Female generally have higher fundamental frequency than male. Therefore, female have higher pitch value than male. But some male have higher frequency and some female have lower pitch. Hence, only pitch is not enough for appropriate classification purpose. Therefore other audio features like Zero Crossing Rate (ZCR), spectral centroid and energy entropy are involved.

II. LITERATURE SURVEY

Researchers have made many attempts over the time to classify audio gender wise using different audio features and classifiers. Hadi Harb, Liming Chen, Jean-yves Auloge have used frequency and ZCR for classification. [1]

BhagyaLaxmi Jena, Abhishek Majhi, Beda Prakash Panigrahi have used short-time average magnitude, short-time energy, short-time auto-correlation and ZCR audio features. [2]

Anchita Goel and Sarthak Ahuja have used features like Mel-Frequency Cepstral Coefficient (MFCC), pitch, short-time energy, energy entropy, ZCR and spectral centroid. [3]

Kunjithapatham Meena, Kulumani Subramaniam, and Muthusamy Gomathy performed classification using short-time energy, ZCR and energy entropy audio features and neural network as a classifier. [4]

Pawan Kumar, Nitika Jakhanwal, Anirban Bhowmick, and Mahesh Chandra classified using pitch and frequency formants. [5]

R. Shiva Shankar, J. Raghaveni, Pravallika Rudraraju, Y.Vineela Sravya have used audio features such as pitch, median, frequency and have tried various classification techniques to find out which one works the best for given dataset [7].

It is observed from the survey that researchers have mostly used pitch and frequency related features for classifying audio into male voice and female voice. Movie songs are not having only solo singers. Duet songs are also there. Duet songs, being a mixer of male solo and female solo, do not possess any specific characteristics. Still an attempt can be made to train a classifier model for duet songs and test whether it can recognize duet songs correctly or not. This paper presents an approach to classify songs into 3 classes: male solo, female solo and duet.

After this, the paper contains sections are follows. In section III, various audio features are explained which are used later in implementation. Workflow of system is explained in section IV. Section V, provides experimental results. Performance analysis of the work is in section IV. Section V concludes the paper followed by possible future extensions in section VI.

III. SPEECH SIGNAL FEATURES

3.1 Pitch

Pitch is a perceptual property of sounds that allows their ordering on a frequency-related scale, or more commonly, pitch is the quality that makes it possible to judge sounds as "higher" and "lower" in the sense associated with musical melodies [14].

Pitch is defined as the fundamental frequency of the excitation source. Fundamental frequency of female is higher than that of male. Hence, pitch is most widely used feature for gender detection.

Pitch value is calculated for 37 male solo songs and 37 female solo songs which are plotted in Figure 1 for comparison. Here series 1 is for male solo songs and series 2 is for female solo songs.

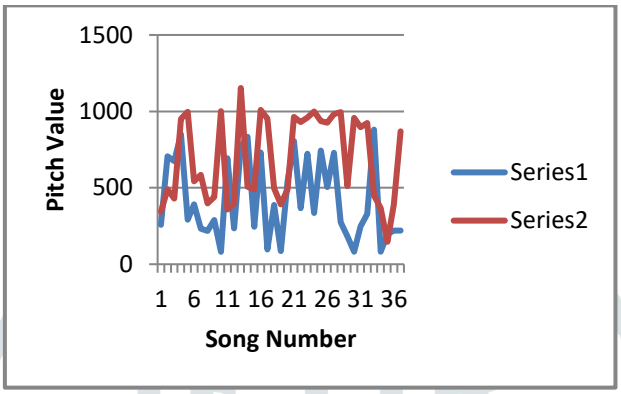

Figure 1. Comparison of pitch

Figure 1 shows that female have higher overall pitch value than male.

3.2 Zero-Crossing Rate (ZCR)

It is the rate of sign-changes along a signal. It is defined as the number of times zero crossed within a frame [9]. ZCR is an important feature for separating signal into voiced and unvoiced. Also it is much useful in gender classification because female have a higher ZCR compared to male. Equation 1 is for calculation of ZCR.

$$\text{ZCR} = \frac{1}{T-1}\sum_{t=1}^{T-1} F(St * St - 1 < 0) \qquad (1)$$

Where, S is a signal of length T

$$F(A) = \begin{cases} 1, & \text{if A=true} \\ 0, & \text{otherwise} \end{cases}$$

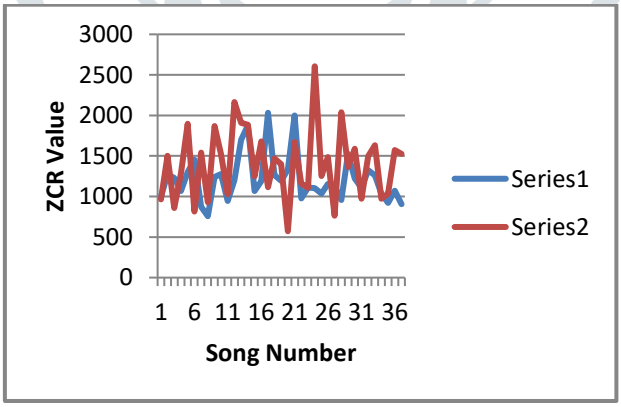

Figure 2. Comparison of ZCR

Figure 2 displays the comparison of Zero Crossing Rate for male solo and female solo songs where series 1 is for male and series 2 is for female. Series 2 is overall higher than series 1.

3.3 Spectral Centroid

The spectral centroid is a measure used in digital signal processing to characterize a spectrum. It indicates where the "center of mass" of the spectrum is. Perceptually, it has a robust connection with the impression of "brightness" of a sound [15].

$$\text{Centroid} = \frac{\sum_{n=0}^{N-1} f(n)x(n)}{\sum_{n=0}^{N-1} x(n)} \qquad (2)$$

Where x(n) is the weighted frequency value or magnitude of bin n and f(n) is the center frequency of that bin.

3.4. Energy Entropy

Energy Entropy (EE) in speech signal is defined as the sudden different changes in the energy level of a speech signal. Energy entropy for males is low and distributed while for females it is high and remains for a short period [3] [4].

External library Music Information Retrieval (MIR) toolbox version 1.5 is used for features calculation [18].

IV. CLASSIFICATION

Proposed method is implemented on MATLAB R2012a 7.14.0.739. Neural network is selected as a classifier for which nntoolbox provided by MATLAB is used. A feed forward neural network with 2 hidden layers having 5 and 2 neurons in each layer respectively is configured for classification.

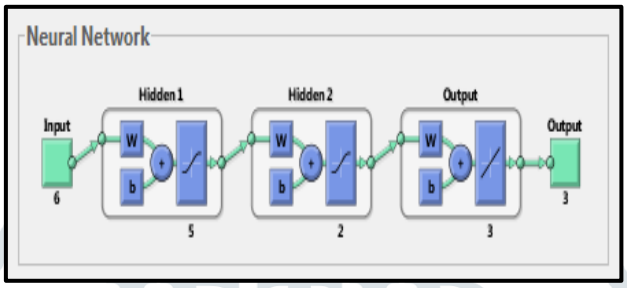

Figure 3. Neural Network

Figure 3 is the network which is trained by 113 total songs containing all the 3 type male solo, female solo and duet songs. 4 features are extracted for each song which makes a [4×113] training matrix used to train the network till the performance goal is met.

Such a trained network is tested for 18 songs different than those of training songs. Table 1 shows the result of 18 testing songs with their actual class and the class recognized by network.

TABLE. 1. Classification result

No.	Name of song	Actual class	Classified as	Result
1	Yaaro dosti badi hi hasin he	Male	Male	PASS
2	Khul kabhi	Male	Male	PASS
3	Bismil	Male	Male	PASS
4	Bhagwan hai kaha re tu	Male	Male	PASS
5	Kyuki tum hi ho	Male	Male	PASS
6	Abhi muj me kahi	Male	Male	PASS
7	Baki mai bhul gae	Female	Female	PASS
8	Dil hai chhota sa	Female	Female	PASS
9	Ishq kabhi kario na	Female	Female	PASS
10	Zindgi me koe kabhi aaye na rabba	Female	Female	PASS
11	Saara jahaan	Female	Female	PASS
12	Zalim	Female	Female	PASS
13	Dil ke badle sanam	Duet	Duet	PASS
14	Ye hum aa gaye hai kaha	Duet	Female	FAIL
15	Pyar ye hota hai kya	Duet	Female	FAIL
16	Pardsi jana nahi	Duet	Duet	PASS
17	Dhire jalna dhire jalna	Duet	Duet	PASS
18	Chori kiya re jiya	Duet	Female	FAIL

Result is plotted in a graph shown in figure 4 for all 18 songs where Series 1 indicates the actual classes of songs and Series 2 indicates the classes labelled by the trained network. Output has 3 classes where, number 1 is for Male, number 2 is for Female and number 3 is for Duet. We can see that series 1 and series 2 differ for song number 6, 9 and 18. All of them are duet songs. This happens with the duet songs because one of the female voice or male voice is more than the other.

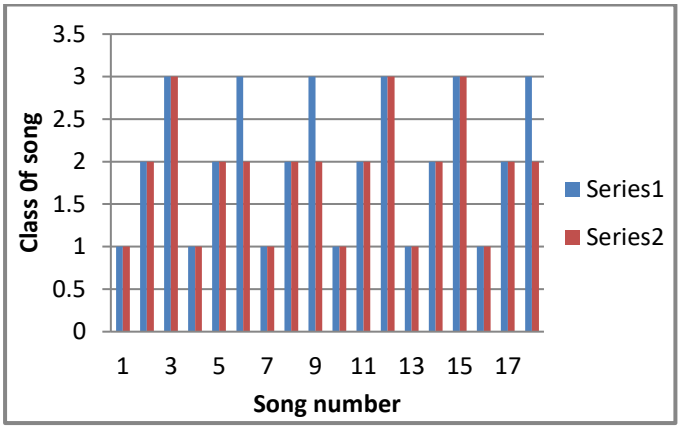

Figure 4. Classification result graph

V. Performance Analysis

TABLE. 2. Performance analysis of song classification

Category	No of songs given for testing	No of songs labeled correctly	Accuracy (%)
Male	6	6	100
Female	6	6	100
Duet	6	3	50
Total	18	15	83.33

Table 2 shows that implementation of song classification work provides overall 83.33% of accuracy. It is analysed that the classifier work fine for male and female songs but it does not give perfect result for duet song due to the fact mentioned earlier.

VI. Conclusion

A song classifier is very much useful in current scenario where popularity and use of movie songs are increasing so high day by day. This paper is an attempt to classify songs into 3 classes: Male solo, Female solo and duet. For which the difference in fundamental frequency of male and female is used by audio features pitch, ZCR, spectral centroid and energy entropy. Neural network was trained and tested for classification. Result analysis shows that male solo and female solo songs are classified successfully but duet songs are difficult to identify accurately as duet song may contain either male or female voice portion greater than the other. Overall precision of 83.33% is achieved.

VII. Future Work

Precision can be improved if duet songs can be classified more precisely. Deep learning approaches can be used for better performance of classification [6]. Songs can also be classified in various ways like emotion wise, genre wise, based on the instruments used etc. Age can also be predicted from the audio [6].

References

1) Hadi Harb, Liming Chen, Jean-yves Auloge, “Speech/Music/Silence and Gender Detection Algorithm”, 7th International conference on Distributed Multimedia Systems DMS01, 2001.
2) A Raahul, “Voice based gender classification using machine learning”, IOP Conf. Ser.: Mater. Sci. Eng. 263 042083, 2017.
3) Musaed Alhussein,1 Zulfiqar Ali,1 Muhammad Imran,2 and Wadood Abdul, “Automatic Gender Detection Based on Characteristics of Vocal Folds for Mobile Healthcare System” Volume 2016, Article ID 7805217, https://doi.org/10.1155/2016/7805217.
4) Kunjithapatham Meena, Kulumani Subramaniam, and Muthusamy Gomathy, “Gender Classification in Speech Recognition using Fuzzy Logic and Neural Network”, The International Arab Journal of Information Technology, Vol. 10, No. 5, September 2013.
5) Pawan Kumar, Nitika Jakhanwal, Anirban Bhowmick, Gender Classification Using Pitch and Formants”, International Conference on Communication, Computing & Security, ACM New York 2011, ISBN: 978-1-4503-0464-1.
6) Sánchez-Hevia, H.A., Gil-Pita, R., Utrilla-Manso, M. et al. Age group classification and gender recognition from speech with temporal convolutional neural networks. Multimed Tools Appl 81, 3535–3552 (2022).Ali, A. 2001.Macroeconomic variables as common pervasive risk factors and the empirical content of the Arbitrage Pricing Theory. Journal of Empirical finance, 5(3): 221–240.
7) R. Shiva Shankar, J. Raghaveni, Pravallika Rudraraju, Y.Vineela Sravya, “Classification of Gender by Voice Recognition Using Machine Learning Algorithms”, International Journal of Advanced Science and Technology, Vol. 29 No. 06, 2020
8) Tursunov, A.; Mustaqeem; Choeh, J.Y.; Kwon, S. Age and Gender Recognition Using a Convolutional Neural Network with a Specially Designed Multi-Attention Module through Speech Spectrograms. Sensors 2021, 21, 5892.
9) Rami S. Alkhawaldeh, “Gender Recognition of Human Speech Using One-Dimensional Conventional Neural Network”, Hindawi, Scientific Programming, Volume 2019, Article ID 7213717

10)BhagyaLaxmi Jena, Beda Prakash Panigrahi, "Gender Classification by Pitch Analysis", International Journal on Advanced Computer Theory and Engineering (IJACTE), ISSN (Print) : 2319 – 2526, Volume-1, Issue-1, 2012.
11)M.C. Darji, Dr. N.M. Patel, Z.H. Shah, "A Review on audio features based extraction of songs from movies", International Journal of Advance Engineering and Research Development (IJAERD) e-ISSN: 2348 – 4470, print-ISSN: 2348-6406, 2015.
12)M.C. Darji, Dr. N.M. Patel, Z.H. Shah, "Extraction of Video Songs from Movies using Audio Features", IEEE, Advanced Computing and Communication (ISACC), Print ISBN: 978-1-4673-6707-3, 2015.
13)S. M. Doudpota, S. Guha, "Mining Movies to Extract Song Sequences", ACM 978-1-4503-0841-0 MDMKDD'11, August 21, 2011.
14)Yakun Hu, Dapeng Wu, Antonio Nucci, "Pitch-based Gender Identification with Two-stage Classification", Security and Communication Networks Volume 5 Issue 2, February 2012.
15)S. Gaikwad, B. Gawali, S. Mehrotra, "Gender Identification Using Svm With Combination Of Mfcc", Advances in Computational Research, ISSN: 0975-3273 & E-ISSN: 0975-9085, Volume 4, Issue 1, 2012.
16)F. Weninger, M. Wollmer, B. Schuller, "Automatic Assessment Of Singer Traits In Popular Music: Gender, Age, Height And Race", 12th International Society for Music Information Retrieval Conference, ISMIR 2011.
17)Carey, M.J., Chepstow, L. Thomas, "A comparison of features for speech, music discrimination" Acoustics, Speech, and Signal Processing, Proceedings, IEEE International Conference in vol. 1, 1999.
18)Music Information Retrieval Toolbox available:
https://www.jyu.fi/hum/laitokset/musiikki/en/research/coe/materials/mirtoolbox