International Journal of Latest Engineering Research and Applications (IJLERA) ISSN: 2455-7137

Volume – 02, Issue – 04, April – 2017, PP – 51-56

A Review: Automatic Singer Identification and Analysis of Musical Performance

Kumari Rambha Ranjan[1], Kartik Mahto[2], Dipti kumari[3]
[1]*(Dept. of Electronics and Communication, Birla Institute of Technology, Mesra,Ranchi, India)*
[2]*(Dept. of Electronics and Communication, Birla Institute of Technology, Mesra, Ranchi, India)*
[3]*(Dept. of Electronics and Communication, Birla Institute of Technology, Mesra,Ranchi, India)*

Abstract: This paper aims to review/survey an automatic singer identification and classification of musical performance and also classified by extracting audio features from real sample sounds. This study provides an overview of various feature and model-based approaches developed in past for robust singer voice identification. The importance and discussion of some standard methods applied for robust singer voice verification/identification tasks have been highlighted. The main focus is to summarily introduce popular state-of-the-art techniques adopted for enhancing singer voice verification performance in noisy conditions.
Keywords: singer identification, MFCC, ANN, SVM, feature extraction etc.

I. Introduction

Automatic singer voice identification is a subtask of for the musical content identification, automatic singeridentification is important for easy to access essential data from vast amount of data. It could be treated as the first step in developing automatic musical transcription and multimedia database annotation. The use of real world recordings should be encouraged in voice recognition [3].Now days there is rapidly growth in amount of digital media, so the need for an effective data management is challenging task. In this context, we present a study aboutautomatically identifying musical instruments from music using classifiers. Information related musical instrumentation is among the most important semantic concepts that humans use for communicate musical meaning [4][5]. Musical instrument belongs to wide range of devices with different characteristics that includes physical aspects, different sound initiation process.So classifying musical instrument becomes a complex issue because of multidimensional nature of musical instruments. Hence the need of automated system is arises to classify musical instruments automatically[6][7] Musical instrument identification using SVM and MLP classifiers helps to reduce manual work to identify and classify musical instruments according to their attributes.

II. Literature Survey

In this segment we presented the literature on the automatic singer identification. We have also introduced the basic concepts that are necessary in the signal processing and the literature survey of the past researches on the singer voice recognition in song sample 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.

B. Whitman, G. Flake [1]In this paper author present an increasing interest in the developmentof frameworks for singer identification and associated topics. Among the earliest of such systems, Minnowmatch mainly focuses on artist rather than singer identification using mel-frequencycepstral coefficients (MFCCs), which is a feature, adapted fromthe classical speech recognition and speaker identification mechanisms.The best identification accuracy achieved on a small dataset, containing a 10 artist set, is approximately 70%. However, with a larger set of 21 artists, the best case accuracy significantly drops to 50%.

A. Berenzweig, D. P. W. Ellis [2], the authors use a neural network trained on radio recordings to similarly segment songs into vocal and non-vocal regions. By focusing on voice regions alone, they improved artist identification by 15%. The system present here also attempts to perform segmentation of vocal regions prior to singer identification. After segmentation, the classifier uses features drawn from voice coding based on Linear Predictive Coding.

C. C. Liu and C. S. Huang [3], a novel scheme is designed and developed to automatically classify music objects according to their singers. First, the coefficients extracted from the output of the polyphase filters are used to compute the music features for segmentation. Based on these features, a music object can be decomposed into a sequenceof notes (or phonemes). Then for each phoneme in the training set, its music feature is extracted and used to train a music classifier which can identify the singer of an unknown music object.

T. Zhang [4] developed a system for automatic singer identification which recognizes the singer of a song by analyzing the music signal. The proposed scheme follows the framework of common speaker identification systems, but special efforts are made to distinguish the singing voice from instrumental sounds in a song. A statistical model is trained for each singer's voice with typical song(s) of the singer. Then, for a song to be identified, the starting point of the singing voice is detected and a portion of the song is excerpted from that point. Audio features are extracted and matched with singers' voice models in the database.

W. H. Tsai and H. M. Wang [5] proposed a solo voice modeling framework to capture singers' vocal characteristics. The technique firstly separates vocal from non-vocal regions and then models the singers' vocal characteristics based on stochastic properties of the background music. However, the main weakness for the method is that it uses only a single type of acoustic feature for vocal portions (20 dimensional Mel-Frequency Ceptral coefficients) to profile differentsingers. Based on their experiment on a small dataset of 230 popular music songs, the accuracy of identification is only 71%. At this point, it is the state-of-the-art in singer identification.

Wei-Ho Tsai and Hsin-Chieh Lee [6]Automatic singer identification (SNID) aims to determine who among a set of singers perform a given music recording. So far, most existing SNID methods follow a framework stemming from speaker identification (SPID) research, which models each person's characteristics using his/her voice data. This framework, however, is impractical in many SNID applications, because acquiring solo a cappella from each singer is usually not as feasible as collecting spoken data from each speaker in SPID applications. In view of the easy availability of spoken data, this work investigates the possibility of modeling singers' voices using spoken data instead of singing data. This study has investigated the feasibility ofcharacterizing singers' voices using their spoken data forSNID. Our experiment found that a GMM derived from aperson's speech utterances usually cannot well characterizehis/her singing voice, owing to the significant differencebetween singing and speech. To overcome this problem, wehave proposed a MAP-adaptation-based method to bridgethe difference, so that a test singing recording can beidentified using the speech-derived singer models.

Namunu Chinthaka Maddage1, Changsheng Xu1, Ye Wang[7]In this paper, we propose a novel method to identify the singer of a query song from the audio database. The database contains over 100 popular songs of solo singers. The rhythm structure of the song is analyzed using our proposed rhythm tracking method and the song is segmented into beat space time frames, where within the beat space time length the harmonic structure is quasi stationary. This inter-beat time resolution of the song is used for both feature extraction and training of the classifiers (i.e. Support Vector Machine (SVM) for vocal/instrumental boundary detection and Gaussian Mixture Models (GMMs) for modeling the singer). Combining the instrumental music similarities in the songs of the same singer with the vocal model can improve the identification of the singer with an accuracy of over 87%.

B Whitman, G Flake, S Lawrence [8]MFCC featurevectors and artificial neural network classifier are used to identify playback singer from a database. An accuracy of 70 % is achieved by this system using 10-artist database. Instrumental and singing sounds were not separated in the system.

TL Nwe, H Li [9]Singer's vibrato-based octave frequency cepstral coefficients (OFCC) are used in singer identification. Experiments were performed using 84 popular songs only from 12-singer database. An average error rate of 16.2 % is achieved in segment level identification.

Shruti, Bharti Chhabra [10] 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%.

III. Methodology Used

A. SVM

SVM is strong classification algorithm because it is simple in structure and it requires less number of features. Fig. 1 describes the Framework of SVM. SVM currently considered the most efficient family of algorithm in machine learning because it is computationally efficient and robust inhigh-dimension .Support vector machine is trained with dataset of musical instrument. During the training of SVM feature extractor converts each input value to feature set and these feature sets capture basic information about each input. And these sets are used for classify the feature unit.

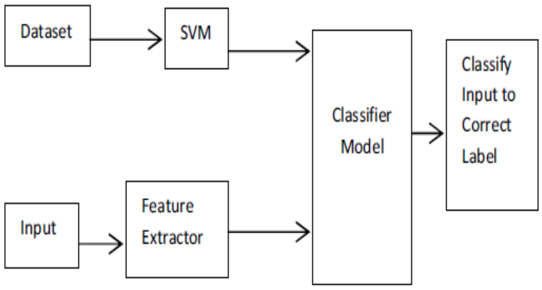

Figure: 1 Framework of SVM Classifier

B. MLP

In this system MLP is used for classification of musical instruments. It consists of two input neurons in first layer and hidden layer consist of six neurons with one output neuron. MLP uses batched back proposition algorithm. Activation functions for MLP. MLP uses supervised training on multiple layers of interconnected perceptrons. MLPs contain at least one layer of hidden neurons.

C. Feature selection and extraction

As commented before, each specific pair of instruments has a feature associated with it. Therefore, the features to be extracted depend directly on which instruments are being considered. Four instruments - alto saxophone (S), violin (V), piano (P) and acoustic guitar (G) - were chosen to validate the strategy. Hence, there are six possible pairs of instruments. The instruments in the pairs SP, SG, VP and VG have considerably different characteristics (temporal waveform, spectral content, etc), while the instruments in the pair SV have some similar characteristics and the instruments in the pair PG are closely related, as discussed in [13]. In this way, the technique can be tested under different levels of difficulty.

D. MFCC

MFCC are the most useful coefficients which are used for speech recognition because of their ability to represent speech amplitude spectrum in a compact form. Figure 2 shows the process of creating MFCC features [7,]. Speech signal is divided into frames by applying a hamming windowing function at fixed intervals. Cepstral feature vectors are generated using each frame.

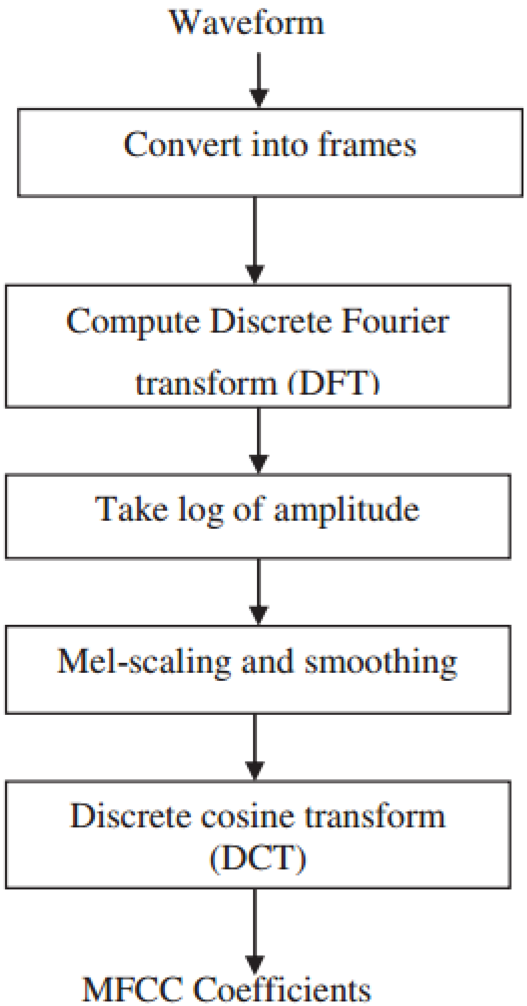

Figure 2: The process of creating MFCC [7] coefficients

The final step is to compute the discrete cosine transform (DCT) of the log filter bank energies in MFCC. But only 12 of the 26 DCT coefficients are kept because the higher DCT coefficients represent fast changes in the filter bank energies which reduce the performance of system. So it gives small improvement by dropping them. Thirteen MFCC coefficients (×28 to ×40) are used by our proposed approach which is extracted for each segment from IVS.

E. ANN (Artificial neural network)

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.

We used MFCC's and its deltas for feature extraction because of its more robustness to noise than other technique. 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.

IV. Importance of Research

Song, Sound, Music, musical Instrument and Voice are very fascinating areas of research for digital signal processing applications.

1. Musical Instrument Identification [1], Singer Identification [2] [3], Speaker Recognition [4], Music Melody Extraction, Music Database Indexing are some promising areas of applications of Music Information Retrieval (MIR)research.
2. In most popular music, the singing voice section captures one of the most important characteristics, and thus detection of the singing voice has attracted a number of researchers in a music information retrieval (MIR) community for many years. Detection of singing voice or vocal/non-vocal discrimination can be used in many other related applications in MIR such as singer identification, singing voice separation, query-by-humming (QBH), structural analysis of music or lyrics-audio alignment, and so on.
3. Music applications, especially for vocal/non-vocal discrimination because it requires a great deal of human labor to manually annotate vocal/non-vocal boundaries as he/she listens to a number of audio files.

V. Discussion

In this paper, we focused on various techniques which are used in identification of singer voice. Two separate ANNs are develop for language and Speech recognition system and trained using back propagation algorithm and radial basis function network. Here use of MFCCs and delta-MFCCs as acoustic features. The present work is limited to small vocabulary and four languages. In future we will develop system for large vocabulary and more languages. The use of vocal and instrumental features and ANN can provide successful music classification. To provide benefit of Information technology in each and everywhere in India, voice based interface for various computer related task is most suitable. To develop a voice based interface for countries like India is very difficult task as here approximate twenty five languages spoken. The Speech signal conveys many levels of information like what is spoken, language, speaker, gender, sentiments etc.

VI. Conclusion

In this review, the basic techniques used in voice identification of singer are discussed and its recent researches and work is highlighted. Some approaches available for developing a Singer Identification system are clearly explained with its merits and demerits, techniques of feature extraction, model training. The performance of the Singer Identification system based on the adapted feature extraction technique and the speaker identification and recognition approach for the particular individual is compared in this paper. In researches it has been seen that ANN approach along with MFCC features is more suitable for these requirements and offers good recognition result. Where ever the above combination will be used to identification will help to generate large powerful systems.

References

[1]. B. Whitman, G. Flake, and S. Lawrence.Artist detection inmusic with minnowmatch.In IEEE Workshop on Neural Networks for Signal Processing, 2001.

[2]. A. Berenzweig, D. P. W. Ellis, and S. Lawrence.Using voice segments to improve artist classification of music. In AES 22nd International Conference on Virtual, Synthetic and Entertainment Audio, pages 119–122, 2002.

[3]. C. C. Liu and C. S. Huang. A singer identification technique for content-based classification of mp3 music objects. In Proc.OfACMCIKMInt. Conference on Information and Knowledge Management, pages 438–445, 2002.

[4]. T. Zhang. Automatic singer identification. In Proc. of IEEEInt. Conference on Multimedia and Expo (ICME), pages 33–36, 2003.

[5]. W. H. Tsai and H. M. Wang.Automatic singer recognition of popular music recordings via estimation and modeling of solo vocal signals. IEEE Trans. on Speech, Audio and Language Processing, 14(1):330–341, 2006.

[6]. Wei-Ho Tsai and Hsin-Chieh Lee "Automatic Singer Identification Based on Speech-Derived Models" *International Journal of Future Computer and Communication, Vol. 1, No. 2, August 2012*

[7]. Namunu Chinthaka Maddage1, 2, Changsheng Xu1, Ye Wang2 "Singer Identification Based on Vocal and Instrumental Models"

[8]. B Whitman, G Flake, S Lawrence, Artist detection in music with minnow match, in Proceeding of the 2001 IEEE Workshop on Neural Networks for Signal Processing, 2001, pp. 559–568

[9]. TL Nwe, H Li, Exploring vibrato-motivated acoustic features for singer identification.IEEE Trans. Audio Speech Lang. Process. 15(2), 519–530 (2007)

[10]. Shruti, Bharti Chhabra "An Approach for Singer Identification Technique Using Artificial Neural Network" Volume 5, Issue 11, November 2015

[11]. Corneliu Octavian DUMITRU, Inge GAVAT, "A Comparative Study of Feature Extraction Methods Applied to Continuous Speech Recognition in Romanian Language", 48th International Symposium ELMAR-2006, 07-09 June 2006, Zadar, Croatia.

[12]. DOUGLAS O'SHAUGHNESSY, "Interacting With Computers by Voice: Automatic Speech Recognition and Synthesis", Proceedings of the IEEE, VOL. 91, NO. 9, September 2003, 0018-9219/03$17.00 © 2003 IEEE.

[13]. D. Fragoulis, C. Papaodysseus, M. Exarhos, G. Roussopoulos, T. Panagopoulos, and D. Kamarotos, "Automated classification of piano-guitar notes," IEEE Trans. Audio Speech Lang. Process., vol. 14, pp. 1040–1050, 2006.