

AUDIO DESCRIPTIVE ANALYSIS OF SINGER AND MUSICAL INSTRUMENT IDENTIFICATION IN NORTH INDIAN CLASSICAL MUSIC

Saurabh H. Deshmukh¹, Divya P. Bajaj², S.G. Bhirud³

¹Research Scholar, IT Dept, G.H Rasoni COE&M, Wagholi, Pune, India

²Student, IT dept, G.H Rasoni COE&M, Wagholi, Pune, India

³Comp Engineering Dept, VJ Tech. Institute, Mumbai, India

Abstract

Music information retrieval (MIR) has reached to a reasonably stable state after advancement in the Low Level audio Descriptors (LLDs) and feature extraction techniques. The analysis of sound has now become simple by the continuous efforts and research of MIR community in the field of signal processing from last two decades. In north Indian classical music, a singer is accompanied by some instruments such as harmonium, violin or flute. These instruments are tuned in the same musical scale (pitch range) in which the singer is signing. Separate researches have been made in recent past to identify a musical instrument and a singer. In this paper, we have analyzed the low level audio descriptors, for singing voice and musical instrument sound together, that appears to human ear as similar with respect to 'timbre', to see if we could treat them same and use identification/ classification routines to classify them into their classes. We have used Hybrid Selection algorithm from wrapper technique (the one that uses classifier also in feature selection process) to identify and extract the features and K-Means and K nearest neighbor classifiers to classify and cross verify the accuracy of classification. The accuracy of classification achieved was 91.1% which clearly proves that musical instruments and singing voice that sounds similar in timbral aspect can be grouped together and classification is possible with mixed database of instruments and singing voices.

Keywords: Music Information Retrieval (MIR), Timbre, Singing Voice, Low level Descriptors (LLD), North Indian Classical music. MIRTOOL BOX

1. INTRODUCTION

Sound, Music, musical Instrument and Voice are very fascinating areas of research for digital signal processing applications. In past four decades, 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. For such applications, and many more, the basic necessity is to extract the useful musical information from the audio files and manipulate (mean, sum, normalize, derivative etc.) or to represent that information in some useful way (transformations, such as Mel domain, Fourier transform, Wavelet transform etc.).

With the advancements in the computational capabilities of MatLab, there are many parallel developments in the supportive toolboxes that speed up the research, [5], [6], [7], [8]. These toolboxes possess strong functions to correctly separate musical information and process it for further applications. Broadly, the research in this field is categorized into voice and instruments. The voice is further classified into speaking voice (speech) and singing voice.

In this paper, we have combined the fields of singing voice identification and musical instrument identification. The aim is to analyze the similarities and/ or the differences in the

audio descriptions of the singing voice and the musical instruments with respect to the timbre.

2. THE EXISTING SYSTEM

Any sound is characterized by an entity called as Timbre. We can identify in a cartoon movie, voices of two characters, by the timbre of the sound of each character, even if both the voices are generated by the same person. There are singers who can emit voices of different singers in one song. How simply is this possible? This becomes possible because of the same entity called timbre. Timbre is multi-dimensional, fuzzy, undefined, unit less attribute of sound that uniquely identifies it [1]. Timbre is refereed in accordance with, texture, light or color. It is also called as Tone color. The typical examples of timbre names are "Rough, Smooth, Dark, Dim, Silky, Heavy" etc.

Let it be a musical instrument or a singing voice, timbre plays very important role in categorizing the sounds. Timbre of a singer depends upon physical anatomy of the singer's body, face, lungs etc. as well as from psychoacoustic studies of singer's psychological state of mind. For North Indian classical music vocal the timbre also depends upon the type of singing such as Bada khayal, Chota Khayal, Lakshangeet, Bandish or Tarana [9] while on the other hand for Musical Instruments playing classical music the timbral aspect of that instrument differs from unit to unit.

In this research, we have considered North Indian Classical Music vocal and musical instrument solo performances. For a monophonic (that includes single melodic line) or Polyphonic (two or more melodic lines with equal importance and different scales) singer identification problem the voice separation is easier. While for North Indian Classical Music which is Homophonic (single melody line accompanied by background instruments such as Tanpura/Harmonium that is played continuously as the melody progresses) the separation of voice from the background voice is challenging task. The output sound is a mixture of all sounds generated together giving a feeling of harmonious sound which is difficult to segregate into parts or layers.

Similar to singing voice there exist timbre for different musical instruments. The North Indian Classical music considers, for example, Flute, Violin, Harmonium, Sarod, Sitar, Santoor, Shehnai, as leading Classical Instruments that are separately played as performing art on stage. These instruments are used in similar way a singer sings a classical raga. From instrument identification perspective a lot of research has been made towards identification of instrument in an orchestra recording, or simple instrument identification by identification of its timbre [1], [10]. The usual categorization or classification of the instruments is done into Flute, Clarinet, Violin, Trumpet etc. or the classification is into woodwind instruments, string instruments, bow instruments etc. Very negligible research has been made into identification of which flute? Which violin? Which trumpet? If, for example, we have five violins available, do we have the capability in our system that uniquely identifies the instrument units (physically separate items) into Violin 1, Violin 2, and Violin 3 etc?

Researches, often announce usage of singing voice as a kind of musical instrument by saying that human voice: the most natural of all musical instruments that is most widely used [11]. But on the other side there is less research and applications to detect separate 'Unit' (body) of instrument rather than just type of musical instrument. The musical instruments are also characterized by their timbers. The flute sounding husky and used for raga containing minor notes of music are definitely different than the flute sounding divine for raga containing major notes. In this research we have used flutes and Violins performances from North Indian Classical Music.

It is to be mentioned here that the classical performances of the musical instruments playing North Indian classical Music are very similar to the performances of classical singers of this kind of music. A classical singer uses Tanpura/harmonium as an accompanying instrument and sings initially long notes contained in the raga. Then, after completion of almost all the notes described by the raga, that prolonged and sung sufficiently, the rhythm starts slowly and then gradually the speed of the rhythm and singing and the rate of change of musical note sung per unit of time increases accordingly. Similar performance is observed with classical performances of musical instruments such as Flute or Violin.

3. THE SYSTEM DESIGN

In the system, we have used two databases each containing 100 files with .wav format recordings each of 5 sec duration, 16 bit PCM with 11,025Hz sampling frequency. DB1 contains 5 different male singers and 5 different flute instrument (different units) recordings with 10 samples of each. DB2 contains 5 different female singers and 5 different violin (different units) recordings with 10 samples of each. Here, we are treating all ten instruments as human voice or in other way we are treating all ten singers as musical instruments. The arrangement of dataset done this way has many purposes to be served. First, the system shall be tested on audio that contains mixture of musical instruments and human voices. Second, the performance, importance and role of timbral audio descriptors [12] can be verified and tested. Then, the comparison of performance of statistical classifiers can be done on the basis of mixed input.

Also, we wish to confuse the recognition system with respect to, whether it is a singer or an instrument. By this way in DB1 or in DB2, if we have total 10 singers and if singer 7 is recognized as singer 7 for example, then that means a violin unit of number 7 has been correctly identified as violin number 7 and singer 6 is recognized as singer 6. Thus by this arrangement of audio input this makes the same system to work as an instrument classifier as well as singer identifier.

In order to study the audio descriptive aspect of this above stated system a hybrid selection algorithm [13] is applied only on the audio descriptors, described under the taxonomy of timber of [14]. As per MPEG standard there are more than 50 audio descriptors and can be classified according to variety of taxonomy depending upon the point of view [10] [15]. In musical instrument identification research [16] taxonomy of audio descriptors is followed that includes following audio descriptors under the title of Timbre in MIRtoolbox [14]. The audio descriptors are: attack time, attack slope, Zero crossing rate, Roll off, brightness, roughness, MFCC (Mel frequency Cepstral Coefficient) and Irregularity. The attack time and slope are not useful for harmonic sound samples hence are neglected giving total 6 audio descriptors including MFCC. Experiments have been carried out on number of coefficients to be considered for MFCC from 10, 20 to 30 and was found that 20 coefficients are sufficient to carry useful information of the audio sample. K means (vector quantization technique) and Statistical classifier, K nearest neighbor have been used to identify the given sound sample into musical instrument or human voice. Here the identification is not merely classification into two classes such as an instrument or a singer. Here the classifiers are supposed to identify the exact entity as which singer number or which musical instrument number, even when they sound timbrally similar to each other. From each database, total 10 singers (and instruments) are trained and tested in which 70% audio samples are used for training and 30% samples are used for testing using Euclidian distance.

4. EXPERIMENTS AND RESULTS

Experiments are carried out for all six audio descriptors using hybrid selection algorithm. For each level of the forward pass the combinations of best audio descriptors are selected and the statistical classifier is trained and tested for performance accuracy. In backward pass the combinations giving less accuracy have been eliminated. Table 2 shows the final combination of audio descriptors giving highest accuracy for both the databases.

Table 1: Audio database and its description

Data	Database Description
DB1	Violin and Female mixed Samples
DB2	Flute and Male mixed Samples

Table 2: Accuracy comparison of K-means and KNN classifiers with contributing audio descriptors

Classifier	Accuracy %	Contributing audio descriptors
K -Means DB1	88.3	MFCC, Roughness, Irregularity, Brightness
KNN DB1	91.1	MFCC, Roughness, ZCR, Roll off
K -Means DB2	69.0	MFCC, ZCR, Rolloff, Irregularity
KNN DB2	90.5	Rolloff, Brightness, MFCC, Roughness

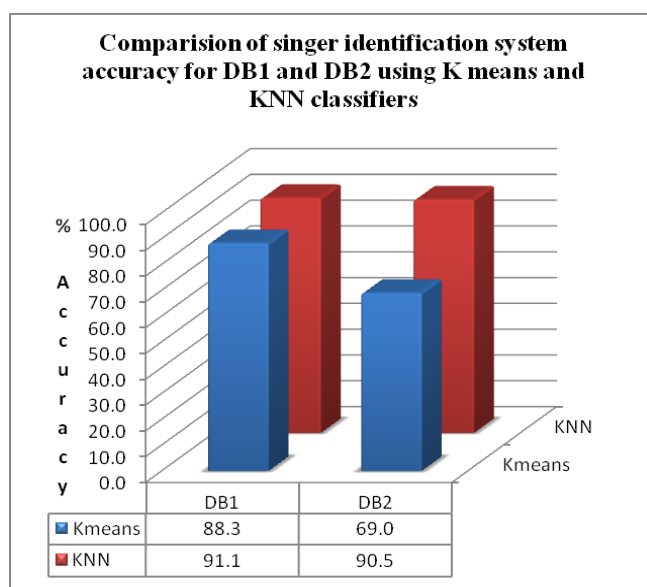


Fig 1: Graphical representation of singer/musical instrument identification accuracies obtained through K means and KNN classifiers.

5. CONCLUSION

Singer identification and Musical Instrument identification system has been implemented using hybrid selection algorithm for the selection of audio descriptors, as explained in MIRTOOL box. Two classifiers are used to identify the contributing audio descriptors and then to test the identification accuracy. The major conclusions drawn from the experiments and the results are as follows:

- The audio descriptors considered under Timbral category of MIRTOOL box [14] are sufficient to identify a singer or a musical instrument.
- MFCC, as in case of [1] and [4] contributes most out of other audio descriptors used under timbre category.
- In case of KNN classifier, for female singer database DB1, ZCR plays vital role in singer identification while for male singer database DB2, brightness plays important role.
- Out of K means and KNN classifiers, the statistical classification of KNN using Euclidian distance is better choice. For k means algorithm, for every new run the centroids are initialized for random values thus making it less effective.
- MFCC, Roughness and roll off are essential audio descriptors that when are used give better identification accuracy.

Overall 91.1% identification accuracy is achieved using KNN classifier and MFCC, Roughness, ZCR and Rolloff audio descriptors for the identification of female singers and violin samples. Similarly 90.5% accuracy is achieved for male singers and flute audio samples using KNN classifier and Rolloff, Brightness MFCC and Roughness audio descriptors.

REFERENCES

- T. H. Park, "Towards Automatic Musical Instrument Timbre Recognition," New Jersey, USA, 2004.
- M. & H. P. Rocamora, "Comparing Audio Descriptors for Singing Voice Detection in Music Audio Files", in 11th Simpósio Brasileiro de Computação Musical (SBCM07), Soo Paulo, Brazil, 2007.
- M. A. Bartsch, "Automatic Singer Identification in Polyphonic Music", Michigan, 2004.
- B. S. Atal, "Automatic Recognition of Speakers from their voices", PROCEEDINGS OF THE IEEE, vol. VOL. 64, no. NO. 4, pp. 460-475, APRIL 1976
- K. Nis, "Audio Database Toolbox", 21 Apr 2009 (Updated 25 Jun 2009).
- G. Peeters, B. L. Giordano, P. Susini, N. Misdariis and S. McAdams, "The Timbre Toolbox: Extracting audio descriptors from musical signals," The Journal of the Acoustical Society of America DOI:10.1121/1.3642604, vol. 5, no. 130, pp. pp 2902-16., March 2011.
- D. Wegmann, "DAFX Toolbox," 06 Nov 2006 (Updated 17 Jul 2007).
- M. Slaney, "Auditory Toolbox"-V2., Palo Alto, CA, 1998.
- Sadhana, 2011. [Online]. Available: <http://raag-hindustani.com/Scales1.html>.
- H.-G. Kim, N. Moreau and T. Sikora, "MPEG-7 Audio and Beyond"- Audio Content Indexing and Retrieval, Chichester, West Susscx: John Wiley & Sons, Ltd, 2005.
- K. Forney and J. Machlis, The Enjoyment of Music: An Introduction to Perceptive Listening, 11 ed., W. W. Norton & Company, 2011.

- [12] P. T. Olivier Lartillot, "A Matlab Toolbox for Musical Feature Extraction From Audio", in International Conference on Digital Audio Effects, Bordeaux, 2007.
- [13] S. Deshmukh and S.G. Bhirud, "A Hybrid Selection Method of Audio Descriptors for Singer Identification in North Indian Classical Music", in Emerging Trends in Engineering and Technology (ICETET) , Himji, Japan, 2012.
- [14] O. Lartillot and P. Toiviainen, "A Matlab Toolbox for Musical Feature Extraction from Audio", in International Conference on Digital Audio Effects, Bordeaux, France, 2007.
- [15] G. Peeters, "A large set of audio features for sound description (Similarity and Classification) in the CUIDADO Project," 75001, Paris, France, 23rd April,2004
- [16] G. R. B. D. Slim ESSID, "HIERARCHICAL CLASSIFICATION OF MUSICAL INSTRUMENTS ON SOLO RECORDINGS," in ICASSP 2006, Paris - FRANCE, 2006.
- [17] S. Cunningham, "Matlab Auditory Demonstrations" - V2, Sheffield, South Yorkshire, UK, May-1998.
- [18] G. Peeters, "A large set of audio features for sound description (similarity and classification) in the CUIDADO project," 23 April 2004.