Volume 5, Issue 11, November 2015 ISSN: 2277 128X

International Journal of Advanced Research in Computer Science and Software Engineering

Research Paper

Available online at: www.ijarcsse.com

An Approach for Singer Identification Technique Using Artificial Neural Network

Shruti
Research Scholar, CSE Department
CEC Landran Chandigarh, India

Bharti Chhabra
Asst Prof. CSE Department
CEC Landran Chandigarh, India

***Abstract** The singer voice recognition system is most prominent technique of identification of singer's voice. The unique qualities of a singer's voice make it relatively easy to identify the particular artist of song, who belongs to that song. This paper focuses on the basic concepts of the feature extraction and classification in speech identification system. There are 9 singers and 5 songs of each singer. So there are total 45 songs in our database. In this paper DCT is applied to derive cepstral features GFCC is used for feature extraction.ANN is applied to classify. The ANN classifier classified 88.9 % of singers correctly and the ERR is 11.1%.*

***Keywords**— Speech Classification, Features extraction, GFCC, Singer Identification, ANN*

I. INTRODUCTION

Singer Identification (SID) is the process of retrieving identity of the singer in a song through features of voice. The voice recognition system is most prominent technique to identify of singers voice. Every person has unique voice quality. The unique qualities of a singer's voice make it relatively easy to identify a song of particular artist. The identity of a singer can be identified by using Artificial Neural Network (ANN). The singer identification comes under speaker identification or voice biometrics

A. Biometrics

The term "biometrics" is derived from the Greek words bio (life) and metric (to measure). Biometrics refers to metrics related to human characteristics. Biometrics authentication is used in computer science as a form of identification and access control. Biometric first came to limelight in 1879.

Types of Biometrics

1. Physiological Characteristics: are related to the shape of the body. Examples include, but are not limited to fingerprint, palm veins, face recognition, DNA, palm print, hand geometry, iris recognition, retina and odour/scent.
2. Behavioural Characteristics: are related to the pattern of behaviour of a person, including but not limited to typing rhythm, gait, and voice identification.
3. Token based: are identification system such as driving licence.

II. PROBLEM FORMULATION

Before embarking on the task of singer identification, it is imperative to define precisely the goals of this work. Voice recognition is a speaker identification technology based on behavioural characteristics and physiological of person's voice. MFCC is a commonly used method in voice/speaker recognition.

- Identification of voice of every individual singer is quite difficult. To make these human computer interaction systems accurate is big challenge.
- To prepare a system for automatic identification of the singer from a set of musical samples.
- There are various other methods like MFCC, FFT ect which are used to get better result in voice/speech/speaker identification but we used GFCC which gave much better result than other.

III. PRESENT WORK

Robust Singer Identification system

Identifying the underlying speaker is known as SID[32]. Verifying if a claimed speaker is indeed the underlying speaker is speaker verification (SV). In addition, speaker recognition is categorized into text-dependent and text-independent recognition based on whether to assume the knowledge of written text. Speaker features encode speaker specific characteristics, and are extracted from time domain signals. Commonly used speaker features include short-time spectral/cepstral features. Short-time features are generally derived from short-time Fourier transform (STFT). Time domain signals are broken into frames with around 20 ms duration. STFT is applied to the frames to obtain magnitude spectrum. First-order and second-order delta features fall into this category. We used GFCC's and its deltas for feature extraction because of its more robustness to noise than MFCC's. Singer models are built from speaker features. After this the features are ready to feed into artificial neural network. The ANN did training and testing and classifies the input voices in different classes of singers.

System design

System design consists of two phases. All two phases are explained in detail in this section. Phases in singer identification are given below.

1. Feature extraction using CMN normalized GFCC's and generation of deltas and double-delta features.
2. Singer classification using ANN.

IV. GFCC FEATURE EXTRACTION

Steps in GFCC feature extraction

1. Pass input signal through a 40-channel gammatone filter bank. There are three stages in this step

- Pre-emphasis stage
- Windowing
- Applying gammatone filter bank

A. Pre-emphasis stage:-

The idea of pre-emphasis is to spectrally flatten the speech signal and equalize the inherent spectral tilt in speech [34]. Pre-emphasis is implemented by a first order FIR digital filter. The transfer function of the pre-emphasis digital filter is given by the following equation (1)

$\mathrm{Hp}(z)=1–a^{-1}$ …………..(1)

Where a is a constant, which has a typical value of 0.97.

After applying Pre-emphasis the signal will become as shown in below figure.

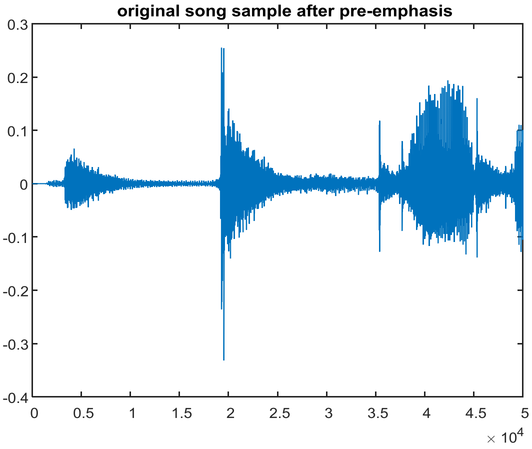


Figure: Song sample after pre-emphasis

B. Windowing:-

This stage consists first to subdivide a speech sequence into frames. The windowing function used is the Hamming window given by equation (2), which aims to reduce the spectral distortion introduced by windowing.

$$w[n] = \begin{cases} 0.54 - 0.46\cos(\frac{2\pi n}{2N-1}), 0 \le n \le N-1 \\ 0 \quad\quad otherwise \end{cases}$$

………2

C. Applying gammatone filter bank

The Gammatone filter bank consists of a series of band pass filters, which models the frequency selectivity property of the human cochlea. The impulse response of each filter was introduced by Patterson [35], as shown in the following equation (3)

$$g(t) = at^{n-1}e^{-2\pi bt}\cos(2\pi f_c t + \varphi)$$

……….(3)

2. At each channel, fully rectify the filter response (i.e. take absolute value) and decimate it to 100 Hz as a way of time windowing. Then take absolute value afterwards. This creates a time frequency
(T-F) representation that is a variant of cochleagram.
3. Take cubic root on the T-F representation
4. Apply DCT to derive cepstral features
5. Take first 2-14 cepstral features as GFCC features
6. Apply CMN normalization

CMN Normalization

- Cepstral Mean normalization ensures that the values in the feature vectors have zero mean and unit variance
- Then to remove noise, we must simply subtract of each cepstral coefficient the average of all cepstral coefficients characterizing the analysed speech signal.

cmn(i,:)=((gfcc1(i,:))mean(gfcc1(i,:)))/std(gfcc1(i,:))

7. Take first derivative and second derivatives of normalized GFCC features and use them as feature set for further process.

$$\sigma^2 = \frac{1}{N}\sum_{n=1}^{N}(c(n)-\mu)^{\ 2}$$

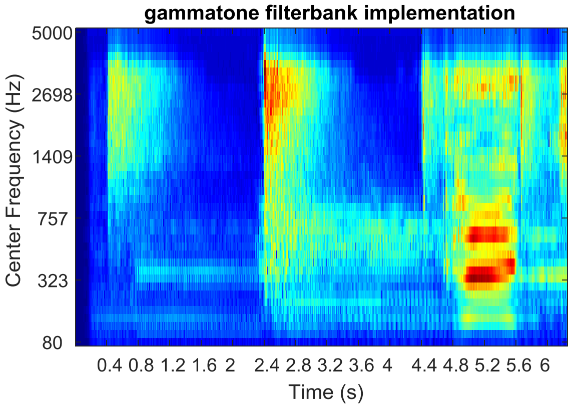


Figure.8: cochleagram showing response after applying gammatone filter bank to signal.

The cochleagram images can be analysed by image processing techniques to extract information that is not so directly accessible through audio analysis. Above is a cochleagram showing the response of gammatone-filter bank on fft channels of the speech. The RED or high intensity tells the portions where there is human voice and its power in terms of energy and blue portions gives the least response to the frequencies and energies present. As we see that in this signal centre frequencies from range about 80-5000 are present in the original signal. The file extraction process is given below.

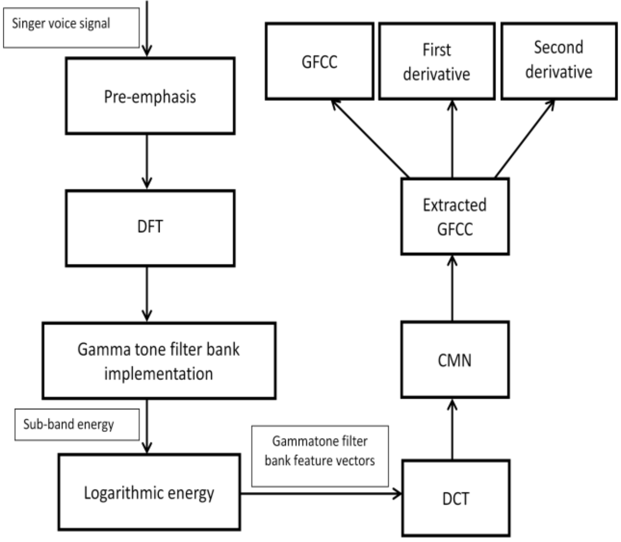


Figure.3: Flowchart of feature extraction process

BACK-PROPAGATION ALGORITHM

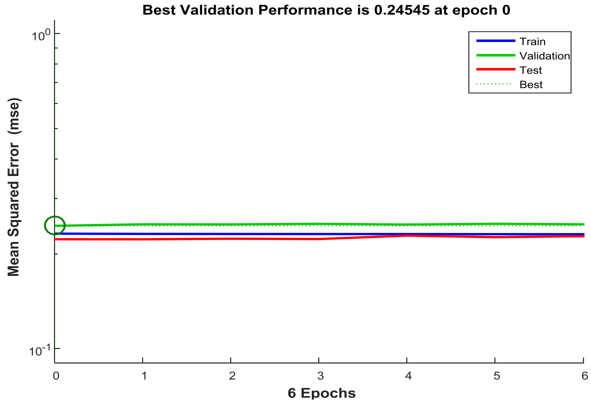
- The process of determining the error rates of each neuron that impact the output is called back propagation.
- The Back Propagation network to be the quintessential Neural Net.
- Back Propagation is the training or learning algorithm rather than the network itself.

The neurons of the input layer are fully connected to the hidden layer and the outputs of the hidden layer are fully connected to the output layer.

V. RESULT

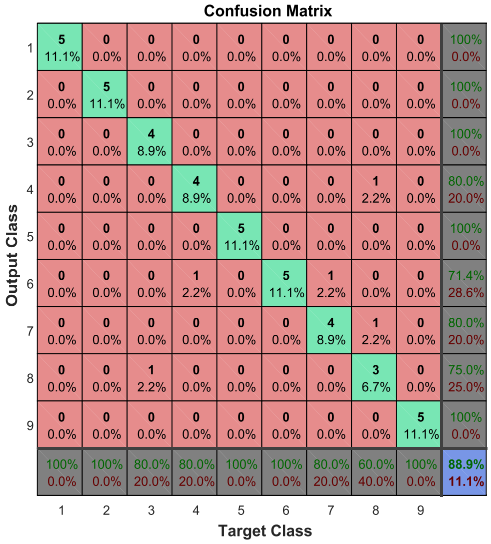
After extracting feature vector having 39 elements, I design a singer recognizer which learns relationships between features using 3-layered artificial neural network system.

Analyses of Neural Network Performance after Training



Confusion matrix

In our work, only five samples out of 45 samples misclassified with wrong singer. Confusion matrix shows 89 % percentage in overall performance.


Confusion Matrix

Output Class	1	2	3	4	5	6	7	8	9	
1	5 11.1%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
2	0 0.0%	5 11.1%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
3	0 0.0%	0 0.0%	4 8.9%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
4	0 0.0%	0 0.0%	0 0.0%	4 8.9%	0 0.0%	0 0.0%	0 0.0%	1 2.2%	0 0.0%	80.0% 20.0%
5	0 0.0%	0 0.0%	0 0.0%	0 0.0%	5 11.1%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
6	0 0.0%	0 0.0%	0 0.0%	1 2.2%	0 0.0%	5 11.1%	1 2.2%	0 0.0%	0 0.0%	71.4% 28.6%
7	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	4 8.9%	1 2.2%	0 0.0%	80.0% 20.0%
8	0 0.0%	0 0.0%	1 2.2%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	3 6.7%	0 0.0%	75.0% 25.0%
9	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	5 11.1%	100% 0.0%
	100% 0.0%	100% 0.0%	80.0% 20.0%	80.0% 20.0%	100% 0.0%	100% 0.0%	80.0% 20.0%	60.0% 40.0%	100% 0.0%	88.9% 11.1%

Target Class

Figure.18: Final Output in the form of confusion matrix

Therefore from confusion matrix above we have plotted the graphs as below

Singer Name	Total	Classified Correctly	Classified wrongly
Alka Yagnik	5	5	0
Arijit Singh	5	5	0
Kumar Sanu	5	4	1
Lata Mangeshkar	5	4	1
Mohammad Rafi	5	5	0
Rahat Fateh Ali	5	5	0
Shreya Ghoshal	5	4	1
Sonu Nigam	5	3	2
Sunidhi Chauhan	5	5	1

A database of 9 singers has been collected in which five sample of five different songs are chosen randomly. First of all 13 GFCC features were extracted using 40 gamma tone filter banks. After that CMN has been applied for normalization of GFCC features. Using extracted GFCC features, first order and second derivative features are also extracted. For classification, ANN has been applied in which training and testing has been done on GFCC feature sets. The ANN classifier classified 88.9 % of singers correctly. As it is text independent data has been used and there is background music also considered. Results can be improved in future work, if one separates the music and vocals of the singers.

VI. LITERATURE SURVEY

Many researchers have been done on extorting the different features of the speech and then classify these features with different classifiers. The common used features are MFFC. This Section explains about the previous research work done in Speaker identification and singer identification System.

Saurabh H. Deshmukhet et al[1] used K-means clustering so that hey can reduce the computational complexity of the system. It has been used for classification of the singers of northen Indian classical vocal music. K-means classifier combined with traditional features of LPC or MFCC giving better singer recognition system.

Ananya Bonjyotsna et al[2] provided a novel method of comparison of three different classifiers for vocal/non-vocal segmentation required for SID. Used three different classifiers: Artificial Neural Network (ANN) with Feed Forward Back propagation algorithm, Gaussian Mixture Model (GMM), and Learning Vector Quantization (LVQ). After using them which showed that LVQ and FFBP have given better results.

P. G Radadia et al[3] proposed SID system on large database of Bollywood Hindi songs using Cepstral Mean Subtracted (CMS) and Mel Frequency Cepstral Coefficients (MFCC) features. They compare the performance of Gaussian Mixture Model (GMM) and 3rd order polynomial classifier which found that CMS-based features are effective to SID.

Vikram C. M et al [5] proposed that GMM-UBM gives better result compared to GMM. GMM will gives accuracy of 61.84%, whereas GMM-UBM gives accuracy 75.29%.

Kekre, H.B. et al [7] proposed the use of a MFCC feature extraction with KMCG for feature vector generation & matching in speaker recognition system. KMCG is fast and simple algorithm. The proposed system gives moderate EER of 84%.

VII. CONCLUSIONS

In this we presented the introduction to the speaker emotion recognition. We have also introduced the basic concepts that are necessary in the speech signal processing and the literature survey of the past researches on the emotion recognition in speech has also been presented in this paper .thus we conclude our survey of the research papers such that many feature extraction techniques have been used by the researchers in combination with many classification techniques ,Still there is lots of work to do in the area of human computer interaction systems.

REFERENCES

[1] 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” In *International Journal of Advanced Research in Computer and Communication Engineering* IJARCCE *Vol. 3, Issue 2, February (2014) pp-5401-5406.*

[2] AnanyaBonjyotsna,ManabendraBhuyan”*Performance Comparison of Neural Networks and GMM for Vocal/Nonvocal segmentation for Singer Identification*” In International Journal of Engineering and Technology (IJET)Vol 6 No 2 Apr-May (2014) pp- 1197-1203.

[3] P.G Radadia, H.A Patil “A Cepstral Mean Subtraction Based Features for Singer Identification” In IEEE (2014) PP-58-61.

[4] A.S.Thakur1, Namrata Sahayam” S*peech Recognition Using Euclidean Distance”* In International Journal of Emerging Technology and Advanced Engineering Volume 3, Issue 3, March (2013) pp-587-590**.**

[5] Vikram C. M and Uma Rani K., ”Text Independent Classification of normal and pathological Voices using MFCCs and GMM-UBM”, I Proceedings of IEEE International conference on Information and Communication Technologies, 2013, NICHE, Kanyakumari, India pp. 980-985.

[6] X Zhao, Y Shao, D Wang, CASA-Based Robust Speaker Identification. IEEE Trans. Audio, Speech Lang. Process 20, 1608–1616 (2012)

[7] Kekre, H.B.; Bharadi, V.A.; Sawant, A.R.; Kadam, O ”Speaker recognition using Vector Quantization by MFCC and KMCG clustering algorithm”, Published in: Communication, Information & Computing Technology (ICCICT), 2012 International Conference on Date of Conference: 19-20 Oct. 2012, Page(s): 1 – 5.

[8]] Martinez, J. et al. "Speaker recognition using Mel frequency Cepstral Coefficients (MFCC) and Vector quantization (VQ) techniques" in Electrical Communications and Computers (CONIELECOMP), IEEE2012.

[9] Pawan K. Ajmera, Raghunath S. Holambe” *Speaker Recognition Using Auditory Features and Polynomial Classifier*” In International Journal of Computer Applications (2010) Volume 1– No. 14 pp-86-91.

[10] P.Chang*”* Pitch Oriented Automatic Singer Identification in Pop Music” In IEEE International Conference on Semantic Computing(2009)DOI 10.1109/ICSC.2009.28pp-161-166.

[11] Wei Cai, Qiang Li, Xin Guan”Automatic Singer Identification Based on Auditory Features” in Seventh International Conference on Natural Computation IEEE (2011) pp-1624-1628

[12] L. Regnier, G. Peeters“*SINGER VERIFICATION: SINGER MODEL .VS. SONG MODEL*”in ICASSP International Conference on Acoustics,Speech,Signal Processing IEEE(2012) pp-437-440

[13] Patil A. H., Purushotam G. R., Basu T. K.”*Combining Evidences from Mel Cepstral Features and Cepstral Mean Subtracted Features for Singer Identification”* In International Conference on Asian Language Processing(2012) IEEE, DOI 10.1109,pp-145-148

[14] Saurabh H. Deshmukh,S.G. Bhirud ” *North Indian Classical Music’s Singer Identification by Timbre Recognition using MIR Toolbox*” In International Journal of Computer Applications Volume 91 – No.4, (April 2014) pp-1-5.

[15] Tsung-Han Tsai, Yu-Siang Huang, Pei-Yun Li ,De-Ming Chen “*Content-based singer classification on compressed domain audio data*” In Springer 10 July (2014) DOI 10.1007 pp-1489–150.