ISSN:0975-9646

A Novel Method to Identify Audio Descriptors, Useful in Gender Identification from North Indian Classical Music Vocal

Saurabh H. Deshmukh
IT dept, G.H.Raisoni COE&M, Wagholi, Pune, India

Dr. S.G. Bhirud
Computer Engineering Dept, VJTI, Mumbai, India

Abstract: **Gender identification application of Music Information Retrieval (MIR) research requires that the audio features that are extracted from the sound files should be sufficiently strong to be able to correctly classify the singer into male or a female. In literature, a lot of such applications are developed for speaker's gender identification. Unfortunately, these applications cannot be applied for gender identification of a singer. For singer identification, concept of sound timbre along with interference of background music has to be considered. Merely applying filters to this background voice may help for polyphonic music recordings, such as, western classical or popular music, but in case of North Indian classical music, which is homophonic, the continuous interference of a supportive musical instrument such as 'Tanpura' or 'Harmonium' cannot be neglected or filtered out. In this paper, we have introduced a novel indirect method to identify the audio descriptors that are helpful in identifying the gender of a singer singing north Indian classical music. The aim is to run a two folded algorithm in which, first, we identify the singer with the help of various combinations of traditional audio descriptors under timbre taxonomy and then to backtrack and look back towards the audio descriptors that made this successful identification of the singer. A traditional K-means classifier is used. The algorithm made run on two databases, viz. DB1 containing all male singers and DB2 containing all female singers. The results showed that, for male singer along with Zero crossing rate and MFCC (Mel Frequency Cepstrum Coefficients), there are two more audio descriptors that take part in achieving highest efficiency of singer identification. These are Roughness and Irregularity. While for female these were Roll off and Brightness along with Zero crossing Rate and MFCC. The accuracy of singer identification for male and female singers is found to be 83.33% and 85% respectively.**

Keywords: *Music Information Retrieval (MIR), Audio descriptor, Gender Identification, K means. MFCC*

I. INTRODUCTION

Gender identification is one of the most rarely touched applications of Music Information Retrieval (MIR) research. The gender detection is itself a challenging task from point of view of which special information/ audio description do we infer from the audio file. In any MIR research, various routine audio feature extractors and classifiers are used to solve typical problems like singer identification, music melody finder, audio data indexing, retrieving a file from audio database on the basis of audio features of a singer and so on.

With the advancement of computational capabilities of MatLab it has become very much simple to extract and then to manipulate the audio information contained in an audio file. To add to its capabilities, there evolved many tool sets that work in accordance with MatLab such as [1], [2], and [3] to name a few. These routines have shown different perspectives and methods to extract the useful musical information from given sound file. The Music Information Retrieval (MIR) Toolbox [4], provides a wide range of energy, timbral, statistical functionality helpful for various applications where audio feature extraction is most challenging and necessary.

In gender detection, various audio features of male singer and female singer differ with respect to the aspects of physical and psychological characteristics of the singer. In many research done, till today, on gender identification of a speaker, major focus is always on the audio information that is extracted and manipulated.

For a singer, singing North Indian classical music, the process becomes much complicated when accompanying musical instruments generate frequencies that are merged in the singing voice such that it becomes impossible to extract exact voice and its features from such homogenous music. The typical accompanying instruments such as 'Tanpura' or violin/harmonium being continuously played in the same scale and almost at the same note of frequency where the singer has taken a long stay on musical note, makes it impossible at some times to know which sound is of singer and which is of Tanpura/violin.

In this paper, we have used a backward approach of recognizing the audio descriptors present in an audio recording that are responsible in deciding the gender of a singer from the results of identification of a singer, instead of directly generating output of classification of male and female singer.

The target is to identify first the singer efficiently and then to backtrack and recognize the audio descriptors that are used in the process of singer identification. The challenges are many faced. As stated above, the general scope of difference of male and female singer does not stop only at physical and psychological aspects of the singer because there are involved many other components especially with respect to the North Indian classical music performer. Following section describes the aspects of gender detection.

II. BACKGROUND

A usual speaker's gender recognition system [5] , after preprocessing of input file, pitch and formants are detected and then are given to the classifier. Similar system is used in singer recognition system where, audio features are extracted from the audio files and then the feature vector is fed to the classifier. The efficiency of the recognition system depends on the type of input music, the feature extractor and the features used and the classifier applied on the extracted features. It makes substantial difference to the accuracy of the system when input dataset is noise free.

The system also performs differently for monophonic, homophonic and polyphonic music input.

In North Indian Classical music, the traditional Indian teacher- student system of inheriting the singing knowledge is observed strictly. Where, the *Gharana* [6] that the guru belongs to, the singing style, the presentation style and most important the stylization of the performer towards throwing the sound notes while singing is strictly observed. For example, a male singer from Gwalior Gharana may adapt to a stylization of singing a raga very softly (like a female) depending upon the stylization its teacher has adopted and taught and there may also exist a female singer from Kirana Gharana that uses a rash or harsh way (sounding similar to a man) of producing the sound for the same raga.

Apart from these technicalities and specialties of North Indian Classical music, there exist huge dependencies of other attributes or circumstances in which the singer is singing. Psychology, physiology, Gharana, stylization, type of song being sung (Bada Khyal, Chota khyal, Tarana, Bandish, Thumri , Dadra etc [7]), the time of singing, the type of raga being sung, the musical scale in which the singer is singing, which accompanying instrument is present and so on, are the aspects that are to be considered while deciding the gender of a singer through computer system.

It is practically impossible to integrate so many aspects of this type of music in one research. We have used here only the timbral aspects of a singer's voice to identify the singer and then to backtracked for those audio descriptors that assured highest efficiency in identifying the singer.

III. DIFFERENCE BETWEEN MALE/FEMALE SINGING VOICE WITH RESPECT TO TIMBRE

Timbre is a non-tangible, multidimensional attribute of a sound that uniquely identifies it [8]. There may be two singers, simultaneously singing same musical note (at a frequency in Hz), with same volume (at some dB amplitude) and of same duration (in seconds of time). In such situation, a human hearing system still recognizes the sound and identifies the singer as well as its gender. This is possible only because of the presence of timbre property of the sound. A lot of research has been done on timbre and there are many attempts to catch this fuzzy word, whose descriptions are sometimes with respect to texture (such as rough, smooth, silky etc) or sometimes with respect to color (such as dull, bright, pink etc) or sometimes with respect to the loudness (loud, soft, harsh etc).

Depending upon the taxonomy used to classify the audio descriptors, the timbre includes different audio descriptors [9] [10] [11]. In MIRtoolbox the timbre includes *Attack time and attack slope, Brightness, Mel-Frequency Cepstrum Coefficients (MFCC), Zero Crossing Rate (ZCR), Roughness, Roll off* and *Irregularity.* The attack time and attack slope audio descriptors are generated from onsets and are of use of identifying the attack time of the audio sample. But since, all the audio recordings used in this data set are pre-edited in such a way that, all the energy of the signal remains present throughout the recording. This reduces the overburden and post processing of identifying the singing portion form an audio file. Thus attack time and hence attack slope are eliminated from the experiment list. All of the rest audio descriptors are used to identify correctly a singer (male/Female) and then backtrack on the efficient contribution of the audio descriptor in identifying the gender.

We have assumed throughout the research that typical frequency range of a singer cannot be restricted to its gender. Usually it is assumed that a male singer sings in low frequency and a female singer sings in high frequency. In North Indian classical music however the categorization is done with respect to the musical scale in which the singer can sing comfortably and produce all frequencies of notes in that scale range. This varies with respect to the singer, hence it cannot be assumed to have a clear bifurcation of male and female singer with respect to the frequency range. The 'pitch' audio descriptor, which does not fit into any typical taxonomy, is not considered since pitch range changes from singer to singer, depending upon the musical scale in which he/she is singing.

From human anatomical aspect, there are huge changes that happen to a boy during or after his puberty. A boy, when before puberty, was sounding like a girl or a woman, suddenly starts sounding like his dad. There are changes in his larynx (the voice box) or throat as a whole. We cannot neglect this difference between male and female in order to study which components/audio characteristics in his/her voice are actually different from each other? The larynx gets larger and thicker as the boy grows. This change is observed in both girl and a boy but it is significantly noticeable with the boy. The tone of the girl's voice only gets deepened for a couple of levels but for boys it deepens to a noticeable level. Two muscles, also called as vocal cords, work as a rubber band. These are stretched across the larynx. [12] .

The vocal cords vibrate by the air passion through it producing sound. The sound quality depends upon, the tightness of the vocal cords, the facial bones, cavities in the sinuses, the nose and back space of the throat. Overall the voice of the boy starts generating events in short durations giving it harshness, deepness and thus roughness. The upcoming section explains these anatomical differences between male and female and proposes a novel method to backtrack for the audio descriptors that are responsible for differently identifying the singers from gender point of view.

IV. THE PROPOSED ALGORITHM AND EVALUATION STRATEGY

In order to identify the audio descriptor(s) responsible for detecting the gender of a singer following algorithm is applied. The algorithm divides the process into two parts. Part A: identification of the singer and calculating the accuracy of complete singer identification process and Part B: backward detection and analysis of audio descriptors responsible for correct singer identification. We use two datasets DB1 and DB2, each containing studio, noise free recordings of popular North Indian classical singers. 10 singers with 10 samples of duration of 5 sec each are considered for DB1, where all singers are Male, and 10 singers with 10 samples of duration of 5 sec each, are considered for DB2, where all singers are female. Each

audio file is PCM compressed, 16 bit, mono channel .wav file sampled at 11025Hz sampling frequency. The experiments are carried out for 10, 5 and 3 singers trained and tested at a time. As a thumb rule each time 70% dataset is used for training and 30% for testing.

The audio descriptor selection process that is followed for forward pass is as follows. Initially, all the single audio descriptors are used in isolation and are extracted from each input file and then the features vectors are fed to the K -means clustering for classification i.e. singer identification. The K- means clustering is used here for its simplicity and effectiveness for small data. Then, efficiency of singer identification by making use of these audio descriptors is calculated and noted.

On the basis of the values from high to low efficiency, first four audio descriptors are considered giving highest accuracy. Further, the combinations of these audio descriptors with each other remaining audio descriptors are considered as feature vector for training and testing. This yielded to another record of best combinations of audio descriptors giving maximum singer identification accuracy. At this stage again, from highest to lowest audio descriptor combination efficiency is considered to move forward for next level according to the descending order of the efficiencies achieved for singer identification. This process is repeated till we reach to a situation where, the maximum efficiency of previous level is greater than the current level or we are exhausted with all the available audio descriptors under the title of timbre.

In the backward pass Pat B: we keep a record of the last successful combination of the audio descriptors that gave the highest accuracy. The last level combination gives best suitable combination of the audio descriptors giving best results of singer identification for 10 singers / 5 singers /3 singers. All these audio descriptors are responsible for identification of singer in DB1, which consists of all male singer recordings.

The entire procedure consisting of Part A and Part B of the proposed algorithm is repeated for another database DB2. For female singer dataset also in forward pass similar process is adopted generating 4 candidate audio descriptors at level 1 that are combined in level 2 and so on. As end result we have best combination of audio descriptors giving maximum accuracy in singer identification process for both, database DB1 for male and DB2 for female.

The indirect method of understanding and inferring the conclusion is this way. For all male singer identification some audio attributes are selected out of timbre group that were giving best singer identification accuracy. The accuracy is calculated for a dataset containing all male singers. While, there also exist another set of audio descriptors that gave maximum accuracy in identifying the singers who are female.

If we closely relate these two sets of results, giving sets of audio descriptors for Male and for Female, there are some common audio descriptors that are used for both male and female database that give highest accuracy and there are some uncommon audio descriptors that took part in the identification process where the genders were different. This proves that the uncommon audio descriptors are the ones which are needed for particular gender to correctly classify the singer. In short, to identify the gender of a singer these audio descriptors can be used. Following section proves this with respect to the results that are generated.

V. EXPERIMENTATION AND RESULTS

By applying the above stated procedure, experiments are carried out for databases, DB1 and DB2, containing audio samples of Male and Female North Indian Classical Vocal respectively. At the end of first experiment on DB1, we found singer identification accuracy to be 40.0463%, 63.75% and 83.3333% for 10, 5 and 3 male singers at a time, respectively. The audio descriptors that gave maximum efficiency were, *Zero crossing Rate (ZCR), Mel-Frequency Cepstral Coefficients (MFCC), Roughness* and *Irregularity.* Also for DB2, the singer identification accuracies were found out to be, 34.4907%, 50.4167% and 85 % for 10,5and 3 female singers respectively. The audio descriptors that were responsible to give these efficiencies were, *Zero crossing Rate (ZCR), Rolloff, Mel-Frequency Cepstral Coefficients (MFCC), and Brightness.* The performance of MFCC was noted for 10, 20 and 30 coefficients and found that there is no significant change in the accuracy of singer identification if we change the number of coefficients. Also, adding the other audio descriptors from the timbre group of the tool box increased the efficiency of singer identification. Table I represents the list of audio descriptors that are useful in singer identification process as stated above.

Table I: List of Audio Descriptors and their requirements for male and female singer identification. ZCR and MFCC are required for both the singers while others are required for either of them.

Audio Descriptor number	Name of Audio Descriptor	Required for Male Singer Identification	Required for Female Singer Identification
1	ZCR	YES	YES
2	ROLLOFF	NO	YES
3	MFCC	YES	YES
4	BRIGHTNESS	NO	YES
5	ROUGHNESS	YES	NO
6	IRREGULARITY	YES	NO

Note that, in Table I, for both, male and female singer identification there exists two audio descriptors that are required namely ZCR and MFCC. This is in line with the traditional systems of singer identification that have already proved that singer identification can be best possible through usage of MFCC as audio feature vector. We have added other audio descriptors to the feature vector along with MFCC with the aim of increasing the efficiency of singer identification. Zero crossing rate (ZCR) is the rate of change of sign of the signal. Low frequency sound usually has less number of sign changes as compared to high frequency sounds. But as stated above, it is not necessary that a female will always sing in high frequency range and male in low frequency range. Hence we cannot consider zero crossing rate as a parameter to differentiate between male and female singer.

Other than these two universally used audio descriptors for singer identification, we found some other audio descriptors that are uncommon in the results of DB1 and DB2. For DB1, Male singers, note that along with ZCR and MFCC, there are two more audio descriptors viz, Roughness and Irregularity.

Roughness is related to the beating phenomenon whenever pair of sinusoids is closed in frequency [4]. It is a sensory dissonance. Roughness is higher when for fixed pulse frequency, short duration events occur. It represents the rapid sequence of events that has shape of its own type and frequency of occurrence. This makes sense with the anatomical aspect of difference between a male and a female. Recall that, during changing years of puberty, the larynx of a boy becomes thick and causes the voice to sound rough and deep.

That means for male singer the roughness values will clearly distinguish the singer from one another since each male will have its own shape of vocal tract after teenage and will produce unique sound , that shall be noticeable than for a female. In case of North Indian Classical music recordings that contain continuous interference of a supportive musical instrument called 'Tanpura', the value of roughness shall clearly identify the singer and it shall be male.

Another audio descriptor that is used in male singer database is irregularity. It is the degree of variation of the successive peaks of the spectrum. Irregularity is calculated by the sum of the square of the difference in amplitudes between neighboring partials. It varies as per the singer. The successive peaks in the spectrum are important attributes from point of view of male singer identification.

Similar to male singer, for female singer identification there are two audio descriptors that are different than the common audio descriptors viz. Roll off and Brightness. Roll off is the estimation of the amount of high frequency in the input signal. Recall that, the audio recordings contain long notes sung by the classical singers. The amount of energy contained in the signal below a certain point (roll off point) is 85% of total signal energy. The female singer singing tonal notes without depth in the voice makes roll off a better tool to distinguish between different female singers.

Brightness is similar to roll off but has different interpretations than roll off. Brightness value is always a value between 0 and 1 and it indicates the amount of energy above a cut off frequency. The cut of frequency in our experiments has been set up of 1500Hz for a simple reason that all the singers are singing near 1 kHz frequency. The entire female singer database has demonstrated that the voice energy levels are brighter than that of male singer database. That means to identify a female singer 'Brightness' timbre attribute contributes to a large extent.

VI. CONCLUSION

A novel method to indentify a singer from north Indian classical music and then to back track and review the audio descriptors that have helped the process for both, male and female singers was proposed. The experiments have been carried out on two datasets DB1 containing all male singer recordings and DB2 containing all female singer recordings. MIRtoolBox provided the tools to extract the timbral audio attributes of the recordings and these attribute feature vectors were fed to K-means clustering. K-means clustering being the simplest method to classify the data provided codebooks which were tested later. The experiments were carried out on 10 singers, 5 singer and 3 singers simultaneously. The efficiencies found were as shown in following Table II, along with the combinations of audio descriptors that gave these results.

Table II. Experimental results showing the major contributory audio descriptors for singer identification from male singer database (DB1) and female singer database (DB2).ROGH is roughness and BRIG is brightness. The highest efficiency achieved for singer identification is 83.33% for male and 85% for female.

Sr	Data base	Audio descriptors used	Results of classification, % efficiency for singers #		
			# 10	# 5	# 3
1	DB1 (Male)	ZCR+MFCC+ROUG+IRRE	40.04	63.75	**83.33**
2	DB2 (Female)	ZCR+ROFF+MFCC+BRIG	34.49	50.41	**85.00**

From the analysis of the Table II, following conclusions are drawn.

A) Singer identification of North Indian Classical music is possible through simple feature extraction of Timbre taxonomy of MIRtoolbox and using K-means classifier.

B) The efficiency of the classifier decreases with increase in total number of singers to be trained and tested.

C) Though MFCC is best suitable method to identify a singer in literature, the efficiency of classification is enhanced by adding other timbral attributes of the sound in the audio feature vector.

D) ZCR and MFCC are useful in singer identification for both the genders.

E) For male singer, Roughness and Irregularity plays vital role in identification process.

F) For female singer identification, Roll off and Brightness are important audio descriptors.

G) Highest efficiencies of singer identification for male and female singers are 83.33% and 85% respectively.

REFERENCES

[1] Dominik Wegmann, "DAFX Toolbox ," MATLAB, MANUAL 06 Nov 2006 (Updated 17 Jul 2007).

[2] Kobi Nis, ""Audio Database Toolbox" ," MATLAB, MANUAL 21 Apr 2009 (Updated 25 Jun 2009).

[3] ""Data Acquisition Toolbox"-V 3.4," MATLAB, Natick, MA, Manual Sept 2013.

[4] Petri Toiviainen Olivier Lartillot, ""A Matlab Toolbox for Musical Feature Extraction From Audio"," in *International Conference on Digital Audio Effects*, Bordeaux, 2007.

[5] Kumar Rakesh, Subhangi Dutta, and Kumara Shama, "GENDER RECOGNITION USING SPEECH PROCESSING TECHNIQUES IN LABVIEW," *International Journal of Advances in Engineering & Technology*, vol. Vol. 1 , no. Issue 2 , pp. pp.51-63, May 2011.

[6] wikipedia. (2014, FEB) http://en.wikipedia.org. [Online]. http://en.wikipedia.org/wiki/Gharana

[7] School of Indian Music - Sangeetalay. (2008) http://www.schoolofindianmusic.com. [Online]. http://www.schoolofindianmusic.com/ivm.htm

[8] Tae Hong Park, ""Towards Automatic Musical Instrument Timbre Recognition"," PRINCETON UNIVERSITY, PRINCETON, Thesis Report Nov 2004.

[9] MPEG-7 Audio. ((2005, October)) http://mpeg.chiariglione.org/standards/mpeg-7/audio. [Online]. MPEG-7 Audio. [Online].

[10] Geoffroy Peeters, "A large set of audio features for sound description (Similarity and Classification) in the CUIDADO Project," Ircam,Anlysis/Synthesis Team, 1Pl Igor, Stravinsky, 75001, Paris, France, Analysis report V 1.0, 23rd April,2004.

[11] M., & Herrera 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.

[12] The Nemours Foundation : Steven Dowshen. (2012,October) http://kidshealth.org. [Online]. http://kidshealth.org/parent/general/body/changing_voice.html